Estimating Seasonality In Time-Series Forecasting Models

Richard G. Fritz
and
James Xander*

INTRODUCTION

The study of seasonality in time series models has resulted in a large volume of published research in this century. The analysis has generally been directed toward either providing more information to facilitate forecasting or the efficient removal of seasonality in order to better analyze the remaining component forces of the historical time series. Our focus is on the improvement of forecast accuracy by efficient evaluation of the seasonal component. This paper reviews the traditional methods and suggests an alternative means for treating seasonality in forecasting models when the pattern of seasonality is varying over time.

Changing seasonal patterns over time are especially important when evaluating regional data where rapid development and transformation occur in the underlying structure of the economic system. In this case abrupt changes in seasonality are rather common when compared with the smooth evolutionary process that is often modeled for aggregate economic data. Changing seasonal patterns impose special difficulties when attempting to separate seasonal and non-seasonal movements. Accordingly, the development and enhancement of techniques to make such distinctions constitutes an important development in time series analysis. Seasonal adjustments of data would be less cumbersome if the seasonal patterns did not shift over time. However seasonal patterns often do change and it is these changes that prove particularly troublesome in forecasting models.

Florida is a regional economy in which significant change has and is continuing to take place, making it very difficult for forecasters and policymakers. This is particularly true for the Florida tourist industry. Seasonal factors, computed by the Census X-11 procedure are presented in Figures I and II for visitor arrivals by air and automobile respectively. Each figure displays the seasonal factors for two years, 1960 and 1981. For automobile

*Department of Economics, University of Central Florida

visitor arrivals the changes in seasonality are readily apparent. The automobile series has become much less seasonal in recent years. However, the air series has retained its highly seasonal nature in the period from 1960 to 1981; although there has been some changes in the seasonal pattern.

Using the Florida tourism data as our sample, we report forecasts based on several competing time series procedures that incorporate alternative seasonal adjustment methods. The results indicate that increased accuracy is gained when adjustments are made for the evolutionary nature of seasonality. The methods presented here are consistent with the Nerlove, Grether and Carvalho (1979) description of optimal seasonal adjustments.

RECENT DEVELOPMENTS IN SEASONAL ADJUSTMENTS

Decomposition methods are one of the oldest forecasting approaches. These methods were first employed in the beginning of this century by economists attempting to identify the business cycle. (Makridakis and Wheelwright, 89) The Census II X-11 program has become the traditional means for seasonally adjusting economic time series. The X-11 routine is based on the common ratio-to-moving average technique. However, X-11 utilizes moving averages of lengths other than the data periodicity as well as orders greater than one. This technique was developed in the 1920's at the National Bureau of Economic Research.

