A Seasonal Forecasting Model for General Use

P. M. Jellett: Charles Sturt University
Wagga Wagga NSW Australia
pjellett@csu.edu.au

Abstract Classical elementary forecasting methods represent a business time series as a time-varying trend with either multiplicative or additive seasonal components. This paper offers an estimation procedure for a nonlinear state-space formulation of a multiplicative time-varying seasonal model. The seasonal component of the model consists of enough fourier frequencies to cover a single seasonal period. The number of these may be selectively reduced. The formulation employs only two parameters to induce time-variation or local adaptability of the trend and seasonal components in contrast with the three usually employed by the Holt-Winters method. The implementation of the model provides for optimisation of these parameters and also includes fixed interval smoothing of the non-linear state-space formulation to produce output which is not only suitable for use in forecasting beyond the data record but also for graphical description of the data record. The method is illustrated with monthly airline passenger arrival data for New Zealand and also with a non-seasonal dataset of market shares for an Australian consumer product. The model is suitable for use by non-specialist users of forecasting methods.

1. INTRODUCTION

The multiplicative formula (1) is commonly used to model seasonal (eg monthly) time series. The *Holt-Winters* extension to the technique of exponential smoothing was designed to estimate such models. In another classical elementary multiplicative method (Makridakis and Wheelwright 1989), a time series can be deseasonalised by computing a time-centred moving average of appropriate length and then using the ratio of the raw data to the moving average to compute average seasonal factors for each month. An extrapolated trend can then be combined with the average seasonal factors to produce forecasts.

$$model_t = trend_t \times (1 + seasonal \quad factor_t) \tag{1}$$

$$model_t = trend_t + seasonal \quad term_t \tag{2}$$

Another approach to multiplicative models is to take logarithms of the data and fit additive seasonal and trend terms by multiple regression. Linear state-space methods, described by Bierman 1977, facilitate estimation of *time-varying* regressions. These models provide flexibility in handling missing values and also provide for a variety of polynomial trends and sinusoidal terms (Ameen and Harrison 1982, Jellett 1989, Jones 1980, Norton 1975 and 1976).

2. USAGE AND APPLICATION

However, if a time-varying regression is used to fit (2) to raw data which exhibit substantial growth and follow a simple multiplicative structure such as (1), this will be reflected in a need for increased parameter variability to account for changes in the seasonal fluctuations (Table 1), whose amplitude may grow with the trend. This idea is demontrated in Figures 1 to 4 which show smoothed multiplicative and additive models together with associated seasonal components. The data were monthly airline passenger arrivals for New Zealand. Figures 5 to 8 show one step ahead forecasts and residuals for the two models. The final 12 months of data were left out of the estimation procedure for later comparison with forecasts (Figure 9). Figures 10 to 13 show residual autocorrelations and partial autocorrelations for the two models. The multiplicative model produced lower residual autocorrelations, lower forecast standard error (Figure 9), lower 1-step-ahead residuals and a more constant (ie slowly changing) seasonal pattern than the additive model. Table 1 gives the estimated noise variance ratios which govern the adaptive response of seasonal and trend components of the models. A side-effect of inappropriate use of the additive model is the apparent lower level of detail in the estimated trend (Table1, Figures 1 and 2).

In order to apply the model class developed in this paper the user need only specify the order of the trend polynomial, the seasonal periodicity and whether the model is additive or multiplicative. Conceptually the models correspond to exponential smoothing and moving average methods. These amount to a structural model for the non-stationary components of a time series and are well established in forecasting practice owing to wide applicability and relative ease of use. Thus non-specialist users of forecasting methods may apply the methods of this paper.

