ISSN 1032-3813
ISBN 0 7326 0765 5

MONASH UNIVERSITY

ANCORA IMPARO

AUSTRALIA

ESTIMATION AND PREDICTION FOR A CLASS OF DYNAMIC NONLINEAR STATISTICAL MODELS

J.K. Ord
A. Koehler
R.D. Snyder

Working Paper 4/95
April 1995

DEPARTMENT OF ECONOMETRICS

Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models

J. K. Ord
Department of Management Science
The Pennsylvania State University

A. Koehler
Department of Decision Science
Miami University

R. D. Snyder
Department of Econometrics
Monash University

Abstract

A class of dynamic, nonlinear, statistical models is introduced for the analysis of univariate time series. A distinguishing feature of the models is their reliance on only one primary source of randomness: a sequence of independent and identically distributed normal disturbances. It is established that the models are conditionally Gaussian. This fact is used to define a conditional maximum likelihood method of estimation and prediction.

A particular member of the class is shown to provide the statistical foundations for the multiplicative Holt-Winters method of forecasting. This knowledge is exploited to provide methods for computing prediction intervals to accompany the more usual point predictions obtained from the Holt-Winters method. The methods of estimation and prediction are evaluated by simulation. They are also illustrated with an application to Canadian retail sales.

Keywords Maximum likelihood estimation, forecasting, Holt-Winters method, nonlinear state space models, bootstrap method.

Correspondence
Associate Professor R. D. Snyder, Department of Econometrics, Monash University, Clayton, Victoria 3168, Australia. e-mail: Ralph.Snyder@BusEco.Monash.Edu.Au

1. INTRODUCTION

The fact that business and economic time series often possess seasonal cycles with increasing amplitudes together with random fluctuations dependent on movements in underlying levels is a compelling reason to study nonlinear statistical models. Most nonlinear behaviour commonly observed in time series plots can be represented by nonlinear state space models. Being dynamic, the latter have the capacity to reflect many forms of inter temporal dependency encountered in economic time series. Like their linear counterparts (Ansley & Kohn,1985; Harvey and Todd, 1983), nonlinear state space models can be adapted to accommodate the non stationary features of most economic time series.

Dynamic nonlinear models are traditionally presented with more than one disturbance source in their measurement and transition equations. They are usually estimated using quasi-maximum likelihood methods in conjunction with an extended Kalman filter - for example, see Harvey (1990, p160). In contrast, a class of nonlinear dynamic models is introduced in this paper with only one primary disturbance source. We exploit this special structural feature to develop an inherently simpler estimation procedure based on regression methods rather than the more complex extended Kalman filter.

The paper is the outcome of research originally directed to uncovering the statistical foundations of the nonlinear version of the Holt-Winters method of forecasting (Winters, 1960), a method originally designed to handle time series displaying level dependent seasonal and irregular components. The aim was to employ such a model in the generation of prediction intervals to accompany the usual point predictions which emerge from the use of the Holt-Winters method. Although an ARIMA model has been found for the additive version, previous attempts to find the model for the nonlinear Holt-Winters method have not been fully successful (McKenzie, 1984;

Abraham & Ledolter, 1986). It is not possible to find an ARIMA model that implies the same forecast function as the same updating equations as those used in the Holt-Winters multiplicative method (Abraham & Ledolter, 1986). The search for procedures to construct statistically correct prediction intervals, including ones for the nonlinear Holt-Winters method, has been thoroughly documented by Chatfield (1993). Even the most recent procedure for the nonlinear Holt-Winters method (Chatfield & Yar, 1991) leaves one with imprecise rules for choosing the variance of the forecast errors.

Our work directed to eliminating this gap eventually succeeded with the discovery of a model which underpins the nonlinear Holt-Winters method. In our attempt to simplify the associated mathematics, we also uncovered a generalisation that forms the dynamic, nonlinear framework of this paper. Both models are introduced in the next section. The associated theory of estimation and prediction is presented in section 3. The results of a simulation study on the special case of the Holt-Winters method to guage the statistical properties of the resultant estimates and predictions is given in section 4. Finally, in section 5, we illustrate the methodology in an application involving the prediction of Canada's total retail sales.

2. THE FRAMEWORK

We adopt the convention, illustrated in Figure 1, that typical period t ends at point of time t. Thus the start of period t occurs at point of time period $t-1$. Whatever happens to the underlying process during period t is assumed to depend on the state of the system at the beginning of period t, denoted by the random p-vector $\mathbf{x}_{t-1}$. It is also assumed to depend on a disturbance e_t from an $NID(0,\sigma^2)$ time series. More specifically, it is envisaged that the outcome y_t of the process in period t is governed by a measurement equation

$$y_t = \mathrm{h}_t(\mathbf{x}_{t-1}) + \mathrm{k}_t(\mathbf{x}_{t-1})e_t, \qquad (2.1)$$

where h_t and k_t are known continuous functions with continuous derivatives mapping from $\Re^p \to \Re$. Transitions of state are governed by the recurrence relationship

$$\mathbf{x}_t = \mathrm{f}_t(\mathbf{x}_{t-1}) + \mathrm{g}_t(\mathbf{x}_{t-1})e_t, \qquad (2.2)$$

where f_t and g_t are known continuous mappings with continuous derivatives from $\Re^p \to \Re^p$. A distinguishing feature of this univariate framework is that both measurement and transition relationships depend on only one primary source of randomness: the e_t. It is assumed that e_t is independent of $\{y_{t-i}, \mathbf{x}_{t-i}, i \geq 1\}$. Linear models with a single random error have been developed systematically by Snyder(1985), as illustrated in the following example.

Example 2.1

When all the functions are time invariant and independent of $\mathbf{x}_{t-1}$, the disturbances are homoscedastic (ie $k_t(\mathbf{x}_{t-1}) = 1$), and the framework reduces to

$$y_t = \mathbf{h}'\mathbf{x}_{t-1} + e_t \qquad (2.3)$$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \alpha e_t \qquad (2.4)$$

h being a fixed p-vector, F a fixed $p \times p$ transition matrix and α a fixed p-vector of parameters. This framework mirrors the most general integrated form of the forecasting functions associated with ARMA and ARIMA processes (Box and Jenkins, 1976). It possesses close connections to the most general linear form of exponential smoothing (Box and Jenkins, 1976). For example, the integrated form of the ARIMA(0,2,2) model is

$$y_t = \ell_{t-1} + b_{t-1} + e_t$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1 e_t$$

$$b_t = b_{t-1} + \alpha_2 e_t\ .$$

Here the state variables can be given a structural interpretation, ℓ_t being an underlying level and b_t an underlying growth rate. Both level and rate drift over time in response to unanticipated structural changes, the magnitude of the drift being governed by the α vector. It is, in fact, a model underpinning trend corrected exponential smoothing (Holt,1957)

Example 2.2

An extension of the linear framework to accommodate heteroscedastic disturbances is

$$y_t = \mathbf{h}'\mathbf{x}_{t-1} + k_t(\mathbf{x}_{t-1})e_t \qquad (2.5)$$

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \mathrm{a}k_t(\mathbf{x}_{t-1})e_t. \qquad (2.6)$$

This is the most general integrated Box-Jenkins model with heteroscedastic disturbances. A heteroscedastic form of the ARIMA(0,2,2), for example, is

$$y_t = \ell_{t-1} + b_{t-1} + (\ell_{t-1} + b_{t-1})^{\gamma} e_t \qquad \text{(measurement equation)}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1(\ell_{t-1} + b_{t-1})^{\gamma} e_t \qquad \text{(level equation)}$$

$$b_t = b_{t-1} + \alpha_2(\ell_{t-1} + b_{t-1})^{\gamma} e_t, \qquad \text{(rate equation)}$$

γ being a parameter determining the magnitude of the heteroscedasticity. An advantage of the integrated form of the ARIMA framework is that it not only explicitly shows the structural features of a time series hidden by the difference equation form, but it enables us to model the commonly occurring form of heteroscedastity where the disturbances of a process depend on its underlying level. Heteroscedasticity is imbedded directly into the model and the parameter γ can be estimated simultaneously with other model parameters using a single estimation procedure. This scheme contrasts with the conventional approach where Box-Cox transformations are relied upon to stabilise the variance before estimation. It is assumed in the Box-Cox transform approach that the model is linear in the transformed variables. The approach outlined here, however, retains the structure of the original variables, like ARCH models. Thus it is preferable because we can model the mean level and variance structures separately if needed (even though we do not do that for the Holt-Winters case).

