SSC Annual Meeting, July 1995
Proceedings of the Survey Methods Section

BENCHMARKING MONTHLY TIME SERIES WITH STRUCTURAL TIME SERIES MODELS

J. Durbin and B. Quenneville[1]

ABSTRACT

We have a monthly series of observations which are obtained from sample surveys and are therefore subject to survey errors. We also have a series of annual values, called benchmarks, which are either exact or are substantially more accurate than the surveys observations; these can be either annual totals or accurate values of the underlying variable at a particular month. The benchmarking problem is the problem of adjusting the monthly series to be consistent with the annual values.

RÉSUMÉ

Nous disposons d'une série d'observations mensuelles obtenues d'une enquête par sondage et, par conséquent, sujette à des erreurs d'échantillonnage. Nous avons également une série de valeurs annuelles, appelée étalon, qui sont soient exactes, soient considérablement plus précises que les observations obtenues par enquête; elles sont ou bien des totaux annuels, ou bien des valeurs précises de la variable sous-jacente à un mois déterminé. Le problème d'étalonnement en est un d'ajustement des valeurs mensuelles de sorte qu'elles soient consistantes avec les valeurs annuelles.

1. INTRODUCTION

A common problem faced by official statistical agencies is the adjustment of monthly time series which have been obtained from sample surveys to make them consistent with more accurate values obtained from other sources. These values can be aggregates or individual values at arbitrary points along the series. The sources can be censuses or more accurate sample surveys or administrative data or some combination of these. This adjustment process is called benchmarking and the more accurate values are called benchmarks. Typically, the benchmarks are either yearly totals or values observed at a particular time-point each year. For simplicity, we assume that the data are monthly and the benchmarks are annual, though the broader interpretation should be borne in mind.

Most previous work (see, for example, Cholette and Dagum (1994) and Mian and Laniel (1993) and the references therein) took the underlying target series as a series of constants to be estimated by some appropriate method such as maximum likelihood. Hillmer and Trabelsi (1987) introduced the idea of using a stochastic model for the target series. Since this takes autocorrelation of the underlying values into account, thus leading to greater efficiency, we follow the latter approach.

In this paper we solve the benchmarking problem by modelling the monthly data by state space structural time series models. We give solutions both for the case where the monthly observations and the annual values behave additively, and also for the case, very common in practice for economic series, where the monthly observations behave multiplicatively but the annual values behave additively, for example yearly totals. It is frequently the case that the sample survey data are biased due to non-response and other factors. The existence of accurate annual values enables this bias to be estimated. We discuss how to derive estimates for the bias in both the additive and multiplicative cases.

We present two methods for solving the benchmarking problem, first a two-stage method in which in the first stage we fit a structural time series model to the monthly values alone and then in the second stage we combine information from the first stage with information from the benchmarks. Secondly, we develop a single-stage method in which we combine the monthly and annual values into a single time series which is fitted to an appropriate state space model. We present the two-stage method first since the two-stage approach to the problem is already familiar to workers in the field through the paper of Hillmer and Trabelsi (1987) who used an ARIMA model to represent the monthly data at the first stage. Our view is that the state space approach is superior to the ARIMA approach for this type of problem because of its greater flexibility

[1] James Durbin, Emeritus Professor, London School of Economics and Political Science, Houghton St., London, WC2 2AE, England, U.K., e-mail: durbin@vax.ise.ac.uk. His work was done at the Time Series Research and Analysis Centre, Statistics Canada.;
Benoît Quenneville, Time Series Research and Analysis Centre, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6.

and comprehensiveness; for example, it permits us to allow for trading day variation in a straightforward way. Other differences from the Hillmer and Trabelsi (1987) treatment are that they assume that non-stationary series have zero means whereas we do not, we allow for heteroscedasticity in the survey errors, we estimate survey bias and finally we treat the multiplicative case.

In section 2, we develop a state space model for the monthly observations which includes trend, seasonal and trading-day components, and which also allows for survey errors. Section 3 describes the fitting of the model and shows how the monthly observations can be combined with the benchmark data for the case where the monthly series behaves additively. In section 4 we consider the case where the monthly observations behave multiplicatively but the benchmarks are additive. Section 5 shows how to estimate survey bias. In section 6 all the methods we developed are applied to a single data set, the Canadian retail trade series from January 1980 to December 1989. The data are particularly interesting since they demonstrate the need for bias estimation and for the use of multiplicative models for the benchmarking problem.

