*Austral. & New Zealand J. Statist.* **42**(1), 2000, 1–23

# THE FOREMAN LECTURE: THE STATE SPACE APPROACH TO TIME SERIES ANALYSIS AND ITS POTENTIAL FOR OFFICIAL STATISTICS

(with Discussion)

J. DURBIN[1]

## 1. Introduction

It is a great honour for me to be invited to give this lecture. I did not know Ken Foreman well personally, but I know about his impressive achievements in improving the methodology in Australian official statistics, and by giving this lecture I am happy to pay tribute to his memory.

I have some reservations about my qualifications for the task before me. I am not and never have been an official statistician. I am an academic who has spent most of his time over the past 50 years working on problems which have very little to do with official statistics. However, I have worked with British Government statisticians occasionally over the years, and recently I have had the good fortune to pay visits to the US Bureau of the Census, Statistics Canada and Statistics New Zealand amounting to over a year altogether, so I cannot deny that I have some knowledge of the official statistical world.

In the first part of this lecture, I compare the state space approach to time series analysis with the Box–Jenkins ARIMA approach. After defining what I mean by the state space and Box–Jenkins approaches, I go back to the historical origins of modern applied time series analysis in exponential smoothing in the 1950s and try to show that both systems can be regarded as evolving naturally from these origins. I then make a wide-ranging comparison of the relative merits of the two systems in the course of which I come down strongly in favour of the state space approach. After this, I briefly describe some recent work by myself and others on non-Gaussian and nonlinear state space models for time series analysis, and I conclude with some speculative comments on the potential that state space methods might have for time series work in official statistics.

The underlying idea behind state space time series modelling is to model separately the various components of the series and to put these sub-models together into a single linear model. The general form of this model is

$$y_t = Z_t\alpha_t + \epsilon_t\,, \qquad \epsilon_t \sim NID(0, H_t) \tag{1}$$

$$\alpha_t = T_t\alpha_{t-1} + \xi_t\,, \qquad \xi_t \sim NID(0, Q_t) \tag{2}$$

for $t = 1, \dots, n$. In equation (1), $y_t$ is the observation, which can be multivariate, $Z_t$ is a known matrix, $\alpha_t$ is an unobserved vector called the state vector, and $\epsilon_t$ is an unobserved error term. It can be seen that if $\alpha_t$ were constant over time, (1) would just be an ordinary linear regression model. Equation (2) shows that the coefficient vector $\alpha_t$ varies over time according to a first-order vector autoregression. Here $T_t$ is a known matrix and $\xi_t$ is an unobserved

Received June 1997; revised September 1999; accepted October 1999.
[1] 31 Southway, London NW11 6RX, UK. e-mail: durbinja@aol.com
*Acknowledgment.* This lecture was written while the author was a Research Fellow at Statistics New Zealand.

error. While the model is conceptually simple, it is very general and includes many other models as special cases, including Box–Jenkins models. The elements of $\alpha_t$ represent the various components of the series, such as trend, seasonal and cyclic effects and the effects of explanatory variables and interventions. A very simple example of a state space model is the random walk plus noise model

$$y_t = \mu_t + \epsilon_t , \qquad \mu_t = \mu_{t-1} + \eta_t . \tag{3}$$

Although state space models are well known to specialists, they are only coming into use gradually in applied work in statistics. Most applied time series analysis over the past 20 years has been based on the Box–Jenkins ARIMA (BJ) approach. Typically, this takes a univariate observation $y_t$ as made up of trend, seasonal effects and an irregular term. The basic idea behind BJ is to eliminate the trend and seasonal effects at the outset of the analysis by differencing, and to treat the differenced series as a stationary time series. Let $\Delta y_t = y_t - y_{t-1}$, $\Delta^2 y_t = \Delta(\Delta y_t)$, $\Delta_s = y_t - y_{t-s}$, $\Delta_s^2 y_t = \Delta_s(\Delta_s y_t)$, and so on, where we are assuming that we have $s$ 'months' per 'year'. In BJ, we continue differencing until trend and seasonal effects have been eliminated, giving a new variable $z_t = \Delta^d \Delta_s^D y_t$ for $d,\ D = 0, 1, \ldots;$ we then model this as a stationary autoregressive-moving average series. A very simple example of a BJ model is

$$\Delta y_t = \nu_t - \theta \nu_{t-1} , \qquad \nu_t \sim NID(0, \sigma_\nu^2). \tag{4}$$

Later, I make some general comparisons between the two approaches. On the face of it, they seem conceptually very different. However, before making these comparisons, I go back to the historical origins of modern time series analysis in exponential smoothing in the 1950s. By doing so, I aim to show that the two approaches can be thought of as evolving from a common origin.

## 2. Historical origins of modern time series analysis

The exponentially weighted moving average (EWMA) was introduced in the 1950s for one-step-ahead forecasting of $y_{t+1}$ given a time series $y_t, y_{t-1}, \ldots$. This has the form

$$\hat{y}_{t+1} = (1-\lambda) \sum_{j=0}^{\infty} \lambda^j y_{t-j} \qquad (0 < \lambda < 1). \tag{5}$$

From (5), we deduce immediately the recursion

$$\hat{y}_{t+1} = (1-\lambda) y_t + \lambda \hat{y}_t , \tag{6}$$

which is used in place of (5) for practical computation. This has a simple structure and requires little storage, so it was very convenient for the primitive computers available in the 1950s. As a result, EWMA forecasting became very popular in industry, particular for sales forecasting of many items simultaneously.

An important contribution was made by Muth (1960) who showed that EWMA forecasts produced by the recursion (6) are minimum mean square error (MMSE) forecasts in the sense that they minimize $\mathrm{E}(\hat{y}_{t+1} - y_{t+1})^2$ for observations $y_t, y_{t-1}, \ldots$ generated by the model

$$y_t = \mu_t + \epsilon_t , \qquad \mu_t = \mu_{t-1} + \eta_t , \tag{7}$$

where $\epsilon_t$ and $\eta_t$ are independent white noise terms, i.e. serially independent random variables with zero means and constant variances. This is the simple state space model (3).

Taking first differences of observations $y_t$ generated by (7) gives

$$\Delta y_t = y_t - y_{t-1} = \epsilon_t - \epsilon_{t-1} + \eta_t \,.$$

Since $\epsilon_t$ and $\eta_t$ are serially uncorrelated the lag-1 autocorrelation of $\Delta y_t$ is non-zero, but all higher autocorrelations are zero. This is the autocorrelation function of the MA(1) model, i.e. the moving average model of order 1,

$$\Delta y_t = \nu_t - \theta \nu_{t-1} \,,$$

which is the simple BJ model (4).

We observe that these two simple forms of the state space and BJ models produce the same one-step forecasts and that these can be calculated by the EWMA (6) which has proven practical value. We can rewrite this in the form

$$\hat{y}_{t+1} = \hat{y}_t + (1-\lambda)(y_t - \hat{y}_t),$$

which is the Kalman filter form for general state space forecasts specialized to this simple case.

The EWMA model was extended by Holt (1957) and Winters (1960) (HW) to series containing trend and seasonal effects. The extension for trend in the additive case is

$$\hat{y}_{t+1} = m_t + b_t \,,$$

where $m_t$ and $b_t$ are level and slope terms generated by the EWMA type recursions

$$m_t = (1-\lambda_1)y_t + \lambda_1(m_{t-1} + b_{t-1}), \qquad b_t = (1-\lambda_2)(m_t - m_{t-1}) + \lambda_2 b_{t-1} \,.$$

In an interesting extension of Muth's (1960) results, Theil & Wage (1964) showed that the forecasts produced by these HW recursions are MMSE forecasts for the state space model

$$y_t = \mu_t + \epsilon_t \,, \qquad \mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \,, \qquad \beta_t = \beta_{t-1} + \zeta_t \,. \tag{8}$$

This extends the random walk plus noise model (7) by adding a slope term $\beta_t$ which enables the model to follow a time-varying trend which is locally linear.

Taking second differences of $y_t$ generated by (8), we obtain

$$\Delta^2 y_t = \zeta_t + \eta_t - \eta_{t-1} + \epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2} \,.$$

This is a stationary series with non-zero autocorrelations at lags 1 and 2 but zero autocorrelations elsewhere. It therefore follows the MA(2) model

$$\Delta^2 y_t = \nu_t - \theta_1 \nu_{t-1} - \theta_2 \nu_{t-2} \,,$$

which is a simple form of the BJ model.

