Non-Gaussian State-Space Modeling of Nonstationary Time Series: Comment: Robustness, Computation, and Non-Euclidean Models

R. Douglas Martin; Adrian E. Raftery

Journal of the American Statistical Association, Vol. 82, No. 400 (Dec., 1987), 1044-1050.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198712%2982%3A400%3C1044%3ANSMONT%3E2.0.CO%3B2-1

The densities in (12)–(14) can be obtained from $p(Bx_{n+1} \mid Y_N)$, $p(Bx_n \mid Y_n)$, and $p(Ax_{n+1} \mid Y_n)$. If B_2 is null, then (13) reduces to

$$p(Bx_n \mid Y_N) = \int p(Bx_n \mid x_{n+1}, Y_N) \times p(Bx_{n+1} \mid Y_N)\, dBx_{n+1} \quad (15)$$

and (14) becomes

$$p(Bx_n \mid x_{n+1}, Y_N) = p(A_1 x_{n+1} \mid x_n) p(Bx_n \mid Y_n) \div p(A_1 x_{n+1} \mid A_2 x_{n+1}, Y_n). \quad (16)$$

If A_2 is null, then Kitagawa's Equations (2.4) and (2.5) apply.

It is quite easy to find suitable transformations of x_n and y_n for the aforementioned formulas. Suppose, for example, that we want to find Ax_n. If we assume for the moment that v_n and w_n are Gaussian, then we can find the conditional covariance matrix $\text{var}(x_n \mid Y_{n-1})$ with the Kalman filter. This has a Cholesky factorization in the form $\text{var}(x_n \mid Y_{n-1}) = L\Lambda L'$, where L is lower triangular with 1s on the diagonal and Λ is diagonal with nonnegative elements. Then a suitable choice for Ax_n is the subvector of $L^{-1}x_n$ whose elements correspond to nonzero elements of Λ. As a second example, we choose $C_1 y_n$ and $C_2 y_n$ to be subvectors of y_n determined by examining the Cholesky factorization of the covariance matrix of $(x_n', y_n')'$ conditional on Y_{n-1}. Although the densities in the model are non-Gaussian, it is nevertheless a linear model, and the procedure we suggest simply makes use of this fact.

Fixed-Point Smoothing. Using the results (9)–(16) it is possible to find a generalization of the fixed-point smoothing recursion formula for $p(Bx_n \mid Y_N)$. Rather than give full details, we consider only scalar y_n and assume that G has full rank and w_n is nonzero so that the degeneracies discussed previously do not occur. If they do, the following steps can be modified along the lines discussed earlier. Similar to Kitagawa's Equation (2.2), we have

$$p(x_{l+1}, x_l \mid Y_l) = p(x_{l+1} \mid x_l) p(x_l \mid Y_l), \quad (17)$$

and for $n > l + 1$

$$p(x_n, x_l \mid Y_{n-1}) = \int p(x_n \mid x_{n-1}) p(x_l, x_{n-1} \mid Y_{n-1})\, dx_{n-1}. \quad (18)$$

Corresponding to (2.3), the updating equation is

$$p(x_n, x_l \mid Y_n) = p(y_n \mid x_n) p(x_n, x_l \mid Y_{n-1}) \div p(y_n \mid Y_{n-1}), \quad (19)$$

and we obtain $p(x_l \mid Y_N)$ from

$$p(x_l \mid Y_N) = \int p(x_N, x_l \mid Y_N)\, dx_N. \quad (20)$$

The recursions (17)–(19) followed by (20) may in some cases be a more efficient way of finding $p(x_l \mid Y_N)$ for a single value of l than the generalized filtering and fixed-interval smoothing formulas (2.2)–(2.5).

ADDITIONAL REFERENCES

Ansley, C. F., and Kohn, R. (1982), "A Geometrical Derivation of the Fixed Interval Smoothing Algorithm," *Biometrika*, 69, 486–487.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

Harvey, A. C., and Todd, P. H. J. (1983), "Forecasting Economic Time Series With Structural and Box–Jenkins Models: A Case Study" (with discussion), *Journal of Business & Economic Statistics*, 1, 299–315.

Jones, R. H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations," *Technometrics*, 22, 389–395.

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.

Sorenson, H. W., and Alspach, D. L. (1971), "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, 7, 465–479.

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Comment

Robustness, Computation, and Non-Euclidean Models

R. DOUGLAS MARTIN and ADRIAN E. RAFTERY*

1. INTRODUCTION

We would like to thank Kitagawa for an interesting article on the analysis of non-Gaussian time series. In the introduction, Kitagawa puts forth the two objectives of his article. The primary objective is to reveal the importance of non-Gaussian models in various problems of time series analysis, and the secondary objective is to present a methodology for non-Gaussian time series models.

