ISSN 1440-771X

MONASH University

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Exponential Smoothing and the Akaike Information Criterion

Ralph D. Snyder and J. Keith Ord

June 2009

Working Paper 4/09

Exponential Smoothing and the Akaike Information Criterion

Ralph D. Snyder
Department of Econometrics and Business Statistics,
Monash University, VIC 3800
Australia.
Email: ralph.snyder@buseco.monash.edu

J. Keith Ord
McDonough School of Business,
Georgetown University,
Washington DC 20057,
United States
Email: ORDK@msb.edu

11 June 2009

JEL classification: C22

Exponential Smoothing and the Akaike Information Criterion

Abstract: Using an innovations state space approach, it has been found that the Akaike information criterion (AIC) works slightly better, on average, than prediction validation on withheld data, for choosing between the various common methods of exponential smoothing for forecasting. There is, however, a puzzle. Should the count of the seed states be incorporated into the penalty term in the AIC formula? We examine arguments for and against this practice in an attempt to find an acceptable resolution of this question.

Keywords: exponential smoothing, forecasting, Akaike information criterion, innovations state space approach.

Introduction

It has been argued (Hyndman et al. 2008) that the exponential smoothing methods of forecasting should be complemented by probabilistic assumptions encapsulated in corresponding linear innovations state space models to enable the derivation of prediction distributions and the use of the principle of maximum likelihood for parameter estimation. Normally a Gaussian distribution is assumed, in which case the maximum likelihood approach is equivalent to the traditional sum of squared errors minimization approach (Holt 1957, 2004) for linear versions of exponential smoothing. The advantage of the likelihood approach, in contrast to the least squares approach, is that it can be extended to other distributions, including the Poisson distribution for count time series (Martin et al. 2008).

A modeling approach to exponential smoothing also admits the possibility of the use of the Akaike information criterion (AIC) for model selection. The Akaike information criterion (Akaike 1974) is defined as

$$\text{AIC} = -2 \log \mathfrak{L}^* + 2p \tag{1}$$

where $\mathfrak{L}^*$ is the maximized value of the likelihood function and p is the number of *free* parameters. $\mathfrak{L}^*$ is an estimator because the unknown true parameter values have been replaced by their maximum likelihood estimates. It can be proven that the asymptotic bias of the estimator of the log-likelihood is p. The AIC is a bias corrected estimate of the log-likelihood scaled by the factor -2. The optimized likelihood alone is a measure of fit which unduly favors models with larger numbers of parameters. The bias factor $2p$ penalizes models on the basis of their parameter count and so encourages the choice of simpler models. This bias factor is called the *penalty*. Note that because of the sign reversal, the aim is to select the model with the lowest AIC.

The AIC has been studied and used with exponential smoothing in three studies.

1. Hyndman et al. (2002) used it with 24 linear and non-linear variations of exponential smoothing to create an automated approach to forecasting. The approach was applied to the time series from the M3 competition (Makridakis & Hibon 2000). It proved to be quite effective compared to the competing approaches, being ranked fourth overall and best for the seasonal time series.

2. Billah et al. (2006) compared a range of approaches, including the AIC, for choosing

between common versions of exponential smoothing. The AIC worked better than other information criteria and even had a slight edge over prediction validation where the mean absolute percentage error (MAPE) is calculated from forecasts against withheld data.

3. Hyndman et al. (2008, Chapter 7) describes a similar study to Billah et al. (2006) with one variation to the way the penalty is calculated.

These three studies used inconsistent interpretations of the penalty term. In Hyndman et al. (2002), p was interpreted literally and set equal to the number of free parameters; in Billah et al. (2006), because exponential smoothing can only be applied if the seed states are treated as fixed and unknown quantities, p was set equal to the number of free seed states; and in Hyndman et al. (2008, Chapter 7) p was set equal to the sum of the number of free parameters and the number of free seed states. The aim of this paper is to thoroughly examine this confusing situation and determine once and for all the appropriate penalty for use with exponential smoothing.