Figure I
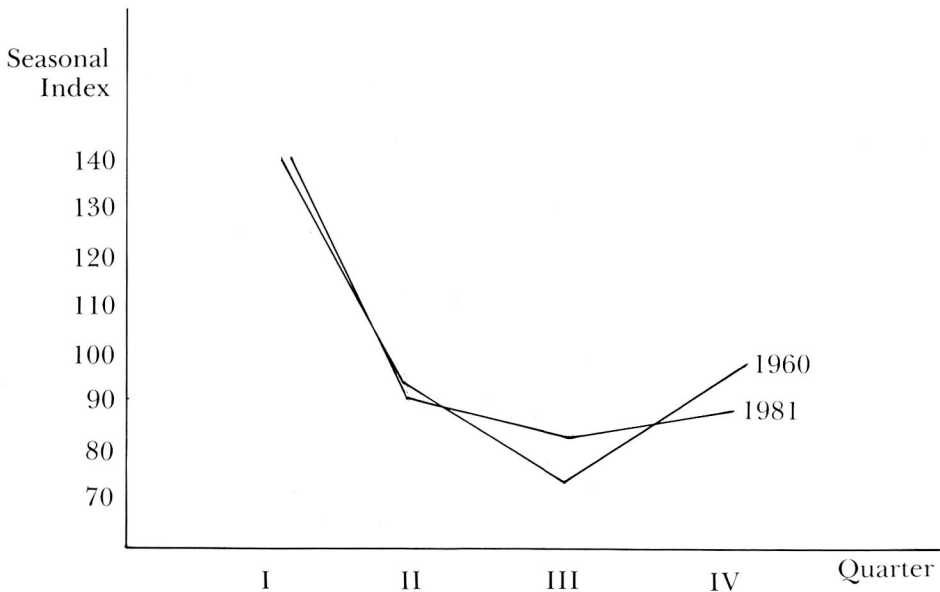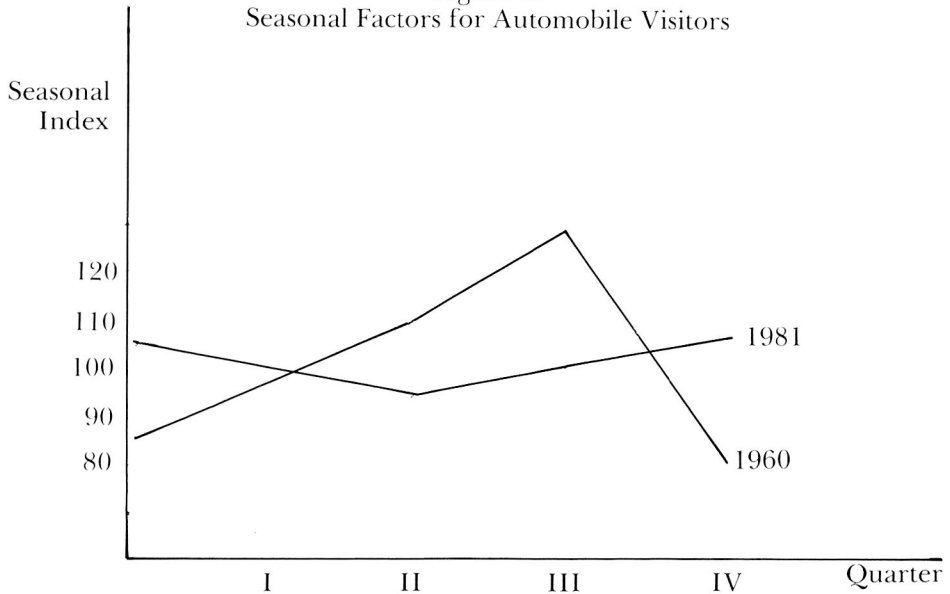Seasonal Factors for Air Visitors

Figure II
Seasonal Factors for Automobile Visitors

The decomposition methods assume that an observed time series, Y_t, can be decomposed into the form (e.g. multiplicative);

$$Y_t = X_t \cdot s_t \cdot e_t, \tag{1}$$

where X_t is the trend cycle, s_t is the seasonal or random component and e_t is the random or irregular component at time period t. The trend or trend-cycle is measured in the original units and the seasonal and irregular components are measured relative to unity reflecting the extent of seasonal or irregular variation.

The X-11 method first accounts for the trend so that the series is stationary. To estimate the remaining changing seasonal pattern, weighted moving averages are computed by seasonal period (month or quarter) to reflect the seasonal factors. Since weighted moving averages are also used to eliminate the trend in the original series, the seasonal factors, s'_t, can be estimated by a single composite moving average of the original series. Thus, the seasonal factors are

$$s'_t = \sum_{k=-m\ldots}^{1} w_k \, Y_{t-k} \tag{2}$$

where w_k are the weights, which extend l periods forward and k periods backward from time period t. The weights are symmetric and X-11 provides numerous options for the weight depending on the stability of the seasonal pattern.

The Census X-11 model provides projections into the future

$$s_{n+1} = s_n + \tfrac{1}{2}[s_n - s_{n-1}] \tag{3}$$

where s_n is the last observation of the seasonal index and n ranges to four or twelve depending on the segmentation of the seasonal observations. These estimates have proven to be troublesome for researchers using both aggregate data and regional data. At the national level, if seasonal factors for monetary aggregates are estimated "poorly," revisions of reported statistics may be large and thus a source of embarrassment for the Federal Reserve (Pierce and Cleveland). Thus various efforts are under way to modify the estimation of seasonal factors building upon the X-11 base and generating new seasonal factor forecasts by alternative means.

