MONASH University

Australia

Department of Econometrics
and Business Statistics

Time Series Forecasting: The Case for the Single Source of Error State Space

**J Keith Ord, Ralph D Snyder, Anne B Koehler,
Rob J Hyndman and Mark Leeds**

I

April 2005

April 2, 2005

Time Series Forecasting: The Case for the Single Source of Error State Space Approach

J. Keith Ord   McDonough School of Business

[Corresponding   Georgetown University, Washington, DC 20057, USA.
Author]
                       Email: ordk@georgetown.edu

Ralph D. Snyder   Department of Econometrics and Business Statistics

                       P.O. Box 11E, Monash University, VIC 3800, Australia

                       Email: ralph.snyder@buseco.monash.edu.au

Anne B. Koehler   Department of Decision Sciences and Management Information Systems

                       Miami University, Oxford, OH 45056, USA.

                       Email: koehleab@muohio.edu

Rob J. Hyndman   Department of Econometrics and Business Statistics

                       Monash University, VIC 3800, Australia

                       Email: rob.hyndman@buseco.monash.edu.au

Mark Leeds   Millenium Partners
                      666 Fifth Avenue, New York, New York, USA.
                      Email: mleeds@mlp.com

Time Series Forecasting: The Case for the Single Source of Error State Space Approach

Abstract

The state space approach to modelling univariate time series is now widely used both in theory and in applications. However, the very richness of the framework means that quite different model formulations are possible, even when they purport to describe the same phenomena. In this paper, we examine the single source of error [SSOE] scheme, which has perfectly correlated error components. We then proceed to compare SSOE to the more common version of the state space models, for which all the error terms are independent; we refer to this as the multiple source of error [MSOE] scheme.

As expected, there are many similarities between the MSOE and SSOE schemes, but also some important differences. Both have ARIMA models as their reduced forms, although the mapping is more transparent for SSOE. Further, SSOE does not require a canonical form to complete its specification. An appealing feature of SSOE is that the estimates of the state variables converge in probability to their true values, thereby leading to a formal inferential structure for the ad-hoc exponential smoothing methods for forecasting. The parameter space for SSOE models may be specified to match that of the corresponding ARIMA scheme, or it may be restricted to meaningful sub-spaces, as for MSOE but with somewhat different outcomes.

The SSOE formulation enables straightforward extensions to certain classes of non-linear models, including a linear trend with multiplicative seasonals version that underlies the Holt-Winters forecasting method. Conditionally heteroscedastic models may be developed in a similar manner. Finally we note that smoothing and decomposition, two crucial practical issues, may be performed within the SSOE framework.

Keywords: *ARIMA, dynamic linear models, equivalence, exponential smoothing, forecasting, GARCH, Holt's method, Holt-Winters method, Kalman filter, prediction intervals.*

1. Introduction

State space models, also known as dynamic linear models, afford considerable flexibility in the specification of the parameter structure. For example, the most common specifications (e.g. Harvey, 1989; West and Harrison, 1997) assume that the errors in the state equations are mutually independent and that these errors are independent of the errors in the observation (or measurement) equation. A complete account of recent developments is provided in the monograph by Durbin and Koopman (2001). This set of conditions provides enough constraints to ensure that the remaining parameters are *estimable* (termed *just identified* in the econometric literature). Clearly there are an infinite number of ways in which the parameter space could be constrained to achieve estimability and the purpose of this paper is to present the case for an option known as the *single source of error [SSOE]* formulation, wherein all the error sources are perfectly correlated. By contrast, we refer to the standard form as having *multiple sources of error [MSOE]*.

Over the years state space models have been formulated in a variety of different ways. Following the seminal work of Kalman (e.g. Kalman and Bucy, 1961) a considerable literature developed in engineering (e.g. Jazwinski, 1970; Anderson and Moore, 1979). Early work in the statistical area included the Markovian representation developed by Akaike (1974). Hannan and Diestler (1988) provided a unifying presentation of work by engineers and statistical time-series analysts for stationary time series. In economics, Beveridge and Nelson (1981) used the state space approach to analyse the components of a time series, and Aoki (1991) looked at multivariate state space models and suggested procedures for both stationary and nonstationary data. For a detailed review of books in the area, see Durbin and Koopman (2001, page 5).

The state space models, both SSOE and MSOE, have their historical roots in the exponential smoothing methods of the 1950's [Brown (1959), Holt (1957), and Winters (1960)]. The most basic of these methods is simple exponential smoothing. We will denote a time series by y_t, $t = 1, 2, \ldots$, a one-period-ahead forecast by $\hat{y}_t(1)$, and a one-period-ahead forecast error by $e_t = y_t - \hat{y}_{t-1}(1)$. Then the one-period-ahead forecast in simple exponential smoothing is

$$\hat{y}_t(1) = \hat{\ell}_t \qquad\qquad (1.1a)$$

where

$$\hat{\ell}_t = \hat{\alpha}_1 y_t + (1 - \hat{\alpha}_1)\hat{\ell}_{t-1} \qquad (1.1b)$$

or in the equivalent error correction form

$$\hat{\ell}_t = \hat{\ell}_{t-1} + \hat{\alpha}_1 e_t \qquad (1.1c)$$

Here $\hat{\ell}_t$ is an estimate of the mean or level of y_{t+1} made at time t, and $\hat{\alpha}_1$ is an estimated smoothing parameter.

An MSOE model that corresponds to simple exponential smoothing is

$$y_t = \ell_t + u_t \qquad (1.2a)$$

$$\ell_t = \ell_{t-1} + w_t \qquad (1.2b)$$

In this MSOE model ℓ_t is an unobserved state variable. It is the level or mean of the time series that can be estimated when the values of $y_1, y_2, \ldots, y_t$ are known. The two sources of error u_t and w_t are assumed to be $N(0, \sigma_u^2)$ and $N(0, \sigma_w^2)$, respectively. They are also assumed to be mutually and serially independent. Muth (1960) was the first to show that the forecasts from simple exponential smoothing are the minimum mean square forecasts for this MSOE model. By applying the Kalman filter to the model, one can see that the steady state equation for the level corresponds to equation (1.1) in simple exponential smoothing.

Muth (1960) also investigated an SSOE model that Snyder (1985) put in the following form

$$y_t = \ell_{t-1} + \varepsilon_t \qquad (1.3a)$$

$$\ell_t = \ell_{t-1} + \alpha_1 \varepsilon_t \qquad (1.3b)$$

The single source of error ε_t is assumed to be $N(0, \sigma^2)$ and serially independent, and α_1 is a parameter. It is clear that this model has the same forecasts as those produced by equation (1.1).

1.1 Comparison with ARIMA models

One of the oldest and most widely known classes of models is the collection of the integrated autoregressive/moving average (ARIMA) models (see Box, Jenkins, and Reinsel, 1994). Linear state space

models correspond to equivalent ARIMA models. For example, taking first differences of the SSOE model (1.3), we obtain

$$(1-L)y_t = \varepsilon_t - (1-\alpha_1)\varepsilon_t \tag{1.4}$$

where L is the lag operator (i.e. $Ly_t = y_{t-1}$). This ARIMA model is referred to as the *reduced form* of the SSOE model (1.3). First differences of the MSOE model (1.2) also produce an ARIMA(0,1,1) model but with a restricted parameter space, as we shall see in Section 5. All three models (ARIMA, SSOE, and MSOE) produce the same forecasts as those of simple exponential smoothing. We will call these models local level models [LLM].

One of the advantages of both state space models over the ARIMA model is that we can clearly see that the time series varies around a mean or level that is changing. This advantage is even more noticeable when we add trend and seasonal components to the time series. In the case of a time series where both the level and growth rate are changing, we have the following models that provide the same forecasts as Holt's trend corrected exponential smoothing (1957). We will call these models the local linear trend models [LLTM].

