ISSN 1032-3813
ISBN 0 7326 1033 8

MONASH UNIVERSITY

AUSTRALIA

EXPONENTIAL SMOOTHING OF SEASONAL DATA: A COMPARISON

Roland G. Shami and Ralph D. Snyder

Working Paper 10/97
November 1997

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

EXPONENTIAL SMOOTHING OF SEASONAL DATA: A COMPARISON

Roland G. Shami* and Ralph D. Snyder*

ABSTRACT

A parsimonious method of exponential smoothing is introduced for time series generated from a combination of local trends and local seasonal effects. It is compared with the additive version of the Holt-Winters method of forecasting on a standard collection of real time series.

Keywords : Time series - Forecasting - Exponential smoothing - Holt-Winters method - Structural models - M-competition.

*Department of Econometrics & Business Statistics
Faculty of Business and Economics - Monash University
Clayton campus - Clayton - Vic 3168

1. Introduction

The appropriate choice of values for the smoothing parameters in exponential smoothing methods can be a vexed question (Gardner, 1985). A traditional approach for forecasting seasonal time series, the additive Holt-Winters method (HW), relies on three smoothing parameters. Determining optimal values for such parameters can be hampered in part by

a) the fact that fitting criteria, such as the likelihood function, are highly nonlinear functions of the parameters and, as such, numerical optimisation methods are required;

b) the likelihood functions need not be concave so that numerical optimisation routines cannot guarantee global optima without a full, three dimensional grid search.

It is particularly attractive, to simplify the associated optimisation process, to explore the possibility of seasonal approaches that make use of fewer smoothing parameters. In this paper we propose a version of seasonal exponential smoothing, called the parsimonious method, that requires only two parameters. Both methods are compared in a study based on the collection of series from Makridakis et al (1982) forecasting competition.

2. State Space Models

Linear exponential smoothing methods may be based on the innovations form of the state space model (Snyder, 1985)

$$\begin{cases} y_t = x'_t\, b_{t-1} + e_t \\ b_t = T\, b_{t-1} + \alpha e_t \end{cases} \qquad (1)$$

where y_t is the series value, x_t is a fixed k-vector, b_t is a random k-vector of time dependent state variables, the e_t's are independent normally distributed disturbances with mean 0 and variance σ^2, T is a kxk transition matrix, and α is a fixed k-vector sometimes called the smoothing parameters vector. The initial state vector b_0 has a diffuse prior distribution (Ansley and Kohn, 1985) and is of the form $b_0 \sim N(0, \tau I)$ where τ is arbitrarily large. The special case, underlying the additive Holt-Winters smoothing method, known as the basic structural model (BSM), involves a level l_t, a growth rate g_t and a seasonal component s_t. It has the form

$$\begin{cases} y_t = l_{t-1} + g_{t-1} + s_{t-p} + e_t \\ l_t = l_{t-1} + g_{t-1} + \alpha_1 e_t \\ g_t = g_{t-1} + \alpha_2 e_t \\ s_t = s_{t-p} + \alpha_3 e_t \end{cases} \tag{2}$$

where α_1, α_2 and α_3 are three smoothing parameters and p is the number of seasons in one year. Equations (2) conform to the linear state space model (1). The state vector b_t consists of $l_t, g_t, s_t, \ldots, s_{t-p+1}$ and α is a $(p+2)$-vector equal to $(\alpha_1\ \alpha_2\ \alpha_3\ 0 \ldots 0)'$.

Using the backshift operator B, all of the elements in the state vector can be eliminated from (2) to yield the equivalent seasonal *ARIMA* process

$$\nabla \nabla_p y_t = \theta(B) e_t \tag{3}$$

where $\nabla = 1-B$, $\nabla_p = 1-B^p$ and $\theta(B)$ is a polynomial of degree $p+1$ in the lag operator B (Roberts, 1982). The relationships between the θ's and α's are given by

$$\begin{cases} \theta_1 = 1 - \alpha_1 - \alpha_2 \\ \theta_i = -\alpha_2,\ i = 2, \cdots, p-1 \\ \theta_p = 1 - \alpha_2 - \alpha_3 \\ \theta_{p+1} = -1 + \alpha_1 + \alpha_3 \end{cases} \tag{4}$$