The first objective is achieved through the treatment of three examples of non-Gaussian models: an artificial ex-

* R. Douglas Martin is Professor of Statistics and Adrian E. Raftery is Associate Professor of Statistics and Sociology, both at the University of Washington, Seattle, WA 98195. This work was supported by Office of Naval Research Contract N00014-84-C-0169. The authors are grateful to Andrew Bruce for computational assistance.

Journal of the American Statistical Association
December 1987, Vol. 82, No. 400, Theory and Methods

ample consisting of independent normal observations with constant variance and switching mean; a real example for which the transformed observations consist of signal plus noise, with the signal being the logarithm of a variance and the noise being distributed as the logarithm of an exponential random variable; and a real example where the problem is to estimate the mean of a nonstationary binary process.

The second objective is achieved by showing that quite reasonable results can be obtained through use of a primitive piecewise linear approximation of the non-Gaussian densities that specify proposed models. The importance of non-Gaussian models in time series has hardly been ignored in the statistical literature. It is true that some of the kinds of behavior exhibited by Kitagawa's example deserve to be taken more seriously by practitioners. The main statistical conclusions, however, are not surprising: use of a correct probability model (or a sufficiently good approximation thereto) for the problem at hand will usually produce good inference results using any one of a number of well-known procedures, for example, quantiles, mode, median, highest posterior density region for the conditional filtering or smoothing densities, conditioned on the observations.

The main difficulty in non-Gaussian state–space modeling, whose importance has been recognized in some contexts for a long time, is the intractability of the conditional filtering and smoothing densities. Although the author's approach to surmounting this difficulty produces pleasing results for his three examples, its computational complexity may prove to be a deterrent for all but simple problems.

Detailed comments follow. In Section 2 we discuss Kitagawa's results in the context of models of a general type into which the first and second examples fall and point out related work concerning outliers and robustness. Section 3 deals with computational issues, and Section 4 considers state–space modeling for general non-Gaussian time series. Section 5 returns to Kitagawa's rainfall example and the general issue of categorical time series. Section 6 concludes with some brief comments concerning Kitagawa's use of the Akaike information criterion (AIC) and the problem of comparing nonnested models.

2. NON-GAUSSIAN MODELS, OUTLIERS, AND ROBUSTNESS

2.1 Non-Gaussian Distributions, Outliers, and Prior Work

For linear state–space models of the form (2.1), one can have non-Gaussian distributions in the state noise only, the observation noise only, or in both. Suppose that the non-Gaussian distributions for the state noise, the observation noise, or both are moderately Gaussian in the middle but have heavier tails than a Gaussian distribution that fits the non-Gaussian distribution fairly well in the middle. Then the data will contain outliers in the state noise, in the observation noise, or in both observation and state noises. Distributions that might be considered nearly Gaussian include some members of the Pearson family used by Kitagawa, Student-t distributions, the logistic distribution, and mixtures of Gaussian distributions of the form

$$CN(\gamma, \sigma_1^2, \sigma_2^2) = (1 - \gamma)N(0, \sigma_1^2) + \gamma N(0, \sigma_2^2),$$

where γ is small and $\sigma_2^2 \gg \sigma_1^2$. The nearly degenerate case in which $\sigma_1^2 = 0$ is also of frequent interest.

It is, by now, common to refer to outliers in the state noise as *innovations outliers* (IO) and to outliers in the observation noise as *additive outliers* (AO). We shall use IO and AO in the exclusive sense and write AO + IO to refer to situations where outliers occur in both the state and observation noise. Kitagawa's first and second examples fall into this framework. The first example is of IO type, with an appropriate state noise distribution being of mixture type with degenerate central component, and perhaps Gaussian contamination component, if one wants to allow random values for the locations and sizes of the jumps. The second example is of IO + AO type, since the level shifts are reasonably well modeled by outliers in the state noise and the negative outliers are accounted for by the observation noise distribution $r(x)$ in (5.6). Although $r(x)$ is not Gaussian in the central part of the distribution, such an approximation is not too crude. Furthermore, the heavy negative tail, which gives rise to the negative outliers, would seem to be the most troublesome feature of $r(x)$.

Given the framework into which these two examples fit, it is unfortunate that Kitagawa's introduction and list of references create the impression that the problem of non-Gaussian state–space modeling has received little attention since the early 1970s. This is hardly the case, as the following partial list of work on particular subtopics indicates: (a) non-Gaussian filters and robust filters for IO and AO—Masreliez (1975), Masreliez and Martin (1977), Ershov and Lipster (1978), West (1981), Guttman and Pena (1985); (b) non-Gaussian smoothers and robust smoothers for AO—Martin (1979); (c) model fitting for AO—Martin (1981), Martin, Samarov, and Vandaele (1983); (d) model fitting for IO + AO—Harrison and Stevens (1976), Hillmer, Bell, and Tiao (1983), Smith and West (1983), Tsay (1986). We note that much of the work in (a) and (b) deals with the general non-Gaussian problem, and not just with the heavy-tailed, nearly Gaussian case. It is also appropriate to mention the important work of Box and Tiao (1975) in connection with "intervention" modeling of additive structures.