MSOE model (Theil and Wage, 1964; Nerlove and Wage, 1964)

$$y_t = \ell_t + u_t \tag{1.5a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + w_{1t} \tag{1.5b}$$

$$b_t = b_{t-1} + w_{2t} \tag{1.5c}$$

SSOE model (Snyder, 1985)

$$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t \tag{1.6a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1\varepsilon_t \tag{1.6b}$$

$$b_t = b_{t-1} + \alpha_2\varepsilon_t \tag{1.6c}$$

ARIMA (Reduced form of the SSOE model (1.6))

$$(1-L)^2 y_t = \varepsilon_t - (2-\alpha_1-\alpha_2)\varepsilon_{t-1} - (\alpha_1-1)\varepsilon_{t-2} \tag{1.7}$$

where ℓ_t is the level of the time series and b_t is the growth rate; the sources of error u_t, w_{1t} and w_{2t} are assumed to be N(0, σ_u^2), N(0, σ_{w1}^2), and N(0,σ_{w2}^2), respectively; u_t, w_{1t}, and w_{2t} are mutually and serially

5

independent; the ε_t are $N(0, \sigma^2)$ and serially independent; and α_1 and α_2 are parameters. The components of level and trend (or growth rate) in the time series are quite obvious in the state space models as the components are estimated as part of the forecasting process. The components are less easily identified in the ARIMA model; we must resort to a decomposition method such as that of Hillmer and Tiao (1982). Thus, state space models are superior to ARIMA models when knowledge of the components is important.

In the Box-Jenkins approach, ARIMA models are selected by the use of the autocorrelation and partial autocorrelation functions. The sample functions are affected by sampling variability, and frequently quite different models can appear to fit the same series. Thus, the wider range of candidate ARIMA models in the Box-Jenkins approach may not be an advantage over the state space approach, which focuses on the components (level, trend, seasonal factors) to choose a model.

1.2 Outline of the paper

Our objective in this paper is to explore a range of topics related to the specification of time series models for forecasting. For each topic we compare the SSOE with the MSOE approach. In Section 2 we present a general form for state space models and show the specialization to the MSOE and SSOE forms. In the same section, we also investigate model equivalence in the sense of West and Harrison (1997) and show that the covariance structure is preserved among equivalent SSOE models but not among equivalent MSOE models. The Kalman filter for state space models is presented in Section 3. It is shown that, unlike the MSOE model, the covariance matrix for the SSOE model converges to 0 so that the estimates for the states converge to their true values. In Section 4, we look at estimation of the parameters in SSOE models using both conditional and exact likelihoods. For model selection with information criteria, it turns out that one should use conditional likelihood rather than exact likelihood. In Section 4, we also look at point forecasts and prediction intervals for the SSOE models with both the classical and Bayesian approaches. In the classical approach we are able to provide analytical formulas the variances of the forecast error for the SSOE models. Invertibility and parameter spaces for the SSOE models are examined in Section 5. We show that the SSOE model has the largest parameter space among models that are special cases of the general state space model of Section 2. We also show that there is a form of the SSOE model for which the parameter space matches that from

classical exponential smoothing. In Section 6 we look at the how the SSOE approach can be stated in a form that includes linear and nonlinear models in the same framework. In section 7, we discuss smoothing using an SSOE model. We look at extensions of the SSOE models to ARCH and GARCH specifications and to models with explanatory variables (including handling missing values) in Section 8. In the final section, we draw some overall conclusions about the relative advantages of SSOE and MSOE models.

2. Formulation and Equivalence of the State Space (Dynamic Linear) Models

We first define a general state space model in which the contemporaneous error terms may be correlated. Then we place restrictions on the covariance structure to obtain the formulations for the two special cases: the MSOE model and the SSOE model, and we discuss the reduced form of these state space models. We end the section by defining model equivalence and showing that equivalent SSOE models retain the same covariance structure while equivalent MSOE models have different covariance structures.

2.1 The state space models

The general state space (dynamic linear) model

The general state space (dynamic linear) model is characterized by an observation equation and a set of state (or transition) equations that are shown below:

$$y_t = h'x_{t-1} + \varepsilon_t, \qquad\qquad \varepsilon_t \sim N[0, \sigma_t^2] \qquad\qquad (2.1a)$$

$$x_t = Fx_{t-1} + w_t, \qquad\qquad w_t \sim N[0, W_t] \qquad\qquad (2.1b)$$

$$\mathrm{cov}\begin{pmatrix} \varepsilon_t \\ w_t \end{pmatrix} = \begin{pmatrix} \sigma_t^2 & V_t' \\ V_t & W_t \end{pmatrix} \qquad\qquad (2.1c)$$

with the initial conditions:

$$(x_0 | I_0) \sim N[m_0, C_0]$$

where ε_t and w_t are serially independent error processes. The random variable y_t is taken to be a scalar and the (unobservable) state vector x_t is of order $(k \times 1)$. The h is a known $(k \times 1)$ vector, and F is a known $(k \times k)$ matrix. Thus, $h'x_{t-1}$ represents the linear combination of the state variables that provides the one-

step-ahead forecasts for y_t. The conditional variance of y_t given x_{t-1} is σ_t^2; likewise, W_t is the $(k \times k)$ conditional covariance matrix for the state vector x_t, and V_t is the $(k \times 1)$ vector of conditional covariances between y_t and x_t. When the time subscripts are dropped on the variance and covariance matrices, they are assumed to be constant. Finally, $I_t = \{y_t, I_{t-1}\}$ denotes the information set available at time t. In the remainder of the paper we will refer to the general state space model by the quintuple $\{h, F, \sigma_t^2, V_t, W_t\}$. More general processes with time-dependent h and F are clearly possible but are not considered in this paper.

A number of authors have examined this more general formulation, notably Koopman (1997). Our present purpose is primarily to embed the SSOE and MSOE formulations within this more general setting so that the results for both cases can be generated simultaneously.

The multiple source of error (MSOE) model

One of the key additional assumptions in the multiple sources of error (MSOE) formulation is that all the contemporaneous error processes are independent. In the context of the covariance matrix given by equation (2.1c), this assumption creates a diagonal matrix by imposing $(1/2)k(k+1)$ conditions. In principal the parameters become estimable provided we impose any such set of $(1/2)k(k+1)$ conditions.

We present the MSOE model in the form of the Harrison-West dynamic linear model (West and Harrison, 1997), which is characterized by the following observation equation and set of state equations:

$$y_t = g'x_t + u_t, \qquad u_t \sim N[0, \sigma_{ut}^2] \tag{2.2a}$$

$$x_t = Fx_{t-1} + w_t, \qquad w_t \sim N[0, W_t] \tag{2.2b}$$

where u_t and w_t are mutually independent error processes; σ_{ut}^2 is the variance of u_t; and initial conditions and other notation are the same as in model (2.1). Further the elements of w_t are assumed to be independent, so that $V_t = 0$ and W_t is a diagonal matrix. We refer to model (2.2) by the quadruple $\{h, F, \sigma_t^2, W_t\}$.

Model (2.2) can be restated in the form of model (2.1) by substituting equation (2.2b) into equation (2.2a) for the state vector x_t. We then have the modified coefficients:

$$h' = g'F, \quad \varepsilon_t = g'w_t + u_t, \quad \sigma_t^2 = \sigma_{ut}^2 + g'W_t g, \text{ and } V_t = W_t g \tag{2.3}$$

We now examine the relationship of the MSOE model in Section 1 for Holt's trend corrected exponential smoothing with model (2.2) and its reformulation as model (2.3).

Example 2.1. The MSOE version of the local linear trend model [LLTM]

By inspection of the equations for model (1.5), we see that in model (2.2)

$$g' = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad x_t = \begin{pmatrix} \ell_t \\ b_t \end{pmatrix}, \quad W = \begin{pmatrix} \sigma_{w1}^2 & 0 \\ 0 & \sigma_{w2}^2 \end{pmatrix}, \text{ and } V = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

whereas in model (2.3), $h' = g'F = \begin{pmatrix} 1 & 1 \end{pmatrix}$ and $V = Wg = \begin{pmatrix} \sigma_{w1}^2 \\ 0 \end{pmatrix}$.