2. A STATE SPACE MODEL FOR BENCHMARKING

Suppose that we have a monthly series of univariate observations y_t, obtained from sample surveys, which are estimates of underlying true values η_t, and which satisfy the relation

$$y_t = \eta_t + k_t u_t, \quad t=1, \ldots, n, \tag{2.1}$$

where $k_t u_t$ is the survey error. We assume that u_t is a unit-variance stationary ARMA(p,q) series and that values of k_t, p and q are available from survey experts. The inclusion of the factor k_t enables us to allow for heteroscedasticity in the survey errors; obviously, k_t is their standard deviation. Suppose that in addition we have a series of annual values $x_i, i=1, \ldots, l$ which are available from another source and which are regarded as more accurate than the y_t's. It is important to note that the term annual values need not refer to calendar years. For example we may have a series of values which are totals of February to January values together with accurate values in particular months. This is the case for the example considered in section 6. Let $\eta = [\eta_1, \ldots, \eta_n]'$ and $x = [x_1, \ldots, x_l]'$. The x_i's are assumed to satisfy the benchmarking relations

$$x = L\eta + e, \quad e \sim N(0, \Sigma_e), \tag{2.2}$$

where L is a known matrix and where we assume that Σ_e is known either because it is known to be a zero matrix or because the x_i's are obtained from accurate annual surveys, in which case survey experts can be expected to provide us with an estimate of Σ_e. The x_i's are called *benchmarks* and when $e = 0$ the benchmarks are said to be *binding*.

We assume that the true values η_t are generated by the structural time series model

$$\eta_t = \mu_t + \gamma_t + \sum_{j=1}^{k} \delta_{jt} w_{jt} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2), \quad t=1, \ldots, n, \tag{2.3}$$

where μ_t is the trend, γ_t is the seasonal pattern, $\sum_{j=1}^{k} \delta_{jt} w_{jt}$ is the trading day (plus leap year) adjustment and ϵ_t is an error term. A choice of models is available for μ_t and γ_t; see, for example, Harvey (1989) section 2.3. In this paper we use the models

$$\mu_t = 2\mu_{t-1} - \mu_{t-2} + \xi_t, \quad \xi_t \sim N(0, \sigma_\xi^2), \tag{2.4}$$

$$\gamma_t = -\sum_{j=1}^{11} \gamma_{t-j} + w_t, \quad w_t \sim N(0, \sigma_w^2). \tag{2.5}$$

The idea behind (2.4) is that if $\xi_t = 0$, a straight line is followed by the trend exactly; inclusion of the error ξ_t allows the slope to change over time. The idea behind (2.5) is that if w_t=0, the seasonal pattern is constant; inclusion of the error w_t allows the pattern to vary. For the trading day coefficients we use the model

$$\delta_{jt} = \delta_{j,t-1} + \zeta_{jt}, \quad \zeta_{jt} \sim N(0, \sigma_\zeta^2), \quad j = 1, \ldots, k. \tag{2.6}$$

This allows the coefficients to change over time, which is a desirable feature in many applications. The case where an adequate fit can be achieved by treating the coefficients as constant can be handled by putting $\zeta_{jt} = 0$; in this case the model-fitting process is simplified substantially. More general models for time-varying trading day adjustments are considered by Dagum, Quenneville and Sutradhar (1992). The error series $\epsilon_t, \xi_t, w_t, \zeta_{1t}, \ldots, \zeta_{kt}$ are assumed to be white noise series independent of each other and of u_t.

By adding in the survey error we obtain the model for the observations y_t,

$$y_t = \mu_t + \gamma_t + \sum_{j=1}^{k} \delta_{jt} w_{jt} + \epsilon_t + k_t u_t, \; t = 1, \ldots, n, \tag{2.7}$$

the state space form of which is described in Durbin and Quenneville (1995).

We wish to estimate the quantity η_t using both monthly and benchmark data. The estimate we shall use is the mean of the posterior distribution of η given all the data, that is,

$$\hat{\eta} = E(\eta \mid z), \tag{2.8}$$

where $y = [y_1, ..., y_n]'$, $\hat{\eta} = [\hat{\eta}_1, ..., \hat{\eta}_n]'$ and $z = [y', x']'$.

