Forecasting Models for Different Horizons

Jamal R.M. Ameen[1,2]*

[1]*University of Kurdistan Hewler, Erbil, Kurdistan Region – F.R. Iraq*
[2]*University of South Wales, United Kingdom*

*Corresponding author's email: jamal.ameem@ukh.edu.krd

ABSTRACT

A general class of models is introduced to provide robust and practically simple forecasts for different horizons. These are obtained in the presence of both high and low frequencies in the data. A submodel is constructed to estimate low-frequency state parameters, and the results obtained are used as conditional information in modelling the high-frequency parameters. All parameters have probability distributions, and the discount principle of Ameen and Harrison (1985) is used in the posterior-prior state distribution transitions. A number of limiting results and special cases are also discussed.

Keywords: Bayesian Modelling, Discounting, Forecasting, Long- and Short-term Forecasting Models

1. INTRODUCTION

One of the main objectives in analyzing data is to learn from past events and to incorporate this knowledge for future planning and decision-making. This involves the use of all the current and past information to forecast future outcomes. In forecasting, the time horizon is determined by the characteristic timescales of the process concerned and the objective in exercising the analysis. For example, the scheduling of production and inventories requires short-term forecasts while longer horizons are required when the long-term prospect ranging from expansions in a plant to environmental factor impacts is investigated (Niu and Harris, 1996; Kwon *et al.*, 2002). Typical periods in the first case are 10–14 weeks and 3–5 years and more for the second.

The problem of short-term forecasting has been addressed extensively in the past. A detailed review of various short-term forecasting models can be found in Harvey (1984). However, long-term forecasting, from a practical viewpoint, is considered more qualitative in nature involving both the forecasters' skills and the available information on an *ad hoc* basis, has suggested by Kendall and Ord (1993), Armstrong (1985), and Martino (1983). Research on theoretical aspects of long-term forecasting models has covered various modifications of short-term forecasting models. Kabaila (1981), Stoica and Soderstorm (1984), and Tiao and Xu (1993) have suggested different least square parameter estimates of the one-step ahead forecasting model to incorporate multiple steps ahead predictions. Different Box Jenkins models have been suggested by Findley (1983), Gersch and Kitagawa (1983), and Lin and Granger (1994). Furthermore, Pillai (1992) has used the maximum entropy criterion for multistep predictions. Other strategies include the direct "plug-in" method where unobserved values are successively replaced by their one-step ahead forecasts. Bhansali (1996, 1999) has provided theoretical grounds for the optimality in using a single model for multiple steps ahead forecasting while Kang (2003) has suggested the use of different models for different spans of the process under investigation.

In general, the possibility of incorporating different horizons forecasts requires the study of two types of variation. These are the inherent structural variation that characterizes the long-term nature of the process and the short-term high-frequency variation. The existing short (long)-term forecasting models are not designed to separate these characteristics, and therefore, not reliable in generating long (short)-term forecasts.

In this paper, a new approach is introduced in which the characteristics of different variational parameters are preserved making forecasting over different horizons possible in a logical manner.

The close link between autoregressive integrated moving averages (ARIMA) and the state space representation is explained in both the statistical and control engineering literature (Box and Jenkins, 1970; Anderson and Moore, 1979).

Assuming that $\{Y_t | \theta_{1,t}\}$ is a time series process with the parameter vector $q_{1,t}$ representing the state of the process, a dynamic linear model (DLM) is shown in the Equation (1):

$$Y_t = F_{1,t}\theta_{1,t} + v_t, \qquad v_t \sim N[0, V_t] \tag{1}$$

And, the state transition is calculated in Equation (2):

$$\theta_{1,t} = G_{1,t}\theta_{1,t-1} + w_t, \quad w_t \sim [0, W] \tag{2}$$

Where $F_{1,t}$ and $G_{1,t}$ are known, and $\{\vartheta_t\}$ and $\{\omega_t\}$ are uncorrelated normal variates with zero means.

