A BAYESIAN ANALYSIS OF MODEL SPECIFICATION UNCERTAINTY
IN FORECASTING AND CONTROL

By
PAUL GEORGE BENSON

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA
1977
Even though you now know the true model of the process and have no need for this dissertation... this is for you, Dad.
ACKNOWLEDGEMENTS

For his technical insights, constant encouragement, unflagging optimism, patience, and personal commitment to me and this dissertation I am deeply indebted to my friend and major advisor, Dr. Christopher B. Barry. Exposure to his enthusiasm for and approach to research has sparked my interest in research and made me aware of career paths I might otherwise have overlooked.

Dr. Ira Horowitz has made numerous helpful technical and editorial suggestions. More importantly, at the time of my greatest personal need his door was always open. His emotional support and understanding will never be forgotten.

I am grateful to the other members of my committee, Dr. Roger D. Blair, Dr. H. Russell Fogler, and Dr. James T. McClave, for not only their comments and constructive criticisms of this dissertation, but for their encouraging words and advice throughout my tenure at Florida. I am fortunate to have such friends.

Dr. William Mendenhall has been both friend and advisor for many years. It was he who originally aroused my interest in combining a major in quantitative management with a minor in statistics. I have benefited both personally and professionally from his advice, support, and concern for my well being.

iii
I particularly want to thank Dr. Max R. Langham for the interest and confidence he showed in me during my first two years at Florida. His words of encouragement made me feel capable of succeeding in a doctoral program and convinced me to continue on for a doctorate.

My colleagues David L. Hill and Ronald E. Shiffler took time from their own dissertations to listen to and comment on my ideas. I particularly want to thank Dave for his many very helpful suggestions.

My thanks to Kathy Jarboe at the University of Florida and Diane Berube at the University of Minnesota for the excellent typing support they provided while I was drafting chapters of this dissertation. Thanks also to Kathy for cheerfully functioning as a go-between and red-tape cutter at the University of Florida after I moved to Minnesota.

I am particularly indebted to Pat Kaluza for professionally creating the final typed draft of this dissertation. Her willingness to meet tight deadlines and ability to smile even after facing the painful notation of Chapter V will be remembered.

My sincere thanks to Elizabeth Wells for patiently listening to and advising me on a potentially all-consuming personal problem. Our many long talks did much to free my mind for work on this dissertation.

The thoughtfulness of friends Sharron K. Duncan and Susan G. Benson, and new friends Thomas R. Lundstedt, Charlotte (Char) A. Lundstedt, and Claris E. Loomis helped me to successfully negotiate the transition from sunny Florida to frozen Minnesota last winter. Claris deserves special mention. Her concern for my happiness and interest in my work helped
speed my return to work on this dissertation. Had we not met when we did, this dissertation might still be a collection of partially finished chapter drafts.

My first course in statistics was at Bucknell University in 1964. The text we used was by Dr. Mendenhall, the instructor was my father. No single course or approach to teaching has influenced me more. Throughout my life, but particularly during the past two and a half years, my father has been a great source of strength and inspiration. Even though absent, he was present. The completion of this dissertation is as much his accomplishment as it is mine.

