BAYESIAN FORECASTING

Mike West

ISDS, Duke University

Bayesian Forecasting encompasses statistical theory and methods in time series analysis and time series forecasting, particularly approaches using dynamic and state space models, though the underlying concepts and theoretical foundation relate to probability modelling and inference more generally. This entry focuses specifically in the time series and dynamic modelling domain, with mention of related areas.

Background

Bayesian forecasting and dynamic modelling has a history that can be traced back to the late 1950s in short-term forecasting and time series monitoring in commercial environments ([19, 20]), and many of the developments since then have retained firm links with the roots in applied modelling and forecasting problems in industrial and socio-economic areas ([21, 22, 23, 24, 46, 71]). Almost in parallel, technically similar developments arose in the systems and control engineering areas ([29, 30]), with a central focus on adaptive estimation and filtering theory for automatic control. Central to the field are state-space models for time-varying systems, and the statistical

1

methodology associated with problems of estimation and inference for time series using these models. These methods are most naturally viewed from a Bayesian perspective, and, indeed, the developments in both forecasting and control environments can be seen as one of the important practical successes of Bayesian methods.

In the time series analysis and forecasting area, Bayesian methods based on dynamic linear models were initiated largely as outgrowths of simple and widely used smoothing and extrapolation techniques, especially exponential smoothing and exponentially weighted moving average methods ([20, 71]). Developments of smoothing and discounting techniques in stock control and production planning areas led to formalisms in terms of linear, state-space models for time series with time-varying trends and seasonal patterns, and eventually to the associated Bayesian formalism of methods of inference and prediction. From the early 1960s, practical Bayesian forecasting systems in this context involved the combination of formal time series models and historical data analysis together with methods for subjective intervention and forecast monitoring, so that complete forecasting systems, rather than just routine and automatic data analysis and extrapolation, were in use at that time ([19, 22]). Methods developed in those early days are still in use now in some companies in sales forecasting and stock control areas. There have been major developments in models and methods since then, especially promulgated by the seminal article ([23]), though the conceptual basis remains essentially unchanged. The more recent history of the field is highlighted in [69], various articles in [61], the books [46, 71], and in commentary below.

Parallel developments in control engineering highlighted the use of sequential statistical learning and optimisation in linear state-space models. From the early foundational work of Wiener, this field developed significantly in the 1960s, and was hallmarked by the key contributions of Kalman, [29] and [30]. The approach to statistical inference based on linear least squares was central, and the associated sequences of

2

updating equations for state parameter estimation are now widely referred to as the Kalman Filter equations. Variants of Kalman Filtering techniques now permeate various engineering areas, and the relationships with Bayesian forecasting in dynamic linear models are obviously intricate ([2, 3, 39, 48, 61] and [71, section 1.4]). The same may be said of the related developments of structural modelling in econometric time series analysis ([26]). In recent times, Bayesian concepts and methods have become more apparent in these related fields.

During the last two decades, the range of developments and applications of Bayesian forecasting and dynamic models has grown enormously. Much of the development involves theoretical and methodological extensions mentioned below. It is simply impossible to do justice to the field here in terms of a comprehensive review of application areas and contexts, even ignoring the vast engineering control arena. To give some of the flavour of the breadth and scope of the field, the results of a scan of the CIS data base in early 1995 reveal application areas running through the traditional fields of commerical and industrial time series forecasting, including sales and demand forecasting (e.g. [17, 28, 50, 71]), energy demand and consumption, advertising market research (e.g. [42]), inventory/stock management (e.g. [21]), and others. This is complemented by ranges of applications in macro-economics and econometrics, especially in connection with problems in financial modelling, structural change, aggregation and combining forecasts/models (e.g. [51, 52, 53, 49]). There are similar ranges of applications in biomedical monitoring contexts (e.g. [16, 59]), and specific kinds of applications in focussed areas including reliability analysis (e.g. [6, 40]), survival analysis in biostatistics and economics ([12, 13, 64]), point processes ([14]), geological data analysis ([67]), hydrological forecasting and river dam management ([54, 55]), competitive bidding ([4]), spectral analysis (e.g. [32, 66]), and many others including traffic accident forecasting, market research, finite population inference, prediction in queues, small area estimation, optimal design and control, novel directions in spatial

3

analysis and imaging, and more. This is a vibrant and rapidly developing field, and one likely to continue to grow well into the 21st century.

Dynamic Bayesian Models

Dynamic models provide the central technical components of Bayesian forecasting methods and systems ([46, 71]). In by far the most widely used incarnation as dynamic linear models, and with associated assumptions of normally distributed components, they also feature centrally in various areas of classical time series analysis and control areas (as state-space models, [2, 3]), and econometrics (as state-space or structural time series models, [26]).

The approach of Bayesian forecasting and dynamic modelling comprises, fundamentally,

- sequential model definitions for series of observations observed over time,

- structuring using parametric models with meaningful parametrisations,

- probabilistic representation of information about all parameters and observables, and hence

- inference and forecasting derived by summarising appropriate posterior and predictive probability distributions.