This has been extended to model correlated variables by adding a third Equation (3) representing transitions on v_t so that:

$$v_t = G_2 v_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N[0; \sigma^2] \tag{3}$$

Where G_2 is assumed known and ε_t is a normal variate with zero mean vector (Anderson and Moor 1979. p. 290). However, the model can be reparametrized as a non-linear predictor model (NPM), Ameen (1992), in which the autoregressive components, G_2, can be assumed known and updated through time.

A main drawback of linear structures in modeling correlated observations is that information is filtered between different state parameter components using the usual Kalman filter. The underlying parameters are updated according to a common one-step forecasting error. This makes causal model parameters to lose their characteristic properties during the sequential updating process. For example, it is not possible to consider one-step forecasting errors that are calculated from a short-term model for the parameters identifying long-term structures and vice versa. This may suggest that the two characteristics are separable and independent modeling structures should be implemented. Earlier efforts have been concentrated on the idea of curve fitting. Gregg *et al.* (1964) and later Harrison and Pearce (1972) have presented comprehensive accounts of the ICI methodology for long-term forecasting. These authors have discussed extrapolation techniques using some specific functions as trend curves. More discussions on these curves and confidence limits on forecasts can be found in the study of Levenbach and Reuter (1976). Harrison *et al.* (1977) have constructed models that consider randomness around predetermined linear/non-linear global trends all constructed to provide long-term forecasts. Although there is a logical connection between forecasts for different horizons, lead time periods separating short-, medium-, and long-term forecasts are rather subjective and difficult to justify.

The discounting principle of Ameen and Harrison (1985) has extensively been used in Bayesian modeling and forecasting (Harrison and Akram, 1983; Migon and Harrison, 1985). In these models, low frequencies are modeled using autoregressive parameters. Although this can accommodate the correlation between the departures from the underlying trend, the identification and estimation of many autoregressive parameters can be tedious. Moreover, different model parameter updates are dominated by one source of information (one-step forecasting errors). The models have no facilities to encapsulate the long-and short-term behavior of the time series. More recently, Danesi *et al.* (2017) have followed a form of variability decomposition (Al-Madfai, 2002; Al-Madfai *et al.*, 2004) extracting some underlying smooth function to capture the long-term behavior of the time series in question and adopt the ARIMA modeling technique on the remainder.

This paper is an attempt to resolve these problems by introducing a fairly general method in which characteristic properties of different state parameter vectors are preserved in updating. This is performed using the idea of submodels. The problem of identification and estimation of autoregressive parameters is overcome through the introduction of a set of easily assessed discount factors. Although the idea is fairly general, we concentrate on the case of normality with random level and growth components. Generalizations to higher dimensions follow easily. The general idea is presented in Section 2 in which full recursive formulas for parameter estimates are given. A number of limiting properties are discussed, and models structural behavior and limiting forecast functions are given in Section 3.

Throughout this paper, $X \sim N[a; b]$ means that X is normally distributed with mean a and variance b. The same holds for random vectors.

2. LOCAL AND GLOBAL FORECASTING MODELS

Let, $\{Y_t | \theta_t\}$ be a process with low- and high-frequency variations. Furthermore, assume that these variations are represented by $\theta_{1,t}$ and $\theta_{2,t}$ as the components of θ_t, respectively. Furthermore, let M_1 and M_2 be subjects modeling the existing low- and high-frequency parameters, respectively. In the above, the parameter vector $\theta_{1,t}$ may be identifying a general polynomial trend approximating the inherent underlying process structure while $\theta_{2,t}$ may represent the causal variability around that of $\theta_{1,t}$.

The common approach for modelling such scenarios has been through the specification of an observation probability distribution $f(y_t | \theta_{1,t}, \theta_{2,t})$ accompanied by a prior state distribution $f(\theta_{1,t}, \theta_{2,t} | D_{t-1}, M_1, M_2)$ where D_{t-1} represents the information available at time $t-1$ including past observations. A before posterior transition in time together with the above specifications allows a complete Bayesian sequential estimation to be conducted for the estimation of model parameters. The general DLMs of Harrison and Stevens (1976) and NPM of Ameen (1992) are examples of such kind. The broad specification above contains both the high and low dimensionalities of the process in a combined form. As a result, the defined models do not allow the two parameter vectors $\theta_{1,t}$ and $\theta_{2,t}$ to preserve their characteristics as they are both updated using one error structure. In such cases, a joint forecasting distribution seems inadequate as the updating structure for $\theta_{1,t}$ and $\theta_{2,t}$ remains less informative about the short- or long-term behavior of the process.