Example 2.3

Level dependent seasonal indexes c_t may be incorporated into a model with relative errors as follows:

$$y_t = (\ell_{t-1} + b_{t-1})c_{t-m}(1 + e_t) \qquad \text{(measurement equation)} \qquad (2.5)$$

$\ell_t = (\ell_{t-1} + b_{t-1})(1+\alpha_1 e_t)$ (level equation) (2.6)

$b_t = b_{t-1} + (\ell_{t-1} + b_{t-1})\alpha_2 e_t$ (rate equation) (2.7)

$c_t = c_{t-m}(1+\alpha_3 e_t)$, (seasonal index equation) (2.8)

m being the number of seasons in a year. It is also assumed that the seed seasonal indexes satisfy the normalising condition

$$\sum_{t=1-m}^{0} c_t = m.$$

Although the equations in this model are not first-order Markovian, they may be converted to this form with appropriate extensions to state variables - for example see Harvey (1990) for the details in the context of linear models.

Example 2.4

A simple seasonal model, belonging to the framework after an appropriate definition of the state vector, is the following generalisation of a random walk:

$$y_t = y_{t-m}(1+e_t). \quad (2.9)$$

This forms a convient benchmark against which to compare the performance of other models in any forecasting exercise.

3. ESTIMATION, PREDICTION AND COMPUTATION

3.1. Point Estimates and Predictions

Consider the general framework (2.1)-(2.2) in period t after observing the sample $Y_{t-1} = \{y_1 \cdots y_{t-1}\}$. The following theorem is crucial to the method of estimation to be outlined below.

Theorem 1

Given the seed state vector $\mathbf{x}_0$ and given the sample $Y_n = \{y_1 \ldots y_n\}$, the state vectors $\mathbf{x}_1 \ldots \mathbf{x}_n$ are all fixed and may calculated with the recurrence relationship

$$\mathbf{x}_t = \mathrm{f}_t(\mathbf{x}_{t-1}) + \mathrm{g}_t(\mathbf{x}_{t-1})\left(\frac{y_t - \mathrm{h}_t(\mathbf{x}_{t-1})}{k_t(\mathbf{x}_{t-1})}\right) \quad (t = 1\ldots n). \qquad (3.1)$$

This recurrence relationship has a closed form solution which depends on the seed state vector $\mathbf{x}_0$ and the sample $Y_n = \{y_1 \ldots y_n\}$. The closed form solution is written accordingly as

$$\mathbf{x}_n = \mathrm{F}_n(\mathbf{x}_0, Y_{n-1}). \qquad (3.2)$$

Proof

When $t = 1$ and $\mathbf{x}_0$ and y_1 are given, equation (2.1) may be solved for a fixed value of e_1. The result may be substituted into (2.2) to obtain a fixed value of $\mathbf{x}_1$. This process may be repeated for $t = 2, \ldots, n$. Equation (3.2) follows from the successive application of (3.1).

Example 3.1

The recursions for Example 2.1 reduce to

$$\mathbf{x}_t = \mathbf{F}\mathbf{x}_{t-1} + \alpha(y_t - \mathbf{h}'\mathbf{x}_{t-1}) \qquad (3.3)$$

This is the most general linear, error correction form of exponential smoothing (Box and Jenkins, 1976). It includes simple exponential smoothing (Brown, 1959), trend corrected exponential smoothing (Holt, 1957) and the extension incorporating additive seasonal effects. The closed form solution of (3.3) is

$$\mathbf{x}_t = \mathbf{D}^t\mathbf{x}_0 + \sum_{j=0}^{t-1} \mathbf{D}^j \alpha y_{t-j} \qquad (3.4)$$

where

$$\mathbf{D} = \mathbf{F} - \alpha\mathbf{h}'. \qquad (3.5)$$

The equation (3.4) is a generalisation of the exponentially weighted average associated with simple exponential smoothing. $\mathbf{D}$ is the matrix generalisation of the discount factor. Provided its characteristic roots all lie within the unit circle the influence of the past behaviour of a time series on the current state of the process declines with age. This is an alternative form of the classical invertibility conditions; see Appendix B.

Example 3.2

The recurrence relationship (3.1) also reduces to (3.3) in Example 2.2. This result indicates that the error correction formulae for exponential smoothing apply without change in the presence of heteroscedastic disturbances. Thus, in this sense, exponential smoothing is a more general methodology than the Box-Jenkins approach. This finding may explain, in part, why the exponential smoothing methods perform so well in forecasting competitions (Makridakis et. al., 1982).

Example 3.3

The error correction equations obtained for Example 2.3 are

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1\left(\frac{y_t}{c_{t-m}} - \ell_{t-1} - b_{t-1}\right) \qquad (3.6)$$

$$b_t = b_{t-1} + \alpha_2\left(\frac{y_t}{c_{t-m}} - \ell_{t-1} - b_{t-1}\right) \qquad (3.7)$$

$$c_t = c_{t-m} + \alpha_3\left(\frac{y_t}{\ell_{t-1} + b_{t-1}} - c_{t-m}\right). \qquad (3.8)$$

Using (3.6), and assuming that $\alpha_1 > 0$, the equation (3.7) can be written as

$$b_t = b_{t-1} + \overline{\alpha}_2(\ell_t - \ell_{t-1} - b_{t-1}) \qquad (3.9)$$

where $\overline{\alpha}_2 = \alpha_2/\alpha_1$. The equations (3.6), (3.9) and (3.8) are almost identical to the error correction equations in Holt-Winters (1960). The only difference is that the trend adjustment term in (3.8) is $(\ell_{t-1} + b_{t-1})$ rather than ℓ_t. The practical difference is slight and (3.8) has the benefit that the trend estimate for time t does not depend on y_t. Thus, for the first time, we effectively have in example 2.3, a model underlying the Holt-Winters method.

The equation (3.2) summarises all the information in the transition equations (2.2). Substituting into (3.1) gives a relationship of the form conditional univariate distributions are normal, $(y_1,\ldots,y_n)$ does not follow the usual multivariate normal distribution because of the nonlinear structure of equations (2.1) and (2.2). Note that equation (3.11) is completely general, given that the errors are independent, so that non-normal structures could be specified.

Since normal random variables are unbounded in both directions, model (3.11) is justified only if we are willing to let the state variables range over the whole real line. In applications, y_t is usually non-negative and the level and seasonal state variables would be taken to be non-negative also. These restrictions do not pose a problem in practice as the Holt-Winters scheme would only be used when the process was well away from the origin. The possibility that the process becomes negative, however, poses a real problem for the simulations described in section 4. Our solution was to eliminate from the analysis any runs for which y_t became non-positive.

