STATISTICS: textbooks and monographs volume 60

BAYESIAN ANALYSIS OF LINEAR MODELS

LYLE D. BROEMELING

BAYESIAN ANALYSIS OF LINEAR MODELS

STATISTICS: Textbooks and Monographs

A SERIES EDITED BY

D. B. OWEN, Coordinating Editor

Department of Statistics
Southern Methodist University
Dallas, Texas

Vol. 1: The Generalized Jacknife Statistic, *H. L. Gray and W. R. Schucany*
Vol. 2: Multivariate Analysis, *Anant M. Kshirsagar*
Vol. 3: Statistics and Society, *Walter T. Federer*
Vol. 4: Multivariate Analysis: A Selected and Abstracted Bibliography, 1957-1972, *Kocherlakota Subrahmaniam and Kathleen Subrahmaniam* (out of print)
Vol. 5: Design of Experiments: A Realistic Approach, *Virgil L. Anderson and Robert A. McLean*
Vol. 6: Statistical and Mathematical Aspects of Pollution Problems, *John W. Pratt*
Vol. 7: Introduction to Probability and Statistics (in two parts), Part I: Probability; Part II: Statistics, *Narayan C. Giri*
Vol. 8: Statistical Theory of the Analysis of Experimental Designs, *J. Ogawa*
Vol. 9: Statistical Techniques in Simulation (in two parts), *Jack P. C. Kleijnen*
Vol. 10: Data Quality Control and Editing, *Joseph I. Naus* (out of print)
Vol. 11: Cost of Living Index Numbers: Practice, Precision, and Theory, *Kali S. Banerjee*
Vol. 12: Weighing Designs: For Chemistry, Medicine, Economics, Operations Research, Statistics, *Kali S. Banerjee*
Vol. 13: The Search for Oil: Some Statistical Methods and Techniques, *edited by D. B. Owen*
Vol. 14: Sample Size Choice: Charts for Experiments with Linear Models, *Robert E. Odeh and Martin Fox*
Vol. 15: Statistical Methods for Engineers and Scientists, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
Vol. 16: Statistical Quality Control Methods, *Irving W. Burr*
Vol. 17: On the History of Statistics and Probability, *edited by D. B. Owen*
Vol. 18: Econometrics, *Peter Schmidt*
Vol. 19: Sufficient Statistics: Selected Contributions, *Vasant S. Huzurbazar (edited by Anant M. Kshirsagar)*
Vol. 20: Handbook of Statistical Distributions, *Jagdish K. Patel, C. H. Kapadia, and D. B. Owen*
Vol. 21: Case Studies in Sample Design, *A. C. Rosander*
Vol. 22: Pocket Book of Statistical Tables, *compiled by R. E. Odeh, D. B. Owen, Z. W. Birnbaum, and L. Fisher*
Vol. 23: The Information in Contingency Tables, *D. V. Gokhale and Solomon Kullback*
Vol. 24: Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods, *Lee J. Bain*
Vol. 25: Elementary Statistical Quality Control, *Irving W. Burr*
Vol. 26: An Introduction to Probability and Statistics Using BASIC, *Richard A. Groeneveld*
Vol. 27: Basic Applied Statistics, *B. L. Raktoe and J. J. Hubert*
Vol. 28: A Primer in Probability, *Kathleen Subrahmaniam*
Vol. 29: Random Processes: A First Look, *R. Syski*

Vol. 30: Regression Methods: A Tool for Data Analysis, *Rudolf J. Freund and Paul D. Minton*
Vol. 31: Randomization Tests, *Eugene S. Edgington*
Vol. 32: Tables for Normal Tolerance Limits, Sampling Plans, and Screening, *Robert E. Odeh and D. B. Owen*
Vol. 33: Statistical Computing, *William J. Kennedy, Jr. and James E. Gentle*
Vol. 34: Regression Analysis and Its Application: A Data-Oriented Approach, *Richard F. Gunst and Robert L. Mason*
Vol. 35: Scientific Strategies to Save Your Life, *I. D. J. Bross*
Vol. 36: Statistics in the Pharmaceutical Industry, *edited by C. Ralph Buncher and Jia-Yeong Tsay*
Vol. 37: Sampling from a Finite Population, *J. Hajek*
Vol. 38: Statistical Modeling Techniques, *S. S. Shapiro*
Vol. 39: Statistical Theory and Inference in Research, *T. A. Bancroft and C.-P. Han*
Vol. 40: Handbook of the Normal Distribution, *Jagdish K. Patel and Campbell B. Read*
Vol. 41: Recent Advances in Regression Methods, *Hrishikesh D. Vinod and Aman Ullah*
Vol. 42: Acceptance Sampling in Quality Control, *Edward G. Schilling*
Vol. 43: The Randomized Clinical Trial and Therapeutic Decisions, *edited by Niels Tygstrup, John M. Lachin, and Erik Juhl*
Vol. 44: Regression Analysis of Survival Data in Cancer Chemotherapy, *Walter H. Carter, Jr., Galen L. Wampler, and Donald M. Stablein*
Vol. 45: A Course in Linear Models, *Anant M. Kshirsagar*
Vol. 46: Clinical Trials: Issues and Approaches, *edited by Stanley H. Shapiro and Thomas H. Louis*
Vol. 47: Statistical Analysis of DNA Sequence Data, *edited by B. S. Weir*
Vol. 48: Nonlinear Regression Modeling: A Unified Practical Approach, *David A. Ratkowsky*
Vol. 49: Attribute Sampling Plans, Tables of Tests and Confidence Limits for Proportions, *Robert E. Odeh and D. B. Owen*
Vol. 50: Experimental Design, Statistical Models, and Genetic Statistics, *edited by Klaus Hinkelmann*
Vol. 51: Statistical Methods for Cancer Studies, *edited by Richard G. Cornell*
Vol. 52: Practical Statistical Sampling for Auditors, *Arthur J. Wilburn*
Vol. 53: Statistical Signal Processing, *edited by Edward J. Wegman and James G. Smith*
Vol. 54: Self-Organizing Methods in Modeling: GMDH Type Algorithms, *edited by Stanley J. Farlow*
Vol. 55: Applied Factorial and Fractional Designs, *Robert A. McLean and Virgil L. Anderson*
Vol. 56: Design of Experiments: Ranking and Selection, *edited by Thomas J. Santner and Ajit C. Tamhane*
Vol. 57: Statistical Methods for Engineers and Scientists. Second Edition, Revised and Expanded, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
Vol. 58: Ensemble Modeling: Inference from Small-Scale Properties to Large-Scale Systems, *Alan E. Gelfand and Crayton C. Walker*
Vol. 59: Computer Modeling for Business and Industry, *Bruce L. Bowerman and Richard T. O'Connell*
Vol. 60: Bayesian Analysis of Linear Models, *Lyle D. Broemeling*

OTHER VOLUMES IN PREPARATION

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

AN IMPORTANT MESSAGE TO READERS...

A Marcel Dekker, Inc. Facsimile Edition contains the exact contents of an original hard cover MDI published work but in a new soft sturdy cover.

Reprinting scholarly works in an economical format assures readers that important information they need remains accessible. Facsimile Editions provide a viable alternative to books that could go "out of print." Utilizing a contemporary printing process for Facsimile Editions, scientific and technical books are reproduced in limited quantities to meet demand.

Marcel Dekker, Inc. is pleased to offer this specialized service to its readers in the academic and scientific communities.

BAYESIAN ANALYSIS OF LINEAR MODELS

Lyle D. Broemeling

Oklahoma State University
Stillwater, Oklahoma

MARCEL DEKKER, INC. New York and Basel

Library of Congress Cataloging in Publication Data

Broemeling, Lyle D., [date]
Bayesian analysis of linear models.

(Statistics, textbooks and monographs ; v. 60)
Includes bibliographical references and index.
1. Linear models (Statistics) 2. Bayesian statistical decision theory. I. Title. II. Series.
QA276.B695 1985 519.5 84-19902
ISBN: 0-8247-8582-7

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Current printing (last digit):
10 9 8 7 6 5 4 3 2

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

This book presents the basic theory of linear models from a Bayesian viewpoint. A course in linear model theory is usually required for graduate degrees in statistics and is the foundation for many other courses in the curriculum. In such a course the student is exposed to regression models, models for designed experiments, including fixed, mixed and random models, and perhaps some econometric and time series models.

The approach taken here is to introduce the reader to a wide variety of linear models, not only those most frequently used by statisticians but those often used by engineers and economists. This book is unique in that time series models such as autoregressive moving average processes are treated as linear models in the same way the general linear model is examined. Discrete-time linear dynamic systems are another class of linear models which are not often studied by statisticians; however, these systems are quite popular with communication and control engineers and are easily analyzed with Bayesian procedures.

Why the Bayesian approach? Because one may solve the inferential problems in the analysis of linear models with one tool, namely Bayes theorem. Instead of learning the large number of sampling theory techniques of estimation, hypothesis testing, and forecasting, one need only remember how to apply Bayes theorem.

Bayes statistics is now rapidly becoming accepted as a way to solve applied statistical problems and has several special features which combine to make it appealing for solving applied problems: First, it can be applied to a wide range of statistical and other problems; second, it allows the formal use of prior information, and thereby it offers a well-defined and straightforward way of analyzing any problem. Also, it assists the understanding and formulation of applied problems in statistical terms, and, finally, it faces squarely the problem of optimal action in the face of uncertainty.

This book is intended for second-year statistics graduate students who have completed introductory courses in probability and mathematical statistics, regression analysis, and the design of experiments. Graduate students of econometrics and communication engineering should also benefit, and the book should serve well as a reference for statisticians, econometricians, and engineers.

Many people have helped me to write this book, and I am indebted to former students Samir Shaarawy, Joaquin Diaz, Peyton Cook, David Moen, Diego Salazar, Muthyia Rajagopalan, Mohamed Al Mahmeed, Mohamed Gharraf, Margaret Land, Juanita Chin Choy, Albert Chi, Yusoff Abdullah, and Don Holbert.

Secretarial and clerical assistance was provided by Leroy Folks of Oklahoma State University and William B. Smith of Texas A & M University. Some of the topics in this book are a result of the research sponsored by the Office of Naval Research, Contract No. N000-82-K-0292.

Lyle D. Broemeling

CONTENTS

INTRODUCTION

There are many who do not accept Bayesian techniques because of the controversy surrounding the use of prior distributions and the rejection, by Bayesian doctrine, that the sampling distribution of statistics is relevant for making inferences about model parameters. See Tiao and Box (1973) for an account of this aspect of the controversy. They essentially say the sampling properties of Bayes estimators are irrelevant, since they refer to hypothetical repetitions of an experiment which in fact will not occur. Lindley (1971) rejects many sampling-theory techniques because they violate the likelihood principle and thus violate the axioms of utility and probability. For more about the advantages and disadvantages of Bayesian inference and other inferential theories, Barnett (1973) gives an unbiased view. Chapter 2 also discusses some of the more controversial topics of Bayesian inference.

Bayes theorem and its use in inference and prediction are introduced with a study of the general linear model. The general linear model is the first model to be encountered in the first linear model course, and from a non-Bayesian viewpoint, Graybill (1961) and Searle (1968) present the sampling theory viewpoint. The treatment given here is quite different, since one theorem on the posterior analysis provides one with a way to make all inferences (estimation and hypothesis testing) about the parameters of the model. The techniques of the prior, posterior, and predictive analyses presented in Chapter 1 will be repeated as each class of models is analyzed.

Chapter 2 gives a brief history of Bayesian inference and explains the subjective interpretation of probability and the implementation of prior knowledge. It is argued that the subjective interpretation of the probability distribution of a parameter is no more subjective than the frequency interpretation of the distribution of an observable random variable. Prior information about the parameters of a model is implemented by fitting the hyperparameters of a proper density either to past data or to prior values of the observations. This chapter also gives a brief discussion of how each of the various models is used in other areas of interest.

The general linear model includes as special cases the regression models and the models of designed experiments, and since they are so important, Chapter 3 gives a detailed account of how a Bayesian would use them.

The mixed linear model is studied in Chapter 4, in which the marginal posterior distribution of the variance components is derived. Using a normal approximation to the posterior distribution of the random effects, one is able to devise Bayes estimators of the primary parameters of the model.

The subject of time series analysis is often given separate treatment; however, with the Bayesian approach to statistical inference, one is able to study ARMA (autoregressive moving average) processes in much the same way one studies the general linear model. First autoregressive processes are considered, in which one is able to find the posterior distribution of the parameters and the predictive distribution of future observations. The marginal posterior distribution of the autoregression coefficients is a multivariate t, and the error precision has a gamma posterior distribution. The predictive distribution of future observations is shown to be a t distribution. The moving average class of models is more difficult to analyze, although the exact analysis of an MA(1) and ARMA(1,1) processes is reported. Other time series models which are studied are the regression model with autocorrelated errors and the so-called lagged variable model in econometrics. This chapter shows that the Bayesian approach to the study of time series models is in its infancy but that the approach will produce some interesting results.

