Analysis of Censored Data
IMS Lecture Notes – Monograph Series (1995) Volume 27

Markov Mesh Models for Filtering and Forecasting with Leading Indicators

Jingxian Chen   Nicholas J. Lynn   Nozer D. Singpurwalla *

Abstract

In this paper, a new approach for forecasting a time series based on a Markov mesh model that incorporates some key features of a leading indicator series is proposed. The features of the leading indicator series are incorporated via a weighting scheme in which the weights are assessed by a Bayesian approach. The Bayesian approach is implemented by a Gibbs sampling technique. The overall scheme proposed here can be viewed as Kalman Filtering in two dimensions using the raster scan method.

Key Words: *Dynamic Linear Models, Gibbs Sampling, Image Processing, Kalman Filtering, Markov Fields, Mine Detection, Raster Scan, Simulation, Warranties.*

1 Introduction and Motivation

For forecasting a single time series, there are a wide variety of techniques that are currently available. Of these, those that are based on the theory of *dynamic linear models* (DLM) have recently gained much popularity. The DLMs, also known as Kalman Filter models, are discussed in the recent books, [9] and [6]. The inference and extrapolation algorithms of the DLMs can be justified via the Bayesian paradigm (cf. [4]). It is by now well known that forecasts from a single series can be greatly improved if certain key patterns and features from an associated series can be incorporated into the inference mechanism. The associated series is typically a series which precedes the series of interest, and [1] p.402, refers to such a series as a *leading indicator series.* The leading indicator series gives us advanced signals about potential changes in the behaviour of the series of interest, and in so doing, enhances our ability to provide improved predictions of the latter. For example, if the series of interest is the number of housing starts per month, then a leading indicator series could be the monthly change in population. Other such examples can be found in [1] pp.407-412.

*The George Washington University, Washington D.C. 20052

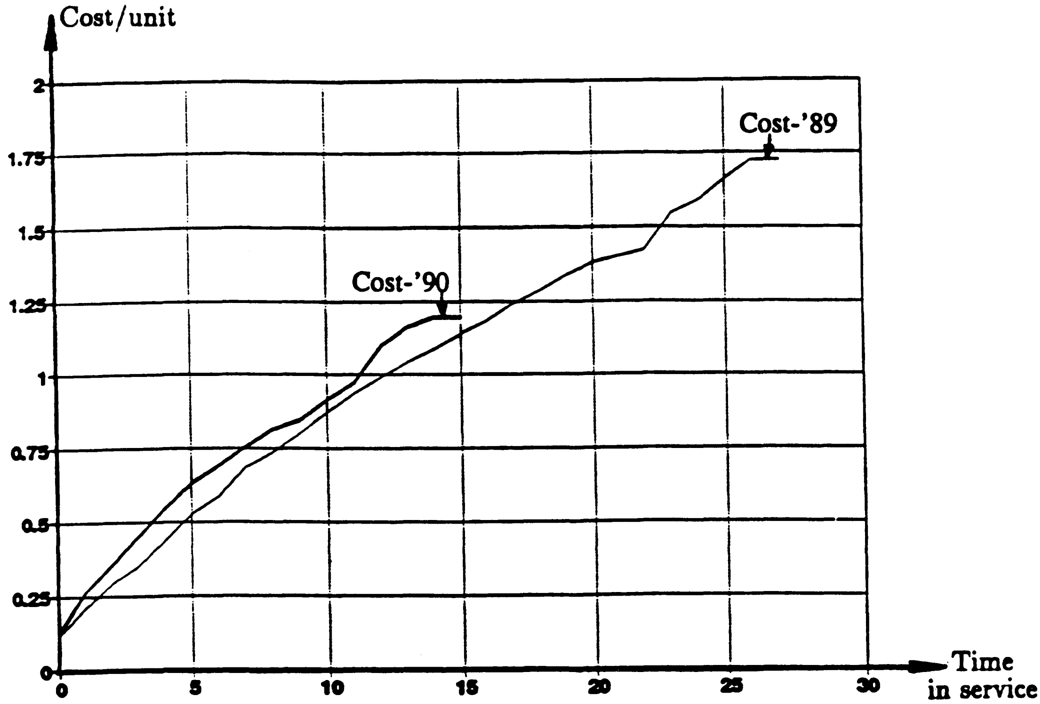

Figure 1: Time Series of Warranty Claims Data

Our interest in the question of forecasting using leading indicators was motivated by the problem of estimating financial reserves to meet (failure) warranty claims on automobiles; see [7]. Warranty claims, in the form of cost/unit, for automobiles of a particular model year, say 1990, arise as a time series indexed by the time in service, months, for that model year. The leading indicator series is the warranty claims data for the previous model year. 1989 in our case. In Figure 1, we show two observed series on warranty claims, the series of interest, labelled "Cost-90," which pertains to the claims for the 1990 model year and the leading indicator series, labelled "Cost-89," which pertains to the claims for the 1989 model year. The series of interest covers the time period of 15 months in service (MIS) whereas the leading indicator series covers data until 27 MIS. Our objective is to forecast the "Cost-90" series for a time horizon of 60 MIS to cover a 5 year warranty on the 1990 model year. The desired forecasts should be based on both, the "Cost-90" series and also the "Cost-89" series.

The traditional approach to forecasting using leading indicators is based on the "transfer function" modelling ideas of [1], p.335. Here, observations from the series of interest are linked with the observations from the leading indicator series by a linear function with unknown coefficients that are estimated from the data. The inference and identification mechanism used by [1] is frequentist, and is predominantly based on the principal of least squares. The point of view of this paper is different. First, we model our series by a "state-space formulation" in which the state parameters of the series of interest are linked with the state parameters of the leading indicator series. Second, the inference and updating mechanism used by us is Bayesian. Our set-up can be viewed as Kalman filtering in two dimensions. This methodology has found applications in areas such as image processing, mine detection, tomography, etc.; see [10] and [11] for an overview.