The functions in the equations (2.1) and (2.2) of the general framework potentially depend on a q-vector θ of parameters in addition to the variance σ^2. To illustrate, in the Example 2.2, θ consists of the four parameters $\alpha_1,\alpha_2,\alpha_3,\gamma$. Hence (3.11) can be used as the basis of a conditional likelihood function given by

$$L(\theta,\mathbf{x}_0|Y_n)=\left(2\pi\sigma^2\right)^{-n/2}\left(\prod_{t=1}^{n}\mathrm{k}_t^*(\mathbf{x}_0,\theta,Y_{t-1})\right)^{-1}\exp\left(-\sum_{t=1}^{n}\frac{\left(y_t-\mathrm{h}_t^*(\mathbf{x}_0,\theta,Y_{t-1})\right)^2}{2\,\mathrm{k}_t^*(\mathbf{x}_0,\theta,Y_{t-1})^2\sigma^2}\right) \quad (3.12)$$

The values of $\mathbf{x}_0$ and θ, denoted by $\hat{\mathbf{x}}_0$ and $\hat{\theta}$, which maximise (3.12) will be called maximum conditional likelihood estimators.

To understand the nature of maximum conditional likelihood estimators, consider the ARMA(1,1) scheme derived from Example 2.1 when $p=1$, $\mathbf{h}=1$ and $\mathbf{F}=f$. The model reduces to:

$$y_t = fy_{t-1}-(f-\alpha)e_{t-1}+e_t .$$

To understand the nature of maximum conditional likelihood estimators, consider the ARMA(1,1) scheme derived from Example 2.1 when $p=1$, $\mathbf{h}=1$ and $\mathbf{F}=f$. The model reduces to:

$$y_t = fy_{t-1} - (f-\alpha)e_{t-1} + e_t.$$

When $|f|<1$, the process is stationary and the marginal density for y_0 may be determined. The conditional maximum likelihood estimator conditions on the value of y_0 but does not make use of the marginal density, whereas the (unconditional) maximum likelihood estimator incorporates the marginal density of the likelihood function. When $f=1$, the process is nonstationary and the marginal density for y_0 is not defined, so we must use a conditional form of the likelihood.

It is assumed that a nonlinear optimisation procedure, such as the Gauss-Newton method, is employed to find those values of $\mathbf{x}_0$ and θ which maximise (3.12). Each stage of the search procedure employed by the optimisation routine conditions on trial values of $\mathbf{x}_0$ and θ. The recurrence relationships (3.1) are employed in conjunction with (3.12) to evaluate the conditional likelihood. In the case of Example 2.3, this is tantamount to embedding the Holt-Winters method in the optimisation routine.

It is not always practicable to implement the estimation procedure exactly as described because of high computational loads. In the case of the Holt-Winters model on monthly time series data, for example, it would be necessary to maximise the conditional likelihood with respect to a total of 17 quantities. Computational loads can be reduced significantly if the seed vector $\mathbf{x}_0$ is chosen using a method that is independent of the parameter vector θ. The Holt-Winters method, for example, is often seeded with quantities obtained with Winters (1960) heuristic; for further details, see Bowerman and O'Connell(1993, pp403-407). This device leaves only the three smoothing parameters $\alpha_1,\alpha_2,\alpha_3$ to be optimised. Traditionally, the optimisation has been done using a sum of squared errors criterion. The model in Example 3.3 clearly indicates, however, that a criterion based on relative errors $\hat{e}_t = (y_t - \hat{y}_t)/\hat{y}_t$ rather than the conventional absolute errors should be used. The conditional likelihood (3.12) has the required orientation in this context.

Once the estimates, denoted by $\hat{\mathbf{x}}_0$ and $\hat{\theta}$, are obtained in this way the standard deviation σ can be estimated with $\hat{\sigma} = \sqrt{\sum_{t=1}^{n} \hat{e}_t^2 \Big/ (n-p)}$, p being the number of elements in both $\mathbf{x}_t$ and θ. Point predictions can be generated with the equations

$$\hat{y}_t = \mathrm{h}_t\left(\hat{\mathbf{x}}_{t-1}, \hat{\theta}\right) \quad (t = n+1, \ldots n+h) \tag{3.12}$$

$$\hat{\mathbf{x}}_t = \mathrm{f}_t\left(\hat{\mathbf{x}}_{t-1}, \hat{\theta}\right) \quad (t = n+1, \ldots, n+h) \tag{3.13}$$

h being the prediction horizon.

3.2 Prediction Intervals

One of our objectives for identifying the model underpinning the Holt-Winters method was to establish a reliable procedure for determining prediction intervals for time series displaying level dependent seasonal and irregular components. The aim, in particular, was to devise a method for obtaining prediction intervals which depend on the seasonal effects, something that is of considerable practical importance (Chatfield,1993). Four strategies are developed and compared.

3.2.1 'Operations Research' Approach

The Holt-Winters method is often used in inventory control to predict aggregate demand over the next h periods (the lead time). The root mean squared error for a single period is estimated by $s = \sqrt{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2 \Big/ n}$. Then the root mean squared error of *total* lead time demand is assumed to be $s\sqrt{h}$. The practice is based on the approximation that the future values of a time series are uncorrelated and that the prediction error does not increase with h. The assumptions are incompatible with the structure of the exponential smoothing methods and the approach leads to serious errors (Johnston & Harrison,1986; Harvey & Snyder,1990) for common *linear* forms of exponential smoothing. We examine this albeit flawed, approach, because of its widespread use in practice.

3.2.2 Linear Approximation Method (LAM)

It is tempting to base estimates of the root mean squared errors on the Hessian of the conditional log-likelihood function. The conditional log-likelihood function, however, is not optimised with respect to the seed factors because they are determined by Winter's seeding heuristic. The Hessian cannot be guaranteed, in these circumstanes, to be positive semi-definite and so this strategy is not viable.

An alternative approach involves the application of first-order Taylor series expansions to the equations of the nonlinear model (2.1) and (2.2) with respect to the vector $\begin{bmatrix}\mathbf{x}_{t-1} & e_t\end{bmatrix}$ about the point formed from the conditional maximum likelihood estimates $\begin{bmatrix}\hat{\mathbf{x}}_{t-1} & \hat{e}_t\end{bmatrix}$ for given $\hat{\theta}$. This yields a so-called linear pseudo-model

$$y_t = d_t + \mathbf{h}_t'\mathbf{x}_{t-1} + k_t e_t \qquad (3.15)$$

$$\mathbf{x}_t = \mathbf{a}_t + \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{g}_t e_t . \qquad (3.16)$$

Example 3.4

We illustrate the procedure by revisiting Examples 2.1-2.3.

Example 2.1

Equations (3.15-3.16) are exact, with $\mathbf{a}_t = \mathbf{0}$, $d_t = 0$, $\mathbf{h}_t = \mathbf{1}$, $\mathbf{F}_t = \mathbf{F}$, $k_t = 1$ where

$$\alpha = \begin{bmatrix}\alpha_1 \\ \alpha_2\end{bmatrix}, \ \mathbf{1} = \begin{bmatrix}1 \\ 1\end{bmatrix} \text{ and } \mathbf{F} = \begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix}.$$

Example 2.2

Going as far as first-order terms in both $m_{t-1} = \ell_{t-1} + b_{t-1}$ and e_t we find, after some reduction, that $d_t = -\gamma \hat{m}_{t-1}^{\gamma} \hat{e}_t$, $\mathbf{a}_t = \alpha d_t$, $\mathbf{h}_t = (1 - d_t)\mathbf{1}$, $\mathbf{F}_t = \mathbf{F} - d_t \alpha \mathbf{1}'$, $k_t = \hat{m}_{t-1}^{\gamma}$ and $\mathbf{g}_t = \alpha k_t$.