Table 1: Noise Variance Ratios

Model	Seasonal (ω_1)	Trend (ω_2)
Multiplicative	0.00296	0.00264
Additive	0.00534	0.000413

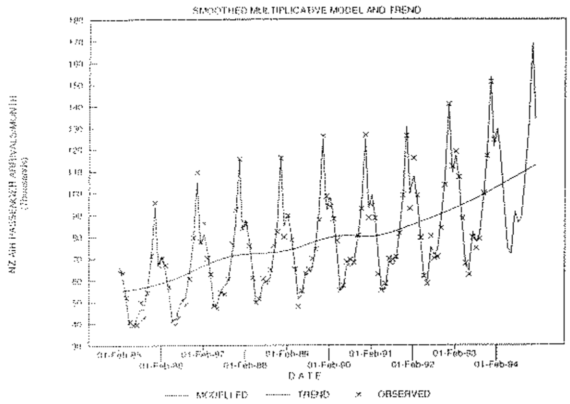

Figure 1: Multiplicative Model for New Zealand

3. MODEL FORMULATION

The present paper offers an iterative algorithm for fitting various time-varying multiplicative models and uses a linear additive state-space formulation as the basis for each iteration, given known stochastic process noise variance ratios (W below and Table 1 above). The model is fitted to the raw data directly and can accommodate missing values and the range of flexibilities inherent in the state-space representation. By representing the likelihood as a weighted sum-of-squares, the unknown stochastic parameters (W below), could then be estimated by non-linear least squares (Box and Jenkins 1976, Marquardt 1970), utilising numerically evaluated derivatives of the weighted errors. This approach is favoured by the author as a simpler and more reliable procedure than carrying out the optimisation via more general mathematical programming techniques (Brockwell 1991, Shumway and Stoffer 1982).

3.1 Additive Model

The model equations are

$$y_t = x_t^T \beta_t + e_t \tag{3}$$

$$\beta_t = G\beta_{t-1} + Q w_t \tag{4}$$

where t runs from 1 to n. In the class of models being covered here, the scalar random errors, e_t, are independent and normally distributed random variables with zero mean and unknown variance, σ^2, uncorrelated with w_t which are independent and normally distributed with zero mean and covariance matrix, $\sigma^2 W$, where W is the diagonal matrix of noise variance ratios. In the implementation of the present class of models only two elements from the diagonal of W can be given non-zero values or estimated. To formulate a particular model, one must specify

$$G,\quad x_t,\quad Q \quad \text{and} \quad W.$$

Useful formulation examples are given by Ameen and Harrison (1982) and Norton (1975 and 1976). For monthly data

$$G = \begin{bmatrix} G_1 & & & & & & \\ & G_2 & & & & \underset{\sim}{0} & \\ & & G_3 & & & & \\ & & & G_4 & & & \\ & & & & G_5 & & \\ & \underset{\sim}{0} & & & & G_6 & \\ & & & & & & G_7 \end{bmatrix}$$

where

$$G_k = \begin{bmatrix} \cos\left(\frac{2\pi k}{12}\right) & \sin\left(\frac{2\pi k}{12}\right) \\ -\sin\left(\frac{2\pi k}{12}\right) & \cos\left(\frac{2\pi k}{12}\right) \end{bmatrix}$$

and k takes values from 1 to 6.

$$G_7 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$Q^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$W = \begin{bmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{bmatrix}$$

$$x_t^T = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]$$

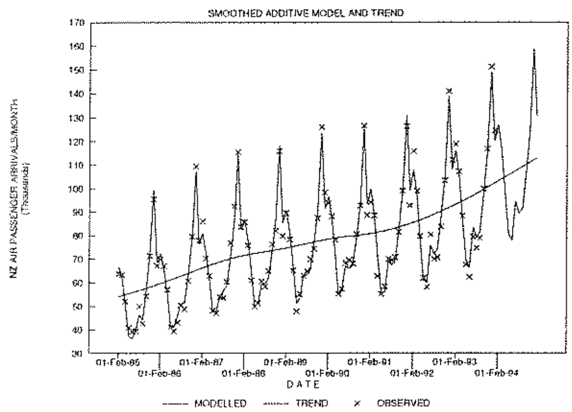

Figure 2: Additive Model for New Zealand

In practice, W may be specified by trial and error in the same way as the smoothing constants of exponential smoothing. Alternatively a procedure for optimising W is described below.