Now consider what happens when seasonality is added to the picture. Starting with the random walk plus noise model (3), add a seasonal term $\gamma_t$ giving

$$y_t = \mu_t + \gamma_t + \epsilon_t \,.$$

If the seasonal pattern were constant over time, the $\gamma_t$s would satisfy the condition $\gamma_t + \gamma_{t-1} + \cdots + \gamma_{t-s+1} = 0$, where $s$ is the number of 'months' per 'year'. To allow the seasonal pattern to change over time, we add a white noise term $\omega_t$ and obtain the structural model

$$y_t = \mu_t + \gamma_t + \epsilon_t, \qquad \mu_t = \mu_{t-1} + \eta_t, \qquad \gamma_t = -\gamma_{t-1} - \cdots - \gamma_{t-s+1} + \omega_t \,. \tag{9}$$

This satisfies the general state space forms (1) and (2).

Now take first differences and first seasonal differences of (9). We find

$$\Delta\Delta_s y_t = \eta_t - \eta_{t-s} + \omega_t - 2\omega_{t-1} + \omega_{t-2} + \epsilon_t - \epsilon_{t-1} - \epsilon_{t-s} + \epsilon_{t-s-1} \,,$$

which is a stationary time series with non-zero autocorrelations at lags $1, 2, s-1, s$ and $s+1$. Consider the BJ model, which in expanded form is

$$\Delta\Delta_s y_t = \nu_t - \theta_1 \nu_{t-1} - \theta_s \nu_{t-1} + \theta_1\theta_s \nu_{t-s-1} \,.$$

This is the famous 'airline model' of Box and Jenkins which has been a good fit to many economic time series containing trend and seasonal effects. It has non-zero autocorrelations at lags $1, s-1, s$ and $s+1$. Now the autocorrelation at lag 2 from model (9) arises only from $\mathrm{var}(\omega_t)$ which in most cases in practice is small. Thus when we add seasonal effects to the models we find again a close correspondence between the state space and the BJ models. A slope term $\beta_t$ can be added to (9) as in (8) without affecting the conclusions.

A pattern is now emerging. Starting with EWMA forecasting, which in appropriate circumstances has been found to work well in practice, we have found that there are two distinct types of models, the state space models and the BJ models, which appear to be very different conceptually but which both give MMSE forecasts from EWMA recursions. The explanation is that when the time series has an underlying structure which is sufficiently simple, the appropriate state space and BJ models are essentially equivalent. It is when we move towards more complex structures that the differences emerge. In the next section I therefore compare the state space and BJ approaches to a broader range of problems in time series analysis.

## 3. Comparison of the state space and Box–Jenkins approaches to time series analysis

The early development of state space methodology took place in the field of engineering rather than statistics, starting with Kalman's (1960) pathbreaking paper. In this paper, Kalman did two crucially important things. He showed that a very wide class of problems could be encapsulated in a simple linear model, essentially the state space model (1) and (2) above. Secondly he showed how, due to the Markovian nature of the model, the calculations needed for practical application of the model could be set up in recursive form in a way that was particularly convenient on a computer. A huge amount of work was done subsequently in development of these ideas in the engineering field.

During the 1960s to the early 1980s, contributions to state space methodology from statisticians were isolated and sporadic. In this paper I do not wish to attempt a historical review of these contributions. The names Jones, Akaike, Duncan, Harrison, Stevens, Kitagawa, Gersch and Harvey are important. Andrew Harvey deserves particular mention, not only for his research contributions to the field but also for his book (Harvey, 1989) and his leadership in the STAMP (Structural Time series Analysis Modeller and Predictor) software project that I refer to later. Harvey (1989) describes the state of the art in state space time series analysis as it was around 1988, and also provides the historical references that I am omitting.

The key advantage of the state space approach is that it is based on a structural analysis of the problem. The various components that make up the series, such as trend, seasonal, cycle and calendar variations, together with the effects of explanatory variables and interventions, are modelled separately before being put together in the state space model. It is up to the investigator to identify and model any features in particular situations that require special treatment. In contrast, BJ is a 'black box' approach in which the model adopted depends purely on data without prior analysis of the structure of the system that generated the data.

A second advantage of state space models is that they are very flexible. Because of the recursive nature of the models and of the computational techniques used to analyse them, it is straightforward to allow for known changes in the structure of the system over time. In contrast, BJ models are homogeneous through time because they are based on the assumption that the differenced series is stationary.

State space models are very general. They cover a very wide range, including all BJ models. Multivariate observations can be handled by straightforward extensions of univariate theory, which is not the case with BJ models.

It is easy to allow for missing observations with state space models. Explanatory variables can be incorporated into the model without difficulty. Moreover, the associated regression coefficients can be permitted to vary stochastically over time if this seems to be called for in the application. Trading-day adjustments and other calendar variations can readily be taken care of.

Because of the Markovian nature of state space models, the calculations needed to implement them can be put in recursive form. This enables increasingly large models to be handled effectively without disproportionate increases in the computational burden.

No extra theory is required for forecasting. All that is needed is to project the Kalman filter forward into the future. This gives the forecasts required and their estimated standard errors with the standard formulae used earlier in the series.

If these are all the advantages of state space modelling, what are the disadvantages relative to the BJ approach? In my opinion, the only disadvantages are the relative lack in the statistical community of information, knowledge and software regarding these models. Box–Jenkins modelling forms a core part of university courses on time series analysis and there are numerous text books on the subject. Software is widely available in major general packages such as SAS, BMDP and MINITAB as well as in many specialist time series packages. In contrast, state space modelling for time series is taught in relatively few universities, and on the statistical side, as distinct from the engineering side, there are only two major text books, those of Harvey (1989) and West & Harrison (1989). There is no state space software on the general statistical packages, and there are only a few specialist packages such as STAMP (Structural Time series Analysis Modeller and Predictor) of Koopman *et al.* (1995).

Now consider some of the disadvantages of the BJ approach. The elimination of trend and seasonal effects by differencing may not be a drawback if forecasting is the only object of the analysis, but in many contexts, particularly in official statistics, knowledge about these components has intrinsic importance. It is true that estimates of trend and seasonal effects can be ‘recovered’ from the differenced series by maximizing the residual mean square, but this seems an artificial procedure which is not as appealing as modelling the components directly.

The requirement that the differenced series should be stationary is a weakness of the theory. In the economic and social fields, real series are never stationary, however much differencing is done. The investigator has to face the question, how close to stationarity is close enough? This is a hard question to answer. There is another point which is intimately related to it. At the outset the observations were assumed to be made up of trend, seasonal and irregular terms. The irregular term could be supposed to behave according to some stochastic model. I find it too remarkable a coincidence that the amount of differencing that is required to eliminate the trend and seasonal term is exactly the same as is needed to provide a nice stationary autoregressive-moving average model for the differenced irregular term.

In the BJ system it is relatively hard to handle matters such as missing observations, adding explanatory variables, calendar adjustments and changes in behaviour over time. These are straightforward to deal with in the state space system.

In practice it is found that the airline model and similar BJ models fit many datasets quite well, but I have already explained the reason for this, namely that they are approximately equivalent to plausible state space models. This point is discussed at length by Harvey (1989 pp. 72, 73). As we move away from airline-type models, the model identification process in the BJ system becomes difficult to apply. The main tool is the sample autocorrelation function which is notoriously imprecise because of its high sampling variability. Examples can be found in which the data appear to be explained equally well by models whose specifications look very different.

A final point in favour of structural models is their transparency. One can examine graphs of trend, seasonal and other components to check whether they behave in accordance with expectation. If, for example, a blip were found in the graph of the seasonal component, one could go back to the original data to trace the source and perhaps make an adjustment to the model accordingly.

To sum up, state space models are based on modelling the observed structure of the data. They are more general, more flexible and more transparent than BJ models. They can deal with features of the data which are hard to handle in the BJ system. Software is available in the STAMP package for state space time series analysis.

## 4. Recent work on SS time series modelling

In this section, I wish to discuss briefly some work on state space modelling that I and others have been engaged in recently, which has some bearing on the potential for official statistics that I consider in the next section.