Finally, my thanks to my mother for not letting me set my goals too low or letting me give up before attaining them. Your many, many sacrifices have been and are appreciated.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>I.1 Statistical Models</td>
<td>2</td>
</tr>
<tr>
<td>I.2 Model Specification Uncertainty</td>
<td>3</td>
</tr>
<tr>
<td>I.3 The Bayesian Approach to Inference and Decision</td>
<td>5</td>
</tr>
<tr>
<td>I.3.1 The Predictive Distribution</td>
<td>5</td>
</tr>
<tr>
<td>I.3.2 The Posterior Distribution</td>
<td>7</td>
</tr>
<tr>
<td>I.4 Chapter Outline and Preview of Results</td>
<td>7</td>
</tr>
<tr>
<td>II. HYPOTHESIS TESTING, BAYESIAN MODEL SELECTION, AND BAYESIAN MODEL COMPARISON</td>
<td>11</td>
</tr>
<tr>
<td>II.1 Harold Jeffreys: Hypothesis Testing</td>
<td>11</td>
</tr>
<tr>
<td>II.2 Harry V. Roberts: Comparing Forecasters</td>
<td>13</td>
</tr>
<tr>
<td>II.3 Martin S. Geisel: Bayesian Model Comparison and Selection</td>
<td>16</td>
</tr>
<tr>
<td>III. FORECASTING WITH AND WITHOUT REGARD FOR MODEL SPECIFICATION UNCERTAINTY</td>
<td>25</td>
</tr>
<tr>
<td>III.1 A Comparison of the Predictive Variances Generated by the Bayesian Mixed Model Distribution and the Bayesian Model Selection Procedure</td>
<td>27</td>
</tr>
<tr>
<td>III.2 Forecasting: Bayesian Model Comparison Versus Bayesian Model Selection and the Maximize-R^2 Rule</td>
<td>36</td>
</tr>
<tr>
<td>III.2.1 The Bayesian Model Selection Procedure (BMS)</td>
<td>37</td>
</tr>
<tr>
<td>III.2.2 The Maximize-R^2 Rule</td>
<td>39</td>
</tr>
<tr>
<td>III.2.3 The Bayesian Model Comparison Procedure (BMC)</td>
<td>42</td>
</tr>
<tr>
<td>III.2.4 Model Space and Assumptions</td>
<td>45</td>
</tr>
<tr>
<td>III.2.5 The Treatment of Model Specification Uncertainty</td>
<td>48</td>
</tr>
<tr>
<td>III.2.6 Risk Specification</td>
<td>50</td>
</tr>
<tr>
<td>III.2.7 A Comparison of Expected Losses</td>
<td>57</td>
</tr>
<tr>
<td>III.2.8 Implications for Point Estimation</td>
<td>59</td>
</tr>
<tr>
<td>III.2.9 Implications for Interval Estimation</td>
<td>62</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>IV. MODEL SPECIFICATION UNCERTAINTY IN SINGLE-PERIOD ECONOMIC CONTROL PROBLEMS.</td>
<td>68</td>
</tr>
<tr>
<td>IV.1 The Economic Control Problem</td>
<td>69</td>
</tr>
<tr>
<td>IV.2 Model Space and Assumptions</td>
<td>74</td>
</tr>
<tr>
<td>IV.3 Single-Period Certainty-Equivalent Control</td>
<td>76</td>
</tr>
<tr>
<td>IV.3.1 Certainty-Equivalent Control Using the B*MMP Distribution</td>
<td>78</td>
</tr>
<tr>
<td>IV.3.2 Risk Specification in Certainty-Equivalent Control</td>
<td>82</td>
</tr>
<tr>
<td>IV.4 Optimal Single-Period Control</td>
<td>87</td>
</tr>
<tr>
<td>IV.4.1 Optimal BMC Control</td>
<td>89</td>
</tr>
<tr>
<td>IV.4.2 Optimal BMC Control When Instrument Use Costs Are Considered</td>
<td>94</td>
</tr>
<tr>
<td>IV.4.3 Certainty-Equivalent BMC Control Solutions When Instrument Use Costs Are Considered</td>
<td>104</td>
</tr>
<tr>
<td>IV.4.4 Risk Specification in Optimal BMC Control</td>
<td>107</td>
</tr>
<tr>
<td>IV.4.5 BMC Control When More Complicated Models Are Included in the Model Space</td>
<td>108</td>
</tr>
<tr>
<td>V. BAYESIAN MODEL SWITCHING</td>
<td>112</td>
</tr>
<tr>
<td>V.1 Bayesian Model Switching Methodology</td>
<td>112</td>
</tr>
<tr>
<td>V.2 Special Case: Two Normal Models</td>
<td>122</td>
</tr>
<tr>
<td>VI. CONCLUDING COMMENTS AND SUGGESTIONS FOR FURTHER RESEARCH</td>
<td>139</td>
</tr>
<tr>
<td>VI.1 Research Difficulties Encountered</td>
<td>141</td>
</tr>
<tr>
<td>VI.2 Shortcomings of the BMC Procedure and Suggestions for Further Research</td>
<td>143</td>
</tr>
<tr>
<td>VI.3 Suggestions for Further Research in the Areas of Economic Control and Model Nonstationarity</td>
<td>145</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>148</td>
</tr>
<tr>
<td>BIOGRAPHICAL SKETCH</td>
<td>151</td>
</tr>
</tbody>
</table>
Abstract of Dissertation Presented to the Graduate Council of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

A BAYESIAN ANALYSIS OF MODEL SPECIFICATION UNCERTAINTY IN FORECASTING AND CONTROL

By

Paul George Benson

August 1977

Chairman: Christopher B. Barry
Major Department: Management

The use of statistical models for forecasting and economic control has received widespread attention in recent years. Most of this attention has been focused on the problems caused by uncertainty concerning the parameters of a given model, whereas little attention has been paid to the problems caused by uncertainty concerning the specification of the model itself. In this dissertation Bayesian methodology is employed to treat model specification uncertainty in forecasting and control environments. The implications of forecasting with and without formal regard for model specification uncertainty are explored via a comparison of the recommended methodology and alternative methods which involve the selection of a single model. The recommended methodology is applied to single-period economic control problems. In particular, certainty-equivalent and optimal analytic solutions are found for problems in which there exist two viable alternative linear models of the data-generating process each with a different instrument and no intercept.
term. Solutions are obtained for situations in which control is cost-free and in which various instrument-use cost functions are known. Finally, a Bayesian procedure for modeling and making inferences about particular nonstationary data-generating processes is introduced. This procedure characterizes data as being generated by different statistical models in different time periods with the switch between models controlled by some random process.
CHAPTER I
INTRODUCTION