The organization of our paper is as follows. In Section 2 we describe our model and Section 3 pertains to inference. Inference, though straightforward in principle, gets complicated because of the ensuing non-linearities in the model. The complications are overcome by simulations involving the Gibbs sampling technique (cf. [2]), which de facto plays a key role in addressing problems of the type considered here. Section 4 describes in more detail the warranty forecasting problem which motivated our work. In the interest of emphasizing the model structure, the overall plan of inference, and relevance to the motivating application, details pertaining to the underlying distributions and their related manipulations have been de-emphasized.

2 Dynamic Linear Models With Leading Indicators

The basic DLM consists of an observable quantity Y_t which is related to an unobservable vector of parameters $\underline{\theta}_t$, via a vector of known coefficients $\underline{F}_t$ and an unobservable error u_t through the *observation equation*

$$Y_t = \underline{F}_t'\underline{\theta}_t + u_t, \tag{1}$$

where t denotes an index, usually time. The vector $\underline{\theta}_t$ is known as the *state vector*, and u_t as an "innovation". The defining feature of a DLM is the notion that $\underline{\theta}_t$ evolves with time with an underlying Markovian structure of the form

$$\underline{\theta}_t = \underline{G}_t\underline{\theta}_{t-1} + \underline{v}_t, \tag{2}$$

where $\underline{G}_t$ is a known matrix, and $\underline{v}_t$ an unknown vector of innovation terms which describes deviations from the Markov structure specified above. The relationship (2) is known as the *system equation.* In the case of a Gaussian set-up, $\underline{\theta}_t$ and $\underline{v}_t$ are assumed to be multivariate normal, and u_t a univariate normal. Also, it is a common practice to assume that the sequences $\{u_t\}$ and $\{\underline{v}_t\}$ are mutually and contempraneously independendent, and that $\underline{v}_t$ is independent of $\underline{\theta}_t$. The DLM with a leading indicator series can be structured as follows.

First, in the interest of simplicity, suppose that $\underline{F}_t$ is a unit vector and that the matrix $\underline{G}_t$ is the identity matrix. Then, if $Y_{1,t}$ denotes the leading indicator series, and $Y_{2,t}$ the series of interest, we would postulate that, for $t = 1, 2, \ldots,$

$$Y_{1,t} = \theta_{1,t} + u_{1,t}, \tag{3}$$

and that

$$\theta_{1,t} = \theta_{1,t-1} + v_{1,t}, \tag{4}$$

with $u_{1,t} \sim \mathcal{N}(\prime,\mathcal{U})$ where "$X \sim \mathcal{N}(\mu, \sigma^{\in})$" denotes the fact that X has a Gaussian distribution with location (scale) μ (σ^2). Similarly, we assume that $v_{1,t} \sim \mathcal{N}(\prime, \mathcal{VU})$ with V *specified*, and $U \sim IG(n_0/2, d_0/2)$ where "$X \sim IG(a, b)$" denotes the fact that X has an inverted gamma distribution with shape (scale) parameter a (b). For the series of interest, $Y_{2,t}$, we postulate the relationships

$$Y_{2,t} = \theta_{2,t} + u_{2,t}, \tag{5}$$

and

$$\theta_{2,t} = \gamma\theta_{2,t-1} + (1-\gamma)\theta_{1,t} + v_{2,t}, \tag{6}$$

for $t = 1, 2, \ldots$, where $u_{2,t} \sim \mathcal{N}(\prime,\mathcal{U})$, $v_{2,t} \sim \mathcal{N}(\prime, \mathcal{VU})$ and the weight coefficient $\gamma \sim Beta(\delta_1, \delta_2)$, where "$X \sim Beta(a, b)$" denotes the fact that X has a beta distribution with parameters a and b.

Contrast the relationships (4) and (6). The former has a Markov chain structure, which in (6) has been extended to two dimensions making the set-up what is known as a "Markov random field." When forecasting with multiple time series, say ℓ, the Markov field is formed by the two dimensions of time, t, and series ℓ, $\ell = 1, 2, \ldots,$. This field forms a four neighbor system, where $\theta_{\ell,t}$, the system level of series ℓ at time t, can be expressed in terms of the corresponding levels of series ℓ at times $(t-1)$ and $(t+1)$, plus the corresponding levels of the two neighboring series $\theta_{\ell-1,t}$ and $\theta_{\ell+1,t}$ at time t. That is, we may write $\theta_{\ell,t} = \gamma_1\theta_{\ell,t-1} + \gamma_2\theta_{\ell,t+1} + \gamma_3\theta_{\ell-1,t} + \gamma_4\theta_{\ell+1,t}$ with the weights $\gamma_i \geq 0$ and $\Sigma_1^4\gamma_i = 1$. Equation (6) is a special case of the above, appropriately modified to accomodate the nuances of a single leading indicator series. Specifically, γ_2 and γ_4 are set equal to zero, and γ_1 is replaced by γ with $\gamma_3 = 1 - \gamma$.

2.1 Markov Random Fields

The foundations of Markov Random Fields (MRF) lie in the physics literature of ferromagnetism; see [3]. We overview here some basic notions and ideas of MRF's.