First I describe some joint work with Benoit Quenneville on the benchmarking problem (Durbin & Quenneville, 1997). Benchmarking is an important problem in official statistics. We have a series of monthly or quarterly observations (for simplicity, say monthly) obtained from sample surveys. These are subject to survey bias and sampling errors. In addition, we have a series of annual values obtained from administrative data, censuses or annual surveys which are regarded as more accurate than the monthly values. The annual values can be either yearly totals or values at a particular month. The problem is to adjust the observations so that the monthly and annual values are consistent. With one exception, we do not discuss earlier work on the problem: for references see Durbin & Quenneville (1997). The exception is

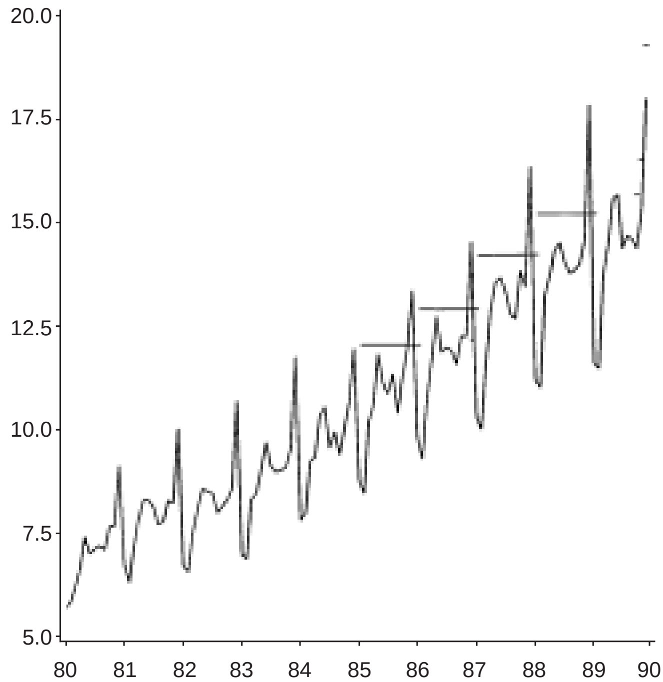


Figure 1. Canadian retail sales series from January 1980 to December 1989 and average values of the benchmarks, graphed as horizontal lines. Values are expressed in millions.

Hillmer & Trabelsi (1987). They developed a two-stage procedure in the first stage of which they fitted a BJ model to the survey data while in the second they added in the annual data.

For reasons given earlier in this lecture, Quenneville and I thought it would be useful to find a state space solution for this problem. By using the state space approach we were able to solve a number of ancillary problems. For example, we were able to allow for trading-day and leap-year effects in a straightforward manner. The trading-day effect reflects the number of Mondays, Tuesdays, etc., in particular months which was important for the retail trade data analysed in the paper. A second problem was that survey data were known to be biased by non-response and other factors. Assuming that the annual values were bias-free, we incorporated a bias coefficient into the model and were thus able to estimate the survey bias. A third problem was that economic data often behave multiplicatively. It is usual to take logs before analysing the data. On the other hand, the addition of monthly values to obtain annual totals is a linear operation. The combination of multiplicative data and annual totals in the same model results in a model that is nonlinear. By estimating the mode of the state space vector given all the observations, monthly and annual, Quenneville and I were able to fit the model, thus ensuring that monthly values satisfied the annual constraints.

Figures 1 and 2, from Durbin & Quenneville (1997), show the effect of applying our methods to Canadian retail sales 1980–89. Figure 1 shows the original series together with seven benchmarks, displayed as horizontal lines; the first four are annual totals and the last three are particular monthly values. Two points are obvious about this diagram. First, the amplitude increases with level so taking logs is indicated; and second, the survey values have a substantial downward bias. Figure 2 shows the effects of benchmarking by the various

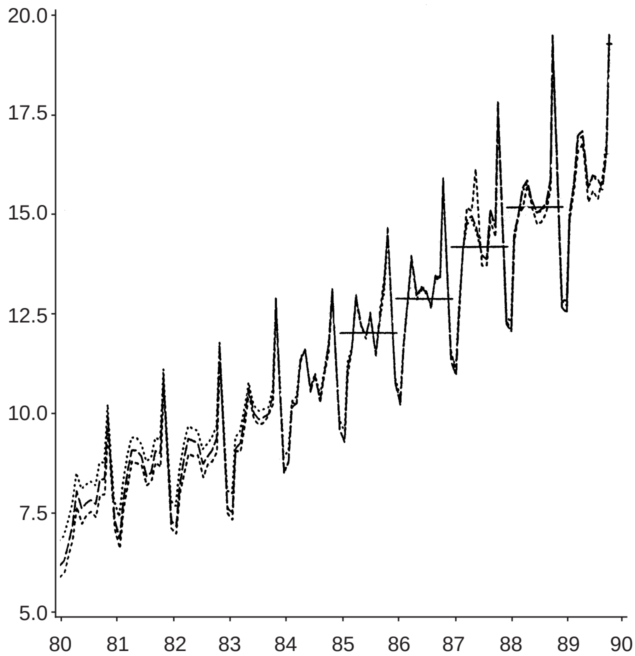


Figure 2. Benchmarked values for the Canadian retail sales series and average values of the benchmarks: Additive Without Bias (AWB) (– – –); Multiplicative with Bias (MB) (- - -); Additive with Bias (AB) $(\cdots)$. Values are expressed in millions.

combinations, additive model with and without bias and multiplicative model with and without bias. Although each combination satisfies the benchmarks it is clear that the best model is the multiplicative model with bias adjustment. Details are given in Durbin & Quenneville (1997).

With relatively minor qualifications, most of the state space techniques discussed in the previous section were based essentially on the assumption that the state space model is Gaussian and linear. In the real world, many observations are far from normal and nonlinearities occur such as the one in the benchmarking problem described above. There is therefore a need for development of techniques to deal with these cases. I now briefly describe some recent work I have done jointly with S.J. Koopman that we believe provides a comprehensive state space method for dealing with non-Gaussian observations. Nonlinear models can be dealt with in a very similar way. My remarks are based on Durbin & Koopman (1997, 2000).

Let $y = [y_1^\top, \dots, y_n^\top]^\top$ and $\alpha = [\alpha_1^\top, \dots, \alpha_n^\top]^\top$ for a series of length $n$. The core of a state space analysis is the estimation of $\alpha$ given the observations $y$; in some cases the estimation of the mean square error matrix of the estimate of $\alpha$ is also important. For the standard linear Gaussian model, $\alpha$ is estimated by the mean of the conditional distribution of $\alpha$ given $y$, $\bar{\alpha} = \mathrm{E}(a \mid y)$. This estimation is efficiently performed by the Kalman filter and smoother (KFS). When the observations are non-Gaussian, analytical calculation of $\mathrm{E}(\alpha \mid y)$ is intractable. However, a suitable alternative is to estimate $\alpha$ by the mode $\hat{\alpha}$ of its conditional distribution given $y$. This is the solution of the vector equation $\partial \log p(\alpha \mid y)/\partial\alpha = 0$ which is nonlinear. Koopman and I showed how the equation can be solved iteratively by linearizing it at each iterative step and solving the linearized equation by KFS, thus obtaining a new estimate. These calculations turn out to be fast, so if parameters of the state space model are known, there is no difficulty whatever in estimating $\hat{\alpha}$ as accurately as desired. However, estimation of the parameters is a hard problem because an analytical expression for the likelihood cannot be obtained.

Rather than use approximate methods for parameter estimation whose accuracy relative to true maximum likelihood cannot be guaranteed, we decided to use a simulation technique in which accuracy can be measured and in which estimates can be made that are as close to the true maximum likelihood estimates as is desired. Consider a model in which the state vector $\alpha_t$ is determined by the linear Gaussian form (2) but $y_t$ has a non-Gaussian distribution depending on $Z_t\alpha_t$. Denote the likelihood by $L$ and let $\theta_t = Z_t\alpha_t$ with $\theta = [\theta_1^\top, \dots, \theta_n^\top]^\top$. Denote non-Gaussian marginal, joint and conditional densities by $p(\cdot)$, $p(\cdot,\cdot)$ and $p(\cdot \mid \cdot)$ and Gaussian marginal, joint and conditional densities by $g(\cdot)$, $g(\cdot,\cdot)$ and $g(\cdot \mid \cdot)$. Then

$$L = p(y) = \int p(y,\theta)\, d\theta = \int p(y \mid \theta) g(\theta)\, d\theta. \tag{10}$$

We have $g(\theta)$ not $p(\theta)$ in (10) because the density of $\alpha$ and hence $\theta$ is Gaussian.

Now consider a Gaussian conditional density $g(y \mid \theta)$ which we use as an approximation to $p(y \mid \theta)$ in the simulation. Together with $g(\theta)$, this defines an approximating Gaussian model. The likelihood for this model is

$$L_g = g(y) = \frac{g(y,\theta)}{g(\theta \mid y)} = \frac{g(y \mid \theta) g(\theta)}{g(\theta \mid y)}\,. \tag{11}$$

Substituting $g(\theta)$ from (11) into (10) gives

$$L = L_g \int \frac{p(y \mid \theta)}{g(y \mid \theta)} g(\theta \mid y)\, d\theta = L_g \mathrm{E}_g\Big[\frac{p(y \mid \theta)}{g(y \mid \theta)}\Big], \tag{12}$$

where $\mathrm{E}_g$ denotes expectation with respect to density $g(\theta \mid y)$.