1 Linear Innovations State Space Models

The issue addressed in this paper will be considered only in the context of the linear innovations state space model (Hyndman et al. 2008). Non-linear variations are ignored, although the conclusions of this paper are also applicable to them. All random quantities associated with the state space model are normally distributed. Random variables are denoted with a tilde to distinguish them from fixed quantities. Vectors and matrices are represented by bold characters.

The linear innovations state space model for a time series $\{\tilde{y}_t\}$ is

$$\tilde{y}_t = \boldsymbol{w}'\tilde{\boldsymbol{x}}_{t-1} + \tilde{\varepsilon}_t \tag{2}$$

$$\tilde{\boldsymbol{x}}_t = \boldsymbol{F}\tilde{\boldsymbol{x}}_{t-1} + \boldsymbol{g}\tilde{\varepsilon}_t \tag{3}$$

$$\tilde{\varepsilon}_t \sim \text{NID}(0, \sigma^2) \tag{4}$$

where $\boldsymbol{w}$ is a fixed k-vector, $\tilde{\boldsymbol{x}}_t$ is a random k-vector of states, $\boldsymbol{F}$ is a fixed $k \times k$ transition matrix, $\tilde{\varepsilon}_t$ is a random variable called the innovation, and $\boldsymbol{g}$ is a fixed k-vector. Normally some

or all of $\boldsymbol{w}$,$\boldsymbol{F}$ and $\boldsymbol{g}$ are dependent on a p-vector of parameters $\boldsymbol{\theta}$.

Example 1. *Damped trend exponential smoothing was first introduced by Gardner & McKenzie (1985). A model underpinning a slight variation of their method is*

$$\tilde{y}_t = \tilde{\ell}_{t-1} + \tilde{b}_{t-1} + \tilde{\varepsilon}_t \tag{5}$$

$$\tilde{\ell}_t = \tilde{\ell}_{t-1} + \tilde{b}_{t-1} + \alpha\tilde{\varepsilon}_t \tag{6}$$

$$\tilde{b}_t = \phi\,\tilde{b}_{t-1} + \beta\tilde{\varepsilon}_t. \tag{7}$$

This model contains random state variables: the local level $\tilde{\ell}_t$ and the local drift $\tilde{b}_t$. It also contains parameters: the damping factor ϕ, the smoothing parameters α and β, and the innovations variance σ^2. These equations can be written in matrix form as

$$\tilde{y}_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \tilde{\ell}_{t-1} \\ \tilde{b}_{t-1} \end{bmatrix} + \tilde{\varepsilon}_t$$

$$\begin{bmatrix} \tilde{\ell}_t \\ \tilde{b}_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} \tilde{\ell}_{t-1} \\ \tilde{b}_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tilde{\varepsilon}_t$$

This confirms that the damped trend model conforms to the structure of the state space model (2) to (4).

The linear state space model has a structure reminiscent of a linear regression. The transition equation (3) can be back-solved to period 0. Given that it is linear, the state $\tilde{\boldsymbol{x}}_t$ is a linear function of the seed state vector $\tilde{\boldsymbol{x}}_0$ and the innovations $\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t-1}, \ldots \tilde{\varepsilon}_1$ from the intervening periods. The typical series value $\tilde{y}_t$ depends linearly on the state vector $\boldsymbol{x}_{t-1}$ and so can also be resolved into a linear function of the seed state vector $\tilde{\boldsymbol{x}}_0$ and the innovations $\tilde{\varepsilon}_t, \tilde{\varepsilon}_{t-1}, \ldots \tilde{\varepsilon}_1$. The random n-vector $\tilde{\boldsymbol{y}}$ formed from $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n$ is related to $\tilde{\boldsymbol{x}}_0$ and the innovations by an equation of the form

$$\tilde{\boldsymbol{y}} = \boldsymbol{A}\tilde{\boldsymbol{x}} + \boldsymbol{L}\tilde{\boldsymbol{\varepsilon}} \tag{8}$$

where $\tilde{\boldsymbol{x}}$ is the seed state vector $\tilde{\boldsymbol{x}}_0$, $\boldsymbol{A}$ is a fixed $k \times k$ matrix, $\boldsymbol{L}$ is a *unit* lower triangular matrix and $\tilde{\boldsymbol{\varepsilon}}$ is a random n-vector formed from the innovations $\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_n$. The innovations vector has a mean of $\boldsymbol{0}$ and a variance matrix $\sigma^2\boldsymbol{I}_n$. The matrices $\boldsymbol{A}$ and $\boldsymbol{L}$ depend on the parameters of the state space model.