We now turn to some small issues concerning Kitagawa's first two examples. As we suggested earlier, a mixture model with degenerate central component and any one of a variety of mixing components would be more appropriate than the Pearson family for the state noise in the first example. The very small scale estimate $\hat{\tau}^2 = 2.2 \times 10^{-7}$ is essentially trying to model the qualitative behavior of the mixture model, since it represents state noise that is nearly 0 most of the time, while the heavy tail is needed, correspondingly, to capture the large jumps of ± 1.

One should recall the general behavior of sums of stable random variables (and here we are talking, roughly, about symmetric stable random variables with index .5). The sums are dominated by an individual term because of the infrequent occurrence of extremely large values. It is thus not surprising that distributions in the domain of attraction of stable laws (e.g., the Pearson family) are reasonable surrogates for mixture distributions with degenerate or nearly degenerate central components.

For the second example, it is not clear why the author makes the assumption that the state noise is Cauchy, particularly in the light of the flexibility demonstrated by the Pearson family in the first example. Again, a mixture family might be more appropriate.

One of the lessons of Kitagawa's work is that when there is non-Gaussian behavior of an outlier-generating, heavy-tailed form, a procedure based on *almost any reasonable* non-Gaussian model will usually do much better than the classical, Gaussian-based procedure. Furthermore, an appropriately constructed, non-Gaussian-based approach will usually suffer by only a small amount if it is used when, in fact, the Gaussian model holds.

This is hardly a new theme, since it has been the motivation for certain robust procedures, with which the statistical literature abounds. The classic example is that of Huber's (1964, 1981) M-estimates. Some quite good M-estimates are simply maximum likelihood estimates (MLE's) for a heavy-tailed distribution, Huber's favorite being the MLE for a distribution that is normal in the middle and exponential in the tails. Furthermore, both the robustness viewpoint and the construction of approximately optimal non-Gaussian filters and smoothers to deal with outliers in time series are well-represented in the references provided earlier. See, for example, Martin (1981) or Martin and Yohai (1985) for an approximate non-Gaussian MLE rationale for dealing with AO problems.

The latter reference also provides a technical definition of robustness for time series, which is due to Boente, Fraiman, and Yohai (1987). In the present discussion we take robustness to mean the simple data-oriented notion of *resistance* to outliers. The technical probability-oriented definition alluded to is quite close in spirit to this simple data-oriented notion.

2.2 IO + AO

Returning to Kitagawa's second example, it must be emphasized that this is indeed an important and interesting example in that it involves both innovations and additive outliers. The early work of Fox (1972) on testing for IO and AO led to subsequent robustification of Fox's tests by Martin and Zeh (1977), and the successful use of non-robustified Fox tests by Hillmer et al. (1983) and Tsay (1986) for dealing with IO + AO situations. In addition, Harrison and Stevens (1976) and Smith and West (1983) used an attractive approach to such problems based on limiting the computational complexity resulting from Gaussian mixture models for the state and observation noises, with a Bayesian flavor. For a remarkable early, and relatively unknown, treatment of the pure IO case with normal mixture distributions for the innovations, see Buxbaum and Haddad (1969).

The problem of dealing with IO + AO is a central one in time series, which occurs in many fields. Proper treatment of the problem is sometimes important for parameter estimation and always crucial for forecasting/prediction. To provide a reasonable forecast, one must really have a good idea whether outliers near the end of the series are IO, AO, or both. Thus Kitagawa's demonstration that an approximating non-Gaussian model will give a reasonable answer for the IO + AO problem is most welcome. It is not clear, however, that the results are any better than those obtainable using the perhaps simpler approaches in some of the references just cited.

We note that Kitagawa's choice for the noise density $r(x)$ does not result in robustness. The reason is that, although the left tail is sufficiently heavy (exponential thickness) to provide protection against negative outliers, the right tail is considerably lighter than Gaussian tails. Thus the model will not provide protection against an occasional unanticipated positive outlier. An easy fix is to modify $r(x)$ so that it has a right tail with (at least) exponential thickness.

Incidentally, it appears that use of the conditional median as a point estimate (bold curves) in Figures 5 and 9 might be considerably better than use of the conditional mean for the posterior distribution. For the conditional means will be affected more by the vagaries of the tails of the conditional distributions. In this connection it is to be noted that the 99.87 and 97.73 percent curves in Figure 5 exhibit wiggles (away from the jumps) that are very similar to the wiggles associated with the linear conditional mean (i.e., Gaussian model) estimates in Figure 3. This similarity was brought to our attention by Andrew Bruce.