The probabilistic view is inherent in the foundation in the Bayesian paradigm, and has important technical and practical consequences. First, the routine analysis of time series data, and the entire process of sequential learning and forecasting, is directly determined by the laws of probability. Joint distributions for model parameters and future values of a time series are manipulated to compute parameter estimates and point forecasts, together with a full range of associated summary measures of

4

uncertainties. Second, and critically in the context of practical forecasting systems that are open to external information sources, subjective interventions to incorporate additional information into data analysis and inference are both explicitly permissible and technically feasible within the Bayesian framework. All probability distributions represent the views of a forecaster, and these may be modified/updated/adjusted in the light of any sources of additional information deemed relevant to the development of the time series. The only constraint is that the ways in which such additional information is combined with an existing model form and the historical information abide by the laws of probability.

The sequential approach focuses attention on statements about the future development of a time series conditional on existing information. Thus, if interest lies in the scalar series Y_t, statements made at time $t - 1$ are based on the existing information set D_{t-1}, whatever this may be. These statements are derived from a representation of the relevant information obtained by structuring the beliefs of the forecaster in terms of a parametric model defining the probability density $f_Y(Y_t | \boldsymbol{\theta}_t, D_{t-1})$, where $\boldsymbol{\theta}_t$ is the defining parameter vector at time t. Through conditioning arguments, the notation explicitly recognises the dependence of y_t on the model parameters and on the available historical information. Parameters represent constructs meaningful in the context of the forecasting problem. For example, $\boldsymbol{\theta}_t$ may involve terms representing the current level of the Y_t series, regression parameters on independent variables, seasonal factors, and so forth. Indexing $\boldsymbol{\theta}_t$ by t indicates that the parametrisation may be dynamic, i.e. varying over time through both deterministic and stochastic mechanisms. In some cases, the parameter may even change in dimension and interpretation.

Normal Dynamic Linear Models

The class of normal dynamic linear models (DLMs) are central to Bayesian forecasting and time series analysis. The basic model over all time t is defined by the observation and evolution equations

$$Y_t = \boldsymbol{F}'_t \boldsymbol{\theta}_t + \nu_t,$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t,$$

with components as follows:

- $\boldsymbol{\theta}_t$ is the state vector at time t;

- $\boldsymbol{F}_t$ is a known vector of regression variables and constants;

- ν_t is an observation noise term, representing measurement and sampling errors corrupting the observation of Y_t, assumed normally distributed with zero mean and known variance v_t, i.e. $N(\nu_t|0, v_t)$;

- $\boldsymbol{G}_t$ is the state evolution matrix, defining the deterministic map of state vectors between times $t-1$ and t; and

- $\boldsymbol{\omega}_t$ is the evolution noise term, or innovation, representing stochastic changes in the state vector and assumed normally distributed with zero mean and known variance matrix $\boldsymbol{W}_t$, i.e. $N(\boldsymbol{\omega}_t|0, \boldsymbol{W}_t)$.

Additionally, the noise sequences ν_t and $\boldsymbol{\omega}_t$, over time t, are assumed independent and mutually independent. Variations on this basic framework allow for correlated noise sequences, non-zero mean noise terms, and other minor modifications.

The model structure is Markovian, with state vectors varying over time according to a linear, Markov evolution equation. The class of DLMs includes many kinds of

time series models, such as models with time varying, "smooth" trends, time varying seasonal/periodic behaviour, regression effects in which the regression parameters may change (usually slowly) over time, transfer responses, and others including stationary and non-stationary variants of ARMA models ([71]).

Routine analysis of the sequentially received data series involves sequentially updating summary statistics that characterise sequences of posterior/predictive distributions for inference about subsets of the $\boldsymbol{\theta}_t$ over all time t and for future values of the Y_t series. Assuming the only information used in updating is the set of observed values $y_1, y_2, \ldots$, we have a closed model in which information updates via $D_t = \{D_{t-1}, y_t\}$ at each time point, given an initial information set D_0 (at an arbitrary time origin $t = 0$). Assuming a normal distribution for the initial state-vector, $\boldsymbol{\theta}_0$, we have complete joint normality of all the $\boldsymbol{\theta}_t$ and Y_s, and the sequential updating of distributions is based, essentially, on the so-called Kalman Filter equations. At time t, we have a current posterior for the current state vector defined by $(\boldsymbol{\theta}_t|D_t) \sim N(\boldsymbol{\theta}_t|\boldsymbol{m}_t, \boldsymbol{C}_t)$ with the updating defined via

$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t e_t \quad \text{and} \quad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t \boldsymbol{A}_t' q_t$$

where $\boldsymbol{a}_t = \boldsymbol{G}_t \boldsymbol{m}_{t-1}$, $\boldsymbol{R}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}_t' + \boldsymbol{W}_t$, $q_t = \boldsymbol{F}_t' \boldsymbol{R}_t \boldsymbol{F}_t + v_t$ and $e_t = y_t - \boldsymbol{F}_t' \boldsymbol{a}_t$. By way of interpretation, $\boldsymbol{a}_t$ and $\boldsymbol{R}_t$ are the defining moments of the state prior, or state prediction, distribution at time $t - 1$, $f_\theta(\boldsymbol{\theta}_t|D_{t-1}) = N(\boldsymbol{\theta}_t|\boldsymbol{a}_t, \boldsymbol{R}_t)$, q_t is the variance of the one-step ahead predictive distribution $f_Y(Y_t|D_{t-1}) = N(Y_t|F_t' a_t, q_t)$, and e_t is the observed forecast error.