Very few statisticians learn about the linear dynamic model and its use by communication engineers with monitoring and navigation problems. The model is called dynamic because the parameters of the model are always changing, from one observation to the next; thus it is natural to study such processes from a Bayesian approach.

There are several important problems in connection with a dynamic linear model. The first is estimation of the parameters of the system, and estimation consists of filtering, smoothing, and prediction. Filtering is estimating the current state (parameter) of the system, while smoothing is estimating the past states of the system. Prediction, unlike time series analysis, is estimating future states of the system. Chapter 6 develops the Bayesian way of estimating the parameters by finding the posterior distribution of the states of the system, and the Kalman filter is shown to be the mean of the posterior distribution of the current state.

Chapter 6 also treats other aspects of the linear dynamic model, including adaptive estimation and control.

Chapter 7 is a complete study of structural change in linear models, in which structural change means some of the parameters of the model will change during the time the observations will be taken. In this type of system the parameters change only occasionally, whereas in linear dynamic systems they change with each additional observation. Thus models with structural change stand between the static class, such as the traditional linear models (regression models, for example), and linear dynamic models.

There are two ways to model structural change: One is with a shift point, and the other is with a transition function. Both are used in Chapter 7, with regression models and time series processes.

Most of the material in Chapter 8 on multivariate processes is given by Box and Tiao (1973) and Zellner (1971). Multivariate linear regression models, models for designed experiments, and multivariate autoregressive processes are introduced. The only new material is a section on structural change of a multivariate regression model.

Many examples illustrate the prior, predictive, and posterior analyses, which are the three main ingredients of a Bayesian inference.

The necessary statistical background is given in the Appendix, in which the following distributions are defined: the univariate normal, the gamma, the normal-gamma, the multivariate normal, the Wishart, the normal-Wishart, the multivariate t, the poly t, and the matrix t. If one learns the material in the Appendix, one will be able to understand the way a Bayesian makes statistical inferences about linear models.

The books by Box and Tiao (1973) and Zellner (1971) have given us the foundation for the Bayesian analysis of linear models; thus this book can best be thought of, first, as a review of the material in those books and, second, as a presentation of new material. This Preface and Chapters 1, 2, and 3 are mostly a review of the Bayesian techniques of estimation, hypothesis, testing, and forecasting as applied to the traditional populations, while the remaining chapters present material which is either new or an extension of existing work.

REFERENCES

Barnett, V. (1973). *Comparative Statistical Inference*. John Wiley and Sons, Inc., New York.

Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Ma.

Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*. McGraw-Hill, New York.

Lindley, D. V. (1971). *Bayesian Statistics, A Review*, Reg. Conf. Ser. Applied Math., 2. SIAM, Philadelphia.

Searle, S. R. (1968). *Linear Models.* John Wiley and Sons, Inc., New York.

Tiao, G. C. and G. E. P. Box (1973). "Some comments on Bayes estimators," The American Statistician, Vol. 27, No. 1, pp. 12-14.

Zellner, Arnold (1971). *An Introduction to Bayesian Inference in Econometrics.* John Wiley and Sons, Inc., New York.

BAYESIAN ANALYSIS OF LINEAR MODELS

1

BAYESIAN INFERENCE FOR THE GENERAL LINEAR MODEL

INTRODUCTION

Many books begin with the general linear model and this is no exception. The general linear model includes many useful and interesting special cases. For example, independent normal sequences of random variables, simple and multiple regression models, models for designed experiments, and analysis of covariance models are all special cases of the general linear model which will be explained in more detail in the following chapters.

This chapter introduces the prior, posterior, and predictive analysis of the linear model and thereby lays the foundation for the remainder of the book where the analyses are reported with other models.

THE PARAMETRIC INFERENCE PROBLEM

Let θ be a $p \times 1$ vector of real parameters, $Y = (y_1, y_2, \ldots, y_n)'$ a $n \times 1$ vector of observations, X a $n \times p$ known design matrix. Then the general linear model is

$$Y = X\theta + e, \tag{1.1}$$

where $e \sim N(0, \tau^{-1} I_n)$, and τI_n is the precision matrix of e, which has covariance matrix $\sigma^2 I_n$, and $\sigma^2 = \tau^{-1} > 0$ is unknown.

This is the general linear model and our objective is to provide inferences for θ and τ when observing $s = (y_1, y_2, \ldots, y_n)$, where y_i is the i-th observation. The word inference is somewhat of a vague word, but it usually implies a procedure which extracts information about θ from the sample s.

Introductory statistics books explain inferential techniques in terms of point and interval estimation, tests of hypotheses, and forecasts.

For the Bayesian, all inferences are based on the posterior distribution of θ, which is given by Bayes theorem.

BAYES THEOREM

Suppose one's prior information about θ is represented by a probability density function $\xi(\theta,\tau)$, $\theta \in R^p$, $\tau > 0$, then Bayes theorem combines this information with the information contained in the sample. The likelihood function for θ and τ is

$$L(\theta,\tau|s) \propto \tau^{n/2} \exp - \frac{\tau}{2}(Y - X\theta)'(Y - X\theta), \qquad (1.2)$$

where $\theta \in R^p$ and $\tau > 0$, where $\propto$ means proportional to (as a function of τ and θ). The likelihood function is one's sample information about the parameters and is the conditional density function of the sample random variables given θ and τ.

Bayes theorem gives the conditional density of θ given s

$$\xi(\theta,\tau|s) \propto L(\theta,\tau/s)\xi(\theta,\tau), \quad \theta \in R^p, \quad \tau > 0. \qquad (1.3)$$

The posterior density of θ is $\xi(\theta,\tau|s)$ and represents one's knowledge of θ and τ after observing the sample s. On the other hand our information about θ and τ before s is observed is contained in the prior density.

Note the posterior density (1.3) is written with a proportional symbol and ξ is used to denote both the prior and posterior densities. If one uses an equality sign, the posterior density is

$$\xi(\theta,\tau|s) = K\cdot L(\theta,\tau|s)\xi(\theta,\tau), \quad \theta \in R^p, \quad \tau > 0 \qquad (1.4)$$

where K is the normalizing constant and is given by

$$K^{-1} = \int_0^\infty \int_{R^p} L(\theta, \tau | s) \xi(\theta, \tau) \, d\theta \, d\tau,$$

which is the marginal probability density of Y.

Obviously it is easier to omit the normalizing constant and use the proportional symbol and this convention will be adhered to in this book.

Before continuing the analysis, one must find the posterior density of θ and τ.

PRIOR INFORMATION

The prior information about the parameters θ and τ is given in two ways. The first is when $\xi(\theta,\tau)$ is a normal-gamma prior density, namely,

$$\xi(\theta,\tau) = \xi_1(\theta|\tau)\xi_2(\tau), \quad \theta \in R^p, \quad \tau > 0, \tag{1.5}$$

where

$$\xi_1(\theta|\tau) \propto \tau^{p/2} \exp - \frac{\tau}{2}(\theta - \mu)' \mathcal{P} (\theta - \mu), \quad \theta \in R^p \tag{1.6}$$

and μ is a $p \times 1$ given vector and $\mathcal{P}$ is a known $p \times p$ positive definite matrix. Thus ξ_1 is the conditional density of θ given τ and is normal with mean vector μ and precision matrix $\tau\mathcal{P}$.

The marginal prior density of τ is gamma with parameters $\alpha > 0$ and $\beta > 0$.

$$\xi_2(\tau) \propto \tau^{\alpha-1} \exp - \tau\beta, \quad \tau > 0. \tag{1.7}$$

What does this imply about the marginal prior density of θ? Since (1.5) is the joint prior density of θ and τ, the marginal density of θ is

$$\begin{aligned}
\xi_1(\theta) &\propto \int_0^\infty \xi(\theta,\tau)\, d\tau \\
&\propto \int_0^\infty \tau^{((p+2\alpha)/2)-1} \exp - \frac{\tau}{2}[2\beta + (\theta - \mu)'\mathcal{P}(\theta - \mu)]\, d\tau \\
&\propto [2\beta + (\theta - \mu)'\mathcal{P}(\theta - \mu)]^{-(p+2\alpha)/2}, \quad \theta \in R^p,
\end{aligned} \tag{1.8}$$

which is a t density with 2α degrees of freedom, location vector μ and precision matrix $(2\alpha)(2\beta)^{-1}P$.

By using the normal-gamma density as a prior for the parameters, one cannot stipulate one's prior information about θ separately from that of τ. The parameters of the marginal distribution of θ involve α and β, which are parameters of the prior distribution of τ, but the marginal prior density of τ does not involve parameters of the marginal of θ.

The parameter μ is one's prior mean for θ, while one's opinion of the correlation between the components of θ is given by $P^{-1}(2\beta)(2\alpha - 2)^{-1}$, which is the marginal prior dispersion matrix of θ. Since this involves α and β, the marginal prior information about τ depends on the choice of the dispersion or precision matrix of θ.

With regard to the information about τ, it is convenient to think of τ as the inverse of the residual variance σ^2 and that

$$E(\tau^{-1}) = \beta(\alpha - 1)^{-1}, \quad \alpha > 1 \tag{1.9}$$

$$\mathrm{var}(\tau^{-1}) = \beta^2(\alpha - 1)^{-2}(\alpha - 2)^{-1}, \quad \alpha > 2 \tag{1.10}$$

or that

$$E(\tau) = \alpha/\beta \quad \text{and} \quad \mathrm{var}(\tau) = \alpha/\beta^2.$$

These two equations together with

$$E(\theta) = \mu \tag{1.11}$$

and

$$D(\theta) = P^{-1}(2\beta)(n + 2\alpha - 2)^{-1}, \tag{1.12}$$

which are the mean vector and dispersion matrix of θ, will assist one in choosing the four hyperparameters for the prior distribution of θ and τ.

As will be seen, the normal-gamma prior density is a member of a conjugate class of distributions, that is the posterior density $\xi(\theta,\tau|s)$ of (2.3) is also a normal-gamma density. Conjugate families have the advantage that one has a scale by which to judge the amount of information added by the sample, beyond the amount given a priori.

Early writers such as Jeffreys (1959) and Lindley (1965) use improper densities for prior information. For example,

$$\xi(\theta,\tau) \propto 1/\tau, \quad \theta \in R^p, \quad \tau > 0 \tag{1.13}$$

is called a vague non-informative prior density for θ and was developed by Jeffreys, because it satisfied certain rules of invariance and conveys "very little" information about the parameters. This density, although improper, produces a normal-gamma posterior density for θ and τ, which of course is proper. Thus, improper prior densities may result in proper posterior densities, but this does not hold in general. The Jeffreys prior implies that, a priori, θ and τ are independent and that θ has a constant density over R^p and that the marginal prior density of τ is $\xi_2(\tau) \propto 1/\tau$, $\tau > 0$.

Both forms (1.5) and (1.13) of prior information will be employed in the posterior analysis which follows.

POSTERIOR ANALYSIS

Using Bayes theorem given by (1.3), then using the normal-gamma prior density (1.6), the posterior density of θ and τ is

$$\xi(\theta,\tau|s) \propto \tau^{((n+2\alpha+p)/2)-1} \exp - \frac{\tau}{2}[2\beta + (\theta-\mu)'P(\theta-\mu) + (Y - X\theta)'(Y - X\theta)] \tag{1.14}$$

for $\theta \in R^p$ and $\tau > 0$.

Now, completing the square on θ gives

$$\begin{aligned}\xi(\theta,\tau|s) \propto{}& \tau^{((n+2\alpha)/2)-1} \exp - \tau \\ &\times \left[\beta + \frac{Y'Y - (X'Y + P\mu)'(X'X + P)^{-1}(X'Y + P\mu)}{2}\right] \\ &\times \tau^{p/2} \exp - \frac{\tau}{2}[\theta - (X'X + P)^{-1}(X'Y + P\mu)]'(X'X + P) \\ &\times [\theta - (X'X + P)^{-1}(X'Y + P\mu)]\end{aligned}$$

which is a normal-gamma density, hence the marginal posterior density of τ is gamma with parameters

$$(n + 2\alpha)/2 \text{ and } \beta + \frac{Y'Y - (X'Y + P\mu)'(X'X + P)^{-1}(X'Y + P\mu)}{2}.$$

The marginal posterior density of θ is found by integrating (1.14) with respect to τ and yields

$$\xi_1(\theta|s) \propto \{2\beta + Y'Y - (X'Y + P\mu)'(X'X + P)^{-1}(X'Y + P\mu) + [\theta - (X'X + P)^{-1}(X'Y + P\mu)]'(X'X + P) \times [\theta - (X'X + P)^{-1}(X'Y + P\mu)]\}^{-(n+2\alpha+p)/2}, \quad \theta \in R^p, \tag{1.15}$$

which is a p-dimensional t density with $n + 2\alpha$ degrees of freedom, location vector

$$\mu^* = (X'X + P)^{-1}(X'Y + P\mu) \tag{1.16}$$

and precision matrix

$$D^*(\theta|s) = (X'X + P)(n + 2\alpha)[2\beta + Y'Y - (X'Y + P\mu)'(X'X + P)^{-1} \times (X'Y + P\mu)]^{-1}. \tag{1.17}$$

The reader is referred to the Appendix for properties of the normal-gamma, multivariate t and gamma distributions.

