BAYESIAN FORECASTING FOR SWITCHING REGRESSION AND AUTOREGRESSIVE PROCESSES

By

MARGARET FOSTER LAND

Bachelor of Science
Northwestern State University of Louisiana
Natchitoches, Louisiana
1963

Master of Science
Northwestern State University of Louisiana
Natchitoches, Louisiana
1967

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
May, 1981

BAYESIAN FORECASTING FOR SWITCHING REGRESSION

AND AUTOREGRESSIVE PROCESSES

Thesis Approved:

Thesis Adviser

Dean of the Graduate College

ACKNOWLEDGMENTS

The author wishes to express her appreciation to her major adviser, Dr. Lyle D. Broemeling, for his patience, guidance, assistance, and encouragement throughout not only this study but also the graduate program. Appreciation also is expressed to the other committee members, Dr. P. Larry Claypool, Dr. William D. Warde, and Dr. Carl R. Anderson, for their encouragement and aid. Ms. Sara Veitl gave valuable assistance in the excellent typing from a very rough copy of the preliminary version of this thesis. Also, the author is grateful to each person on the staff of the Department of Experimental Statistics, New Mexico State University, for encouragement. Also, various graduate students have been of invaluable help, especially Peyton Cook for computer assistance, Linda Willson for encouraging me, and Kate Mellina for being in my corner. Also, thanks go to the staff of TOP Services Unlimited for cheerful help in the face of deadlines.

Special gratitude is reserved for members of my family for their continued supportiveness over the long term. And, a subset of the family, my children, Peter, David, and Stephanie, have endured a role-reversal and lovingly encouraged their mother to complete her school work.

Finally, this study is dedicated to my grandmother, Ona Ruth Heltcel Foster, who shared her dreams, inspired a joy of learning, and gave that of which she herself had so little--opportunity.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The history of mankind is replete with efforts to predict the future, the outcome of the roll of the dice, the outcome of an experiment yet to be performed, the sex of an unborn child. Fortune telling, reading of palms or tea leaves, consultation of horoscopes, and sortilege are among the methods used to predict the future and, hence, to lessen uncertainty and increase control over one's life.

Although prediction is intended to give one a glimpse of the future, various methods have been used to warp that glimpse. For example, David (1955) mentions faked dice for cheating used in Egypt about 300-30 B.C. These dice were used in some form of divination rite, with prediction biased to suit the user.

Statistics also attempts to predict the outcome of future experiments but utilizes mathematical techniques and estimates the goodness of the resulting prediction.

Frequently statistics is concerned with design of experiments, collection of sample data, and subsequently the estimation of parameter values of the population distribution, calculation of confidence intervals, or testing of hypotheses. Nearly all standard statistical texts discuss such topics. However, with the exception of linear regression, few explicitly explore prediction. Yet the point of many statistical analyses is prediction of the outcomes of similar future experiments.

According to Lindley (in the Discussion section of the paper by Harrison and Stevens, 1976) most statistical tasks are concerned with predicting. The implicit assumption is that one can consider what has occurred and project a similar outcome for a future experiment. But no specific consideration is given to the impact of variability of sample data on the precision of that implicit prediction (Aitchison and Dunsmore, 1975). In recent years, increasing interest has come to be focused on Bayesian techniques for solutions to a wide range of statistical problems.

Some of the statistical problems concern the switching regression model for a sequence of random variables. That regression model involves a system which switches among several regression regimes. If exactly one switch occurs, the two-phase regression model is the case at hand. Previous studies of two-phase regression have emphasized estimation of any unknown parameters: the switch point, any of several regression coefficients, or the variance about a regression line.

The model employing a sequence of multivariate random vectors has been explored by Salazar (1980) and the posterior distribution of the change point was derived. This model has been applied (Austin and Heghinian, 1977) to traffic deaths in Illinois and estimations made. Prediction for this model has not been studied in the literature.

Autoregressive time series models are frequently used in applications, for example in econometrics. Prediction has been achieved by several methods, including extrapolation of the model obtained by estimating the parameters in that model. Box and Jenkins (1970) and Kendall (1973) outline some of the more frequently used forecasting techniques. Fuller and Hasza (1981) studied the properties of predictors for both stationary and non-stationary processes. They looked only at

non-Bayesian predictors and hence did not incorporate prior information.

Formal statistical prediction is built upon appropriate design, data collection and analysis, and parameter estimation. Since much of that foundation has been laid for the cases mentioned above, our problem is to proceed to the stage of prediction.

CHAPTER II

BAYESIAN ROUTE TO PREDICTION

An overview of the Bayesian approach to be employed in this study follows so that the details and rationale can be examined in a simpler setting.

Suppose Y is a random sample $\{y_1, y_2, \ldots, y_n\}$ of size n from $N(\mu, \sigma^2)$. Let π_0 denote an appropriate prior density for the unknown parameters. For example, suppose μ is known and σ^2 unknown, $\sigma^2 > 0$. To indicate vague prior knowledge of the value of σ^2, assign

$$\pi_0(\sigma^2) = 1/(\sigma^2).$$

Incidentally, Lindley (1972) commented that σ^{-1} is the only prior density for σ which is invariant under changes of scale and origin. Note that $\pi_0(\sigma^2)$ is an improper prior density, that is,

$$\int_0^\infty \sigma^{-2} d\sigma^2 \text{ diverges.}$$

Also, the term "prior" denotes the density assigned to σ^2 before the sample data are analyzed; that is, the density is prior to the data being collected. Dickey (1976) commented that one of the reasons Bayesian methods were partially discredited and fell into disuse for over a century, well into the 1900's, was a tendency for early writers such as Laplace to select an arbitrary unique prior. The basis for

selecting a uniform prior to express vague information, Bayes Principle of Insufficient Reason, was formally expressed first in the writings of Thomas Bayes (1958) but much earlier a similar idea was advanced by Jacob Bernoulli (Barnett, 1973). Hence vague prior knowledge is represented by a vague or noninformative prior which allows the data to dominate the results. This is especially appropriate in cases where the amount of a priori information is small relative to the amount of information from the experiment (Box and Tiao, 1973). More specific a priori knowledge can be coded appropriately in the prior density. Winkler (1967) examined approaches to eliciting informative priors.

Next, the likelihood function of the sample, denoted $L(Y|\sigma^2)$ where σ^2 is treated as a constant, is calculated. Let $U_n = (\mu, \ldots, \mu)'$ be an nxl vector.

$$L(Y|\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\{[1/(2\sigma^2)](Y-U_n)'(Y-U_n)\}.$$

Then the posterior probability density function (p.d.f.) for σ^2, the density arising posterior to knowledge of the sample data, is obtained by combining the prior density for σ^2 with the likelihood function for the sample data Y. Denote the posterior by $\pi_1(\sigma^2|Y)$.

$$\pi_1(\sigma^2|Y) = \frac{L(Y|\sigma^2)\pi_0(\sigma^2)}{\int_0^\infty L(Y|\sigma^2) \cdot \pi_0(\sigma^2) d\,\sigma^2}, \quad \sigma^2 > 0.$$

The expression on the right is an application of Bayes' Theorem of Inverse Probability or, simply, Bayes' Theorem. Note that the denominator is a normalizing constant with the function of producing a value of $\pi_1(\sigma^2|Y)$ such that,

$$\int_0^\infty \pi_1(\sigma^2|Y)\, d\sigma^2 = 1.$$

Thus $\pi_1(\sigma^2|Y)$ is a proper density function. For notational simplicity, π_1 is frequently expressed by

$$\pi_1(\sigma^2|Y) \propto L(Y|\sigma^2) \cdot \pi_0(\sigma^2)$$

$$\propto [(\sigma^2)^{-(n+2)/2}] \exp \{-1/(2\sigma^2)(Y - U_n)'(Y- U_n)\}.$$

$\pi_1(\sigma^2|Y)$ can be used, for example, to estimate σ^2, calculate a region of highest probability density (HPD), or test a hypothesis about σ^2.

For a more general discussion of Bayesian statistics, prior densities and their selection, and applications of Bayes' Theorem, see Chapter 1 of Box and Tiao (1973) or DeGroot (1970).

The posterior density can be a means for estimation but also is needed for the predictive density. For a future sample of size k $\{y_{n+1}, \ldots, y_{n+k}\}$, for clarity let $Y_{n+i}=W_i$ and then denote $W = \{w_1, w_2, \ldots, w_k\}$. Then likelihood is calculated. The likelihood is the conditional density of the k future values, W, conditioned on σ^2. Using the posterior p.d.f. of σ^2 and this conditional density, the predictive density of W is obtained. Thus, letting $U_k=(\mu, \ldots, \mu)'$ be a kx1 vector,

$$g(W|Y) = \int_0^\infty f(W|\sigma^2) \cdot \pi_1(\sigma^2|Y)d\sigma^2 \propto \int_0^\infty (\sigma^2)^{-(n+k+2)/2}$$

$$x \exp \{-1/(2\sigma^2)[(Y - U_n)'(Y - U_n) + (W - U_k)'(W - U_k)\}d\sigma^2, \ W\epsilon R^k.$$

Let $1/\sigma^2 = r$, then $d\sigma^2 = -r^{-2}dr$, and letting $\sigma^2 \to 0$, we get $r \to \infty$, and as $\sigma^2 \to \infty$, we get $r \to 0$. So,

$$g(W|Y) \propto \int_0^\infty r^{(n+k-2)/2} \exp \{(-1/2)[(Y - U_n)'(Y - U_n) +$$

$$(W - U_k)'(W-U_k)]r\}dr, \ W\epsilon R^k.$$

Then using the fact that this integral is in the form of a gamma density, one obtains

$$g(W|Y) \propto \left[(Y - U_n)'(Y - U_n) + (W - U_n)'(W - U_k)\right]^{-(n+k)/2}, \quad W \epsilon R^k.$$

Thus we have an expression for the prediction density of the k future values. Guttman and Tiao (1964) call this the density of future observations. It is also labeled the predictive density of W given Y. For a general discussion of predictive distributions, see Chapter 2 of Aitchison and Dunsmore (1975).

CHAPTER III

REVIEW OF PREVIOUS WORK

Switching Regression and Time Series

Before proceeding to a review of work on the problem, one must define the problem itself.