3.2 Multiplicative Model

In the additive model for monthly data the 13-th element of β_t, corresponds to the trend. Now define the multiplicative model by replacing x_t in the *observation* equation (3) with

$$\chi_t = [\tau_t \quad 0 \quad \tau_t \quad 0 \quad \tau_t \quad 0 \quad \tau_t \quad 0 \quad \tau_t \quad 0 \quad \tau_t \quad 0 \quad 1 \quad 0]^T$$

$$\text{and } \tau_t = (\beta_t)_{13}$$

The model is now nonlinear and the observation equation for the multiplicative model is

$$y_t = \chi_t^T \beta_t + e_t \tag{5}$$

The *system* equation (4) remains unchanged and with (5) now constitutes the multiplicative model.

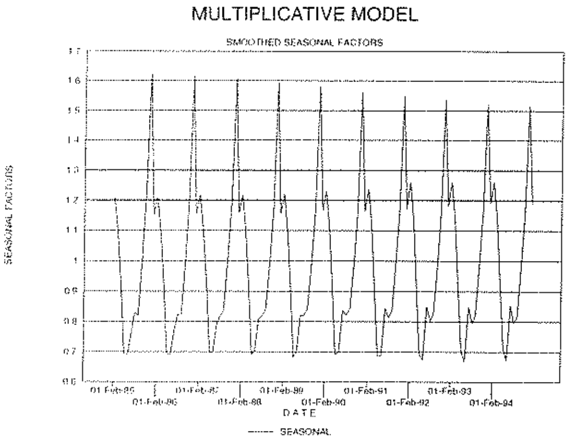

Figure 3: Multiplicative Seasonal Factors

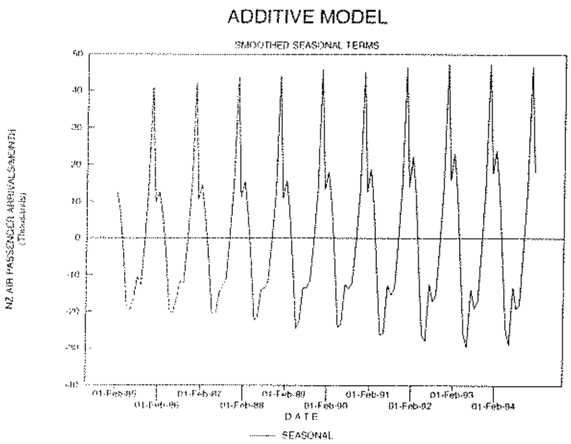

Figure 4: Additive Seasonal Terms

3.3 Relation to Classical Time Series Models

Partition x_t, χ_t and β_t as:

$$x_t = [x_{S.t} \quad x_{T.t}]^T,$$

$$\chi_t = [\chi_{S.t} \quad \chi_{T.t}]^T,$$

$$\beta_t = [\beta_{S.t} \quad \beta_{T.t}]^T$$

with

$$\chi_{S.t} = [\tau_t \quad 0 \quad \tau_t \quad 0 \quad \tau_t \quad 0 \quad \tau_t \quad 0 \quad \tau_t \quad 0 \quad \tau_t \quad 0]^T$$

$$x_{S.t} = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]^T$$

$$\chi_{T.t} = x_{T.t}$$

$$= [1 \quad 0]^T$$

Now the multiplicative model equation is

$$y_t = \chi_{S.t}^T \beta_{S.t} + \chi_{T.t}^T \beta_{T.t} + e_t \tag{6}$$

$$= trend_t \times seasonal factor_t + trend_t + e_t$$

$$= (seasonal factor_t + 1) \times trend_t + e_t$$

and the additive equation is

$$y_t = x_{S.t}^T \beta_{S.t} + x_{T.t}^T \beta_{T.t} + e_t \tag{7}$$

$$= seasonal \quad term_t + trend_t + e_t$$

corresponding to the classical models, equations (1) and (2), respectively.