The single source of error (SSOE) model

For the SSOE model, we impose the $(1/2)k(k+1)$ conditions on the covariance matrix in equation (2.1c) by following Snyder (1985) and assuming that all the $(k+1)$ random errors in model (2.1) are perfectly correlated. Given the assumption of normality, this implies that there is effectively only one error term. Thus, we describe the SSOE formulation as follows:

$$y_t = h'x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N[0, \sigma_t^2] \tag{2.4a}$$

$$x_t = Fx_{t-1} + \alpha \varepsilon_t \tag{2.4b}$$

with the same initial conditions as model (2.1) and where α is a $(k \times 1)$ vector often referred to as the permanent effects vector. The elements of α may either be positive or negative and the covariance matrix for the state equations is:

$$W_t = Cov(\alpha \varepsilon_t) = \sigma_t^2 \alpha \alpha'$$

By construction, the elements of the error vector in the state equations are perfectly correlated. The (i,j)th element of the correlation matrix is $\text{sign}(\alpha_i \alpha_j)$

We also have that $V_t = \sigma_t^2 \alpha$. With this notation and the model as defined, the SSOE model may be characterized by the quadruple $\{h, F, \sigma_t^2, \alpha\}$. We now relate the SSOE model for Holt's trend corrected smoothing from Section 1 to model (2.4).

Example 2.2. The SSOE version of the LLTM

Looking at the equations for model (1.6), we see that in model (2.4)

$$h' = \begin{pmatrix} 1 & 1 \end{pmatrix}, \quad F = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad x_t' = \begin{pmatrix} \ell_t \\ b_t \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

$$W = \sigma^2 \begin{pmatrix} \alpha_1^2 & \alpha_1 \alpha_2 \\ \alpha_1 \alpha_2 & \alpha_2^2 \end{pmatrix}, \quad \text{and} \quad V = \sigma^2 \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$$

ARIMA reduced form of the state space model

We may use the lag operator, defined as $Ly_t = y_{t-1}$, to eliminate the state variables from the general state space model and derive the ARIMA reduced form (c.f. Harvey, 1989, pp. 66-69). For the SSOE model, we note that we can write equation (2.4b) as

$$(I - FL)x_t = \alpha \varepsilon_t \tag{2.5}$$

We can multiply both sides of equation (2.5) by the adjoint of $I - FL$ to obtain

$$\det(I - FL)x_t = adj(I - FL)\alpha \varepsilon_t \tag{2.6}$$

If the eigenvalues of F do not exceed 1, then $\det(I - FL) = \Phi(L)\Psi(L)$ where $\Psi(L)$ is a polynomial whose roots are all the unit eigenvalues of F, and $\Phi(L)$ is a polynomial that has an inverse. Then equation (2.6) can be written as

$$\Psi(L)x_t = \frac{adj(I - FL)}{\Phi(L)}\alpha \varepsilon_t \tag{2.7}$$

Multiplying equation (2.4a) by $\Psi(L)$ and using equation (2.7), we find the integrated moving average model

$$\Psi(L)y_t = \left(h' \frac{adj(I - FL)}{\Phi(L)}\alpha L + \Psi(L) \right)\varepsilon_t = \Theta(L, \alpha)\varepsilon_t \tag{2.8}$$

(Snyder, Ord, and Koehler, 2001)

See equations (1.6) and (1.7) for an example using the LLTM. Some care is needed in applying expression (2.8) since the resulting model may be overdifferenced, as shown by the following example.

Example 2.3: Converting the LLTM model with additive seasonal to an ARIMA form

The additive seasonal version of the SSOE model (2.4) with m periods may be expressed as:

$$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t \qquad\qquad (2.9a)$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1 \varepsilon_t \qquad\qquad (2.9b)$$

$$b_t = b_{t-1} + \alpha_2 \varepsilon_t \qquad\qquad (2.9c)$$

$$s_t = s_{t-m} + \alpha_3 \varepsilon_t \qquad\qquad (2.9d)$$

where s_t is the seasonal factor and m in the number of seasons per year. Direct application of (2.8) to (2.9) yields:

$$\Psi(L) = (1-L)^2(1-L^m)$$
$$\Theta(L,\alpha) = (1-L)^2(1-L^m) + L(1-L)(1-L^m)\alpha_1 + L(1-L^m)\alpha_2 + L^m(1-L)^2\alpha_3$$

Inspection of the right hand side expression clearly indicates the presence of a unit root. Elimination of the unit root yields the final model:

$$\Psi(L) = (1-L)(1-L^m)$$
$$\Theta(L,\alpha) = (1-L)(1-L^m) + L(1-L^m)\alpha_1 + (L+\ldots+L^m)\alpha_2 + L^m(1-L)\alpha_3$$

This model contains *(m+1)* moving average terms but only three parameters, so it differs from the usual seasonal ARIMA scheme. When $\alpha_2 = 0$, this model is close to the ARIMA $(0, 1, 1)(0, 1, 1)_m$ airline model, differing only by the factor $\alpha_1\alpha_3$ in the coefficient of L^{m+1}.

2.2 The concept of equivalence for state space models

The flexibility of the state space approach means that different models may give rise to the same forecasts in all circumstances. We now explore the implications of such equivalences for the MSOE and SSOE models.

Definition 2.1 (West and Harrison, 1997, page 144). A state space model is *observable* if and only if the $(k \times k)$ matrix T has full rank k, where

11

$$T = \begin{bmatrix} h' \\ h'F \\ \vdots \\ h'F^{k-1} \end{bmatrix}$$

Definition 2.2(West and Harrison, 1997, page 150). Two state space models, denoted by $M = \{h, F, \sigma_t^2, W_t\}$ and $\breve{M} = \{\breve{h}, \breve{F}, \breve{\sigma}_t^2, \breve{W}_t\}$, are *similar* if the state matrices F and $\breve{F}$ are similar matrices, so that for some non-singular $(k \times k)$ similarity matrix H,

$$F = H\breve{F}H^{-1}.$$

Definition 2.3 (West and Harrison, 1997, page 152, adapted). Consider two similar state space models *M* and $\breve{M}$ with similarity matrix $H = T^{-1}\breve{T}$ and initial moments $m_0 = H\breve{m}_0$ and $C_0 = H\breve{C}_0H'$. Then *M* and $\breve{M}$ are said to be *equivalent*, denoted by $M \equiv \breve{M}$, if

$$\sigma_t^2 = \breve{\sigma}_t^2 \quad \text{and} \quad W_t = H\breve{W}_t H'$$

for all *t*.

Two models that are observable and equivalent will produce the same forecast distributions (West and Harrison, Section 5.3, 1997).

A feature of the MSOE scheme is that if model *M* has a diagonal matrix *W* then $\breve{M}$ will typically have $\breve{W}$ with non-zero elements in the off-diagonal positions. That is, *M* has a state space formulation with independent errors, but $\breve{M}$ does not. West and Harrison (1997, section 5.4) go on to develop canonical models chosen, in effect, to guarantee special structure for *W*.

The required structure for the SSOE schemes is much simpler, as indicated by the following result.

Theorem 2.1. Consider the SSOE models $M = \{h, F, \sigma_t^2, \alpha\}$ and $\breve{M} = \{\breve{h}, \breve{F}, \breve{\sigma}_t^2, \breve{\alpha}\}$. If $H = T^{-1}\breve{T}$ with T and $\breve{T}$ given by definition 1, $M \equiv \breve{M}$ if and only if $\alpha = H\breve{\alpha}$.

Proof. Follows by direct evaluation of the covariance matrix for $\alpha\varepsilon_t$ (c.f. Leeds, 2000).

Corollary 2.1.1. If two SSOE schemes are equivalent , they both have all state error processes perfectly correlated.

The immediate implication of the Corollary is that, in contrast to MSOE models, we do not need to concern ourselves with the specification of a canonical form for SSOE models.

3. The Kalman Filter and Convergence

The following theorem, similar to that of West and Harrison (1997, pages 103-104), gives the updating equations in the Kalman filter for the general state space model (2.1).