Consider an experiment in which data are taken from a population such that

$$y_i \text{ is distributed as } N(\alpha_1 + \beta_1 x_i, \sigma^2),\ i = 1, 2, \ldots, m;$$

$$\text{or distributed as } N(\alpha_2 + \beta_2 x_i, \sigma^2),\ i = m + 1, \ldots, n, \text{ where}$$

$$\alpha_1 \varepsilon R,\ \beta_1 \varepsilon R,\ \alpha_2 \varepsilon R,\ \beta_2 \varepsilon R,\ \sigma^2 > 0, \text{ and } (\alpha_1, \beta_1) \neq (\alpha_2, \beta_2)$$

$$\text{also, } m = 1, 2, \ldots, n - 1.$$

Assume that α_1, β_1, α_2, β_2, m, and σ^2 are a priori independent. This model is a form of two-phase or switching regression; that is, at some point m, $1 \leq m < n$, the y_i's switch linear regression regimes. Various cases may be encountered and considered, with any or all of (α_1, β_1), (α_2, β_2), m and σ^2 unknown. It is assumed that n is known, as is the sample $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. Among the labels that have been applied to m are switch point, change point, and shift point.

Over the past few decades, switching regression problems have been

studied by several methods. The earlier work relied upon classical statistical methods whereas more recent papers have included Bayesian approaches. Several methods have been used for prediction for time series.

Prediction efforts have focused on the simple extrapolation of ordinary least squares for linear regression and more sophisticated methods for time series; more recently, prediction work has included Bayesian methods. None of these has been applied specifically to forecasting for two-phase regression. Furthermore, Bayesian prediction for autoregressive time series processes has not been achieved.

Classical Estimation

The early work with switching sequences of random variables and two-phase regression used non-parametric tests. D. V. Lindley (1952) used what he termed the method of minimum unlikelihood as a solution to the estimation of the change point m. Page (1954) proposed continuous inspection schemes and decision rules to detect a change in a sequence. Then (1955) he used control charts to detect a change in μ and σ^2 for a sequence from a normal population. He used cumulative sums as the basis for the charts. Later (1957) he considered a discrimination approach with restricted sequential procedures.

Quandt (1958) assumed two simple linear regression regimes with one switch and continuity at the switch between the regimes. He used maximum likelihood (ML) to estimate the parameters of the regression regimes. To estimate the switch point he calculated the value of the likelihood function for all possible values of the switch point. Then he took as his estimate of the switch point the value corresponding to

the maximum among the likelihood values. To test the hypothesis that no switch occurred, he suggested a likelihood ratio test. Also, he tested that the two regression regimes are the same by a method employing the sums of squares of regression from the two groups and from a single regression based on the data as a whole. He later (1960) emphasized classical methods to test that the linear regression followed two regimes and suggested an F-test.

Robison (1964) used a maximum likelihood estimate for the point of intersection of two polynomial regressions. He also derived a confidence interval for that point when the time of the switch was known. He encountered practical difficulties in the implementation of his estimation methods. He commented that sometimes the a priori information is not used but it is much better than the information received from the sample (p. 219). However, he made no attempt to use Bayesian methods and incorporate the a priori information. Hudson (1966) proposed a method for calculating the maximum likelihood estimate of the abscissa of the point of intersection when the switch point was unknown. Hinkley (1969) derived an asymptotic distribution for the abscissa of the point of intersection and produced an improved approximation for moderate size samples. Later (1971a) he worked with the application of the estimation procedures, especially on the intercepts and point of intersection in two-phase regression. Also (1971b) he tried a cumsum scheme, that is, one using cumulative sums, for detecting the change point in a sequence of random variables, not for two-phase regression.

Farley and Hinich (1970) were interested in the problem of a shifting slope coefficient in a linear time series, where they assumed at most one shift. When the shift point (if any) was known, they compared

the pooled regression squared residuals to weighted squared residuals of the assumed pre- and post-shift regression and tested to see if a shift had occurred. Under the assumption of shift point unknown, they calculated the regression for the observations by groups and used a Chi-square test on the likelihood of a switch. When a shift was indicated, they used Quandt's likelihood ratio to estimate the location. Although they did not use Bayesian methods, they did assume no prior knowledge and that probability of shift was equal at each of the possible points. Brown, Durbin, and Evans (1975) used the cumulative sums (cumsum) of least-squares residuals to test for constancy in regression relationships. This cumsum technique was mathematically intractable and thus they made a transformation to recursive residuals and standardized cumsums of $N(0, \sigma^2)$ variables. Regression coefficients and their stability were examined by two methods of Garbade (1977). His time series oriented paper emphasized variable parameters regression (VPR) in the linear regression model and the cumsum of squares test of Brown, Durbin, and Evans. Garbade used both methods on the problem of detecting a change and estimating the variation of the coefficients.

Bayesian Estimation

One of the earlier papers using a Bayesian approach was Guttmann and Tiao (1964) which considered a sequence of random variables distributed as either normal or exponential random variables. They used vague priors with $\pi(\sigma) \propto 1/\sigma$ and $\pi(\mu) \propto k_1$, a constant. Their work achieved the estimation of a tolerance interval for the population.

A tracking problem was motivation for the study of a sequence of normally distributed random variables where the mean of the distribution

is subject to a change. Chernoff and Zacks (1964) used a Bayesian viewpoint to obtain an estimation of the second mean and also obtained a minimum variance linear unbiased estimator as well as a simpler "at most one change" (AMOC) Bayes estimator. They then considered an ad hoc procedure, a combination of AMOC and a sequence of tests to locate the point of the last change in mean. This approach is similar to one used by Page but the latter did not involve Bayesian aspects. A paper by Kander and Zacks (1966) continued the work of Chernoff and Zacks by looking at the properties of the test statistic they had proposed, a weighted cumsum value. Kander and Zacks considered in particular the case of any positive upward change in the mean of the sequence of $N(\mu_i, \sigma^2)$ random variables and also generalized to the one parameter exponential family. They also used a Bayesian approach with a vague prior for the time of change,

$$\pi_0(M) = 1/(n-1),\ m = 1, 2, \ldots, n-1,$$

$$= 0 \text{ otherwise.}$$

Bacon and Watts (1971) tackled the same problems as Hinkley, Robison, and Quandt but used Bayesian procedures rather than maximum likelihood estimation to estimate the parameters of the model and the join point. Their Bayesian method led to a bimodal posterior density for the join point and transition parameter. They assumed a smooth transition from one regression regime to the next. Broemeling (1972) used Bayesian procedures to detect the change in distribution parameters in a sequence of random variables. Also (1974) he looked at an observed sequence of independent random variables of the regular exponential class and considered Bayesian procedures to estimate the time point at

which a parameter change occurred. He also generalized the work to the two-phase regression problem. Holbert (1973) assumed exactly one switch and used improper priors and Bayesian methods to estimate the switch point. For a sequence of normally distributed random variables, the other parameters were estimated. Extended to the case of two-phase regression, the work produced an estimate of the shift point, the abscissa of the point of intersection, and the posterior distributions of the regression parameters. Chin Choy (1977) used natural conjugate priors and the general linear model to obtain results analogous to Holbert's. She gave point estimates for the unknown parameters in the two regression regimes and gave a point estimate as well as made inferences about the abscissa of the point of intersection of the regimes.

Swamy and Mehta (1975) discussed a Bayesian solution and, to simplify quite lengthy calculations, proposed an approximate Bayesian solution for estimation. The switching regression problem where the number of changes is known but not necessarily two was the subject of work by Ferriera (1975). The change points were assumed unknown and Bayesian estimates for these points were given for each of three different priors and also for maximum likelihood. Also Holbert and Broemeling (1977) discuss Bayesian results for inference and estimation with shifting sequences and especially for two-phase regression. These results are an aspect of earlier research (Holbert, 1973).

Bayesian Prediction

The preceding papers deal with estimation. An early paper dealing with Bayesian prediction was Aitchison (1964) which discussed the relationship of guaranteed coverage predictors to a Bayesian form of

interval prediction. With a linear regression model Kakwani (1965) used Bayesian forecasting to incorporate prior information on the regression coefficients into the forecasting procedures. He also gave the sampling error of the resulting forecast value. Comparison of the variance of the forecast using no prior information with that incorporating a prior showed a gain in efficiency for the Bayesian forecast.

Forecasting based on Bayesian principles but applied to time series was the subject of a paper by Harrison and Stevens (1971). They looked at changes in slope and trend over a short term. Lindley's (1972) review of Bayesian statistics included a discussion of tolerance regions and predictive densities. Green and Harrison (1973) applied Bayesian forecasting to the problem of a mail order company. This work involved time series and included estimates for the number of returned items.

Bayesian prediction analysis was treated thoroughly in a book by Aitchison and Dunsmore (1975). They gave examples, lists of conjugate priors, posteriors, and predictive densities. They did not treat two-phase regression, nor even simple linear regression as such, but the discussions are illuminating. Brown and Payne (1975) discussed a modified ridge regression approach to election night forecasting and used a vague prior for the intercept parameter. They used Bayesian methods for forecasting but did not use a linear regression model.

Statistical Forecasting is the title and subject of a book by Gilchrist (1976). He made the point that

> . . . common error in use of forecasts is to treat the forecast in the same way as one would treat the true future observation, if only one had a crystal ball to see it. However, . . . number out of a crystal ball is a known constant . . . the number one obtains in statistical forecasting is essentially the value taken by a random value . . . distinction between constant and random variable (p. 308).

A significant and comprehensive extension of past work was presented by Harrison and Stevens in the report "Bayesian Forecasting" (1976). They proposed a dynamic linear model (DLM) which uses Kalman filter recurrence relations and a Bayesian approach. Linear regression, exponential smoothing, and linear time series models as well as the static case are special cases of the DLM. As Chatfield (1975, p. 231) commented, ". . . Adaptive Forecasting could be a better description of the method. . . . subjective information can be incorporated as one goes along as well as in the priors."