This set of equations provides a computationally efficient algorithm for sequentially updating the distributional summaries. Related forecasting and smoothing algorithms ([71, chapter 4]) provide for computation of the practically relevant distributions

- $f_Y(Y_{t+k}|D_t)$, for $k > 0$, the $k-$step ahead forecast distribution at time t, and

7

- $f_\theta(\boldsymbol{\theta}_{t-k}|D_t)$, for $k > 0$, the $k-$step filtered distribution at time t.

Forecasting is based on the former, and retrospective time series analysis and decomposition, especially in connection with the evaluation of changes over time in elements of the state vector, is based on the latter.

Various practically important extensions of this basic model are discussed in [71]. Two of key practical relevance include estimation of the observation variances v_t, possibly constant though often slowly varying over time, perhaps with additional weights ([71]), and the use of uninformative or reference initial prior distributions for the initial state vector ([44]).

Practical forecasting systems based on dynamic models require consideration of issues of model specification and structuring, forecast monitoring, and intervention. These kinds of developments, discussed below, are covered in detail in [71] and [46].

Component DLMs

Model building and structuring relies heavily on the development of overall DLMs from simpler and specific components. Refer to the DLM described earlier by the model quadruple $\{\boldsymbol{F}_t, \boldsymbol{G}_t, v_t, \boldsymbol{W}_t\}$, defined for all t. DLMs are built from individual components describing features such as slowly varying trends, usually local polynomials, seasonal components, regression components, and possibly others including residual AR or ARMA components. Each component can be viewed as a sub-DLM, and the linear combination of components provides an overall DLM for the series. Suppose a set of $m > 1$ independent DLMs is defined by individual elements model quadruples $\{\boldsymbol{F}_{i,t}, \boldsymbol{G}_{i,t}, v_{i,t}, \boldsymbol{W}_{i,t}\}$, for all t and $i = 1, \ldots, m$; write $Y_{i,t}$ for the observation on the ith series at time t. Add the series following these models, to obtain a series $Y_t = \sum_{i=1}^{m} Y_{i,t}$. Then Y_t follows a DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, v_t, \boldsymbol{W}_t\}$, in which: (a) the state vec-

tor, regression vector, and evolution innovation vector are obtained via the catenation of the corresponding elements of the individual models, so that $\boldsymbol{\theta}_t = (\theta'_{1,t}; \cdots; \theta'_{m,t})'$, with a similar form for $\boldsymbol{F}_t$ and $\boldsymbol{\omega}_t$ in the overall model; (b) the observation error variances are added, $v_t = \sum_{i=1}^{m} v_{i,t}$; and (c) the evolution matrices are built from the components in block diagonal form, namely $\boldsymbol{G}_t = \text{block diag}(\boldsymbol{G}_1; \cdots; \boldsymbol{G}_m)$ and $\boldsymbol{W}_t = \text{block diag}(\boldsymbol{W}_1; \cdots; \boldsymbol{W}_m)$. The model for Y_t is said to be formed by the superposition of the component DLMs.

Key model components for many applications are those for smooth trends and seasonal patterns, in which the $\boldsymbol{F}$ and $\boldsymbol{G}$ elements are time-independent. For example, a second-order polynomial (or locally-linear) trend is defined via

$$\boldsymbol{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{G} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

with higher order, local polynomial forms by extension ([71, chapter 7]).

Fourier representations of seasonal or cyclical components are based on the time-varying sinusoidal DLM with

$$\boldsymbol{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{G} = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix},$$

where $\alpha = 2\pi/p$ is the frequency and p the period or wavelength. In typical cases, p is an integer, e.g. $p = 12$ for a monthly seasonal pattern, and the above cyclical DLM describes a sine wave of varying amplitude and phase but fixed period p. More complicated seasonal patterns require additional periodic components at harmonic frequencies, thus adding components like that above but with frequencies $2\alpha, 3\alpha$, etc. Often a complete description of an arbitrary seasonal pattern of integer period p requires the full set of harmonic components ([24] and [71, chapter 8]).

There are often various alternative DLM representations of specific component forms, choice among which is something of a matter of taste, convenience, and interpretation. For example, the above cyclical component is alternatively modelled

via

$$\boldsymbol{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{G} = \begin{pmatrix} 2\cos(\alpha) & -1 \\ 1 & 0 \end{pmatrix},$$

a special case of an autoregressive DLM, here defining an AR(2) component with a unit root (e.g. [71, chapter 5], and [66]). Extensive discussion of component representation and this issue of similar models with differing $\boldsymbol{F}$ and $\boldsymbol{G}$ appears in [71, chapter 5 and 6]. This reference also includes wider development of the connections with linear systems theory.