Theorem 3.1. In the general state space model (2.1), the one-step ahead forecast and posterior distributions are given are given as follows:

(a) Posterior at time t-1:

$$(x_{t-1}|I_{t-1}) \sim N[m_{t-1}, C_{t-1}]$$

(b) Prior at time t:

$$(x_t|I_{t-1}) \sim N[Fm_{t-1}, R_t], \quad \text{where} \quad R_t = FC_{t-1}F' + W_t$$

(c) One-step ahead forecast:

$$(y_t|I_{t-1}) \sim N[f_t, Q_t], \quad \text{where} \quad f_t = h'm_{t-1} \text{ and } Q_t = h'C_{t-1}h + \sigma_t^2$$

(d) Posterior at time t:

$$(x_t|I_t) \sim N[m_t, C_t], \text{ where } m_t = Fm_{t-1} + a_t e_t \text{ and } C_t = R_t - a_t Q_t a_t', \text{ with } a_t = (V_t + FC_{t-1}h)Q_t^{-1} \quad \text{and}$$

$e_t = y_t - f_t.$ a_t *is known as the* <u>*Kalman gain*</u>.

Proof. Minor extension of West and Harrison (1997, pages 103-104). See also Koopman (1997).

Consider a constant variance state space model $\{h, F, \sigma^2, V, W\}$. As the number of observations increases, the posterior variance, C_t, defined in part (d) of Theorem 2 converges to a constant value, C say. A proof of this result is given, for example, by West and Harrison (1997, pages 162-163). The SSOE version has the following interesting extension of this result.

Theorem 3.2. For any observable constant variance SSOE model $\{h, F, \sigma^2, \alpha\}$, the limiting posterior variance converges to zero; i.e. $\lim_{t\to\infty} C_t = 0$.

Further, the posterior covariance converges to zero only for the SSOE form of the general model (2.1).

Proof. The proof of the first part follows the approach introduced by Caines and Mayne (1970 and correction in 1971), after correcting a significant error in their papers. Burridge and Wallis (1988) prove the same result for ARIMA models, using a different approach. A detailed presentation of the corrected Caines and Mayne approach is given in Leeds (2000, pages 63-73). The uniqueness property is proved in Leeds (2000, pages 78-79).

Corollary 3.2.1. If we take the limiting value of the posterior covariance and substitute back into the expression for the one-step ahead forecast in Theorem 3.1, we obtain:

$$(y_t \mid I_{t-1}) \sim N[f_t, Q_t],$$

where $f_t = h'm_{t-1}$ reduces to $h'x_{t-1}$ and $Q_t = h'C_{t-1}h + \sigma_t^2$ reduces to σ_t^2.

Although this result seems innocuous enough, its practical implications are most interesting. The implication of this corollary is that, after the series has "settled down" the forecasting updates are given precisely by the state equations. Thus, heuristic procedures such as Holt's linear trend (Holt, 1957) and the additive seasonal Holt-Winters method (Winters, 1960) can retain their forecasting structures and gain a stochastic modeling framework by adding an observation equation. Once the observation equation is specified, we can generate the predictive distribution and suitable prediction intervals. For further discussion of these issues, see Chatfield, Koehler, Ord and Snyder (2001).

The above results for the SSOE model are only useful if convergence takes place at a reasonable rate, so that the simpler updating expressions implied by the steady state can be used in applications. Detailed simulation results for the local level model [LLM] and the local level plus trend model [LLTM] are given in Leeds (2000, pages 80-85).

Referring back to Theorem 3.1 and taking $\sigma^2 = 1$, without loss of generality, we have the following recurrence relations for the LLM:

$$C_t = \frac{C_{t-1}(1-\alpha)^2}{1+C_{t-1}} \text{ and } a_t = \frac{\alpha + C_{t-1}}{1+C_{t-1}}.$$

Starting from $C_0 = 1000$ and $a_0 = 1$, the results in Table 1 show that the limits are approached rapidly unless α is very close to the limiting values of 0 or 2. Even then, convergence is still achieved.

Table 1: Convergence of Kalman gain and posterior variance for the Locally Linear Model [LLM]

Start values   C(0) = 1000   a(0) = 1

alpha	0.1 or 1.9		0.2 or 1.8		0.5 or 1.5		1	
t	C(t)	a(t)	C(t)	a(t)	C(t)	a(t)	C(t)	a(t)
1	0.809	0.999	0.639	0.999	0.250	1.000	0	1
2	0.362	0.503	0.250	0.512	0.050	0.600		
3	0.215	0.339	0.128	0.360	0.012	0.524		
4	0.144	0.260	0.073	0.291	0.003	0.506		
5	0.102	0.213	0.043	0.254	0.001	0.501		
6	0.075	0.183	0.027	0.233	0.000	0.500		
7	0.056	0.163	0.017	0.221	0.000	0.500		
8	0.043	0.148	0.010	0.213	0.000	0.500		
9	0.034	0.137	0.007	0.208	0.000	0.500		
10	0.026	0.129	0.004	0.205	0.000	0.500		

4. Inference with SSOE models

One of the results of using SSOE models is that sound statistical estimation of the parameters α and σ can be done with exponential smoothing rather than a Kalman filter. We describe both conditional and exact likelihood approaches to this estimation problem.

4.1 Conditional likelihood

A feature of the SSOE model is that we may replace the Kalman filter of Section 3.1 with exponential smoothing (Snyder, Ord and Koehler, 2001). If we solve Equation (2.4a) for ε_t and substitute it into Equation (2.4b), we obtain the smoothing equation

$$x_t = Fx_{t-1} + \alpha(y_t - h'x_{t-1}) \qquad\qquad (4.1)$$

This smoothing equation implies that fixed successive values of x_t can be computed recursively if we condition on a trial value for x_0 and assume the sample $Y_n = (y_1, y_2, \ldots, y_n)'$ is known. Strictly speaking, the x_t in (4.1) should be read as $(x_t | Y_t, x_0, \theta)$ where θ denotes the vector of unknown parameters contained in (h, F, α, σ). The SSOE form then implies that the observations are conditionally independent, or

$$(y_t|Y_{t-1}, x_0, \theta) \sim D(h'x_{t-1}, \sigma^2).$$

Thus, the likelihood function has the form

$$L(x_0, \theta|Y_n) = \prod_{t=1}^{n} p(y_t|Y_{t-1}, x_0, \theta)$$

where $p()$ is the pdf for ε_t. Since x_0 is treated as a fixed vector of unknown parameters, we have a conditional, rather than exact likelihood function.

The Kalman filter can only be used as part of the maximum likelihood procedure when the error terms in the SSOE are normally distributed. However, the exponential smoothing method outlined in this section can be applied for any error term distribution. When the ε_t are normally distributed, the conditional likelihood function for the SSOE model is given by

$$L(x_0, \theta|Y_n) = \frac{1}{(2\pi\sigma)^2} \exp\left(-\frac{\sum (y_t - h'x_{t-1})^2}{2\sigma^2}\right)$$

and x_t is found by using the smoothing equation (4.1).

4.2 Exact likelihood

The exact likelihood function is based on the distribution $p(y|\alpha, \sigma^2)$. The conditional likelihood, considered in the previous section, was based on the conditional distribution $p(y|\alpha, \sigma, x_0)$. The exact likelihood is obtained by averaging the conditional likelihood with respect to the seed state x_0. Full details are given in Snyder (2005).

4.3 Model selection

The Akaike information criterion (AIC) (Akaike, 1973) and other information criteria have become a common way of adjusting the likelihood to avoid overfitting in a model selection process. It is tempting to calculate an information criterion with the exact likelihood, but this does not work because of comparability problems. Since the state variables in the models are generated by non-stationary processes, the seed state vector has an improper unconditional distribution. One is confronted with a situation that is similar to

Bartlett's paradox (Bartlett (1957)) in Bayesian statistics. Exact likelihood values for models with different state dimensions are non-comparable and information criteria, such as the AIC, based on them will also be non-comparable. This is not an issue for the conditional likelihood because the use of an improper unconditional distribution of the seed vector is avoided.