The two types of variability discussed above are conditionally independent. That is, given the observation y_t, $\theta_{1,t}$ is independent on $\theta_{2,t}$ and this can be exploited to learn about these parameters in one model. The resulting model can be used to bridge the gap between forecasts for different horizons and make data analysis more efficient in some practical cases.

The practical need for estimating parameters of different characteristics has led by Smith *et al.* (1994) to construct a two-stage sales forecasting procedure using discounted least squares. The low-frequency variations are modelled in one stage followed by the modeling of the departures from this trend as a second stage.

Using the above terminology, we have the following Equation (4):

$$Z_{t+1} = Y_{t+1} - E(Y_{t+1} | D_t, M_1) \tag{4}$$

Moreover, the general forecast function may be expressed as follows:

$$E(Y_{t+k} | D_t, M) = E(Y_{t+k} | D_t, M_1) + E(Z_{t+k} | D_t, M) \tag{5}$$

Where $M = \{M_1, M_2\}$.

Furthermore, the departures Z_t are modelled so that $E\{Z_{t+k} | D_t, M\} \to 0$ as $k \to \infty$. This property ensures a logical transition between forecasts for different horizons from a single model as the lead-time increases.

The underlying modeling strategy is described as follows:

Given the observation distribution:

$$Y_t | \theta_{1,t}, V_t \sim N[F_{1,t}\theta_{1,t}; V_t^{-1}] \tag{6}$$

And, posterior distributions at time $t-1$:

$$V_{t-1} | D_{t-1}, M_1 \sim \Gamma(\frac{r_{t-1}}{2}, \frac{s_{t-1}}{2})$$

$$\theta_{1,t-1} | D_{t-1}, V_{t-1}, M_1 \sim N[m_{1,t-1}; V_{t-1}^{-1}R_{t-1}]$$

The prior distributions for time t are defined as follows:

$$V_t | D_{t-1}, M_1 \sim \Gamma(a\frac{r_{t-1}}{2}, b\frac{s_{t-1}}{2}) \tag{7}$$

$$\theta_{1,t} | D_{t-1}, V_t, M_1 \sim N[m_{1,t-1}; V_t^{-1}R_{1,t}]$$

Where

$$R_{1,t} = \beta_1^{-1}G_1 C_{1,t-1} G_1^T$$

And, $\beta_1 < 1$ is a discount factor with a value subjectively selected to be close to 1. The constants a and b are introduced to accommodate trends in the evolution of the likelihood variance since $E(V_t | D_{t-1}) = ar_{t-1}/bs_{t-1}$.

The above model incorporates correlated inputs and can be extended to multiple discount factors as necessary (Ameen and Harrison, 1985; Ameen, 1992 for more elaboration on the choice of the discount factor and the constants a and b). A sequential updating procedure is then obtained representing the evolution of the low-frequency parameters with time.

The resulting forecasting distribution is as follows:

$$Y_t | V_t, D_{t-1}, M_1 \sim N[\hat{y}_{1,t}; V_t^{-1}Q_{1,t}] \tag{8}$$

Where

$$\hat{y}_{1,t} = F_{1,t} G_{1,t} m_{1,t-1}, \tag{9}$$

$$Q_{1,t} = F_{1,t} R_{1,t} F^T_{1,t} + 1. \tag{10}$$

The posterior state distribution for time t is as follows:

$$\theta_{1,t} | V_t, D_{t-1}, y_t, M_t \sim N[\, m_{1,t}; V_t^{-1} C_{1,t} \,] \tag{11}$$

Where

$$m_t = G_1 m_{1,t-1} + A_{1,t} e_{1,t}$$

$$C_{1,t} = (I - A_{1,t} F) R_{1,t}$$

Moreover, the updated variance distribution is as follows:

$$V_t | D_t \sim \Gamma\left(\frac{r_t}{2}, \frac{s_t}{2}\right) \tag{12}$$

Where

$$r_t = a r_{t-1} + 1 \text{ and } s_t = b s_{t-1} + e_{t-1}^2 / Q_{1,t}$$

Note that, all the above distributions that are conditional on V_t, their unconditional distributions can be found using the rules of probability leading to t-distributions with appropriate parameter values.

