NUSC Reprint Report 8407
6 October 1988

Signal Model Analysis Via Model-Critical Methods

G. R. Swope
Submarine Sonar Department

A. S. Paulson
Rensselaer Polytechnic Institute

Naval Underwater Systems Center
Newport, Rhode Island / New London, Connecticut

Reprint of an article published in the Proceedings of the IEEE-sponsored " Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling," Spring Hill Conference Center, Minneapolis, MN, August 3-5, 1988.

(1)

SIGNAL MODEL ANALYSIS VIA MODEL-CRITICAL METHODS

GERALD R. SWOPE
Naval Underwater Systems Center
New London, CT 06320

ALBERT S. PAULSON
Dept. Operations Research and Statistics
Rennsselaer Polytechnic Institute
Troy, NY 12180

ABSTRACT

Model-critical procedures provided a means to scrutinize an assumed parametric statistical model by varying the way the data is processed for repeated fits to the model. The criticism of the data is accomplished using the generalized likelihood function for the assumed probability density of the data. The degree of criticism is controlled by a user specified constant c. The model-critical parameter estimates are obtained by maximization of the generalized likelihood function. When $c = 0$, no criticism is performed and maximum likelihood estimates are obtained. These procedures can indicated if any model assumptions have been violated. Model critical estimation procedures are presented for autoregressive (AR) models. The analysis of an AR example is presented.

Keywords: [illegible] (KR)

INTRODUCTION

Signal models are as the name implies only models of reality. As such, they only approximate the true process and in general, more than one model can be used to describe a given set of data. Under these conditions, the experimenter must determine models which adequately describe the observed data. Depending on how the model is to be used, one of the candidate models is chosen. The selected model is then analyzed as to how well it describes the observed data. In this paper, the term model refers to the structural and distributional description of the data. The nominal distributional model is the Gaussian distribution; the structural models are parametric models such as an autoregression. This paper examines model-critical procedures which scrutinize the data and the assumed model by varying the way the data is processed during model fitting. Some aspects of model-critical analysis have been presented in [11, 12, 13, 14, 15, 16].

Model-critical analysis is based on the generalized likelihood [12] for a random sample $y_1, y_2, \ldots y_n$

$$L_c(\theta) = (1/c) \sum_{i=1}^{n} \left[f^c(y_i;\theta)/Q^a(\theta) \quad -1 \right] \tag{1}$$

where

$a = c/(1+c)$, $f(y_i;\theta)$ is the assumed probability density of the data evaluated at observation y_i, θ is a parameter vector that characterizes the data and the density f,

$$Q(\theta) = \int_{R_p} f^{1+c}(y;\theta)\, dy \tag{2}$$

is the information generating function for $f(y;\theta)$ [5, 12], and c is the model-critical parameter. The model-critical estimate for θ is the value of $\theta(c)$ which maximizes (1). The estimate $\theta(c)$ is a robust estimate for θ with the degree of robustness controlled by c. Using L'Hospitals rule, the limit of $L_c(\theta)$ as c approaches zero is $L_0(\theta)$ the usual log likelihood for θ. Differentiating $L_c(\theta)$ with respect to θ and setting the result equal to zero yields

$$\frac{\partial L_c(\theta)}{\partial\theta} = 0 = \tag{3}$$

$$\sum_{i=1}^{n} \frac{f^c(y_i;\theta)}{Q^a(\theta)} \left[\frac{\partial \log f(y_i;\theta)}{\partial\theta} - \frac{1}{(1+c)} \frac{\partial \log Q(\theta)}{\partial\theta} \right]$$

which is a necessary condition that must be satisfied by $\theta(c)$. For the models discussed here, equation (3) is used to obtain $\theta(c)$. From (3) it can be seen that each term in the sum is weighted by the c-th power of the assumed probability density of the data evaluated at y_i. The value of c determines the amount and type of weighting used in (3). Positive values of c downweight outlying observations and negative values of c downweight inlying observations. This weighting produces a criticism of the data and the assumed model $f(y;\theta)$ that can indicate if any model assumptions have been violated.