It would be wrong to conclude from this that the conditional likelihood should be used in preference to exact likelihood with SSOE models. The estimators obtained with exact likelihood are less biased (Snyder, 2005). So it seems that estimation for SSOE models should utilize both types of likelihoods: the exact likelihood for estimation, and the conditional likelihood for model selection in conjunction with the AIC and other information criteria. See Billah, King, Snyder, and Koehler (2005) for more information on model selection.

4.4 Prediction intervals

There are three major approaches when using SSOE models to find prediction intervals for future values of a time series. One approach is to find analytical formulas for the variances of the forecast errors. Since it was shown in Section 2.1 that an SSOE model may be put into an ARIMA reduced form, one could find prediction intervals for an SSOE model by first converting it to its reduced form and then finding the ψ - weights in the Box-Jenkins approach. However, in contrast to the MSOE approach, one can directly derive a simple general analytical formula from the SSOE model; see Hyndman et al, (2005) for details. A second approach is to use an SSOE model to simulate the entire prediction distribution of $y_{n+\tau}$, given information through time period n. This approach enables us to account for the estimation of the parameters in θ (Ord, Koehler, and Snyder, 1997; Snyder, Ord, and Koehler et al, 2001). A third way to find prediction intervals is part of the Bayesian approach to estimation and prediction for SSOE models that is presented in Forbes, Snyder and Shami, 2000).

5. Specification of the Parameter Space

We now turn to a comparison of restrictions on the parameters for the MSOE and SSOE models. One desired restriction is that the parameters meet the condition that the corresponding ARIMA reduced model be

invertible. In addition the form of the state space model itself can restrict the size of the parameter space. We first investigate these latter restrictions. Then we look at the requirements for an SSOE model to be invertible.

5.1 Limitations on the parameter space imposed by the model

ARIMA models can be written in the general state space form (c.f. West and Harrison, 1997, pp. 296-299). Conversely, by following a sequence of operations employed in the introduction and Section 2.1, a general state space model with constant values in its quadruple can be converted to an ARIMA reduced form. These results are rather beguiling since it should be borne in mind that, although the model classes are formally equivalent in this sense, important differences remain:

1.   The state space models do not usually make an assumption of stationarity.

2.   Models selected by standard identification procedures in one regime may not be selected under the other (e.g. the seasonal state space model that corresponds to Holt-Winters additive seasonal exponential smoothing would not be identified under a Box-Jenkins approach for ARIMA models).

3.   The implied parameter space of a state space model may not be the same as the parameter space of the corresponding ARIMA scheme.

It is the third issue that we now consider. We proceed by examining the LLTM process in its three forms: the ARIMA(0,2,2) model

$$(1-L)^2 y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

and its two state space counterparts, as outlined in the Section 1. Assuming stationarity, the autocorrelations $[\gamma_k, k = 1, 2, \ldots]$ for $z_t = (1-L)^2 y_t$ in the three forms are summarized in Table 3. As noted in Section 1, we can set $\theta_1 = 2 - \alpha_1 - \alpha_2$ and $\theta_2 = \alpha_1 - 1$ to match up the ARIMA and SSOE forms exactly. It follows that these two formulations yield the same parameter space. However, the requirement that the variances in the MSOE scheme must be non-negative restricts its parameter space. It follows from the entries in Table 3 that the MSOE scheme imposes the restrictions:

$$-0.667 < \gamma_1 < 0 \qquad \text{and} \qquad 0 < \gamma_2 < 0.167,$$

in contrast to an invertible ARIMA or SSOE model that imposes the restrictions:

$$-0.707 < \gamma_1 < 0.707 \quad \text{and} \quad -0.5 < \gamma_2 < 0.5.$$

Table 3. The autocorrelation functions for the three forms of the LLTM model

Lag	ARMA	MSOE	SSOE
1	$\dfrac{(-\theta_1 + \theta_1\theta_2)}{(1+\theta_1^2+\theta_2^2)}$	$\dfrac{(-4\sigma_u^2 - \sigma_{w1}^2)}{(6\sigma_u^2 + 2\sigma_{w1}^2 + \sigma_{w2}^2)}$	$\dfrac{-(2-\alpha_1-\alpha_2)(2-\alpha_1)}{1+(2-\alpha_1-\alpha_2)^2+(\alpha_1-1)^2}$
2	$\dfrac{(-\theta_2)}{(1+\theta_1^2+\theta_2^2)}$	$\dfrac{\sigma_u^2}{(6\sigma_u^2 + 2\sigma_{w1}^2 + \sigma_{w2}^2)}$	$\dfrac{(1-\alpha_1)}{1+(2-\alpha_1-\alpha_2)^2+(\alpha_1-1)^2}$
≥ 3	0	0	0

In general, the form of the parameter space will depend upon the correlations among the errors in the state and observations equations even though, as noted Section 2.1, these coefficients are not estimable. For other models, such as that proposed by Beveridge and Nelson (1981), some of the correlations may be estimable; see Morley, Nelson and Zivot (2002).

In defence of the MSOE scheme, it is sometimes argued that the region defined by the non-negative variances defines the natural boundaries of the parameter space and that any extension beyond those limits is artificial. However, we would argue that a more natural approach is to consider the error adjustment mechanism specified in the SSOE form and then to decide whether the parameter space should be restricted.

A natural question that arises from the foregoing analysis is whether some other pattern of correlation among the state equation error variables would lead to a larger parameter space. To answer this question, we may consider the general LLTM given by space model (2.1).

To determine the boundaries of the parameter space, we may consider fractional programming problem (Charnes and Cooper, 1962) to maximise the extreme values of the autocorrelations, subject to the choice of correlation structure among the errors. The details are given in Leeds (2000, pages 50-56) and lead to the following result:

Theorem 5.1. The SSOE version of the Local Linear Trend Model yields the largest parameter space for the autocorrelations, independent of the values of α_1 and α_2 .

Since there is a one-to-one mapping between (γ_1, γ_2) and (θ_1, θ_2) the parameter space is the largest possible when the SSOE scheme is used and, as we have seen, exactly matches the parameter space under the usual ARMA assumptions.   The theorem is capable of extension to higher order schemes:

Theorem 5.2. Consider an SSOE version of the general state space model, whose stationary component may be defined in terms of the first k autocorrelations. This form maximizes the parameter space of the equivalent ARIMA model in the sense that no other general state space model, as defined in (2.1), for the same ARIMA model can have a larger parameter space.

In the general case, we were not able to prove that the largest parameter space was uniquely determined by the SSOE scheme, although we believe this to be so. In any event, no reduction in the parameter space could ever arise from using the SSOE version. The practical impact of using the SSOE version rather than the MSOE standard is that the best estimates of the parameters may lie in the "forbidden territory" of the MSOE scheme   For example, Garcia-Ferrer and del Hoyo (1992) compared the forecasting performance of the ARIMA and MSOE approaches and found ARIMA to be slightly better. The restricted parameter space for MSOE could be a partial explanation for such results.

5.2 Invertibility and classical conditions for parameter space of the SSOE model

Because of the historical roots in classical exponential smoothing, one of the interesting questions for the state space models is the relationship between the classical restrictions on the parameters and the invertibility conditions on these same parameters. For example, consider the three local level models for simple exponential smoothing that were presented in Section 1. The classical restriction on the parameter α_1 is that it lie between 0 and 1, while the invertibility conditions for the corresponding ARIMA model translates into the condition that α_1 may lie anywhere between 0 and 2.   Hence, we now consider the parameter spaces for invertible SSOE models and their relationship to the classical parameter spaces for the common exponential smoothing methods (Hyndman, Akram, and Archibald, 2003).

We define invertibility as follows.

Definition 5.1. An SSOE model is said to be *invertible* if there exists a sequence of constants $\{\pi_j\}$ such that

$\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j y_{t-j}$$

This is analogous to the definition of invertibility for an ARMA model. See, for example, Brockwell and Davis (1991).

Theorem 5.3. Let y_t be defined by the state SSOE model (2.4), and let $D = F - \alpha h'$. Then the model is

invertible if and only if all the eigenvalues of D lie inside the unit circle (Snyder, Ord, and Koehler, 2001).