Alternatively, V_t values can be replaced by their estimates keeping the distributions normal.

Now, given the joint likelihood for the high and low frequencies as:

$$Y_t | \theta_{1,t} \theta_{2,t} \sim N[F_{1,t} \theta_{1,t} + F_{2,t} \theta_{2,t}; V_{2,t}], V_{2,t} N[0; V_2] \tag{13}$$

Where the $V_{2,t}$'s are all uncorrelated, the information in the Equation (6) and Equation (11) can be used to construct the posterior distribution for the high-frequency system parameter vector $\theta_{2,t-1}$:

$$\theta_{2,t-1} | \theta_{1,t-1}, D_{t-1}, M \sim N[m_{2,t-1} + P_{t-1}(\theta_{1,t-1} - m_{1,t-1}); C_{2,t-1}] \tag{14}$$

Where $m_{2,t-1} = E\{\theta_{2,t-1} | M, D_{t-1}\}$ and P_{t-1} is a known $m \times n$ matrix similar to the usual regression coefficients and whose components are functions of the adaptive factors.

The prior conditional distribution for $\theta_{2,t}$ is as follows:

$$\theta_{2,t} | \theta_{1,t}, D_{t-1}, M \sim N[G_2 m_{2,t-1} + P^*_{t-1}(\theta_{1,t} - G_1 m_{1,t-1}); R_{2,t}] \tag{15}$$

Where P^*_{t-1} is obtained from P_{t-1} to reflect the added degree of uncertainty due to the transition in time. For example, a direct calculation based on the normal discount Bayesian models setting of Ameen and Harrison (1985) leads to $P^*_{t-1} = \delta G_2 P_{t-1} G_1^{-1}$, where $\delta = (\beta_2 / \beta_1)^{1/2}$ and $R_{2,t}$ is obtained using the discount principle described earlier. The discount factor β_2 is similar to β_1 but smaller in magnitude to increase the adaptivity for the estimation of high-frequency parameters.

Combining the Equation (13) and Equation (15) and using Bayes theorem, the updating conditional distribution of $\theta_{2,t}$ is obtained.

$$\theta_{2,t} | \theta_{1,t}, D_{t-1}, y_t, M \sim N[m_{2,t} + P_t(\theta_{1,t} - m_{1,t}); C_{2,t}] \tag{16}$$

Where

$$C_{2,t}^{-1} = R_{2,t}^{-1} + V_2^{-1} \tag{17}$$

$$m_{2,t} = G_2 m_{2,t-1} + A_{2,t} e_{2,t} + P_t A_{1,t} e_{1,t} \tag{18}$$

$$P_t = P^*_{t-1} - A_{2,t} F_2 P^*_{t-1} - A_{2,t} F_1 \tag{19}$$

And,

$$e_{2,t} = y_t - F_1 G_1 m_{1,t-1} - F_2 G_2 m_{2,t-1} \tag{20}$$

The forecast distribution is given by

$$Y_t | D_{t-1}, M_t \sim N[\, \hat{y}_{2,t}; H_t \,] \tag{21}$$

Where

$$\hat{y}_{2,t} = F_1 G_1 m_{1,t} + F_2 G_2 m_{2,t} \tag{22}$$

And,

$$H_t = (F_1 - F_2 P^*_{t-1}) R_{1,t} (F_1 - F_2 P^*_{t-1})^T + F_2 R_{2,t} F^T_2 + V_2 \tag{23}$$