The marginal posterior moments of τ^{-1} are

$$E(\tau^{-1}|s) = [2\beta + Y'Y - (X'Y + P\mu)'(X'X + P)^{-1} \times (X'Y + P\mu)](n + 2\alpha - 2)^{-1} \tag{1.18}$$

and

$$\mathrm{Var}(\tau^{-1}|s) = E(\tau^{-1}|s)^2(n + 2\alpha - 4)^{-1}. \tag{1.19}$$

The posterior analysis of the general linear model reveals the joint distribution of θ and τ is a normal-gamma distribution, the marginal distribution of θ is a multivariate t, and the marginal of τ a gamma if the prior of the parameters is a normal-gamma.

What is the posterior analysis if the improper density $\xi(\theta,\tau) \propto 1/\tau$, $\theta \in R^p$, $\tau > 0$ is used as prior information? The following theorem should be verified.

Theorem 1.1. If θ and τ are the parameters of the general linear model (1.1) where $X'X$ is of full rank and if the prior density of θ and τ is Jeffreys' improper density

$$\xi(\theta,\tau) \propto 1/\tau, \quad \tau > 0, \quad \theta \in R^p, \tag{1.20}$$

then the posterior density of θ and τ is normal-gamma where the marginal posterior density of τ is gamma with parameters $(n - p)/2$ and $[Y'Y - Y'X(X'X)^{-1}X'Y]/2$, and the conditional posterior density of θ given τ is normal with mean vector $(X'X)^{-1}X'Y$ and precision matrix $\tau X'X$. Also, the marginal posterior density of θ is a p-dimensional t distribution with $n - p$ degrees of freedom, location vector

$$E(\theta|s) = (X'X)^{-1}X'Y \tag{1.21}$$

and precision matrix

$$D^*(\theta|s) = (n - p)(X'X)[Y'Y - Y'X(X'X)^{-1}X'Y]^{-1}. \tag{1.22}$$

These results can be obtained by assuming the prior distribution of θ and τ is a normal-gamma with hyperparameters α, β, $\mathfrak{P}$, and μ, then letting $\beta \to 0$, $\alpha \to -p/2$, and $\mathfrak{P} \to 0(p \times p)$ in the joint posterior distribution of θ and τ. Note, this is not equivalent to letting $\beta \to 0$, $\alpha \to -p/2$, and $\mathfrak{P} \to 0(p \times p)$ in the joint prior distribution of θ and τ because the prior density would not be Jeffreys improper prior, (1.20).

If one uses the normal gamma density as a prior for θ and τ, $X'X$ may be singular, but if one uses Jeffreys improper prior, $X'X$ must be nonsingular, otherwise the posterior density of θ and τ is improper.

AN EXAMPLE

Consider a special case of the general linear model, where θ is a scalar $(p = 1)$ and the design matrix is a $n \times 1$ column vector of ones, then

$$y_i = \theta + e_i \tag{1.23}$$

where the e_i are n.i.d. $(0,\tau^{-1})$, $\tau > 0$, $\theta \in R$.

The y_i, $i = 1, 2, \ldots, n$ represent a random sample of size n from a normal population with mean θ and precision τ. Suppose the prior information about θ and τ is a normal-gamma with parameters α, β, $\mathfrak{P}$, and μ, where α, β, and $\mathfrak{P}$ are positive scalars and μ is any real number, then the marginal posterior density of θ is a univariate t with $n + 2\alpha$ degrees of freedom, mean $(n + \mathfrak{P})^{-1}(\Sigma Y_i + \mathfrak{P}\mu)$ and precision $(n + \mathfrak{P})(n + 2\alpha)$

$\times\ [2\beta + \Sigma Y_i^2 - (\Sigma Y_i + P\mu)^2(n + P)^{-1}]^{-1}$. The marginal posterior density of τ^{-1} is gamma with mean

$$E(\tau^{-1}|s) = [2\beta + \Sigma Y_i^2 - (\Sigma Y_i + P\mu)^2/(n + p)](n + 2\alpha - 2)^{-1},$$

and variance

$$\mathrm{Var}(\tau^{-1}|s) = [E(\tau^{-1}|s)]^2(n + 2\alpha - 4)^{-1}.$$

Using the improper Jeffreys prior density, the marginal posterior density of θ is a t with $n - 1$ degrees of freedom, mean $\bar{y}$ and precision $n(n - 1)/\Sigma(Y_i - \bar{y})^2$, and the marginal posterior distribution of τ is gamma with parameters $(n - 1)/2$ and $\Sigma(Y_i - \bar{y})^2/2$. The results are obvious from Theorem 1.1.

INFERENCES FOR THE GENERAL LINEAR MODEL

From the Bayesian viewpoint, all inferences are based on the joint posterior distribution of θ and τ, but what is meant by inference? Within the Bayesian approach, there are two ways to study the parameters. The first is called inference and consists of plotting the posterior distribution of the parameters and computing certain characteristics, such as the mean, variance, mode, and median of the distribution. Usually, one wants to estimate the parameters with point and interval estimates or test hypotheses about the parameters. These methods are to be explained and will be illustrated in the book.

The second way to study the parameters of the model is called Bayesian decision theory approach which utilizes a loss function $L(d,\phi)$, which measures the loss when d is the decision and ϕ is the value of the parameter. The loss d* which minimizes the average loss $EL(d,\phi)$ with respect to the posterior distribution of ϕ is called the Bayes decision.

This more formal method of studying the parameters will not be used, however the reader is referred to the end of the chapter, where the current literature is discussed.

Point Estimation

This section will present point and interval estimates of the parameters θ and τ of the general linear model. Joint estimates of θ and τ will be based on the joint posterior distribution of θ and τ. If τ is a nuisance parameter and θ is the parameter of interest, estimates of θ will be based on the marginal posterior distribution of θ. In the same way, if θ is the nuisance parameter and τ is of interest, estimates of the latter will be based on the marginal posterior distribution of τ.

Consider the joint estimation of θ and τ and the joint marginal distribution of θ and τ, assuming a normal-gamma prior density with hyperparameters α, β, P, and μ. Recall from the section on posterior analysis (pp. 5-7) that the joint posterior distribution of θ and τ is also normal-gamma with parameters

$$\alpha^* = (n + 2\alpha)/2$$

$$\beta^* = \frac{2\beta + Y'Y - (X'Y + P\mu)'(X'X + P)^{-1}(X'Y + P\mu)}{2}$$

$$P^* = (X'X + P)\tau \tag{1.24}$$

and

$$\mu^* = (X'X + P)^{-1}(X'Y + P\mu).$$

If one estimates θ and τ jointly, what characteristic of the joint distribution should one use as an estimate? There is no definitive answer, but since θ is a normal random variable with mean μ^* and μ^* does not depend on τ, and since α^* and β^* do not depend on θ, then $[\mu^*, \beta^*(\alpha^* - 1)^{-1}]$ appears to be a reasonable choice with which to estimate (θ, τ^{-1}).

The marginal distribution of τ is $G[\alpha^*, \beta^*]$, hence the mean of τ^{-1} is

$$E(\tau^{-1}|s) = \beta^*(\alpha^* - 1)^{-1}$$

and its mode is

$$M(\tau^{-1}|s) = \beta^*(\alpha^* + 1)^{-1}.$$

Since the gamma distribution is asymmetric, whether one takes the mean or mode to estimate τ^{-1} is a matter of personal choice.

Suppose τ is a nuisance parameter and θ is to be estimated. The marginal mean vector of θ is μ^* and the dispersion matrix of θ is, from (1.17),

$$D(\theta|s) = (X'X + P)^{-1}(n + 2\alpha - 2)^{-1}[2\beta + Y'Y - (X'Y + P\mu)'$$
$$\times (X'X + P)^{-1}(X'Y + P\mu)]. \quad (1.25)$$

Since the posterior density of θ is symmetric, the mean, median, and mode coincide and μ^* is the natural estimate of θ.

Suppose θ is partitioned into two subvectors, θ_1 which is $p_1 \times 1$ and θ_2 which is $(p - p_1) \times 1$, where $\theta = (\theta_1', \theta_2')'$ and we want to estimate θ_1, ignoring θ_2 and τ. Using the properties of the t distribution (see the Appendix), the marginal posterior distribution of θ_1 is a p_1 dimensional t distribution with mean vector μ_1^* and precision matrix $D_1^*(\theta_1|s)$, where $\mu^* = (\mu_1^{*\prime}, \mu_2^{*\prime})'$, where μ_1^* is $p_1 \times 1$, and

$$D_1^*(\theta|s) = D_{11}^*(\theta|s) - D_{12}^*(\theta|s)D_{22}^{*-1}(\theta|s)D_{21}^*(\theta|s) \quad (1.26)$$

and

$$D^*(\theta|s) = \begin{bmatrix} D_{11}^*(\theta|s) & D_{12}^*(\theta|s) \\ D_{21}^*(\theta|s) & D_{22}^*(\theta|s) \end{bmatrix}, \quad (1.27)$$

where $D^*(\theta|s)$ is the precision matrix of θ, given by (1.17), and $D_{11}^*(\theta|s)$ is a $p_1 \times p_1$ matrix.

The natural estimator of θ_1 is the mean of the marginal posterior distribution of θ_1, which is μ_1^*.

If one uses a point estimate to estimate θ and τ, the mean of θ, μ^*, and either the mean or mode of τ will suffice, however these estimates are special characteristics of the joint posterior distribution of the parameters, hence they convey only partial information about these parameters. It is important to remember that the whole of the posterior distribution should be used to make inferences and that one should not be restricted to a few special properties of the distribution.

A distinctive feature of the Bayesian approach to inference is that one is not interested in the sampling properties of estimators since the posterior distribution of the parameters is conditional on the observed values s of the sample random variables y_i, $i = 1, 2, \ldots, n$. Future values of the estimators in hypothetical repetitions of the experiment are not relevant, and neither are the sampling distributions of the estimators, therefore the sampling properties of the estimators such as variance, mean square error, bias, correlation, etc., will not be pursued.

Nevertheless, there are interesting connections between those estimators which are employed by non-Bayesians, and certain moments of the posterior distribution of θ and τ, when one assumes a Jeffreys prior density.

For example, the "usual" estimate of θ is the least squares estimate $\hat{\theta} = (X'X)^{-1}X'Y$, which has a $N[\theta, \tau(X'X)]$ sampling distribution, but the marginal posterior distribution of θ is a t density with mean $\hat{\theta}$ and dispersion matrix $(X'X)^{-1}(n - p - 2)^{-1}(Y - X\hat{\theta})'(Y - X\hat{\theta})$. In other words, the least squares estimate is the mean vector of the marginal posterior distribution of θ and the dispersion matrix of θ is the estimated dispersion matrix (except for a constant) of the sampling distribution of $\hat{\theta}$. Also, the dispersion matrix of the posterior distribution of θ is a subjective measure of how the posterior probability of θ is concentrated about the mean $\hat{\theta}$, whereas the estimated dispersion matrix of the sampling distribution of $\hat{\theta}$ is a subjective measure of the variability of future values of $\hat{\theta}$ in hypothetical repetitions of the experiment. There is a close connection between the two ways of estimating θ, because on the one hand, probability is interpreted with regard to hypothetical repetitions, and on the other, posterior probability is interpreted in the subjective sense of confidence, that is in one's personal confidence, but either way the interpretation is subjective.

Using a proper prior for the parameters such as the normal-gamma will not produce estimators which are "close" (except in a limiting sense) to their sampling-theory counterparts, because the Bayesian estimators are functions of the hyperparameters α, β, P, and μ.

The early workers in Bayesian inference employed Jeffreys' prior distribution and the corresponding Bayes estimators were closely related in their sampling properties to the sampling theory counterparts, thus Bayes estimators were more or less tolerated. With the introduction of proper prior distributions, Bayes estimators were viewed with more suspicion since they differed more and more from the traditional estimators.

Interval and Region Estimation

In 1965, Box and Tiao defined regions of highest posterior density or HPD regions as they are called and used them to estimate the parameters θ and τ of the linear model. Their formulation refers to a posterior probability density $\xi(\phi|s)$, $\phi \in \Phi$ of some parameter ϕ, where $s = (y_1, y_2, \ldots, y_n)$ are the observed values of the sample random variables y_i, $i = 1,2, \ldots, n$. A subset R of Φ is called a $1 - \gamma$, $0 < \gamma < 1$ HPD region for ϕ if

(i) $P[\phi \in R|s] = 1 - \gamma$
(ii) If $\phi_1 \in R$ and $\phi_2 \notin R$, then $\xi(\phi_1|s) \geq \xi(\phi_2|s)$.