2.3 Non-Gaussian Core and Outliers

Our preceding comments have focused on non-Gaussian situations where the central part, or "core," of the distributions are nearly Gaussian and heavy-tailed deviations give rise to outliers. Similar comments apply to situations where the core of the distribution is substantially non-Gaussian. If the core is quite non-Gaussian and the tails are sufficiently heavy, then the optimal non-Gaussian smoother and parameter estimates will often already be robust toward outliers. On the other hand, a model with a substantially non-Gaussian core but with light tails will need to be suitably robustified—one way to do so being to construct optimal or nearly optimal procedures for modified distributions with sufficiently heavy tails. For example, consider the problem of estimating scale for an exponential distribution with iid observations. To obtain a bounded score function and thereby achieve robustness, one needs to use a distribution with tails at least as heavy as Cauchy tails. For a min–max robust solution to this problem, see Huber (1981) and Thall (1979). Incidentally, inliers can also pose a robustness problem when estimating scale parameters (see Martin and Zamar 1986).

3. COMPUTATIONAL CONSIDERATIONS

3.1 Computational Complexity of Kitagawa's Method

One wonders, as does the author in his concluding remarks, whether the piecewise linear approximation method he proposes is really practical. Andrew Bruce has implemented Kitagawa's procedure on the (quite fast) Symbolics computer in our department. For the first example, the running time, using 100 knots, was nearly 1 hour!

Furthermore, it appears that the computing time for Kitagawa's procedure is about $O(m^k)$, where k is the dimension of the state, and in the examples given, m is on the order of 400. This is because the prediction, filtering, and smoothing equations (2.2), (2.3), and (2.5) involve k-dimensional integrals. Thus the method seems unlikely to be of use in the real-time applications that arise in communications and control theory. In addition, even if it can be made to work on higher-dimensional problems where the data analyst has the time to analyze the data many ways, the run time may still be prohibitive. We hope that the author will provide us with some guidance and running time information in his rejoinder.

We also wonder how much approximation error occurs when one includes both the linear approximation and accumulated round-off error. Since one is forced to work with some degree of approximation, one may be much better off using alternative approximations that are more attractive computationally.

3.2 Approximations for AO Filters

Experience shows that for AO models, the computationally attractive conditional mean non-Gaussian filters of Masreliez (1975) and Masreliez and Martin (1977) work quite well in practice (Kleiner, Martin, and Thompson 1979; Martin et al. 1983; Martin and Thompson 1982). In addition, see Martin and Su (1985) for a detailed comparison of robust filters and guidelines for selecting tuning constants. It is to be emphasized that the conditional mean filters may be constructed for rather arbitrary non-Gaussian observation noise [see, e.g., fig. 2 of Masreliez (1975)].

It should also be emphasized that the result of Masreliez (1975) contains essentially only the Gaussian case and the Kalman filter as an *exact* result [see the proposition in sec. 5 of Martin (1979)]. We believe, however, that use of Masreliez's theorem produces estimates that are quite good approximations to the exact conditional mean in non-Gaussian AO situations. Some evidence in support of this claim may be found in the Monte Carlo results in figure 3 of Masreliez (1975) and in the theorem of section 5 in Martin (1979).

The key approximation property used to construct these filters is that the state prediction density is approximately Gaussian. The discovery that this approximation yielded intuitively appealing non-Gaussian filter recursions, with *data-dependent* covariance (unlike the Gaussian case) was due to Masreliez (1975), and his clever derivation also provides one of the nicest ways of establishing the standard Kalman filter recursions. Some theoretical justification for use of the Masreliez approximation is provided by the "continuity of state prediction densities" theorem in Martin (1979).

For somewhat different approximations, the reader is referred to Ershov and Lipster (1978), West (1981, 1982), and Guttman and Pena (1985). A thorough analysis of the quality of all of the approximations suggested to date remains to be carried out.

3.3 Approximations for AO Smoothers

It is possible to use Masreliez's Gaussian approximation for state prediction density to derive an approximately optimal conditional mean smoother. This is done in the approximate conditional mean smoother theorem of Martin (1979), where the resulting smoother has the "standard" form. One early derivation of the optimal linear smoother in standard form may be found in Meditch

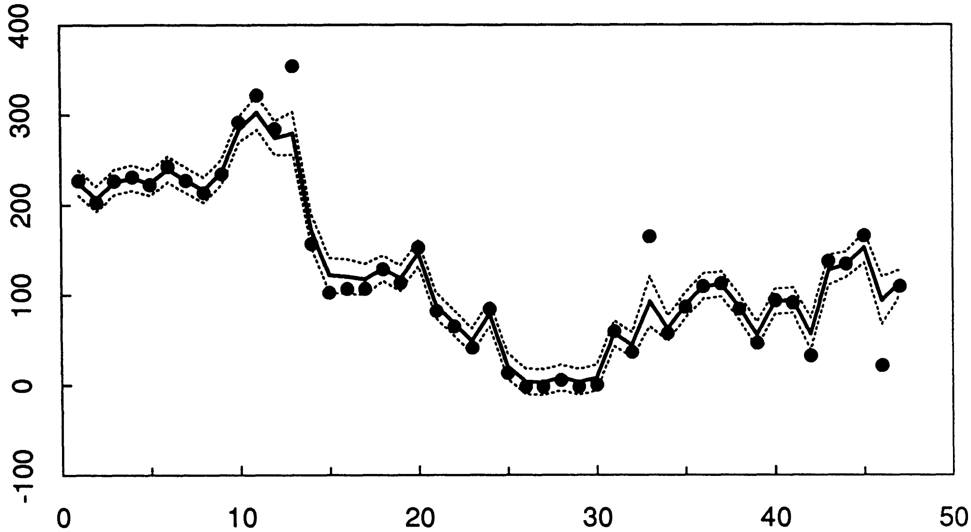

Figure 1. Suspended Deposits (dots) and Robust Smooth (solid line), With Standard Error Bands.