The use of statistical models for forecasting and economic control has received widespread attention in recent years. Most of this attention has been focused on the problems caused by uncertainty concerning the parameters of a given model. As a result, much has been written about parameter specification and estimation and their decision-making implications, whereas little analytical attention has been paid to the problems caused by uncertainty concerning the specification of the model itself. The implications of this type of uncertainty for forecasting and control are virtually unexplored. That these implications are significant and worth exploring has been expressed by Pierce:

Another area of uncertainty has to do with our models. . . . The problem lies not only with uncertainty concerning the true value of model parameters, but also with the structure of models themselves. . . . We have found that with some relatively minor changes in the specification of our quarterly model. . . . we can importantly alter its policy multipliers.\(^1\)

In this dissertation the Bayesian Model Comparison procedure developed by Geisel\(^1\) from the work of Roberts\(^2\) is advocated as a method for formally treating model specification uncertainty in forecasting and control problems. The implications for forecasting with and without regard for model specification uncertainty are examined, and the Bayesian Model Comparison procedure is applied to simple single-period economic control problems.

The following sections of this chapter introduce definitions and discuss concepts that will be referred to throughout the remainder of the dissertation.

I.1 Statistical Models

Throughout this dissertation the term "model" refers to a parametric statistical characterization of a data-generating process composed of both deterministic and random components. The general linear model used in regression analysis is an example of such a characterization. Each such model describes the data-generating process via a family of probability density functions in which each member of the family depends on a finite number of parameters, probability density functions over the parameters, and predetermined values of a specified set of variables upon which it has been hypothesized that the data-generating process depends.


Statistical models are used to describe the stochastic behavior of a data-generating process. Decision makers use them "as if" they were actually generating the data of interest. Any reference to a model as being the "true" or "correct" model of a data-generating process should not be taken literally. A model is referred to as being the "true" model only insofar as it behaves "as if" it were generating the observed data.

I.2 Model Specification Uncertainty

The statistical models discussed in the previous section explicitly admit uncertainty about the data-generating process through their parameters and random error terms. These two sources of uncertainty will be referred to as parameter uncertainty and random error (or residual) uncertainty. Random error is present in each model since the deterministic component of the model cannot realistically be expected to account for all factors influencing a realization of the data-generating process. Parameter uncertainty is present since a model's parameters are typically not observable and must be estimated from sample data. Being explicitly present in a statistical model, these two types of uncertainty and their implications for decision making have received considerable attention in the literature.¹

Thus, it is well-known that the appropriate use of a statistical model in decision making requires the consideration and treatment of both parameter and random error uncertainty.

¹References are provided in Chapters III and IV.
When a decision maker is uncertain as to the functional form of his model and/or is uncertain as to the set of variables upon which the data-generating process depends, model specification uncertainty is said to be present. Model specification uncertainty and its decision-making implications have received little attention in the literature. As a result, model specification uncertainty is typically ignored or assumed away in the statistical analysis of data-generating processes that precedes decision-making.¹ The usual procedure is for the decision-maker to utilize sample information to aid in the selection of a model from a set of models he believes to be viable alternative representations of the data-generating process. The chosen model is then assumed to appropriately represent the data-generating process, and the decision maker bases his decisions on the information provided by this model. Such a procedure can formally consider only parameter and random error uncertainty. Depending on the particular model selection procedure utilized, model specification uncertainty is either completely ignored or suboptimally treated. The result is that some or all of the information provided about the data-generating process by the set of models which were not chosen, but were believed to be viable, is lost. This loss is analogous to the information loss that would occur if the decision maker assumed he knew the parameters of a given model and made his decisions without acknowledging parameter uncertainty. Chapters III and IV will discuss in detail the decision-making implications of the information loss caused by failing to treat model specification uncertainty.

¹An interesting exception is the recent paper by M. Brenner, "The Effect of Model Misspecification on Tests of the Efficient Market Hypothesis," Journal of Finance, 32 (1977), 57-66. There are other exceptions as well.
I.3 The Bayesian Approach to Inference and Decision Making

In this dissertation uncertainty is dealt with via Bayesian inferential procedures. This section briefly reviews the methodology of Bayesian inference.

1.3.1 The Predictive Distribution

Decisions frequently hinge on the future outcome of a data-generating process. In such cases decision makers typically use a statistical model to characterize the data-generating process. If model specification uncertainty is negligible and the parameters of the model are known, then the decision maker can feel secure in basing his decision on the information provided him by his model. However, if the parameters of the model are unknown the model should be altered to reflect the decision-maker's uncertainty concerning the parameters. This can be accomplished by treating the parameters as random variables, utilizing a probability distribution over the parameters to reflect the decision-maker's parameter uncertainty, and computing the marginal distribution of future realizations of the data-generating process, i.e., the distribution of future realizations which is not conditioned on the model's parameters.