From the definition, one may prove that (a) among all regions of Φ which have posterior probability $(1 - \gamma)$, R has minimum volume, (b) R is a unique $(1 - \gamma)$ HPD region if the posterior density $\xi(\phi|s)$ is not uniform over every region of the parameter space, and (c) if $\xi(\phi_1|s) = \xi(\phi_2|s)$, then either both ϕ_1 and ϕ_2 belong to R or both are not contained in R; also, the converse holds.

We now give the $(1 - \gamma)$ HPD region for θ, the parameter vector of the linear model.

Since the marginal distribution of θ is $t_p[\theta, n + 2\alpha, \mu^*, D^*(\theta|s)]$, i.e., a p-dimensional t distribution with $n + 2\alpha$ degrees of freedom, location vector μ^* and precision matrix $D^*(\theta|x)$, the random variable

$$F(\theta) = p^{-1}(\theta - \mu^*)'D(\theta|s)(\theta - \mu^*), \qquad \theta \in R^p \tag{1.28}$$

has an F distribution with p and $n + 2\alpha$ degrees of freedom and a $1 - \gamma$ HPD region for θ is

$$C_{1-\gamma}(\theta) = \{\theta: F(\theta) \leq F_{\gamma;p,n+2\alpha}\} \tag{1.29}$$

To prove this the reader should refer to the properties of the t distribution and the definition of an HPD region.

Letting $\alpha \to -p/2$, $\beta \to 0$, and $P \to 0(p \times p)$ in the posterior distribution of θ is equivalent to using the Jeffreys prior density, thus under these conditions the $1 - \gamma$ HPD region for θ is given by (1.29) with the necessary adjustments.

An HPD region for τ, based on $\xi_2(\tau|s)$, which is gamma with parameters α^* and β^*, must be done numerically, because the gamma density is asymmetric. An approximate HPD region is easily found by using the chi-square tables. Note, since $(\tau|s) \sim G(\alpha^*, \beta^*)$, then $2\beta^*\tau \sim \chi^2(2\alpha^*)$, and from this one may find an interval estimate for τ and τ^{-1}.

Testing Hypotheses

The Neyman-Pearson approach is based on sampling theory and is the prevailing way to test hypotheses. Within the Bayesian framework, one has several techniques with which to test hypotheses about the parameters θ and τ of the linear model.

Consider a null hypothesis of the form H_0: $A\theta = b$ versus the alternative hypothesis H_1: $A\theta \neq b$, where θ is the vector of parameters in the linear model, A is a known $m \times k$ matrix, and b a known $m \times 1$ vector. If the null hypothesis is true, then m linear functions of θ are determined to be given constants. Most hypotheses in regression analysis and those arising in the design of experiments are of this type and will be explained in later chapters.

The approach taken here is based on the HPD region for θ and was initiated by Lindley (1965) and Box and Tiao (1965). This approach is chosen because it produces the usual confidence region for θ when a Jeffreys' prior density is appropriate.

Suppose prior information for θ and τ is given by a normal-gamma density $\xi(\theta,\tau)$, (1.5), then the HPD region for θ of content $1 - \gamma$ was derived in the previous section. Since the null hypothesis H_0 is given in terms of $U(\theta) = A\theta$ (=b), we now derive a $1 - \gamma$ HPD region for $U(\theta)$ and reject the null hypothesis whenever b is not a member of the region. What is the HPD region for $U(\theta)$?

Now $U(\theta)$ is a linear function of θ and since θ has a t distribution with $n + 2\alpha$ degrees of freedom, mean μ^*, and precision matrix D^*, $U(\theta)$ also has a t distribution and the parameters are $n + 2\alpha$ degrees of freedom, mean $A\mu^*$, and $m \times m$ precision matrix $(AD^*A')^{-1}$, where A is chosen so that AD^*A' is nonsingular. If AD^*A' is singular the posterior distribution of $U(\theta)$ is not proper. The distribution of $U(\theta)$ is denoted by

$$U(\theta) \sim t_m[u(\theta); n + 2\alpha, A\mu^*, (AD^*A')^{-1}], \tag{1.30}$$

where the m indicates an m-dimensional distribution.

Since $U(\theta)$ has a t distribution, the random variable

$$G[U(\theta)] = m^{-1}[U(\theta) - Au^*]'(AD^*A')^{-1}[U(\theta) - Au^*] \tag{1.31}$$

has an F distribution with m and $n + 2\alpha$ degrees of freedom and a $1 - \gamma$ HPD region for $U(\theta) = U$ is given by

$$C_{1-\gamma}(u) = \{u: G(u) \leqslant F_{\gamma;m,n+2\alpha}\}, \tag{1.32}$$

where $F_{\gamma;m,n+2\alpha}$ is the upper $100\gamma\%$ point of the F distribution with m and $n + 2\alpha$ degrees of freedom. The null hypothesis is rejected if $b \notin C_{1-\gamma}(u)$ or when $G(b) > F_{\gamma;m,n+2\alpha}$.

If the prior density is improper $\xi(\theta,\tau) \propto \tau^{-1}$ for $\theta \in R^p$ and $\tau > 0$, the $1 - \gamma$ HPD region for θ is the set of all θ, where $Q(\theta) = (\theta - \hat{\theta})'X'X(\theta - \hat{\theta}) \leqslant ps^2 F_{\gamma;p,n-p}$, where $\hat{\theta} = (X'X)^{-1}X'Y$ is the least squares estimate of θ and the estimate of σ^2 is $s^2 = (Y - X\hat{\theta})' \times (Y - X\hat{\theta})(n - p)^{-1}$. This is the standard confidence region of content $1 - \gamma$ for θ, and if the null hypothesis is H_0: $\theta = \theta_0$ and the alternative is H_1: $\theta \neq \theta_0$, then H_0 is rejected if $Q(\theta_0) > ps^2 F_{\gamma;p,n-p}$.

This method of testing hypotheses is not the only way and other tests are available. For example, one may use posterior odds ratios or a decision theory approach, when a loss function is available to describe the costs of making a wrong decision. DeGroot (1970) presents various Bayesian tests of hypotheses while Zellner (1971) has a good introduction to posterior odds ratios.

With the HPD approach, one must specify the hyperparameters μ, P, α, and β of the normal gamma prior density and this information must reflect the statistician's opinion about the parameters. One's opinion about θ and τ, when the null hypothesis is true, is perhaps different than when the alternative is true, so one must be careful in choosing the prior density.

PREDICTIVE ANALYSIS

Predictive analysis is the methodology that is developed in order to forecast future observations. Of course, forecasting is a very important activity in business and economics, where sophisticated time series techniques are used.

The Bayesian uses the so-called Bayesian predictive density to forecast future observations, so this will be defined in this section. There are many ways to predict observations, however the Bayesian approach is natural in that one bases one's prediction on the conditional distribution of the future given the past. In order to do this, one must treat the parameters of the model as random, that is, as will be seen, the joint distribution of the future observations and the parameters, given the past observations, is averaged over the posterior distribution of the parameters.

In this section the Bayesian predictive density is defined and illustrated with linear models, then in the later parts of the book the same approach will be employed with time series models.

Suppose $y_1, y_2, \ldots, y_n$ is a random sample from a normal population with mean θ and precision τ and that $w_1, w_2, \ldots, w_k$ are future observations, how does one forecast the values of the future observations?

The Bayesian predictive density is the conditional density of $w = (w_1, w_2, \ldots, w_k)'$ given the sample $s = (y_1, y_2, \ldots, y_n)'$. The density of w given θ, τ, and s is

$$f(w \mid \theta, \tau, s) \propto \tau^{k/2} \exp -\frac{\tau}{2} \sum_{i=1}^{k} (w_i - \theta)^2 \tag{1.33}$$

where $w \in R^k$ and does not depend on s, since given θ and τ, w and s are independent.

Now suppose θ and τ have a prior normal-gamma density, say

$$\xi(\theta, \tau) \propto \tau^{\alpha-1} e^{-\tau\beta} \tau^{1/2} e^{-(\tau/2)q(\theta-\mu)^2}, \quad \theta \in R, \quad \tau > 0, \tag{1.34}$$

where $\alpha > 0$, $\beta > 0$, $\mu \in R$, and $q > 0$ are the hyperparameters.

The product of the posterior density of θ and τ and (1.33) is the joint conditional density of w, θ, and τ given s, which when averaged with respect to θ and τ produces the Bayesian predictive density of w.

The posterior density of θ and τ is

$$\xi(\theta, \tau \mid s) \propto \tau^{(n+2\alpha)/2-1} \exp -\frac{\tau}{2}[2\beta + \Sigma(y_i - \bar{y})^2 + n(\theta - \bar{y})^2] \tag{1.35}$$

for $\theta \in R$ and $\tau > 0$, and when this is multiplied by (1.33), gives

$$g(w, \theta, \tau \mid s) \propto \tau^{(n+2\alpha+k)/2-1} \exp -\frac{\tau}{2}\left[2\beta + \sum_{1}^{n} (y_i - \bar{y})^2 + n(\theta - \bar{y})^2 + \sum_{1}^{k} (w_i - \theta)^2\right] \tag{1.36}$$

for $w \in R^k$, $\theta \in R$, and $\tau > 0$, which is to be integrated with respect to θ and τ over $R \times (0, \infty)$.

Completing the square on θ and integrating with respect to θ over R, then integrating (1.36) with regard to τ over $(0, \infty)$, then completing the square on w, results in

$$h(w \mid s) \propto \{[w - A^{-1}B]'A[w - A^{-1}B] + C - B'A^{-1}\}^{(n+2\alpha+k)/2}, \quad w \in R^k \tag{1.37}$$

as the predictive density of w, where A is a $k \times k$ matrix with diagonal elements $(n + k - 1)(n + k)^{-1}$ and off-diagonal elements $-(n + k)^{-1}$. The column vector B has elements $n\bar{y}(n + k)^{-1}$, thus we see one would forecast w with a k-dimensional t distribution with $n + 2\alpha$ degrees of freedom, location vector $A^{-1}B$, and precision matrix $(n + 2\alpha)A(C - B'A^{-1}B)^{-1}$, where $C = 2\beta + \Sigma(y_i - \bar{y})^2 + n\bar{y}^2 - n^2\bar{y}^2(n + k)^{-1}$.

It is important to remember one's prediction of w is to be based on the whole of the predictive distribution, that is to say one has available an entire distribution with which to forecast w. A point forecast of w is given by $A^{-1}B$, the mean of the distribution and the confidence or the forecast error of this prediction is given by the dispersion matrix of w namely $A^{-1}(C - B'A^{-1}B)(n + 2\alpha - 2)^{-1}$. Note, the degrees of freedom do not depend on the number of observations to be predicted but on the number of observations in the sample and the prior parameter α.

An ad hoc way to make a forecast about w_1, one future observation, is to treat $w_1 \sim n(\theta, \tau)$ and estimate θ by $\bar{y}$ and τ^{-1} by $(n - 1)^{-1} \Sigma_1^n (y_i - \bar{y})^2$, then predict w_1 with $\bar{y}$, since $\bar{y}$ is the mean of the conditional distribution of w_1, given θ and τ.

The same prediction arises from the Bayesian method if one uses a Jeffreys' vague prior density for θ and τ, and uses the mean $\bar{y}$ of the predictive density of w_1 given s; however the ad hoc method is based on a $n(\theta, \tau)$ distribution, while the Bayesian forecast is based on a t distribution; but the latter is equivalent to the former, when the number of observations is large, in which case the t distribution approaches the normal.

In the more general case of a general linear model, how does one forecast future observations w, where

$$w = Z\theta + \varepsilon, \tag{1.38}$$

w is a $k \times 1$ vector of future observations, Z a known $k \times p$ matrix, θ the $p \times 1$ parameter vector of the linear model

$$Y = X\theta + e \tag{1.1}$$

and ε a $k \times 1$ vector of future errors?

The forecasting principle is the same here as in the previous example, one must find the conditional distribution of w given y, thus assuming ε and e are independent normal errors, where $\varepsilon \sim N(0, \tau^{-1}I_k)$

and using a normal-gamma prior density for θ and τ, namely (1.5), it is easy to derive the Bayesian predictive density of w. The previous example is a special case of forecasting w in the general linear model.