Suppose that $\mathcal{L}$ is a lattice of points. Then a neighborhood system $\mathcal{N}$ on $\mathcal{L}$ is defined as

$$\mathcal{N} = \{\eta_{\dashv} \subset \mathcal{L} : \dashv \in \mathcal{L}\}, \tag{7}$$

such that for any $a \in \mathcal{L}$, $a \notin \eta_a$ and $b \in \eta_a \Leftrightarrow a \in \eta_b$. The set η_a thus consists of the neighbors of a. It is clear that $\mathcal{N}$ can be varied, describing elaborate spatial dependence. In two dimensions, relatively simple neighborhood systems are usually adequate, the two simplest being the four neighbor system $\mathcal{N}^{(\infty)}$ and the eight neighbor system $\mathcal{N}^{(\in)}$, or Markov mesh. The four neighbor system describes the case when a point in the field is entirely determined by its neighbors: above, below, left and right. The model $\theta_{\ell,t} = \gamma_1\theta_{\ell,t-1} + \gamma_2\theta_{\ell-1,t} + \gamma_3\theta_{\ell-1,t} + \gamma_4\theta_{\ell+1,t}$ mentioned before is based on the four neighbor system. The eight neighbor system is a four neighbor system to which the four diagonal points are added. Adjustments must be made at the boundaries of the field. The choice of a neighborhood system is motivated by the application. With image re-construction one typically uses the four or the eight neighbor system. In (6) we have used a two neighbor system. In applications of tomography and mine detection one considers lattices in three dimensions and then chooses an appropriate neighborhood system. Our motivation for considering a MRF to describe the evaluation of $\theta_{2,t}$ — see (6) — is the belief that the series of interest and the leading indicator series have a common trend, for otherwise we could treat the two series seperately.

2.2 The Raster Scan Method

Since our series of interest $\theta_{2,t}$ and the leading indicator series $\theta_{1,t}$ are chronological, we choose to transverse the MRF using the "raster scan method." This essentially means that we first treat the series which occurs first, namely the leading indicator series, and then the series of interest $\theta_{2,t}$. This method of traversal prompts the following set-up of the two-dimensional dynamic linear model:

$$\begin{aligned} Y_{\ell,t} &= \theta_{\ell,t} + u_{\ell,t}, & \ell = 1,2,\ldots \\ \theta_{\ell,t} &= \gamma_1\theta_{\ell,t-1} + \gamma_2\theta_{\ell,t+1} + \gamma_3\theta_{\ell-1,t} + \gamma_4\theta_{\ell+1,t} + v_{\ell,t}, & \ell = 2,3,\ldots \\ \theta_{1,t} &= \theta_{1,t-1} + v_{1,t}, & \end{aligned} \tag{8}$$

for $t = 1,2,\ldots,$.

For series one, that is, for $\ell = 1$, the above equations are the standard ones given by (1) and (2) — with appropriate simplifications. The equations for the other series are extended to reference the previous series. For our application $\ell = 2$, $\gamma_2 = \gamma_4 = 0$, and $\gamma_3 = 1 - \gamma_1$ with $\gamma_1 = \gamma$. Thus the model given above simplifies as:

$$\begin{aligned} Y_{\ell,t} &= \theta_{\ell,t} + u_{\ell,t}, & \ell = 1,2, \\ \theta_{2,t} &= \gamma\theta_{2,t-1} + (1-\gamma)\theta_{1,t} + v_{2,t}, & \\ \theta_{1,t} &= \theta_{1,t-1} + v_{1,t}, & \end{aligned} \tag{9}$$

for $t = 1, 2, \ldots$, with γ interpreted as a weight which captures the amount of dependence of $\theta_{2,t}$ on its previous value $\theta_{2,t-1}$ and upon its corresponding value in the leading indicator series, namely $\theta_{1,t}$.

To complete the specification of the model, we need to describe our prior beliefs surrounding the errors, the initial value of the system variable $\theta_{1,0}$, and the weighting parameter γ. Prior beliefs about $u_{1,t}$, $u_{2,t}$, $v_{1,t}$ and $v_{2,t}$ were specified before, via the Gaussian/inverted gamma model. The parameter $\theta_{1,0}$ is assumed to be such that $\theta_{1,0} \sim \mathcal{N}(\dashv, \lfloor\mathcal{U})$ with a and b specified. Finally, as was stated previously, $\gamma \sim Beta(\delta_1, \delta_2)$ with δ_1 and δ_2 specified.

We remark that there are three mechanisms via which information from the leading indicator series is incorporated into the series of interest. The first is via the weighting coefficient γ; the second is via the scaling constant U, which is unknown but is assumed to be such that $U \sim IG(n_0/2, d_0/2)$ with n_0 and d_0 specified. The third is via $\theta_{2,0}$, where $\theta_{2,0}$ has the same distribution as the posterior distribution of $\theta_{1,0}$ given the data from the leading indicator series. To conclude, the set-up (9) proposed here requires a specification of the constants V, a, b, δ_1, δ_2, n_0 and d_0. It is said to be non-linear because in the second equation of (9), both γ and $\theta_{2,t-1}$ are unknown. Were γ to be known (specified) then standard results from the DLM theory — see, for example, [4] — can be adapted for inference and predictions. With γ unknown, the closed form results of the DLM are not possible to obtain, and inference and predictions are to be undertaken via a simulation based technique such as the Gibbs sampler. Details are given in Section 3 below.

3 Inference and Predictions

Suppose that the leading indicator series consists of t_1 observations, $\underline{Y}_1 = (Y_{1,1}, \ldots, Y_{1,t_1})'$, and the series of interest consists of t_2 observations $\underline{Y}_2 = (Y_{2,1}, \ldots, Y_{2,t_2})'$. The following notation helps to simplify the ensuing text: $\underline{U}_1 = (u_{1,1}, \ldots, u_{1,t_1})'$, $\underline{V}_1 = (\theta_{1,0}, v_{1,1}, \ldots, v_{1,n})'$, $\underline{\Theta}_i = (\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,n})'$, $i = 1, 2$, where n $(\geq t_1)$ is the *forecast horizon*, and $\underline{a} = (a, 0, \ldots, 0)'$ is an $(n+1) \times 1$ vector.