A variant of the BSM, called the parsimonious model, avoids the explicit use of a seasonal equation. The seasonal effect is instead directly incorporated into the level equation. The new level m_t, which depends on the level in the corresponding period a year earlier, is augmented by the total growth in all seasons during the intervening year. The parsimonious model (PARS) is therefore defined as

$$\begin{cases} y_t = m_{t\text{-p}} + \sum_{i=1}^{p} g_{t-i} + e_t \\ m_t = m_{t\text{-p}} + \sum_{i=1}^{p} g_{t-i} + \beta_1 e_t \\ g_t = g_{t\text{-}1} + \beta_2 e_t \end{cases} \tag{5}$$

where the state vector b_t consists of m_t, ..., $m_{t\text{-}p+1}$, g_t, ..., $g_{t\text{-}p+1}$, α is a $2p$-vector comprising the two smoothing parameters β_1 and β_2 and equal to $(\beta_1\ 0\ \ldots\ 0\ \beta_2\ 0\ \ldots\ 0)'$. The state vector is much larger than its counterpart in BSM. But the number of smoothing parameters is reduced from three to two. Because PARS is linear in the state vectors, it is still more tractable than BSM. It can be proved that the parsimonious model also has an equivalent *ARIMA* process of the form (3) where

$$\begin{cases} \theta_1 = 1 - \beta_2 \\ \theta_2 = \cdots = \theta_{p-1} = -\beta_2 \\ \theta_p = 1 - \beta_1 - \beta_2 \\ \theta_{p+1} = \beta_1 - 1 \end{cases} \tag{6}$$

A comparison of (4) and (6) indicates that the second *ARIMA* representation is less flexible than the first.

3. Estimation and Prediction

The estimation objective, for both BSM and PARS, was to select α and b_0 to minimise the $SSE = \sum_{t=1}^{n} \tilde{e}_t^{\,2}$ where $\tilde{e}_t$ denotes the one-step ahead prediction error in period t. The computational strategy was to employ a general form of exponential smoothing, Ord, Koehler and Snyder (1997), for trial values of b_0 and α, to obtain the one-step ahead prediction errors. The best values of b_0 and α were found using the constrained optimiser in Gauss. The variance was then estimated with $\tilde{\sigma}^2 = SSE / n$. It can be established that the estimate of α obtained in this way also maximises the marginal likelihood function, Ansley and Kohn (1985), for the state space model (1).

On finding the estimates of α and b_0, the point predictions are generated with the equation

$$\tilde{y}_n(h) = x'_{n+h} b_{n+h-1|n} \tag{7}$$

where $\tilde{y}_n(h)$ is the forecasted value of the series for h-steps ahead and $b_{n+h|n}$ is the state vector estimate of b_{n+h} based on sample size n. The state vector estimates are obtained with $b_{n+h|n} = Tb_{n+h-1|n}$. For the two state space models in this study, (7) reduces to

BSM
$$\tilde{y}_n(h) = l_{n|n} + hg_{n|n} + s_{n-p+k|n} \tag{8}$$

PARS
$$\tilde{y}_n(h) = m_{n-p+k|n} + hg_{n|n} + \sum_{i=1}^{p-k} g_{n-i|n} \tag{9}$$

where $h \equiv k \mod(p)$ and $1 \le k \le p$.

4. Comparison using M-competition Data

The two methods were compared on real seasonal time series from the M-competition (Makridakis et al, 1982). There were 23 quarterly and 68 monthly series with forecasting horizons of 8 quarters and 18 months respectively. Accordingly, a corresponding number of observations were withheld from the fitting process at the end of each series and reserved to calculate forecast errors. The two methods, HW and PARS, were compared using the median absolute percentage errors (medAPE) where the absolute percentage errors are given by $APE\ (h) = 100(|\tilde{y}_n(h) - y_{n+h}|/y_{n+h})$.