The general k steps forecast function is given by

$$F_t(k) = F_1 G_1^k m_{1,t} + F_2 G_2^k m_{2,t} \tag{24}$$

With a forecasting variance of

$$H_t(k) = (F_1 - F_2 P^*_{t-1}(k)) R_{1,t}(k) (F_1 - F_2 P^*_{t-1}(k))^T + F_2 R_{2,t}(k) F^T_2 + V_2 \tag{25}$$

Where

$$P_{t-1}(k) = \delta^k G_2^k P_{t-1} G_1^{-k}, \quad R_{1,t}(k) = \beta_1^{-k} G_1^k C_{1,t-1}(G_1^T)^k \text{ and } R_{2,t}(k) = \beta_2^{-k} G_2^k C_{2,t}(G_1^T)^k.$$

In practice, G_2 can be selected so that $G_2^k \rightarrow 0$ for large values of k so that the limiting forecast function

$$F_t(k) = \hat{y}_{2,t+k} = F_1 G_1^k m_{1,t-1}. \tag{26}$$

3. LIMITING RESULTS

Given that, the eigenvalues of $\beta_1 G_1$ and $\beta_2 G_2$ are outside the unit circle, it can be seen that (Ameen and Harrison, 1985)

$$[C_{i,t}, R_{i,t}, A_{i,t}, Q_{i,t}] \rightarrow [C_i, R_i, A_i, Q_i], i=1,2 \tag{27}$$

Given a posterior to prior transition for P_{t-1} so that $P^*_{t-1} = \delta P_{t-1}$ where $\delta \leq 1$ is a known constant, we have Equation (28).

$$P_t = \delta^t (I - A_2 F_2)^t P_0 - \left(\sum_0^{t-1} \delta^i (I - A_2 F_2)^i \right) A_2 F_1 \tag{28}$$

$$= -(I - \delta^t (I - A_2 F_2)^t)(I - \delta(I - A_2 F_2))^{-1} A_2 F_1 \tag{29}$$

$$\rightarrow -[(1-\delta)I + \delta A_2 F_2]^{-1} A_2 F_1 \tag{30}$$

$$= -(I - \frac{\delta A_2 F_2}{\delta F_2 A_2^T + 1 - \delta}) A_2^T F_1 / (1 - \delta) \tag{31}$$

To discuss the limiting predictor functions described here, three simple but rather popular model settings are considered.

3.1. The Steady State M_2 and a Static Regression M_1 Submodel

This restricts the models to scalar forms with $F_1 = F_2 = G_1 = G_2 = \beta_1 = 1$ and δ and β_2 are as before.

Therefore, under M_1,

$$m_{1,t} = \frac{1}{t} \sum_0^{t-1} y_{t-i} = \bar{y}_t, \tag{32}$$

And,

$$A_2 = 1 - \beta_2, P = -\frac{1 - \beta_2}{1 - \delta \beta_2}, \tag{33}$$

It can also be seen which gives

$$m_{2,t} = \beta_2 m_{2,t-1} + (1 - \beta_2)(y_t - m_{1,t-1}) - \frac{1 - \beta_2}{(1 - \delta \beta_2)t}(y_t - m_{1,t-1}) \tag{34}$$

And, as t increases,

$$m_{2,t} = \beta_2 m_{2,t-1} + (1 - \beta_2)(y_t - m_{1,t-1}) \tag{35}$$

That is, the short-term departures are the exponentially weighted moving average (EWMA) of the past departures of the long-term predictions from the observed values.

The k-steps forecasting function is as follows:

$$F_t(k) = \hat{y}_{t+k} + m_{2,t+k} = \bar{y}_t + \beta_2^k m_{2,t} \tag{36}$$

Clearly, as k increases, the k-steps forecast function approaches the mean value of Y at time t.

3.2. The Steady State M_1 and a Steady State M_2 Submodel

This restricts the models to scalar forms with $F_1 = F_2 = G_1 = G_2 = 1$, and β_1, δ, and β_2 are as before.

Under this formulation, $A_i = 1 - \beta_i$, $i = 1, 2$ and $P = -(1 - \beta_2)/(1 - \delta \beta_2)$.