We also need to define the following matrices:

$$\underline{A}_1 = \begin{pmatrix} 1 & & & 0 \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & 1 & \cdots & 1 \end{pmatrix}_{((n+1)\times(n+1))}$$

$$\underline{D}_1 = \begin{pmatrix} b & \underline{0} \\ \underline{0} & V\underline{I}_n \end{pmatrix}_{((n+1)\times(n+1))} \qquad \underline{B}_1 = (\underline{0}, \underline{I}_{t_1}, \underline{0}, \cdots, \underline{0})_{(t_1\times(n+1))}$$

where $\underline{I}_k$ is an $(k \times k)$ identity matrix and $\underline{0}$ is a $(t_1 \times 1)$ vector whose entries are identically zero.

Then, in matrix notation, the $Y_{1,t}$'s and $\theta_{1,t}$'s of (9), for $t = 1, 2, \ldots, t_1$, and our prior beliefs about the errors and the state vector Θ_1 can be written as:

$$\begin{aligned} \underline{\Theta}_1 &= \underline{A}_1 \underline{V}_1, \\ \underline{Y}_1 &= \underline{B}_1 \underline{\Theta}_1 + \underline{U}_1, \end{aligned}$$

with $(\underline{V}_1|U) \sim \mathcal{N}(\underline{\dashv}, \mathcal{U}\underline{\mathcal{D}}_\infty)$, $(\underline{U}_1|U) \sim \mathcal{N}(\underline{\imath}, \mathcal{U}\underline{\mathcal{I}}_{\sqcup_\infty})$, and $(\underline{\Theta}_1|U) \sim \mathcal{N}(\underline{\ell}, \mathcal{U}\underline{\mathcal{C}})$, where $\underline{\ell} = \underline{A}_1\underline{a}$ and $\underline{C} = \underline{A}_1\underline{D}_1\underline{A}_1'$.

Using standard prior-to-posterior manipulations [cf. [5]] it can be shown that

$$(\underline{\Theta}_1|U, \underline{Y}_1) \sim \mathcal{N}(\underline{\Updownarrow}_\infty, \mathcal{U}\underline{\mathcal{D}}), \tag{10}$$

where

$$\begin{aligned} \underline{m}_1 &= \underline{\ell} + \underline{C}\underline{B}_1'(\underline{B}_1\underline{C}\underline{B}_1' + \underline{I}_{t_1})^{-1}(\underline{Y}_1 - \underline{B}_1\underline{\ell}), \\ \underline{D} &= \underline{C} - \underline{C}\underline{B}_1'(\underline{B}_1\underline{C}\underline{B}_1' + \underline{I}_{t_1})^{-1}\underline{B}_1\underline{C}. \end{aligned}$$

Furthermore, the posterior distribution of U is given as

$$(U|\underline{Y}_1) \sim IG(\frac{n_1}{2}, \frac{d_1}{2}), \tag{11}$$

where

$$\begin{aligned} n_1 &= n_0 + t_1, \\ d_1 &= d_0 + (\underline{Y}_1 - \underline{B}_1\underline{\ell})'(\underline{B}_1\underline{C}\underline{B}_1' + \underline{I}_{t_1})^{-1}(\underline{Y}_1 - \underline{B}_1\underline{\ell}). \end{aligned}$$

Hence, unconditional on U,

$$\begin{aligned} (\underline{\Theta}_1|\underline{Y}_1) &\sim \int (\underline{\Theta}_1|U, \underline{Y}_1)(U|\underline{Y}_1) \\ &\sim T_{n_1}(\underline{m}_1, \tfrac{d_1}{n_1}\underline{D}), \end{aligned} \tag{12}$$

The notation "$\underline{X} \sim T_n(\underline{a}, \underline{b})$" denotes the fact that the vector $\underline{X}$ has a multivariate Student's-t distribution with location vector $\underline{a}$, scale matrix $\underline{b}$ and degrees of freedom n.

As mentioned before, the above inferential results are very standard; the reader may refer to [9] for more details on the derivation of (12) and (11). Equation (12) can be used to forecast $Y_{1,t_1+1}, Y_{1,t_1+2}, \ldots$, the unobserved values of the series of interest.

If the vector $\underline{\Theta}_1$ is partitioned as $[\underline{\Theta}_{11}, \underline{\Theta}_{12}]$, where $\underline{\Theta}_{11} = (\theta_{1,0}, \ldots, \theta_{1,t_1})'$ and $\underline{\Theta}_{12} = (\theta_{1,t_1+1}, \ldots, \theta_{1,n})'$, and the corresponding partitions of $\underline{m}_1$ and $\underline{D}$ are as follows:

$$\underline{m}_1 = \begin{pmatrix} \underline{m}_{11} \\ \underline{m}_{12} \end{pmatrix}, \underline{D} = \begin{pmatrix} \underline{D}_{11} & \underline{D}_{12} \\ \underline{D}_{21} & \underline{D}_{22} \end{pmatrix},$$

then the predictive distribution of the vector $\underline{Y}_{1,n} = (Y_{1,t_1+1}, \ldots, Y_{1,n})$ is of the form

$$(\underline{Y}_{1,n}|\underline{Y}_1) \sim T_{n_1}\left(\underline{m}_{12}, \frac{d_1}{n_1}(\underline{D}_{22} + \underline{I}_{(n-t_1)})\right). \tag{13}$$

