Dynamic Generalized Linear Models and Bayesian Forecasting: Comment
Author(s): I. Guttman and D. Pena
Source: *Journal of the American Statistical Association*, Vol. 80, No. 389 (Mar., 1985), pp. 91-92
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2288046
Accessed: 24-11-2015 10:14 UTC

Comment

Robust Filtering

I. GUTTMAN and D. PENA*

We congratulate West, Harrison, and Migon (WHM) on a stimulating and interesting article. This article undoubtedly will be the focus for much discussion on the many issues raised by WHM in the area of Bayesian forecasting in general and Kalman filtering in particular. Because of limitations of time and space, we will comment only on the aspect of filtering in the presence of outliers. We agree with WHM that this is a very serious and crucial point, since it is well known that extreme observations can make the standard Kalman-filter procedure highly unstable.

Therefore, it is important to provide, within the filtering algorithm, an automatic method for "downweighting" extreme observations, if and when they occur. Hence, we feel it is important to emphasize that such (robust) filters be derived using formal statistical models that allow for the appearance of spuriously generated observations so that resulting filtering procedures may be compared.

To achieve this, we believe that the simplest way to proceed is to assume a mixture of two distributions for the noise in the observation equation. One of these will represent the generation of observations as intended, whereas the other would represent a departure from the former. This approach has been useful in Bayesian estimation, and the idea was first used by Jeffreys (1961) and Box and Tiao (1968).

Using this approach, the posterior of the state $\boldsymbol{\theta}_t$ turns out to be a mixture of two distributions. To arrive at the next step in a Kalman-filter algorithm, we believe that his latter mixture should be approximated by a single distribution, *uniquely* determined by the first and second moments of the posterior. Here we differ with WHM, who would favor the procedure of Harrison and Stevens (1976a), for our procedure avoids the arbitrariness in the collapsing of N^2 posterior distributions to N distributions when $N > 2$ that is necessary in the Harrison–Stevens procedure. Furthermore, our procedure leads to a unique filtering procedure that is relatively simple and easy to implement.

To illustrate our procedure, consider the dynamic linear model; the generalization for non-normality is straightforward. Suppose the observation equation is of the form

$$\mathbf{y}_t = F_t\boldsymbol{\theta}_t + \mathbf{v}_t, \tag{1}$$

where

$$\mathbf{v}_t \sim a_1 N(\mathbf{0}, V_1) + a_2 N(0, V_2), \tag{2}$$

$a_1 \geq .90$, $a_2 = 1 - a_1$, and

$$V_2 = k^2 V_1, \qquad k^2 > 1. \tag{3}$$

Then the filter is as depicted in Figure 1. The derivation and application to various different problems is given in Guttman and Pena (1984).

In summary, the advantages of this approach are as follows:

1. The derived filter is unique.
2. There is a simple algorithm for updating results.
3. The approach encompasses, in a straightforward manner, the multivariate approach.