Equation (8) is reminiscent of a regression where the vector of regression coefficients $\tilde{\boldsymbol{x}}$ is not fixed but random. The distribution of $\tilde{\boldsymbol{y}}$ derives from the distributions of $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{\varepsilon}}$, its mean and variance being given by $\mathbb{E}(\tilde{\boldsymbol{y}}) = \boldsymbol{A}\mathbb{E}(\tilde{\boldsymbol{x}})$ and $\mathbb{V}(\tilde{\boldsymbol{y}}) = \boldsymbol{A}\mathbb{V}(\tilde{\boldsymbol{x}})\boldsymbol{A}' + \sigma^2\boldsymbol{I}$. A problem is that $\tilde{\boldsymbol{x}}$ may have only a partially defined probability distribution.

Assuming that the process has operated from the infinite past, $\tilde{\boldsymbol{x}}$ can be resolved into a linear function of the past innovations $\tilde{\boldsymbol{\varepsilon}}_0, \tilde{\boldsymbol{\varepsilon}}_{-1}, \ldots$. Provided that the coefficients in this linear relationship are absolutely summable, $\tilde{\boldsymbol{x}}$ has a proper probability distribution. For example, the local drift in Equation (7) in the damped trend model can be written as $\tilde{b}_0 = \sum_{j=0}^{\infty} \phi^j \tilde{\varepsilon}_{-j}$. Provided the damping parameter satisfies the condition $1 < \phi < 1$, $\tilde{b}_0$ is normally distributed with a mean of 0 and a variance $\sigma^2/(1-\phi^2)$. However, when $\phi = 1$, the variance of $\tilde{\boldsymbol{x}}$ is undefined despite the fact that the resulting model makes sense and in fact underpins the Holt method of trend corrected exponential smoothing.

The moments of the density $p(\boldsymbol{y})$ of a time series can be derived from the moments of $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{\varepsilon}}$ provided that the distribution of $\tilde{\boldsymbol{x}}$ is known. Since most business and economic time series contain a unit root (Nelson & Plosser 1982), at least one of the states associated with a series has to be non-stationary, so in most applications the distribution of $\tilde{\boldsymbol{x}}$ is not completely known. There are typically two resolutions of this problem.

1. Assume that the offending states have an infinite variance (Ansley & Kohn 1985) and use a suitably adapted variant of the Kalman filter such as a fast Givens filter (Snyder & Saligari 1996). It can be established that the effect of this, when there are k unit roots, is to base the analysis on the conditional distribution $p(y_{k+1}, \ldots, y_n | y_1, \ldots, y_k)$.

2. Condition the offending seed vector on a fixed but unknown value $\boldsymbol{x}$ and effectively treat its elements as additional parameters.

Ignoring the technical details of the arguments, the main message to take away is that unit roots make it necessary to condition on something, one possibility being to condition on an initial run of series values, another being to condition on seed states with infinite variances.

2 Exponential Smoothing Approach

In both cases stationary states can be the cause of difficulties. The damped drift has a finite variance provided the damping factor does not equal 1. As soon as $\phi = 1$, a second unit root emerges, the consequence being that the variance of the drift becomes infinite. An analysis based on the conditional distribution $p(y_2,\ldots,y_n|y_1)$ must suddenly switch to an analysis based on the conditional distribution $p(y_2,\ldots,y_n|y_1,y_2)$ in the case of the first option. It is this basic discontinuity problem that makes one question whether it really makes sense to incorporate information about the process from periods prior to the sample period into the likelihood function. It is for this reason Hyndman et al. (2008) recommend ignoring 'prior' information by treating *all* seed states as fixed but unknown parameters, or in mathematical terms set $\tilde{\boldsymbol{x}}_0 = \boldsymbol{x}$ where $\boldsymbol{x}$ is a fixed but unknown vector. The likelihood function is then formed from the distribution $p(y_1,\ldots,y_n|x)$ and is maximized with respect to both the parameters and the seed states. Another advantage is that the relatively complex fast Givens filter can be replaced by exponential smoothing to generate the errors needed for evaluating the likelihood function. More specifically, the most general linear form of exponential smoothing relies on an application of the equations:

$$\varepsilon_t = y_t - \boldsymbol{w}'\boldsymbol{x}_{t-1} \tag{9}$$

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{g}\varepsilon_t \tag{10}$$

This results in a one-to-one transformation between the original series and the innovations. In other words, the innovations are an alternative way of summarizing all pertinent information about the original series. The distribution of the series can be rewritten in terms of the innovations as $p(y_1,\ldots,y_n|x) = p(\varepsilon_1,\ldots,\varepsilon_n)$. The reason for undertaking this transformation is that the independence of the innovations allows us to rewrite the distribution of the series as a product of univariate distributions as follows:

$$p(y_1,\ldots,y_n|x) = p(\varepsilon_1)\ldots,p(\varepsilon_n). \tag{11}$$

The likelihood function can then be written in terms of the innovations as

$$\mathfrak{L}(\boldsymbol{x},\boldsymbol{\theta}) = (2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\varepsilon_t^2\right). \tag{12}$$

The primary function of an exponential smoothing algorithm is to calculate the errors necessary for the evaluation of this likelihood function.

3 The AIC Penalty

The AIC appears to be a good criterion for model selection (Hyndman et al. 2008, Chapter 7). In this study the penalty was determined from both the free parameters and the free seed states. The question remains whether it is legitimate to treat the seed states as parameters. We now examine the arguments both ways.

3.1 Arguments against including the seed state count

1. The seed states are random quantities and should be integrated out of the likelihood function. It is an inappropriate tactic to condition on fixed and unknown values of them.

2. When certain stability conditions akin to the invertibility conditions for Box-Jenkins models are satisfied, the prediction distributions effectively become independent of the seed states and so a model's capacity to predict should not be gauged with a criterion that depends on their count. For example, the unknown future series value $\tilde{y}_{n+1}$ for a time series governed by a local level state space model that underpins simple exponential smoothing is related the seed level $\tilde{\ell}_0$ by the equation $\tilde{y}_{n+1} = \delta^n \tilde{\ell}_0 + \alpha \sum_{j=0}^{n-1} \delta^j y_{n-j} + \tilde{\varepsilon}_{t+1}$ where α is the smoothing parameter and $\delta = 1 - \alpha$. Its mean, the point prediction, is the usual exponentially weighted average of past series values and the seed level when $-1 < \delta < 1$. The actual series value, which remains to be observed, deviates from the mean by the innovation $\tilde{\varepsilon}_{n+1}$. The seed state is discounted by the factor δ^n. Its effect on this future series value is negligible for moderate sample sizes.

3.2 Arguments for including the seed state count

1. When applying exponential smoothing, the seed states are treated as fixed quantities and Equation (8) becomes a conventional generalized regression. It is then necessary to fall into step with theory of the AIC for regression analysis and include them in the

count.

2. As the number of states increase, so does the *complexity* of a model. In the case of monthly time series, it may be necessary to have twelve states to represent a seasonal cycle, as is the case in the model underpinning the Winters method. However, a more parsimonious possibility is a model with only two seasons (low and high), that is inherently simpler than the usual twelve state version. If the seed states count is not included, both models would employ the same penalty [1] and the model with 12 seasons would always be favored.

4 Confutations

We examine each of the above arguments in turn.

4.1 Random seed states...

By construction most states are random, but does this mean they should not be counted in the penalty? Consider, for example, the local level model:

$$\tilde{y}_t = \tilde{\ell}_{t-1} + \tilde{\varepsilon}_t \tag{13}$$

$$\tilde{\ell}_t = \tilde{\ell}_{t-1} + \alpha\tilde{\varepsilon}_t \tag{14}$$

This model, which provides the statistical foundations for simple exponential smoothing, has the two parameters α and σ^2. If the seed level is ignored because of its randomness, the penalty for the AIC is 4. If it is counted, then the penalty is 6.