The discussion thus far has centered around inference and prediction for the leading indicator series. For inference and predictions for the series of interest, we need to introduce, in addition to the vector $\underline{Y}_2 = (Y_{2,1}, \ldots, Y_{2,t_2})'$, the following additional vectors and matrices: $\underline{V}_2 = (\underline{\Theta}_1', v_{2,1}, \ldots, v_{2,n})'$, $\underline{U}_2 = (u_{2,1}, \ldots, u_{2,t_2})'$, $\underline{B}_2 = [\underline{0}, \underline{I}_{t_2}, \underline{0}, \cdots, \underline{0}]_{(t_2\times(n+1))}$, $\underline{A}_2(\gamma) = [\underline{A}_{21}(\gamma)|\underline{A}_{22}(\gamma)]$, where

$$\underline{A}_{21}(\gamma) = \begin{pmatrix} 1 & & & & \\ \gamma & (1-\gamma) & & & \\ \gamma^2 & \gamma(1-\gamma) & (1-\gamma) & & \\ \vdots & \vdots & \vdots & \ddots & \\ \gamma^n & \gamma^{n-1}(1-\gamma) & \gamma^{n-2}(1-\gamma) & \cdots & (1-\gamma) \end{pmatrix}_{(n+1)\times(n+1)}$$

$$\underline{A}_{22}(\gamma) = \begin{pmatrix} 0 & & & & \\ 1 & 0 & & & \\ \gamma & 1 & 0 & & \\ \vdots & \vdots & \ddots & \ddots & \\ \gamma^{n-2} & \gamma^{n-3} & \cdots & 1 & 0 \\ \gamma^{n-1} & \gamma^{n-2} & \cdots & \gamma & 1 \end{pmatrix}_{(n+1)\times n}$$

Then, in matrix notation, the $Y_{2,t}$'s and the $\theta_{2,t}$'s of (9), for $t = 1, 2, \ldots, t_2$, can be written as

$$\begin{aligned} \underline{\Theta}_2 &= \underline{A}_2(\gamma)\underline{V}_2 = \underline{A}_{21}(\gamma)\begin{pmatrix} \theta_{2,0} \\ \theta_{1,1} \\ \theta_{1,2} \\ \vdots \\ \theta_{1,n} \end{pmatrix} + \underline{A}_{22}(\gamma)\begin{pmatrix} v_{2,1} \\ v_{2,2} \\ v_{2,3} \\ \vdots \\ v_{2,n} \end{pmatrix} \\ \underline{Y}_2 &= \underline{B}_2\underline{\Theta}_2 + \underline{U}_2 \end{aligned}$$

with prior beliefs about the errors given by

$$
\begin{aligned}
(\underline{U}_2|U,\underline{Y}_1,\gamma) &= (\underline{U}_2|U) \sim \mathcal{N}(\iota, \mathcal{U}\underline{I}_{\sqcup_\in}) \\
(\underline{V}_2|U,\underline{Y}_1,\gamma) &= (\underline{V}_2|U,\underline{Y}_i) \sim \mathcal{N}\left(\left(\begin{array}{c} \underline{m}_1 \\ 0 \end{array}\right), \mathcal{U}\left(\begin{array}{cc} \underline{D} & 0 \\ 0 & V\underline{I}_n \end{array}\right)\right)
\end{aligned}
$$

Our prior opinion on $\underline{\Theta}_2$ should be based on $\underline{Y}_1$ since the leading indicator series gives us information on the state vector of the series of interest. By the assumption that $\theta_{2,0} = (\theta_{1,0}|\underline{Y}_1)$,

$$
\begin{aligned}
(\underline{\Theta}_2|U,\underline{Y}_1,\gamma) &= (\underline{A}_{21}(\gamma)\underline{\Theta}_1 + \underline{A}_{22}(\gamma)\left(\begin{array}{c} v_{21} \\ \vdots \\ v_{22} \end{array}\right)|U,\underline{Y}_1,\gamma) \\
\Rightarrow (\underline{\Theta}_2|U,\underline{Y}_1,\gamma) &\sim \mathcal{N}(\underline{\nu}(\gamma), \mathcal{U}\pm(\gamma)),
\end{aligned}
$$

where

$$
\begin{aligned}
\underline{\nu}(\gamma) &= \underline{A}_{21}(\gamma)\underline{m}_1, \\
\underline{\Sigma}(\gamma) &= \underline{A}_{21}(\gamma)\underline{D}\underline{A}'_{21}(\gamma) + V\underline{A}_{22}(\gamma)\underline{A}'_{22}(\gamma).
\end{aligned}
$$

Once again, standard prior-to-posterior manipulations show that the posterior distribution for $\underline{\Theta}_2$ given $\underline{Y}_2$ (and also $\underline{Y}_1$), were γ to be known, is of the form

$$
(\underline{\Theta}_2|U,\gamma,\underline{Y}_1,\underline{Y}_2) \sim \mathcal{N}(\underline{\Updownarrow}_{\in}(\gamma), \mathcal{U}\underline{\mathcal{E}}(\gamma)), \tag{14}
$$

where

$$
\begin{aligned}
\underline{m}_2(\gamma) &= \underline{\nu}(\gamma) + \underline{\Sigma}(\gamma)\underline{B}'_2(\underline{B}_2\underline{\Sigma}(\gamma)\underline{B}'_2 + \underline{I}_{t_2})^{-1}(\underline{Y}_2 - \underline{B}_2\underline{\nu}(\gamma)), \\
\underline{E}(\gamma) &= \underline{\Sigma}(\gamma) - \underline{\Sigma}(\gamma)\underline{B}'_2(\underline{B}_2\underline{\Sigma}(\gamma)\underline{B}'_2 + \underline{I}_{t_2})^{-1}\underline{B}_2\underline{\Sigma}(\gamma).
\end{aligned}
$$