Suppose the decision-maker's statistical model describes the data-generating process via the sampling distribution $f(y_F|\theta)$, where $y_F$ is a future value of the data-generating process ($y_F \in Y$) and the

---

parameters of the data-generating process are represented by
\[ \theta \ (\theta \in \Theta) \]. Then, if the decision-maker's parameter uncertainty can be described by a probability distribution \( g(\theta) \), the decision maker can compute the marginal distribution of future realizations of the data-generating process as follows:
\[
f(y_F) = \int g(\theta) f(y_F | \theta) d\theta. \tag{1.1}
\]
This distribution is referred to as a predictive distribution.

If the decision maker is able to obtain a sample from the data-generating process of interest he may update his distribution of \( \theta \) to reflect the sample information. Then, utilizing his revised distribution of \( \theta \), he may recompute his predictive distribution of \( y_F \) so that it too reflects the sample information. The revision of \( f(\theta) \) is accomplished via Bayes' Rule:
\[
f''(\theta | y) = \frac{g'(\theta) f(y | \theta)}{\int g'(\theta) f(y | \theta) d\theta}. \tag{1.2}
\]
The function \( g'(\theta) \) is called the decision-maker's prior distribution of \( \theta \) since it was established prior to obtaining the sample \( y \). The function \( f(y | \theta) \) is a likelihood function. It describes the likelihood of the given sample result, \( y \), for different values of \( \theta \). The function \( f''(\theta | y) \) is the decision-maker's revised distribution of \( \theta \). It is called a posterior distribution since it was computed following the receipt of sample information. The posterior distribution reflects all the information about \( \theta \) currently available to the decision maker. This information may be incorporated into his predictive distribution of \( Y_F \) as follows:
\[ f(y_F|y) = \int_\theta f''(\theta|y)f(y_F|\theta)d\theta. \] (1.3)

It is from this distribution that needed information about future observations of the data-generating process should be extracted. As more sample and/or subjective information about the process becomes available, the decision maker can formally revise his predictive distribution to reflect that information by repeatedly applying the above procedure.

1.3.2 The Posterior Distribution

There are three inputs to Bayes' Rule: (1) the decision-maker's prior information about \( \theta \) expressed via \( g'(\theta) \); (2) sample observations from the data-generating process; and (3) the choice of the functional form of the data-generating process, i.e., the choice of a likelihood function. The output of Bayes' Rule is an inferential statement about \( \theta \) in the form of a probability distribution, \( f''(\theta|y) \). A decision maker interested in obtaining information about a parameter of the data-generating process should compute \( f''(\theta|y) \). The function \( f''(\theta|y) \) can stand alone as an inferential statement about \( \theta \), or it can be used to determine point and interval estimates of \( \theta \). As more sample and/or subjective information about the data-generating process becomes available, Bayes' Rule can be reapplied to revise \( f''(\theta|y) \). The sequential application of Bayes' Rule permits the decision maker to formally learn about \( \theta \) over time.

1.4 Chapter Outline and Preview of Results

Typically econometric forecasting and control models are developed
and used without formally considering the full impact of model specification uncertainty. The usual procedure is to (1) utilize a model selection technique to choose one model from a set of alternative competing models to characterize the data-generating process, and (2) assume the chosen model to be the correct model of the data-generating process and use it to forecast and/or control the process. Such procedures either ignore or do not fully consider the information about the data-generating process contributed by the models that were proposed as being viable but were not selected by the model selection procedure. Further, in assuming the chosen model is the correct model of the process, the forecaster or controller is behaving as though he faces a lesser degree of uncertainty than is really the case. Thus, in utilizing model selection procedures, forecasters and controllers are simultaneously discarding relevant information about the data-generating process and behaving as if they have more information than is actually possessed. This dissertation advocates the use of the Roberts/Geisel Bayesian Model Comparison Procedure as a means of comprehensively treating model specification uncertainty and avoiding such contradictory behavior. The Bayesian Model Comparison Procedure and its origins are described in Chapter II. Chapter II also describes a Bayesian model selection procedure referred to herein as the Bayesian Model Selection Procedure.

In Chapter III, the effects of forecasting with and without regard for model specification uncertainty are examined by comparing forecasts determined via the Bayesian Model Comparison procedure (BMC) with those yielded by a Bayesian procedure which fails to appropriately consider
model specification uncertainty, the Bayesian Model Selection procedure (BMS). The following results are derived:

1. If the variance of the decision-maker's predictive distribution is used to measure forecast-risk, and a decision maker forecasts via the BMS procedure rather than the BMC procedure, the risk he takes in predicting future values of the data-generating process is misspecified.

2. A decision-maker's posterior expected loss from using a forecast derived via the BMC procedure is less than his posterior expected loss from forecasting via the BMS procedure.

3. Point estimates derived via BMS are frequently misplaced.

4. The reliability of credible intervals derived via the BMS procedure may be misspecified.

In Chapter IV, the BMC procedure is applied to simple single-period economic control problems. In particular, certainty-equivalent and optimal analytic solutions are found for the case of two competing linear models each with a different instrument (controllable variable) and no intercept term. The following results are obtained:

1. The BMC certainty-equivalent control solution is to set both instruments as if each instrument's respective model were in fact the true model of the data-generating process.