Expression (12) is in a form which is very convenient for simulation. The Gaussian likelihood (11) is easily calculated by KFS. We then need the simulation only to compute the adjustment from $L_g$ to $L$. In simulation parlance, $g(\theta \mid y)$ is an importance density. It has to be a Gaussian density because this is the only type of density for which simulation samples can be drawn. We choose it to resemble its non-Gaussian analogue $p(\theta \mid y)$ as closely as possible. This is done by making its mode equal to that of $p(\theta \mid y)$. This latter mode is a simple function of the mode $\hat{\alpha}$ whose calculation we have just described. Remarkably, if $p(y_t \mid \theta_t)$ belongs to the exponential family

$$p(y_t \mid \theta_t) = \exp[\theta_t^\top y_t - b(\theta_t) + c(y_t)],$$

the curvature at the modes of the two densities and the modes are automatically equal. This covers important cases such as the Poisson, binomial and multinomial densities and provides a useful enhancement of accuracy. As well as for the exponential family, Koopman and I provided details for the t-distribution and normal mixtures which are useful for dealing with heavy-tailed densities.

We obtained further increases of efficiency by employing two antithetic variables. Suppose $\theta$ is a draw from $g(\theta \mid y)$. Then take also $\tilde{\theta}$ such that $\tilde{\theta} - \hat{\theta} = -(\theta - \hat{\theta})$ where $\hat{\theta}$ is the mean of $g(\theta \mid y)$. Since $\theta$ is normally distributed $\tilde{\theta}$ has the same distribution. We called this 'balancing for location' and we developed an analogous antithetic variable by balancing distance of $\theta$ from $\hat{\theta}$ by another vector $\acute{\theta}$ with the same direction as $\theta$. We called this 'balancing for scale'. The effects of the two antithetic variables were comparable and complementary and in the examples we examined they accounted together for a reduction in variance due to simulation of about 95%.

For a particular draw $\theta^{(i)}$ from $g(\theta \mid y)$, the two antithetic variables together provide four values. Denote the mean of these by $g_i$. For a simulation sample of size $N$ we then estimate the likelihood by

$$\hat{L} = L_g \frac{1}{N} \sum_{i=1}^{N} g_i \, .$$

This is maximized numerically to obtain maximum likelihood estimates of model parameters.

The process is very efficient computationally. We found that in the cases we looked at, sample sizes of $N = 200$ were sufficient to bring the standard error due to simulation to less than 10% of the asymptotic standard error of parameter estimates. We believe this is accurate enough for practical purposes. Computing times were surprisingly low. For example, in a particular case discussed in our 1997 paper, using non-simulation starting values that were obtained in a way described in the paper, less than 73 seconds with $N = 200$ were needed on a Pentium PC to fit with high accuracy a model in which the observation error has a t-distribution with an unknown number of degrees of freedom and in which there are four other unknown parameters.

At this point it would have been feasible to draw a line under our work. We had shown how to estimate the parameters of a model efficiently and we had provided an efficient procedure for calculating the conditional mode $\hat{\alpha}$ of the state vector. This is an intuitively reasonable estimate because it is the most probable value of $\alpha$ given the observations. However, we decided to extend the work to the estimation of conditional mean $\bar{\alpha}$ for two reasons. The first was that no estimate was available of the mean square error matrix of $\hat{\alpha}$; secondly, we did not know how big the discrepancy might be between $\hat{\alpha}$ and $\bar{\alpha}$. No analytical methods are available for calculating $\bar{\alpha}$ or its mean square error. We therefore developed simulation methods and these turned out to be very close to those used for the likelihood in Durbin & Koopman (1997) which also were very efficient computationally.

Figures 3–6 show some of the results we obtained for two of the four illustrative examples we considered in Durbin & Koopman (1998). Figures 3 and 4 refer to British gas consumption 1961–86. The model used a t-distribution for the observation error and a Gaussian basic structural model for the state; the number of degrees of freedom was treated as an unknown parameter and estimated by maximum likelihood. Figure 3a graphs the conditional mode and conditional mean on the same diagram and demonstrates that the difference between them can barely be detected. Figure 4 compares the performances of the Gaussian model and the t-model. The left-hand diagrams show that the t-model handles the outliers in 1970 better than the Gaussian model. The right-hand diagrams show that if the Gaussian model is used, the seasonal pattern is distorted by the outliers in 1970, whereas if the t-model is used it is not.

Figures 5 and 6 refer to monthly numbers of British drivers killed and seriously injured in road accidents from 1969 to 1984. The wearing of seat belts was made compulsory in February 1983 and this resulted in a substantial drop in casualties. We wanted to see how well this drop was picked up by a model which has a t-distribution for the error term in the trend element of the state equation as compared with a Gaussian distribution. No information was included in the model about the time point at which the drop in casualties occurred. Figure 6a

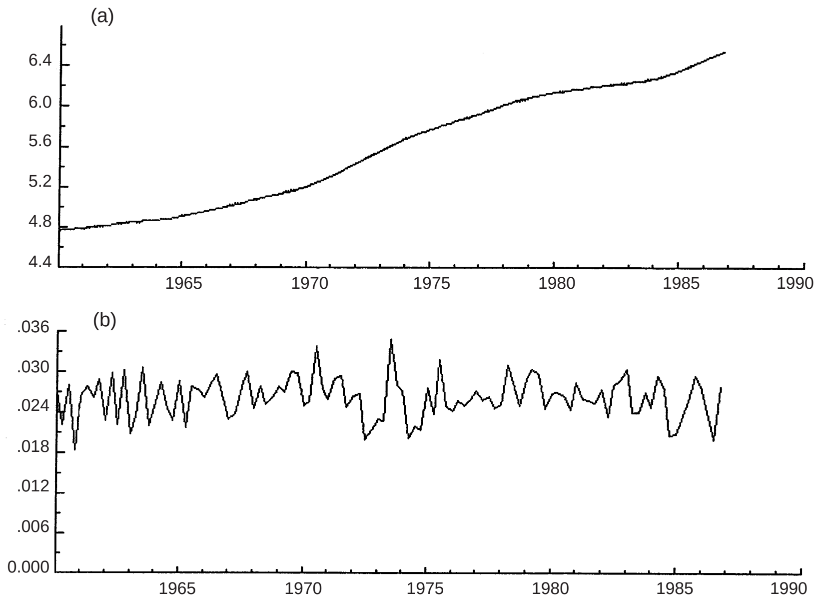


Figure 3. (a) The mode (—) and mean $(\cdots)$ of the trend $(N = 200)$;
(b) the RMSE of the trend $(N = 200)$

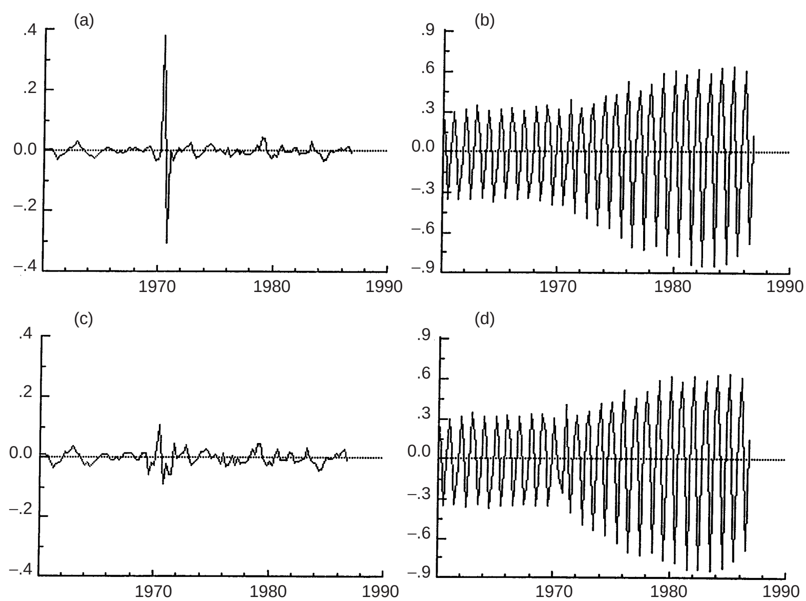


Figure 4. (a) Irregular t-model; (b) seasonal t-model;
(c) irregular Gaussian model; (d) seasonal Gaussian model

shows that the t-model copes entirely satisfactorily with the drop in level in 1983 whereas the Gaussian model does not. This is confirmed by Figure 6b where the t-model rightly gives a large negative outlier at the point where the level shift occurred, whereas the response of the Gaussian model is feeble.

