http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

Working Paper 8507

FORECASTING AND SEASONAL ADJUSTMENT

by Michael L. Bagshaw

Thanks are due to Gordon Schlegel for programing support and 8111 Gavln and Klm Kowalewski for helpful comments.

Working papers of the Federal Reserve Bank of Cleveland are preliminary materials, circulated to stimulate discussion and critical comment. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System.

December 1985

Federal Reserve Bank of Cleveland

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

FORECASTING AND SEASONAL ADJUSTMENT

Key words: Seasonal adjustment, forecasting performance, multivariate time series models.

Abstract

There have been many studies and papers written about the effects of seasonal adjustment on the relationships among variables. However, there has been a dearth of studies about the effects of seasonal adjustment on the problem of forecasting. Since the development of time serles models often has forecasting as a major product, it is essential to study the effects of seasonal adjustment on forecasting in these models. In this paper, we present an application of multivariate time series forecasting applied to five economic time series, in which we compare forecasts developed from seasonally adjusted data with forecasts from seasonally not-adjusted data. The results of this exercise are mixed. For some forecasting situations, using not-seasonally adjusted data provides better forecasts for most of the variables In this study. However, in other instances, using seasonally adjusted data provides better forecasts for most of the variables in this study. The results appear to depend on the length of the forecast period. Also, it appears that the best solution in some instances might be to develop models for both seasonally adjusted data and not-seasonally adjusted data.

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

I. Introduction

The goal of this research is to compare forecasts from two models developed for an earlier study (see Bagshaw and Gavin [1983]) to obtain an indication of whether it is better to seasonally adjust data when developing multivariate time series models for forecasting. There have been many studies indicating that seasonally adjusting data will affect the relationships among the variables. Bell and Hillmer (1984) provide references for many of these studies. However, there has been little empirical evidence concerning the effects of seasonal adjustment on forecasting accuracy. The question of whether to use seasonally or not-seasonally adjusted data is especially important in time series analysis, because these models are often developed mainly, if not entirely, for forecasting purposes. Even if the seasonal adjustment procedure changes the relationships among variables, this will not matter for forecasting, if the new relationships provide as accurate, or even more accurate, forecasts than those developed from not-seasonally adjusted data. Makridakis and Hibon (1979) compared forecasts of seasonally and not-seasonally adjusted data using several popular univariate forecasting methods. Their conclusion was that using seasonally adjusted data provided somewhat better forecasts than using not-seasonally adjusted data. However, these results may have been influenced by their choice of constant seasonal factors in the development of models for the not-seasonally adjusted data (see Bell and Hillmer [1984]). Plosser (1979) forecasts five unadjusted economic time series with univariate seasonal autoregressive integrated moving average (ARIMA) models and the same series after seasonal adjustment with univariate nonseasonal ARIMA models. He found that the nonseasonal ARIMA models

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

performed substantially better on two series, slightly better on two series, and slightly worse on one series. Thus, the results on whether to seasonally adjust or not when developing models for forecasting are mixed and limited. In particular, they are limited to univariate models.

The present study adds to the information concerning the advisability of seasonal adjustment before forecasting by examining the forecast accuracy of five economic variables in a multivariate time series model. This is in contrast to the abovementioned papers, which deal only with univariate methods of forecasting. Because there is much evidence that seasonal adjustment affects the relationships among variables (see Bell and Hillmer [1984]), it is critical to test whether this effect carries over to forecast accuracy. If the seasonal adjustment is such that the relationships remain stable over time in the seasonally adjusted data, then seasonally adjusted data might provide better forecasts than not-seasonally adjusted data. However, if the seasonal adjustment process is not stable, then worse forecasts may be obtained using the seasonally adjusted data. This latter conclusion was reached by Plosser (1979) in the univariate case.