Suppose the prior density for θ and τ is the normal-gamma density (1.5) with hyperparameters $\alpha > 0$, $\beta > 0$, $\mu \in R^p$, and $\mathcal{P}$, a positive definite $p \times p$ matrix, then the reader may verify that the predictive density of w is

$$g(w|y) \propto \{[w - A^{-1}B]'A[w - A^{-1}B] + C - B'A^{-1}B\}^{(n+k+2\alpha)/2},$$

$$w \in R^k, \qquad (1.39)$$

where

$$A = I - Z(Z'Z + X'X + \mathcal{P})^{-1}Z',$$

$$B = Z(Z'Z + X'X + \mathcal{P})^{-1}(X'Y + \mathcal{P}\mu),$$

and

$$C = Y'Y + \mu'\mathcal{P}\mu - (X'Y + \mathcal{P}\mu)'(Z'Z + X'X + \mathcal{P})^{-1}(X'Y + \mathcal{P}\mu) + 2\beta.$$

Thus w has a k-dimensional t distribution with $n + 2\alpha$ degrees of freedom, location vector $A^{-1}B$, and precision matrix $(n + 2\alpha)A(C - B'A^{-1}B)^{-1}$, and it may be confirmed that this density reduces to the special case (1.37).

We have seen that if one has the t density with which to forecast w that one may use the mean $A^{-1}B$ as a point forecast of the k future values, but one may also want an interval or region prediction and some idea of the accuracy of the forecast.

Since w has a t distribution each component of w has a scalar t distribution (see the Appendix), hence to predict w_1 for example, one could construct a $1 - \gamma$ confidence interval for w_1 or one can perform simultaneous predictions for the k future observations by constructing a $1 - \gamma$ prediction region of w, with highest predictive density.

Let

$$Q(w) = (n + 2\alpha)(w - A^{-1}B)'A(w - A^{-1}B)(C - B'A^{-1}B)^{-1}/k,$$

$$(1.40)$$

then it may be shown that $Q(w)$ has a predictive distribution which is F with k and $n + 2\alpha$ degrees of freedom, and that $\{w: Q(w) \leqslant F(\gamma;k,n + 2\alpha)\}$ is a $1 - \gamma$ region for w with highest predictive density. Thus, regions of highest predictive density for w are constructed in the same way as regions of highest posterior density for θ. A future value of w will be contained in $Q(w)$ with a probability of $1 - \gamma$, giving one an idea of the error of the forecast.

The Bayesian predictive density will be used in a variety of ways, not only for forecasting future values, although this is its most important function. Another use of the predictive density is that it can be used to determine the parameters α, β, μ, and P of the prior density of the parameters θ and τ of the linear model. By employing the so-called prior predictive density, one may predict hypothetical future values w of the experimental values y, where $Y = X\theta + e$ is the model of the experiment, then fit the w values to the hyperparameters of the normal-gamma conjugate prior. These techniques will be introduced in the chapter on regression models and experimental design models.

Another use of the Bayesian predictive density is the control problem, which is encountered in regression and time series analysis and linear dynamic systems.

Consider a general linear model

$$Y = X_1\theta_1 + X_2\theta_2 + e \tag{1.41}$$

where X and θ have been partitioned into $X = (X_1,X_2)$ and $\theta = (\theta_1',\theta_2')'$, and suppose X_2 is a $n \times p_2$ matrix of p_2 control variables; that is variables that can be set at one's discretion. The control problem is to choose X_2 in such a way so that the future values of $w = T$, where T is some target value and $w = z_1\theta_1 + x_2\theta_2 + \varepsilon$.

The predictive density of w will depend on x_2, hence one may choose x_2 so that w in some sense is close to the target value. These ideas will be explored in the other chapters.

One may use a loss function in conjunction with the Bayesian predictive density. Suppose $L(w,w^*)$ is the loss in predicting w by w^*, then w^* can be chosen to minimize the expected loss with respect to the predictive distribution of w. This decision-theory way of forecasting will not be used in this book.

SUMMARY AND GUIDE TO THE LITERATURE

We have seen that this chapter introduces the Bayesian analysis of the general linear model. An analysis consists of identifying the model

$Y = X\theta + e$, specifying a prior distribution for the parameters θ and τ, either by Jeffreys' vague prior or by a normal-gamma conjugate prior, deriving the posterior densities of the parameters so one may estimate or test hypotheses about the parameters, and then, if one wishes, forecast future values by the Bayesian predictive density. The parameters are estimated with point estimates, which are particular characteristics of the posterior distribution or with interval estimates constructed by regions of highest posterior density, which also provide a way to test hypotheses about the parameters. In the same way, regions of highest predictive density give simultaneous forecasts of future values.

It has been shown that only five distributions occur in the study of linear models, namely the normal, gamma, normal-gamma, the multivariate t, and the F distributions. The normal distribution specifies the errors of the model and the conditional prior and posterior distributions of θ given τ, while the gamma is the marginal prior and posterior distribution of τ. The joint marginal prior and posterior distribution of θ and τ is normal-gamma, which is also the conjugate family of distributions to the normal linear model. Lastly, the F distribution occurs as a transformation of the t distribution and gives one a method to construct regions of highest posterior and predictive density for θ and w respectively, where w is a vector of future values. The t distribution is the marginal prior and posterior distribution of θ and also that of the predictive density of w.

These five distributions will suffice until the posterior analysis of the mixed model is developed. At that time, it will be necessary to introduce a generalization of the t distribution, called the poly-t distribution (see the Appendix).

An advantage of the Bayesian approach is that Bayes theorem provides a way to solve all problems in the analysis of the normal linear model. We have seen that the inferences for the parameters of the model are all based on the posterior distribution of the pertinent parameters and that these posterior distributions are a consequence of Bayes theorem. Forecasting is accomplished by the predictive distribution which is derived on the basis of Bayes theorem, hence if one knows how to apply Bayes theorem, one is well on the way to learning the theory and methodology of analyzing data which can be modeled in terms of linear models.

The reader should be familiar with some of the literature which deals with the Bayesian analysis of linear models.

First, there are several books which introduce the reader to general Bayesian inferential procedures as well as those methods which are peculiar to the linear model. The earliest book is by Jeffreys (1939) who introduced the vague prior distribution, which we use in this chapter. Later Lindley (1965) wrote a book which contains very useful information about the linear model, but prior information is restricted to improper vague prior distributions. Box and Tiao's (1973) book actually deals exclusively with all types of linear models including mixed and random models. Zellner's (1971) book is not only an excellent introduction to linear model theory, but also provides the Bayesian analysis of time series and econometric models, including simultaneous equation models. He also discusses hypothesis testing techniques and control theory problems. The above four books are not oriented to the Bayesian decision theory approach, but instead, focus on informal inferential procedures, which is the way this book treats the subject.

For a more formal decision-theory approach, one should refer to Savage (1972), Raiffa and Schlaifer (1961), DeGroot (1970), and Berger (1980). The Savage book is one of the earliest formulations to present the axiomatic development of decision theory, where one postulates axioms of utility and probability, and as a consequence, proves that Bayesian decisions are the only optimal ones. This tradition is continued by Ferguson (1967), DeGroot (1970), and Berger (1980), and the former is a good introduction to Bayesian decision theory as well as other theories of decision. DeGroot (1970) has a short section on the theory of linear models and good accounts of conjugate family distributions and the distribution theory which is involved in the posterior and predictive analysis of linear models. This book is highly recommended to the reader.

In order to learn more about the implementation of prior information, the reader should consult Bernardo (1979), Jeffreys (1939), Box and Tiao (1973), Winkler (1977), and Zellner (1971, 1980). The Winkler and Zellner (1980) papers are about choosing the hyperparameters of the normal-gamma prior distribution, while the remaining references deal with improper prior distributions, which convey little or no information.

With regard to the posterior analysis of linear models, Box and Tiao's (1965) paper on regions of highest posterior density shows how to test hypotheses about the means and variances of normal theory models and their book (1973) is by far the best account of Bayesian methodology as applied to the linear model. Also, this book is the only one which studies random and mixed models and it gives many numerical examples of the analysis of all types of models. Needless to say, this book is highly recommended to those who want to know the detailed analysis of the traditional statistical models, that is regression models and models (fixed and random) employed to analyze designed experiments.

Both Box and Tiao (1973) and Zellner (1971) use regions of highest posterior density to estimate parameters and test hypotheses, but the latter specializes in posterior odds ratios to test hypotheses. For forecasting problems in economics, Zellner's (1971) book is the only one to give a detailed account of the Bayesian treatment of time series and econometric models.

Other Bayesian references on predictive analysis are Aitchison and Dunsmore (1975), Harrison and Stevens (1979) and Geisser (1971).

The controversy surrounding the Bayesian approach has somewhat subsided and it is interesting to read some of the earlier discussions on the subject. For example, Lindley and Smith's (1972) paper on exchangeable prior distributions provoked a heated debate about Bayesian methodology, but for a more balanced approach to the comparison of Bayesian and non-Bayesian theories of inference, read Barnett (1973) and Lindley's (1971) monograph.

Lindley's monograph is a review of Bayesian theory and methodology written up to 1971, and although much material has appeared since then, the serious student cannot afford not to read it. This reference contains criticism of some sampling theory methodologies, such as the Neyman-Pearson theory of testing hypotheses and the confidence interval approach to estimation, and Barnett presents an unbiased comparison of Bayesian and other theories of inference.

EXERCISES

1. Prove Theorem 1.1.
2. Referring to Theorem 1.1, show that if $X'X$ is singular and if the prior density of the parameters is vague (1.20), then the joint posterior density of the parameters is improper.
3. Verify equation (1.28) and show (1.29) yields a $1 - \gamma$ HPD region for θ.
4. Show the region given by (1.32) is a $1 - \gamma$ HPD region for $A\theta$.
5. Show the random variable $Q(w)$, equation (1.40), has an F distribution with k and $n + 2\alpha$ degrees of freedom.
6. Assume $k = 2$ in problem 5 above and find an interval of highest predictive density for the second future observation.
7. The general linear model is given by (1.1). How would you represent a normal-theory simple linear regression model, that is, explain what Y, X, θ, and e are? Now, using the notation of the linear model (1.1), define a normal theory multiple regression model with $p - 1$ regressors.
8. Compare the classical and non-Bayesian approaches to the analysis of a linear model. For example, compare the way a Bayesian would estimate θ to the way a non-Bayesian would do it. What are the similarities and differences in the two approaches? How would a non-Bayesian predict future observations assuming the future observations are generated with the general linear model?
9. Let

$$Y = X\theta + e$$

be a linear model with the usual assumptions, except $e \sim N(0, P^{-1})$, where P is a $n \times n$ unknown precision matrix. Assume θ is a $p \times 1$ unknown parameter vector and that θ and P have a normal-Wishart distribution. (See the Appendix.) Find the marginal posterior distribution of θ.

10. Let

$$Y = X\theta + e$$

be a linear model with the usual assumptions where $e \sim N(0, P^{-1})$ and P is a $n \times n$ *known* precision matrix. Using a conjugate prior density for θ, find the marginal posterior distribution of θ.

11. Let

$$Y = X\theta + e$$

with the usual assumptions except that θ is known and $e \sim N(0, \tau^{-1} I_n)$, and $\tau > 0$ is an unknown precision with a prior gamma density. Find the posterior density of τ.

REFERENCES

Aitchison, J. S. and J. R. Dunsmore (1975). *Statistical Prediction Analysis*, Cambridge University Press, London.

Barnett, V. (1973). *Comparative Statistical Inference*, John Wiley and Sons, Inc., New York.

Berger, James O. (1980). *Statistical Decision Theory, Foundations, Concepts, Methods*, Springer-Verlag, New York.

Bernardo, J. M. (1979). "Reference posterior distributions for Bayesian inference" (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 41, No. 2, pp. 113-147.

Box, G. E. P. and G. C. Tiao (1965). "Multiparameter problems from a Bayesian point of view," Annals of Mathematical Statistics, Vol. 36, pp. 1468-1482.

Box, G. E. P., and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*, Addison Wesley, Reading, Mass.

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

Ferguson, Thomas S. (1967). *Mathematical Statistics, A Decision Theoretic Approach*, Academic Press, New York and London.

Geisser, S. (1971). "The inferential use of predictive distributions," in *Foundations of Statistical Inference*, edited by V. P. Godambe and A. D. Sprott, Holt, Rinehart and Winston, Toronto.

Harrison, P. J. and C. F. Stevens (1976). "Bayesian forecasting" (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 38, pp. 205-247.

Jeffreys, H. (1939/1967). *Theory of Probability*, Third Edition, Clarendon Press, Oxford.
Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, University Press, Cambridge.
Lindley, D. V. (1971). *Bayesian Statistics, A Review*, Reg. Conf. Ser. Appl. Math., 2. SIAM, Philadelphia.
Lindley, D. V. and A. F. M. Smith (1972). "Bayes estimates for the linear model," Journal of the Royal Statistical Society, Series B, Vol. 34, pp. 1-42.
Raiffa, Howard and Robert Schlaifer (1961). *Applied Statistical Decision Theory*, Division of Research, Harvard Business School, Boston.
Savage, L. J. (1972). *The Foundation of Statistics*, Dover, Inc., New York.
Winkler, Robert, L. (1977). "Prior distributions and model building in regression analysis," in *New Developments in the Applications of Bayesian Methods*, edited by Ahmet Aykac and Carlo Brumat, North-Holland, Amsterdam.
Zellner, Arnold (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, Inc., New York.
Zellner, Arnold (1980). "On Bayesian regression analysis with g-prior distributions." Technical Reprint, H. G. B. Alexander Research Foundation, Graduate School of Business, University of Chicago.