Outliers are an example of a violation of the model assumptions on the data. It is noted that an observation is an outlier only with respect to the assumed underlying model; if a different model is used, the observation may no longer be an outlier. Unlike unstructured data, where an outlier "sticks out", the structure of a model can hide the outlier. If multiple outliers are present, they can compensate each other [2]. In time series where the observations are not independent, outliers need only be large with respect to the error process to seriously affect the parameter estimates [8]. With outliers of this magnitude, they may not show up in plots of the data. Fox [4] considers two outlier models for time series. The additive outlier is a gross error at a single observation. The innovations outlier is a large value in the error process due to a heavy tailed error distribution. It is noted that both types of outliers can occur with independent as well as dependent observations. With dependent observations, the innovative outlier will affect subsequent observations due to the correlation between observations. A host of robust or resistant procedures are available to reduce the effect of outliers on the parameter estimates [1,10]. The previous discussion has focused on outlying contamination; however, inlying or short-tailed contamination can also be a concern [6].

From the above discussion, it can be seen that robustness and goodness of fit are related. Model-critical analysis uses this relationship to examine models for a given set of data. The analysis compares maximum likelihood and robust parameter estimates. Clearly, the robust and maximum likelihood estimates must estimate the same quantity if they are to be comparable. From the derivation of $L_c(\theta)$, the estimate $\theta(c)$ is a consistent estimate of θ [3]. Thus, $\theta(c)$ and $\theta(o)$ are two consistent estimates of θ. If the data and the assumed model are internally consistent, then $\Theta(o)$ and $\Theta(c)$ should be approximately equal over a range of c values. However, if the data and the assumed model are not consistent, then $\Theta(c)$ will change considerably as c increases. Large changes in parameter estimates $\Theta(c)$ indicate that the model requires closer examination. As an M-estimator [7], robust weights are obtained as part of the estimation process. For $c \neq 0$, these critical weights can be used to flag questionable observations. Examination of the weights aids the analyst in evaluating the model. For example, small weights indicate outlying contamination when $c > 0$.

MODEL-CRITICAL ESTIMATION

Let the density $f(y;\Theta)$ in (1) be the r-variate Gaussian density

$$f(y) = \left|2\pi D\right|^{-1/2} \exp[-(y-m)^T D^{-1}(y-m)/2] \quad (4)$$

with mean vector m and covariance matrix D. For the Gaussian density in (4), it is straight forward to evaluate (2) and obtain

$$Q(m,D,c) = [\,|2\pi D|^{c}(1+c)^{r}]^{-1/2}, \; c > -1. \quad (5)$$

Using (3), (4) and (5), the following set of implicit estimation equations for m and D are obtained.

$$m = \sum_{i=1}^{n} w_i y_i \quad (6)$$

$$D = (1+c)\sum_{i=1}^{n} w_i (y_i - m)(y_i - m)^T \quad (7)$$

where

$$w_k = f^c(y_k)/\sum_{i=1}^{n} f^c(y_i) \quad (8)$$

The estimates for m and D are the values m(c) and D(c) which satisfy (6) and (7). Clearly, when $c = 0$, m(0) and D(0) are the usual maximum likelihood estimates. The specification of the user-provided constant c is based on sample size n, dimension r, and the character of the sample.

For models with structure in addition to a mean vector and covariance matrix, the errors can be expressed as

$$\epsilon_i = y_i - h(x_i;\Theta) \quad (9)$$

where

ϵ_i is a r x 1 vector of errors,
y_i is a r x 1 vector of observations,
x_i is a p x 1 vector of concomitant variables, and
Θ is a p x 1 vector of parameters.

The errors ϵ_i are assumed to be independent and identically distributed Gaussian random variables with zero mean and covariance matrix D. For the model given by (9), the generalized likelihood without the constant term denoted L(c) is

$$L(c) = (1/c)\sum_{i=1}^{n} A\exp(-c\, e_i D^{-1} e_i/2) \quad (10)$$

where

$$A = |(1+c)/2\pi D|^{\,0.5c/(1+c)},$$

$$e_i = y_i - h(x_i;\Theta).$$

For the proposed model $h(x;\Theta)$ in L(c), the model-critical estimates of Θ and D are obtained by maximizing (10) over Θ and D. For many models, setting equal to zero the derivatives of L(c) with respect to Θ and D yields a set of implicit equations which can be solved via a fixed point algorithm.