Under some circumstances, it is useful to have a weaker notion of invertibility which we shall call

forecastability, formally introduced by Hyndman, Akram and Archibald (2003), although the basic concept

was originally identified by Sweet (1985) and independently in Lawton (1998).

Definition 5.2. Let (λ_i, v_i) denote an eigenvalue-eigenvector pair of *D*. Then an SSOE model is said to be

forecastable if, for all i, either $|\lambda_i| < 1$ or $h'F^{\tau-1}v_i = 0$ where $h'F^{\tau-1}x_n = \mu_n(\tau) = E(y_{n+\tau} | Y_n, x_0, \theta)$.

The notion of forecastability is motivated by the idea that a non-invertible model can still produce stable

forecasts provided the eigenvalues which cause the non-invertibility have no effect on the forecasts.

Obviously, any model that is invertible is also forecastable.

For SSOE models that have the same point forecasts as the nonseasonal classical exponential methods, such

as Holt's trend corrected exponential smoothing and damped trend exponential smoothing, the classical

parameter space is contained in the parameter space for invertibility. For example, the classical restrictions

are $0 < \alpha_1 < 1$ and $0 < \alpha_2 < \alpha_1$, and the invertibility conditions are $0 < \alpha_1 < 2$ and $0 < \alpha_2 < 4 - 2\alpha_1$.

The classical requirements for the smoothing parameters are determined by requiring that all the parameters

lie between 0 and 1 for the classical form of the exponential smoothing equations and then looking at the

translation of these conditions to the parameters in the error correction form of the exponential smoothing

equations. The SSOE models are written in the same form as the error correction form of the exponential smoothing equations (See equations (1.1b), (1.1c), (1.3b) for example). Holt-Winters additive seasonal exponential smoothing is another matter. The corresponding SSOE model was given in equations (2.9 a-d). It has no invertible region for its parameters. However, the parameter space determined by the classical restrictions, $0 < \alpha_1 < 1$, $0 < \alpha_2 < \alpha_1 < 1$, and $0 < \alpha_3 < 1 - \alpha_1$, intersects the parameter space that assures forecastability. The lack of an invertibility region can be corrected by replacing the state equation for the seasonal factor with an equation that normalizes the seasonal factors. For more information see Hyndman, Akram, and Archibald (2003).

In cases where we want a state space model that matches a classical exponential smoothing method, we can always restrict the parameter model for the SSOE model in the estimation procedure to match the classical restrictions. Another possibility involves the MSOE model (2.2) where the independence assumption for the disturbances is replaced by a perfect correlation assumption. It is shown in Snyder (2005) that the persistence parameters of the resulting SSOE model correspond to the classical restrictions from exponential smoothing.

6. Nonlinear Models

An important advantage that SSOE models have over MSOE models is that there is a general form for SSOE models (Ord, Koehler, and Snyder, 1997) that encompasses both the linear SSOE model in (2.4) and nonlinear SSOE models. These nonlinear SSOE models can combine the error term and the state vector in a nonlinear relationship. They can also include a nonlinear function of the components of the state vector. The single source of randomness, ε_t, is the key to evaluating likelihood functions with exponential smoothing rather than an extended Kalman filter.

The general nonlinear SSOE state space model

$$y_t = h(x_{t-1}, \alpha) + k(x_{t-1}, \alpha)\varepsilon_t \qquad \varepsilon_t \sim IID(0, \sigma^2) \qquad (6.1a)$$

$$x_t = f(x_{t-1}, \alpha) + g(x_{t-1}, \alpha)\varepsilon_t \qquad (6.1b)$$

The state vector x_t is of order $(k \times 1)$, α is a vector of parameters, h and k are known continuous functions with continuous derivatives from $\mathfrak{R}^k \rightarrow \mathfrak{R}$, and f and g are known continuous mappings with continuous derivatives from $\mathfrak{R}^k \rightarrow \mathfrak{R}^k$. We now present examples to illustrate the inclusiveness of this general framework.

Example 6.1. Linear SSOE model: the heteroscedastic case

This example can also be considered to be a form of the SSOE model in (2.4) in which σ_t varies with time.

$$y_t = h'x_{t-1} + k(x_{t-1})\varepsilon_t \qquad\qquad (6.2a)$$

$$x_t = Fx_{t-1} + \alpha k(x_{t-1})\varepsilon_t \qquad\qquad (6.2b)$$

Example 6.2. LLTM model: the heteroscedastic case

This model is an extension of the SSOE model in (1.6) and a special case of the model in Example 6.1.

$$y_t = \ell_{t-1} + b_{t-1} + (\ell_{t-1} + b_{t-1})\varepsilon_t \qquad\qquad (6.3a)$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1(\ell_{t-1} + b_{t-1})\varepsilon_t \qquad\qquad (6.3b)$$

$$b_t = b_{t-1} + \alpha_2(\ell_{t-1} + b_{t-1})\varepsilon_t \qquad\qquad (6.3c)$$

Example 6.3. Model for multiplicative Holt-Winters

This SSOE model has a nonlinear function of the components of the state vector in the observation equation. It is a nonlinear SSOE model that underpins the multiplicative Holt-Winters method.

$$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t) \qquad\qquad (6.4a)$$

$$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha_1\varepsilon_t) \qquad\qquad (6.4b)$$

$$b_t = b_{t-1} + \alpha_2(\ell_{t-1} + b_{t-1})\varepsilon_t \qquad\qquad (6.4c)$$

$$s_t = s_{t-m}(1 + \alpha_3\varepsilon_t) \qquad\qquad (6.4d)$$

In the general nonlinear model (6.1), the observation equation may be used to eliminate ε_t from the state equation to give the following exponential smoothing form of the state (transition) equation:

$$x_t = f(x_{t-1}, \alpha) + g(x_{t-1}, \alpha) \left(\frac{y_t - h(x_{t-1}, \alpha)}{k(x_{t-1}, \alpha)} \right) \qquad (6.5)$$

Using the same arguments as in Section 4.1, the conditional likelihood function follows directly.

7. Smoothing and Filtering

The updating equations provide a recursive procedure for generating estimates of x_t given the observations up to and including time *t-1*, which we now denote by $x_{t|t-1}$ and refer to as the *filtered* estimator. We may also define an improved estimator, $x_{t|T}$ which is based upon the entire sample of *T* observations. This is the *smoothed* estimator and may be derived as the least squares estimator. De Jong (1989) provided computationally efficient procedures for evaluating these smoothed estimates.

Harvey and Koopman (2000) point out that the MSOE scheme leads to optimal symmetric two-sided smoothers (for an infinite series, clearly applications will involve truncation after a finite number of terms). Further, this smoother corresponds to the Wiener-Kolmogorov (WK) filter. They also show that when the components are correlated, as for the SSOE scheme, the resulting signal extraction filter is asymmetric. Indeed the perfect correlation among the components of the SSOE model led us to Theorem 3.2, which may be restated as:

$$x_{t|t-1} \xrightarrow{P} x_t \text{ as } t \to \infty. \qquad (7.1)$$

For any vector *a*, $V(a'x_{t|T}) \leq V(a'x_{t|t-1})$ which implies that $x_{t|t-1} \xrightarrow{P} x_{t|T}$. In other words, using the rest of the series does not improve the estimation of the state vector.

This result has been interpreted as a weakness of the SSOE approach, as it appears to suggest that smoothing and filtering operations are indistinguishable for such models. However, this argument is not correct. We know that any SSOE model may be expressed in ARIMA form, so that an appropriate WK filter could be

developed, at least in principle. The key point is that the one-sided state vector x_t is not the relevant set of components when we turn to consider smoothing, since it is defined in a one-sided context. The following example illustrates how an appropriate WK smoother can be constructed.