2. If the variance of the controller's predictive distribution is used to measure control-risk, and certainty-equivalent control is utilized, it can be shown that under certain circumstances the BMS approach to control always understates
the control-risk involved.

3. The optimal BMC control solution is to set both instruments as if each instrument's respective model were in fact the true model of the process. Since optimal BMC control treats model specification uncertainty, parameter uncertainty, and residual uncertainty, whereas certainty-equivalent control treats only model specification uncertainty, the optimal BMC control solution differs from the BMC certainty-equivalent control solution.

Certainty-equivalent and optimal BMC control solutions for cases where instrument use costs are known are also derived in Chapter IV.

In Chapter V, a procedure for handling model nonstationarity is introduced. Called Bayesian Model Switching, this procedure was suggested by anomalies observed in sequences of posterior model probabilities generated by the BMS and BMC procedures. The Bayesian Model Switching procedure characterizes the data-generating process in a manner similar to Quandt's switching regression regimes.¹

Chapter VI contains an overview of the dissertation, a discussion of the shortcomings of the Bayesian Model Comparison and Bayesian Model Switching procedures, and suggestions for future work in the area of model specification uncertainty.

CHAPTER II

HYPOTHESIS TESTING, BAYESIAN MODEL SELECTION, AND BAYESIAN MODEL COMPARISON

The Bayesian Model Comparison approach to handling model specification uncertainty in decision-making problems has its origins in the hypothesis testing work of Harold Jeffreys\(^1\) and is a direct spin-off of a Bayesian procedure developed by Harry Roberts\(^2\) for combining expert opinions. Martin Geisel\(^3\) adapted Roberts' work for use in econometrics and in so doing formalized the Bayesian Model Comparison and Bayesian Model Selection procedures. The contributions of Jeffreys, Roberts, and Geisel to the existing Bayesian Model Comparison and Bayesian Model Selection procedures are discussed in this chapter.

II.1 Harold Jeffreys: Hypothesis Testing\(^4\)

In considering two mutually exclusive and exhaustive hypotheses about the parameter vector \(\theta\) of a probability density function, Jeffreys suggests that the decision maker should place prior probability masses on each of the hypotheses. The probabilities should be consistent with the decision maker's prior information and, consequently,

\(^2\)Roberts, pp. 50-62.
\(^3\)Geisel, pp. 1-45.
\(^4\)Jeffreys, Chapters 4 and 5.
prior beliefs about the appropriateness of each of the hypotheses. Thus, if the two hypotheses \( H_0 \) and \( H_1 \) are exhaustive and nonoverlapping their prior probabilities \( P'(H_0) \) and \( P'(H_1) \) would be assessed, and must sum to one. If \( H_0 \) and \( H_1 \) are a priori equally likely, \( P'(H_0) = P'(H_1) \).

It is assumed that given \( H_0 \) a future sample result \( y \) has probability density function \( f(y|H_0) \), and that given \( H_1 \) is true, \( y \)'s probability density function is \( f(y|H_1) \). Then, using Bayes' Rule, the posterior probability that, say, \( H_0 \) is the appropriate hypothesis is

\[
P''(H_0|y) = \frac{P'(H_0)f(y|H_0)}{P'(H_0)f(y|H_0) + P'(H_1)f(y|H_1)}, \tag{2.1}
\]

and \( P''(H_1|y) = 1-P''(H_0|y) \). After determining \( P''(H_0|y) \) and \( P''(H_1|y) \), the decision maker can choose as the more appropriate hypothesis the one with the higher posterior probability. Or, if the decision maker can economically determine the losses involved from choosing an incorrect hypothesis, he can use \( P''(H_0|y) \) and \( P''(H_1|y) \) to determine the expected loss of choosing \( H_0 \) or \( H_1 \) and then select as being the more appropriate the hypothesis that minimizes his expected loss.

More formally, if \( H_0 \) is \( \theta = \theta_0 \) and \( H_1 \) is \( \theta = \theta_1 \), where \( \theta_0 \) and \( \theta_1 \) are particular values of the parameter vector (i.e., two simple hypotheses), then (2.1) would be

\[
P''(H_0|y) = P''(\theta = \theta_0|y) = \frac{P'(\theta = \theta_0)f(y|\theta = \theta_0)}{P'(\theta = \theta_0)f(y|\theta = \theta_0) + P'(\theta = \theta_1)f(y|\theta = \theta_1)}. \tag{2.2}
\]

If \( H_0 \) is \( \theta \in \psi_1 \) and \( H_1 \) is \( \theta \in \psi_2 \) where \( \psi_1 \) and \( \psi_2 \) \( (\psi_1 \cup \psi_2 = \psi) \) are mutually exclusive and exhaustive sets (i.e., \( H_0 \) and \( H_1 \) are two composite hypotheses), then it is necessary for the decision maker to
assess a prior pdf for $\theta$ over $\Psi_1$, $P'(\theta|\theta \in \Psi_1)$, and another for $\theta$ over $\Psi_2$, $P'(\theta|\theta \in \Psi_2)$. Then (2.1) would be

$$P''(H_1|y) = P''(\theta \in \Psi_1|y) = \frac{P'(\theta \in \Psi_1)f(y|\theta \in \Psi_1)}{\sum_{\Psi_1} f(y|\theta \in \Psi_1)}$$