Averaging over the posterior distribution of U, where

$$
(U|\gamma,\underline{Y}_1,\underline{Y}_2) \sim IG(\frac{n_2}{2}, \frac{d_2(\gamma)}{2}), \tag{15}
$$

with

$$
\begin{aligned}
n_2 &= n_1 + t_2, \\
d_2(\gamma) &= d_1 + (\underline{Y}_2 - \underline{B}_2\underline{\nu}(\gamma))'(\underline{B}_2\underline{\Sigma}(\gamma)\underline{B}'_2 + I_{t_2})^{-1}(\underline{Y}_2 - \underline{B}_2\underline{\nu}(\gamma)),
\end{aligned}
$$

we have

$$
(\underline{\Theta}_2|\gamma,\underline{Y}_1,\underline{Y}_2) \sim T_{n_2}(\underline{m}_2(\gamma), \frac{d_2(\gamma)}{n_2}\underline{E}(\gamma)). \tag{16}
$$

3.1 Inference and Prediction Based on the Gibbs Sampler

Equation (16) provides inference about the $(n+1)$ dimensional state vector $\underline{\Theta}_2$ when γ is known. When γ is unknown, as in the case assumed here, the Kalman filter model becomes non-linear and closed form inference is not possible. However, since we have all the necessary full conditional distributions for implementing the Gibbs sampler (see [2]), we will use this technique for inference about $\underline{\Theta}_2$, U and γ, conditional on $\underline{Y}_1$ and $\underline{Y}_2$ alone. The full conditionals needed to generate a sample from the joint distribution of $(\underline{\Theta}_2, U, \gamma | \underline{Y}_1, \underline{Y}_2)$ are:

$$\begin{aligned} &a) \quad (\underline{\Theta}_2 | U, \gamma, \underline{Y}_1, \underline{Y}_2) \\ &b) \quad (U | \gamma, \underline{\Theta}_2, \underline{Y}_1, \underline{Y}_2) = (U | \gamma, \underline{Y}_1, \underline{Y}_2) \\ &c) \quad (\gamma | U, \underline{\Theta}_2, \underline{Y}_1, \underline{Y}_2) \end{aligned} \tag{17}$$

The equality in part b) follows since U is judged independent of $\underline{\Theta}_2$. Equation (14) provides us with the distribution of the quantity labelled a) above and Equation (15) provides the distribution of the quantity labelled b). To assess the distribution of the quantity labelled c) we invoke Bayes' Law by which

$$(\gamma | U, \underline{\Theta}_2, \underline{Y}_1, \underline{Y}_2) \propto \left(\begin{array}{c} \underline{\Theta}_2 \\ \underline{Y}_2 \end{array} | U, \gamma, \underline{Y}_1 \right) (\gamma | U, \underline{Y}_1) \tag{18}$$

The second term on the right hand side of (18) is simply the prior distribution of γ. The first term is obtained via standard techniques [cf. [5]] as

$$\left(\begin{array}{c} \underline{\Theta}_2 \\ \underline{Y}_2 \end{array} | U, \gamma, \underline{Y}_1 \right) = \left(\begin{array}{cc} \underline{A}_2(\gamma) & 0 \\ \underline{B}_2 \underline{A}_2(\gamma) & \underline{I}_{t_2} \end{array} \right) \left(\begin{array}{c} \underline{V}_2 \\ \underline{U}_2 \end{array} | U, \gamma, \underline{Y}_1 \right)$$

$$\sim \mathcal{N}\left(\left(\begin{array}{c} \underline{\nu}(\gamma) \\ \underline{B}_2 \underline{\nu}(\gamma) \end{array} \right), \mathcal{U} \left(\begin{array}{cc} \underline{\Sigma}(\gamma) & \underline{\Sigma}(\gamma) \underline{B}_2' \\ \underline{B}_2 \underline{\Sigma}(\gamma) & \underline{B}_2 \underline{\Sigma}(\gamma) \underline{B}_2' + \underline{I}_{t_2} \end{array} \right) \right) \tag{19}$$

Thus Equations (14), (15), (19) and the prior beta distribution of γ can be used to generate samples from the joint distribution of $(\underline{\Theta}_2, U, \gamma | \underline{Y}_1, \underline{Y}_2)$. Suppose that M such samples are generated, where M is large, say 100. Let the i-th generated sample be denoted by the vector

$$(\theta_{2,0}^{(i)}, \theta_{2,1}^{(i)}, \ldots, \theta_{2,t_2}^{(i)}, \theta_{2,t_2+1}^{(i)}, \ldots, \theta_{2,n}^{(i)}, U^{(i)}, \gamma^{(i)})', \tag{20}$$

for $i = 1, \ldots, M$. Since $(u_{2,j} | U) \sim \mathcal{N}(\prime, \mathcal{U})$, for $j = 1, 2, \ldots$, we shall use each $U^{(i)}$ to generate the n innovation terms $u_{2,1}^{(i)}, u_{2,2}^{(i)}, \ldots, u_{2,n}^{(i)}$, $i = 1, \ldots, M$, via the distributional relationship $(u_{2,j} | U^{(i)}) \sim \mathcal{N}(\prime, \mathcal{U}^{()})$, $j = 1, 2, \ldots, n$.