Bayesian Inference for the General Linear Model

Aitchison, J. S. and J. R. Dunsmore (1975). Statistical Prediction Analysis, Cambridge University Press, London.

Barnett, V. (1973). Comparative Statistical Inference, John Wiley and Sons, Inc., New York.

Berger, James O. (1980). Statistical Decision Theory, Foundations, Concepts, Methods, Springer-Verlag, New York.

Bernardo, J. M. (1979). “Reference posterior distributions for Bayesian inference” (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 41, No. 2, pp. 113–147.

Box, G. E. P. and G. C. Tiao (1965). “Multiparameter problems from a Bayesian point of view,” Annals of Mathematical Statistics, Vol. 36, pp. 1468–1482.

Box, G. E. P. , and G. C. Tiao (1973). Bayesian Inference in Statistical Analysis, Addison Wesley, Reading, Mass.

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

Ferguson, Thomas S. (1967). Mathematical Statistics, A Decision Theoretic Approach, Academic Press, New York and London.

Geisser, S. (1971). “The inferential use of predictive distributions,” in Foundations of Statistical Inference, edited by V. P. Godambe and A. D. Sprott , Holt, Rinehart and Winston, Toronto.

Harrison, P. J. and C. F. Stevens (1976). “Bayesian forecasting” (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 38, pp. 205–247.

Jeffreys, H. (1939/1967). Theory of Probability, Third Edition, Clarendon Press, Oxford.

Lindley, D. V. (1965). Introduction to Probability and Statistics from a Bayesian Viewpoint, University Press, Cambridge.

Lindley, D. V. (1971). Bayesian Statistics, A Review, Reg. Conf. Ser. Appl. Math., 2. SIAM, Philadelphia.

Lindley, D. V. and A. F. M. Smith (1972). “Bayes estimates for the linear model,” Journal of the Royal Statistical Society, Series B, Vol. 34, pp. 1–42.

Raiffa, Howard and Robert Schlaifer (1961). Applied Statistical Decision Theory, Division of Research, Harvard Business School, Boston.

Savage, L. J. (1972). The Foundation of Statistics, Dover, Inc., New York.

Winkler, Robert L. (1977). “Prior distributions and model building in regression analysis,” in New Developments in the Applications of Bayesian Methods, edited by Ahmet Aykac and Carlo Brumat , North-Holland, Amsterdam.

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

Zellner, Arnold (1980). “On Bayesian regression analysis with gprior distributions.” Technical Reprint, H. G. B. Alexander Research Foundation, Graduate School of Business, University of Chicago.

Linear Statistical Models and Bayesian Inference

Aykac, Ahmet and Carlo Brumat (1977). New Developments in the Applications of Bayesian Methods. North-Holland Publishing Company, New York.

Barnett, Vic (1982). Comparative Statistical Inference, second edition. John Wiley and Sons, New York.

Bayes, T. (1963). “An essay toward solving a problem in the doctrine of chances,” Philosophical Transactions Royal Society, Vol. 53, pp. 320–418.

Berger, James O. (1980). Statistical Decision Theory, Foundations, Concepts, and Methods. Springer-Verlag, New York.

Bernardo, J. M. (1980). “A Bayesian analysis of classical hypothesis testing,” in Bayesian Statistics: Proceedings of the First International Meeting, Valencia, 1979, edited by Bernardo , DeGroot , Lindley and Smith .

Bernoulli, J. (1713). Ars Conjectandi. Basel.

Box, G. E. P. (1980). “Sampling and Bayes inference in scientific modeling and robustness” (with discussion), Journal of the Royal Statistical Society, Series A, Vol. 143, part 4, pp. 383–430.

Box, G. E. P. and Gwilym M. Jenkins (1970). Time Series Analysis, Forecasting, and Control, Holden-Day, San Francisco.
Box, G. E. P. and G. C. Tiao (1973). Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Ma.
Box, G. E. P. and G. C. Tiao (1981). “Modeling multiple time series,” Journal of the American Statistical Association, Vol. 76, No. 376, pp. 802–816.
Broemeling, Lyle (1982). “Introduction,” Journal of Econometrics, Vol. 19, pp. 1–5.
Davies, O. L. (1967). Statistical Methods in Research and Production, third edition, Oliver and Boyd, London.
Dawid, A. P. , Stone, M. and J. W. Zedek (1973). “Marginalization paradoxes in Bayesian structural inference” (with discussion), Journal of the Royal Statistical Society, B, Vol. 35, pp. 189–233.
deFinetti, B. (1930). “Sulla proprieti conglomerativa delle probabilita subordinate,” Rend R. Inst. Lombardo (Milano), Vol. 63, pp. 414–418.
deFinetti, B. (1937). “Foresight: its logical laws, its subjective sources,” Annals de l’Institute Henri Poincare. Reprinted (in translation) in Kyburg and Smokler (1964).
DeGroot, M. H. (1970). Optimal Statistical Decisions, McGraw-Hill, New York.
DeMorgan, August (1847). Formal Logic, Taylor and Walton, London.
Fienberg, S. E. and A. Zellner (1975). Studies in Bayesian Econometrics and Statistics. North-Holland Publishing Co., New York.
Fisher, R. A. (1922). “On the mathematical foundations of theoretical statistics,” Phil. Transactions Roy. Soc. London, A, Vol. 222, pp. 309–362.
Ferguson, Thomas, S. (1967). Mathematical Statistics, A Decision Theoretic Approach, Academic Press, New York.
Geisser, S. (1980). “A predictivistic primer,” in Bayesian Analysis in Econometrics and Statistics, edited by A. Zellner , North-Holland, Amsterdam.
Good, I. J. (1950). Probability and the Weighing of Evidence, Griffin, London.
Holbert, Don (1982). “A Bayesian analysis of a switching linear model,” Journal of Econometrics, Vol. 19, pp. 77–87.
Jaynes, Edwin T. (1980). “Marginalization and prior probabilities,” in Bayesian Analysis in Econometrics and Statistics, edited by A. Zellner , North Holland Publishing Co., Amsterdam.
Jazwinski, Andrew H. (1970). Stochastic Processes and Filtering Theory, Academic Press, New York.
Jeffreys, Harold (1939/1961). Theory of Probability, Oxford at the Clarendon Press, first and third editions.
Jenkins, Gwilym M. (1979). Practical Experiences with Modeling and Forecasting Time Series, Gwilym Jenkins and Partners (overseas) Ltd.
Kadane, J. B. (1980). “Predictive and structural methods for eliciting prior distributions,” in Bayesian Analysis in Econometrics and Statistics, edited by A. Zellner , North - Holland, Amsterdam.
Kadane, J. B. , J. M. Dickey , R. I. Winkler , W. S. Smith , and S. C. Peters (1980). “Interactive elicitation of opinion for a normal linear model,” Journal of American Statistical Association, Vol. 75, pp. 845–854.
Kalman, R. E. (1960). “A new approach to linear filtering and prediction problems,” Transactions of ASME, Series D, Journal of Basic Engineering, Vol. 83, pp. 95–108.
Kyburg, Jr. Henry E. , (1970). Probability and Inductive Logic. The Macmillan Company, Collier-Macmillan Limited, London.
Kyburg, Jr. H. E. , and Smokier, H. E. (eds.) (1964). Studies in Subjective Probability, John Wiley and Sons, New York.
de Laplace, P. S. (1951). A Philosophical Essay on Probabilities. (An English translation by Truscott and Emory of 1820 edition.) Dover, New York.
Leonard, T. (1972). “Bayesian methods for binomial data,” Biometrika, Vol. 59, pp. 581–589.
Lindley, D. V. (1965). Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2, Inference. Cambridge University Press.
Lindley, D. V. (1971). Making Decisions, Wiley-Interscience, John Wiley and Sons Ltd., London.
Lindley, D. V. and A. F. M. Smith (1972). “Bayes estimates for the linear model,” Journal of the Royal Statistical Society, B, Vol. 34, pp. 1–42.

McGee, V. E. and W. T. Carleton (1970). “Piecewise regression,” Journal of the American Statistical Association, Vol. 65, pp. 1109–1124.
Monahan, J. F. (1983). “Fully Bayesian analysis of ARMA time series models,” Journal of Econometrics, Vol. 21, pp. 307–331.
Pearson, K. (1920). “Note on the fundamental problem of practical statistics,” Biometrika, Vol. 13, pp. 300–301.
Poirier, D. J. (1976). The Econometrics of Structural Change, North-Holland Publishing Company, Amsterdam.
Press, S. James (1982). Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference, second edition, Robert E. Krieger Publishing Company, Malabar, Florida.
Raiffa, H. and R. Schlaifer (1961). Applied Statistical Decision Theory, Division of Research, Graduate School of Business Administration, Harvard University, Cambridge, Massachusetts.
Rajagopalan, Muthiya (1980). Bayesian Inference for the Variance Components in Mixed Linear Models. Ph.D. dissertation, December, 1980, Oklahoma State University, Stillwater, Oklahoma.
Ramsey, F. P. (1931/1964). “Truth and probability,” in Kyburg and Smokier (1964), Reprinted from the Foundations of Mathematics and other Essays, Kagan, Paul, Trench, London, 1931.
Savage, L. J. (1954). The Foundation of Statistics. John Wiley and Sons, Inc., New York.
Stigler, Stephen M. (1982). “Thomas Bayes and Bayesian inference,” Journal of the Royal Statistical Society, A, Vol. 145, part 2, pp. 250–258.
Winkler, R. L. (1972). Introduction to Bayesian Inference and Decision, Holt, Rinehart and Winston, New York.
Winkler, R. L. (1980). “Prior distributions and model building in regression analysis,” in New Developments in the Applications of Bayesian Methods, edited by Aykac and Brumat , North-Holland Publishing Company, Amsterdam.
Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, Inc., New York.
Zellner, Arnold (1980). Bayesian Analysis in Econometrics and Statistics, Essays in Honor of Harold Jeffreys, North-Holland Publishing Company, New York.

The Traditional Unear Models

Box, G. E. P. and G. C. Tiao (1973). Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Mass.
Cochran, W. G. and G. M. Cox (1957). Experimental Designs, 2nd ed., John Wiley and Sons, New York.
Cox, D. R. (1958). Planning of Experiments, John Wiley and Sons, New York.
de Finetti, Bruno (1972). Probability, Induction, and Statistics, the Art of Guessing, John Wiley and Sons, London.
Draper, N. R. and H. Smith (1966). Applied Regression Analysis, John Wiley and Sons, Inc., New York.
Fisher, R. A. (1951). Design of Experiments, 6th ed., Oliver and Boyd, Edinburgh.
Graybill, F. A. (1961). An Introduction to Linear Statistical Models, McGraw Hill, New York.
Kadane, Joseph B. , James M. Dickey , Robert L. Winkler , Wayne S. Smith and Stephen C. Peters (1980). “Interactive elicitation of opinion for a normal linear model,” Journal of American Statistical Association, Vol. 75, pp. 845–854.
Kempthorne, O. (1952). Design and Analysis of Experiments, John Wiley and Sons, Inc., New York.
Lindley, D. V. (1965). Introduction to Probability and Statistics from a Bayesian Viewpoint, University Press, Cambridge.
Lindley, D. V. (1968). “The choice of variables in multiple regression,” Journal of the Royal Statistical Society, Series B, Vol. 30, pp. 31–66.
Lindley, D. V. and A. F. M. Smith (1972). “Bayes estimates for the linear model,” Journal of the Royal Statistical Society, Series B, Vol. 34, pp. 1–42.
Patil, V. H. (1964). “The Behrens-Fisher problem and its Bayesian solution,” Journal of the Indian Statistical Association, Vol. 2, pp. 21–31.

Scheffe, Henry (1943). "On solutions to the Behrens-Fisher problem based on the t-distribution," Annals of Mathematical Statistics, Vol. 14, pp. 35–44.
Searle, S. R. (1971). Linear Models, John Wiley and Sons, Inc., New York.
Sedory, S. A. (1981). Pursuing the Poly-t. Unpublished Masters Report, Department of Statistics, Oklahoma State University.
Winkler, Robert L. (1977). "Prior distributions and model building in regression analysis," in New Developments in the Applications of Bayesian Methods, edited by Ahmet Aykac and Carlo Baumat , North-Holland, Amsterdam.
Yusoff, Mat Abdullah (1982). Bayesian Inference with the Poly-t Distribution. Ph.D. dissertation, Oklahoma State University.
Zellner, Arnold (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, Inc., New York.
Zellner, Arnold (1980). "On Bayesian regression analysis with g prior distributions," Technical reprint, H. G. B. Alexander Research Foundation, Graduate School of Business, University of Chicago.