If the seed states are to be treated as random variables, it is not then possible to use exponential smoothing: the associated recursions must be started with fixed seed states. If random seed states are to be retained it is necessary to resort to a Kalman filter to generate the errors necessary for likelihood evaluation. The existence of non-stationary states muddies the waters because the associated seed states then have infinite variances and their distributions are degenerate. The Kalman filter does not work in this context so it is necessary to resort to an algorithm such as the fast Givens filter (Snyder & Saligari 1996) where the associated formulae

[1] The seed seasonal effects are restricted to sum to 0. If there are twelve seasons, there are only eleven *free* seed seasonal effects; if there are two seasons, there is only one *free* seasonal effect.

have well defined limiting forms in the presence of the seed states with infinite variances. By ignoring the randomness of the seed states and conditioning on fixed but unknown values of them, it is possible to apply exponential smoothing in place of the more complex fast Givens filter, and hence greatly simplify the estimation process. There is a price paid for this tactic, it being the loss of efficiency of the maximum likelihood estimators because information about the process prior to the sample period is ignored. However, this price cannot be particularly high if we are primarily interested in forecasting, because prediction distributions are effectively unaffected by the seed states in moderate to large samples provided the invertibility conditions, needed for the discounting of old data, are satisfied.

4.2 Long run independence of the seed state

We have just indicated that provided the invertibility condition is satisfied by the parameters, in moderate samples, the seed state has very little impact on the prediction distribution. It suggests that maybe our conclusion in the previous sub-section is wrong: maybe the seed state count should be excluded from the penalty?

4.3 Consistency with the AIC for regression

The state space models here are linear in the states but non-linear in the parameters. It is possible to do a first-order linear Taylor's expansion of the non-linear components of the model around the maximum likelihood values of the parameters to eventually emerge with a model of the form (8) where $\tilde{\boldsymbol{y}}$ is now a transformation of the original series and $\tilde{\boldsymbol{x}}$ contains both the seed states and the parameters. This approximate model is a linear regression provided we replace the random elements of $\tilde{\boldsymbol{x}}$ by fixed but unknown values as is done with exponential smoothing. It has been established that the penalty of the AIC for a regression model should be based on the number of regression coefficients (length of $\tilde{\boldsymbol{x}}$) and the variance. Since the vector $\tilde{\boldsymbol{x}}$ is constructed from both the seed states and the parameters, this argument implies that the seed states should be counted.

This, however, is inconsistent with the argument presented above about the diminishing importance of the seed states. If we pre-multiply the regression equation (8) by L^{-1} we obtain a classical regression

$$\tilde{\boldsymbol{y}}^* = \boldsymbol{A}^* \boldsymbol{x} + \tilde{\boldsymbol{\varepsilon}} \tag{15}$$

where $\tilde{\boldsymbol{y}}^* = L^{-1}\tilde{\boldsymbol{y}}$ and $\boldsymbol{A}^* = L^{-1}\boldsymbol{A}$. It may be established (Snyder 1985) that this action is equivalent to applying exponential smoothing, and that as a consequence, the elements of $\boldsymbol{A}^*$ in columns corresponding to the seed states converge to zero. The transformed regression equations eventually become independent of the seed states. It then does not appear to make sense to count them.

4.4 Complexity of seasonal models

It was argued that for monthly data, a model with two seasons is simpler than a model with twelve seasons, so the number of seasonal effects in the state variable of a seasonal state space model reflects the level of the complexity. This is an argument in favor of including the number of seed states in the penalty. However, in the two season model, dummy variables must be used to select the season that is appropriate to the period under consideration. More specifically, the two season model based on a local level model with drift b is given by

$$\begin{aligned}
\tilde{y}_t &= \tilde{\ell}_{t-1} + b + z_{1t}\tilde{s}_{1t-1} + z_{2t}\tilde{s}_{2t-1} + \tilde{\varepsilon}_t \\
\tilde{\ell}_t &= \tilde{\ell}_{t-1} + b + \alpha\tilde{\varepsilon}_t \\
\tilde{s}_{it} &= \tilde{s}_{i,t-1} + z_{it}\gamma\tilde{\varepsilon}_t \ (i = 1, 2)
\end{aligned}$$

where the z_{it} are seasonal dummy variables. These dummy variables represent additional information that is not needed in the twelve season model. They suggests that the number of states and parameters are not the sole determinant of the level of complexity of a model.