MODELING ALTERNATIVES

As an alternative to the X-11 method, seasonal models can be either naive or causal. The major competitors are ARIMA models (autoregressive integrated moving-average models) and regression models.

ARIMA MODELS

ARIMA models are naive in the sense that they attempt to explain a given series in terms of its past values. The general description of an ARIMA model is

$$Y_t = \emptyset_{1Y_{t-1}} + \emptyset_{2Y_{t-2}} + \ldots + \emptyset_{pY_{t-p}} + e_t +$$
$$\theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}. \tag{4}$$

The lags, p, for the autoregressive part of the model and q for the moving average error part of the model are determined through the identification process. Deseasonalized data, using the X-11 routine for instance, can be input directly into an ARIMA model as defined above in equation (4). This procedure, however, is not recommended by all since it may substantially distort the basic structure of the stochastic process. (Judge, et al, 700). Thus, it is generally recommended that a seasonal ARIMA model be estimated by differencing so as to reduce the autocorrelations at the seasonal intervals.

REGRESSION MODELS

Regression models have the advantage of being causal in the sense that changes in the dependent variable are statistically "explained" by a set of independent variables. However, if the regression model is to employ or measure the seasonal component in the time series, it is typically done so through the use of seasonal dummies with very little information concerning the underlying causes of the seasonal variation. However, it is pos-

sible to integrate the measurement of seasonal variation into a causal model despite the fact that no causal relationships are made explicit for seasonlity.

A relatively recent procedure that integrates the X-11 program and the ARIMA models has been developed and is being tested at the Federal Reserve. (Dagum) The seasonal adjustment routine performs the following:

1) The construction of an ARIMA model for the original series.
2) The extrapolation of one or two years of seasonal data at the beginning as well as the end of the original series utilizing the ARIMA model identified in 1).
3) The seasonal adjustment of the "extended" series constructed by various linear filters equivalent to those of the X-11 program.

The X-11 routine projects seasonal factors based upon asymmetric moving averages. The Fed's interest in X-11-ARIMA is centered around whether it can more accurately measure current as well as future seasonal factors to reduce revisions. The evidence does suggest that the X-11-ARIMA procedure does reduce the revisions significantly as compared to the regular X-11 method.

MIXED MODELS

Most previous attempts to adjust time series for seasonality have either explicitly or implicitly assumed the series' components may be evaluated with either a deterministic model or as a stochastic process. The small collection of exceptions include papers by Pierce and Porter (1973), Pierce (1978), and a recent addition to this list, Havenner and Swamy (1981). The latter develop separate estimates of four components: deterministic trend, deterministic seasonal, stochastic trend and stochastic seasonal. Their method admits interactions between the stochastic trend and the stochastic seasonal components, simultaneously estimating trend and seasonality.

This paper offers an alternative to the complicated simultaneous estimation procedure by developing a step-wise method of integrating the deterministic and stochastic models. The deterministic trend and deterministic seasonal components are first filtered out using the traditional Census X-11 multiplication method. If the resulting time-series of seasonal indexes are time invariant, then the model remains in this deterministic version, identified by some order of moving average. However, if the resulting X-11 seasonal indexes compose a stochastic process, then an ARIMA model is fitted to the process as the second step in identifying future values of the seasonal components.

Our method differs from the Dagum/Federal Reserve X-11-ARIMA method by reversing the step-wise procedures. In principle simultaneous methods should be superior to step-wise procedures. However optimal estimation always requires a trade-off between precision and cost. The classic solution to specifying the costs or loss functions when no precise knowledge of how the prediction is going to be used has been to minimize the expected value of the squared error between the predicted and the true value. Alternative forecasts are compared using the criterion of minimum mean-square error.