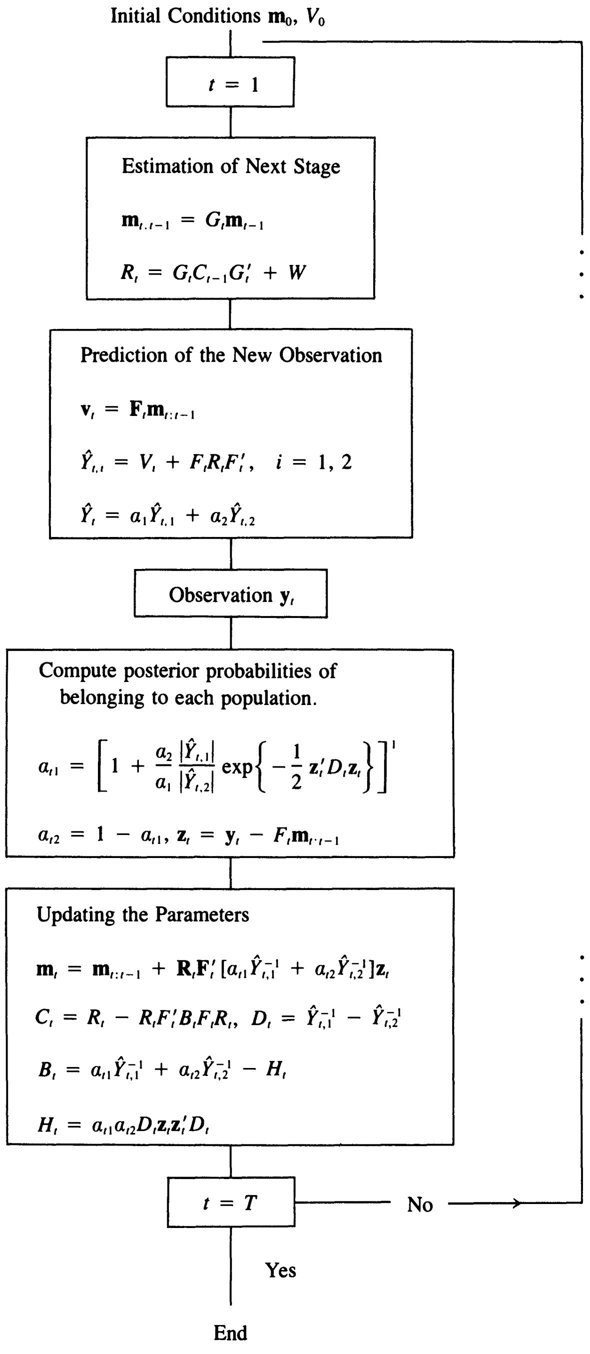

Figure 1. The Recursive Scheme for a Robust Kalman Filter.

* I. Guttman is Professor, Department of Statistics, University of Toronto, Ontario M5S 1A1, Canada, and D. Pena is Professor of Statistics, Universidad Politécnica de Madrid, Castellana 88, Madrid 6, Spain.

Journal of the American Statistical Association
March 1985, Vol. 80, No. 389, Theory and Methods

Consulting the "Updating the parameters" box of Figure 1, we note that the updating formula for $\mathbf{m}_{t:t} = \mathbf{m}_t$ of our knowledge of $\boldsymbol{\theta}_t$ involves a weighted combination of two filtering procedures, the weights being a_{t1} and $a_{t2} = 1 - a_{t1}$. If the posterior evidence is high that $\mathbf{y}_t$, the current observation, is from the intended source $N(F_t\boldsymbol{\theta}_t, V_1)$, that is, if $a_{t1} \approx 1$, then the residual $\mathbf{z}_t = (\mathbf{y}_t - F_t\mathbf{m}_{t:t-1})$ is filtered by $R_tF_t^{-1}Y_{t,1}^{-1}$, which is the appropriate Kalman gain matrix if indeed $\mathbf{y}_t$ is from the intended source, whereas the opposite is true if $a_{t2} \approx 1$, and so forth.

All of this is to say that the use of a mixture of two normals in the observation equation leads to sensible results that are easy to implement in the resulting recursive scheme. This requires the simple approximation of a two-termed posterior by a single normal, matched by means and variances. This gives the unique approximation talked about earlier.

ADDITIONAL REFERENCES

Box, G. E. P., and Tiao, G. C. (1968), "A Bayesian Approach to Some Outlier Problems," *Biometrika*, 55, 119–129.

Guttman, Irwin, and Pena, Daniel (1984), "Robust Kalman Filtering and Its Applications," technical report, University of Wisconsin, Mathematics Research Center.

Jeffreys, H. (1961), *Theory of Probability* (3rd ed.), Oxford: Clarendon Press.

Comment

CORNELIS A. LOS*

As a non-Bayesian econometrician, I wholeheartedly endorse West, Harrison, and Migon's (WHM) methodology of specifying structural parametric (state-space) models. The following six comments on their interesting and thought-provoking article are made with this perspective in mind.