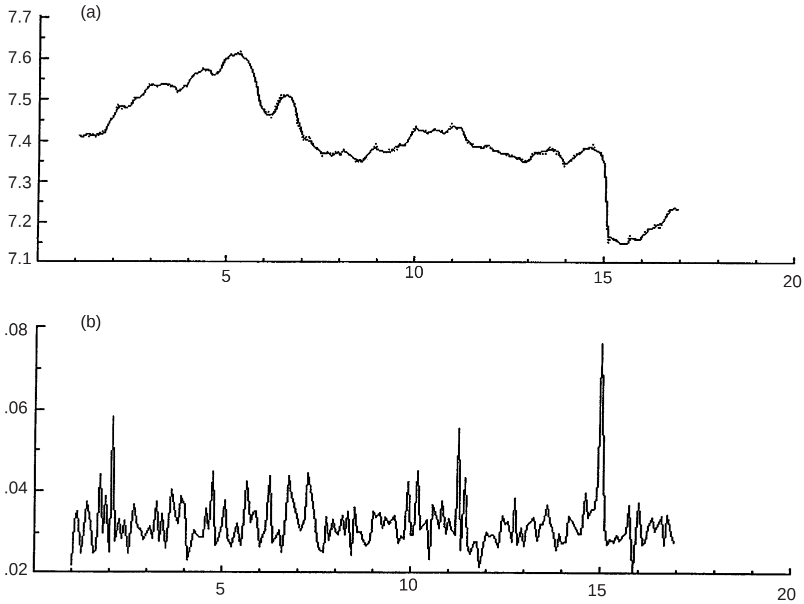


Figure 5. (a) The mode (—) and mean $(\cdots)$ of the trend $(N = 200)$;
(b) the RMSE of the trend $(N = 200)$

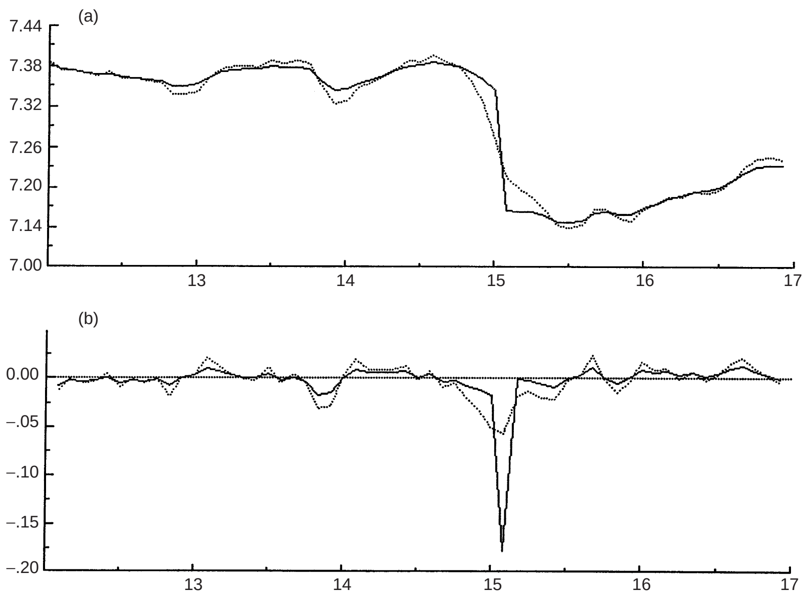


Figure 6. (a) Trends t-model (—) and Gaussian model $(\cdots)$ for subsample;
(b) irregular t-model (—) and Gaussian model $(\cdots)$ for subsample

To sum up, Koopman and I have developed methods for dealing with non-Gaussian time series which are both computationally efficient and feasible and which give good results in applications. We anticipate the development of software to make these methods widely accessible to applied statisticians.

## 5. Potential for official statistics

In this section I consider the potential that state space methods might have for time series work in official statistics. For everyday work in time series model building, I hope that in Section 3.1 I have succeeded in demonstrating that state space models offer greater scope, flexibility and relevance to practical needs than BJ models.

The work on benchmarking raises a number of points. Benchmarking is, of course, an important problem in its own right. Consumers of statistics rightly do not like to see published sets of monthly data which do not add up to published annual totals. The methods that Quenneville and I worked out for the basic benchmarking problem clearly need extending to the multivariate case, with constraints requiring that regional figures add to national totals and figures for different sectors of industry add to overall totals. These constraints are additive; however, as we showed in Durbin & Koopman (1998), this does not preclude the use of multiplicative models for the individual series, where appropriate. The resulting non-linear models can be handled by the techniques we developed.

I would like to see an increase in the use of state space time series models on a routine basis for handling data from monthly and quarterly surveys. The treatment that Quenneville and I gave of bias estimation suggests possibilities for improving the accuracy of survey estimates by linking the survey results to data from other sources, particularly administrative data.

Much of the observational data that we meet in the economic and social fields comes from distributions with heavy tails. The effect of this is that if we analyse these by methods based on the normal distribution, the large-deviation observations have an excessive influence on the analysis. The usual way to combat this is to identify large-deviation observations as outliers or extreme values and scale them down so that they can be accommodated more comfortably within the normal distribution framework. I think that this approach is wrong in principle. My view is that the correct procedure is to write down an appropriate non-Gaussian model that fits the data properly and base the analysis on that. Of course, this means that new theory sometimes has to be worked out. This is what Koopman and I have done for state space time series models. We have shown that when there appear to be so-called outliers or structural shifts in the data, these can be accommodated quite naturally in the analysis by using the t-distributions or normal mixtures for the error densities in the model. I appreciate that there are occasional catastrophes in the world and gross recording errors in the data; such eventualities have to be taken care of. However, most of what are usually called outliers are not really outliers but naturally occurring values which should be modelled as such. I hope that this approach to the so-called ‘outlier problem’ is increasingly accepted in the official statistical world.

The most important use of time series analysis in official statistics, by a very long way, is for seasonal adjustment. In most government agencies, this is done by some variant of X-11 or by a similar intuitively-based smoothing technique. Most academic time series specialists are surprised by this when they first encounter these methods. They feel that someone should have worked out a model-based technique based on modern statistical ideas that would outperform any intuitively based procedure. Yet the basic methods of X-11 were worked out over 30 years ago and have not been changed in essence since. One has to admire the success of the American statisticians in creating a method that has been so popular for so long, and which has also survived a revolution in computing.

Part of the explanation is that neither trend nor seasonal effects can be objectively defined. How smooth should a trend be? No objective criterion can decide this; ultimately, the answer must be determined by the consumer. How rapidly should a seasonal pattern change over time? The same applies. It seems that the designers of X-11 must have got the answers about right, presumably by trial and error and the exercise of judgement.

Nevertheless, X-11 and its derivatives such as X-12 are by no means perfect. In particular, they do not make the most efficient use of recent data at the end of the series. With the development of state space models I believe that the way is open for the development of suitable model-based methods that would make efficient use of the data. I have discussed earlier the flexibility and generality of these models. Koopman and I, and others, have shown how non-normality and nonlinearities can be handled. Explanatory variables such as trading-day, calendar variations and weather variables can be accommodated relatively easily. Use of such models would provide a better base for subsequent analysis because they are analytically based.

The task of developing state space models for seasonal adjustment would be large since I think it unlikely that the existing rather simple trend and seasonal component models in existing structural time series models could be accepted by official statisticians without further investigation. A range of models would probably have to be explored to ensure that just the right amount of smoothness of the trend and change over time of the seasonal is achieved. But though the task is large, I very much hope that someone can be persuaded to undertake it.

## *References*

DURBIN, J. & KOOPMAN, S.J. (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika* **84**, 669–684.

DURBIN, J. & KOOPMAN, S.J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *J. Roy. Statist. Soc. Ser. B* **62**, 3–56.

DURBIN, J. & QUENNEVILLE, B. (1997). Benchmarking by state space models. *Int. Statist. Rev.* **65**, 23–98.

HARVEY, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

HILLMER, C.H. & TRABELSI, A. (1987). Benchmarking of economic time series. *J. Amer. Statist. Assoc.* **82**, 1064–1071.

HOLT, C.C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. ONR Research Memorandum 52. Carnegie Institute of Technology.

KALMAN, R.E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Engineering, Trans. ASME ser. D* **82**, 95–108.

KOOPMAN, S.J., HARVEY, A.C., DOORNIK, J.A. & SHEPHARD, N. (1995). *STAMP 5.0. Structural Time Series Analyser, Modeller and Predictor.* London: Chapman & Hall.

MUTH, J.F. (1960). Optimal properties of exponentially weighted forecasts. *J. Amer. Statist. Assoc.* **55**, 299–305.

THEIL, H. & WAGE, S. (1964). Some observations on adaptive forecasting. *Management Science* **10**, 198–206.