The Mixed Model

Box, G. E. P. and G. C. Tiao (1973). Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Massachusetts.
Daniels, H. E. (1939). "The estimation of components of variance," Journal of the Royal Statistical Society (Supp.), Vol. 6, pp. 186–197.
Dreze, Jacque H. (1977). "Bayesian regression analysis using poly-t-distributions," from New Developments in the Application of Bayesian Methods, edited by Aykac and Brumat , North-Holland Publishing Co., Amsterdam.
Gharrah, M. K. (1979). A General Solution to Making Inferences about the Parameters of Mixed Linear Models. Ph.D. dissertation, Oklahoma State University.
Goel, Prem K. and Morris H. DeGroot (1981). "Information about hyperparameters in hierarchical models," Journal of the American Statistical Association, Vol. 76, pp. 140–147.
Graybill, F. A. (1954). "On quadratic estimates of variance components," Annals of Math. Statistics, Vol. 25, pp. 367–372.
Graybill, F. A. , and R. A. Hultquist (1961). "Theorems on Eisenhart's model II," Annals of Math Statistics, Vol. 32, pp. 261–269.
Graybill, F. A. and A. W. Wortham (1956). "A note on uniformly best unbiased estimators for variance components," Journal of the American Statistical Association, Vol. 51, pp. 266–268.
Hartley, H. O. and J. N. K. Rao (1967). "Maximum likelihood estimation for the mixed analysis of variance model," Biometrika, Vol. 54, pp. 93–108.
Hemmerle, W. J. and J. A. Lorens (1976). "Improved algorithm for the W-transform in variance component estimation," Technometrics, Vol. 18, No. 2, pp. 207–212.
Henderson, C. R. (1953). "Estimation of variance and covariance components," Biometrics, Vol. 9, pp. 226–252.
Hill, B. M. (1965). "Inference about variance components in the oneway model," Journal of the American Statistical Association, Vol. 60, pp. 806–825.
Hill, B. M. (1967). "Correlated errors in the random model," Journal of the American Statistical Association, Vol. 62, pp. 1387–1400.
Lindley, D. V. and A. F. M. Smith (1972). "Bayes estimates for the linear model," Journal of the Royal Statistical Society, Series B, Vol. 34, pp. 1–18.
Rajagopalan, M. (1980). "Bayesian inference for the variance components in mixed linear models," Ph.D. dissertation, Oklahoma State University, Stillwater, Oklahoma.
Rajagopalan, M. and Lyle Broemeling (1983). "Bayesian inference for the variance components of the general mixed models," Communications in Statistics, Vol. 12, No. 6, pp. 701–724.
Ruben, H. (1960). "Probability content of regions under spherical normal distributions," Annals of Math. Statistics, Vol. 31, pp. 598–619.
Ruben, H. (1962). "Probability contents of regions under spherical normal distributions, IV: the distribution of homogeneous and non-homogeneous quadratic function of normal variables," Annals of Math. Statistics, Vol. 33, pp. 542–570.

Searle, S. R. (1971). Linear Models, John Wiley and Sons, Inc., New York.
Searle, S. R. (1978). “A summary of recently developed methods of estimating variance components,” Technical Report, BU-338, Biometrics Unit, Cornell University, New York.
Smith, A. F. M. (1973). “Bayes estimates in one-way and two-way models,” Biometrika, Vol. 60, pp. 319–329.
Stone, M. and B. G. F. Springer (1965). “A paradox involving quasi prior distributions,” Biometrika, Vol. 52, p. 623.
Winsor, C. P. and G. L. Clarke (1940). “A statistical study of variation in the catch of plankton nets,” Sears Foundation Journal of Marine Research, Vol. 3, pp. 1–34.

Time Series Models

Abraham, Bovas and William W.S. Wei (1979). “Inferences in a switching time series,” American Statistical Association Proceedings of the Business and Economic Statistics Section, pp. 354–358.
Akaike, H. (1979). “A Bayesian extension of the minimum AIC procedure of autoregressive model building,” Biometrika, Vol. 66, No. 2, pp. 237–242.
Ali, M. M. (1977). “Analysis of autoregressive moving average models: estimation and prediction,” Biometrika, Vol. 64, pp. 535–545.
Ansley, C. F. (1979). “An algorithm for the exact likelihood of mixed autoregressive moving average process,” Biometrika, Vol. 66, pp. 59–65.
Aoki, Masanao (1967). Optimization of Stochastic Systems, Topics in Discrete Time Systems, Academic Press, New York.
Box, George E. P. and Gwilym M. Jenkins (1970). Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco.
Broemeling, Lyle D. (1977). “Forecasting future values of changing sequences,” Communications in Statistics, Theory and Methods, Vol. A6(1), pp. 87–102.
Chatfield, C. (1977). “Some recent developments in time series analysis,” Journal of the Royal Statistical Society, Series A, Vol. 140, pp. 492–510.
Chin Choy, J. H. and L. D. Broemeling (1980). “Some Bayesian inferences for a changing linear model,” Technometrics, Vol. 22, No. 1, pp. 71–78.
Current Index to Statistics, Appplications, Methods, and Theory, Volume 5, 1979 (1980).
Dent, W. T. (1977). “Computation of the exact likelihood function for an ARIMA process,” Journal of Statistical Computations and Simulation, Vol. 5, pp. 193–206.
Diaz, Jaoquin and Jose Luis Farah (1981). “Bayesian identification of autoregressive processes,” 22nd NBER-NSF Seminar on Bayesian Inference in Econometrics.
Granger, C. W. J. and Paul Newbold (1977). Forecasting Economic Time Series, Academic Press, New York.
Haavelmo, T. (1953). “Methods of measuring the marginal propensity to consume,” Journal of the American Statistical Assoc., Vol. 42, pp. 105–122.
Harrison, P. J. and Stevens, C. F. (1976). “Bayesian forecasting” (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 38, pp. 205–247.
Hillmer, Steven C. , and George C. Tiao (1979). “Likelihood function of stationary multiple autoregressive moving average models,” Journal of the American Statistical Association, Vol. 74, No. 367, pp. 652–660.
Lahiff, Maureen (1980). “Time series forecasting with informative prior distributions,” Technical Report No. 111, Department of Statistics, The University of Chicago.
Land, Margaret Foster (1981). Bayesian Forecasting for Switching Regression and Autoregressive Processes, Ph.D. dissertation, Oklahoma State University, Stillwater, Oklahoma.
Lindley, D. V. and A. F. M. Smith (1972). “Bayes estimates for the linear model,” Journal of the Royal Statistical Society, Series B, Vol. 34, pp. 1–18.
Litterman, Robert B. (1980). “A Bayesian procedure for forecasting with vector autoregressions,” An unpublished photocopy paper, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Ljung, G. M. and G. E. P. Box (1976). “Studies in the modeling of discrete time series—maximum likelihood estimation in the autoregressive moving average model,” Technical Report No. 476, Department of Statistics, University of Wisconsin, Madison.
Monahan, John F. (1981). “Computations for Bayesian time series analysis,” a paper presented at the 141st Annual Meeting of the American Statistical Association, August 12, 1981.
Newbold, Paul (1973). “Bayesian estimation of Box Jenkins transfer function noise models,” Journal of the Royal Statistical Society, Series B, Vol. 35, No. 2, pp. 323–336.
Newbold, Paul (1974). “The exact likelihood function for a mixed autoregressive moving average process,” Biometrika, Vol. 61, pp. 423–426.
Nicholls, D. F. and A. D. Hall (1979). “The exact likelihood function of multivariate autoregressive moving average model,” Biometrika, Vol. 66, No. 2, pp. 259–264.
Perterka, V. (1981). “Bayesian system identification,” Automatica, Vol. 17, No. 1, pp. 41–53.
Poirier, Dale J. (1976). The Econometrics of Structural Change, North-Holland Publishing Company, Amsterdam.
Salazar, D. , Broemeling, L. , and A. Chi (1981). “Parameter changes in a regression model with autocorrelated errors,” Communications in Statistics, Part A—Theory and Methods, Vol. A10, Number 17, pp. 1751–1758.
Smith, J. O. (1979). “A generalization of the Bayesian steady forecasting model,” Journal of the Royal Statistical Society, Series B, Vol. 41, No. 3, pp. 375–387.
Tsurumi, H. (1980). “A Bayesian estimation of structural shifts by gradual switching regression with an application to the US gasoline market,” in Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys, edited by A. Zellner , North-Holland, Amsterdam.
Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, Inc., New York.

Linear Dynamic Systems

Suppose u(l),u(2),…; V(1),V(2), … are independent and develop a test of the null hypothesis θ2(t) = 0, t = 1,2, …, k, where k is a fixed positive integer.

Alspach, D.L. (1974). “A parallel filtering algorithm for linear systems with unknown time varying noise statistics,” IEEE Transactions on Automatic Control, Vol. AC-19, 555–556.
Aoki, Masanao (1967). Optimization of Stochastic Systems, Topics in Discrete Time Series, Academic Press, New York, London.
Box, G. E. P. and G. C. Tiao (1973). Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Mass.
Downing, D.J. , D. H. Pike , and G. W. Morrison (1980). “Application of the Kalman filter to inventory control,” Technometrics, Vol. 22, No. 1, pp. 17–22.
Dreze, Jacques H. (1977). “Bayesian regression analysis using poly-t densities,” in New Developments in the Application of Bayesian Methods, edited by Ahmet Aykac and Carlo Bremat , North-Holland Publishing Company, Amsterdam.
Harrison, P. J. and C. F. Stevens (1976). “Bayesian forecasting” (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 38, 205–247.
Hawkes, R. M. (1973). Demodulation of Pulse Modulated Signals Using Adaptive Estimation, unpublished master’s thesis, University of Newcastle.
Jazwinski, Andrew H. (1970). Stochastic Processes and Filtering Theory, Academic Press, New York.
Kailath, T. (1968). “An innovations approach to least squares estimation–Part I: Linear filtering in addative white noise,” IEEE Transactions on Automatic Control, Vol. AC-13, 639–645.
Kalman, R. E. (1960). “A new approach to linear filtering and prediction problems,” Transactions, ASME, J. of Basic Engineering, Vol. 82-D, pp. 34–45.
Lainiotis, D. G. (1971). “Optimal adaptive estimation: structure and parameter adaptation,” IEEE Transactions on Automatic Control, Vol. AC-16, pp. 160–170.
Maybeck, Peter S. (1979). Stochastic Models, Estimation, and Control, Volumes I and II, Academic Press, New York.
Mehira, R. A. (1970). “Approaches to adaptive filtering,” IEEE Transactions on Automatic Control, Vol. AC-17, pp. 693–698.

Mehira, Raman K. (1970). “On the identification of variances and adaptive Kalman filtering,” IEEE Transactions on Automatic Control, Vol. AC-15, pp. 175–184.
Plackett, R. L. (1950). “Some theorems in least squares,” Biometrika, Vol. 37, pp. 149–157.
Press, James S. (1972). Applied. Multivariate Analysis, Holt, Rinehart, and Winston, Inc., New York.
Sallas, William M. and David A. Harville (1981). “Best linear recursive estimation for mixed linear models,” Journal of American Statistical Association, Vol. 76, pp. 860–869.
Sawaragi, Yashikazu , Yoshifumi Sunaharo , and Takayoshi Nakamizo (1967). Statistical Decision Theory in Adaptive Control Systems, Academic Press, New York, London.
Shumway, R. H. , D. E. Olsen and L. J. Levy (1981). “Estimation and tests of hypotheses for the initial mean and covariance in the Kalman filter model,” Communications in Statistics, Part A–Theory and Methods, Vol. A10, No. 16, pp. 1625–1642.
Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, Inc., New York.