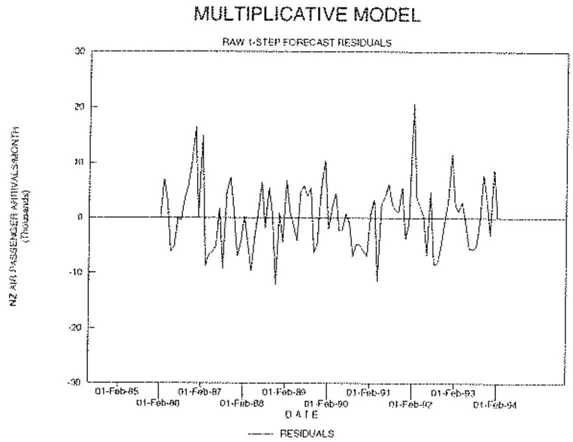

Figure 5: Multiplicative Forecast Errors

4. ESTIMATION

The additive model, equations (3) and (4), can be estimated by fixed interval smoothing (Bierman 1977, Norton 1975 and 1976, Sarris 1974, Shumway and Stoffer 1982) if W is known. Otherwise some nonlinear optimisation method will be required to estimate W. In the case of the multiplicative model the estimation is nonlinear whether or not W is given. The following two schemes, led to estimates for the multiplicative model, while for the additive model the second scheme alone is sufficient:

- Estimation of β_t in the multiplicative model given known W: An Iterative Instrumental Variable method (eg Young 1984) is given below. $\hat{\chi}_t^i$, below, approximates the gradient of the model residuals with respect to $\hat{\beta}_t$. Thus satisfactory statistical properties and numerical performance are well established.

- Estimation of W, assuming in the case of the multiplicative model, that the trend is known.

4.1 Estimation of β_t given known W

Write T_t for the optimally smoothed trend estimate at time t. Then, for the additive model equation (3),

$$T_t = (\beta_{t/n})_{13}$$

where the *hat* notation indicates the computed estimate of the corresponding *population* parameter, $(\beta_t)_{13}$. *Fixed interval smoothing* is used to obtain $\hat{\beta}_{t/n}$ with the additive model equations (3) and (4).

At each iteration in the nonlinear estimation of the multiplicative model equations, (5) and (4), fixed interval smoothing is applied with χ_t replaced at the i-th iteration by its estimate from the (i-1)-th iteration:

$$\hat{\chi}_t^i = [T_t^{i-1} \quad 0 \quad T_t^{i-1} \quad 0 \quad T_t^{i-1} \quad 0 \quad T_t^{i-1} \quad 0 \quad T_t^{i-1} \quad 0 \quad T_t^{i-1} \quad 0 \quad 1 \quad 0]^T$$

$$\hat{\chi}_t^1 = [1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0]^T$$

$$T_t^{i-1} = (\hat{\beta}_{t/n})_{13}^{i-1}$$

At convergence the trend estimate for the multiplicative model is $T_t = T_t^{\infty}$.