1. Let me first introduce some simple concepts to facilitate the discussion. In the standard generalized linear model (GLM) the relationship between the dependent variable and the predetermined variables, or regressors, is assumed to be constant over time. In its simplest form this assumption can be represented by the regression equation

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + e_t, \qquad e_t \sim N(0, \sigma^2), \tag{1}$$

where y_t is the dependent variable, $\mathbf{x}_t$ the $(1 \times k)$ vector of predetermined variables, $\boldsymbol{\beta}$ the $(k \times 1)$ vector of constant, unknown, *structural* regression parameters, and e_t the measurement error (which, for simplicity, is assumed to be normally distributed).

In a nonstandard, evolutionary, or dynamic, general linear model (DGLM), the structural parameters vary over time, are stochastic, and, possibly, are subject to various policy influences. These assumptions can be represented by the following equations:

$$y_t = \mathbf{x}_t\boldsymbol{\beta}_t + e_t, \qquad e_t \sim N(0, \sigma^2), \tag{2}$$

where $\boldsymbol{\beta}_t \sim N(\hat{\boldsymbol{\beta}}_t, \Sigma_t)$ and $\hat{\boldsymbol{\beta}}_t = E\{\boldsymbol{\beta}_t\}$, the evolutionary mean (mode) $(k \times 1)$ vector of the evolutionary structural parameters $\boldsymbol{\beta}_t$ and $\Sigma_t = E\{(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_t)(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_t)'\}$, the time-varying covariance matrix of $\boldsymbol{\beta}_t$. The logic of modeling requires a specification of the processes of these structural parameters to be able to track their evolution.

A very general model specification is a controlled, multidimensional Markov process that allows for an almost infinite number of possible evolution trajectories for the elements of $\boldsymbol{\beta}_t$. For example, let the elements of $\boldsymbol{\beta}_t$ be selected from a $(n \times 1)$ vector of dynamic states $\mathbf{z}_t$ by the equation

$$\boldsymbol{\beta}_t = H\mathbf{z}_t, \tag{3}$$

where H is a known $(k \times n)$ selection matrix of zeros and ones, which contains design information and exhibits the property that $HH' = I$, a $(k \times k)$ identity matrix. Equations (2) and (3) can be written as one equation as in WHM's article:

$$y_t = \bar{\mathbf{x}}_t\mathbf{z}_t + e_t, \tag{4}$$

where $\bar{\mathbf{x}}_t = \mathbf{x}_tH$. Let the state vector $\mathbf{z}_t$ evolve according to a transition equation in state-phase form:

$$\mathbf{z}_t = F\mathbf{z}_{t-1} + H'G\mathbf{u}_t + H'\mathbf{v}_t, \qquad \mathbf{v}_t \sim N(0, Q), \tag{5}$$

where $\mathbf{u}_t$ is the $(m \times 1)$ vector of known structural policy variables (possibly a subset of the predetermined variables) and $\mathbf{v}_t$ the $(k \times 1)$ vector of process disturbances.

The constant *hyperstructural* parameters F $(n \times n)$, G $(k \times m)$, Q $(k \times k)$, and σ^2 specify the evolutionary processes of the elements of the state vector $\mathbf{z}_t$ and, therefore, also of the parameter vector $\boldsymbol{\beta}_t$. This model is a high-order DGLM with constant hyperstructural parameters, which is better adapted to the modeling situation in econometrics than the ones WHM discuss in their article.

2. The proper statistical problem of inference is to estimate and test, simultaneously, $\boldsymbol{\beta}_t$, Σ_t and F, G, Q, and σ^2. But WHM assume the values of F, (G), Q, and σ^2 to be known a priori and proceed accordingly. As so many other statistical practitioners, they uncritically adopt the Kalman filter from control engineering to track $\boldsymbol{\beta}_t$ and Σ_t or, more precisely, to track $\mathbf{z}_t$ and P_t, the covariance matrix of $\mathbf{z}_t$. WHM confuse tracking with inferring values of the structural parameters, and they leave the rabbit of statistical inference in their respective Bayesian hats. Kalman (1960) assumed for his filters the values of

* Cornelis A. Los is Economist, Domestic Research Department, Federal Reserve Bank of New York, New York, NY 10045.

Journal of the American Statistical Association
March 1985, Vol. 80, No. 389, Theory and Methods