Other useful components include regression effects, in which $F_{i,t}$ is a regressor variable and $G_{i,t} = 1$, and more general regression transfer function models ([47, 71]). ARMA components may also be developed, in various forms ([3, 26, 33, 34, 66]), and the use of, in particular, AR component DLMs is becoming more widespread with the advent of appropriate computational methods (see below).

Discounting

In dynamic modelling generally, and with DLMs in particular, the notion of information discounting, and its use in structuring models, has been usefully exploited. Given a DLM form defined by $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$, model structuring is completed by specification of the evolution variance matrix $\boldsymbol{W}_t$, a key ingredient that controls and determines the extent and nature of stochastic changes in state vectors over time. In practical models, the dimension of $\boldsymbol{\theta}_t$ may range up to 12-15, or more, so that this is a challenging specification task unless $\boldsymbol{W}_t$ is structured in terms of just a few basic parameters. Traditional concepts of information discounting in time series led to the development of methods of DLM component discounting that rather neatly and naturally provide practitioners with interpretable ways of specifying evolution variance matrices ([71, chapter 6]).

For example, the various DLMs implemented in the BATS package ([46]) associate a single discount with each of the polynomial trend, seasonal, and regression components of an overall model. These determine individual evolution variances $\boldsymbol{W}_{i,t}$ in the block diagonal form of $\boldsymbol{W}_t$, and provide opportunity to model varying degrees of smoothness of, or, reciprocally, variation in, the individual components over time. This can be absolutely critical in terms of both short-term forecasting performance and in adapting to changes in component structure over time in practical time series analysis.

Intervention

In practice, statistical models are only components of forecasting systems, and interactions between forecasters and models is necessary to adequately account for events and changes that go beyond the existing model form. One of the major features of Bayesian forecasting is the facility for integrating intervention information into existing DLMs, and other models ([70]). Interventions can be classified as either feed-forward or feed-back. The former is anticipatory, providing opportunity to anticipate changes in a series as a result of control actions or changes in the environment and context generating the series. This is very common in commercial environments, for example. Feeb-back interventions are, on the other hand, corrective actions taken to adjust models for recent events that had not been foreseen nor adequately catered for within the existing model.

Various modes of intervention in forecasting models are described in [70] and [71], including a range of practical contexts and exploration of technical aspects of intervening in specified models with existing probabilistic summaries. Key concepts involved include: the deletion of observations viewed as suspect or possible outliers; the imposition of constraints on particular model components to protect against per-

haps spurious adjustments in updating the distributions of state vectors; the extension of an existing DLM form to include additional components representing perhaps transient intervention effects; and, perhaps most importantly, the simple input of additional uncertainty about state vector components, usually by momentarily inflating the relevant elements of the evolution variance matrix. Ranges of applications and illustrations are given in [46, 70] and [71].

Monitoring and Adaptation

The principle of "management by exception" has been a guide to the development of some forecasting systems. This suggests that, ideally, an adopted model, or set of models, is used routinely to process data and information, providing forecasts and inferences that are used unless exceptional circumstances arise. In exceptional cases, either identified in advance or by the detection of deterioration in model/forecasting performance, some kind of intervention is called for. This raises the need for monitoring of forecast performance and model fit in order to signal the need for feed-back, or retrospective, interventions in attempting to cater for exceptional events that occur at unforeseen times.

Typical applications of dynamic models involve various sources of exceptional events. Outliers in the data may occur singly or in groups. Elements of the state vector $\boldsymbol{\theta}_t$, at a single time point t, may change in ways that are greater than predicted under the existing evolution model, being consistent with larger variance elements and perhaps different covariance structure for the $\boldsymbol{W}_t$ matrix. These changes may be quite abrupt, or they may be more subtle and difficult to detect, with parameters "drifting" in unpredicted ways over the course of a few observation periods. Further departures from model form may be due to new external effects that are best explained by additional model components. Whatever the exception, monitoring of

forecast performance to "flag" deteriorating forecast accuracy is the starting point for model adaptation via feed-back interventions. Various monitoring techniques exist and are in routine application, including variants of traditional cusum methods, and more general Bayesian sequential monitoring methods that are sensitive to forecast deterioration from both single observations, that may be outliers, or groups of a few recent observations, that may represent perhaps more subtle changes in state vector values ([46, 63, 69, 71]). These kinds of techniques are central to the recent development of applied Bayesian forecasting systems. Aspects of model monitoring and assessment related to traditional residual and regression diagnostics, including influence analysis, have also been developed ([25, 34]), though are of more interest and utility in contexts of retrospective time series analysis than in progressive, sequential analysis with a forecasting and updating focus.