Conclusions

In the end it seems sensible to use the AIC defined as

$$\text{AIC} = -2log\mathfrak{L}^* + 2(p + k - r)$$

where r is the of *active* constraints in the optimal solution. The parameter count p should include the unknown smoothing parameters, damping factors, and the variance σ^2. Despite the fact that prediction distributions are usually independent of the seed states, the seasonal example suggests that the count of the unknown seed states should not be ignored. It also motivates the adjustment for the active restrictions adjustment.

One interesting point is that no restrictions are placed on the use of the AIC. The various versions of exponential smoothing encompass approaches such as simple exponential smoothing (Brown 1959) which implicitly involves a unit root and trend corrected exponential smoothing (Holt 1957) which involves a double unit root. We advocate the use of the AIC to choose between them. This can be contrasted with the Box-Jenkins approach (Box et al. 1994) where unit root tests (Dickey & Fuller 1981) rather than the AIC are advocated for determining the appropriate level of differencing for a time series. The interesting question is whether the Box-Jenkins approach can be adapted to forgo the use of unit root tests in favor of the AIC? It is this question that awaits further research.

References

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**, 716–723.

Ansley, C. F. & Kohn, R. (1985), 'Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions', *The Annals of Statistics* **13**, 1286–1316.

Billah, B., King, M. L., Snyder, R. D. & Koehler, A. B. (2006), 'Exponential smoothing model selection for forecasting', *International Journal of Forecasting* **22**, 239–247.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994), *Time series analysis: forecasting and control*, 3rd edn, Prentice-Hall, Englewood Cliffs, New Jersey.

Brown, R. G. (1959), *Statistical forecasting for inventory control*, McGraw-Hill, New York, USA.

Dickey, D. A. & Fuller, W. A. (1981), 'Likelihood ratio statistics for autoregressive time series with a unit root', *Econometrica* **49**, 1057–1071.

Gardner, Jr, E. S. & McKenzie, E. (1985), 'Forecasting trends in time series', *Management Science* **31**(10), 1237–1246.

Holt, C. C. (1957), Forecasting trends and seasonals by exponentially weighted averages, O.N.R. Memorandum 52/1957, Carnegie Institute of Technology.

Holt, C. C. (2004), 'Forecasting seasonals and trends by exponentially weighted moving averages', *International Journal of Forecasting* **20**, 5–10.

Hyndman, R. J., Koehler, A. B., Ord, J. K. & Snyder, R. D. (2008), *Forecasting with exponential smoothing: the state space approach*, Springer.
URL: *http://www.robhyndman.info/expsmooth/*

Hyndman, R. J., Koehler, A. B., Snyder, R. D. & Grose, S. (2002), 'A state space framework for automatic forecasting using exponential smoothing methods', *International Journal of Forecasting* **18**(3), 439–454.
URL: *http://www.robhyndman.info/papers/hksg.htm*

Makridakis, S. & Hibon, M. (2000), 'The M3-competition: Results, conclusions and implications', *International Journal of Forecasting* **16**, 451–476.

Martin, G. P., Gould, P., Feigin, P. & Snyder, R. D. (2008), 'Feasible parameter regions for alternative discrete state space models', *Statistics and Probability Letters,* **78**, 2963–2970.

Nelson, C. R. & Plosser, C. I. (1982), 'Trends and random walks in macroeconomic time series', *Journal of Monetary Economics* **10**, 132–162.

Snyder, R. D. (1985), Estimation of a dynamic linear model - another approach, Working Paper 15/85, Department of Econometrics and Operations Research, Monash University.

Snyder, R. D. & Saligari, G. (1996), 'Initialisation of the Kalman filter with partially diffuse initial conditions', *Journal of Time Series Analysis* **17**, 409–424.