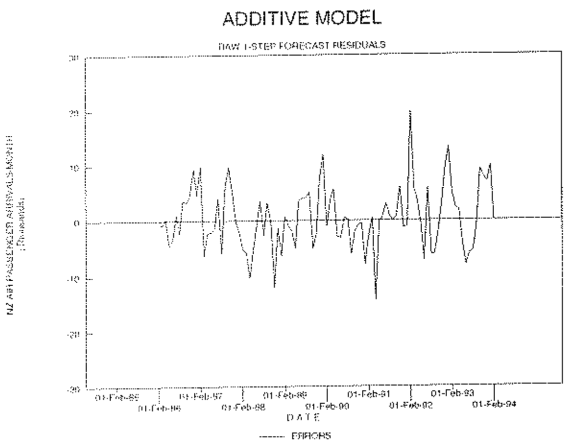

Figure 6: Additive Forecast Errors

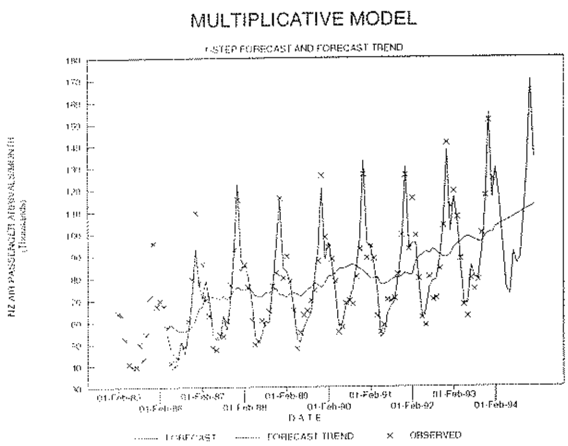

Figure 7: Multiplicative 1-Step Forecasts

4.2 Estimation of W

For a multiplicative model T_t, the estimate of τ_t was determined from the scheme given in section 4.1. The likelihood was evaluated for given W in both additive and multiplicative cases. It was expressed as a weighted sum-of-squares minimisation, permitting use of non-linear least squares.

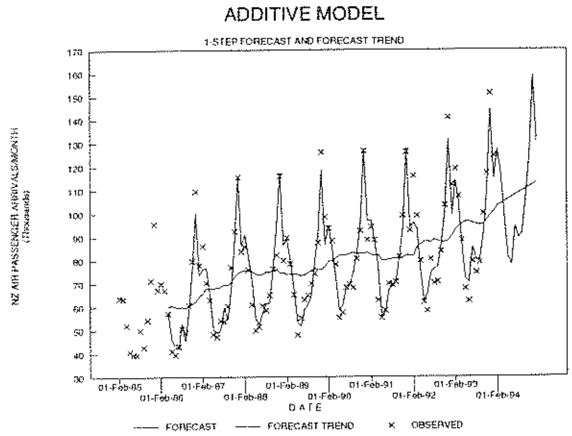

Figure 8: Additive 1-Step Forecasts

Much literature, (eg Shumway and Stoffer 1982), consider the likelihood of a state-space model as a function of unknown G, with full-rank QW (not the case here), possibly containing off-diagonal elements. Multivariate dependent variables are often included in such models. General mathematical programming techniques are then applied to determine values of the unknowns. Existing estimation literature does not appear to consider the simple multiplicative model here where only two elements in W remain to be estimated. The likelihood maximisation is equivalent to minimisation (Brockwell 1991, Sarris 1972, Shumway and Stoffer 1982) of

$$L = -\frac{1}{2}\sum_{t=1}^{n} \ln[D_t\hat{\sigma}^2] \quad -\frac{1}{2}\sum_{t=1}^{n} \frac{(y_t - \hat{y}_{t/t-1})^2}{D_t\hat{\sigma}^2} \tag{8}$$

$\hat{y}_{t/t-1}$ are the one-step predictions obtained from the Kalman filter. $D_t\sigma^2$ are the variances of the prediction errors, $y_t - \hat{y}_{t/t-1}$. Differentiating (8) partially with respect to $\hat{\sigma}^2$ led to (Jones 1980)

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \frac{(y_t - \hat{y}_t)^2}{D_t} \tag{9}$$

Next (9) was substituted into (8). The quantity to be minimised was then

$$L^* = \sum_{t=1}^{n}\left[(y_t - \hat{y}_t)\sqrt{\frac{\overline{D}}{D_t}}\right]^2$$

where

$$\overline{D} = \left[\prod_{t=1}^{n} D_t\right]^{\frac{1}{n}}$$

L^* was minimised by nonlinear least squares with numerical derivatives of the weighted errors in the square brackets. L^* is the geometric mean of prediction error variances multiplied by the sum of squared innovations.