WEST, M. & HARRISON, P.J. (1989). *Bayesian Forecasting and Dynamic Models.* New York: Springer-Verlag.

WINTERS, P.R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science* **6**, 324–342.

## Discussion of the paper by J. Durbin

**Geoff Lee** (*Australian Bureau of Statistics*): I am honoured to have been invited to speak tonight, especially as a discussant for a presentation by Professor Durbin. My own reservations about my qualifications are not related to my experience with official statistics. Rather they centre on the relative level of my statistical abilities compared to Professor Durbin's expertise. I would like to thank several people at the Australian Bureau of Statistics (ABS), Philip Bell and Andrew Sutcliffe in particular, who have helped me prepare for this evening.

I have the advantage of having known and worked for Ken Foreman, albeit at that stage as a rather junior member of his staff. I recall Ken as being a real gentleman, who took genuine interest in the staff working for him, both as people, and as professional statisticians and methodologists in the making. I can record that even in those early days of my career, I recognized and appreciated the contribution Ken had made to establishing a system for recruiting and training subsequent generations of mathematical statisticians in the ABS. At that time I had much less of a sense of what Ken and those who worked with him had achieved prior to my arrival at the ABS, in introducing such 'innovative' techniques as probability sampling, seasonal analysis, and in establishing the household survey system. In my innocence I took many of these achievements as being just part of the way things naturally ought to be. It is only more recently, meeting visitors from some other agencies who did not share the ABS's good fortune in having a 'Ken Foreman', that I really began to appreciate the totality of what Ken and his colleagues had achieved. Ken, throughout his career, helped develop in the ABS an appreciation and enthusiasm for soundly based methodology and practice that is a very real part of his legacy to the ABS. I, and many others, are heirs to that legacy and we are indebted to him for that.

I intend to follow the basic outline of Professor Durbin's paper, and then conclude with some comments about recent experiences we at the ABS have had with state space modelling.

I found Professor Durbin's exposition, linking the history both of the Box–Jenkins models and of the state space approach back to the exponentially weighted moving average technique, very illuminating. It helped me to understand much better the commonalities and differences between the two techniques. It also reminded me of some work which Professor Danny Pfefferman brought to my attention when he visited us several years ago. Cleveland & Tiao (1976) posed the question 'For what class of models is X-11 in some sense appropriate, or optimal?'. They answered this by providing the auto-correlation function for 'appropriate models'. Maravall (1985) examined the auto-correlation function of the Basic Structural Model (a simple state space model similar to the one presented here by Professor Durbin). Professor Pfefferman pointed out that an auto-correlation function very similar to that of X-11 could be obtained by suitable choice of the variances of the trend term, the seasonal term, and the irregular term in the Basic Structural Model. Put one way, if the data are suitable for X-11 (and in practice many data can be handled adequately by X-11), then there is also a Basic Structural Model that fits it equally well. Or, put another way, if we tell the Basic Structural Model how flexible our trend is allowed to be, and how rapidly seasonality can evolve in the series, we can reproduce the X-11 outcome. It is important for official agencies to understand these links between X-11 and the better known Box–Jenkins and state space model approaches because we make substantial use of X-11-style approaches. Professor Durbin's exposition and the work I have just mentioned show that for many practical series the approaches are not as different as they appear on the surface.

Professor Durbin gave a fairly convincing set of arguments for state space modelling being better than a simple Box–Jenkins approach. For an official statistical agency involved in publishing numerous statistical series each day, the real challenge is between an underlying model-based approach, of whatever variety, and a more pragmatic or heuristic filtering approach based on X-11. One key reason the ABS uses the X-11 approach is the sheer volume of the workload. We seasonally adjust something like 4000 series every year. We simply could not afford to do an adequate job of selecting models for each and every one of those series. There are also concerns that there is an element of subjectivity in selecting an underlying model, although I think this argument can be applied in both directions because even within the X-11 approach there are a number of parameters which have to be set based on judgement and experience. X-11 shares with the state space approach the flexibility to deal with trading-day, calendar-varying holiday proximity effects, and tuned end weights.

If a model-based approach were to be used in an official agency, Professor Durbin has done a good job of convincing me that the state space approach offers many advantages. The transparency of the models being fitted is certainly a big advantage, and the flexibility and ability to incorporate knowledge of the structure of the data being modelled are also very attractive features. I agree that the main drawbacks are lack of information and knowledge within official agencies about the state space approach, and also the lack of well-developed robust easily-operable software.

In commenting on the drawbacks of the Box–Jenkins approach, Professor Durbin mentioned that the differencing approach in essence removes the trend and seasonality, and the modelling then focuses only on what is left. He comments that in some cases the trend and seasonal components are of intrinsic importance. That is certainly our experience at the ABS, where much of the interest of users is focused on exactly those parts of the data; techniques which provide information about these components are therefore of great significance.

Professor Durbin commented that the model identification process was a drawback in the Box–Jenkins approach because many different models seem to fit the data well. It would seem to me, possibly in my ignorance, that an equivalent charge could be laid against the state space approach; there may be many different structures which could be fitted to the data which would adequately explain the observed points.

Professor Durbin mentioned making monthly or quarterly series agree with annual benchmarks. This work is certainly of interest, because the same requirement has arisen in a number of situations within the ABS. I would be interested to learn from Professor Durbin how difficult it would be to modify this approach to the situation in which the benchmarks were also known to have sampling errors, but were based on a much larger sample size and therefore had a smaller error than the sub-annual series. I presume this would be a relatively simple adaptation to make.

I was interested to note, in the example Professor Durbin gave, that three of the benchmark points were not actually annual surveys. Rather they were the pilot surveys for a new collection which was going to update and eventually replace the older collection that had provided the historical data with which he was working. The approach Professor Durbin presented helped link the old survey with the new one. This task of linking a series, when a collection method is changed or updated or a survey is replaced by one that is newer and more appropriately designed, is an important job for statistical agencies. It does not occur frequently, but when it does the task of parallel running the old and new systems is substantial and important.

I was interested also by Professor Durbin's comment about treatment of outliers in the analysis of seasonally adjusted data. While I certainly agree that philosophically it is better to fit a correct distribution to the data rather than declare awkward observations to be outliers, I would still observe that in X-11 a fairly pragmatic approach appears to operate reasonably successfully. Nonetheless this is a topic of interest to us in the ABS, and we have some active research underway, although it is in the X-11 arena rather than in an explicit model arena.

I do not claim to be competent to review Professor Durbin's description of his technique for fitting the state space model to obtain the mode and the mean. I do, however, have several

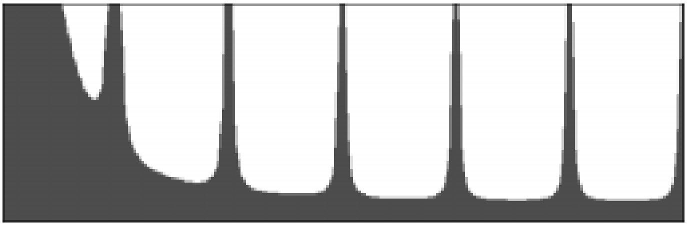

Figure I. Idealized spectrum of a time series (the airline model)

people working for me at the ABS who have undertaken some very preliminary experiments using the Bayesian approach and are looking forward to quizzing him in detail when he visits us. Our experience was that the likelihood function tended to be flat, and values quite far removed from the mode or mean were still plausible.

As Professor Durbin notes, seasonal adjustment and trend extraction are by a long way the most substantial use of time series analysis techniques in official statistical agencies. I believe, despite Professor Durbin's plea, that X-11 will continue to be used for routine seasonal adjustment and trend extraction for some time. As he notes, it is accepted by consumers, and, despite its less impressive academic pedigree, it seems to work in many cases. In addition, there is an element of inertia. We have something like 30 years of experience and development invested in the X-11 approach, and the plain vanilla X-11 that many know about is not exactly the system that is used in practice. For example, Professor Durbin mentions that X-11 does not make efficient use of the recent data. This would certainly be a very serious criticism of X-11 if true, because it seems on occasion that the consumer's sole interest is in the most recent points in a time series. However, at ABS we have investigated and implemented an approach that tailors the end weights to match the observed characteristics of key series, to obtain better initial estimates of the trend at the end of a series. We are in the process of looking at the weights, and particularly the end weights, used for extracting the latest seasonal factors, for exactly the same reason: that is, to obtain more efficient use of the latest data.

Perhaps for routine situations the differences between the X-11 and state space techniques are not that important. As noted earlier, for many practical series the results of analysis by an X-11 or an ARIMA or a state space approach are quite similar. In this sense perhaps it would not be quite as large a task as Professor Durbin suggests, to introduce state space modelling into routine work.