Furthermore, under M_1, the low-frequency state parameter is the EWMA of the past observations:

$$m_{1,t} = (1 - \beta_1) \sum_0^{t-1} \beta_1^i y_{t-i} + \beta_1^t m_{1,0} \tag{37}$$

While the high-frequency parameter estimate is

$$m_{2,t} = \beta_2 m_{2,t-1} + (1 - \beta_2) \frac{\beta_1 - \delta \beta_2}{1 - \delta \beta_2}(y_t - m_{1,t-1}). \tag{38}$$

The k-steps forecasting function is:

$$F_t(k) = m_{1,t+k} + m_{2,t+k} = m_{1,t} + \beta_2^k m_{2,t} \tag{39}$$

$$F_t(k) \rightarrow m_{1,t} \quad as \quad k \rightarrow \infty \tag{40}$$

3.3. Linear Growth M_1 and a Steady State M_2 Submodel
This defines β_1, β_2, G_2, and F_2 as in example (i) and

$$G_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, F_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}, \tag{41}$$

$$m_{1,t} = G_1 m_{1,t-1} + A e_{1,t}. \tag{42}$$

With the limiting adaptive coefficients are as follows:

$$A_1^T = [(1 - \beta_1^2), (1 - \beta_1)^2], A_2 = 1 - \beta_2, \tag{43}$$

And, also using the relation

$$P_t = \lambda P_{t-1} G_1^{-1} - (1 - \beta_2) F_1 \tag{44}$$

Where $\lambda = \beta_2 \delta$, the limiting P_t can be found as

$$P = \frac{1-\beta_2}{1-\lambda}\left[-1 \quad \frac{1+\lambda}{1-\lambda}\right] \qquad (45)$$

$$m_{2,t} = m_{2,t-1} + (1-\beta_2)e_{2,t} + Be_{1,t} \qquad (46)$$

The limiting short-term systems parameter is then updated according to the departures from the estimated trend under M_1.

Or

$$m_{2,t} = \beta_2 m_{2,t-1} + (1+B)e_{1,t} \qquad (47)$$

Where

$$B = \frac{(1-\beta_2)(1-\beta_1)}{1-\lambda}\left(\frac{(1+\lambda)(1-\beta_1)}{1-\lambda} - \beta_1 - 1\right) \qquad (48)$$

The k-steps forecasting function is:

$$F_t(k) = F_1 G_1^k m_{1,t} + m_{2,t+k} \rightarrow F_1 G_1^k m_{1,t} \quad as \quad k \rightarrow \infty \qquad (49)$$

Furthermore, with a diminishing prior-posterior regression coefficients matrix P_t such that

$$P_t = \lambda P_{t-1} - (1-\beta_2)F_1 \qquad (50)$$

$$P_t = -\frac{1-\beta_2}{1-\lambda}F_1 \qquad (51)$$

Leading to a modification in the above recurrence relation with

$$B = -\frac{(1-\beta_2)(1-\beta_1^2)}{1-\lambda} \qquad (52)$$

REFERENCES

Al-Madfai, H. (2002). Weather corrected electricity demand forecasting. Unpublished PhD Thesis, University of Glamorgan (Currently South Wales).

Al-Madfai, H., Ameen, J. R. M., & Ryley, A. (2004). *The Hierarchical Profiling Approach to STLF of Multi-Year Daily Electricity Demand in South Wales*. Sydney, Australia: 24th International Symposium on Forecasting.

Ameen, J. R. M. (1992). Nonlinear predictor models. *Journal of Forecasting,* 11, 309-324.

Ameen, J. R. M., & Harrison, P. J. (1985). Normal discount bayesian models. In: Bernardo, J. M., editor. *Bayesian Statistics 2.* Amsterdam: North Holland.

Anderson, B. D. O., & Moor, J. B. (1979). *Optimal Filtering.* Englewood Cliffs, NJ: Prentice-Hill.

Armstrong, J. S. (1985). *Long-Range Forecasting from Crystal Ball to Computer.* 2nd ed. New York: Wiley.