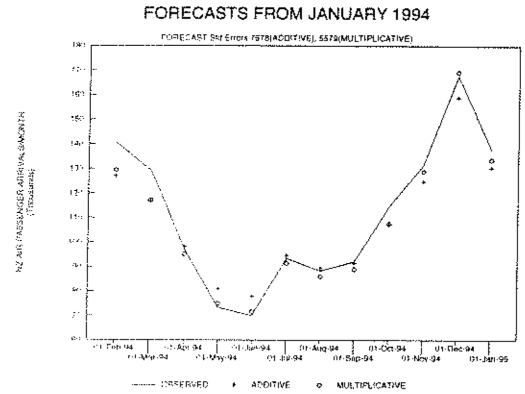

Figure 9: Forecasts Up to 1 Year Ahead

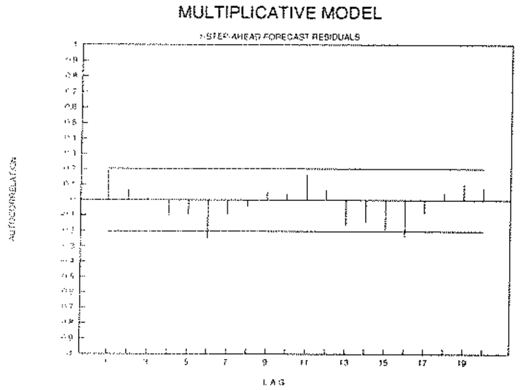

Figure 10: Multiplicative Model ACF

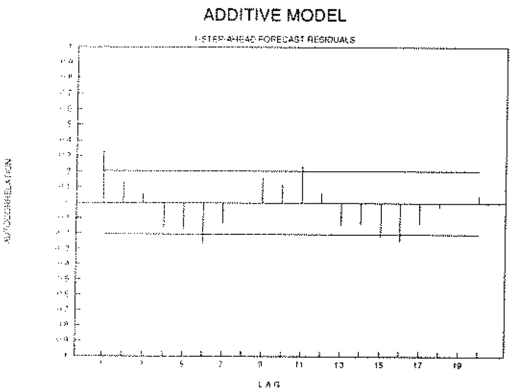

Figure 11: Additive Model ACF

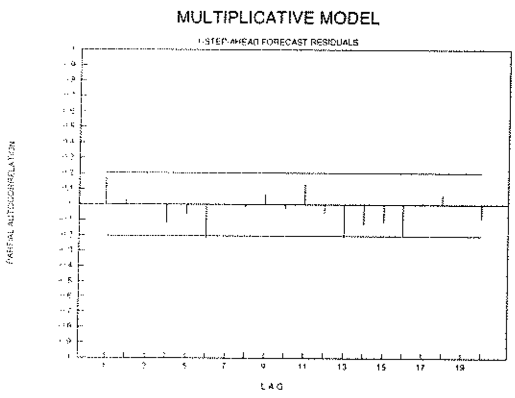

Figure 12: Multiplicative Model PACF

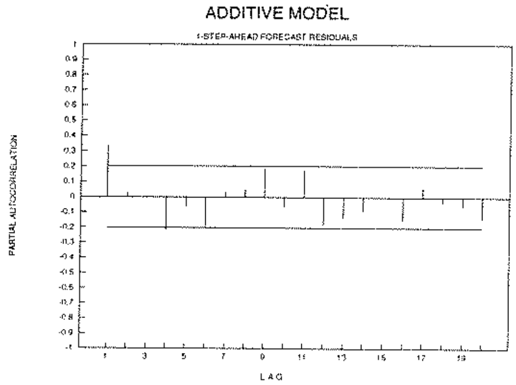

Figure 13: Additive Model PACF

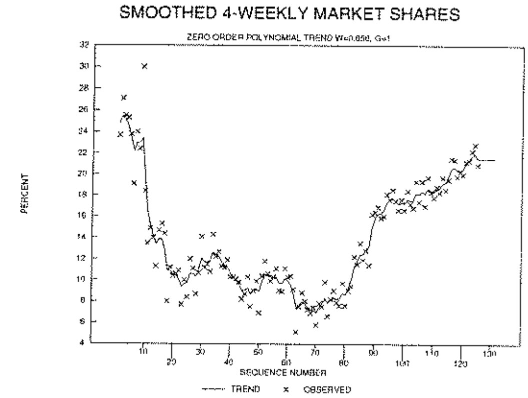

Figure 14: Simple Exponential Smoothing

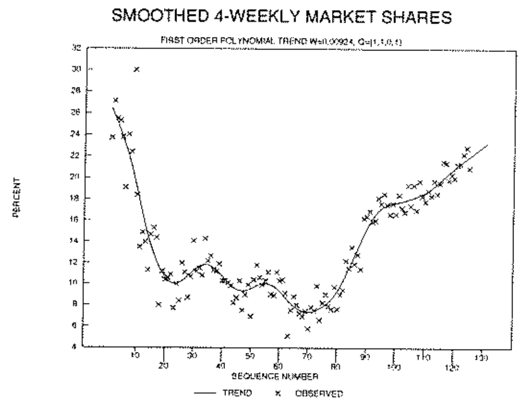

Figure 15: Linear Growth

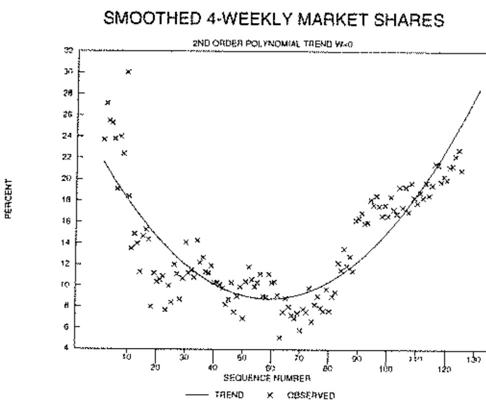

Figure 16: Quadratic Fit

5. NON-SEASONAL DATA

Figures 14 to 18 show results for a non-seasonal dataset. With the foregoing seasonal application they show the flexibility of the present model class in fitting polynomial trends and sinusoidal terms. Figure 14 shows an adaptive zero order trend which corresponds with simple exponential smoothing and is the model of choice. However the graphed trend was fixed-interval smoothed, rather than exponentially smoothed (ie filtered). Thus it was calculated from both future and past data as in a time-centred moving average but according to a model. Figure 15 shows a first order polynomial trend as in the trend component of the model for the New Zealand data. Figures 16 and 17 show non-adaptive and adaptive quadratic trends respectively.