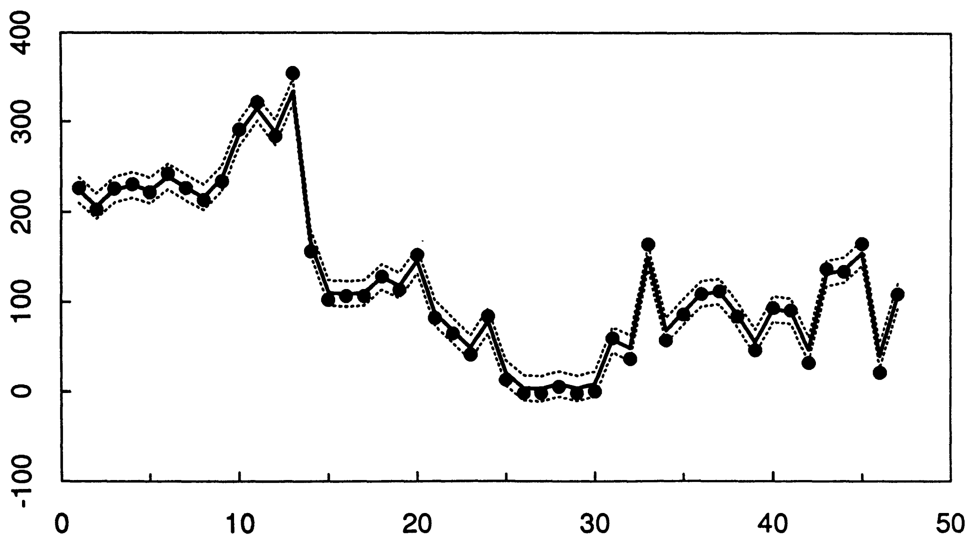

Figure 2. Suspended Deposits (dots) and Kalman Smooth (solid line), With Standard Error Bands.

(1967). A considerably simpler derivation was given more recently by Ansley and Kohn (1982). In figure 2 of Martin (1979), a robust smoother of this type is applied to "suspended deposits" data, and the result indicates that certain kinds of jumps in a process are handled rather nicely. Note that the approximately optimal smoother may be constructed for any non-Gaussian observation noise distribution, not just those that are nearly normal.

Actually, a "two-filter" form of approximate conditional mean smoother [see Solo (1982) for the linear/Gaussian case] is preferred, both with regard to ease of interpretation and jump handling ability. This form of smoother is constructed as a weighted sum of forward and backward approximate conditional mean filters of Masreliez type (or robust versions thereof), with the data-dependent weights having a natural Bayesian interpretation. We illustrated the efficacy of a two-filter version of robust smoother on the "suspended deposits" data. Figure 1 shows $\hat{x}_t$ and $\hat{x}_t \pm s_t$, along with the original data, where $\hat{x}_t$ is the robust smoother and s_t^2 is a data-dependent estimate of mean squared error for $\hat{x}_t$, and we arbitrarily assumed a Gaussian observation noise with mean zero and standard deviation 20 (details will be provided in a forthcoming technical report). We see that $\hat{x}_t$ has reasonable resistance toward outliers and handles the jump around $t = 12, 13$ well. By way of comparison, results for the usual linear Kalman smoother are shown in Figure 2, where three outliers exert undue influence.

Following Kitagawa's results, we are motivated to construct a conditional median or conditional mode variant of the approximate conditional mean type estimate $\hat{x}_t$ used in Figure 1. Even nicer point estimates may result.

3.4 Approximate Conditional Mean Filters and Smoothers for Non-Gaussian State Noise

Masreliez (1975) and Masreliez and Martin (1977) provided theoretical results concerning exact conditional mean and robust filters for the case of non-Gaussian state noise. However, it turns out that, unlike the case of non-Gaussian observation noise, these results do not appear to immediately provide a filter structure via the simplifying assumption of Gaussianity of the state prediction density. When the state noise is non-Gaussian, the state prediction density can be quite non-Gaussian, a fact we had overlooked by not paying much attention to the non-Gaussian state noise case in the past. Thus Kitagawa's work motivates us to study further the problem of obtaining computationally attractive filters and smoothers for non-Gaussian state noise problems and for simultaneous non-Gaussian state noise and non-Gaussian observation noise. For heavy-tailed nearly Gaussian distributions, such filters and smoothers will, it is hoped, provide fast robust procedures for IO and IO + AO cases. Many practical situations make this an important area for study.