THE MODEL

Specifically we test the relative power of the mixed deterministic-stochastic model.

$$Y_t^* = FA_t = \frac{S_t \cdot T_t \cdot C_t \cdot E_t}{S_t} \qquad (5)$$

(Deterministic)

Here is the classic Census X-11 decomposition of the original time-series Y_t, where Y_t^* is the final series (FA_t) adjusted for seasonlity. The deterministic method of identifying S_t, the seasonal indexes, is the ratio to moving average technique, with adjustments for replacing extreme values. (Makridakis and Wheelwright, 1978, 106-139) This first step yields a time series of estimated seasonal indices, S_t. Thus, the next step is to analyze this time series with an ARIMA model.

$$\hat{s}_t = \emptyset_1 S_{t-1} + \emptyset_2 S_{t-2} + \ldots + \emptyset_p S_{t-p} + U_t +$$

$$\theta_1 U_{t-1} + \theta_2 U_{t-2} + \ldots + \theta_q U_{t-q} \qquad (6)$$

(Stochastic)

Equation (6) represents the general stochastic ARIMA model used to identify the historical pattern of the changes in the deterministic seasonal indices, S_t found in equation (5). The forecast of future seasonal changes, $\hat{s}_t$, may be made following the Box-Jenkins method described in equation (4). This forecast of seasonal pattern must be linked with a forecast of the trend-cycle component, Y_t^*.

$$\hat{Y}_t = \emptyset_1 Y_{t-1}^* + \emptyset_2 Y_{t-2}^* + \ldots + \emptyset Y_{t-p}^* + E_t +$$

$$\theta_1 E_{t-1} + \theta_2 E_{t-2} + \ldots + \theta_q E_{t-q} \qquad (7)$$

Future values of the trend-cycle component, $\hat{y}_t$, can be made by an ARIMA model suggested in equation (7). However, with the deterministic values of y_t^*, it may be preferable to propose an econometric model for $\hat{y}_t$. At this point in the mixed model, any user developed forecasting tool may be employed. We choose to analyze our sample tourism data using the stochastic time-series model in equation (7). The accuracy of the forecast may indeed be improved by adopting a transfer function technique (Courchesne, et al., 1981) or a combined model using optimal weights. (Theil; Bates and Granger; Brandon, Fritz and Xander) The final step is simply to combine steps one, two and three:

$$\hat{\hat{y}}_t = \hat{s}_t \cdot \hat{y}_t \cdot \mu_t \qquad (8)$$

where $\hat{\hat{y}}_t$ is the forecast value of the original time series using deterministic and stochastic analysis of seasonality and μ is a remaining random term. Our method contrasts a similar attempt to consider a dynamic (stochastic)

model for the seasonal effects by the procedure Vera and Guerrero (1981) used as a filter to extract white noise from the seasonal series. They chose to apply a power transformation to stabilize the variance of the original series, prior to estimating the deterministic seasonal indices. (Vera and Guerrero, pp. 609-613). Our results are consistent with the Vera and Guerrero findings.

COMPARISON OF THE ACCURACY OF THE FORECASTS

To examine the performance of the step-wise procedure, the average number of hotel/motel rooms per night in Central Florida were forecasted using traditional techniques and the results were compared to the step-wise technique. The data were collected and reported monthly by the accounting firm of Laventhal and Horwath and adjusted for seasonality by the authors with the X-11 program.

Two traditional techniques for forecasting seasonal time series were employed. First, deseasonalized data (via X-11) were used as input into a non-seasonal (stochastic) Box-Jenkins model. After identification of the appropriate form of the model, parameter estimates were made for the period from January 1972 to December 1979. Based upon the estimated model, forecasts were made for a test period defined as the subsequent twelve months of 1980. The seasonally adjusted forecasts for 1980 were re-seasonalized via X-11 seasonal factors so that a forecast error could be computed relative to the seasonal actual data. A second traditional method employed was to identify and then estimate a seasonal model through 1979 and again forecast hotel/motel occupancies for the 1980 test period. Again the actual and forecasted results were examined to determine forecasting error to compare with other alternative methods.

The step-wise procedure was initiated by again inputting seasonally adjusted data into a non-seasonal Box-Jenkins model for the period from 1972 through 1979. The estimated model was then used to generate seasonally adjusted forecasts for the same test period. However, rather than using X-11 forecasts of seasonal factors, the seasonal factors determined by X-11 for the period from 1972 to 1979 were utilized as data and a Box-Jenkins seasonal model was estimated for the factors. The seasonal factors were forecast for the test period and then utilized to re-seasonalize the previously de-seasonalized Box-Jenkins forecasts.