Broemeling (1977) extended his previous work with Bayesian estimation and sequences of random variables to the case of forecasting when exactly one change occurs at an unknown point in the sequence. He analyzed cases for normal, exponential, and binomial sequences. Bibby and Toutenburg (1977) produced a comprehensive view of recent work in prediction and estimation in linear models, especially that which had appeared in the German Democratic Republic and USSR. Thus results which had not been available in English were summarized and related to work published in English. Although Bibby and Toutenburg's volume is not Bayesian, in it they made references to incorporating prior information.

CHAPTER IV

TWO-PHASE REGRESSION PREDICTIVE DENSITIES FROM VAGUE PRIORS

Overview

Suppose that sample data, denoted $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, are from two regression regimes. Suppose also that

$$y_i \sim N(\alpha_1 + \alpha_2 x_i, \tau) \text{ for } i = 1, 2, \ldots, m,$$

$$y_i \sim N(\alpha_3 + \alpha_4 x_i, \tau) \text{ for } i = m + 1, m + 2, \ldots, n;$$

$$\text{and } m = 1, 2, \ldots, n - 1, (\alpha_1, \alpha_2) \neq (\alpha_3, \alpha_4).$$

Suppose $\tau > 0$, $-\infty < \alpha_i < \infty$, $i = 1, 2, 3, 4$; and m is the switch point of the system. Assume $1 \leq m < n$, that is, that the switch has occurred at one of the first n-1 observed values. Also assume α_1, α_2, α_3, α_4, M, and τ are a priori independent. It is desired to predict, using prior knowledge and the sample data, the next k values of the system, $\{(x_{n+1}, y_{n+1}), \ldots (x_{n+k}, y_{n+k})\}$, denoted for clarity as $\{(v_1, w_1), (v_2, w_2), \ldots, (v_k, w_k)\}'$. Let $V = [(1, \ldots, 1)', (v_1, \ldots, v_k)']$ and $W = (w_1, w_2, \ldots w_k)'$. We assume V is known. Also let $X_1 = [(1, \ldots, 1)', (x_1, \ldots, x_m)]$, $X_2 = [(1, \ldots, 1)', (x_{m+1}, \ldots, x_n)']$, $Y_1 = (y_1, \ldots, y_m)$, $Y_2 = (y_{m+1}, \ldots, y_n)'$.

One can consider various cases, some using a vague or non-informative prior distribution as well as others using a conjugate prior

distribution, where some or all of the parameters α_1, α_2, α_3, α_4, M, and τ are unknown. Let $\beta_1 = (\alpha_1, \alpha_2)'$ and $\beta_2 = (\alpha_3, \alpha_4)'$. Note that if β_1, β_2, and τ are known and M is unknown then one can obtain the prediction equation when $M \varepsilon \{1, 2, \ldots, n-1\}$. If τ is unknown then one must estimate τ. In this case, if M is unknown, one can consider $M \varepsilon \{2, 3, \ldots, n-2\}$, since the posterior distribution of M does not exist at M=1, n-1, n for $\pi_0(\tau) = \tau$. If τ is known, one can let $\tau = 1$ without loss of generality. This simplification will be followed.

There are 16 cases possible, differing in which of the parameters are unknown. It can be shown that for cases involving known precision, the resulting predictive density involves the k-dimensional normal density. For cases with unknown precision the predictive density involves a k-dimensional t-density. Recall that k represents the number of steps ahead for which a prediction is to be obtained. Also, cases involving unknown M result in a form with summation over M. Hence, if M and the precision are unknown, the resulting predictive density is a mixture of k-dimensional t's. And, if the precision is known and M unknown, the result is a mixture of multivariate normals. The general case with τ known will be examined in some detail and then the corresponding case with the precision unknown.

Precision Known, General Case

Suppose β_1, β_2, and M are unknown. The precision τ is known and without loss of generality, let $\tau = 1$. Also, $-\infty < \alpha_i < \infty$, $M \varepsilon \{1, 2, \ldots, n-1\}$, $i = 1, 2, 3, 4$. Assume that α_1, α_2, α_3, α_4, and M are a priori independent with prior densities

$$\pi_0(\beta_1, \beta_2) \propto \text{constant}$$

$$\pi_0(M) = \frac{1}{n-1}, \quad M\varepsilon\{1, 2, \ldots, n-1\}.$$

The likelihood function for the sample S, given the unknown parameters, is

$$L(S|\beta_1, \beta_2, M) = (2\pi)^{(-n/2)} \exp\{(-1/2)[(Y_1 - X_1\beta_1)'(Y_1 - X_1\beta_1) + (Y_2 - X_2\beta_2)'(Y_2 - X_2\beta_2)]\}, \quad \beta_1\varepsilon R^2, \ \beta_2\varepsilon R^2.$$

The joint posterior p.d.f. of the unknown parameters is proportional to the product of π_0 and L. The density of the future values W, conditional on the parameters, is normal and the predictive density of W is proportional to the product of the joint posterior and the conditional density. After summation over the possible values of M and integration with respect to β_1 and β_2, one obtains as the predictive density of W

$$g(W|S) \propto \sum_{m=1}^{n-1} K_m^* \cdot N(\mu^*, \tau^*), \quad W\varepsilon R^k,$$

where

$$A = X_2'X_2 + V'V$$

$$\tau^* = I - VA^{-1}V'$$

$$\mu^* = (\tau^*)^{-1} VA^{-1} X_2' Y_2$$

and the normal density $N(\mu^*, \tau^*)$ is k-dimensional with mean μ^* and precision τ^* as above. Also, the mixing coefficients K_m^* are

$$K_m^* = |X_1'X_1A\tau^*|^{-(1/2)} \exp\{(-1/2)(Y_1'Y_1 + Y_2'Y_2)\}$$

$$\times \exp\{(-1/2)\mathrm{tr}[Y_2'X_2A^{-1}X_2'Y_2 - (X_1'X_1)^{-1}X_1'Y_1[(X_1'Y_1)^{-1}X_1'Y_1]'$$

$$-VA^{-1}X_2'Y_2B^{-1}(VA^{-1}X_2'Y_2)'\}$$

Thus, we have proved the following theorem.

Theorem 4.1. Given sample data from regression regimes,

$$y_i \sim N(\alpha_1 + \alpha_2 x_i, \tau),\ i = 1, 2, \ldots, m,$$

$$y_i \sim N(\alpha_3 + \alpha_4 x_i, \tau),\ i = m+1, \ldots, n.$$

Suppose β_1, β_2, and M are unknown but τ is known. Assign a vague prior to β_1 and β_2 and assign $\pi_0(M)=1/(n-1)$. Then the predictive density of W, the next k values, is

$$g(W|S) \propto \sum_{m=1}^{n-1} K_m^* N(\mu^*, \tau^*),\ W\epsilon R^k,$$

with K_m^*, μ^*, and τ^* as defined above.

Corollary 4.1. For k=1, the predictive density in Theorem 1.1 is

$$g(w_1) \propto \sum_{m=1}^{n-1} K_m^* N(w_1;\ \mu^*, \tau^*),\ w_1\epsilon R,$$

where, letting $v=x_{m+1}$ and $w_1=y_{n+1}$,

$$A = X_2'X_2 + v^2$$

and the precision is

$$\tau^* = 1 - v^2 A^{-1}$$

$$\mu^* = (\tau^*)^{-1}\, vA^{-1}X_2'Y_2$$

and the mixing coefficients are

$$K_m^* = |X_1'X_1 A\tau^*|^{(-1/2)} \exp\{(-1/2)(Y_1'Y_1 + Y_2'Y_2)\}$$

$$\times \exp\{(-1/2)\mathrm{tr}[Y_2'X_2A^{-1}X_2'Y_2 - (X_1'X_1)^{-1}X_1'Y_1[(X_1'X_1)^{-1}(X_1'Y_1)]'$$

$$-vA^{-1}X_2'Y_2(\tau^*)^{-1}(vA^{-1}X_2'Y_2)']\}.$$

This case is used as an example and computer programs utilized to obtain the mode, mean, higher moments, and variance of the predictive density as well as a highest forecast density (HFD) region of content $1-\alpha$, where HFD region is defined analogously to the definition of HPD region given by Box and Tiao (1965, p. 1469).

Letting R denote the region,

$$\text{(i) } \Pr(\theta \varepsilon R|S) = 1 - \alpha \text{ and}$$

$$\text{(ii) for } \theta_1 \varepsilon R \text{ and } \theta_2 \varepsilon R,\ P(\theta_1|S) \geq P(\theta_2|S).$$

This is the $1-\alpha$ region of highest predictive (or forecast) density, that is, the one of smallest volume with content $1-\alpha$.

For the example, k=1 is used. For k greater than one, multiple integration techniques would be required and computer costs would escalate. Also the results could not so easily be compared with known data sets.

Numerical Example

The data of Quandt (1958), used by Holbert (1973) and Chin Choy

(1977) are used in this example. The data are given in Table I of Appendix A. These data were generated by the model

$$y_i = 2.5 + 0.7\, x_i + e_i,\ i = 1, \ldots, 12$$

and

$$y_i = 5 + 0.5\, x_i + e_i,\ i = 13, \ldots, 20,$$

where the e_i's are i.i.d. N(0, 1).

Computer subroutines MODE and HPD developed by Cook (1980) were used with the function for the predictive density. The results are as follows:

Point Estimates

mode	11.6200
median	11.7170
mean	11.7242
variance	131.2489
std. dev.	11.4564

Moments

n	$E(w^n)$	$E[w - E(w)]^n$
1	11.7242	0
2	268.7066	131.2489
3	6233.4971	5.5316
4	179094.9370	51693.8639

H.F.D. Regions

90%	-4.7152, 28.1502
95%	-6.0436, 29.4837
98%	-6.8917, 30.3353
99%	-7.1538, 30.5984

Next the most general case, that is, all parameters unknown, will be examined in some detail.