3.5 Special Approaches for Level Shifts, Variance Changes, Etcetera

Kitagawa's first two examples involve two basic kinds of structural change that occur frequently in time series and must be dealt with adequately: level shifts and changes in variance that may or may not be abrupt. Robustness deals with outliers but not, in general, with such structural changes (other structural changes, such as change in slope of trend, etc., may also need to be dealt with; see Smith and West 1983). The best overall strategy is probably to combine robustness with special methods for dealing with other structural changes. Although, as Kitagawa demonstrates, it may be possible to deal with both problems by a single non-Gaussian model, this may not always be possible or desirable, from a computational point of view, among others.

Box and Tiao's (1975) intervention analysis indicates that certain additive structures may be nicely dealt with by dummy variables, when the time of onset is known. Furthermore, Harrison and Stevens (1976) provided a nice framework for dealing with a variety of structural changes, including the types mentioned previously. Further developments are given by West (1986a) and West and Harrison

(1986). Our question to Kitagawa is: Would it not be preferable, from the computational point of view at the very least, to develop an approach using parametric structures for dealing with such special effects as level shifts, change in variance? In addition, for slowly changing variances, should one approach the problem nonparametrically [e.g., as did Carroll (1982) in the linear model setting]?

4. STATE–SPACE MODELING FOR GENERAL NON-GAUSSIAN TIME SERIES

Kitagawa's approach is restricted to situations where both observation and state take values in Euclidean space; in addition, the underlying model (2.1) is linear. Thus there are many kinds of non-Gaussian time series that do not fit easily into his framework. These include unordered, or partially ordered, categorical time series; time series of angles; time series of compositional data, defined on the simplex; and time series of contingency tables.

We shall now discuss a class of non-Gaussian state–space models that may be able to deal with such problems. We first review the model of Raftery (1985a,b) for a stationary time series $\{Z_n\}$ taking values in an arbitrary space $\mathbf{Z}$. It was introduced by Raftery (1985a) in the context of discrete-valued time series as a model for high-order Markov chains; however, the approach is much more general.

Suppose that (V_i, W_i) $(i = 1, \ldots, p)$ is a set of bivariate random vectors taking values in $\mathbf{Z} \times \mathbf{Z}$, with conditional densities $f_i(v \mid w)$ with respect to some measure, where the marginal distribution of V_i is the same as that of W_i for each $i = 1, \ldots, p$. Suppose that the conditional density of Z_n given $Z_{n-1}, \ldots, Z_{n-p}$ is given by

$$p(z_n \mid z_{n-1}, \ldots, z_{n-p}) = \sum_{i=1}^{p} \lambda_i f_i(z_n \mid z_{n-i}), \quad (1)$$

where $\Sigma\lambda_i = 1$. This is called a *linear conditional probability* (LCP) time series model. In the discrete-valued case it fits data well, can be physically motivated, is flexible and easy to generalize to other dependence patterns, and is analogous in several ways to the standard autoregressive model.

Consideration of (1) suggests the following system equation in the non-Gaussian state–space model as an alternative to Kitagawa's (2.1). Suppose that $\mathbf{x}_n = (x_{n1}, \ldots, x_{nk})^T$, where k is the dimension of the state, and that $g_i(\mathbf{x}_n \mid x_{n-1,i})$ $(i = 1, \ldots, k)$ are conditional densities, with obvious meanings. Then we consider models having state transition densities of the form

$$p(\mathbf{x}_n \mid \mathbf{x}_{n-1}) = \sum_{i=1}^{k} \lambda_i g_i(\mathbf{x}_n \mid x_{n-1,i}), \quad (2)$$

where $\Sigma\lambda_i = 1$. The observation equation, which specifies $p(\mathbf{y}_n \mid \mathbf{x}_n)$, could have a similar form, but this is not necessary. Transition densities of the form (2) clearly can represent time series defined on a wide variety of spaces. They also include all of the models that Kitagawa considers, with $k = 1$.

Another advantage of (2) is that it may render Kitagawa's numerical technique computationally feasible when k exceeds 1. The reason is that the multiple integrals required by Kitagawa's primitive technique now reduce to sums of single integrals. For example, the prediction equation (2.2) becomes

$$p(\mathbf{x}_n \mid Y_{n-1}) = \sum_{i=1}^{k} \lambda_i \int g_i(\mathbf{x}_n \mid x_{n-1,i}) p(x_{n-1,i} \mid Y_{n-1})\, dx_{n-1,i},$$

which can be approximated using the one-dimensional technique in Section 3 of Kitagawa's article.

LCP state–space models other than (2) can be defined and may be more useful. For example, a kth order Markovian model with observation error is obtained when the right side of (2) is multiplied by $(k - 1)$ singular measures. This is analogous to the state–space form of the standard kth-order autoregressive model with observation error (Harvey 1981, chap. 4).