Example 7.1: local level model

Consider the LLM in (1.3), and re-write the ARIMA(0,1,1) model in (1.4) as

$$(1-L)y_t = [1-(1-\alpha_1)L]\varepsilon_t$$

The (doubly infinite)WK filter is given by:

$$\hat{\ell}_t = \frac{\alpha_1^2 y_t}{[1-(1-\alpha_1)L][1-(1-\alpha_1)L^{-1}]} = \frac{\alpha_1}{2-\alpha_1}\sum_{j=-\infty}^{\infty}(1-\alpha_1)^{|j|}y_{t-j} \qquad (7.2)$$

This smoother also corresponds to the two-sided Beveridge-Nelson (BN) filter given by Proietti and Harvey (2000), although it should be noted that the filter is admissible only for $0 < \alpha_1 < 1$. The WK and BN filters often do not have the same form.

As pointed out by Gijbels, Pope and Wand (1999), when exponential smoothing is interpreted as a kernel estimate, simple exponential smoothing is the natural forecast and filter (7.2) is the natural smoother.

7.1 Smoothed components

The approach just described provides a smoothed estimator for the mean of the process and we now turn to consider the individual components. Key elements in the analysis of economic time series are the creation of the deseasonalized series and the creation of a smoothed trend. Bell (1984) and Burridge and Wallis (1988) extended the WK filter to non-stationary series to enable the extraction of unobserved components.

One way to develop a WK filter for the components of the SSOE scheme would be to generate the corresponding ARIMA model and then apply a canonical decomposition, such as that developed by Hillmer and Tiao (1982). However, if we recall result (7.1), a much simpler approach is possible: we may construct

the seasonally adjusted or detrended series directly and then smooth the remaining components, as illustrated in the following example.

Example 7.2: seasonal adjustment and smoothed trend

Consider the SSOE model:

$$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$$

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

$$s_t = s_{t-m} + \alpha_3 \varepsilon_t$$

Given result (7.1) , we may generate an approximately detrended series as:

$$z_{1t} = y_t - \ell_{t|T} \approx s_{t-m} + \varepsilon_t \tag{7.3}$$

It follows from Example 7.1 that the smoothed seasonal components may be computed as:

$$\hat{s}_t \approx \frac{\alpha_3^2 z_{1t}}{[1-(1-\alpha_3)L^m][1-(1-\alpha_3)L^{-m}]} = \frac{\alpha_3}{2-\alpha_3} \sum_{j=-\infty}^{\infty} (1-\alpha_3)^{|j|} z_{1,t-jm} \tag{7.4}$$

In turn, the smoothed seasonal components lead to the deseasonalized series:

$$z_{2t} = y_t - \hat{s}_t \approx \ell_t + \varepsilon_t$$

The smoothed trend is then given by:

$$\hat{\ell}_t \approx \frac{\alpha_1^2 z_{2t}}{[1-(1-\alpha_1)L][1-(1-\alpha_1)L^{-1}]} = \frac{\alpha_1}{2-\alpha_1} \sum_{j=-\infty}^{\infty} (1-\alpha_1)^{|j|} z_{1,t-j} \tag{7.5}$$

We may iterate between (7.4) and (7.5) until the results converge, although the differences may be expected to be small provided the series is of reasonable length. Finally, if we suspect the presence of a cyclical component, we could examine the series:

$$z_{3t} = y_t - \hat{\ell}_t - \hat{s}_t \approx \varepsilon_t$$

7.2 Non-additive models

As noted earlier, one advantage of the SSOE scheme is that the single source of error allows us to model non-additive processes and still retain closed-form results. In particular, result (7.1) will continue to hold under

fairly mild conditions, so that we may define $\{z_{jt}; j = 1, 2, 3\}$ as before, but using ratios or other appropriate operations to define the modified series.

8. Extensions of the State Space Models

Some of the extensions of state space models are more easily implemented with the SSOE form of the state space model than in the MSOE form. We consider two such extensions in this section; one is modelling the changes in the variance of the series over time and the other is adding explanatory variables.

8.1 Autoregressive conditional heteroscedastic (ARCH) models

Following Harvey et al. (1992), we define the MSOE model with ARCH(1) effects.

MSOE model with ARCH(1) effects

Consider the extended model with heteroscedastic errors:

$$y_t = g'x_t + u_t, \tag{8.1a}$$

$$x_t = Fx_{t-1} + \Psi \eta_t, \tag{8.1b}$$

$$u_t = q_t^{1/2} u_t^* \qquad\qquad \eta_t = r_t^{1/2} \eta_t^* \tag{8.1c}$$

where $\varepsilon_t^* \sim N(0,1)$ and $\eta_t^* \sim N(0,1)$ and

$$q_t = \beta_0 + \beta_1 u_{t-1}^2 \tag{8.1d}$$

$$r_t = \gamma_0 + \gamma_1 \eta_{t-1}^2 \tag{8.1e}$$

Ψ is a vector of ones so that each component in the state equation experiences the same ARCH effect.

An ARCH assumption is that u_t and η_t are conditionally normal , given u_{t-1} and η_{t-1}. In order to deal with this assumption, Harvey et al. (1992) extend the state equation to include η_t. This extension is necessary because an estimate $\hat{\eta}_t$ is not observable. This problem will not be encountered in the SSOE model because there is only one source of error ε_t, and its estimate $y_t - \hat{y}_t$ can be obtained. Adding one more component to the state equation may not be very complex, but an ARCH(m) model will require the state equation to be

27

augmented by *m* components. Another advantage of the SSOE formulation is that state equations can have different ARCH effects that are implicitly modelled through the different components of the vector α.

SSOE model with ARCH(1) effects

The SSOE version of the conditionally heteroscedastic model may be written as:

$$y_t = h'x_{t-1} + \varepsilon_t, \tag{8.2a}$$

$$x_t = Fx_{t-1} + \alpha\varepsilon_t \tag{8.2b}$$

$$\varepsilon_t = q_t^{1/2}\varepsilon_t^* \tag{8.2c}$$

where $\varepsilon_t^* \sim N(0,1)$ and

$$q_t = \beta_0 + \beta_1\varepsilon_{t-1}^2 \tag{8.2d}$$

See Leeds (2000) for more details on ARCH, extensions to GARCH, and comparisons of the MSOE and SSOE formulations.

8.2 Explanatory Variables and Missing Values

Explanatory variables can be easily added to SSOE and MSOE models. The advantage of the SSOE model is that the extended model can readily be put in an equivalent regression framework and conventional least squares methods can be used in combination with general exponential smoothing to estimate the parameters and initial values in the model.

SSOE model with explanatory variables

We may incorporate explanatory variables into the observation equation of the SSOE scheme, as follows:

$$y_t = h'x_{t-1} + z_t'\gamma + \varepsilon_t \qquad \varepsilon_t \sim N(0,\sigma^2) \tag{8.3a}$$

$$x_t = Fx_{t-1} + \alpha\varepsilon_t \tag{8.3b}$$

where z_t is a vector of explanatory variables, and γ is a fixed effects vector of parameters.

Model (8.3) can be transformed into an equivalent regression

$$\widetilde{y}_t = \widetilde{z}'_t \beta + \varepsilon_t \qquad (8.4)$$

The dependent vector $\widetilde{y}$ is an appropriate transformation the raw series vector y; $\widetilde{z}'_t$ is an augmented transformation of h and z_t; and $\beta = \begin{bmatrix} x'_0 & \gamma' \end{bmatrix}'$. Conventional least squares methods may be applied to the regression (8.4).

9. Conclusions

The primary focus of this paper has been the comparison of MSOE and SSOE state space models. However, we should commence our summary by observing that the state space approach generally has advantages which both models share. In particular, the natural formulation of a model as the combination of components affords both ease of interpretation and a more straightforward decomposition of the series into such components (sections 1.1 and 7). Further, both schemes provide a basis for computationally efficient estimation procedures, although SSOE does not require explicit use of the Kalman filter (sections 4.1 and 4.2).

Turning to points of difference between MSOE and SSOE, we may note the following points:

1. The reduction to an ARIMA is more direct for SSOE models (section 2.1).

2. SSOE does not require the selection of a canonical form (section 2.2).

3. The estimates of the SSOE state variables converge in probability to their true values, a property not shared by MSOE (section 3). This result provides practical validation for many empirical forecasting methods that generate forecasts using the state equations without distinguishing estimates from true values.