Precision Unknown, General Case

Consider the most general case in two-phase regression. β_1, β_2,

M, and τ are unknown. Assume a priori that they are independent and assign priors as follows:

$$\pi_0(M) = 1/(n-3),\ M = 2, 3, \ldots, n-2;$$

$$\pi_0(\beta_1, \beta_2) \propto C,\ C > 0,\ \beta_i \epsilon R^2,\ i = 1, 2;$$

$$\pi_0(\sigma^2) \propto \tau,\ \tau > 0.$$

The likelihood function for S is

$$L(S|\beta_1, \beta_2, M, \tau) = (2\pi\tau^{-1})^{-(n/2)} \exp\{(-\tau/2)$$

$$[(Y_1 - X_1\beta_1)'(Y_1 - X_1\beta_1) + (Y_2 - X_2\beta_2)'(Y_2 - X_2\beta_2)]\}.$$

The joint posterior p.d.f. is

$$\pi_1(\beta_1, \beta_2, M, \tau) \propto \tau^{(n+2)/2} \exp\{(-\tau/2)[(Y_1 - X_1\beta_1)'(Y_1 - X_1\beta_1) +$$

$(Y_2 - X_2\beta_2)'(Y_2 - X_2\beta_2)]\}$, $(\beta_1, \beta_2)\epsilon R^2$, $\tau > 0$, and $M = 2, 3, \ldots, n-2$.

The conditional density of the k future values W is

$$f(W|\beta_1, \beta_2, M, \tau) = (2\pi\tau^{-1})^{-k/2} \exp\{(-\tau/2)(W - \beta_2 V)'(W - \beta_2 V)\},\ W\epsilon R^k.$$

Combining the posterior p.d.f. and the conditional density, one obtains the predictive density of W as

$$g(W|S) \propto \sum_{m=2}^{n-2} |A|^{-1/2} B^{-(n+k+2)/2} |T^*|^{-1/2} \cdot T(W;\ n + k - 3,\ \mu^*,\ T^*),\ W\epsilon R^k,$$

where

$$A = X_2'X_2 + V'V$$

$$B = I - VA^{-1}V$$

$$F = K - G'B^{-1}G$$

$$K = Y_2'Y_2 - Y_2'X_2A^{-1}X_2'Y_2$$

$$G = VA^{-1}X_2Y_2.$$

Also, $T(W; n+k-3, \mu^*, T^*)$ is a t density with degrees of freedom $n+k-3$, location parameter μ^* and precision T^* where

$$\mu^* = B^{-1}G$$

$$T^* = (n + k - 3)F^{-1}B.$$

Thus the predictive density of W is a mixture of k-dimensional multivariate t densities with degrees of freedom $n+k-3$, location vector μ^*, and precision matrix T^*.

Thus, we have proved the following theorem.

Theorem 4.2. Suppose

$$y_i \sim N(\alpha_1 + \alpha_2 x_i, \tau), \quad i = 2, \ldots, m,$$

and

$$y_i \sim N(\alpha_3 + \alpha_4 x_i, \tau), \quad i = m + 1, \ldots, n, \quad M \in \{2, \ldots, n - 1\}.$$

Suppose β_1, β_2, M and τ are unknown and a priori independent. Assign priors as above. Then the predictive density of W, the next k observations, is

$$g(W|S) \propto \sum_{m=2}^{n-2} |A|^{-1/2} B^{-(n+k+2)/2} |T^*|^{-1/2} \cdot T(W; n + k - 3, \mu^*, T^*), \quad W \in R^k,$$

where the parameters and mixing coefficients are as defined above.

CHAPTER V

PREDICTIVE DENSITIES FOR GENERAL LINEAR MODEL

Conjugate Priors

Complementing the work with vague priors, the predictive density with conjugate priors for the general linear model for two-phase regression will show some similarities. Bayesian methods and, in this case, conjugate priors will be used. Basically, this will take Chin Choy's (1977) work and extend it to prediction.

The general linear model (GLM) of interest is

$$y_i \sim N(x_i'\beta_1, \sigma^2),\ i = 1, 2, \ldots, m$$
$$\text{and} \sim N(x_i'\beta_2, \sigma^2),\ i = m+1, \ldots, n,$$

where $S = x_i$ is a pxl vector of known fixed quantities on p regressors for the i-th observation. Also, β_1 and $\beta_2 \epsilon R^{(2)}$ and $\beta_1 \neq \beta_2$.

Let $Y_1=(y_1, \ldots, y_m)'$ be the mxl vector of observations on the first linear model and let $Y_2=(y_{m+1}, \ldots, y_n)'$ be the (n-m)xl vector of observations on the second linear model. Also, $Y=(Y_1, Y_2)'$ is the nxl vector of all observations.

$$X_1 = \begin{pmatrix} x_1' \\ \cdot \\ \cdot \\ \cdot \\ x_m' \end{pmatrix}$$

is an mxp matrix, and

$$X_2 = \begin{pmatrix} x_{m+1}' \\ \cdot \\ \cdot \\ \cdot \\ x_n' \end{pmatrix}$$

is an (n-m)xp matrix.

Define

$$X = \begin{pmatrix} X_1 & \phi \\ \phi & X_2 \end{pmatrix}$$

where the ϕ's are zero matrices and X is an nx2p matrix.

The objective of this effort is the predictive density of the next k values, denoted by $(v_1, w_1), \ldots, (v_k, w_k)$ where v_i is pxl and known, w_i is lxl for each i. Also, $(v_1', \ldots, v_k')' = V$ and $(w_1, \ldots, w_k)' = W$.

For the general case, assume β_1, β_2, σ^2, and M are unknown and a priori independent and assign prior densities as follow:

M has marginal prior probability mass function (p.m.f.)

$$\pi_0(M) = \frac{1}{n-1}, \quad M = 1, \ldots, n-1.$$

The parameters β_1, β_2, and σ^2 are assigned a joint conjugate prior density. Let $R = 1/\sigma^2$, $\beta = (\beta_1', \beta_2')$. Then the conditional distribution of β when $R = r$, $r > 0$, is a 2p-variate normal distribution with mean vector β_μ and precision matrix $r\tau$ where $\beta_\mu \epsilon R^{2p}$, τ is a given 2p x 2p symmetric positive definite matrix. The marginal distribution of R is a gamma distribution with parameters a and b, $a > 0$, $b > 0$.
If $\pi_0(\beta|r)$ and $\pi_0(r)$ denote the conditional p.d.f. of β when R=r and the marginal p.d.f. of R, respectively, then, according to Chin Choy (1977),

$$\pi_0(\beta|r) = (2\pi)^{-p}|r\tau|^p \exp\,[(-r/2)(\beta - \beta_\mu)'\tau(\beta - \beta_\mu)],\ \beta\epsilon R^{2p},$$

and

$$\pi_0(r) = [(b^a)/\Gamma(a)]r^{a-1}e^{-br}.$$

It can be shown that the joint posterior of β, M, and R is

$$\pi(\beta,\ M,\ R) \propto r^{(2a+2p+n-2)/2}$$

$$x \exp\,\{(-r)[b + (1/2)(\beta - \beta_\mu)'\tau(\beta - \beta_\mu) + (1/2)\ (Y - X\beta)'(Y - X\beta)]\}$$

for $M = 1, 2, \ldots, n - 1$, $\beta\epsilon R^{2p}$, $r > 0$.

One obtains the conditional density of the k future values $f(W|\beta,$ M, r), and uses that and the joint posterior p.d.f. above to derive the predictive density of W. For the case of σ^2 known, and any of the other parameters unknown, one finds the predictive density of W to be a mixture of multivariate normals. For σ^2 unknown, one notes the predictive density is a mixture of multivariate-t densities. Since the latter case with σ^2 unknown and the other parameters also unknown is the most general, its development will be outlined.

From the posterior p.d.f. of β, M, and R, we obtain the joint marginal p.d.f. of β_2 and R. Since the conditional density of W depends on β_2, not β, this marginal p.d.f. is needed. It is found by integration of the joint marginal of β, M, and R over β_1, and summation over M. This marginal p.d.f. of β_2 and R is a mixture of p-variate normals.

$$\pi(\beta_2,\ R) \propto \sum_{m=1}^{n-1} r^{(2a+n-2)/2}|U|^{-p} \exp\,\{(-r)[b + (1/2)(Y'Y +$$

$$\beta_\mu'\tau\beta_\mu - Q'U^{-1}Q)]\}\ x\ N_p(\beta_2;\ \mu_2^*,\ T_2^*),\ \beta_2\epsilon R^p,\ r > 0,$$

where

$U = \tau + X'X$ which is a 2px2p matrix

$Q = \tau\beta_\mu + X'Y$ which is a 2px1 vector

and

$$\mu^* = U^{-1}Q = \begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix} \text{ and } T^* = rU = \begin{pmatrix} rU_1^* \\ rU_2^* \end{pmatrix} = \begin{pmatrix} T_1^* \\ T_2^* \end{pmatrix}.$$

It is clear μ_2^* consists of the (p+1)th through (2p)th element of the 2px1 vector $U^{-1}Q$ and T_2^* is as given. Let $x_{n+j}=v_j$, j=1, ..., k and $V=(v_1, \ldots, v_k)'$ and $y_{n+j}=w_j$, j=1, ..., k, and $W=(w_1, \ldots, w_k)'$. The conditional density of the k future values, W, is

$$f(W|\beta_2, M, R) = (2\pi)^{-k/2}r^{k/2} \exp\{(-r/2)(W - V\beta_2)'(W - V\beta_2)\}, \; W\epsilon R^k.$$

Thus the predictive density of W is

$$g(W|Y) \propto \sum_{m=1}^{n-1} |U^{-2}U_2^*|^{p/2}|AD|^{-1/2}|A_1|^{-(n+p+k+2a-2)/2}$$

$$x\; T_k(W;\, 2a + p + n - 1,\, C,\, P),\; W\epsilon R^k,$$

where T_k is a k-dimensional general t density with 2a+p+n-1 degrees of freedom, location parameter C and precision P=(2a+p+n-1)AD, with

$A = V'V + U_2^*$

$D = I - VA^{-1}V'$

$C = D^{-1}VA^{-1}U_2^*\mu_2^*$

and

$$A_1 = b + (1/2)[Y'Y + \beta_\mu'\tau\beta_\mu - Q^1U^{-1}Q + \mu_2^{*\prime}U_2^*\mu_2^* -$$

$$(U_2^*\mu_2^*)'A^{-1}U_2^*\mu_2^* - (VA^{-1}U_2^*\mu_2^*)'C]$$

Thus we have proved the following theorem.