Since $Y_{2,j} = \theta_{2,j} + u_{2,j}$, $j = 1, 2, \ldots, n$, we can approximate the mean of the predictive distribution of Y_{2,t_2+k}, $k = 1, 2, \ldots, (n - t_2)$ via the Gibbs sample estimate

$$E(Y_{2,t_2+k}|\underline{Y}_1, \underline{Y}_2) \approx \frac{1}{M}\sum_{i=1}^{M}(\theta^{(i)}_{2,t_2+k} + u^{(i)}_{2,t_2+k}). \tag{21}$$

The predictive distribution of Y_{2,t_2+k}, $k = 1, 2, \ldots, (n - t_2)$, is estimated via the histogram of $(\theta^{(i)}_{2,t_2+k} + u^{(i)}_{2,t_2+k})$, $i = 1, \ldots, M$. Similarly, the histogram of $\theta^{(i)}_{2,j}$, $i = 1, \ldots, M$ estimates the posterior distribution of $\theta_{2,j}$, $j = 0, 1, \ldots, n$ and the mean of this distribution is estimated by

$$\frac{1}{M}\sum_{i=1}^{M}\theta^{(i)}_{2,j}. \tag{22}$$

Finally, the posterior distribution of the weighting constant γ is estimated by the histogram of $\gamma^{(i)}$, $i = 1, \ldots, M$, and an estimate of its mean is simply the average values of $\gamma^{(i)}$. This estimate of the mean gives us a clue as to the extent of the impact of the leading indicator series on the series of interest.

3.2 Extensions to Censored Data

In the discussion so far, we have limited our analysis to the case of complete data. In practice, data is often censored, so the inference of the previous section must be extended to deal with incomplete data. We consider two types of censoring. The first type is inherent in the warranty claims data, wherein the data is cost/unit and refers to the average cumulative claims per unit. The data for t months in service (MIS), Y_t, can be written as $Y_t = \frac{1}{r_t}(X_{1t} + X_{2t} + \cdots + X_{r_t t})$ where r_t is the number of units for which we have data at time t, and X_{it} is the claims data for unit i at time t. Censoring comes about because units are manufactured at different times with the consequence that $r = r_0 \geq r_1 \geq r_2 \geq \ldots$. The number of censored units at time t is $c_t = r - r_t$.

We propose two methods of handling this censoring. The first method is based on the observation that claims data for t-MIS is based on fewer units than claims data for $(t-1)$-MIS. Since the data is decreasingly robust, a simple solution is to adjust the observational variance accordingly. Specifically, we would define the observational variance at time t to be normally distributed with variance rU/r_t.

An alternative solution employs data augmentation (cf. [8]), wherein unobserved data is introduced into the Gibbs sampler as further unknowns. The resulting structure of the full-conditional distributions is typically very simple, and leads to a straightforward implementation of the Gibbs sampler.

For the problem discussed, the unobserved data is the future repair data for the units. Hence, we introduce the variables $Y_1', Y_2', \ldots$, into the Gibbs sampler, where these represent the observations that we would have observed in the absence of censoring. The Gibbs sampling mechanism is easily adjusted to deal with these extra variables.

The second form of censoring considered is the censoring of observations due to clerical errors or irregular data collection. This is otherwise known as *missing data.* Again, the solution to this problem is to employ data augmentation, wherein the missing observations of $\underline{Y}_\ell$ are introduced into the Gibbs sampler as further unknowns. As in the case above, the simple structure of the full-conditional distributions leads to a straightforward implementation of the Gibbs sampler.

4 Forecasting Warranty Claims

The model of Equation (4) is known as the "steady model". It is suitable for forecasting a time series that is generally level; that is, the observations tend to fluctuate around a constant value. However, an examination of Figure 1 shows us that both, the leading indicator series and the series of interest, have an upward trend which tend to be generally linear. In order to describe such series (4) needs to be extended as follows:

$$\begin{array}{rcllcl} Y_{1,t} & = & \theta_{1,t} + u_{1,t}, & \text{with} & u_{1,t} & \sim \mathcal{N}(0,\mathcal{U}) \\ \theta_{1,t} & = & \theta_{1,t-1} + \beta_{1,t-1} + v_{1,t}, & \text{with} & v_{1,t} & \sim \mathcal{N}(0,\mathcal{VU}) \\ \beta_{1,t} & = & \beta_{1,t-1} + w_{1,t}, & \text{with} & w_{1,t} & \sim \mathcal{N}(0,\mathcal{WU}) \end{array} \tag{23}$$

where V and W are specified constants, and as before, $U \sim IG(\frac{n_0}{2}, \frac{d_0}{2})$. The parameter $\theta_{1,t}$ denotes the level of the series, and the parameter $\beta_{1,t}$, the trend. Prior information about $\theta_{1,0}$ and $\beta_{1,0}$, the initial values of the parameters is described via the distribution

$$\left(\begin{array}{c} \theta_{1,0} \\ \beta_{1,0} \end{array} |U \right) \sim \mathcal{N}\left(\left(\begin{array}{c} a_{1,0} \\ b_{1,0} \end{array} \right), \mathcal{U} \left(\begin{array}{cc} \sigma_0^2 & \rho\sigma_0\tau_0 \\ \rho\sigma_0\tau_0 & \tau_0^2 \end{array} \right) \right), \tag{24}$$

with σ_0, τ_0 and ρ specified.