Professor Durbin certainly picked out one very noteworthy issue, though, when he observed that the trend and seasonal effects are not objectively defined; they are unobservable components of the data. What the analysis delivers depends on our initial underlying assumptions and requirements. The following examples based on the simple 'airline model' might help explain this.

Figure I shows the spectrum of an idealized time series which follows the airline model. The peaks correspond to strong seasonality in the data; on the left-hand side of the diagram you can see the longer frequency cycles which represent the trend, and on the right-hand side the remainder after seasonality and trend have been removed; this constitutes the residual or irregular. Figure II shows an idealized picture of the X-11 approach to dividing the time series into trend, seasonal factors, and irregular components. First, all the power in the seasonal frequencies is removed. Next, the remaining longer frequency cycles are defined as the trend, and the residual is, as its name suggests, what is left. Figure III shows the outcome of using

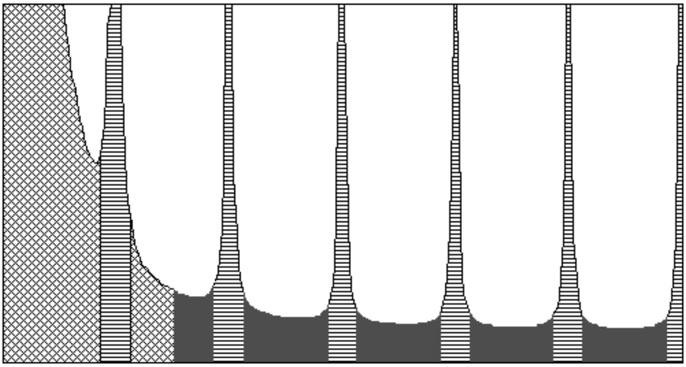

Figure II. X-11 decomposition of spectrum into seasonal factor, trend, and irregular components

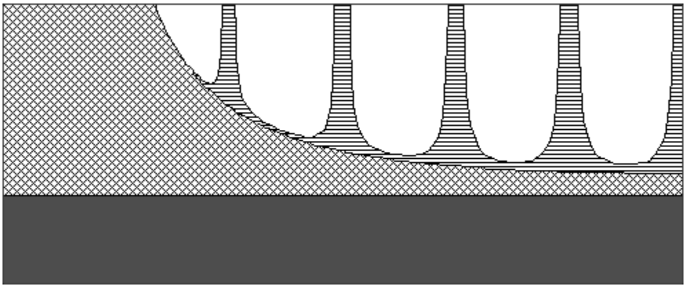

Figure III. State space model decomposition of spectrum into noise, trend and seasonal components

a basic structural model to extract the trend and seasonal terms. First, the model has a white-noise term for the irregular component, and this you can see being removed across the bottom of the diagram. Second, it extracts the trend, and what is left constitutes the seasonal component. You can note that there is an element of arbitrariness in this particular presentation. We could have increased the variance on the irregular component, i.e. given more weight to it, and this would have left us with correspondingly less power in the trend, even though the basic 'shapes' of the three components would have remained largely the same.

Although I believe that X-11 will be used for routine analysis in official agencies for some time, there is already a very useful and valuable role for state space time series modelling in official agencies. It can be used as an investigatory tool, for difficult cases, where there is some knowledge of the underlying structure and the problem warrants the effort of looking into the detail. The advantage of state space modelling is that it allows us to join a time series view of the data with our knowledge of sample design and cross-sectional studies conducted on unit record data. We can also incorporate our knowledge about external influences on the data, for example a change in the method of collection, or some external influence on the economy or series. It offers an alternative to a basic 'composite estimation' approach which is a more traditional tactic used by sample survey statisticians for drawing strength from neighbouring time points.

Two examples might help put this into context. In conjunction with Professor Danny Pfefferman a couple of years ago, we investigated the use of state space models to produce time series of labour force estimates for small areas (Pfefferman, Bell & Signorelli, 1996; Bell & Carolan, 1998). The labour force survey is very accurate at national and state levels, but is subject to substantial sampling error at the small area level. By small area I mean regions such as the Hunter Valley, the Illawarra, the Western Suburbs of Sydney, or even the Northern Territory. In this instance the known structure which we wished to include in our time series analysis was the sampling error. Using state space models we were able to extract far more plausible estimates of time series behaviour, particularly trend, for these regions. One problem we encountered while doing this work was setting the variance parameters. We were using a maximum likelihood estimation approach to fit the model, and it tended, based on the data for each small area alone, to estimate the variance term for the trend as very small, i.e. in essence to make very ‘stiff’ almost straight line trends. This did not accord with our underlying belief about what the trend ought to be doing, even in these small areas, and we ended up having to pre-specify the relative magnitudes of several of the variance terms in the model. I would be interested to discuss with Professor Durbin how he dealt with this issue, in his work on the non-Gaussian state space models.

The second example is a case for which we used state space modelling to investigate a methodological change in a major survey. Recently we introduced telephone interviewing for our labour force survey. You may have seen comments in the labour force publication, or in the press, noting that during the implementation period there was an effect on the published estimates due to the changeover process. We used several quite different analysis techniques to monitor the statistical impact during the transition process. All told the same basic story, but in my opinion, state space modelling was clearly the most flexible approach. Our analysis (Bell, 1998) had to be done in real time, literally within a few hours, because of the exceedingly tight timetables under which the monthly survey is collated and published. The way information unfolded, month by month, as the transition proceeded, made the state space and Kalman filter approaches for updating earlier estimates almost a natural setting in which to operate.

Earlier testing, and overseas analysis, had suggested that we should expect some effect on the estimates, albeit a relatively minor one. From previous work and using unit level data, we had good working models to describe the sampling induced variance covariance structure of the labour force time series (Pfefferman, Feder & Signorelli, 1998). We also knew what I might term the experimental design for the transition process — we gradually phased in the telephone interviewing technique for 1/8th of the sample each month, over a seven-month period (the first contact with each household remains a face to face interview).

Given this, the initial form of the state space model was fairly easy to specify. However, there are a number of other small effects which we know occur in the survey and which could easily have confounded our analysis. We began with a fairly simple allowance for a ‘first time in survey’ effect. In the course of examining the results of the initial analyses it became apparent that it would be better to allow for a broader range of ‘time in survey’ effects. We had conducted previous studies of this, and we could have gone back to the files, reviewed the earlier work, picked up the earlier estimates, and incorporated them into our analyses. But in the state space formulation we simply updated the description of this part of the structure in the state space model and re-ran the whole process, within a few hours. That speed and flexibility was important when decisions had to be made in advance of a publication print deadline less than 12 hours away.

Similarly the questions we wanted to ask about the telephone effect itself changed, as the picture unfolded from month to month during the phase-in period. Initially, we specified the telephone effect as a simple constant. Individual monthly estimates of the telephone effect were quite volatile, and it was important to react only when a consistent picture had emerged.

About half-way through the process it appeared that an effect had been detected, and we then wanted to know more about its behaviour over time, with the key question being 'was the effect temporary or permanent?' As it turns out, it was a transitory effect which appeared in the early stages of the implementation and subsequently decayed. In the state space formulation, asking and answering these questions was as simple as it is in any other model-building approach, and, moreover, we could get estimates of the confidence intervals under the model at the same time. This was far more complicated under the other analytical approaches that we also used. Finally, the state space modelling approach proved very useful when we were searching for possible operational reasons that might be generating the effect we were observing. It was fairly straightforward to look for subgroups which were displaying the effect more noticeably than others. For example, was the effect more prominent in particular states, age groups, full-time or part-time employment, etc? The results from these analytical studies of the data using state space modelling helped guide and corroborate a considerable amount of practical investigation, field studies, interviewer debriefing and the like, which together enabled the ABS to keep users of the data informed about the statistical impact of the methodological changes which were occurring.

I'd like to thank Professor Durbin for his presentation. It has been an interesting talk on a topic of considerable interest to me personally, and of considerable relevance to the ABS where I work. We have already made a start by using some of the techniques he has described. His presentation has put these techniques into perspective with approaches which have preceded them, and has mapped out a range of future possibilities for us to consider and explore.

### *References*


BELL, P.A. (1998). Using state space models and composite estimation to measure the effects of telephone interviewing on labour force estimates. In *Working Papers in Econometrics and Applied Statistics*, Cat. no. 1351.0, no. 98/2. Canberra: ABS.

BELL, P.A. & CAROLAN, A. (1998). Trend estimation for small areas from a continuing survey with controlled sample overlap. In *Working Papers in Econometrics and Applied Statistics*, Cat. no. 1351.0, no. 98/1. Canberra: ABS.

CLEVELAND, W.P. & TIAO, G.C. (1976). Decomposition of seasonal time series: a model for the X-11 program. *J. Amer. Statist. Assoc.* **71**, 581–587.