Example 2.3

The equations for the Holt-Winters model may also be written in the form of (3.15-3.16), but is more instructive to write them down directly. The derivation for (3.15) is as follows. Write the observation equation (2.7) as

$$y_t = \left[\hat{m}_{t-1} + (m_{t-1} - \hat{m}_{t-1})\right]\left[\hat{c}_{t-1} + (c_{t-1} - \hat{c}_{t-1})\right]\left[1 + \hat{e}_{t-1} + (e_{t-1} - \hat{e}_{t-1})\right] \qquad (3.17)$$

Collecting terms as far as the first order, we obtain

$$y_t = -\hat{m}_{t-1}\hat{c}_{t-m}(1 + 2\hat{e}_t) + m_{t-1}\hat{c}_{t-m}(1 + \hat{e}_t) + c_{t-m}\hat{m}_{t-1}(1 + \hat{e}_t) + \hat{m}_{t-1}\hat{c}_{t-m}e_t .$$

Likewise

$$\ell_t = -\alpha_1 \hat{m}_{t-1}\hat{e}_t + m_{t-1}(1 + \alpha_1 \hat{e}_t) + \alpha_1 \hat{m}_{t-1} e_t$$

$$b_t = -\alpha_2 \hat{m}_{t-1}\hat{e}_t + b_{t-1} + m_{t-1}(1 + \alpha_2 \hat{e}_t) + \alpha_2 \hat{m}_{t-1} e_t$$

$$c_t = -\alpha_3 \hat{c}_{t-m}\hat{e}_t + c_{t-m} + c_{t-m}(1 + \alpha_3 \hat{e}_t) + \alpha_3 \hat{c}_{t-m} e_t .$$

As expected, the leading term in the zero'th order Taylor series expansion in equations like (3.17) corresponds to replacing each term by its estimate; in effect, this is the procedure followed by Winters (1960) and many others from the operations research profession.

If we ignore the prediction errors in the coefficients d_t, $\mathbf{h}_t$, k_t, $\mathbf{a}_t$, $\mathbf{F}_t$, $\mathbf{g}_t$ in equations (3.15-3.16), on the grounds that they will be $0_p(n^{-1/2})$ for long series, we may determine expressions for the j-step ahead prediction mean squared error of y_t., PSME(t,j), as follows. Let $\mathbf{h}_t(j)$ denote the j-step ahead point prediction of $\mathbf{h}_{t+j}$, with similar notation for the other unknowns in (3.15-3.16). Then, from (3.16), the j-step ahead variance for $\mathbf{x}_{t+j}$, $\mathbf{V}(t,j)$ say, is given recursively by

$$\mathbf{V}(t,j) = \mathbf{F}_t(j)\mathbf{V}(t,j-1)\mathbf{F}_t'(j) + \sigma^2 \mathbf{g}_t(j)\mathbf{g}_t'(j),$$

whence

$$PSME(t,j) = \mathbf{h}_t'(j)\mathbf{V}(t,j-1)\mathbf{h}_t(j) + \sigma^2 k_t^2(j);$$

where $\mathbf{V}(t,0) \equiv \mathbf{0}$.

The scalars d_t, k_t, the vectors $\mathbf{h}_t$, $\mathbf{g}_t$ and the matrix $\mathbf{F}_t$ in equations (3.15-3.16) can be generated numerically with computing routines for constructing first-order Taylor approximations of the nonlinear equations. This is inherently simpler than the analytical approach just described.

Using the theory associated with Theorem 3.1 and equation (3.10), this linear system can resolved into a linear 'regression' with the general form

$$y_t^* = \mathbf{h}_t^{*\prime}\mathbf{x}_0 + k_t e_t \,. \tag{3.18}$$

The details may be found in Appendix A. The vector of 'regression' coefficients $\mathbf{x}_0$ is stochastic. Duncan and Horn (1972) indicate, when considering regressions with stochastic coefficients, that the conventional weighted linear least squares solution denoted by $\breve{x}_0$ satisfies a generalised version of the Gauss-Markov theorem . If $\breve{\sigma}$ denotes the estimate of σ obtained using conventional weighted least squares methods and $\mathbf{H}$ is the matrix formed by stacking the row vectors $\mathbf{h}_t^{*\prime}$ the covariance matrix is given

$$E(\mathbf{x}_0 - \breve{\mathbf{x}}_0)(\mathbf{x}_0 - \breve{\mathbf{x}}_0)' = \breve{\sigma}^2(\mathbf{H}'\mathbf{H})^{-1} \,. \tag{3.19}$$

Root mean squared prediction errors, conditional on $\hat{\theta}$, depend on the covariance matrix in (3.19) and can be obtained using the standard methods associated with the conventional theory of linear regression. We adopted a different but equivalent strategy to expedite the development of the associated computer programs. The strategy is based on the observation that the measurement equations associated with the future periods $t = n+1, \ldots n+h$ can be rewritten as

$$0 = d_t - y_t + \mathbf{h}_t'\mathbf{x}_{t-1} + k_t e_t \,. \tag{3.20}$$

The state vectors were augmented to include the unknown future series values and the pseudo-linear model expanded to

$$\tilde{y}_t = d_t + \tilde{\mathbf{h}}_t'\tilde{\mathbf{x}}_{t-1} + k_t e_t \tag{3.21}$$

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{a}}_t + \tilde{\mathbf{F}}_t\tilde{\mathbf{x}}_{t-1} + \tilde{\mathbf{g}}_t e_t \,. \tag{3.22}$$

where

$$\tilde{y}_t = \begin{cases} y_t & \text{for } t = 1,2,\ldots,n \\ 0 & \text{for } t = n+1,\ldots,n+h \end{cases}$$

$$\tilde{\mathbf{h}}'_t = \begin{cases} [0 \quad \mathbf{h}'_t] & \text{if } t = 1,2,\ldots,n \\ [-\delta_{t-n} \quad \mathbf{h}'_t] & \text{if } t = n+1,\ldots,n+h \end{cases}$$

$$\tilde{\mathbf{x}}'_t = [y_{n+1} \cdots y_{n+h} \quad \mathbf{x}'_t]$$

$$\tilde{\mathbf{a}}_t = \begin{bmatrix} \mathbf{0} \\ \mathbf{a}_t \end{bmatrix} \quad \tilde{\mathbf{g}}_t = \begin{bmatrix} \mathbf{0} \\ \mathbf{g}_t \end{bmatrix} \quad \tilde{\mathbf{F}}(t) = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}(t) \end{bmatrix}$$

δ_i being a unit h-vector with the i^{th} element equal to 1. This enlarged model is converted to its equivalent regression form (3.18). Then conventional weighted least squares methods are used to estimate the enlarged seed state vector conditional on $\hat{\theta}$. The top h elements of the error vector $\tilde{\mathbf{x}} - \breve{\tilde{\mathbf{x}}}$ correspond to the unknown prediction errors $y_{n+i} - \hat{y}_n(i)$ for i steps ahead from origin n. The first h diagonal elements of the enlarged covariance matrix (3.19) therefore conveniently correspond to the required prediction mean squared errors. Because of the conditioning on $\hat{\theta}$, however, these results ignore the effect of errors associated with the estimation of θ, as do the interval estimates for ARIMA models.

3.2.3 Bootstrap Method

Bootrapping is another alternative for obtaining estimation and prediction intervals. This technique, devised by Efron (1979, 1982), has been applied in a time series context to ARIMA processes, the most recent example being McCullough (1994). Its use for exponential smoothing remains largely unexplored.