II. Multivariate ARMA Time Series Models

The following is a very brief description of multivariate ARMA time series models; Tiao and Box (1981) provide a more detailed description. The general multivariate ARMA model of order (p,q) is given by:

$$(1)\quad \underline{\Phi}_P(B^s)\underline{\phi}_p(B)\underline{Z}_t = \underline{\Theta}_Q(B^s)\underline{\theta}_q(B)\underline{a}_t + \underline{\theta}_0,$$

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

where

(2) $$\underline{\Phi}_P(B) = \underline{I} - \underline{\Phi}_1 B - \ldots - \underline{\Phi}_P B^P,$$

$$\underline{\phi}_p(B) = \underline{I} - \underline{\phi}_1 B - \ldots - \underline{\phi}_p B^p,$$

$$\underline{\Theta}_Q(B) = \underline{I} - \underline{\Theta}_1 B - \ldots - \underline{\Theta}_Q B_Q,$$

$$\underline{\theta}_q(B) = \underline{I} - \underline{\theta}_1 B - \ldots - \underline{\theta}_q B^q,$$

where

s = the length of the seasonal, for example, for quarterly data, $s=4$,

B = backshift operator (i.e., $B^s z_{i,t} = z_{i,t-s}$),

$\underline{I}$ = k x k Identity matrix,

$\underline{z}$ = vector of k variables in the model,

$\underline{\Phi}_j$'s, $\underline{\phi}_j$'s, $\underline{\Theta}_j$'s and $\underline{\theta}_j$'s = k x k matrixes of unknown parameters,

$\underline{\theta}_0$ = k x 1 vector of unknown parameters, and

$\underline{a}$ = k x 1 vector of random errors that are identically and independently distributed as $N(0,\Sigma)$.

Thus, it is assumed that the $a_{i,t}$'s at different points in time are independent, but not necessarily that the elements of $\underline{a}_t$ are independent at a given point in time.

The n-period-ahead forecasts from these models at time t ($\underline{z}_t(n)$) are given by:

(3) $$\underline{z}_t(n) = \underline{\phi}_1[\underline{z}_{t+n-1}] + \ldots + \underline{\phi}_p[\underline{z}_{t+n-p}] + [\underline{a}_{t+n}] - \underline{\theta}_1[\underline{a}_{t+n-1}] - \ldots - \underline{\theta}_q[\underline{a}_{t+n-q}],$$

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

where, for any value of t,n,m, $[\underline{x}_{t+n-m}]$ implies the conditional expected values of the random variables $\underline{x}_{t+n-m}$ at time t. If n-m is less than or equal to zero, then the conditional expected values are the actual values of the random variables and the error terms. If n-m is greater than zero, then the expected values are the best forecasts available for these random varfables and error terms at time t. Because the error terms are uncorrelated with present and past information, the best forecasts of the error terms for n-m greater than zero are their conditional means, which are zero. The forecasts can be generated iteratively with the one-period-ahead forecasts that depend only on known values of the variables and error terms. The longer-length forecasts, in turn, depend on the shorter-length forecasts. .

III. Development of Models For Forecasting

The Tiao-Box procedure was used to estimate multivariate time series models for the following five variables: the money supply (M1), credit is funds raised by the nonfinancial sector (NFD) including private and government debt, the quantity of goods is GNP in constant (1972) dollars (GNP72), the price of output is the implicit GNP deflator (PGNP), and the price of credit is the yield on three-month Treasury securities (RTB3).

Two models were estimated, one using seasonally adjusted data (except for RTB3, which is not-seasonally adjusted) and one with not-seasonally adjusted data (except for, PGNP which is not available not-seasonally adjusted). These models were estimated over the time period from the first quarter of 1959 through the fourth quarter of 1979. The results presented here may be slightly biased in favor of the seasonally adjusted model, because

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

the latest revised seasonal adjusted data was used in estimating these models. The seasonal adjustment procedure is a two-sided filter; therefore, some of the data being forecast in this study were used in developing seasonal adjustment factors for the data in the estimation period. To be completely comparable, we should really use the seasonally adjusted data that were available at the time of the forecast. In this way, the seasonal adjustment factors would not be modified by using data from the forecast period. However, as Young (1968) has indicated, the asymmetric filters used to adjust the ends of a series are chosen with the objective of minimizing the revision necessary after new data becomes available. The effects of using the revised seasonally data should thus be minimal. The model estimated uslng the not-seasonally adjusted data is given in table 1. The model estimated using seasonally adjusted data is given in table 2.

From the estimation results, we would expect that the seasonally adjusted model would forecast better than the not-seasonally adjusted model for four of the five variables (PGNP, M1, NFD, GNP72) because the within-sample estimated variances are smaller for the seasonally adjusted model than for the not-seasonally adjusted model. Thls dlfference ranges from 19 percent to 81 percent. For RTB3, which is not seasonally adjusted in either model, the within-sample variance is slightly smaller for the not-seasonally adjusted data.