Theorem 5.1

Given the general linear model for two-phase regression as above and quasi-conjugate priors as specified, all parameters unknown with $R > 0$, $\beta \epsilon R^{2p}$, and $M = 1, 2, \ldots, n-1$, then the predictive density of the next k observations W is a mixture of k-dimensional general t densities, with mixing coefficients and parameters as given above.

CHAPTER VI

PREDICTION FOR AUTOREGRESSIVE TIME SERIES

The recognition that a set of observations may be ordered by time was made many centuries ago. For example, the Romans knew that the apparent motion of the sun was time dependent and used that idea in construction of a calendar. Others also utilized time dependency in studying data. Lagrange in 1777 used rational functions to analyze the periods and thus the orbit of a comet, according to Bloomfield (1976). Although the recognition of time dependent processes has existed for centuries, only in the past forty years have significant strides been made in the investigation of the description and analysis of time series. According to Kendall (1973), the gains have been possible at least in part due to the development of electronic computers, which have made feasible the extensive computations necessary for study of time series. Kendall (1973) also commented that work by Udny Yule in 1927 brought out the non-deterministic nature of processes and this led toward the ideas of stochastic processes. A stochastic process is a set of random variables which have an order in time. Time series work is one aspect of the developments resulting from Yule's work. Chatfield (1975) notes that around 1920 Yule introduced autoregressive-type processes. In fact, second order autoregressive processes are sometimes called Yule autoregressive processes (Kendall, 1973).

An autoregressive process of order m is one in which y_t is

regressed not upon an independent variable but upon the preceding m values of y. Thus, a first order autoregressive process, denoted AR(1), can be written

$$y_t = \alpha_o + \alpha_1 y_{t-1} + u_t,\ t = 1, 2, \ldots, T;$$

and the u_t are i.i.d. $N(0, \tau)$, where τ is the precision, $\tau > 0$,

α_0 and α_1 are real numbers.

Frequently interest centers on the case with zero mean. Differencing, that is, subtracting the mean, is used to obtain the AR process in a form with zero mean, $y_t = \beta y_{t-1} + u_t$; $t = 1, \ldots, T$. The AR(1) process is also known as a Markov process, named for the Russian A. A. Markov who made contributions to the developing theory. Fuller (1976) asserts that the AR(1) model is not only one of the simplest but also one of the most frequently used models in time series. This fact is due in part to its being an appropriate representation in economic models for error time series.

The AR(2) process, with zero mean, is written

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + u_t,\ t = 1, \ldots, T;\ u_t \text{ are i.i.d. } N(0, \tau),$$

$$\beta_1 \varepsilon R,\ \beta_2 \varepsilon R.$$

And, the m-th order autoregressive process AR(m) is

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_m y_{t-m} + u_t,\ t = 1, \ldots, T,\ \beta_i \varepsilon R,$$

and the u_t are i.i.d. $N(0, \tau)$, $\tau > 0$.

It is assumed that the time intervals are equally spaced.

In econometrics as in other fields that use time series, one goal may be prediction of the next k observations of the AR(m) process where k = 1, 2, ..., but clearly less than m + 1 as the observations separated by more than m steps are independent.

General Result, AR(m), k Steps

Consider the AR(m) process already given. Sample data can be expressed

$$y_t - (\beta_1 y_{t-1} + \ldots + \beta_m y_{t-m}) = u_t,\ t = 1, \ldots, T, \text{ and the}$$

$$u_t \text{ are i.i.d. } N(0, \tau),\ \tau > 0.$$

Adoption of convenient notation will make results clearer. Hence, let

$$Y_T = (y_1, \ldots, y_T)',\ \beta = (\beta_1, \ldots, \beta_m)',$$

$$Y = \begin{bmatrix} y_0 & y_{-1} & \cdots & y_{-m+1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ y_{T-1} & y_{T-2} & \cdots & y_{T-m} \end{bmatrix}, \text{ which is a Txm matrix.}$$

And for the future values $(y_{T+1}, \ldots, y_{T+k})$ to be predicted, let $w_j = y_{T+j}$, and $w_0 = y_T$; also let

$$W_k = (w_1, \ldots, w_k)'$$

and

$$W = \begin{bmatrix} w_0 & w_{-1} & \cdots & w_{-m+1} \\ w_1 & w_0 & \cdots & w_{-m+2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ w_{k-1} & w_{k-2} & \cdots & w_{k-m} \end{bmatrix}, \text{ which is a kxm matrix.}$$

Assume initial values $Y_0 = (y_0, y_{-1}, \cdots, y_{-(m-1)})$ are given. Assume the β_i are unknown coefficients. Clearly β_m is assumed non-zero as the AR process is of order m. Also, assume little prior information is available about β and τ, and use a diffuse prior

$$\pi_0(\beta, \tau) \propto 1/\tau, \ \tau > 0, \ \beta \epsilon R^m.$$

The likelihood for Y_T is

$$L(Y_T|\beta, \tau) \propto \tau^{T/2} \exp\{(-\tau/2)(Y_T - Y\beta)'(Y_T - Y\beta)\}, \ y_T \epsilon R^T.$$

The posterior p.d.f. π_1 is thus proportional to the product of π_0 and L. The likelihood of W_k, the k future values, conditional on β, τ and Y, is

$$f(W_k|\beta, \tau, Y) \propto \tau^{k/2} \exp\{(-\tau/2)(W_k - W\beta)'(W_k - W\beta)\}, \ W_k \epsilon R^k.$$

So, the Bayesian predictive density for the future values W_k is the product of π_1 and f, integrated over β and τ. This is

$$g(W_k|Y_T) \propto \int_\tau \int_\beta \tau^{(T+k-2)/2} \exp\,[(-\tau/2)E] d\beta d\tau, \ W_k \epsilon R^k,$$

where

$$E = (Y_T - Y\beta)'(Y_T - Y\beta) + (W_k - W\beta)'(W_k - W\beta).$$

Note that $Y_T - Y\beta$ is a Txl vector, so $(Y_T - Y\beta)'(Y_T - Y\beta)$ is lxl, a scalar. Also $(W_k - W\beta)'(W_k - W\beta)$ is a scalar. Thus E=trE. This facilitates rewriting E in terms of β to effect the integration.

So,

$$E = tr\ E$$

$$= tr[Y_T'Y_T - 2\beta'Y'Y_T + \beta'Y'Y\beta + W_k'W_k - 2\beta'W'W_k + \beta'W'W\beta]$$

$$= \text{tr}\ [Y_T'Y_T + W_k'W_k + Y'Y\beta\beta + W'W\beta\beta' - 2Y'Y_T\beta' - 2W'W_k\beta']$$

$$= \text{tr}[Y_T'Y_T + W_k'W_k + (Y'Y + W'W)\beta\beta' - 2(Y'Y_T + W'W_k)\beta']$$

$$= \text{tr}\{Y_T'Y_T + W_k'W_k + (Y'Y + W'W)[\beta\beta' - 2(Y'Y + W'W)^{-1}$$

$$(Y'Y_T + W'W_k)\beta']\}.$$

Let $A=Y'Y+W'W$. Note that A is an mxm matrix. Then continue

$$E = \text{tr}\{Y_T'Y_T + W_k'W_k - (Y'Y_T + W'W_k)[A^{-1}(Y'Y_T + W'W_k)]' +$$

$$A[\beta\beta' - 2A^{-1}(Y'Y_T + W'W_k)\beta' + A^{-1}(Y'Y_T + W'W_k)[A^{-1}(Y'Y_T + W'W_k)]']\}$$

Let

$$B=Y_T'Y_T + W_k'W_k - (Y_T'Y + W_k'W)A^{-1}(Y'Y_T + W'W_k)$$

and then

$$E = \text{tr}\{B + A[\beta - A^{-1}(Y'Y_T + W'W_k)][\beta - A^{-1}(Y'Y_T + W'W_k)]'\}$$

$$= \text{tr}\{B +[\beta - A^{-1}(Y'Y_T + W'W_k)]'A[\beta - A^{-1}(Y'Y_T + W'W_k)]\}.$$

So, if $W_k \varepsilon R^k$,

$$g(W_k|Y_T) \propto \int_\tau \int_\beta \tau^{(T+k-2)/2} \exp\ [(-1/2)\tau B]$$

$$x \exp\ \{(-\tau/2)[\beta - A^{-1}(Y'Y_T + W'W_k)]'A[\beta - A^{-1}(Y'Y_T + W'W_k)]\}d\beta d\tau.$$

Using the form of the normal density to integrate with respect to β,

$$g(W_k|Y_T) \propto \int_\tau \tau^{(T+k-2)/2}\ |\tau A|^{-1/2} \exp\ [-\tau(A/2)]d\tau,\ W_k \varepsilon R^k.$$

By the gamma density with parameters $a > 0$, $b > 0$, it can be shown

$$g(W_k|Y_T) \propto |A|^{(-1/2)}\ |B|^{-(T+k-1)/2},\ W_k \varepsilon R^k.$$

Hence, the predictive density for $(w_1, w_2, \ldots, w_k)$ is given and the following theorem is proven.

Theorem 6.1

Given initial values $y_0, y_{-1}, \ldots, y_{m+1}$, and vague prior $\pi_0(\beta, \tau) \propto \tau^{-1}$, the next k values of the m-th order autoregressive process given by

$$y_t = \beta_1 y_{t-1} + \ldots + \beta_m y_{t-m} + u_t,$$

$$u_t\text{'s are i.i.d. } N(0, \tau),\ 0 < \tau < \infty,\ t = 1, 2, \ldots, T,$$

have a predictive density

$$g(W_k|Y_T) \propto |A|^{-1/2} |B|^{-(T+k-1)/2},\ W_k \epsilon R^k,$$

where

$$A = Y'Y + W'W$$

$$B = Y_T'Y_T + W_k'W_k - (Y_T'Y + W_k'W)A^{-1}(Y'Y_T + W'W_k).$$

Note that A^{-1} involves $w_1, \ldots, w_{k-1}$, and $W_k'W$ involves terms with products and powers of the w's. It appears not to be feasible to write $g(W_k|Y)$ in the form of a known density when k is more than one. But, that lack affects only the elegance and simplicity of the representation of the predictive density, not its utility. In fact, the function of the density generally will be written in matrix or summation form before computer programs can be implemented to compute the actual values of the mode, mean, higher order moments, the plot of the density, or the highest forecast density (HFD) regions, and obtain usable values for W_k.