ANALYSIS OF AUTOREGRESSIVE MODELS

For the multivariate autoregressive (AR) model of order p, (9) can be expressed as

$$\epsilon_i = y_i - \sum_{k=1}^{p} A_k\, y_{i-k} \quad (11)$$

for $i = p+1, p+2, \ldots, n$. The model-critical estimates for D and A_k, $k=1, 2, \ldots, p$ are obtained by differentiating L(c) with respect to D and A_k, $k=1, 2, \ldots, p$; setting the derivatives equal to zero; and solving for D and A_k, $k=1, 2, \ldots, p$. The model-critical estimation equations are

$$A(c) = G^{-1}b \quad (12)$$

and

$$D(c) = ((1+c)/w.)\sum_{i=p+1}^{n} w_i\, e_i(c)\, e_i(c)^T \quad (13)$$

where

$$A(c) = [A_1(c), A_2(c), \ldots, A_p(c)]^T, \text{ and} \quad (14)$$

$$e_i(c) = (y_i - \sum_{k=1}^{p} A_k(c)\, y_{i-k}). \quad (15)$$

The block matrix entries of G are

$$g_{ij} = \sum_{k=p+1}^{n} w_k y_{k-i} y_{k-j}^T, \quad i, j = 1,2, \ldots, p$$

and the block matrix entries of b are

$$b_i = \sum_{k=p+1}^{n} w_k y_k y_{k-i}^T , \quad i = 1,2, \ldots, p$$

where

$$w_k = \exp(- c\, e_k(c)^T D(c)^{-1} e_k(c)/2) \qquad (16)$$

and

$$w. = \sum_{k=p+1}^{n} w_k .$$

The estimates A(0), D(0) and A(c), D(c), $c \neq 0$ are conditional maximum likelihood and conditional model-critical estimates since they are conditioned on the first p observations; however, the estimates will still be referred to as maximum likelihood or model-critical estimates.

As an illustration, model-critical estimates are presented for a simulated univariate, Gaussian AR(4) process with representation given by (11) where $A_1 = 2.0625$, $A_2 = -2.4325$, $A_3 = 1.5845$, $A_4 = -0.652$, and the ϵ_t are independent identically distributed with zero mean and unit variance. Figure 1 is a plot of the realization used to the obtain parameter estimates. Table 1 contains the parameter estimates for c = 0, 0.1, 0.2, 0.3, and 0.4. It can be seen that the model-critical and maximum likelihood estimates are approximately equal. Next, additive outliers were added to four observations selected at random in the realization shown in Figure 1. The outliers are independent, identically distributed Gaussian random variables with zero mean and variance 2; also, the outliers are independent of y_t. The additive outlier model [9] is given by $z_t = y_t + v_t$ where y_t is the AR(4) process and v_t is the outlier. For this example, $v_t \neq 0$ for only four observations. Figure 2 is a plot of the data in Figure 1 with outliers added to four observations; without a priori knowledge, one would not suspect that outliers are present. Martin [9] notes that for time series, the outliers need only be large relative to the innovations process ϵ_t to seriously affect the parameter estimates. With outliers of this magnitude, they may not stand out as they do when the observations are independent. This is clearly seen in Table 2, which presents the maximum likelihood and model-critical parameter estimates for the data in Figure 2. As c increases, the model-critical estimates approach the true values; the improvement in the estimates follows from the downweighting of the outliers in model-critical estimation. For the data with outliers, Figure 3 contains plots of the model-critical and maximum likelihood spectrum. The critical spectrum is closer to the true spectrum (the solid line) than the maximum likelihood spectrum estimate. For the data with outliers, the maximum likelihood spectrum contains more high frequency components than the critical spectrum. This is the case because maximum likelihood estimation fits all the data, whereas critical estimation fits the bulk of the data without outliers.

For c = 0.4, Figure 4 is a plot of the model-critical weights

$$w_t = \exp(- c(z_t - \sum_{k=1}^{p} a_k(c) z_{t-k})^2 / 2s^2(c))$$

where $s^2(c)$ and $a_k(c)$ are calculated using (12) and (13). Since the critical weights are a measure of fit between the data and the model, analysis of the weights is an integral part of the modeling process. In Figure 4, the small weights at observations 5, 6, 11, and 25 indicate that the model which describes the bulk of the data does not give a good fit to these observations. In fact, some of the subsequent weights are small due to the dependence between observations. Small weights alert the experimenter to the fact that further analysis of the data and model may be necessary.