IV. Forecasting Results

The two models were used to forecast the levels of the variables in three different situations: 1) one-quarter ahead, 2) one-year ahead, and 3) a

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

combination of one- through four-quarters ahead. For one-quarter-ahead forecasts, one-quarter ahead forecasts were generated for a given year. The resulting forecast errors were then averaged over the year. In this manner, both the seasonally and not-seasonally adjusted models were forecasting the same values because the seasonally adjusted data and the not-seasonally adjusted data must sum to the same value for a year. Similarly, the year-ahead forecasts were averaged over the year. That is, forecasts were generated from the first quarter of the previous year for the first quarter of the forecast year, from the second quarter for the second quarter, etc. These forecast were then averaged. In the combination forecasts, one-, two-, three-, and four-quarter-ahead forecasts were generated from the fourth quarter of the year prior to the forecast year and then the forecast errors were averaged for a given year. In order to have consistent forecast periods for the three types of forecasting, one-year-ahead forecasts were generated for 1980 starting in the first quarter of 1979. Thus, for four of the series (PGNP, M1, NFD, and RTB3) there were five years of forecast error data. For GNP72, the not-seasonally adjusted data for 1984 were not available at the time of the study. To be consistent, the results for GNP72 for both models is reported only for 1980 through 1983. Thus, there are four years of data for GNP72 forecast errors. Consequently, there are either five or four observations in the analysis presented in this paper.

The mean error, mean absolute error, and the root mean square error (RMSE) for the three forecast horizons and the two models are presented in tables 3 through 5. The following discussion is based on the analysis of the RMSE from these forecasts.

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

Examining the one-quarter-ahead forecasts (Presented in table 3), we see that the not-seasonally adjusted model forecasts better for three of the series (PGNP, RTB3 and GNP72), and the seasonally adjusted model forecasts better for the other two series (M1 and NFD). The differences in the RMSE are very substantial for several of these series. The ratios of the not-seasonally adjusted models RMSE to the seasonally adjusted models RMSE are 0.60 for PGNP, 1.16 for M1, 1.32 for NFD, 0.65 for RTB3, and 0.58 for GNP72. Given that the within-sample standard deviation ratios were 1.09, 1.22, 1.19, 0.98, and 1.35 (in terms of logarithms of PGNP, M1, NFD, RTB3, and GNP72, respectively), this result is somewhat unexpected. The seasonally adjusted model provides a better within-sample fit for four of the five series. The fifth series is essentially tied, while it provides better forecast for only two series. This appears to imply that the relationship among seasonally adjusted data may not be as stable as that among not-seasonally adjusted data.

When we examine the year-ahead forecasts (presented in table 4), we obtain different results. Here, the seasonally adjusted model forecasts four of the series (PGNP, M1, NFD, and GNP72) better than the not-seasonally adjusted model. However, three of these four have essentially the same RMSEs for the two models. The ratios of the corresponding RMSEs are 1.30, 1.01, 1.01, 0.59, 1.02 for PGNP, M1, NFD, RTB3, and GNP72, respectively. Thus, "on average", these two models perform roughly the same for the five series considered as a group when forecasting one year ahead. This may be related to the fact that we are forecastfng here one season ahead. Thus, the seasonally adjusted model may have a built-in advantage for this forecast length.

E amining the combined one- to four-quarters-ahead forecasts (presented in table 5), we again arrive at a different result. Here, the not-seasonally

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

adjusted model forecast four of the five series better than the seasonally adjusted model. The corresponding RMSE ratios are 0.83, 0.81, 1.01, 0.81, and 0.80, for PGNP, M1, NFD, RTB3, and GNP72, respectively. The only series for which the seasonally adjusted model had a smaller RMSE than the not-seasonally adjusted model for this combination forecast was NFD, a series constructed such that (for the technique used in this paper of averaging forecast over a year), the combination forecast result is the same as the one-year-ahead forecast result. Thus, this result may agaln be attributed to the seasonal model's advantage in forecasting one season ahead.