The conventional bootstrap approach, when applied to the model (2.1) and (2.2), involves the following steps:

1. The estimation of $\mathbf{x}_0$, α and the disturbances $e_1,\ldots,e_n$ using the method outlined in section 3.1, the results being denoted by $\mathbf{x}_0^{(0)}$, $\alpha^{(0)}$ and $e_1^{(0)},\ldots,e_n^{(0)}$.
2. The repetition of the following steps for trials $\tau = 1,\ldots,T$

where k is the number of quantities being estimated in the state vector $\mathbf{x}_0$ and the parameter vector α.

(e) The recursive generation for lead times $j = 1,\ldots,h$ of the time series predictions, state vector estimates and the prediction errors respectively using the equations:

$$y_n^{(\tau)}(j) = h_{n+j}\left(\mathbf{x}_n^{(\tau)}(j-1)\right)$$
$$\mathbf{x}_n^{(\tau)}(j) = f_{n+j}\left(\mathbf{x}_n^{(\tau)}(j-1)\right)$$
$$e_n^{(\tau)}(j) = \left(y_{n+j}^{(\tau)} - y_n^{(\tau)}(j)\right)/y_n^{(\tau)}(j)$$

where $\mathbf{x}_n^{(\tau)}(0) = \mathbf{x}_n^{(\tau)}$.

3. The selection of the prediction intervals for lead times $j = 1,\ldots,h$ as follows:

   (a) the ranking of each sample $e_n^{(1)}(j),\ldots,e_n^{(T)}(j)$ to give $e_n^{[1]}(j),\ldots,e_n^{[T]}(j)$ where $e_n^{[1]}(j) \le e_n^{[2]}(j),\ldots,e_n^{[T-1]}(j) \le e_n^{[T]}(j)$.

   (b). the selection of the lower and upper limits of the prediction intervals using $L_j = e_n^{[\tau_1]}(j)$ and $U_j = e_n^{[\tau_2]}(j)$ where $\tau_1 = [(1-p_r)T/2]$ and $\tau_2 = [(1+p_r)T/2]$, p_r being the nominal interval probability.

3.2.4 Semi Parametric Method

A semi parametric method for interval estimation represents a fourth possibility. It is similar to the bootstrap method. Letting

$$s_n^{(0)} = \sqrt{\frac{\sum_{t=1}^{n}\left(e_t^{(0)}\right)^2}{n-k}}$$

denote the standard deviation of the errors generated in step 1, instead of step 2a we generate the sample of disturbances $e_1^{[\tau]},\ldots,e_n^{[\tau]}$ from a normal distribution with mean 0 and standard deviation $s_n^{(0)}$. The rest of the algorithm remains unchanged.

4. SIMULATION STUDY OF HOLT-WINTERS METHOD

4.1. General Simulation Strategy

We undertook simulation studies on an 80486 desktop computer to gauge the effectiveness of the above methods. All programs were written in Microsoft Fortran 5.1, the final results being collated with Microsoft Excel 5. Each study consisted of 1000 repetitions of an experiment. The typical experiment κ began with the generation of the disturbances using the normal pseudo-random number generator RNOR from the NSWC Fortran subroutine library (Morris, 1993). The formulae in Example 2.3 were then used with the parameter and seed values shown in the "Actual Value" column in Table 1 to compute a synthetic quarterly time series $y_1^{(\kappa)}, y_2^{(\kappa)}, \ldots y_n^{(\kappa)}, y_{n+1}^{(\kappa)}, y_{n+2}^{(\kappa)}, \ldots, y_{n+h}^{(\kappa)}$. Time series with non-positive values were discarded in accordance with our discussion in section 3.1. Estimates of the seed values, denoted by $\hat{\mathbf{x}}_0^{(\kappa)}$, were obtained using the Holt-Winters heuristic (Bowerman and O'Connell, 1993) applied to only the first three years of the series.

Estimates of the smoothing parameters α, denoted by $\hat{\alpha}^{(\kappa)}$, were obtained with the unconstrained nonlinear optimization routine OPTF from the NSWC Fortran subroutine library using the first n values of the synthetic time series. The nonlinear nature of the model precluded us from identifying and imposing stability conditions on α. It was found, however, that the optimization routine could fail to converge without the imposition of nonnegative conditions on the elements of α. This was done in the context of the unconstrained optimizer by optimizing with respect to their square roots. To discourage redundant, distant probes of the parameter space by the optimizer, a penalty of 10^{20} was incorporated into minus the log-likelihood function when the sum of the values of the elements of α exceeded of 6.0. The final optimized values never seemed to reach this bound. The predictions were obtained for periods $n+1, \ldots, n+8$ and compared with the corresponding synthetic series values.

nature of the model precluded us from identifying and imposing stability conditions on α. It was found, however, that the optimisation routine could fail to converge without the imposition of nonnegative conditions on the elements of α. This was done in the context of the uncontrained optimiser by optimising with respect to their square roots. To discourage redundant, distant probes of the parameter space by the optimiser, a penalty of 10^{20} was incorporated into minus the log-likelihood function when the sum of the values of the elements of α exceeded of 6.0. The final optimised values never seemed to reach this bound. The predictions were obtained for periods $n+1,\ldots,n+8$ and compared with the corresponding synthetic series values.

4.2. *Simulation of Point Estimates and Predictions*

The first simulation focussed on the statistical preperties of the point estimates and predictions that emerge from the nonlinear optimisation routine. The results are shown in Table 1 for three values of the sample size n. The average and root mean squared errors of the predictions are defined relative to the averages of the generated series values by the formulae:

$$\frac{\sum_{\kappa=1}^{1000}\left(y_t^{(\kappa)} - \hat{y}_t^{(\kappa)}\right)\Big/1000}{\sum_{\kappa=1}^{1000} y_t^{(\kappa)}\Big/1000} \text{ and } \frac{\sqrt{\sum_{\kappa=1}^{1000}\left(y_t^{(\kappa)} - \hat{y}_t^{(\kappa)}\right)^2\Big/1000}}{\sum_{\kappa=1}^{1000} y_t^{(\kappa)}\Big/1000} \quad (t = n+1,\ldots,n+h).$$

The corresponding formulae for the smoothing parameters are:

$$\frac{\sum_{\kappa=1}^{1000}\left(\alpha_i - \hat{\alpha}_i^{(\kappa)}\right)\Big/1000}{\alpha_i} \text{ and } \frac{\sqrt{\sum_{\kappa=1}^{1000}\left(\alpha_i - \hat{\alpha}_i^{(\kappa)}\right)^2\Big/1000}}{\alpha_i} \quad (i = 1,\ldots,3)$$

while similar formulae apply for the seed estimates. Note that all numbers in the table are presented in relative terms correct to two decimal places. The zeros in the table therefore correspond to numbers that are smaller than 0.005.

A striking feature of the results is that the predictions for y_t are effectively unbiased; further, as expected, the PSME increases with the forecasting horizon. The root mean squared errors also increase with sample size. At first sight, this result is paradoxical. There are two sampling effects, however, to be considered. As the sample size increases, the estimated state values converge to their true values, so that the prediction errors are purely driven by the e_t term at time t. Secondly, in large samples, the risk of the process taking on negative values is increased, as we can see from the numbers of the rejected trials in Table 1. Thus it would appear that larger values of n served only to make the assumptions of normality and non-negativity come more sharply into conflict.

The results for the seed level are reasonable. Given that rates are usually relatively more unstable than levels in economic forecasting, the poor results for the seed rate are not surprising. It should be recalled that this figure is the rate at $t = 0$, so that inaccurate estimates will have little impact on predictions unless n is small. The smoothing parameter estimates are biased but the biases, together with their relative root mean squared errors, decline with increases in the sample size. The seasonal index estimates were surprisingly reliable. Estimates of the standard deviation of the disturbances were also effectively unbiased, becoming more accurate with increases in sample size. Overall, one interesting finding is that the initialisation heuristic recommended by Winters (1960) provides, with the exception of the seed rate, remarkably good results. Optimisation of the seed values appears to be unnecessary from a practical perspective.