Mixtures of Dynamic Models

Probabilistic mixtures of sets of normal DLMs provide practically useful classes of (explicitly non-normal) forecasting models, and have been in use since at least the early 1970s in both the Bayesian forecasting and engineering areas ([2, 22, 60]). There are many variants, and much recent work related to model mixing in time series, though the kinds of applications can be broadly classified into two groups (sometimes referred to as multi-process models, classes one and two ([23] and [71, chapter 12])). In the first case, an assumed DLM form is accepted as appropriate for a time series, but there are defining parameters (such as variance components, discount factors, elements of the state evolution matrix G, and so forth) that are uncertain and assumed to belong to a specified, fixed, and finite set of values. Sequential analysis then formally includes these parameters; the DLM mixture structure can be viewed as induced by marginalisation with respect to these parameters. This is precisely the

traditional approach to Bayesian model selection and model averaging.

The second class of mixtures models assumes, in contrast, that, at any time point t, the model generating the observation is drawn from a set of DLMs, with the choice of model at any time determined by a probability distribution across the set. In general, the probabilities over models at any time may depend on the past selection of models, and on past data, so this is a rather general framework for model switching ([58]). The early work with this approach in connection with the modelling of time series outliers and abrupts changes in time series structure ([23, 59]) is precursor to more recent work in related areas and with other model classes (e.g. [7, 11] and [71, chapter 12]).

Historically, computational constraints have limited wider development and application of mixtures of models, particularly in this second class. Sequential and retrospective analysis becomes quite challenging computationally in even rather modest mixture models, though various analytic approximation methods have been very useful. More recently, the advent of efficient simulation methods have provided potential to overcome these constraints ([10, 11]) and much more can be done in terms of realistic modelling of the complexities of problems. This is currently a vibrant research area that will surely grow substantially in coming years.

Non-Normal/Non-Linear Models

From the 1960s, analytic techniques of approximation for analyses of dynamic non-linear models have been very widely used in engineering areas, typically under the names of extended Kalman filtering and its variants ([2]). Similar methods have been used in Bayesian forecasting, though here wider generalisation of DLMs has focussed on non-normal (non-Gaussian) models. A variety of areas have generated interest in dynamic models with non-normal distributions for either the time series

14

observations (non-normality of observation models $f_Y(Y_t|\boldsymbol{\theta}_t, D_{t-1})$), or the evolution equation (non-normality of state evolution models $f_\theta(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, D_{t-1})$), or both.

In addition to the developments using mixtures of DLMs, there has been some interest in other approaches to modelling outliers and change-points using heavy-tailed error distributions, such as t-distributions, in place of normal forms ([31, 32, 41, 43, 62]). Perhaps the major developments of non-normal models relate to the synthesis of dynamic and Bayesian modelling concepts with the widely used class of generalised linear models, for problems of time series analysis and forecasting with discrete data, such as binary or Poisson counts, binomial response data, and other error models in the exponential family class ([42, 43, 68]). Related models, concerning multivariate extensions for compositional data series, appear in [18] and [8].

A rather novel field of development of Bayesian dynamic models is the survival analysis arena ([12, 13, 14, 64]). Here the data arise in the traditional survival context, as observed or censored failure times for collections of units, such as diseased and treated individuals, over time. The development of dynamic models for time-varying hazard functions and effects of explanatory variables is a direct extension of dynamic generalised linear modelling. The utility of these models as tools for exploring non-proportional hazards structure, and time-variation in the effects of explanatory variables, is illustrated in biostatistical and economic applications in some of the references just noted.

Analysis in non-normal/non-linear models has historically involved ranges of creative analytic/numerical approximation to overcome computational problems arising in implementation. However, during the last few years the revolution in Bayesian statistics through the development of simulation methods of analysis has provided impetus to develop non-standard models as these kinds of computational methods are very widely applicable. Some early examples of non-normal/non-linear dynamic modelling, with analysis made feasible via simulation methods, appear in [9, 27].

15

Multivariate Models

Various theoretical classes of dynamic linear models for multivariate time series exist, though the kinds of multivariate applications of note have tended to develop models quite specific to the application context (e.g. [50, 71]). Probably the most widely used models are those based on matrix-variate extensions of basic DLMs ([51, 52]). These are particularly appropriate in contexts of modelling several or many univariate series that are viewed as structurally similar, or perhaps exchangeable, in the sense of sharing common defining elements $\boldsymbol{F}_t$, $\boldsymbol{G}_t$, and $\boldsymbol{W}_t$. These models have been, and continue to be, usefully applied in macro-economics and, especially, finance (e.g. [53]), among other areas, where the multivariate structure of most importance is contemporaneous, rather than predictive. Modifications of these multivariate models, as well as quite distinct, non-normal dynamic models, have been developed for problems in which data series represent proportions or compositions ([18, 52] and [8]). One area that has seen some exploration to date, and is likely to be a growth area in future, is the development of multivariate dynamic models using traditional Bayesian hierarchical modelling concepts and methods ([72, 73]). Though not explicitly based in dynamic linear models, the Bayesian forecasting approaches using vector autoregression and structured prior distributions ([35, 36]) represent a significant area of development, and an important branch of the wider Bayesian forecasting arena.