MARAVALL, A. (1985). On structural time series models and the characterization of components. *J. Bus. Econom. Statist.* **3**, 350–355.

PFEFFERMANN, D., BELL, P. & SIGNORELLI, D. (1996). Labour force trend estimation in small areas. In *Proc. US Bureau of the Census Annual Research Conference.*

PFEFFERMANN, D., FEDER, M. & SIGNORELLI, D. (1998). Estimation of autocorrelations of survey errors with application to trend estimation in small areas. *J. Bus. Econom. Statist.* **16**, 339.


**Des Nicholls** (*The Australian National University*): It is a pleasure and an honour to comment on Jim Durbin's paper which he has presented as the 1997 Foreman Lecture to the Statistical Society of Australia. His overview of the state space approach and its relevance to the modelling of time series is impressive. Like so much of his research over half a century, the state space approach which he and his colleagues have been advocating has been shown to be of importance in its application to the analysis of time series data.

Kalman (1960) introduced the state space approach in the engineering literature. During the next 20 years, the methodology was considered by statisticians and econometricians who recognized the advantages of such an approach to overcome problems such as the handling of missing data in time series. Jones (1980, 1993), Harvey (1984, 1989) and Carter & Kohn

(1997) are examples of just a few of the researchers who have made significant contributions in the application of the state space approach to the solution of time series problems.

What is of interest, like much of the application of time series methods, is just how dependent the development of the state space approach has been on advances in computer technology, both in hardware and in software. This is, of course, also true of the development and acceptance of the Box–Jenkins approach to the modelling of time series. The widespread application of that methodology is greatly assisted by the availability of easy-to-use software packages which incorporate it; so much so, that packages such as SPSS and MINITAB allow for application of the Box–Jenkins procedure entirely on a menu-driven basis. Unfortunately, this is not so in the case of the application of the state space modelling procedures. The package STAMP, developed by Harvey and his colleagues and referred to by Professor Durbin, is a purpose-specific package for the application of the structural time series approach based on the state space methodology. This package, while impressive when first developed, now needs upgrading to make it more user friendly in a Windows (rather than DOS) environment. Once the methodology is incorporated into menu-driven software packages, there is no doubt that the structural time series approach using the state space methodology will achieve the widespread acceptance it so rightly deserves.

In his paper, Professor Durbin compares the Box–Jenkins approach with the state space approach, presenting a strong case in support of the state space approach. It seems natural to ask why, if this is so, the Box–Jenkins approach is still the dominant methodology taught in universities. One reason alluded to earlier is that the software is readily available for application of the Box–Jenkins technique; this is not the case for the state space approach as has been recognized by Professor Durbin who refers us to the STAMP package for application of the state space approach.

I believe there is a second reason for the state space approach being neither widely accepted nor included in curricula although the underlying structural approach is not difficult; indeed it is very logical to consider a time series composed of a trend, seasonal, cyclic and calendar variations, together with explanatory variables (including interventions) and for these to be modelled separately. But, when applying the state space approach using the Kalman filter method, many students have problems putting the non-parametric models used for each of the structural components into a state space form for application of the Kalman recursion formulae, and in actually applying these recursion formulae. The students' basic training results in them relating much more comfortably to a parametric modelling approach. Therefore, the extension of standard regression modelling approaches to autoregressive, moving average and autoregressive moving average models is natural and well accepted, giving support to the Box–Jenkins approach. There is no doubt that the state space approach will become more accepted in the future, particularly when its enormous flexibility is more fully recognized and if it is supported by software which is well documented, regularly updated and easily applied.

In the Box–Jenkins approach, an integrated series is one which, after differencing, becomes stationary. Stationarity is important because it is one of the basic assumptions made in modelling and forecasting. If a series is not stationary, its mean, variance and covariances may be changing, and models which assume that these are constants will be misleading.

One drawback of the procedure of differencing is that it results in a loss of valuable 'long-run' information in the data. A regression on the differences of time series provides no information about the long-run relationship; that information can only be provided by a regression estimated on the levels of the data. The concept of co-integrated series has been suggested as one solution to this problem; it considers relationships between integrated processes and the concept of (steady state) equilibrium. It was introduced in the econometric context by Granger (1981) and extended in Granger & Weiss (1983). Granger (1986) gives an elementary overview of the basic concepts, while Engle & Granger (1987) consider the representation, estimation and testing in the case of co-integration and error correction. The investigation of long-run relationships has received considerable attention in the (economics and econometric) literature as a result of the development of co-integration and error correction models, while the state space approach, which also has the flexibility to look at long-run effects, has not received the attention it deserves.

The latter part of Professor Durbin's paper relates to recent work he and his colleagues have undertaken in applying the state space approach to benchmarking and in extending the approach to deal with non-Gaussian observations. In each case the relevance of the Kalman filter and smoother is evident. Finally, he has discussed the potential for the use of the methodology for official statistics purposes and has suggested a rich source of research problems, both theoretical and applied.

This thoughtful contribution gives an excellent overview of the state space approach to time series modelling and its advantages over the more recognized approach introduced some 30 years ago by Box and Jenkins. The discussion on the flexibility and adaptability of the state space approach is compelling and leaves little doubt that it must be fully recognized and utilized. Incorporation of the methodology into widely-used menu-driven software packages would go a long way to achieving this.

### *References*

CARTER, C.K. & KOHN, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *J. Roy. Statist. Soc. Ser. B* **59**, 259–268.

ENGLE, R.F. & GRANGER, C.W.J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica* **55**, 251–276.

GRANGER, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *J. Econometrics* **16**, 121–130.

GRANGER, C.W.J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bull. Econom. Statist.* **48**, 213–228.

GRANGER, C.W.J. & WEISS, A.A. (1983). Time series analysis of error correction models. In *Studies in Econometrics, Time Series and Multivariate Statistics,* eds S. Karlin, T. Amemiya & L.A. Goodman, pp . 255–278. New York: Academic Press.

HARVEY, A.C. (1984). A unified view of statistical forecasting procedures (with discussion). *J. Forecasting* **3**, 245–283.

HARVEY, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press.

JONES, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22**, 389–395.

JONES, R.H. (1993). *Longitudinal Data Analysis with Serial Correlation: a State Space Approach.* New York: Chapman & Hall.

**Author's rejoinder:** I thank both discussants for their generous remarks about my lecture and congratulate them on their pragmatic approach to practical aspects of time series analysis.

Des Nicholls rightly points out that acceptance of state space methodology for practical use depends crucially on the availability of suitable user-friendly software. There have been significant developments in this respect since I delivered the lecture. Software for state space methodology will be freely available on the Internet by the time these words are published, as mentioned in a discussion paper on non-Gaussian applications read to the Royal Statistical

Society in May 1999 (Durbin & Koopman, 2000a; see also Koopman *et al.*, 1999). An updated Windows version of the STAMP package is also being developed. Thus the software barrier should be removed very soon.

Des Nicholls also discusses the important question of teaching time series analysis to students. It is true that Box–Jenkins analysis can be fun to teach. However, problems arise in applications, particularly in the identification process, which can sometimes be quite baffling, particularly when models which look very different appear to fit the data equally well. If one thinks hard about the underlying theoretical basis of Box–Jenkins analysis instead of taking it at face value, it is found to be rather shaky, as I indicated in my On the other hand, the theoretical basis of state space is very simple because it consists only of matrix manipulation plus elements of multivariate normal regression. In Durbin & Koopman (2000b), we try to keep presentation of theory simple by introducing all the essential ideas in an introductory chapter in terms of an elementary special case, the random walk plus noise model. We hope this will help teachers to get across to students the basic elements of the approach without them having to confront formidable-looking matrix products that occur in the general theory.

Geoff Lee asks whether the case in which benchmarks have sampling errors can be treated. The answer is that in Durbin & Quenneville (1997) the state space theory is presented for the general case in which benchmarks as well as observations are subject to sampling variability.

I found Geoff Lee's description of his use of state space methods for small-area problems and telephone interviewing particularly interesting. These applications reinforce the remarks in my lecture about the possibilities for using state space methods in official statistics.

The contributions of the two discussants provide a valuable extension of the subject matter of my lecture.

## *References*

DURBIN, J. & KOOPMAN, S.J. (2000a). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *J. Roy. Statist. Soc. Ser. B* **62**, 3–56.

DURBIN, J. & KOOPMAN, S.J. (2000b). *Time Series Analysis by State Space Models*. Oxford: Oxford University Press (to appear).

KOOPMAN, S.J., SHEPHARD, N. & DOORNIK, J.A. (1999). Statistical algorithms in state space using SsfPack 2.2. *Econom. J.* (to appear).