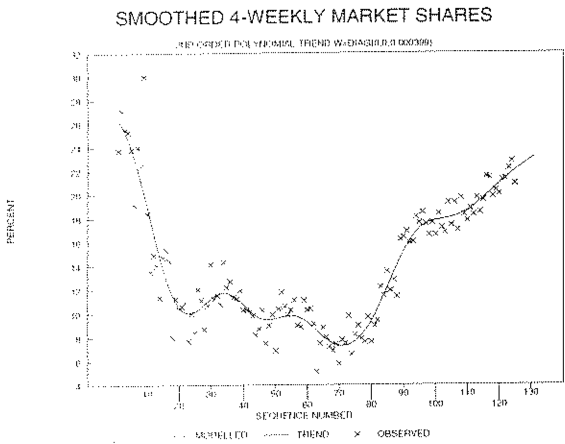

Figure 17: Adaptive Quadratic Fit

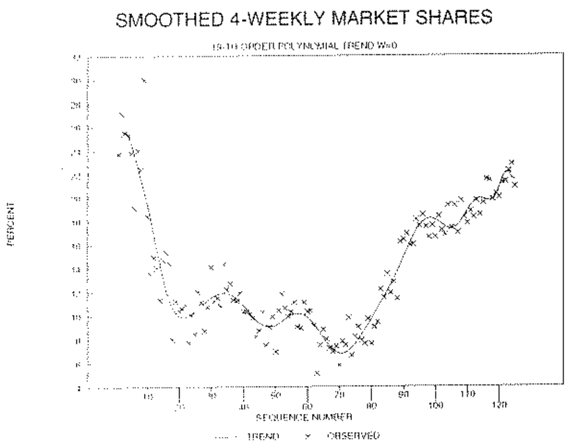

Figure 18: Non-adaptive Polynomial

6. CONCLUSION AND ACKNOWLEDGEMENT

This paper described a class of univariate time series models. The literature express Kalman Filter and fixed-interval smoothing formulae in terms of matrices. The implementation here required extensive repetition of the calculations in order to evaluate derivatives based on central differences in an optimisation. However many of the matrices contained zero's, one's or repeated trend entries, permitting execution-time efficiencies.

State-space literature recommend use of square-root methods for filtering and smoothing calculations (Kaminski et al 1971, Bierman 1977) for accuracy. These were dispensed with here in favour of speed. Instead two concessions were needed towards accuracy. Firstly the data were scaled to have a mean of one and then, with results, unscaled after completion of the calculations. Secondly the calculations were arranged so that inner products generally led to sums of squares rather than sums of powers of 4. This was achieved at the cost of some extra square root evaluations. Figure 18 shows that accuracy was adequate in the fitting of a non-adaptive 19-th order polynomial trend to a non-seasonal dataset. Despite the usual recent improvements in computing performance the calculations are sufficiently arduous so that computing time remains practically important.

A multiplicative time-varying seasonal model and associated estimation procedures were developed. The adaptive response of the trend and seasonal components were driven by two noise variance ratios in contrast with the three of the Holt-Winters method. The model was implemented in a general format for a range of polynomial trends and seasonal periodicities. Application to New Zealand short-term airline passenger arrival data showed that the multiplicative model was indeed appropriate and produced reasonable forecasts up to one year beyond the data record. The New Zealand airline passenger arrival data were kindly provided by Peter Thomson of the University of Wellington, New Zealand.

7. REFERENCES

Ameen, J.R.M., and P.J. Harrison, Discount weighted estimation, *Warwick University* Report no. 28, October, 1982.

Bierman, G.J., Factorization methods for discrete sequential estimation, *Academic Press*, 1977.

Box, G.E.P., and G.M. Jenkins, Time series analysis: forecasting and control, *Holden-Day*, 1976.

Brockwell, P. J. and R. A. Davis, Time series: theory and methods, *Springer-Verlag*, 1991.

Jellett, P.M., Missing values and parametric variation in time series regressions, *Proceedings 8th Biennial Conference*, Simulation Society of Australia, A.N.U., Canberra, 184-189, 1989.

Jones, R. H., Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics*, Vol 22, No. 3, 389-395, 1980.

Kaminski, P. G., A. E. Bryson Jr and S. F. Schmidt, Discrete square root filtering: a survey of current techniques, *IEEE Trans AC*, Vol, AC-16, No. 6, 727-735, 1971.

Makridakis, S., and S. Wheelwright, Forecasting methods for management, 5th edition, *Wiley*, Brisbane, 1989.

Marquardt, D.W., Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation, *Technometrics*, Vol 12, no 3, 591-612, August, 1970.

Norton, J.P., Optimal smoothing in the identification of linear time-varying systems'', *Proc. IEE*, Vol 122, No 6, 663-668, June, 1975.

Norton, J.P., Identification by optimal smoothing using integrated random walks, *Proc. IEE*, Vol 123, No 5, 451-452, May, 1976.

Sarris, A., A general algorithm for simultaneous estimation of constant and randomly-varying parameters in linear regressions, *NBER working paper* 38, 20pp, April, 1974.

Shumway, R.H., and D.S. Stoffer, An approach to forecasting and smoothing using the EM algorithm'', *Jnl Time Series Anal*, Vol 3, No 4, 253-264, 1982.

Young, P., Recursive estimation and time-series analysis, *Springer-Verlag*, 1984.