(2.3)

where $f(y|\theta \in \Psi_1) = \int P'(\theta|\theta \in \Psi_1)f(y|\theta, \theta \in \Psi_1)d\theta$

and $f(y) = \sum_{i=1}^{2} P'(\theta \in \Psi_i)f(y|\theta \in \Psi_i)$.

The next section discusses Harry Roberts' important extension of Jeffreys' work.

II.2 Harry V. Roberts: Comparing Forecasters

Roberts was concerned with "reconciling conflicting expert interpretations of the same data." Building on Jeffrey's work, Roberts devised a method for discriminating among a set of alternative parametric statistical models each of which purports to describe some random process of interest. This Bayesian discrimination procedure will be discussed in detail in the next section.

It will be assumed that person C knows nothing about a particular data-generating process $f(y|\theta)$, but wishes, for example, to predict future $y$ values and is, therefore, interested in learning about the process. Persons A and B possess knowledge about the same process. A and B express their knowledge about $f(y|\theta)$ via the data distributions

---

1 Roberts, pp. 50-62.

2 Ibid., p. 55.
f(y|θ, A) and f(y|θ, B), respectively, and their prior distributions on the parameter θ, g(θ|A) and g(θ|B). θ may be a vector. For expository purposes, only two individuals will be assumed to possess knowledge about the process, and all probability distributions of this section will be assumed to be discrete.¹

C's prior distribution for the parameter θ may be expressed as:

\[ g'(θ) = P'(A)g(θ|A) + P'(B)g(θ|B). \]  (2.4)

P'(A) and P'(B) sum to one and may be thought of as C's probability assessment of the accuracy of A's judgment and B's judgment, respectively. If C had some knowledge about the reliability of opinions expressed by A and B he might tend to respect the opinion of one, say A, more than the other, and so assign P'(A) > P'(B). If C knew nothing about either A or B it would be appropriate for him to assess P'(A) = P'(B) = .5. C can then learn about f(y|θ) by combining his thoughts (if any) about A and B (reflected in P'(A) and P'(B)) with the opinions expressed by A and B about f(y|θ) (represented by g(θ|A), f(y|θ, A), g(θ|B), and f(y|θ, B)) as in (2.4), and by using sample information to revise (2.4). Thus it is C's posterior distribution of θ, g"(θ|y), that C should use in predicting y. Roberts' development of g"(θ|y) is outlined in the next paragraph.

Following Roberts, let $\lambda$ index the opinions of A and B, i.e., when $\lambda = \lambda_A$ reference is being made to person A, and when $\lambda = \lambda_B$ reference is being made to person B. With C's prior distribution for $\lambda$ denoted by $P'(\lambda)$, C's joint prior distribution for $\lambda$ and $\theta$ is denoted:

$$g'(\theta, \lambda) = P'(\lambda)g(\theta|\lambda).$$

(2.5)

Accordingly, C's marginal prior distribution for $\theta$ is denoted by

$$P'(\theta) = \sum_{\lambda} P'(\lambda)g(\theta|\lambda).$$

(2.6)

Equations (2.6) and (2.4) are equivalent. C's joint posterior distribution of $\lambda$ and $\theta$ is obtained via Bayes' Rule as follows:

$$h''(\lambda, \theta|y) = \frac{h'(\lambda, \theta)f(y|\lambda, \theta)}{\sum_{\lambda, \theta} h'(\lambda, \theta)f(y|\lambda, \theta)} = \frac{P'(\lambda)g'(\theta|\lambda)f(y|\lambda, \theta)}{f(y)}.$$  

(2.7)

$f(y|\lambda, \theta)$ represents the likelihood of observing the sample result $y$ given particular values for $\lambda$ and $\theta$. $f(y)$ is the marginal distribution of the data. Then, recognizing that $g(\theta|\lambda)f(y|\lambda, \theta) = f(y|\lambda)g(\theta|y, \lambda)$, $g''(\theta|y)$ is obtained from (2.7) as follows:

$$g''(\theta|y) = \sum_{\lambda} \frac{P'(\lambda)g'(\theta|\lambda)}{f(y)} f(y|\lambda, \theta)$$

$$= \sum_{\lambda} \frac{P'(\lambda)f(y|\lambda)}{f(y)} g(\theta|\lambda, \theta)$$

$$= \sum_{\lambda} P(\lambda|y)g(\theta|\lambda, y).$$

(2.8)

Thus C's posterior distribution of $\theta$ is a weighted average of A's posterior distribution of $\theta$ and B's posterior distribution of $\theta$.