AR(m), k = 1

Consider the one-step ahead forecast. Then the following can be

proved.

Corollary 6.1. The predictive density for w_1 is a univariate general t distribution with T-1 degrees of freedom, location parameter μ, and precision τ, where

$$\mu = D^{-1}WA^{-1}Y'Y_T$$

$$\tau = (T-1)C^{-1}D$$

with

$$A = Y'Y + W'W$$

and

$$C = Y_T'Y_T - Y_T'YA^{-1}Y'Y_T - Y_T'YA^{-1}W'D^{-1}WA^{-1}Y'Y_T$$

$$D = 1 - WA^{-1}W'.$$

Proof: For $k=1$, $W_k=w_1$ and $W=(w_0, w_{-1}, \ldots, w_{-m+1})$ and thus $W=(y_T, y_{T-1}, \ldots, y_{T-m+1})$.

Note that

$$A = Y'Y + W'W$$

$$= Y'Y + \sum_{t=T-m+1}^{T} y_t^2.$$

Hence, A has no term involving w_1.

B in Theorem 6.1 can be expressed as

$$B = Y_T'Y_T + w_1^2 - (Y_T'Y + w_1W)A^{-1}(Y'Y_T + w_1W)$$

$$= w_1^2 - 2w_1WA^{-1}Y'Y_T - w_1^2WA^{-1}W' + Y_T'Y_T - Y_TYA^{-1}Y'Y_T$$

$$= w_1^2(1 - WA^{-1}W') - 2w_1 (WA^{-1}Y'Y_T) + (Y_T'Y_T - Y_T'YA^{-1}Y'Y_T)$$

$$= C\{1 + C^{-1}(1 - WA^{-1}W')[w_1 - (1 - WA^{-1}W')^{-1}WA^{-1}Y'Y_T]^2\}$$

where C is as given in the statement of this corollary. Note that A contains only y's and so A can go into the constant of proportionality.

Also, C contains only y's and similarly can be disposed of. Thus, we are left with

$$g(w_1|Y_T) \propto \{1 + C^{-1}(1 - WA^{-1}W)[w_1 - (1 - WA^{-1}\ W')^{-1}WA^{-1}Y'Y_T]2\},\ w_1 \varepsilon R,$$

which is in the form of a general t distribution.

Hence, $g(w_1|y_T)$ is a general t density of the random variable w_1, $w_1 \varepsilon R$, and has degrees of freedom T-1, location parameter μ and precision τ^*,

$$\mu = (1 - WA^{-1}W')^{-1}WA^{-1}Y'Y_T$$

$$\tau^* = (T - 1)C^{-1}(1 - WA^{-1}W').$$

Note that the degrees of freedom depend only on T, the number of observations. Hence, as T increases, the degrees of freedom also increase. Recall that for a t distribution, the variance is

$$(T - 1)/[(T - 3)\tau^*].$$

Now, suppose updating is of interest, that is, w_1 is predicted then observed and used as observation y_{T+1}, so there are T + 1 observations upon which to base the prediction for w_2. This clearly results in w_2 having a predictive density which is univariate t with T degrees of freedom and analogous values for the location parameter and precision. But, suppose the predictive density with its mean is available for w_1 but not the observed value and one needs to find the predictive density for w_2 conditional on w_1. This process is an example of the chain rule of forecasting (Litterman, 1980). Then, k = 2 is needed. Thus, proceed to the next case.

AR(m), k = 2

$$g(W_2) \propto |Y'Y + W'W|^{-1/2} \; B^{-(T+1)/2}, \; W_2 \varepsilon R^2.$$

It is necessary to update the prediction and $g(w_2|w_1)$ is needed. An approach to finding an expression for $g(w_2|w_1)$ is to represent $g(w_1, w_2)$ as the product of two functions, one of which, say f_1, is a function of w_1 only while the other, call it f_2, is a function of both w_1 and w_2. This product form shows that $f_2(w_1, w_2)$ is proportional to $f(w_2|w_1)$, the conditional density of w_2 given w_1. To do this, we let k=2 and use the results of Theorem 6.1.

$$W = \begin{bmatrix} w_0 & w_{-1} & \cdots & w_{-m+1} \\ w_1 & w_0 & \cdots & w_{-m+2} \end{bmatrix} = \begin{bmatrix} y_T & y_{T-1} & \cdots & y_{T-m+1} \\ w_1 & y_T & & y_{T-m+2} \end{bmatrix}$$

Thus, W contains w_1 and y's but not w_2. Since Y is a matrix of y's only, $Y'Y+W'W$ is a function of w_1 and y's.

Let $d = |Y'Y+W'W|$

and

$$(Y'Y + W'W)^{-1} = (1/d) \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix} = (1/d)a.$$

Note that the a_{ij} do not involve w_2. Also, B can be written

$$B = Y_T'Y_T + w_1^2 + w_2^2 - (1/d)[Y_T'YaY'Y_T + 2W_2'WaYY_T' + W_2'WaW'W_2]$$

In order to express B in terms of w_1, w_2, and y's, we examine the expressions in the brackets.

$$W_2'WaY'Y_T = (w_1 \; w_2) \begin{bmatrix} y_T & y_{T-1} & \cdots & y_{T-m+1} \\ w_1 & y_T & \cdots & y_{T-m+2} \end{bmatrix} a \begin{bmatrix} y_0 & \cdots & y_{T-1} & y_1 \\ y_{-1} & \cdots & y_{T-2} & y_2 \\ \cdot & & \cdot & \\ \cdot & & \cdot & \\ y_{-m+1} & & y_{T-m} & y_T \end{bmatrix}$$

$$= (w_1y_T + w_1w_2,\; w_1y_{T-1} + w_2y_T,\; \ldots,\; w_1y_{T-m+1} + w_2y_{T-m+2})\; a \begin{bmatrix} \sum_1^T y_{t-1}y_t \\ \cdot \\ \cdot \\ \cdot \\ \sum_1^T y_{t-m}y_t \end{bmatrix}$$

With further matrix multiplication, it can be shown that

$$W_2'WaY'Y_T = w_1 \sum_{j=1}^{m} \Big[\sum_{i=1}^{m} y_{T-1}\, a_{j,i} \Big(\sum_{t=1}^{T} y_{t-j}\, y_t \Big) \Big]$$

$$+ w_2 \sum_{j=1}^{m} \Big[\sum_{i=1}^{m} y_{T+1-j}\, a_{j,i} \Big(\sum_{t=1}^{T} y_{t-j}\, y_t \Big) \Big].$$

A similar expansion of the matrix products leads to

$$W_2'WaW'W_2 = w_2^2 \Big[\sum_{j=1}^{m} y_{T+2-j} \Big(\sum_{i=1}^{m} y_{(T+2-i)} a_{ij} \Big) \Big]$$

$$+ w_2 \{ w_1 \Big[\sum_{j=1}^{m} y_{(T+2-j)} \Big(\sum_{i=1}^{m} y_{T+1-i} a_{ij} \Big)$$

$$+ \sum_{j=1}^{m} y_{T+1-j} \Big(\sum_{i=1}^{m} y_{T+2-i} a_{ij} \Big) \Big] \}$$

$$+ \{ w_1^2 \Big[\sum_{j=1}^{m} y_{T+1-j} \Big(\sum_{i=1}^{m} y_{T+1} a_{ij} \Big) \Big] \}.$$

Recall B and write it in terms of w_2. It can be shown that

$$B = w_2^2[1 - C/d] - w_2\, D/d + E$$

where

$$C = \sum_{j=1}^{m} y_{T+2-j}\Big(\sum_{i=1}^{m} y_{T+2-i}a_{ij}\Big)$$

$$D = 2\sum_{j=1}^{m}\Big[\sum_{i=1}^{m} y_{T+2-i}a_{ij}\Big(\sum_{t=1}^{T} y_{t-j}y_t\Big)\Big]$$

$$+ w_1\Big[\sum_{j=1}^{m} y_{T+2-j}\Big(\sum_{i=1}^{m} y_{T+1-i}a_{ij}\Big) + \sum_{j=1}^{m} y_{T+1-j}\Big(\sum_{i=1}^{m} y_{T+2-i}a_{ij}\Big)\Big]$$

$$E = Y_T'Y_T + w_1^2 - (Y_T'YaY'Y_T/d) - 2(w_1/d)\Big[\sum_{j=1}^{m}\Big(\sum_{i=1}^{m} y_{T+1-i}a_{ij}\sum_{t=1}^{T} y_{t-j}y_t\Big)\Big]$$

$$- (w_1^2/d)\Big[\sum_{j=1}^{m} y_{T+1-j}\Big(\sum_{i=1}^{m} y_{T+1-i}a_{ij}\Big)\Big].$$

Note that C, D, and E involve the y's and w_1 but not w_2. Define F and G by

$$F = E - D^2/[d(d - C)]$$

$$G = 1 - (C/d).$$

Then,

$$B = G\{w_2^2 - [D/(d - C)]w_2 + [D^2/(d - C)^2]\} + F.$$

Thus,

$$g(w_1, w_2) \propto |Y'Y + W'W|^{(-1/2)}|F|^{-(T+1)/2}$$

$$x\ \{1 + F^{-1}G[w_2 - D/(d - C)]^2\}^{-(T+1)/2}.$$

Let $g_1(w_1) = |Y'Y + W'W|^{-1/2} F^{-(T+1)/2}$ and $g_2(w_1, w_2)$ be the remaining expression. Then, $g(W_2)$ can be written as the product of g_1 and g_2 which indicates that g_2 is proportional to the conditional density of w_2 given w_1. Therefore, $g(w_2|w_1)$ is a t density with degrees of freedom T, location $D/(d-C)$, precision $TF^{-1}G$. Hence, the following corollary has been proved.

Corollary 6.2. Given an m-th order autoregressive process, and observations $y_1, y_2, \ldots, y_T$, the predictive density of y_{T+2}, conditioned on y_{T+1}, is a univariate t with T degrees of freedom, location parameter μ, and precision τ, where μ and τ are as given above.

In summary, note that w_1 has a t distribution with degrees of freedom depending only on the number of observations T, not on the order of the process. Furthermore, while $g(W_2)$ is not a t, the conditional density of w_2 given w_1 is a general univariate t distribution.