Alternative I—Non-Seasonal Box-Jenkins
Re-Seasonalized with X-11 Seasonal Factors

Given seasonally adjusted Central Florida hotel/motel occupancies, a non-seasonal Box-Jenkins model was identified with first order autoregressive and second order moving average terms. To ensure stationarity first differencing was required. Thus, the model identified and estimated was (112). The results of this forecast strategy are developed and presented in Table I.

TABLE I.
Alternative I Results for Test Period

Date	Actual Seasonal Data	X-11 Seasonal Factors	Forecast B-J Non-Seasonal*	Forecast Error
1980 Jan	29024	.884	29208	− 184
Feb	38848	1.230	40270	− 1422
Mar	42420	1.346	42715	− 295
Apr	37509	1.192	37204	305
May	27238	.889	28233	− 995
Jun	31704	1.046	32164	− 460
Jul	36169	1.231	37715	− 1546
Aug	33936	1.141	33889	47
Sep	23666	.680	20520	3146
Oct	23000	.740	25470	− 2470
Nov	22773	.756	22855	− 82
Dec	24113	.777	24066	47

MSE** = 1349

* Forecasts re-seasonalized with X-11 derived seasonal factors.

**MSE = mean squared error

Alternative II—Seasonal Box-Jenkins

The seasonal hotel/motel occupancy data was used to identify and estimate a seasonal Box-Jenkins model. In the identification stage second differencing was required to generate stationarity as well as first order seasonal differencing. The full ARIMA model contained third order autoregression and a second order moving average error term. The seasonal component of the model only contained a first order moving average term. The forecasts and errors from this technique are displayed in Table II. Note that no seasonal index is employed since the model directly accounts for seasonality in the estimation and forecasting procedures.

Alternative III—Non-Seasonal Box-Jenkins
Re-Seasonalized with Box-Jenkins Seasonal Factors

Seasonally adjusted Central Florida hotel/motel occupancies are forecast with a non-seasonal Box-Jenkins model identical to that used in Alternative I. However, the re-seasonalization of the data was accomplished through seasonal factors generated as forecasts from a Box-Jenkins model developed for the time series of seasonal factors through 1979. A seasonal Box-Jenkins model was identified and estimated for the seasonal factors and included first differencing as well as second differencing at the sea-

TABLE II.
Alternative II Forecasts for Test Period

Date	Actual Seasonal Data	Forecast B-J Seasonal	Forecast Error
1980 Jan	29024	28655	369
Feb	38848	38212	636
Mar	42420	41210	1210
Apr	37509	37692	− 183
May	27238	29460	− 2222
Jun	31704	31350	354
Jul	36169	38624	− 2455
Aug	33936	34158	− 222
Sep	23666	20625	3041
Oct	23000	25612	− 2612
Nov	22773	22485	288
Dec	24113	23893	200
			MSE = 1564.7

sonal frequency of 12. The non-seasonal and seasonal components contained first order autoregressive but no moving average terms. The forecast results for Alternative III are displayed in Table III.

TABLE III
Alternative III Forecasts for Test Period

Date	Actual Seasonal Data	B-J Seasonal Factors	Forecast B-J Non-Seasonal*	Forecast Error
1980 Jan	29024	.884	28849	175
Feb	38848	1.230	39785	− 937
Mar	42420	1.346	42151	269
Apr	37509	1.192	37393	116
May	27238	.889	28170	− 932
Jun	31704	1.046	32011	− 307
Jul	36169	1.231	37532	− 1363
Aug	33936	1.144	33624	312
Sep	23666	.680	20282	3384
Oct	23000	.740	25232	− 2232
Nov	22773	.756	22705	68
Dec	24113	.777	23851	262
				MSE** = 1304

* Forecasts re-seasonalized with seasonal factors predicted via a Box-Jenkins model.