Roberts points out that if (2.7) is summed over $\theta$ instead of $\lambda$, as was done in (2.8), the marginal posterior distribution of $\lambda$ is obtained:
\[ P(\lambda | y) = \frac{P'(\lambda) f(y | \lambda)}{f(\lambda)} . \] (2.9)

Roberts notes that in statistical discrimination problems where it is assumed that \( y \) is generated by either \( f(y | \lambda_A) \) or \( f(y | \lambda_B) \), \( P'(\lambda_i) \) (\( i = A, B \)) may be interpreted as the discriminator's prior probability that \( f(y | \lambda_i) \) generates \( y \), and \( P(\lambda_i | y) \) may be interpreted as the discriminator's posterior probability that \( f(y | \lambda_i) \) generates \( y \). Roberts suggests that discrimination between these two alternative generating processes should be accomplished via examination of the posterior odds ratio, \( \frac{P(\lambda_A | y)}{P(\lambda_B | y)} \).

Roberts' interpretation of \( P(\lambda_i | y) \), and his suggested procedure for discriminating among alternative statistical models, were formalized by Geisel in his Bayesian Model Selection and Bayesian Model Comparison schemes. Geisel's extension of Roberts' work is discussed in the next section.

II.3 Martin S. Geisel: Bayesian Model Comparison and Selection

Geisel's work was concerned with Bayesian procedures for comparing and choosing among parametric statistical models. His procedure for comparing models will be referred to as the Bayesian Model Comparison (BMC) approach. His procedure for choosing one model from among a set of competing models uses the same methodology as the BMC approach but for different purposes. Consequently, the latter procedure will be referred to here as the Bayesian Model Selection (BMS) procedure.

---

1 Geisel, pp. 1-45.
Suppose the decision maker feels that any one of N alternative models could represent the data-generating process of interest to him. Denote by $P'(M_i)$, $i=1,2,...,N$ the decision-maker's prior probability that $M_i$, the ith model, is an accurate representation of the data-generating process. If the decision maker assesses $P'(M_i) > 0$, then the model should be included in the set of N models. It follows that $\sum_{i=1}^{N} P'(M_i) = 1$. The unknown vector of parameters of $M_i$ is denoted by $\theta_i$, $i=1,...,N$ where $\theta_i \in \Theta$. The decision maker's knowledge about $\theta_i$ is described via a prior density function, $g'(\theta_i|M_i)$.

If $M_i$ were known to be the true model and its parameters were known to be $\theta_i^0$, the data-generating process could be completely characterized by the density function $f(y|\theta_i^0, M_i, D_i)$, where $y$ is the random variable of interest to the decision maker. In the forecasting problems of Chapter III, $D_i$, which may be a vector, will be the explanatory variables of $M_i$ and will be used to help forecast future values of $y$. In the economic control problems of Chapter IV, $D_i$ will be the independent variables of $M_i$ and will be under the control of the decision maker. Once $y$ has been observed, $f(y|\theta_i, M_i, D_i)$, viewed as a function of $\theta_i$, $M_i$ and $D_i$, is a likelihood function and can be used to make inferences from the data about the correct model and about the parameters of all the models.

---

1 $y$ may be vector-valued, but in order to simplify the notation and discussion to follow it is assumed that $y$ is a scalar.
As new information is received about the data-generating process being modeled, i.e., as \( y \) is observed, the prior distribution on the parameters of \( M_i \) should be revised to reflect this new information. Revision for a single model is accomplished exactly as if the parameter distribution of a known data-generating process with unknown parameters were being revised. Applying Bayes' Rule yields

\[
g''(\theta_i | M, y, D) = g'(\theta_i | M) f(y | \theta_i, M, D) / f(y | M, D) \tag{2.10}
\]

where

\[
f(y | M, D) = \int g'(\theta_i | M) f(y | \theta_i, M, D) d\theta_i. \tag{2.11}
\]

The function \( g''(\theta_i | M, y, D) \) is the posterior distribution of \( \theta_i \).

Given \( M_i \) and \( D_i \), and before observing \( y \), \( f(y | M, D) \) is commonly called the predictive density function of \( y \). It is the distribution of future realizations of the data-generating process conditioned on \( M_i \) being the correct model of the process and unconditioned on \( \theta_i \), the parameters of \( M_i \). Having observed \( y \), \( f(y | M, D) \) may be thought of as a "model likelihood" since it compares the relative likelihood of the data, \( y \), across models. Utilizing these model likelihoods Bayes' Rule is invoked a second time to revise the prior model probabilities:

\[
P''(M_i | y, D) = P'(M_i) f(y | M_i, D) / f(y | D) \tag{2.12}
\]

where

\[
f(y | D) = \sum_{i=1}^{N} P'(M_i) f(y | M_i, D). \tag{2.13}
\]

\( P''(M_i | y, D) \) is the posterior probability that \( M_i \) is the correct model. \( f(y | D) \) is a predictive distribution, a distribution of future
realizations of the data unconditioned on a particular model being the correct model. \( D \), written without a subscript, is a vector comprised of the set of decision variables, \( D_i \), \( i=1,2,...,N \) from all \( N \) models.