V. Summary

In this study, we have examined whether one should seasonally adjust data before developing multivariate time series models to provide forecasts. The results are mixed; that is, performance of each model seemed to depend on the length of the forecast. For one-period-ahead forecasts, the evidence of this study suggests that perhaps it would be best to develop models for both seasonallly adjusted and not-seasonally adjusted data. The forecasts from these models would then be evaluated to determine which series are better forecast using the seasonally adjusted model, and which using the not-seasonally adjusted model. The within-sample fit is not a good deciding factor in this choice. since the within-sample fits indicated that the seasonally adjusted model provided a better fit for four of the five series (with a virtual tie for the fifth), while forecasts indicate that the not-seasonally adjusted model did better for three of the five series.

If one wishes to forecast for more than one period ahead, then the results are mixed. If one wants to forecast one year ahead, then the results appear ’ ’ cate that one should use seasonally adjusted models. The only series for which the not-seasonally adjusted model provided a better forecast for one year ahead was RTB3, which is not-seasonally adjusted. Relationships among variables change more drastically if some series are seasonally adjusted and others are not, than if all series are treated equally, which could explain this result. For the case where it is desirable to forecast a combination of lengths ahead, the results appear to indicate that the not-seasonally adjusted data are the best choice, because the not-seasonally adjusted model forecast four of the five series better. The fifth was a special case, which naturally favored using seasonally adjusted data.

Because of the small out-of-sample forecast period used here, and the small number of series studied. there is obviously no way that the results presented here can be conclusive. Thus, more study of this very important area is called for.

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

able 1 Estimated Model Using Not-Seasonally Adjusted Data: 1959:IQ - 1979:IVQ

$$\begin{bmatrix} 1-.978B & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -.210B & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & -.278B & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \nabla \ln PGNP_t \\ \nabla\nabla_4 \ln MINS_t \\ \nabla\nabla_4 \ln NFDNS_t \\ \nabla \ln RTB3_t \\ \nabla\nabla_4 \ln GNP72NS_t \end{bmatrix} = \begin{bmatrix} 1-.441B & 0 & 0 & & 0 \\ 0 & (1+.504B)(1-.558B^4) & 0 & 012B & 0 \\ 0 & 0 & (1+.359B)(1-.459B^4) & 0 & 0 \\ 0 & 3.913(1-.558B^4) & 0 & 1+.617B & 0 \\ -.408B & 0 & 0 & 0 & 1-.823B^4 \end{bmatrix} a_t$$

$$\underset{\sim}{\theta}_0 = 0, \quad \underset{\sim}{\hat{\rho}}_a = \begin{bmatrix} .138\times10^{-4} & & & & \\ .146 & 367\times10^{-4} & & & \\ .194 & .492 & .116\times10^{-4} & & \\ -.023 & .127 & .183 & .895\times10^{-2} & \\ -.085 & .203 & .362 & .141 & .131\times10^{-3} \end{bmatrix}$$

Note: Diagonal terms in $\underset{\sim}{\hat{\rho}}_a$ are the variances of the a's.

∇ represents the first difference of the series: $\nabla Z_t = Z_t - Z_{t-1}$

∇_4 represents the annual first difference of the series: $\nabla_4 Z_t = Z_t - Z_{t-4}$

Table 2 Estimated Model Using Seasonally Adjusted Dat   1959:IQ - 1979:IVQ

$$\begin{bmatrix} 1-.851B & -.220B & 0 & 0 & 0 \\ 0 & 1-.433B & -.244B & 0 & 0 \\ 0 & 0 & 1-.850B & 0 & 0 \\ 0 & -4.366B & 0 & 1 & 0 \\ -.354B & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \nabla \ln PGNP_t \\ \nabla \ln M1_t \\ \nabla \ln NFD_t \\ \nabla \ln RTB3_t \\ \nabla \ln GNP72_t \end{bmatrix} = \begin{bmatrix} 1-.470B & 0 & -.254B & 0 & 0 \\ 0 & 1 & 0 & -.01[illegible] & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1+.[illegible] & 0 \\ -.347B & 0 & 0 & 0 & 1+.183B \end{bmatrix} a_t + \begin{bmatrix} -.0009 \\ .0018 \\ .0032 \\ -.0358 \\ .0126 \end{bmatrix}$$

$$\hat{\underset{\sim}{\Omega}}^{a} = \begin{bmatrix} .116\times10^{-4} & & & & \\ .151 & .247\times10^{-4} & & & \\ .006 & .558 & .817\times10^{-5} & & \\ -.175 & -.033 & .183 & .931\times10^{-2} & \\ -.176 & .253 & .362 & .141 & .723\times10^{-4} \end{bmatrix}$$