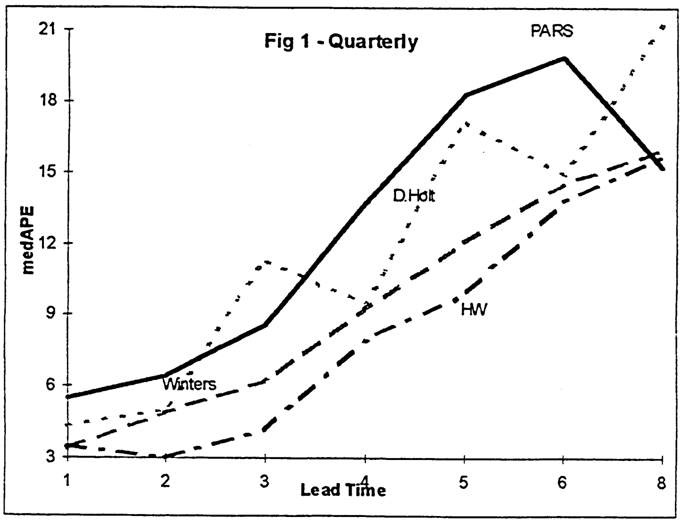
Table 1 shows the number of times each smoothing method is best for the one-step ahead forecasts. Overall, HW performs better than PARS. It is better 55% of the time.

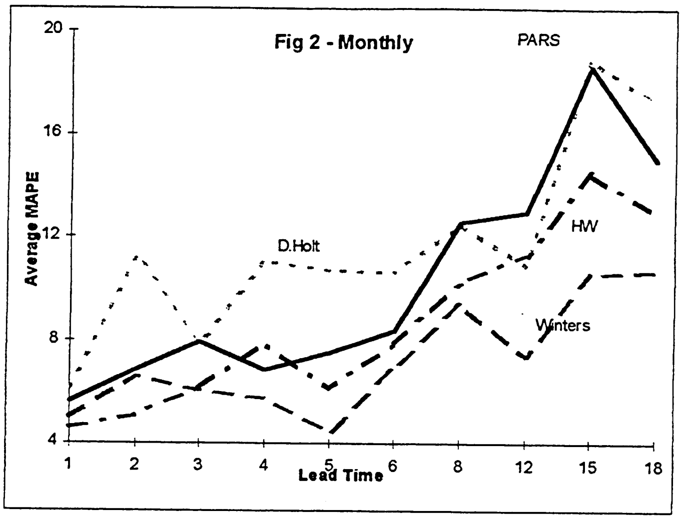
	H-W		PARS		Total
Monthly	38	56%	30	44%	68
Quarterly	12	52%	11	48%	23
Total	61	55%	50	45%	111

Table 1 - Number of times each smoothing method yields better APE

Comparison between our results and M-competition results

An aim of the study was to compare the forecast performance of the two procedures HW and PARS with some of those in the M-competition study (D-Holt and Winters). The D-Holt method consists of deseasonalising the seasonal series and applying Holt's trend corrected exponential smoothing method. The Winters method is the multiplicative form of the Holt-Winters method. In Figures 1 and 2, their medAPE's are compared for quarterly and monthly series respectively.





In interpreting the results, it should be noted that the time axis of Figures 1 and 2 contains gaps beyond horizon 6 (Makridakis did not report the errors at horizons 7, 9-11, 13-14, and 16-17). For quarterly series in Fig 1 and monthly series in Fig 2, the HW method performs better than PARS at most time horizons. Note that HW is the best method for the quarterly data. However, for the monthly data, HW is best for only the first two steps ahead. Beyond that Winters method performs better.

5. Conclusion

As expected the additive Holt-Winters method performed better than the parsimonious approach. Nevertheless, we were surprised by the magnitude of the differences in the M-competition series. Our original hope had been that the gap between the methods would be small enough to justify opting for the computational advantages of the parsimonious method at the expense of marginally less accurate forecasts. It appears that the additional seasonal equation and the associated parameter are critical for the generation of better forecasts of seasonal time series.

References

Ansley C. F. and Kohn R., 1985, Estimation, Filtering, and Smoothing in State Space Models with Incompletely Specified Initial Conditions, *The Annals of Statistics*, 13, 1286-1316.

Gardner E. S. Jr., 1985 Exponential Smoothing : The State of the Art, *Journal of Forecasting*, 4, 1-28.

Makridakis S. and al., 1982, The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition, *Journal of Forecasting*, 1, 111-153.

Ord J. K., Koehler A. and Snyder R. D., 1997, Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models, *Journal of the American Statistical Association*, forthcoming.

Roberts S. A., 1982, A General Class of Holt-Winters Type Forecasting Model, *Management Science,* 28, 808-820.

Snyder R. D., 1985, Recursive Estimation of Dynamic Linear Models, *The Journal of the Royal Statistical Society - Series B,* 47, 272-276.