Computation and Simulation

Early work on numerical methods of Bayesian computation for dynamic models include methods using mixtures ([22, 23, 60] and references in [2] and [71, chapter 12]), developments based on numerical quadrature ([31, 45, 45]), and traditional Monte Carlo simulation approaches ([65]).

Since the early 1990s, the field has been given dramatic impetus by the development of iterative simulation methods, particularly Markov chain Monte Carlo using Gibbs sampling, Metropolis-Hastings methods, and variants. The entire field of dynamic modelling (state space and structural time series modelling) is likely to experience much wider use of such methods in coming years. As in other areas of applied Bayesian statistics, researchers can now focus very much more on developing realistic models of specific contexts, in the knowledge that very general and powerful computational techniques are at hand to implement resulting analyses of some complexity.

Basic work in this area stems from ([9, 10, 11]), with recent extensions and applications in ([57, 66]), and elsewhere. In the context of a rather generic dynamic model, suppose a sequence of defining observation and evolution equations are described in terms of conditional densities, i.e.

$$f_Y(Y_t|\boldsymbol{\theta}_t, \boldsymbol{\phi}, D_{t-1}) \quad \text{and} \quad f_\theta(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \boldsymbol{\phi}, D_{t-1}),$$

where now additional parameters $\boldsymbol{\phi}$ are made explicit in conditionings; $\boldsymbol{\phi}$ may contain, for example, observation and evolution variance components in a normal model, elements of state evolution matrices $\boldsymbol{G}$ in a linear model, and other parameters. These distributions may involve non-normal and non-linear components, as in [9, 57], for example. Inference is desired for collections of state vectors, current and past, future observations on the Y_t series, and the defining model parameters $\boldsymbol{\phi}$. Iterative simulation methods are focussed on the evaluation of posterior and predictive distribution inferences by sampling representative values of the state vectors, parameters, and observations, from the relevant distributions. The utility of this approach is illustrated in various interesting contexts in the above references, indicating that this novel area is one with with promise for the future.

Related methodological and computational issues, with interesting complications,

arise in chaotic time series modelling ([5, 56]).

Related Areas

The foregoing discussion is rather specifically focussed on the Bayesian dynamic modelling area. Bayesian time series analysis and forecasting is a much wider field than so far represented, and, from the outset, no attempt has been made to cover the field in any generality. To offset this somewhat, some references are made here to provide access to areas of time series modelling and Bayesian analysis more widely. As noted, there are strong connections with (largely non-Bayesian) structural times series modelling in econometrics (e.g. [26]), with the use of state-space models and Kalman filtering techniques in ARMA modelling and other areas (e.g. [3, 33, 34]), and at the interface of statistics and the control engineering fields (e.g. [3, 61]). In connection, in particular, with multivariate time series forecasting, important Bayesian work has been developing in econometrics and finance. Notable areas include the applications in multivariate forecasting using vector autoregressions and related models (e.g. [35, 36]), and ranges of applications of models combining dynamic linear model components with hierarchical/shrinkage methods (e.g. [15, 72, 73]). Further, based on the advent of simulation methods for Bayesian analysis, there has recently been much activity in developments of computational Bayesian methods in other, closely related, areas of time series (e.g. [1, 37, 38]).

References

[1] Albert, J.H., and Chib, S. (1993) Bayesian inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of*

Business and Economic Statistics, **11**, 1-15.

[2] Anderson, B.D.O., and Moore, J.B. (1979) *Optimal Filtering.* Prentice-Hall, New Jersey.

[3] Aoki, M. (1990) *State Space Modelling of Time Series.* Springer-Verlag, New York.

[4] Attwell, D.N., and Smith, J.Q. (1991) A Bayesian forecasting model for sequential bidding. *Journal of Forecasting,* **10**, 565-577.

[5] Berliner, L.M. (1992) Statistics, probability, and chaos (with discussion). *Statistical Science,* **7**, 69-122.

[6] Blackwell, L.M., and Singpurwalla, N.D. (1988) Inference from accelerated life tests using filtering in coloured noise. *Journal of the Royal Statistical Society,* (Series B), **50**, 281-292.

[7] Bolstad, W.M. (1988) Estimation in the multiprocess dynamic generalized linear model. *Communications in Statistics, Series A–Theory and Methods,* **17**, 4179-4204.

[8] Cargnoni, C., Müller, P., and West, M. (1996) Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association,* (to appear).

[9] Carlin, B.P., Polson, N.G., and Stoffer, D.S. (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *Journal of the American Statistical Association,* **87**, 493-500.

[10] Carter, C.K., and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika* **81**, 541-553.

[11] Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis,* **15**, 183-102.

[12] Gamerman, D., and West, M. (1987) A time series application of dynamic survival models in unemployment studies. *The Statistician,* **36**, 269-174.