Consider now the special case where m=1 and k=1. Hence, interest is on the AR(1) or Markof process with prediction for one step ahead. With no great difficulty the following corollary can be proved.

Corollary 6.3. Given a first order autoregressive process and T observations $y_1, \ldots, y_T$, the predictive density of $w_1=y_{T+1}$, a one step ahead forecast, is given by

$$w_1 \sim t(w_1;\ T - 1,\ \mu,\ \tau)$$

where t is a univariate general t distribution with T-1 degrees of freedom, location parameter μ and precision τ, where

$$\mu = y_T\Big(\sum_{t=1}^{T} y_t y_{t-1}\Big)\Big/\Big(\sum_{t=1}^{T} y_{t-1}^2\Big)$$

$$\tau = (T-1)\Big(\sum_{t=1}^{T} y_{t-1}^2\Big)^2 \Big/ \Big\{\Big(\sum_{t=1}^{T} y_{t-1}^2\Big)\Big(\sum_{t=1}^{T} y_t^2\Big)\Big(\sum_{t=0}^{T} y_t^2\Big) - \Big(\sum_{t=0}^{T} y_t^2\Big)\Big(\sum_{t=1}^{T} y_t y_{t-1}\Big)^2\Big\}.$$

Nonzero Mean Case

Consider the AR(1) process with nonzero mean and explore this case to contrast with the preceeding one. This process is defined by

$$y_t = \beta_1 + \beta_2 y_{t-1} + u_t$$

and the u_t are i.i.d. $N(0, \tau)$, $0 < \tau < \infty$, $t = 1, 2, \ldots, T$.

Assume β_1, β_2 and τ are unknown and a priori independent and assign the prior

$$\pi_0(\beta_1, \beta_2, \tau) \propto 1/\tau, \quad -\infty < \beta_1 < \infty, \quad -\infty < \beta_2 < \infty, \quad 0 < \tau < \infty.$$

Assume the initial value y_0 is known. Let $Y=(y_1, \ldots, y_T)$. Then

$$L(Y|\beta_1, \beta_2, \tau) \propto \tau^{T/2} \exp\Big[(-\tau/2) \sum_{t=1}^{T} (y_t - \beta_1 - \beta_2 y_{t-1})^2\Big].$$

The posterior p.d.f. is a product of π_0 and L. Let the future values $y_{T+1}, \ldots, y_{T+k}$ be denoted by $W_k=(w_1, \ldots, w_k)$, with, as previously defined, $w_j=y_{T+j}$, $j=1, \ldots, k$. The predictive density can be shown to be

$$g(W_k) \propto |A|^{-1/2} (D - C'A^{-1}C)^{-(T+k-1)/2}, \quad W_k \varepsilon R^k,$$

where

$$A = \begin{bmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{1}^{T} y_{t-1}^2 \end{bmatrix} + \begin{bmatrix} k & \sum_{q=1}^{k} w_{q-1} \\ \sum_{1}^{k} w_{q-1} & \sum_{1}^{k} w_{q-1}^2 \end{bmatrix}$$

$$C = \left(\sum_{1}^{T} y_t + \sum_{1}^{k} w_q, \sum_{1}^{T} y_t y_{t-1} + \sum_{1}^{k} w_q w_{q1}\right)$$

$$D = \sum_{1}^{T} y_t^2 + \sum_{1}^{k} w_q^2.$$

Numerical Example

The IMSL subroutine FTGEN was used to generate values for autoregressive processes. A model used by Zellner (1971) for the AR(1) process is

$$y_t = 0.8y_{t-1} + u_t, \; t = 1, 2, \ldots, T,$$

where the u_t are i.i.d. N(0, 4), that is τ=4, σ=0.5, and initial value y_0=2.0 is given. The generated data are given in Table II, Appendix. Then subroutines MODE and HPD (Cook, 1980) were used with the predictive density function given by Corollary 6.3 to obtain values for mode, median, mean and higher order moments, as well as variance for the prediction. Using x_8 as y_0 and x_9 through x_{28} as sample data, the results for predicting y_{21} are

Point Estimates

mode	-0.2358
median	-0.2349
mean	-0.2349

HFD regions of content $1-\alpha$

content	end points of region	
.90	-.8123	0.3424
.95	-.8585	0.3883
.98	-.8866	0.4167
.99	-.8959	0.4261

Next, consider updating a prediction for the same AR(1) model. Given $\hat{y}_{T+1}$, predict $\hat{y}_{T+2}$. Again, using MODE and HPD we obtain the point estimates

mode	-.1500
median	-.1525
mean	-.1525.

CHAPTER VII

SUMMARY

The main objective of this paper is the development of predictive densities for several cases of interest, including switching regression with vague priors, the general linear model with quasi-conjugate priors and with one switch in regression parameters, and autoregressive processes of order m, AR(m). The last is a special case of a time series model.

For each case, the Bayesian predictive density is obtained. Certain results are demonstrated with numerical examples. All calculations for examples were done in double precision on an IBM 370/158 Computer at Oklahoma State University Computer Center. In general, if the precision is known, then the predictive density involves a normal density. If the precision is unknown, then the predictive density involves a general t density. If the switch point is known (or not applied, as in the AR process), then the result is a simple density, the particular one being as indicated above. If there is a switch point and it is unknown, then the predictive density is a mixture of densities (normal or t, as seen above) and the mixing coefficients are functions of the observations. Furthermore, from numerical work, it appeared that a mixture of t densities is, at least approximately, also a t.

For k steps ahead, the predictive density for the AR(m) process is derived. For k more than one, the joint density of the k future values

is not a known form. But for one step (k=1) ahead and also for a second step, conditioned on the predicted value of the first step, we obtained a Bayesian predictive density in the form of a univariate general t density, with the degrees of freedom being T-1 or T, respectively, where T observations on the process are available. We note that the degrees of freedom depend on the number of observations, not the order of the process. Thus, as noted, T observations produce T-1 for the degrees of freedom parameter. However, although the so-called initial values y_0, y_{-1}, $\cdots$, $y_{-(m-1)}$ are not considered observations as such, they are values obtained from the process and do enter into the calculations for the predictive density and its parameters.

Although not an addition to theory, as are the results delineated above, but merely an addition to technique, that which made obtaining the most general case, AR(m), k steps, a possibility was defining the matrix of autoregressive values for y_t as

$$Y = \begin{bmatrix} y_0 & y_{-1} & \cdots & y_{-(m+1)} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ y_{T-1} & y_{T-2} & \cdots & y_{T-m} \end{bmatrix}$$

and for W similarly. As Pigno remarked (Kolata, 1981), "Historically, it's always turned out that new methods have more significance than the original problem" (p. 31). Perhaps this technique of representing the process can be adapted to help solve the problem of finding a predictive density in a number of other interesting and useful cases. Among these unsolved cases is prediction for moving average (MA) processes, that is, a time series model of the form

$$y_t = u_t - \theta_1 u_{t-1} - \cdots - \theta_q u_{t-q},$$

where the u_t are i.i.d. $N(0, \tau)$, for finite precision τ, $\tau > 0$, and the θ_i, $i = 1, 2, \ldots, q$ are constants. Also, the MA(q) process could be studied with quasi-conjugate as well as with vague priors for the prediction of future values. Also, prediction for the AR(m) process could be approached from the standpoint of conjugate or quasi-conjugate priors on the unknown parameters. Lahiff (1980) has tackled this particular problem, assuming stationarity, and has found the solution mathematically intractable and has proceeded numerically to seek an approximate solution. But, perhaps that problem merits further study and with different restrictions on the priors assigned.

Similarly, the ARMA(p, q) process with vague and with quasi-conjugate priors, and the respective predictive densities would be of interest as would the autoregressive integrated moving average (ARIMA) process and the corresponding predictive densities.

Furthermore, each of these time series models deserves attention from the standpoint of a switch in the parameters occurring at some point and yet a need exists to predict future values, considering the switch and various priors on the parameters. The switch could be abrupt or gradual, as in the work of Salazar (1980).

In any of the problems suggested, the predictive density derived mathematically could be compared, mathematically or numerically, with non-Bayesian results, to explore under what conditions each approach is superior to other ways. In addition, forecasts for both stationary and non-stationary time series processes deserve to be explored, with suitable priors in each case. Hence, many problems remain for resolution in the future, but it is hoped that this study completes at least a step toward better understanding of forecasting and provides a sound theoretical base for work on predictive densities in other cases.

A SELECTED BIBLIOGRAPHY

Aitchison, J. "Bayesian Tolerance Regions." Journal of the Royal Statistical Society, Series B, Vol. 26 (1964), 161-175.

Aitchison, J. and I. R. Dunsmore. Statistical Prediction Analysis. Cambridge, London: Cambridge University Press, 1975.

Austin, F. S. Lee and Sylva M. Heghinian. "A Shift in the Mean Level in a Sequence of Independent Normal Random Variables--A Bayesian Approach." Technometrics, Vol. 19, No. 4 (1977), 503-506.

Bacon, D. W. and D. G. Watts. "Estimating the Transition Between Two Intersecting Straight Lines." Biometrika, Vol. 58 (1971), 525-534.

Barnett, V. Comparative Statistical Inference. New York: John Wiley and Sons, 1973.

Bayes, T. "Studies in the History of Probability and Statistics. IX. Thomas Bayes' Essay 'Toward Solving a Problem in the Doctrine of Chances.'" Ed. G. A. Barnard. Biometrika, Vol. 45 (1958), 293-315.

Bibby, J. and H. Toutenburg. Prediction and Improved Estimation in Linear Models. New York: John Wiley and Sons, 1977.

Bloomfield, Peter. Fourier Analysis of Time Series: An Introduction. New York: John Wiley and Sons, 1976.

Box, G. E. P. and George C. Tiao. "Multiparameter Problems from a Bayesian Point of View." The Annals of Mathematical Statistics, Vol. 36 (1965), 1468-1482.

Box, G. E. P. and G. M. Jenkins. Time Series Analysis, Forecasting and Control. San Francisco: Holden-Day, 1970.