CONCLUSION

Model-critical procedures have been presented for analyzing the joint character of the data and the assumed model. The critical estimation equations for multivariate autoregressive models were presented. The model-critical weights which are obtained during parameter estimation can be used to identify observations which violate the model assumptions. It is noted that model-critical procedures can be applied to wide variety of models. These procedures can assist the engineer in model identification, and analysis of the error process.

REFERENCES

[1] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey, Robust Estimators of Location, Princeton, NJ: Princeton University Press, 1972.

[2] V. Barnett, and T. Lewis, Outliers in Statistical Data, New York, NY:John Wiley and Sons, 1978.

[3] T. A. Delehanty, "Sensitivity Analysis in Statistics Based on Self-Critical Estimation," Unpublished Ph.D. Thesis, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY, 1983.

[4] A. J. Fox, "Outliers in Time Series," J. R. Statist. Soc. B, 34, pp. 350-363, 1972.

[5] S. W. Golomb, "The Information Generating Function of a Probability Distribution," IEEE Transactions on Information Theory, pp. 75-77, 1966.

[6] R. V. Hogg, "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory," J. Amer. Statist. Assoc., 69, pp. 909-927, 1974.

[7] P. J. Huber, Robust Statistics, New York:John Wiley and Sons, 1981.

[8] B. Kleiner, R. D. Martin, and D. J. Thomson, "Robust Estimation of Power Spectra," J. R. Statist. Soc. B, 41, pp. 313-351, 1979.

[9] R. D. Martin, "Robust Methods for Time Series," in Findley, D. F., (ed.), Applied Time Series Analysis II, New York:Academic Press, 1981.

[10] R. D. Martin, and P. J. Thomson, Robust-Resistant Spectrum Estimation," *Proceeding of the IEEE*, 70, pp. 1097-1115, 1982.

[11] A. S. Paulson, and N. J. Delaney, "Self-Critical, and Robust, Procedures for the Analysis of Multivariate Normal Data," Unpublished Technical Report, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY, 1981.

[12] A. S. Paulson, and T. A. Delehanty, "Sensitivity Analysis in Experimental Design," Computer Science and Statistics: Proceedings of the 14th Symposium on the Interface," pp. 52-57, 1983.

[13] A. S. Paulson, M. A. Presser, and C. E. Lawrence, "Self-Critical, and Robust Procedures for the Analysis of Binary Data," *Technical* Report 37-81-P1, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY, 1981.

[14] A. S. Paulson, M. A. Presser, and E. H. Nicklin, "Self-Critical and Robust Procedures for the Analysis of Univariate Complete Data," Unpublished Technical Report, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY, 1982.

[15] A. S. Paulson, and G. R. Swope, "Autoregressive Order Selection via Model-Critical Methods," IEEE Journal of Oceanic Engineering, OE-12, pp. 75-79, 1987.

[16] M. A. Presser, "Self-Critical and Robust Estimation Procedures for a Wide Class of Probability Models," Unpublished Ph.D. Thesis, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY, 1980.

TABLE 1

Maximum Likelihood and Model-Critical Parameter Estimates for a Simulated Univariate AR(4) Process with Sample Size = 100

c	a_1	a_2	a_3	a_4	s^2
0.0	2.1346	-2.5685	1.7113	-0.7092	0.9100
0.1	2.1190	-2.5438	1.6906	-0.7094	0.9137
0.2	2.0984	-2.5040	1.6527	-0.6981	0.8915
0.3	2.0841	-2.4820	1.6353	-0.6999	0.8663
0.4	2.0631	-2.4400	1.5962	-0.6914	0.8241

TABLE 2

Maximum Likelihood and Model-Critical Parameter Estimates for a Simulated Univariate AR(4) Process with Four Additive Outliers, and Sample Size = 100

c	a_1	a_2	a_3	a_4	s^2
0.0	1.7661	-1.7643	0.9633	-0.4409	2.0344
0.1	1.8477	-1.9116	1.0799	-0.4651	1.7752
0.2	1.9556	-2.1520	1.2968	-0.5410	1.4180
0.3	2.0342	-2.3064	1.4354	-0.5912	1.1620
0.4	2.0391	-2.3498	1.4765	-0.6102	1.0740