4. The parameter space for an SSOE scheme matches that for the corresponding ARIMA scheme, whereas the space for the comparable MSOE scheme is smaller (section 5.1).

5. The formulation of an SSOE model may lead to a scheme that is *forecastable* but not invertible (section 5.2). Such schemes may still generate valid forecasts.

6. Smoothing may be achieved in the SSOE model using an adaptation of the Wiener-Kolmogorov filter. Smoothed estimates of individual components may be obtained by estimating the components and then smoothing.

7. The single source of error formulation allows straightforward extensions to certain classes of nonlinear model, such as the widely used model with linear trend and multiplicative seasonal effects (section 6).This structure also allows for a straightforward development of conditionally heteroscedastic models (section 8).

References

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (eds B. N. Petrov and F. Csaki), pp. 267-281). Budapest: Akdemiai Kiado.

Akaike, H. (1974) Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics*, **26**, 363-387.

Anderson, B. D. O. and Moore, J. B. (1979) *Optimal Filtering.* Englewood Cliffs: Prentice-Hall.

Aoki, M. and Havenner, A. (1991) State space modelling of multiple time series. *Economic Reviews*, 10, 1-59.

Bartlett, M. S. (1957) A comment on D. V. Lindley's statistical paradox. *Biometrika*, **44**, 533-534.

Bell, W. R. (1984) Signal Extraction for nonstationary time series. *Annals of Statistics*, **13**, 646-664.

Beveridge, S. and Nelson, C. R. (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics,* **7**, 151-174.

Billah, B., King, M. L., Snyder, R. D. and Koehler, A. B. (2005) *Exponential Smoothing Model Selection for Forecasting.* Department of Econometrics and Business Statistics Working Paper, Monash University, Australia.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*. **31**, 307-327.

Box, G. E. P., Jenkins, Jenkins, G. M., and Reinsel, G. C. (1994) *Time Series Analysis, Forecasting and Control, Third Edition*. Englewood Cliffs, NJ: Prentice-Hall.

Brockwell, P. J. and Davis, R. A. (1996) *Introduction to Time Series and forecasting*. Springer-Verlag: New York.

Brown, R. G. (1959) *Statistical Forecasting for Inventory Control*. New York: McGraw-Hill.

Burridge, P. and Wallis, K. (1988) Prediction theory for autoregressive moving average processes. *Econometric Reviews,* **7**, 65-69.

Caines, P. and Mayne, D. (1970) On the discrete time matrix Riccati equation of optimal control. *International Journal of Control,* **12**, 785-794.

Caines, P. and Mayne, D. (1971) On the discrete time matrix Riccati equation of optimal control –a correction. *International Journal of Control*, **14**, 205-207.

Charnes, A. and Cooper, W. (1962) Programming with linear fractional functionals. *Naval Research Logistics Quarterly,* **9**, 181-186.

Chatfield, C., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2001) A new look at models for exponential smoothing. *The Statistician*, **50**, 147-159.

De Jong, P. (1989) Smoothing and interpolation with the state space model. *Journal of the American Statistical Association,* **84**, 1085-1088.

Durbin, J. and Koopman, S.J. (2001) *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Engle, R. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. Inflation. *Econometrica*, **50**, 987-1008.

Forbes, C. S., Snyder, R. D. and Shami, R. G. (2000) *Bayesian Exponential Smoothing*. Department of Econometrics and Business Statistics Working Paper, Monash University, Australia.

Garcia-Ferrer, A. and del Hoyo, J. (1992) On trend extraction models: interpretation, empirical evidence and forecasting performance (with discussion). *Journal of Forecasting*, **11**, 645-665.

Gijbels, I., Pope, A. and Wand, M. P. (1999) Understanding exponential smoothing via kernel regression. *Journal of the Royal Statistical Society, Series B*, **61**, 39-50.

Hannan, E. J. and Diestler, M. (1988) *the Statistical Theory of Linear Systems*. New York: John Wiley and Sons.

Harvey, A. (1989) *Forecasting Structured Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Harvey, A. and Koopman, S. J. (2000) Signal extraction and the formulation of unobserved components models. *Econometrics Journal*, **3**, 84-107.

Harvey, A., Ruiz, E. and Sentana, E. (1992) Unobserved components models with ARCH disturbances. *Journal of Econometrics*, **52**, 129-157.

Hillmer, S. C. and Tiao, G. C. (1982) An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association,* **77**, 63-70.

Holt, C. E. (1957) *Forecasting Seasonals and Trends by Exponentially Weighted Averages*. ONR Memorandum No. 52, Carnegie Institute of Technology, Pittsburgh, USA. Available from the Engineering Library, University of Texas at Austin, 2003. Published in 2004 in the *International Journal of Forecasting*, 20, 1-13, with additional commentaries.

Hyndman, R. J., Akram, M. and Archibald, B. (2003) *The Admissible Parameter Space for Exponential Smoothing Models*. Department of Econometrics and Business Statistics Working Paper, Monash University, Australia.

Hyndman, R. J., Koehler, A. B., Ord, J. K. and Snyder, R. D. (2005) Prediction intervals for exponential smoothing state space models. *Journal of Forecasting,* **24**, 17-37.

Jazwinski, A. H. (1970) *Stochastic Processes and Filtering Theory*. New York: Academic Press.

Kalman, R. E. and Bucy, R. S. (1961) New results in linear filtering and prediction theory. *Transactions, ASME Journal of Basic Engineering,* D83, 95-108.

Koehler A. B., Snyder, R. D., and Ord, J. K. (2001) Forecasting models and prediction intervals for the multiplicative Holt-Winters method. *International Journal of Forecasting* 17, 269-286.

Koopman, S. J. (1997) Exact initial Kalman filtering and smoothing for non-stationary time series models. *Journal of the American Statistical Association,* **92**, 1630-1638.

Lawton, R. (1998) How should additive Holt-Winters estimates be corrected? *International Journal of Forecasting*, **14**, 393-403.

Leeds, M. (2000). *Error Structures for Dynamic Linear Models: Single Source versus Multiple Source.* Ph.D. Thesis, Department of Statistics, The Pennsylvania State University, University Park, PA.

Morley, J. C., Nelson, C. R. and Zivot, E. (2002) *Why are Beveridge-Nelson and Unobserved Component Decompositions of GDP so Different?* Department of Economics Working Paper, University of Washington, Seattle.

Muth, J. R. (1960) Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, **55**, 299-306.

Nerlove, M. and Wage, S. (1964) On the optimality of adaptive forecasting. *Management Science*, **10**, 207-229.

Ord, J. K., Koehler, A. B. and Snyder, R. D. (1997) Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association,* **92**, 1621-1629.

Proietti, T. and Harvey, A. (2000) A Beveridge-Nelson smoother. *Economics Letters*, **67**, 139-146.

Snyder, R. D. (1985) Recursive estimation of dynamic linear statistical models. *Journal of the Royal Statistical Society,* **B47**, 272-276.

Snyder, R. D. (2004) *Exponential Smoothing: A Prediction Error Decomposition Principle*. Department of Econometrics and Business Statistics Working Paper, Monash University, Australia

Snyder, R. D. (2005) *A Pedant's Approach to Exponential Smoothing*. Department of Econometrics and Business Statistics Working Paper, Monash University, Australia.

Snyder, R. D., Ord, J. K. and Koehler, A. B. (2001) Prediction intervals for ARIMA models. *Journal of Business & Economic Statistics*, **19**, 1-9.

Sweet, A.L. (1985) Computing the variance of the forecast error for the Holt-Winters seasonal models. Journal of Forecasting, **4**, 235–243.

Theil, H. and Wage, S. (1964) Some observations on adaptive forecasting. *Management Science*, **10**, 198-206.

West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models, Second Edition*. New York: Springer-Verlag.

Winters, P. R. 1960 Forecasting sales by exponentially weighted moving averages. *Management Science*, **6**, 324-342.