For the series of interest $Y_{2,t}$, we postulate a relationship analogous to (6) except that the weighting constant γ is applied to the parameter $\beta_{1,t}$ instead of the parameter $\theta_{1,t}$. There are two reasons behind this choice. The first is that forecasts obtained via weighting $\beta_{1,t}$ turn out to be superior to those obtained via weighting $\theta_{1,t}$. The second reason is more structural. It is more meaningful to suppose that the underlying trend in the series of interest would mimick the underlying trend of the leading indicator series. Accordingly, we have

$$
\begin{aligned}
Y_{2,t} &= \theta_{2,t} + u_{2,t}, && \text{with} \quad u_{2,t} \sim \mathcal{N}(0,\mathcal{U}) \\
\theta_{2,t} &= \theta_{2,t-1} + \beta_{2,t-1} + v_{2,t}, && \text{with} \quad v_{2,t} \sim \mathcal{N}(0,\mathcal{VU}) \qquad (25) \\
\beta_{2,t} &= \gamma\beta_{2,t-1} + (1-\gamma)\beta_{1,t} + w_{2,t}, && \text{with} \quad w_{2,t} \sim \mathcal{N}(0,\mathcal{WU})
\end{aligned}
$$

As before, given U, the innovation terms $u_{1,t}$, $v_{1,t}$, $w_{1,t}$, $u_{2,t}$, $v_{2,t}$ and $w_{2,t}$ are serially and contemporaneously independent. The prior on γ is again a beta distribution with parameters δ_1 and δ_2. The prior on $\theta_{2,0}$ and $\beta_{2,0}$ is determined by the relationship

$$
\begin{pmatrix} \theta_{2,0} \\ \beta_{2,0} \end{pmatrix} = \begin{pmatrix} \theta_{1,0} \\ \beta_{1,0} \end{pmatrix} + \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \qquad (26)
$$

where the constants s_1 and s_2 are specified by the user, and in our application depend on the nature of the difference between the automobiles manufactured in the 1989 and the 1990 model years.

Inference and forecasts based on the model Equations (23) and (25) proceed along the lines indicated in Section 3 with appropriate modifications to include the trend term $\beta_{1,t}$ and its weighting by γ. Application to the data of Figure 1 are summarized in the section below.

4.1 Application to Data and Summary of Results

For the cost/unit data on warranty claims, we apply the models (23) and (25) with $n_0 = 5$, $d_0 = 2$, $\delta_1 = \delta_2 = 1$, $\rho = 0$, $\sigma_0^2 = 0.01$, $\tau_0^2 = 0.1$, $V = 0.01$ and $W = 0.2$. This choice of values is arbitrary. The constants $a_{1,0}$ and $b_{1,0}$ — the initial values of $\theta_{1,0}$ and $\beta_{1,0}$ — are chosen from data on "pre-warranty claims". that is, claims made by the automobile dealers of the manufacturer, prior to the delivery of the vehicles. Because of reasons of confidentiality, these numbers are not revealed here. However, since the leading indicator series $Y_{1,t}$ consists of 27 observations the effect of the initial values on the rest of the analysis is negligible.

In Table 1, we summarize the results of our analysis of the series of interest, by way of forecasts of $Y_{2,t}$ based on the 27 observations of the leading indicator series (i.e. $t_1 = 27$) and 5 observations of the series of interest (i.e. $t_2 = 5$) and also 10 observations of the series of interest. With each case, the forecast horizon is 5. In order to show the effectiveness of incorporating information from the leading indicator series on forecasts of the series of interest, we also show — see columns 4 and 6 of Table 1 — forecasts that are based on previous values of the $Y_{2,t}$ series alone. That is,

Index t	Observed Values $Y_{2,t}$	Predicted Values of Y_{2t}			
		Based on First 5		Based on First 10	
		27 Obs of $Y_{1,t}$	No Obs of $Y_{1,t}$	27 Obs of $Y_{1,t}$	No obs of $Y_{1,t}$
6	0.693	0.688	0.729		
7	0.752	0.760	0.819		
8	0.814	0.829	0.909		
9	0.845	0.895	0.998		
10	0.915	0.959	1.088		
11	0.974			0.987	0.968
12	1.098			1.046	1.021
13	1.162			1.101	1.074
14	1.196			1.152	1.127
15	1.196			1.203	1.181
Cumulative Squared Error		0.005	0.068	0.009	0.019
Moments of the Posterior Distribution of γ (when its prior distribution is uniform)					
Mean		0.419		0.315	
Variance		0.069		0.043	

Table 1: Comparison of Actual and Predicted Values

when the model (23) is applied to the series $Y_{2,t}$ and the forecast horizon is 5.

An inspection of the cumulative squared errors of the forecasts shows that incorporating the 27 observations from the leading indicator series results in a substantial reduction of forecast error from that which is obtained via an extrapolation of $Y_{2,t}$ alone. The mean of the posterior distribution of γ averages to be about 0.37 suggesting that the leading indicator series is assigned a weight of about 0.63. Recall that the mean of the prior of γ was 0.5.

Acknowledgements

Research supported in part by the Army Research Office Grant DAAH04-93-G-0020, the Air Force Office of Scientific Research Grant AFOSR-F49620-95-1-0107 and the U.S. National Science Foundation Grant SES-9122494.

References

[1] Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control* (Revised ed.) San Fransisco: Holden-Day.

[2] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398-409.

[3] Kindermann, R. and Snell, J.L. (1980). *Markov Random Fields and Their Applications.* Providence, RI: American Mathematical Society.

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

[5] Meinhold, R.J. and Singpurwalla, N.D. (1987). A Kalman Filter smoothing approach for extrapolations in certain dose-response, damage-assessment, and accelerated life-testing studies. *The American Statistician* **41**: 101-106.

[6] Pole, A., West, M., and Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis.* London: Chapman and Hall.

[7] Singpurwalla, N.D. and Wilson, S.P. (1993). The warranty problem: Its statistical and game theoretic aspects. *SIAM Review* **35**: 17-42.

[8] Smith, A.F.M. and Roberts, G.O. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *J. R. Statist. Soc. B* **55**: 3-23.

[9] West, M. and Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models.* New York: Springer-Verlag.

[10] Woods, J.W. (1972). Two-dimensional discrete Markovian fields. *IEEE Transactions on Information Theory* **V. IT-18**: 232-240.

[11] Woods, J.W. and Radewan, C.H. (1977). Kalman filtering in two dimensions. *IEEE Transactions on Information Theory* **V. IT-23**: 473-482.