4.3 Simulation of Estimation and Prediction Intervals

A simulation study was also undertaken of the four methods for establishing estimation and prediction intervals described in section 3.2. All intervals in the study were based on nominal confidence levels and probabilities of 90 percent. The simulated proportions, shown in Table 2, should consequently be as close as possible to 0.9.

It transpired that the simulated proportions for the traditional operations research approach were quite low. Although it was anticipated that prediction intervals based on this approach would prove to be too narrow, the seriousness of the problem was not expected. This result complements and reinforces the findings of Johnston & Harrison (1986) and Harvey & Snyder (1990). There is, as a consequence, an urgent need to review current practices in sales forecasting for inventory control.

The linear approximation method (LAM), as described in this paper, conditions on the α's. Thus, no estimation intervals could be established for these parameters, a situation reflected by corresponding blank cells in the LAM columns of Tables 2. The simulated proportions display an upward drift with increases in sample size. The prediction intervals appear to become quite reliable in large samples.

The simulation of the bootstrap method yielded disappointing results. Although matters improved with increases in sample size, this method displayed a tendency to produce intervals that are too narrow.

With the exception of the α's, the results for the semi-parametric method were quite plausible. The improved performance of this method over the bootstrap method can be traced to the fact that (a) the simulated time series, as described in section 4.1, are generated from a normal density, and (b) the method uses a normal distribution in place of an empirical distribution of the residuals in its internal simulation. It is unlikely that this performance could be maintained in the presence of disturbances from other distributions without a corresponding adjustment to the distribution used in the internal simulation.

Despite the qualifications, the last three methods for interval predictions were superior to the traditional OR approach. The differences became quite stark with increases in the forecast horizon.

5. AN APPLICATION

Canadian Quarterly Total Retail Sales(source:OECD Main Economic Indicators Database) are plotted from 1960 to 1991 in Figure 2. The time series contains seasonal and irregular components which grow with the underlying level. The change in trend after 1988 suggests the need for a model that allows for structural change. The time series is therefore a suitable candidate for the Holt-Winters method and the extensions outlined in this paper.

To help gauge the performance of the Holt-Winters method, we used the naive, relative seasonal random walk model in Example 2.4 as a benchmark. Predictions from this model, based on a sample of size n, are given by $y_n(im+j) = y_{n+j-m}$. for $i = 0,1,\ldots$ and $j = 1,\ldots,m$. The associated prediction mean squared errors are computed quite simply with $PMSE(n,im+j) = y_n(im+j)i\sigma^2$ for $i = 0,\ldots$ and $j = 1,\ldots,m$.

The available data extended beyond 1991 to 1993. The final two years of data were reserved for evaluating the forecasting capacitites of the models. Only the data for the years 1960 to 1991 were used for the estimation. Predictions for 1992 and 1993 were therefore conditioned on the sample up to 1991. These together with the associated prediction errors are shown in Table 3. On this basis the seasonal random walk model performs best. However, the situation is clouded by the fact that the 90 percent prediction intervals associated with the Holt-Winters approach, obtained by the linear approximation method, are smaller. It is likely that the apparent advantage of the seasonal random walk would be short lived in the longer term. The estimates obtained from the Holt-Winters method are shown in Table 4.

6. CONCLUDING REMARKS

We have, in the paper, outlined the advantages to be gained from considering general integrated forms of ARIMA processes. The integrated form not only facilitates model specification and identification by enabling us to incorporate explicitly structural features that are hidden in the difference equation form. We also have demonstrated that they provide a convenient, flexible basis for heteroscedastic generalisations. We have shown in this more general context that exponential smoothing transcends the usual ARIMA with constant error variance processes, thereby identifying a potential reason for the success of these simple forecasting methods in practice.

The main contribution of this paper has been the discovery of a model underlying the Holt-Winters method. Another framework involving contemporaneously independent multiple disturbance sources $\varepsilon_t, \xi_t, \eta_t, \omega_t$:

$$y_t = (\ell_{t-1} + b_{t-1}) c_{t-m} (1+\varepsilon_t) \qquad \text{(measurement equation)} \qquad (6.1)$$

$$\ell_t = (\ell_{t-1} + b_{t-1})(1+\xi_t) \qquad \text{(level equation)} \qquad (6.2)$$

$$b_t = b_{t-1} + (\ell_{t-1} + b_{t-1})\eta_t \qquad \text{(rate equation)} \qquad (6.3)$$

$$c_t = c_{t-m}(1+\omega_t)\,. \qquad \text{(seasonal index equation)} \qquad (6.4)$$

is another possibility. The extended Kalman filter for the pseudo-linear counterpart of this model seems to converge to a steady state. When it does, it can be shown that the steady state form of the algorithm corresponds to the Holt-Winters method. A preliminary exploration of these ideas may be found in Ord & Koehler (1990). We have excluded this alternative from current consideration because the approach is inherently more complex and involves greater levels of approximation than the single source of error scheme.

It has been demonstrated how the model for the Holt-Winters method enables us to obtain more reliable prediction intervals than methods currently used in practice. It is anticipated that it will also provide the basis for the development of proper diagnostic test procedures in the future. The model therefore presents opportunities for an improvement forecasting practice. Without it such an opportunity would not exist.

We conclude the paper with a possible improvement to the Holt-Winters method. A problem with the method is that the seasonal indexes, as they adapt over time, loose their normalisation property. The traditional response to this problem has been rather ad hoc, users of the method being encouraged to occasionally renormalise the indexes when deemed to be necessary. A better possibility, which ensures that the seasonal indexes automatically remain normalised in terms of conditional means, is to replace the seasonal equation (2.8) by

$$c_t = \left(m - \sum_{j=1}^{m-1} c_{t-j} \right) (1 + \alpha_3 e_t).$$

This modification yields an enhanced model which warrants serious consideration as a replacement for the traditional approach.

APPENDIX A

In this appendix the details of the derivation of (3.18) from (3.15) and (3.16) are presented. The relationship (3.1) for (3.15) and (3.16) specialises to

$$\mathbf{x}_t = \mathbf{a}_t + \mathbf{D}_t \mathbf{x}_{t-1} + \mathbf{g}_t (y_t - d_t)/k_t \qquad \text{(A1)}$$

where

$$\mathbf{D}_t = \mathbf{F}_t - \mathbf{g}_t \mathbf{h}_t'/k_t \qquad \text{(A2)}$$

The closed form solution of (A1) corresponding to (3.2) has the general linear form

$$\mathbf{x}_t = \mathbf{P}_t \mathbf{x}_0 + \mathbf{z}_t \qquad \text{(A3)}$$

where

$$\mathbf{P}_t = \mathbf{D}_t \mathbf{P}_{t-1}, \qquad \text{(A4)}$$

the matrix $\mathbf{P}_0 = \mathbf{I}$, the vector

$$\mathbf{z}_t = \mathbf{a}_t + \mathbf{F}_t \mathbf{z}_{t-1} + \mathbf{g}_t \, y_t^*/k_t \qquad \text{(A5)}$$

and

$$y_t^* = y_t - d_t - \mathbf{h}_t' \mathbf{z}_{t-1}. \qquad \text{(A6)}$$

That (A3) is the solution to (A1) can be seen by induction. Substitution of a lagged (A3) into (3.15) gives (3.18) where

$$\mathbf{h}_t^{*\prime} = \mathbf{h}_t' \mathbf{P}_{t-1}. \qquad \text{(A7)}$$

APPENDIX B: STABILITY CONDITIONS

Using the backward shift operator, **B**, we may solve equation (3.3) for $\mathbf{x}_t$ to give $\mathbf{x}_t = (\mathbf{I}-\mathbf{DB})^{-1}\alpha y_t$. This may be substituted into (2.3) to give

$$\left[\mathbf{I}-\mathbf{h}'\mathbf{B}(\mathbf{I}-\mathbf{DB})^{-1}\right]y_t = e_t \qquad \text{(B1)}$$

which is the ARIMA representation.