By the properties of the multivariate normal distribution, $\hat{\eta}$ is a linear function of z. Durbin and Quenneville (1995) show that the estimate $\hat{\eta}$ that is obtained by assuming normality has the minimum mean square error property in the class of linear estimate of η even when the random variables in the model are non-normal, that is, $\hat{\eta}$ is the linear estimate that minimizes $E[(\hat{\eta}-\eta)(\hat{\eta}-\eta)']$. This is a Gauss-Markov-like property but it does not follow from the Gauss-Markov theorem since η is a random vector.

3. BENCHMARKING WHEN THE MONTHLY MODEL IS ADDITIVE

Two different approaches can be used to calculate $\hat{\eta}$ given by (2.8). In the first approach the problem is solved in two stages. At the first stage we compute $\tilde{\eta} = E(\eta \mid y)$ by applying standard Kalman filtering and smoothing (KFS) to the monthly series alone, ignoring the benchmark values. Then at the second stage we combine the information about η in $\tilde{\eta}$ with that in the benchmark values x to obtain the final estimate $\hat{\eta}$.

Let $\Omega = E[(\tilde{\eta}-\eta)(\tilde{\eta}-\eta)']$. Our final estimate $\hat{\eta}$ of η based on both monthly and benchmark values is given by

$$\hat{\eta} = \tilde{\eta} + \Omega L'[L\Omega L' + \Sigma_e]^{-1}(x - L\tilde{\eta}), \qquad (3.1)$$

with MSE matrix

$$E[(\hat{\eta}-\eta)(\hat{\eta}-\eta)'] = \Omega - \Omega L'[L\Omega L' + \Sigma_e]^{-1}L\Omega. \quad (3.2)$$

Formula (3.1) is due to Hillmer and Trabelsi (1987), equation (2.16). However, their proof is complicated due to the fact that they do not relate the result directly to standard regression theory. Moreover, they make the unnecessary and, in our case, invalid assumption that a nonstationary series has zero mean. Durbin and Quenneville (1995) give an alternative proof.

Assuming that Σ_e is known, we see that implementation of (3.1) and (3.2) require only a knowledge of $\tilde{\eta}$ and Ω, which are computed by KFS. As a consequence, the two-stage method has the advantage that the structural time series model could be fitted at the first stage by existing software, so only the second stage would require new software.

In the second approach we form a single series, $y_1, ..., y_{12}, x_1, y_{13}, ..., y_{24}, x_2, y_{25}, ...$, in the case of annual benchmarks, and apply a specially designed KFS to this series. Of course, the requisite model, given in Durbin and Quenneville (1995), is somewhat more complicated than the model of section 2. However, the advantages of the single-stage method are that it fits the structural time series model using all the data, including the benchmarks. This is important for trend estimation and seasonal adjustment. A further point is that from the standpoint of state space theory this is the natural way to solve the benchmarking problem. From this perspective the single-stage solution represents an advance in state space methodology which has substantial potential for further development, for example to the multivariate case where several series are related. It has the additional advantage, characteristic of state space models, that it is relatively easy to update the system each time a new observation comes in.

4. BENCHMARKING WHEN THE MONTHLY MODEL IS MULTIPLICATIVE

In many practical situations, social and economic variables behave multiplicatively, and when such variables are measured by sample surveys the survey errors also usually operate multiplicatively. Thus what we actually observe in such cases is

$$Y_t = N_t U_t, \quad t = 1, ..., n, \qquad (4.1)$$

where N_t is the underlying true value that we wish to estimate and U_t is the survey error. Let $y_t = \log Y_t$, $\eta_t = \log N_t$ and $k_t u_t = \log U_t$, where we make the same assumptions about u_t as in section 2 but k_t is now the coefficient of variation. Then (4.1) transforms into

$$y_t = \eta_t + k_t u_t,$$

which is the same as (2.1). We assume that η_t and y_t satisfy the same models as in (2.3) and (2.7).

The benchmarking relations are, however, expressed in terms of the true values N_t and not their logs η_t. Thus the benchmark values we observe are given by

$$x = LN + e, \qquad (4.2)$$

where $x = [x_1, ..., x_l]'$ as before, $N = [N_1, ..., N_n]'$ and L and e are the same as in (2.2). The problem we shall consider is to estimate $N_1, ..., N_n$ given both the observations Y_t and the benchmarks x_t. The difficulty is that the model is now nonlinear so the solution of section 3 does not apply. Moreover, because the model is nonlinear we cannot estimate N_t straightforwardly by its posterior mean $E(N_t \mid y, x)$ since an analytical solution is intractable. What we suggest is that we estimate η_t or N_t by the *modes* of their posterior densities given y and x instead of their *means*. We call these estimates *posterior mode estimates* (PME's). Such estimates have been used by Fahrmeir (1992), Durbin and Koopman

(1994) and Durbin and Cordero (1994) for estimating the state vector in state space models for non-Gaussian observations. Of course, when the posterior density is symmetric the mode is equal to the mean, so since calculation of the mode is relatively straightforward, it gives an easy way of computing the mean. If the posterior density is not symmetric, use of the mode can be justified on the intuitive ground that it is the estimate of an unknown which is most probable given the data. We assume that the density is unimodal which it will usually be in the situations under consideration.