Structural Change in Linear Models

Bacon, D. W. and D. G. Watts (1971). “Estimating the transition between two intersecting lines,” Biometrika, Vol. 58, pp. 524–534.
Bhattacharyya, G. K. and B. A. Johnson (1968). “Non parametric tests for shift at an unknown time point,” Annals of Mathematical Statistics, Vol. 39, pp. 1731–1734.
Box, G. E. P. and G. C. Tiao (1965). “Multiparameter problems from a Bayesian point of view,” Annals of Mathematical Statistics, Vol. 36, pp. 1468–1482.
Box, G. E. P. and G. C. Tiao (1973). Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Mass.
Broemeling, L. D. (1972). “Bayesian procedures for detecting a change in a sequence of random variables,” Metron, Vol. XXX-N-1-4, pp. 1–14.
Broemeling, L. D. (1974). “Bayesian inference about a changing sequence of random variables,” Communications in Statistics, Vol. 3 (3), pp. 243–255.
Broemeling, L. D. (1977). “Forecasting future values of a changing sequence,” Communications in Statistics, Vol. A6(1), pp. 87–102.
Broemeling, L. D. and J. H. Chin Choy (1981). “Detecting structural change in linear models,” Communications in Statistics, Vol. A10(24), pp. 2551–2562.
Brown, R. L. , J. Durbin , and J. M. Evans (1975). “Techniques for testing the constancy of regression relations over time” (with discussion), Journal of the Royal Statistical Society, Series B, pp. 49–92.
Chernoff, H. and S. Zacks (1964). “Estimating the current mean of a normal distribution which is subjected to changes over time,” Annals of Mathematical Statistics, Vol. 35, pp. 999–1018.
Chi, Albert Yu-Ming (1979). The Bayesian Analysis of Structural Change in Linear Models, doctoral dissertation, Oklahoma State University, Stillwater, Oklahoma.
Chin Choy, J. H. (1977). A Bayesian Analysis of a Changing Linear Model, doctoral dissertation, Oklahoma State University, Stillwater, Oklahoma.
Chin Choy, J. H. and L. D. Broemeling (1980). “Some Bayesian inferences for a changing linear model,” Technometrics, Vol. 22(1), pp. 71–78.
Chow, G. (1960). “Tests of the equality between two sets of coefficients in two linear regressions,” Econometrica, Vol. 28, pp. 591–605.
Farley, J. U. , M. J. Hinich , and T. W. Mcguire (1975). “Some comparisons of tests for a shift in the slopes of a multivariate linear time series,” J. of Econometrics, Vol. 3, pp. 297–318.
Feder, P. (1975). “The log likelihood ratio in segmented regression,” Annals of Statistics, Vol. 3, pp. 84–97.
Ferreira, R. E. (1975). “A Bayesian analysis of a switching regression model: known number of regimes,” Journal of the American Statistical Association, Vol. 70, pp. 370–374.
Goldfield, S. M. and R. E. Quandt (1973a). “The estimation of structural shifts by switching regressions,” Annals of Economic and Social Measurement, Vol. 2, pp. 475–485.
Goldfield, S. M. and R. E. Quandt (1973b). “A Markov model for switching regressions,” Journal of Econometrics, Vol. 1, pp. 3–16.

Goldfield, S. M. and R. E. Quandt (1974). “Estimation in a disequilibrium model and the value of information,” unpublished manuscript.
Green, C. V. and E. Fekette (1933). “Differential growth in the mouse,” J. Experimental Zoology, Vol. 66, pp. 351–370.
Hartley, M. J. and P. Mallela (1975). “The asymptotic properties of a maximum likelihood estimator for a model of markets in disequilibrium,” Discussion paper No. 329, State University of New York at Buffalo.
Hinkley, D. V. (1969). “Inference about the intersection in two- phase regression,” Biometrika, Vol. 56, pp. 495–504.
Hinkley, D. V. (1971). “Inference in two-phase regression,” J. of American Statistical Association, Vol. 66, pp. 736–793.
Holbert, D. and L. D. Broemeling (1977). “Bayesian inference related to shifting sequences and two phase regression,” Communications in Statistics, Vol. A6(3), pp. 265–275.
Hudson, D. (1966). “Fitting the segmented curves whose endpoints have to be estimated,” Journal of the American Statistical Association, Vol. 61, pp. 1097–1109.
Kander, Z. and S. Zacks (1966). “Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points,” Annals of Mathematical Statistics, Vol. 37, pp. 1196–1210.
Lindley, D. V. (1965). Introduction to Probability and Statistics from a Bayesian Viewpoint, Cambridge University Press, Cambridge.
Madala, G. S. and F. D. Nelson (1974). “Maximum likelihood methods for models of markets in disequilibrium,” Econometrica, Vol. 42, pp. 1013–1030.
Needham, A. E. (1935). “On relative growth in the jaws of certain fishes,” Proceedings of Zoological Society of London, Vol. 2, pp. 773–784.
Ohki, K. (1974). “Manganese nutrition of cotton under two boron levels. II. Critical M level,” Agronomy Journal, Vol. 66, pp. 572–575.
Page, E. S. (1954). “Continuous inspection schemes,” Biometrika, Vol. 41, pp. 100–114.
Page, E. S. (1957). “On problems in which a change in parameter occurs at an unknown point,” Biometrika, Vol. 44, pp. 248–252.
Poirier, D. J. (1976). The Econometrics of Structural Change, North-Holland Publishing Company, Amsterdam.
Pool, J. and C. F. Borchgrevinck (1964). “Comparison of rat liver response to coumarin administered in vivo versus in vitro,” American Journal of Physiology, Vol. 206(1), pp. 229–238.
Quandt, R. E. (1958). “The estimation of the parameters of a linear regression system obeying two separate regimes,” Journal of the American Statistical Association, Vol. 53, pp. 873–880.
Quandt, R. E. (1960). “Tests of hypotheses that a linear regression system obeys two separate regimes,” Journal of the American Statistical Association, Vol. 55, pp. 324–330.
Quandt, R. E. (1972). “A new approach to estimating switching regressions,” Journal of the American Statistical Association, Vol. 67, pp. 306–310.
Quandt, R. E. (1974). “A comparison of methods for testing nonnested hypotheses,” Review of Economics and Statistics, Vol. 56, pp. 92–99.
Salazar, Diego (1980). The Analysis of Structural Changes in Time Series and Multivariate Linear Models, doctoral dissertation, Oklahoma State University, Stillwater, Oklahoma.
Salazar, Diego , L. Broemeling , and A. Chi (1981). “Parameter changes in a regression model with autocorrelated errors,” Communications in Statistics, Vol. A10(17), pp. 1751–1758.
Sen, A. K. and M. S. Srivastava (1973). “On multivariate tests for detecting change in mean,” Sankhya, Vol. A35, pp. 173–185.
Sen, A. K. and M. S. Srivastava (1975a). “On tests for detecting a change in mean,” Annals of Statistics, Vol. 3, pp. 98–108.
Sen, A. K. and M. S. Srivastava (1975b). “On tests for detecting change in mean when variance is unknown,” Annals of the Institute of Statistical Mathematics, Vol. 27, pp. 479–486.
Sen, A. K. and M. S. Srivastava (1975c). “Some one-sided tests for change in level,” Technometrics, Vol. 17, pp. 61–64.
Smith, A. F. M. (1975). “A Bayesian approach to inference about a change point in a sequence of random variables,” Biometrika, Vol. 62, pp. 407–416.
Smtih, A. F. M. (1977). “A Bayesian analysis of some time-varying models,” in Recent Developments in Statistics (edited by J. R. Barra et al.), North-Holland, Amsterdam, pp. 257–267.

Smith, A. F. M. (1980). “Change-point problems: approaches and applications,” in Bayesian Statistics, edited by J. M. Bernardo , University Press, Valencia, pp. 83–98.
Sprent, P. (1961). “Some hypotheses concerning two-phase regression lines,” Biometrics, Vol. 17, pp. 634–645.
Srivastava, M. S. (1980). “On tests for detecting change in the multivariate mean,” Technical Report no. 3, University of Toronto.
Tsurumi, H. (1977). “A Bayesian test of a parameter shift with an application,” J. of Econometrics, pp. 371–380.
Tsurumi, H. (1978). “A Bayesian test of a parameter shift in a simultaneous equation with an application to a macro savings function,” Economic Studies Quarterly, Vol. 24(3), pp. 216–230.
Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons, Inc., New York.

Multivariate Linear Models

Al Mahmeed, M. (1982). The Analysis of Autoregressive Processes: The Identification and the Prior, Posterior, and Predictive Analysis, Ph.D. Dissertation, Oklahoma State University, Stillwater, Oklahoma.
Anderson, T. W. (1951). An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc., New York.
Box, G. E. P. and G. C. Tiao (1973). Bayesian Inferences in Statistical Analysis, Addison Wesley, Reading, Massachusetts.
Broemeling, L. D. (1982). The Econometrics of Structural Change, Special issue of Journal of Econometrics, July 1982.
Chow, Gregory C. (1975). “Multiperiod predictions from stochastic difference equations by Bayesian methods,” in Studies in Bayesian Econometrics and Statistics, In Honor of Leonard J. Savage, edited by Fienberg and Zellner .
DeGroot, Morris H. (1970). Optimal Statistical Decisions, McGraw-Hill, New York.
Glass, Gene V. , Victor L. Willson , and John M. Gottman (1975). Design and Analysis of Time Series Experiments, Colorado Associated University Press.
Granger, C. W. J. and P. Newbold (1977). Forecasting Economic Time Series, Academic Press, New York.
Litterman, Robert B. (1980). “A Bayesian procedure for forecasting with vector autoregressions,” Massachusetts Institure of Technology.
Moen, David H. (1982). “The Bayesian analysis of structural change in multivariate linear models,” Dissertation proposal, Department of Statistics, Oklahoma State University, Stillwater, Oklahoma.
Morrison, Donald F. (1967). Multivariate Statistical Methods, McGraw-Hill, New York.
Newbold, Paul (1973). “Bayesian estimation of Box-Jenkins transfer function-noise models,” Journal of the Royal Statistical Society, Series B, pp. 323–336.
Press, James S. (1980). “Bayesian inference in MANOVA,” in Handbook of Statistics, Vol. I, edited by P. B. Krishnaiah , North-Holland Publishing Company, New York.
Press, James S. (1982). Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference, Krieger Publishing Co., Melbourne, Florida.
Press, James S. and Kazuo Shigemasu (1982). “Bayesian MANOVA and MANOCOVA under exchangeability,” Technical report no. 97, August 1982.
Rothenberg, T. J. (1963). “A Bayesian analysis of simultaneous equations systems,” Report 6315, Econometric Institute, Netherlands School of Economics, Rotterdam.
Salazar, Diego (1980). The Analysis of Structural Changes in Time Series and Multivariate Linear Models, Ph.D. dissertation, Oklahoma State University, Stillwater, Oklahoma.
Salazar, Diego , L. D. Broemeling , and A. Chi (1981). “Parameter changes in a regression model with autocorrelated errors,” Communications in Statistics, Vol. A10, pp. 1751–1758.
Salazar, Diego (1982). “Structural changes in time series models,” in The Econometrics of Structural Change, a special issue of the Journal of Econometrics, edited by L. D. Broemeling , July 1982.

Sen, A. K. and M. S. Srivastava (1973). “On multivariate tests for detecting change in mean,” Sankhya A, Vol. 35, pp. 173–185.
Sen, A. K. and M. S. Srivastava (1975a). “On tests for detecting change in mean,” The Annals of Statistics, Vol. 3, pp. 98–108.
Sen, A. K. and M. S. Srivastava (1975b). “On tests for detecting change in mean when variance is unknown,” Annals of the Institute of Statistical Mathematics, Vol. 27, pp. 479–486.
Sen, A. K. and M. S. Srivastava (1975c). “Some one-sided tests for change in level,” Technometrics, Vol. 17, pp. 61–64.
Tsurumi, H. (1978). “A Bayesian test of a parameter shift in a simultaneous equation with an application to a macro savings function,” Economic Studies Quarterly, Vol. 29, pp. 216–230.
Tsurumi, H. (1980). “A Bayesian estimation of structural shifts by gradual switching regressions with an application to the U.S. gasoline market,” in Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys, edited by A. Zellner , North-Holland Publishing Company, New York.
Zellner, Arnold (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons Inc., New York.

Looking Ahead

Akaike, H. (1979). “A Bayesian extension of the minimum AIC procedure of autoregressive model building,” Biometrika, Vol. 66, No. 2, pp. 237–242.
Abraham, Bovas and William W. S. Wei (1979). “Inferences in a switching time series,” American Statistical Association Proceedings of the Business and Economic Statistics Section, pp. 354–358.
Barnet, Vic and Toby Lewis (1978). Outliers in Statistical Data, John Wiley and Sons, Inc., New York.
Box, G. E. P. and G. C. Tiao (1973). Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Mass.
Broemeling, Lyle (1982). “Introduction,” in Structural Changes in Econometrics, pp. 1–5, edited by Lyle Broemeling , Annals of Applied Econometrics 1982–2, A Supplement to the Journal of Econometrics, Vol. 19, No. 1.
Diaz, Joaquin and Jose Farah (1981). “Bayesian identification of autoregressive processes,” 22nd NBER-NSF Seminar on Bayesian Inference in Econometrics.
Dreze, Jacque H. (1977). “Bayesian regression analysis using poly-t distributions,” in New Developments in the Application of Bayesian Methods, edited by Aykac and Brumat , North-Holland Publishing Co., Amsterdam.
Graybill, F. A. (1961). An Introduction to Linear Statistical Models, McGraw-Hill, New York.
Kalman, R. E. (1960). “A new approach to linear filtering and prediction problems,” Transactions, ASME, Journal of Basic Engineering, Vol. 82-D, pp. 34–45.
Press, James S. (1982). Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference, Kreiger Publishing Co, Melbourne, Florida.
Zellner, Arnold (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley and Sons Inc., New York.