Box, G. E. P. and G. C. Tiao. Bayesian Inference in Statistical Analysis. Reading, MA: Addison-Wesley, 1973.

Broemeling, L. D. "Bayesian Procedures for Detecting a Change in a Sequence of Random Variables." Metron, Vol. xxx-N-1-4 (1972), 1-14.

Broemeling, L. D. "Bayesian Inference About a Changing Sequence of Random Variables." Communications in Statistics, Vol. A3 (1974), 234-255.

Broemeling, L. D. "Forecasting Future Values of Changing Sequences." Communications in Statistics, Vol. A6 (1977), 87-102.

Brown, P. and C. Payne. "Election Night Forecasting." Journal of the Royal Statistical Society, Series A, Vol. 138 (1975), 463-498.

Brown, R. L., J. Durbin, and J. M. Evans. "Testing Constancy of Regression Relationships." Journal of the Royal Statistical Society, Series B, Vol. 37 (1975), 149-192.

Chatfield, Christopher. The Analysis of Time Series: Theory and Practice. London: Chapman and Hall, 1975.

Chernoff, H. and S. Zacks. "Normal Distribution Which is Subjected to Changes in Time." Annals of Mathematical Statistics, Vol. 35 (1964), 999-1018.

Chin Choy, J. "A Bayesian Analysis of a Changing Linear Model." (Unpub. Ph.D. dissertation, Oklahoma State University, Stillwater, Oklahoma, 1977.)

Cook, Peyton James. "A Manual for Subroutine MODE and HPD." (Unpub. M.S. report, Oklahoma State University, Stillwater, Oklahoma, 1980.)

David, F. N. "Studies in the History of Probability and Statistics: Dicing and Gaming (A Note on the History of Probability)." Biometrika, Vol. 42 (1955), 1-15.

DeGroot, M. H. Optimal Statistical Decisions. New York: McGraw-Hill Book Company, 1970.

Dickey, James M. "Appropriate Posterior Distributions." Journal of the American Statistical Association, Vol. 71 (1976), 680-689.

Farley, J. U. and M. J. Hinich. "A Test for a Shifting Slope Coefficient in a Linear Model." Journal of the American Statistical Association, Vol. 65 (1970), 1320-1329.

Ferriera, P. E. "A Bayesian Analysis of a Switching Regression Model: Known Number of Regimes." Journal of the American Statistical Association, Vol. 70 (1975), 370-374.

Fuller, Wayne A. Introduction to Statistical Time Series. New York: John Wiley and Sons, 1976.

Fuller, Wayne A. and David P. Hasza. "Properties of Predictors for Autoregressive Time Series." Journal of the American Statistical Association, Vol. 76 (1981), 155-161.

Garbade, K. "Two Methods for Examining the Stability of Regression Coefficients." Journal of the American Statistical Association, Vol. 72 (1977), 54-63.

Gilchrist, W. Statistical Forecasting. New York: John Wiley and Sons, 1976.

Green, M. and P. J. Harrison. "Fashion Forecasting for a Mail Order Company Using a Bayesian Approach." Operational Research Quarterly, Vol. 24 (1973), 193-205.

Guttmann, I. and G. C. Tiao. "A Bayesian Approach to Some Best Population Problems." Annals of Mathematical Statistics, Vol. 35 (1964), 825-835.

Harrison, P. J. and C. F. Stevens. "A Bayesian Approach to Short-term Forecasting." Operational Research Quarterly, Vol. 22 (1971), 341-362.

Harrison, P. J. and C. F. Stevens. "Bayesian Forecasting (With Discussion)." Journal of the Royal Statistical Society, Series B, Vol. 38 (1976), 205-247.

Hinkley, D. V. "Inference About the Intersection in Two-Phase Regression." Biometrika, Vol. 56 (1969), 495-504.

Hinkley, D. V. "Inference in Two-Phase Regression." Journal of the American Statistical Association, Vol. 66 (1971a), 736-742.

Hinkley, D. V. "Inference About the Change-Point From Cumulative Sum Tests." Biometrika, Vol. 58 (1971a), 509-523.

Holbert, D. "A Bayesian Analysis of Shifting Sequences With Applications to Two-Phase Regression." (Unpub. Ph.D. dissertation, Oklahoma State University, 1973.)

Holbert, D. and L. D. Broemeling. "Bayesian Inference Related to Shifting Sequences and Two-Phase Regression." Communications in Statistics, Vol. A6 (1977), 265-275.

Hudson, D. J. "Fitting Segmented Curves Whose Join Points Have to be Estimated." Journal of the American Statistical Association, Vol. 61 (1966), 1097-1129.

Kakwani, N. C. "Note on the Use of Prior Information in Forecasting With a Linear Model." Sankhya: The Indian Journal of Statistics, Vol. 27 (1965), 101-104.

Kander, Z. and S. Zacks. "Test Procedures for Possible Changes in Parameters of Statistical Distributions Occurring at Unknown Time Points." Annals of Mathematical Statistics, Vol. 37 (1966), 1196-1210.

Kendall, M. G. Time-Series. London: Charles Griffin and Company, Limited, 1973.

Kolata, Gina Bari. "Littlewood Conjecture Resolved." Science, Vol. 212, No. 4490 (1981), 30-31.

Lahiff, Maureen. "Time-Series Forecasting with Informative Prior Distributions," Technical Report No. 111, Department of Statistics, The University of Chicago, May, 1980.

Lindley, D. V. "The Theory of Queques With a Single Server." Proceedings of the Cambridge Philosophical Society, Vol. 48 (1952), 277-289.

Lindley, D. V. Bayesian Statistics, A Review. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1972.

Litterman, Robert B. "A Bayesian Procedure for Forecasting with Vector Autoregressions." (Unpub. Photocopies Paper, Massachusetts Institute of Technology, Boston, Massachusetts, 1980), 36 pp.

Page, E. S. "Continuous Inspection Schemes." Biometrika, Vol. 41 (1954), 100-115.

Page, E. S. "Control Charts With Warning Lines." Biometrika, Vol. 42 (1955), 243-257.

Page, E. S. "On Problems in Which a Change in a Parameter Occurs at an Unknown Time Point." Biometrika, Vol. 44 (1957), 248-252.

Quandt, R. E. "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes." Journal of the American Statistical Association, Vol. 53 (1958), 873-880.

Quandt, R. E. "Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes." Journal of the American Statistical Association, Vol. 55 (1960), 324-330.

Robison, D. E. "Estimates for the Points of Intersection of Two Polynomial Regressions." Journal of the American Statistical Association, Vol. 59 (1964), 214-224.

Salazar, Diego. "The Analysis of Structural Changes in Time Series and Multivariate Linear Models." (Unpub. Ph.D. dissertation, Oklahoma State University, Stillwater, Oklahoma, 1980.)

Swamy, P. A. V. B. and J. S. Mehta. "Bayesian and Non-Bayesian Analysis of Switching Regressions and of Random Coefficient Regression Models." Journal of the American Statistical Association, Vol. 70 (1975), 593-602.

Winkler, R. L. "The Assessment of Prior Distributions in Bayesian Analysis." Journal of the American Statistical Association, Vol. 62 (1967), 776-800.

Zellner, Arnold. An Introduction to Bayesian Inference in Econometrics. New York, New York: John Wiley and Sons, 1971.

APPENDIX

TABLE I

DATA SET GENERATED BY QUANDT

Obs. No. i	1	2	3	4	5	6	7	8	9	10
x_i	4	13	5	2	6	8	1	12	17	20
y_i	3.473	11.555	5.714	5.710	6.046	7.650	3.140	10.312	13.353	17.197

Obs. No. i	11	12	13	14	15	16	17	18	19	20	21
x_i	15	11	3	14	16	10	7	19	18	19	13
y_i	13.036	8.264	7.612	11.802	12.551	10.296	10.014	15.472	15.65	9.871	

TABLE II

AR(1), $x_t = 0.8\ x_{t-1} + u_t$, $u_t \sim$ i.i.d. N(0, 4), $y_0 = 2.0$

i	1	2	3	4	5	6	7	8	9	10
x_i	.7046	-.0748	-.4559	-.4650	-.0348	.2017	.0201	-.1317	-.0248	.1289
i	11	12	13	14	15	16	17	18	19	20
x_i	.1466	-.0966	-.5410	-.6694	-.7044	-.5654	-.5354	-.1964	-.8373	-.3751
i	21	22	23	24	25	26	27	28	29	30
x_i	.4762	.9060	.0195	.4863	.6922	.4337	.2119	-.3618	-1.2544	-.8735

VITA

Margaret Foster Land

Candidate for the Degree of

Doctor of Philosophy

Thesis: BAYESIAN FORECASTING FOR SWITCHING REGRESSION AND AUTOREGRESSIVE PROCESSES

Major Field: Statistics

Biographical:

Personal Data: Born in Norman, Oklahoma, February 20, 1939, daughter of Agnes Scott; reared by maternal grandparents Henry C. and Ona Ruth Foster.

Education: Graduated from Norman High School, Norman, Oklahoma, in 1957; attended University of Oklahoma, 1957-58 and 1959-60; attended Concord College, Athens, West Virginia, 1960-62; graduated from Northwestern State University of Louisiana, Natchitoches, Louisiana, in 1963 with a Bachelor of Science degree in Mathematics with Honors; received a Master of Science degree in Mathematics in 1967 from Northwestern State University of Louisiana; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in May, 1981.

Professional Experience: Graduate Teaching Assistant, Northwestern State University of Louisiana, 1963-67; Instructor, 1967-68; Assistant Professor, 1969-70. Participant, National Science Foundation program for College Teachers at the University of Georgia, Athens, Georgia, June-August, 1970; Graduate Teaching Assistant, Departments of Mathematics and Statistics, Oklahoma State University, 1970-77; Assistant Professor, University Statistics Center and Department of Experimental Statistics, New Mexico State University, Las Cruces, New Mexico, 1977-80; participated in statistics research project, August, 1980, Instituto de Investigaciones en Matematicas Aplicadas y Sistemas, Universidad Nacional Autonoma de Mexico, in Mexico City; Graduate Associate, Department of Statistics, Oklahoma State University, 1980-81.

Professional Associations: American Statistical Association, The Biometric Society, Mathematical Association of America, and American Association for the Advancement of Science.