We have the choice between taking $\hat{\eta}_t$ as the mode of its density and then taking $\hat{N}_t = \exp(\hat{\eta}_t)$ or taking $\hat{N}_t$ as the mode of its own density. However, our expectation is that the difference between these estimates will be small.

The basic steps of the methodology are to obtain the log conditional density of η given y and x; to differentiate this with respect to η_t and equate to zero for $t = 1, ..., n$; to linearise these equations and put the linearised equations into a form analogous to the additive case; to use the theory for the additive case at each step of the iteration; and finally, to estimate the true values N_t by $\exp(\hat{\eta}_t)$ where $\hat{\eta}$ denotes the estimate of η. The details are provided in Durbin and Quenneville (1995).

Since the final estimate $\hat{\eta}$ is not linear in y and x, it does not have the minimum mean square error linear property discussed at the end of Section 2.

For the single-stage solution, we linearise the estimating equations for the unknowns concerned and put them into a form which can be solved by a Kalman filtering and smoothing operation analogous to that used for the additive case. This enables us to develop an iterative method for the estimation of the posterior mode of the target series.

5. ESTIMATION OF SURVEY BIAS

It is frequently the case that estimates obtained from surveys are known to be biased due to non-response and other factors. The existence of benchmarks free from bias enables us to estimate the bias in the survey observations. Assume that in the additive case considered in Sections 2 and 3, the observations y_t contain a constant additive bias b and in the multiplicative case considered in Section 4, the observations Y_t contain a constant multiplicative bias $B = \exp(b)$. Then in the additive case directly and in the multiplicative case after taking logs we have the relation

$$y_t = \eta_t + b + k_t u_t,$$

where the survey error $k_t u_t$ is assumed to have zero mean. Let $\breve{\eta}_t = \eta_t + b$. Then in both the additive and multiplicative cases, KFS applied to the y_t series provides us with the estimate $\tilde{\eta} = E(\breve{\eta}|y)$ where $\breve{\eta} = [\breve{\eta}_1, ..., \breve{\eta}_n]'$.

Our approach, described in Durbin and Quenneville (1995), is first to obtain an expression for b in terms of the unknown $\breve{\eta}_t$. Next, we concentrate out b from the log posterior density of $\breve{\eta}_t$ given y and x. Finallly, we use the previous theory to estimate $\breve{\eta}_t$ and hence b in both the additive and multiplicative cases.

6. APPLICATION TO CANADIAN RETAIL SALES DATA

The methods developed earlier will now be illustrated by applying them to monthly values of the Canadian Retail Trade Sales series from January 1980 to December 1989. These values are set out in Table A1 of Durbin and Quenneville (1995). The first four benchmarks are annual totals from February to the following January 1985-88. Retail trade data are totalled in this way because of the special nature of the transition from December to January in the retail trade. The last three benchmarks are values for October, November and December 1989 observed in a separate and redesigned monthly survey intended to replace the previous one and they form part of a linking exercise. The benchmark values are set out in Table A2 of Durbin and Quenneville (1995). Sources and other features of the data are discussed in Mian and Laniel (1993). The monthly data and the benchmarks are displayed in Figure 1.

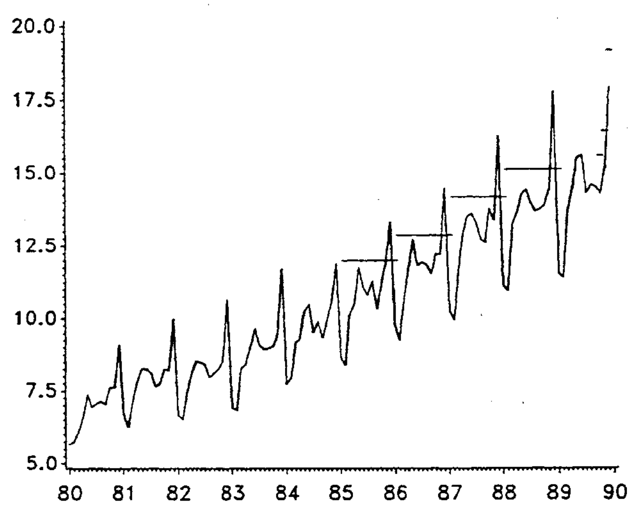

Figure 1: Canadian retail sales series from January 1980 to December 1989 and average values of the benchmarks, graphed as horizontal lines. Values are expressed in millions.