5. THE RAINFALL DATA AND CATEGORICAL TIME SERIES

The analysis of the rainfall data in Kitagawa's third example is based on the assumption that, conditional on the slowly changing p_n, which is essentially the seasonal effect, the occurrence of rainfall on successive days is independent. We wonder if this is, in fact, the case. Analyses of rainfall data from many parts of the world indicate that it is not [see, e.g., Stern and Coe (1984) and references therein].

If it is not, the dependence could itself be modeled, perhaps using a Markovian assumption; the LCP model might be helpful there. This would involve analyzing the actual data, rather than the aggregated data in Kitagawa's Table 2. We conjecture that this would lead to smoother estimates of p_n and, perhaps, to broader confidence bands as well.

Techniques for binary time series do not always generalize to unordered, categorical time series with more than two categories. For example, the mean is well-defined in the binary case but not in the general categorical case. West, Harrison, and Migon (1985) and West (1986b) proposed interesting new models for binary time series, which are similar to those of Kitagawa. It is not yet clear, however, how such techniques carry over to general categorical time series, for which the LCP approach may be useful.

6. MODEL COMPARISON

We are surprised to see that Kitagawa bases his model comparisons in the first two examples on the AIC. He justifies this by referring to Akaike's interpretation of likelihood from which the AIC is naturally derived, citing Akaike (1973). This derivation of the AIC, however, is based on the quadratic approximation to the likelihood, which is precisely what Kitagawa's article is trying to get away from. In addition, use of the AIC seems problematical, even asymptotically, when nonnested models are being compared, as in the first example. This is because the sampling properties of the minimum AIC procedure seem to be known only when the models being compared are nested (see, e.g., Findley 1987, sec. 5).

Bayes factors, by contrast, provide an exact, finite-sample solution to the model-comparison problem and should be available as by-products of the procedure proposed in Kitagawa's article. They have a precise inferential interpretation, which applies whether or not the models being compared are nested. Bayes factors have been used successfully to compare nonnested models in different situations, including point processes (Raftery and Akman 1986), contingency tables (Raftery 1986), and problems involving unknown population size (Raftery 1987).

Akaike (1983) pointed out that, asymptotically, model comparisons based on the AIC are approximately equivalent to those based on Bayes factors. This is true, however, only in the rather special situation in which the information in the prior increases at the same rate as the information in the likelihood. If prior information is small relative to the information provided by the data, then Bayes factors based on such priors, or approximating vague priors, should be used, rather than the AIC. If this is the case, an analysis based on Bayes factors will often favor models simpler than those selected by the minimum AIC procedure.

ADDITIONAL REFERENCES

Akaike, H. (1983), "Information Measures and Model Selection," *Bulletin of the International Statistical Institute*, 50, 277–290.

Ansley, C. F., and Kohn, R. (1982), "A Geometrical Deviation of the Fixed Interval Smoothing Algorithm," *Biometrika*, 69, 486–487.

Boente, G., Fraiman, R., and Yohai, V. J. (1987), "Qualitative Robustness for General Stochastic Processes," *The Annals of Statistics*, 15, 1293–1312.

Box, G. E. P., and Tiao, G. C. (1975), "Intervention Analysis With Applications to Economic and Environmental Problems," *Journal of the American Statistical Association,* 70, 70–79.

Buxbaum, P. J., and Haddad, R. A. (1979), "Recursive Optimal Estimation for a Class of Non-Gaussian Processes," in *Computer Processing in Communications Proceedings,* ed. J. Fox, Brooklyn, NY: Polytechnic Press, pp. 375–399.

Carroll, R. J. (1982), "Adapting for Heteroscedasticity in Linear Models," *The Annals of Statistics*, 10, 1224–1233.

Ershov, A. A., and Lipster, R. S. (1978), "Robust Kalman Filter and Discrete Time," *Automation and Remote Control*, 3, 60–69.

Findley, D. F. (1987), "An Analysis of AIC and the Entropy Maximization Principle for Comparing Exact or Approximate Models for Random Samples, Including Non-nested and Conditional Models," technical report, U.S. Bureau of the Census, Washington, DC.

Fox, A. J. (1972), "Outliers in Time Series," *Journal of the Royal Statistical Society*, Ser. B, 34, 350–363.

Guttman, I., and Pena, D. (1985), Comment On "Dynamic Generalized Linear Models and Bayesian Forecasting," by M. West, P. J. Harrison, and H. S. Migon, *Journal of the American Statistical Association*, 80, 91–92.

Harrison, P. J., and Stevens, C. F. (1976), "Bayesian Forecasting" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

Harvey, A. C. (1981), *Time Series Models*, New York: Halsted Press.

Hillmer, S. C., Bell, W. R., and Tiao, G. C. (1983), "Modeling Considerations in the Seasonal Adjustment of Economic Time Series," in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, DC: U.S. Bureau of the Census, pp. 74–100.