[13] Gamerman, D. (1991) Dynamic Bayesian models for survival data. *Applied Statistics,* **40**, 63-79.

[14] Gamerman, D. (1992) A dynamic approach to the statistical analysis of point processes. *Biometrika,* **79**, 39-50.

[15] Garcia-Ferrer, A., Highfield, R.A., Palm, F., and Zellner, A. (1987) Macroeconomic forecasting using pooled international data. *Journal of Business and Economic Statistics,* **5**, 53-67.

[16] Gordon, K. and Smith, A.F.M. (1990) Modeling and monitoring biomedical time series *Journal of the American Statistical Association,* **85**, 328-337.

[17] Green, M, and Harrison, P.J. (1973) Fashion forecasting for a mail order company. *Operations Research Quarterly,* **24**, 193-205.

[18] Grunwald, G.K., Raftery, A.E., and Guttorp, P. (1993) Time series of continuous proportions. *Journal of the Royal Statistical Society,* (Series B), **55**, 103-116.

[19] Harrison, P.J. (1965) Short-term sales forecasting. *Applied Statistics,* **15**, 102-139.

[20] Harrison, P.J. (1967) Exponential smoothing and short-term forecasting, *Management Science,* **13**, 821-842.

[21] Harrison, P.J. (1988) Bayesian forecasting in O.R. In *Developments in Operational Research 1988*, N.B. Cook and A.M. Johnson, eds. Pergamon Press, Oxford.

[22] Harrison P.J., and Stevens, C. (1971) A Bayesian approach to short-term forecasting. *Operations Research Quarterly*, **22**, 341-362.

[23] Harrison P.J., and Stevens, C. (1976) Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society*, (Series B), **38**, 205-247.

[24] Harrison, P.J., and West, M. (1987) Practical Bayesian forecasting. *The Statistician*, **36**, 115-125.

[25] Harrison, P.J., and West, M. (1991) Dynamic linear model diagnostics. *Biometrika*, **78**, 797-808.

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

[27] Jaquier, E., Polson, N.G., and Rossi, P.E. (1994) Bayesian analysis of stochastic volatility. *Journal of Business and Economic Statistics*, **12**, 371-417.

[28] Johnston, F.R., and Harrison, P.J. (1980) An application of forecasting in the alcoholic drinks industry. *Journal of the Operational Research Society*, **31**, 699-709.

[29] Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82**, 35-45.

[30] Kalman, R.E. (1963) New methods in Wiener filtering theory. In *Proceedings of the First Symposium of Engineering Applications of Random Function Theory and Probability*, J.L. Bogdanoff and F. Kozin, eds. Wiley, New York.

[31] Kitagawa, G. (1987) Non-Gaussian state-space modelling of nonstationary time series (with discussion). *Journal of the American Statistical Association,* **82**, 1032-1063.

[32] Kleiner, B., Martin, R.D., and Thompson, D.J. (1979) Robust estimation of power spectra (with discussion). *Journal of the Royal Statistical Society,* (Series B), **41**, 313-351.

[33] Kohn, R., and Ansley, C.F. (1986) Estimation, prediction and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association,* **81**, 751-761.

[34] Kohn, R., and Ansley, C.F. (1993) A fast algorithm for signal extraction, influence and cross-validation in state-space models. *Biometrika,* **76**, 65-79.

[35] Litterman, Robert B. (1986) Specifying vector autoregressions for macroeconomic forecasting. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti,* P.K. Goel and A. Zellner, eds. North-Holland, Amsterdam.

[36] Litterman, Robert B. (1986) Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business and Economic Statistics,* **4**, 25-38.

[37] McCulloch, R., and Tsay, R. (1993) Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association,* **88**, 968-978.

[38] McCulloch, R., and Tsay, R. (1994) Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis,* **15**, 235-250.

[39] Meinhold, R.J., and Singpurwalla, N.D. (1983) Understanding the Kalman filter. *The American Statistician,* **37**, 123-127.

[40] Meinhold, R.J., Singpurwalla, N.D. (1985) A Kalman filter approach to accelerated life testing. In *Theory of Reliability,* A. Serra and R.E. Barlow, eds. North-Holland, Amsterdam.

[41] Meinhold, R.J., and Singpurwalla, N.D. (1989) Robustification of Kalman filter models. *Journal of the American Statistical Association,* **84**, 479-486.

[42] Migon, H.S., and Harrison, P.J. (1985) An application of non-linear Bayesian forecasting to television advertising. In *Bayesian Statistics 2,* J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds. North-Holland, Amsterdam, and Valencia University Press.

[43] Pole, A., West, M., Harrison, P.J. (1988) Non-normal and non-linear dynamic Bayesian modelling. In *Bayesian Analysis of Time Series and Dynamic Models,* J.C. Spall, ed. Marcel Dekker, New York.

[44] Pole, A., and West, M. (1989) Reference analysis of the DLM. *Journal of Time Series Analysis,* **10**, 131-147.