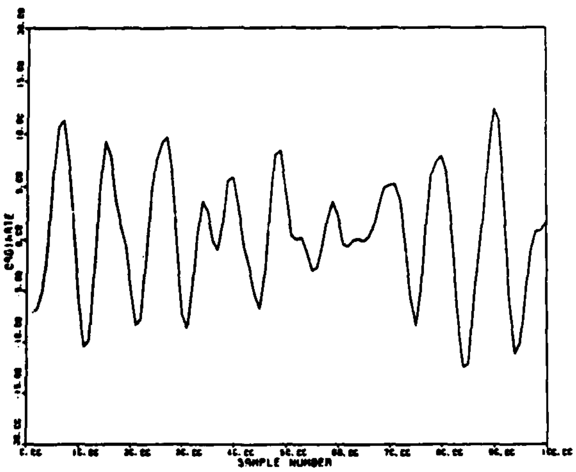

Figure 1. A Simulated AR(4) Process with Innovations Distributed Normal(0,1).

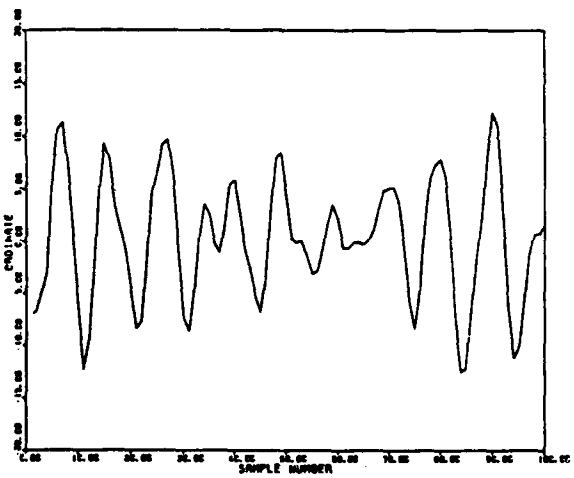

Figure 2. A Simulated AR(4) Process with Innovations Distributed Normal(0,1) and Four Outliers.

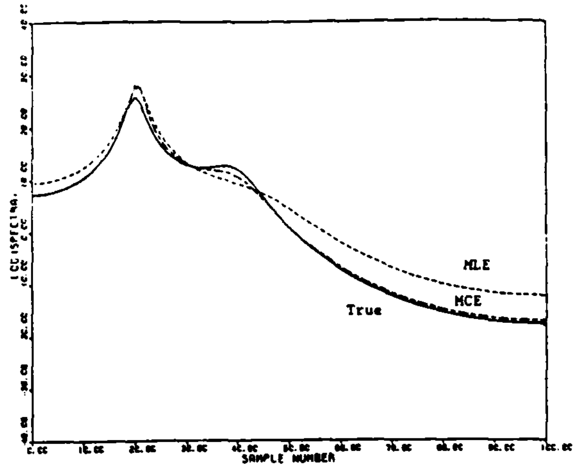

Figure 3. The Maximum Likelihood and Model-Critical (c = 0.4) Spectrum Estimates for the Simulated AR(4) Process with Four Outliers, and the True Spectrum.

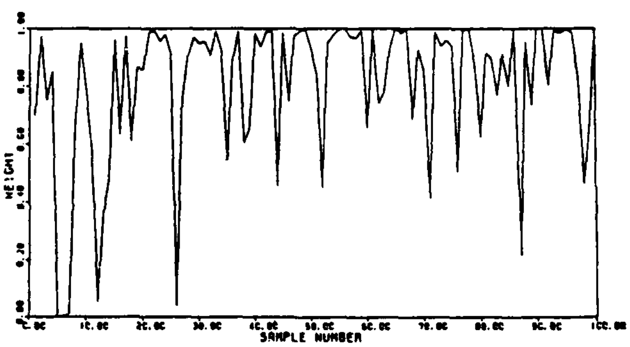

Figure 4. The Model-Critical Weights for the Simulated AR(4) Process with Innovations Distributed Normal(0,1) and Four Outliers; c = 0.4.