The graph of the monthly data shows an upward trend with increasing amplitude, which is typical behaviour of series that behave multiplicatively. Also, a comparison of the data with the benchmarks demonstrates clearly that the observations are negatively biased. These characteristics suggest strongly that the most suitable model for benchmarking these data is the multiplicative model with bias. However, since our objective is illustrative we have in fact analysed the data with all the methods we developed.

We modelled the survey error term u_t in (2.7) by the seasonal autoregressive model

$$(1\text{-}0.9387B)\ (1\text{-}0.8927B^{12})u_t = \chi_t, \qquad (6.1)$$

with (χ_t) chosen to make $var(u_t)=1$. We calculated the coefficients of the model from the autocorrelations of the survey errors $k_t u_t$ provided by Mian and Laniel (1993), Table 2. Since k_t is regarded as non-stochastic the autocorrelations of u_t are the same as those of $k_t u_t$.

For the trading-day and leap year coefficients we used model (2.6); we took $\sigma_\zeta^2=0$ since work by Dagum, Quenneville and Sutradhar (1992) indicated that a fixed trading day pattern is suitable for this type of data. The method of maximum likelihood of Engle and Watson (1981) was used for the estimation of the hyperparameters σ_ϵ^2, σ_ξ^2 and σ_w^2. The estimates obtained are $\sigma_\xi^2=2.5267\times10^8$, $\sigma_w^2=1.8382\times10^{10}$, and $\sigma_\epsilon^2=5.0083\times10^9$ for the additive model, and $\sigma_\xi^2=3.293\times10^{-4}$, $\sigma_w^2=1.10\times10^{-8}$, and $\sigma_\epsilon^2=1.2195\times10^{-4}$ for the multiplicative model. The estimate of the bias parameter under the additive model is $\hat{b}=-1215099$ with a coefficient of variation $cv(\hat{b})=4.0833$ per cent. For the multiplicative model, the estimate of the bias paramater is $\hat{B}=0.9140659$ with a coefficient of variation $cv(\hat{B})=0.02386$ per cent. The fact that the coefficient of variation (cv) in the multiplicative case is much smaller than in the additive case indicates the superiority of the multiplicative model for these data.

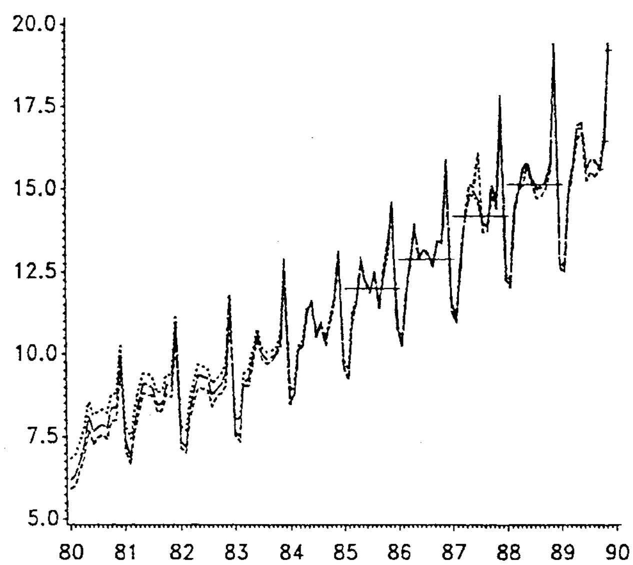

Figure 2: Benchmarked values for the Canadian retail sales series and average values of the benchmarks: Additive Without Bias (AWB) is the dashed line; Multiplicative with Bias (MB) is the broken line; Additive with Bias (AB) is the dotted line. Values are expressed in millions.