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

Table 3 Out-of-Sample Forecasts: One-Quarter-Ahead Forecast Errors

Series	Mean error	Mean absolute error	RMSE
PGNP			
Seasonally adjusted	-.0052	.0080	.0092
Not-seasonally adjusted	-.0011	.0041	.0055
M1			
Seasonally adjusted	.7189	1.7102	1.8264
Not-seasonally adjusted	.5603	1.8078	2.1150
NFD			
Seasonally adjusted	20.3930	20.5370	31.0380
Not-seasonally adjusted	30.9550	30.9550	41.1150
RTB3			
Seasonally adjusted	-.5276	.5276	.6522
Not-seasonally adjusted	-.2558	.2821	.4258
GNP72			
Seasonally adjusted	-17.0460	17.0460	20.3460
Not-seasonally adjusted	8.7582	9.1400	11.7290

*RMSE is the root mean square error of the forecast.

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

Table 4 Out-of-Sample Forecasts: Year-Ahead Forecast Errors

Series	Mean error	Mean absolute error	RMSE
PGNP			
Seasonally adjusted	.0156	.0204	.0282
Not-seasonally adjusted	.0271	.0320	.0368
M1			
Seasonally adjusted	11.2550	11.4440	16.2230
Not-seasonally adjusted	10.1300	10.8250	16.4240
NFD			
Seasonally adjusted	169.3800	169.3800	205.6400
Not-seasonally adjusted	150.5500	150.5500	207.3700
RTB3			
Seasonally adjusted	-.7289	1.5213	2.0494
Not-seasonally adjusted	.6652	1.0853	1.2092
GNP72			
Seasonally adjusted	10.9460	48.5080	51.9720
Not-seasonally adjusted	-4.4079	47.9280	53.0900

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

Table 5 **Out-of-Sample Forecasts: Combined One- to Four-Quarters Forecast Errors**

Series	Mean error	Mean absolute error	RMSE
PGNP			
Seasonally adjusted	.0071	.0426	.0489
Not-seasonally adjusted	.0190	.0362	.0404
M1			
Seasonally adjusted	15.6510	15.6510	18.9070
Not-seasonally adjusted	11.2780	11.2780	15.3530
NFD			
Seasonally adjusted	169.3800	169.3800	205.6400
Not-seasonally adjusted	150.5500	150.5500	207.3700
RTB3			
Seasonally adjusted	-1.5847	2.4767	3.1615
Not-seasonally adjusted	-.1101	2.4485	2.5517
GNP72			
Seasonally adjusted	31.4150	49.0840	64.7170
Not-seasonally adjusted	-1.5364	48.6520	51.4900

http://clevelandfed.org/research/workpaper/index.cfm
Best available copy

References

Bagshaw, Michael L., and William T. Gavin. "Forecasting the Money Supply in Time Series Models," Working Paper 8304, Federal Reserve Bank of Cleveland, December 1983.

Bell, William R., and Steven C. Hillmer. "Issues Involved with the Seasonal Adjustment of Economic Time Series," Journal of Business & Economic Statistics, vol. 2, no. 4 (October 1984), pp. 291-320.

Box, George E.P., and Gwilym M. Jenklns. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1976.

Makridakis, Spyros, and Michele Hibon. "Accuracy of Forecasting: An Empirical Investigation," Journal of the Royal Statistical Society, Series A (General), vol. 142, part 2 (1979), pp. 97-125.

Plosser, Charles I. "Short-term Forecasting and Seasonal Adjustment," Journal of the American Statisticial Association, vol. 74, no. 365 (March 1979), pp. 15-24.

Tiao, G.C., and G.E.P. Box. "Modeling Multiple Time Series with Applications," Journal of the American Statistical Association, vol. 76, no. 376 (December 1981), pp. 802-16.

Young, Allan H. "Linear Approximations to the Census and BLS Seasonal Adjustment Methods," Journal of the American Statisticial Association, vol. 63, no. 322 (June 1968), pp. 445-71.