For example, the level-only model has $\mathbf{D}=1-\alpha$, $h=1$ (scalars) and (B1) reduces to the ARIMA(0,1,1) scheme. The stability condition is $|1-\alpha|<1$ or $0<\alpha<2$. The Holt (level-plus-rate) scheme has

$$\mathbf{D}=\begin{bmatrix}1-\alpha_1 & 1-\alpha_1\\ -\alpha_2 & 1-\alpha_2\end{bmatrix},\ \mathbf{h}=\begin{bmatrix}1\\1\end{bmatrix}$$

and (B1) reduces to the ARIMA(0,2,2) scheme. The stability conditions are $\alpha_1>0$, $\alpha_2>0$ and $2\alpha_1+\alpha_2<4$.

APPENDIX C: ROOT MEAN SQUARED PREDICTION ERRORS

If we ignore the prediction errors in the coefficients d_t, $\mathbf{h}_t$, k_t, $\mathbf{a}_t$, $\mathbf{F}_t$, $\mathbf{g}_t$ in equations (3.15-3.16), on the grounds that they will be $0_p(n^{-1/2})$ for long series, we may determine expressions for the j-step ahead prediction mean squared error of y_n, PSME(n,j), as follows. Let $\mathbf{h}_n(j)$ denote the j-step ahead point prediction of $\mathbf{h}_{n+j}$, with similar notation for the other unknowns in (3.15-3.16). Then, from (3.16), the j-step ahead variance for $\mathbf{x}_{n+j}$, $\mathbf{V}(n,j)$ say, is given recursively by

$$\mathbf{V}(n,j) = \mathbf{F}_n(j)\mathbf{V}(n,j-1)\mathbf{F}_n'(j) + \sigma^2\mathbf{g}_n(j)\mathbf{g}_n'(j),$$

whence

$$PSME(n,j) = \mathbf{h}_n'(j)\mathbf{V}(n,j-1)\mathbf{h}_n(j) + \sigma^2 k_n^2(j).$$

If we compute the PMSE conditionally upon $\hat{\theta}$ values, an appropriate starting value is to set $\mathbf{V}(n,0) \equiv \mathbf{0}$. It is more correct, however, to use the extended Kalman filter covariance matrix at time n as the start value for $\mathbf{V}(n,0)$.

Efron, B. (1979), "Bootstrap methods: another look at the jackknife", *Annals of Statistics,* 7, 1-26.

Efron, B. (1982), "The Jackknife, Bootstrap, and other Resampling Plans", *SIAM, CMBS-NSF Monographs, 38*, Philadelphia.

Harvey, A. C. (1990*), Forecasting, Structural Time Series Models, and the Kalman Filter*, Cambridge: Cambridge University Press.

Harvey, A. C. and R. D. Snyder (1990), "Structural Time Series Models in Inventory Control", *International Journal of Forecasting*, 6, 187-198.

Harvey, A. C. and P. H. J. Todd(1983), "Forecasting Economic Time Series with Structural and Box-Jenkins Models: A Case Study", *Journal of Business and Economic Statistics*, 1, 299-307.

Holt, C. E. (1957), "Forecasting Trends and Seasonals by Exponentially Weighted Averages", *ONR Memorandum No. 52*, Carnegie Institute of Technology, Pittsburgh.

Johnston, F. R. and P .J. Harrison (1986), "The Variance of Lead Time Demand", *Journal of the Operational Research Society*, 37, 303-308.

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

McCullough, B. D. (1994), "Bootstrapping forecast intervals: an application to AR(p) models", *Journal of Forecasting*, 13, 51-66.

McKenzie, E. (1984), "General exponential smoothing and the equivalent ARMA process", *Journal of Forecasting*, 3, 333-344.

Morris,A.H. (1993), *NSWC Library of Mathematics Subroutines*, Dahlgren, Virginia: Naval Suface Warfare Center.

Ord, J. K. and A. B. Koehler (1990), "A Structural Model for the Multiplicative Holt-Winters Method", *Proceedings, Decision Sciences Institute*, 555-57.

Snyder, R. D. (1985), "Recursive Estimation of Dynamic Linear Statistical Models", *Journal of the Royal Statistical Society*, B, 272-276.

Winters, P. R. (1960), "Forecasting Sales by Exponentially Weighted Moving Averages", *Management Science*, 6, 324-42.