Figure 2 shows the results of benchmarking using the additive model without bias (AWB). All other methods without bias are indistinguishable on the graph from this case. The results of benchmarking for the additive model with bias (AB) and the multiplicative model with bias (MB) are also shown. These exhibit differences between each other and from the AWB case, particularly at the beginning of the series. Differences here are to be expected since there are no benchmarks at the beginning of the series. To explain the differences we note that as we move away from the benchmarks, the benchmarking correction converges to the bias parameter and this is zero for AWB, a constant for AB and a multiple of the level for MB.

There are further differences in years '87, '88 and '89 between the methods with and without bias. These are explained by the exceptionally large cv for the monthly observation in July '87, as shown in Figure 3. This is a correct value and is due to an intentional adjustment to the sample that was made by survey statisticians for technical reasons. The effect has been to distort the benchmark adjustment in '87 and, because of autocorrelation in the survey errors, in subsequent years also. In a real-life analysis the value concerned would have been treated as an outlier in the cv series and would have been adjusted accordingly. We left the value unchanged in order to illustrate what happens in such cases.

The graphs show that even with a biased series, benchmarking with a model without bias can work well in years containing benchmarks. However, the models with bias handle the outlier in the cv series better in the sense that the monthly pattern in the benchmarked series when bias is included is closer to the pattern in the original series than with AWB.

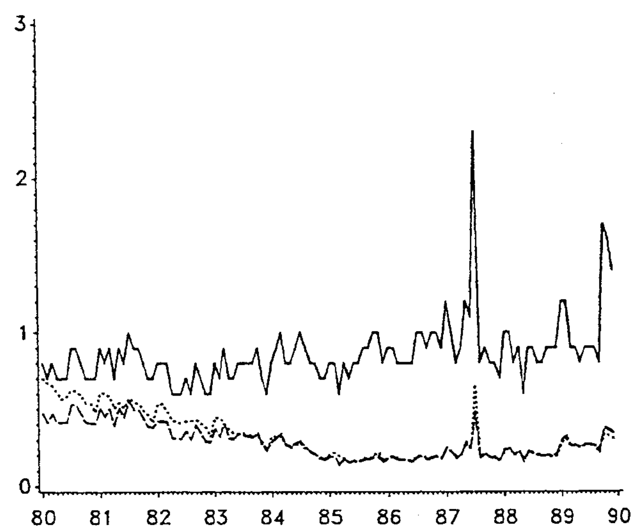

Figure 3: Coeffients of variation (cv) expressed as percentages: original data is the solid line; AB is the dotted line; MB is the broken line.

Figure 3 shows the cv's of the original series (also provided in Table A3 of Durbin and Quenneville, 1995) and of AB and MB; all other methods give graphs which are indistinguishable from that of MB. It is clear that benchmarking reduces the cv by an amount which has considerable intrinsic value, quite apart from the general desirability of adjusting the monthly values to satisfy the benchmark constraints. The fact that in the first part of the series the cv of AB is higher than that of MB is presumably due to the higher cv of the estimate of the bias parameter for the additive model compared with that for the multiplicative model.

For this particular series our conclusion is that the most appropriate model is the multiplicative model with bias correction. Our overall conclusion is that state space models provide an excellent basis for benchmarking time series, which has substantial potential for further development.

7. REFERENCES

Cholette, P.A. and Dagum, E.B. (1994). "Benchmarking time series with autocorrelated survey errors", *International Statistical Review*, 62, 365-377.

Dagum, E.B., Quenneville, B. and Sutradhar, B. (1992). "Trading-day variations multiple regression models with random parameters", *International Statistical Review*, 60, 57-73.

Durbin, J. and Cordero, M. (1994). "Handling structural shifts, outliers and heavy-tailed distributions in state space time series models". To appear.

Durbin, J. and Koopman, S.J. (1994). "Filtering, smoothing and estimation for time series when the observations come from exponential family distributions". To appear.

Durbin, J. and Quenneville, B. (1995). "Benchmarking by state space models". To appear.

Engle, R. and Watson, M. (1981). "A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates", *Journal of the American Statistical Association*, 76, 774-781.

Farmeir, L. (1992). "Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalised linear models", *Journal of the American Statistical Association*, 38, 37-60.

Harvey, A.C. (1989). *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.

Hillmer, C.H. and Trabelsi, A. (1987). "Benchmarking of economic time series", *Journal of the American Statistical Association*, 82, 1064-1071.

Mian, I.U.H. and Laniel, N. (1993). "Maximum likelihood estimation of constant multiplicative bias benchmarking model with application", *Survey Methodology*, 19, 165-172.