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73–101.

——— (1981), *Robust Statistics*, New York: John Wiley.

Kleiner, B., Martin, R. D., and Thompson, D. J. (1979), "Robust Estimation of Power Spectra," *Journal of the Royal Statistical Society*, Ser. B, 41, 313–351.

Martin, R. D. (1979), "Approximation Conditional Mean Type Smoothers and Interpolators," in *Smoothing Techniques for Curve Estimation*, eds. T. Gasser and M. Rosenblatt, Berlin: Springer-Verlag, pp. 117–143.

——— (1981), "Robust Methods for Time Series," in *Applied Time Series II*, ed. D. F. Findley, New York: Academic Press, pp. 683–759.

Martin, R. D., Samarov, A., and Vandaele, W. (1983), "Robust Methods for ARIMA Models" (with discussion), in *Applied Time Series Analysis of Economic Data*, ed. A Zellner, Economic Research Report ER-t Washington, DC: U.S. Bureau of the Census, pp. 153–177.

Martin, R. D., and Su, K. Y. (1985), "Robust Filters and Smoothers: Definition and Design," Technical Report 58, University of Washington, Seattle, Dept. of Statistics.

Martin, R. D., and Thompson, D. J. (1982), "Robust Resistant Spectrum Estimation," *IEEE Proceedings*, 70, 1097–1115.

Martin, R. D., and Yohai, V. J. (1985), "Robustness in Time Series and Estimating ARMA Models," in *Handbook of Statistics* (Vol. 5), eds. E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, New York: Elsevier, pp. 119–155.

Martin, R. D., and Zamar, R. (1986), "Min-Max Bias *M*-Estimates of Scale," Technical Report 72, University of Washington, Seattle, Dept. of Statistics.

Martin, R. D., and Zeh, J. E. (1977), "Determining the Character of Time Series Outliers," in *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pp. 818–823.

Masreliez, C. J. (1975), "Approximate Non-Gaussian Filtering With Linear State and Observation Relations," *IEEE Transactions on Automatic Control*, 20, 107–110.

Masreliez, C. J., and Martin, R. D. (1977), "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter," *IEEE Transactions on Automatic Control*, 22, 361–371.

Meditch, J. S. (1967), "Orthogonal Projections and Discrete Optimal Linear Smoothing," *SIAM Journal on Control and Optimization*, 5, 74–89.

Raftery, A. E. (1985a), "A Model for High-Order Markov Chains," *Journal of the Royal Statistical Society*, 47, 528–539.

——— (1985b), "A New Model for Discrete-Valued Time Series: Autocorrelations and Extensions," *Rassegna di Metodi Statistici ed Applicazioni*, 3–4, 149–162.

——— (1986), "A Note on Bayes Factors for Log-Linear Contingency Table Models With Vague Prior Information," *Journal of the Royal Statistical Society*, Ser. B, 48, 249–250.

——— (1987), "Inference and Prediction for a General Order Statistic Model With Unknown Population Size," *Journal of the American Statistical Association*, 82, 1163.

Raftery, A. E., and Akman, V. E. (1986), "Bayesian Analysis of a Poisson Process With a Change-Point," *Biometrika*, 73, 85–89.

Smith, A. F. M., and West, M. (1983), "Monitoring Renal Transplants: An Application of the Multiprocess Kalman Filter," *Biometrics*, 39, 867–878.

Solo, V. (1982), "Smoothing Estimation of Stochastic Processes: Two-Filter Formulas," *IEEE Transactions on Automatic Control*, 27, 473–476.

Stern, R. D., and Coe, R. (1984), "A Model Fitting Analysis of Daily Rainfall Data" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 147, 1–34.

Thall, P. F. (1979), "Huber-Sense Robust *M*-Estimation of a Scale Parameter, With Application to the Exponential Distribution," *Journal of the American Statistical Association*, 74, 147–152.

Tsay, R. S. (1986), "Time Series Model Specification in the Presence of Outliers," *Journal of the American Statistical Association*, 81, 132–141.

West, M. (1981), "Robust Sequential Approximate Bayesian Estimation," *Journal of the Royal Statistical Society*, Ser. B, 43, 157–166.

——— (1982), "Aspects of Recursive Bayesian Estimation," unpublished Ph.D. dissertation, University of Nottingham, England.

——— (1986a), "Bayesian Model Monitoring," *Journal of the Royal Statistical Society*, Ser. B, 48, 70–78.

——— (1986b), Discussion of "Influence Functionals for Time Series," by R. D. Martin and V. J. Yohai, *The Annals of Statistics*, 14, 838–840.

West, M., and Harrison, P. J. (1986), "Monitoring and Adaption in Bayesian Forecasting Models," *Journal of the American Statistical Association*, 81, 741–750.

West, M., Harrison, P. J., and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting" (with comments and rejoinder), *Journal of the American Statistical Association*, 80, 73–97.