[45] Pole, A., and West, M. (1990) Efficient Bayesian learning in non-linear dynamic models. *Journal of Forecasting,* **9**, 119-136.

[46] Pole, A., West, M. and Harrison, P.J. (1994) *Applied Bayesian Forecasting and Time Series Analysis.* Chapman-Hall, New York. (An applied treatment of the central classes of dynamic models with many examples and applications, together with the BATS computer software, manual and user guide.)

[47] Pole, A., Harrison, P.J., and West, M. (1995) Forecasting and assessing structural change: the case of promotional campaigns. In *Advances in Econometrics (Vol. 11B): Bayesian Methods Applied to Time Series Data.* R. Carter Hill and T. Fomby, eds. JAI Press Inc.

[48] Priestley, M.B. (1980) System identification, Kalman filtering, and stochastic control. In *Directions in Time Series,* D.R. Brillinger and G.C.Tiao, eds. Institute of Mathematical Statistics.

[49] Putnam, B.H., and Quintana, J.M. (1994) New Bayesian statistical approaches to estimating and evaluating models of exchange rates determination. In *1994 Proceedings of the Section on Bayesian Statistical Science,* Alexandria: American Statistical Association.

[50] Queen, C.M., and Smith, J.Q. (1992) Dynamic graphical models. In *Bayesian Statistics 4,* J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith, eds. Oxford University Press.

[51] Quintana, J.M., and West, M. (1987) Multivariate time series analysis: new techniques applied to international exchange rate data. *The Statistician,* **36**, 275-281.

[52] Quintana, J.M., and West, M. (1988) Time series analysis of compositional data. In *Bayesian Statistics 3,* J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds. Oxford University Press.

[53] Quintana, J.M. (1992) Optimal portfolios of forward currency contracts. In *Bayesian Statistics 4,* J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith, eds. Oxford University Press.

[54] Rios Insua, D., and Salewicz, K. (1995) The operation of Kariba Lake: a multi-objective analysis. *Journal of Multicriteria Decision Analysis,* (to appear).

[55] Rios Insua, D., Salewicz, K., and Bielza, C. (1995) Bayesian methods in reservoir operations. *Case Studies in Bayesian Analysis,* (eds: S. French and J.Q. Smith), (to appear).

[56] Scipione, C.M., and Berliner, L.M. (1993) Bayesian statistical inference in non-linear dynamical systems. In the *1993 Proceedings of the Bayesian Statistical Science Section of the ASA.*

[57] Shephard, N. (1994) Partial non-Gaussian state space models. *Biometrika,* **81**, 115-131.

[58] Shumway, R.H., and Stoffer, D.S. (1991) Dynamic linear models with switching. *Journal of the American Statistical Association,* **86**, 763-769.

[59] Smith, A.F.M., and West, M. (1983) Monitoring renal transplants: an application of the multi-process Kalman filter. *Biometrics,* **39**, 867-878.

[60] Sorenson, H.W., and Alspach, D.L. (1971) Recursive Bayesian estimation using Gaussian sums. *Automatica,* **7**, 465-479.

[61] Spall, J.C. (ed.) (1988) *Bayesian Analysis of Time Series and Dynamic Models,* Marcel Dekker, New York.

[62] West, M. (1981) Robust sequential approximate Bayesian estimation. *Journal of the Royal Statistical Society,* (Series B), **43**, 157-166.

[63] West, M. (1986) Bayesian model monitoring. *Journal of the Royal Statistical Society,* (Series B), **48**, 70-78.

[64] West, M. (1992) Modelling time-varying hazards and covariate effects (with discussion). In *Survival Analysis: State of the Art,* J.P. Klein and P.K. Goel, eds. Kluwer.

[65] West, M. (1993) Mixture models, Monte Carlo, Bayesian updating and dynamic models. In *Computing Science and Statistics,* **24**, 325-333.

[66] West, M. (1995) Bayesian inference in cyclical component dynamic linear models. *Journal of the American Statistical Association*, (to appear).

[67] West, M. (1995) Some statistical issues in Palæoclimatology (with discussion). In *Bayesian Statistics 5*, J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith, eds. Oxford University Press.

[68] West, M., Harrison, P.J., and Migon, H.S. (1985) Dynamic generalised linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association*, **80**, 73-97.

[69] West, M., and Harrison, P.J. (1986) Monitoring and adaptation in Bayesian forecasting models. *Journal of the American Statistical Association*, **81**, 741-750.

[70] West, M., and Harrison, P.J. (1989) Subjective intervention in formal models. *Journal of Forecasting*, **8**, 33-53.

[71] West, M. and Harrison, P.J. (1989) *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York. (A full treatment of theory and methods of Bayesian time series analysis and forecasting with dynamic models.)

[72] Zellner, A., and Hong, C. (1989) Forecasting international growth rates using Bayesian shrinkage and other procedures. *Journal of Econometrics*, **40**, 183-202.

[73] Zellner, A., Hong, C., and Min, C. (1991) Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter, and pooling techniques. *Journal of Econometrics*, **49**, 275-304.