Figure 1 Time Conventions

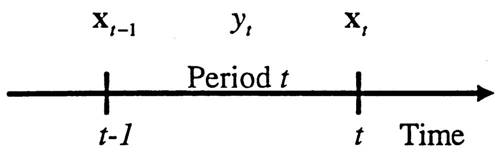

Figure 2. Quarterly Total Retail Sales: Canada

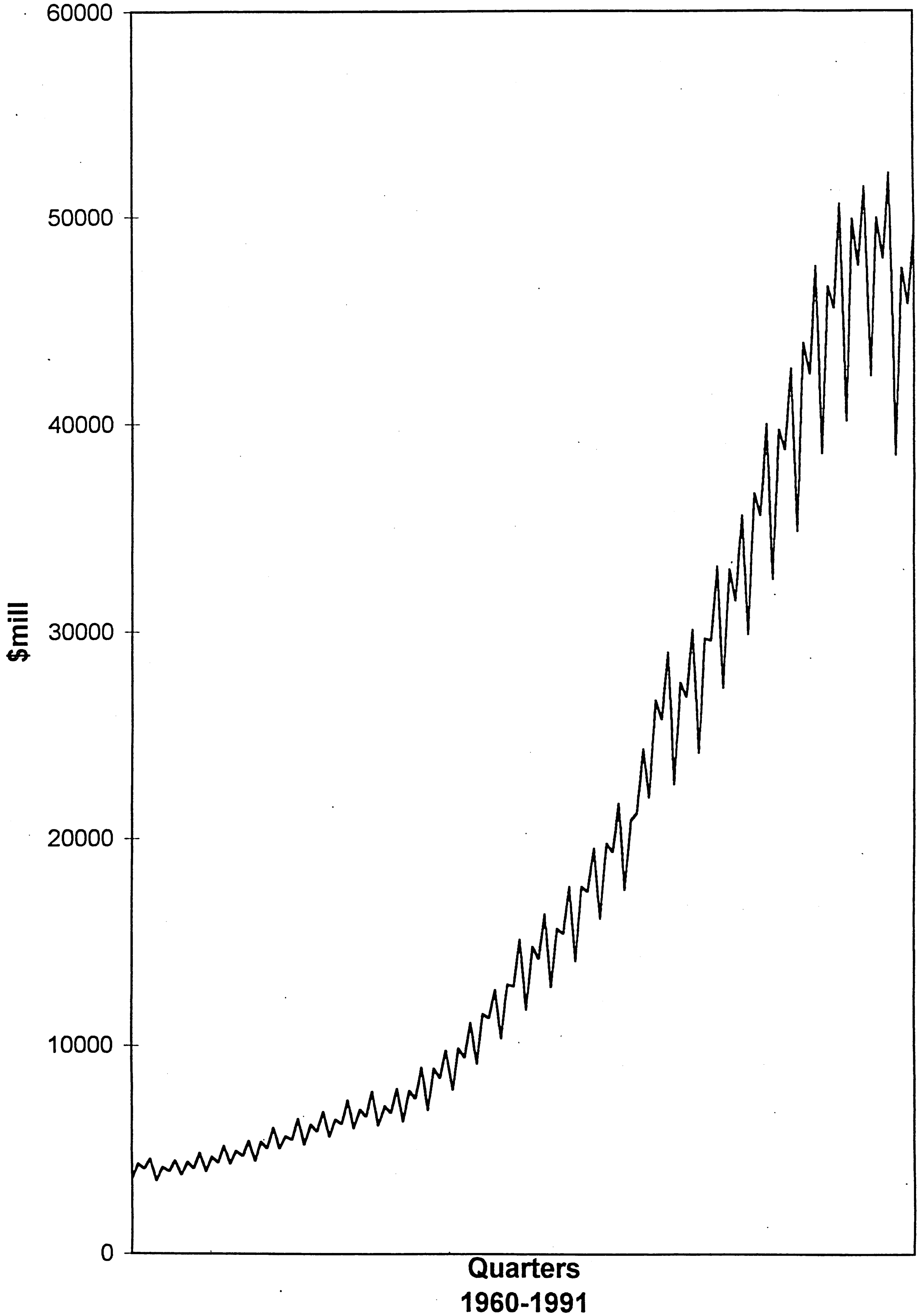

Table 1. Point Predictions and Estimates: Simulation Results

Quantity	Actual Value	Sample Size 32 Average Error	32 Standard Error	80 Average Error	80 Standard Error	200 Average Error	200 Standard Error
Prediction 1		0.00	0.04	0.00	0.04	0.00	0.05
Prediction 2		0.00	0.04	0.00	0.05	0.00	0.06
Prediction 3		0.00	0.05	0.00	0.06	0.00	0.07
Prediction 4		0.00	0.06	0.00	0.06	0.00	0.08
Prediction 5		0.00	0.07	0.00	0.08	0.00	0.10
Prediction 6		0.00	0.08	0.00	0.09	0.00	0.11
Prediction 7		0.00	0.09	0.00	0.10	0.00	0.12
Prediction 8		0.00	0.10	0.00	0.11	0.00	0.14
Seed Level	100	-0.01	0.03	0.00	0.03	0.00	0.03
Seed Rate	1.00	0.04	1.08	-0.13	1.01	-0.23	1.08
Seed Index 1	1.20	0.00	0.01	0.00	0.01	0.00	0.01
Seed Index 2	0.95	0.00	0.01	0.00	0.01	0.00	0.01
Seed Index 3	0.80	0.00	0.01	0.00	0.01	0.00	0.01
Seed Index 4	1.05	0.00	0.01	0.00	0.01	0.00	0.01
Alpha 1	0.50	0.23	0.52	0.05	0.24	0.01	0.14
Alpha 2	0.10	0.21	0.94	0.17	0.50	0.10	0.29
Alpha 3	0.10	-0.11	1.21	-0.13	0.63	-0.06	0.36
Std Devn	0.03	-0.02	0.15	-0.01	0.09	0.00	0.05
Rejected Trials		1		191		1434	

Table 2. Ninety Percent Intervals: Simulation Results

	Sample Size											
	32				80				200			
Quantity	**Trad**	**LAM**	**Boot**	**Semi**	**Trad**	**LAM**	**Boot**	**Semi**	**Trad**	**LAM**	**Boot**	**Semi**
Prediction 1	0.67	0.85	0.81	0.91	0.73	0.89	0.88	0.92	0.64	0.90	0.88	0.90
Prediction 2	0.70	0.84	0.79	0.88	0.74	0.88	0.86	0.90	0.66	0.90	0.88	0.91
Prediction 3	0.72	0.82	0.77	0.88	0.76	0.86	0.86	0.89	0.70	0.90	0.88	0.91
Prediction 4	0.54	0.81	0.75	0.84	0.59	0.87	0.85	0.88	0.55	0.89	0.86	0.88
Prediction 5	0.40	0.77	0.73	0.85	0.50	0.84	0.85	0.88	0.45	0.88	0.86	0.91
Prediction 6	0.44	0.78	0.71	0.81	0.53	0.84	0.83	0.86	0.46	0.89	0.84	0.89
Prediction 7	0.46	0.77	0.70	0.82	0.56	0.83	0.84	0.85	0.48	0.88	0.84	0.88
Prediction 8	0.32	0.73	0.70	0.80	0.42	0.82	0.83	0.86	0.38	0.86	0.82	0.91
Alpha 1			0.78	0.76			0.83	0.86			0.87	0.89
Alpha 2			0.78	0.71			0.75	0.69			0.80	0.76
Alpha 3			1.00	1.00			1.00	0.99			0.98	0.98
Seed Level		0.79	0.71	0.84		0.88	0.82	0.84		0.88	0.86	0.86
Seed Rate		0.61	0.65	0.74		0.82	0.83	0.85		0.85	0.86	0.88
Seed Index 1		0.81	0.81	0.92		0.79	0.88	0.94		0.77	0.90	0.91
Seed Index 2		0.81	0.81	0.91		0.79	0.90	0.92		0.79	0.91	0.89
Seed Index 3		0.80	0.82	0.91		0.78	0.87	0.93		0.81	0.91	0.89
Seed Index 4		0.79	0.85	0.94		0.79	0.87	0.90		0.76	0.90	0.91

Trad = Traditional Stationary Approximation Method
LAM = Linear Approximation Method
Boot = Bootstrap Method
Semi = Semi-Parametric Method

Table 3. Predictions of Canadian Total Retail Sales

		Holt-Winters Method					Naive Seasonal Method				
Date	Actual	Prediction	Lower Limit	Upper Limit	Error Margin	Actual Error	Prediction	Lower Limit	Upper Limit	Error Margin	Actual Error
Mar Qtr 1994	39416	38817	37020	40614	5%	2%	38567	32264	44869	16%	2%
Jun Qtr 1994	47683	47373	44585	50161	6%	1%	47548	39739	55356	16%	0%
Sep Qtr 1994	46901	45221	41957	48484	7%	4%	45869	38575	53164	16%	2%
Dec Qtr 1994	51049	48729	44528	52931	9%	5%	49225	41486	56963	16%	4%
Mar Qtr 1995	41004	38409	34303	42516	11%	7%	38567	29654	47480	23%	6%
Jun Qtr 1995	50009	46875	41170	52579	12%	7%	47548	36505	58590	23%	5%
Sep Qtr 1995	49405	44744	38598	50889	14%	10%	45869	35553	56186	22%	8%
Dec Qtr 1995	53398	48214	40799	55628	15%	11%	49225	38280	60169	22%	8%

Note: Error Margin = 100*(UpperLimit-LowerLimit)/(2*Prediction)

Table 4. Canadian Total Retail Sales: Holt-Winters Model Estimates

Quantity	Estimate	Lower Limit	Upper Limit	Error Margin
Seed Level	4089	3918	4259	4%
Seed Rate	18	-59	96	422%
Seed Index (Mar)	0.88	0.87	0.90	2%
Seed Index (Jun)	1.04	1.02	1.06	2%
Seed Index (Sep)	0.98	0.96	0.99	2%
Seed Index (Dec)	1.10	1.08	1.12	2%
Alpha 1	0.67			
Alpha 2	0.11			
Alpha 3	0.23			
Std Devn	0.03			

Note: Error Margin = 100*(UpperLimit-LowerLimit)/(2*Estimate)