**MSE = mean squared error

SUMMARY AND CONCLUSIONS

Forecasting is still an art and likely will always be so. However, scientific advancements continue in the field of forecasting and some of the more recent and promising work appears to be in the area of forecasting seasonal variation. Deterministic decomposition of the economic time series has become a reliable and often employed method of adjusting seasonal data. However, when the very structure of the economy undergoes change that yields a stochastic pattern in the seasonal component of the time series, deterministic decomposition reduces forecast efficiency. The recent work at the Federal Reserve has rather convincingly shown that some new procedures for handling this seasonal component are extremely powerful and promise to increase forecast accuracy. The contribution of Havenner and Swamy separating deterministic and stochastic components not only for the trend-cycle but also for the seasonal component is possibly the most versatile and advanced of all the new forecast methods.

This paper offers an alternative means for handling seasonality in time series forecasting models. It is similar to the X-11-ARIMA Method developed by Dagum in that it combines the deterministic X-11 procedure with a stochastic ARIMA process. However, it reverses the order in which they are combined.

Using the tourism industry of the State of Florida, these findings indicate that increased efficiency in forecasting hotel occupancies can be achieved by using the step-wise seasonal method. When compared with two other traditional methods of time series forecasting, the step-wise technique had minimum mean square error. Given that the step-wise procedure is easily understood, inexpensive to use, and extremely flexible—additional research into its accuracy appears warranted.

REFERENCES

Bates, J. and Granger, C. W., 1969, "The Combination of Forecasting," *Operations Research Quarterly*, Vol. 20, pp. 451-468.

Brandon, C., Fritz, R. and Xander, J., "Econometric Forecasts: Evaluation and Revision," *Applied Economics*, forthcoming.

Courchesne, C., Fontenay, A. and Pourier, J., 1981, "An Empirical Study of Seasonality in Econometric Modeling," in O. Anderson and M. Perrymore, eds., *Time Series Analysis*, North-Holland, Amsterdam, pp. 109-131.

Dagum, Estela Bee, 1978, "Modeling, Forecasting & Seasonality: Adjusting Economic Time Series with the X-11-ARIMA Method," *The Statistician*, Vol. 27, pp. 203-216.

Havenner, Arthur & Swamy, P.A.V.B., 1981, "A Random Coefficient Approach to Seasonal Adjustment of Economic Time Series," *Journal of Econometrics*, Vol. 15, pp. 177-210.

Judge, G., Hill, R. C., Griffiths, W. E., Lutkepohl, H., Lee, T., 1982, *Introduction to the Theory and Practice of Econometrics*, Wiley, New York.

Makridakis, Spyros & Wheelwright, Steven C., 1978, *Forecasting Methods and Applications*, Wiley, New York.

Nerlove, M., Grether, D., and Carvalho, J., 1979, *Analysis of Economic Time Series*, Academic Press, New York.

Pierce, D. A., 1978, "Seasonal Adjustment When Both Deterministic and Stochastic Seasonality are Present," in: A. Zellner, ed., *Seasonal Analysis of Economic Time Series* (U.S. Department of Commerce, Bureau of the Census, Washington, D.C.), pp. 242-269.

Pierce, D. A. and Cleveland, William P., 1981, "Seasonal Adjustment Methods for the Monetary Aggregates," *Federal Reserve Bulletin*, Vol. 67, pp. 875-887.

Pierce, D. A., Porter, R. D., 1973, "Linear Models and Linear Filters in the Analysis of Seasonal Time Series," in: *Proceeding of the Business and Economic Statistics Section* (American Statistical Association), pp. 517-522.

Stephenson, James A. and Farr, Helen T., 1972, "Seasonal Adjustment of Economic Data by Application of the General Linear Statistical Model," *Journal of the American Statistical Association*, Vol. 67, pp. 37-45.

Theil, H., 1971, *Applied Economic Forecasting*, North-Holland, Amsterdam.

U.S. Department of Commerce, Bureau of Census, 1967, "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Technical Paper No. 15, U.S. Government Printing Office, Washington, D.C.

Vera, G. and Guerrero, V., 1981, "Estimation of Seasonal Factors Using Both Traditional Methods and Box-Jenkins Techniques," in: O. Anderson and M. Perryman, eds. *Time Series Analysis*, North-Holland, Amsterdam, pp. 607-616.