Bhansali, R. J. (1996). Assymptotically efficient autoregressive model selection for multistep prediction. *Annals of the Institute of Statistical Mathematics,* 48(3), 577-602.

Bhansali, R. J. (1999) Autoregressive model selection for multistep prediction. *Journal of Statistical Planning and Inference*, 78(1), 295-305.

Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control.* San Francisco: Holden Day.

Danesi, A., Mantecchini, L., & Paganelli, F. (2017). Long-term and short-term forecasting techniques for regional airport planning. *ARPN Journal of Engineering and Applied Sciences*, 12(2), 739-745.

Findley, D. F. (1983). On the use of multiple models for multiperiod forecasting. *In: Proceedings of the Business and Economic Statistics Section.* Washington, D. C.: American Statistical Association. p528-531.

Gersch, W., & Katigawa, G. (1983). The prediction of a time series with trends and seasonalities. *Journal Business and Economic Statistics,* 1, 253-264.

Gregg, J. V., Hossell, C. H., & Richardson, J. T. (1964). *Mathematical Trend Curves: An Aid to Forecasting.* London: Oliver and Boyed.

Harrison, P. J., & Akram, M. (1983). Generalized exponentially weighted regression and parsimonious dynamic linear modelling. In: Anderson, O. D., editor. *Time Series Analysis: Theory and Practice.* 3rd ed. Amsterdam: North Holland. pp. 19-50.

Harrison, P. J., & Pearce, S. F. (1972). The use of trend curves as an aid to market forecasting. *Industrial Marketing Management*, 2, 149-170.

Harrison, P. J., & Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society: Series B,* 38, 205-247.

Harrison, P. J., Leonard, T., & Gazard, T. N. (1977). An application of multivariate hierarchical forecasting. Research Report 15, Department of Statistics, University of Warwick.

Harvey, A. C. (1984), A unified view of statistical forecasting procedures. *Journal of Forecasting*, 3, 245-275.

Kabaila, P. V. (1981). Estimation based on one step Ahead prediction versus estimation based on multi-step ahead prediction. *Stochastics,* 6, 43-55.

Kang, I. B. (2003). Multi-period forecasting using different models for different horizons: An application to U.S. economic time series data. *International Journal of Forecasting*, 19(3), 387-400.

Kendall, M., & Ord, J. K. (1993). *Time Series.* 3rd ed. England: Edward Arnold.

Kwon, H. J., Cho, S. H., Chun, Y., Lagarde, F., & Pershagen, G. (2002). Effects of Asian dust events on daily mortality in Seoul. *Environmental Research*, 90(1), 1-5.

Levenbach, H., & Reuter, B. E. (1976). Forecasting trending time series with relative growth rate models. *Technometrics*, 18(3), 216-272.

Lin, J. L., & Granger, C. W. J. (1994). Forecasting from nonlinear models in practice. *Journal of Forecasting,* 13, 1-9.

Martino, J. P. (1983). *Technological Forecasting for Decision Making.* New York and Amsterdam: North Holland.

Migon, H., & Harrison, P. J. (1985). An application of nonlinear baysian forecasting to television advertising. In: Bernardo, J. M., editor. *Bayesian Statistics* 2. `Amsterdam: North Holland.

Niu, W. Y., & Harris, W. M. (1996). China: The forecast of its environmental situation in the 21st century. *Journal Environmental Management,* 47(2), 101-114.

Pillai, S. U. (1992). A new spectrum extension method that maximizes the multistep minimum prediction error-generalization of the maximum entropy concept. *IEEE Transactions on Signal Processing,* 40, 142-158.

Smith, S. A., Micentyre, S. H., & Achabal, D. D. (1994). A two-stage sales forecasting procedure using discounted least square. *Journal of Marketing Research,* 31, 44-56.

Stoica, P., & Soderstorm, T. (1984). Uniqueness of estimated k-step prediction models of ARMA processes. *System and Control Letters,* 4, 325-331.

Tiao, G. C., & Xu, D. (1993). Rubustness of MLE for multistep predictions: The exponential smoothing case. *Biometrica,* 80, 623-641.