After observing \( y \) and revising the prior distributions on \( M_i \) and \( \theta_i \), the posterior probability distributions reflect all the information the decision maker has about the set of models and their parameters. Any prior information is reflected in the prior distributions, \( P'(M_i) \) and \( g'(\theta_i|M_i) \). The sample evidence, \( y \), is incorporated through the likelihood function, \( f(y|M_i,\theta_i,D_i) \). As additional information in the form of further observations of \( y \) is obtained, it may be reflected in new posterior distributions that are obtainable via revision of the existing posteriors (which, relative to the latest data, are called priors) derived in (2.10) and (2.12) above.

As long as the data-generating process does not change over time, the application of (2.10) and (2.12) to successive sets of new data permits the decision maker to "learn from experience" about which model of the process is the most appropriate. When the data may be generated by different models in different time periods, successive application of the probability revision procedures in this section would be inappropriate. This problem and an approach to handling it are discussed in Chapter V.

The above procedure can be used to select a single model to represent a random process by a decision maker who is uncertain about the appropriate form of that process. He can accomplish this by
choosing from his original set of $N$ competing models the one with the highest posterior probability or, if losses associated with choosing the incorrect model are known or can be estimated, by selecting the model that minimizes his posterior expected loss. The use of posterior model probabilities for model selection is the procedure referred to in this dissertation as Bayesian Model Selection (BMS).

The decision-maker's posterior model probabilities indicate that he is uncertain of the form of the random process. Thus, any decision procedure based on a chosen model fails to appropriately treat model specification uncertainty. Geisel points out that if the posterior probability of a model is positive, then the model contributes to our knowledge of future observations of the random process of interest and there is no theoretical reason to neglect this contribution. Hence, any decision procedure that involves selecting a single model from among a set of competing models ignores relevant information, and, computation costs and other complexities aside, ¹ can only be viewed as an approximation to an optimal procedure.

The key to utilizing all the information contained in the set of competing models relative to future observations of the random process lies in the use of the predictive density function derived in (2.13) above and repeated here in more detail:

\[
f(y|D) = \sum_{i=1}^{N} P'(M_i) \int_{\Theta} f(y|\theta_i, M_i, D_i) g'(\theta_i|M_i) d\theta_i
\]

(2.14)

\[
= \sum_{i=1}^{N} P'(M_i) f(y|M_i, D_i).
\]

\[f(y|D)\] is a weighted average of the predictive densities of \(y\) for each of the \(N\) models (referred to below as model predictives). It is this distribution that the decision maker should use to characterize the random process upon which his decision hinges and about whose form he is uncertain. This distribution will herein be referred to as a Bayesian Mixed Model Predictive (BMMP) distribution. The process of computing and analyzing posterior probabilities and the associated BMMP distribution is called "comparing models" by Geisel and is referred to herein as the Bayesian Model Comparison (BMC) procedure.

Suppose that \(y\) has been observed and that the decision maker is interested in making a decision that relates to some future value, \(y_F\), of the random variable. If the decision maker knew the correct model, say \(M_i\), and its parameters, say \(\theta_i\), then his distribution of \(y_F\) would be \(f(y_F|\theta_i, M_i, D_{Fi})\) and his decision would depend on this distribution. \(D_{Fi}\) is used to denote values of the decision variables of model \(i\) associated with \(y_F\). But the decision maker knows neither the correct model nor its parameters. What he does know is summarized in \(P'(M_i|y,D)\) and \(f(y_F|M_i, y, D_i)\). Thus, his distribution of \(y_F\) should be a BMMP conditioned on the data already observed, \(y\) and \(D\):
\[ f(Y_F|y,D,D_F) = \sum_{i=1}^{N} p''(M_i|y,D) \left[ \int_{\Theta} f(y_F|\Theta_i,M_i,D_F_i)g(\Theta_i|M_i,y,D_i)d\Theta_i \right] \tag{2.15} \]

This BMMP is a function of all \( N \) competing models and thus enables the decision maker to choose a course of action in light of all available information relating to \( Y_F \).

Even when the BMMP is the distribution (model) that the decision maker should use to characterize the random process in question, there can be at least three reasons for selecting a single model via the Bayesian Model Selection procedure:

1) In comparing alternative theories or hypotheses it may be desirable to choose the one with the most substantive content.

2) It may be more convenient to approximate the random process with a simple model.

3) The use of a BMMP may prove too costly. In general, the computation of a BMMP involves the combination of its components via extensive numerical methods.

Geisel shows that under certain assumptions, Bayesian Model Selection provides a Bayesian interpretation for the classical procedure of choosing from a set of models the one with the lowest estimated residual variance, \( s^2 \), or highest coefficient of determination, \( R^2 \). Given a set of normal regression models each of which
has the same number of parameters, given diffuse prior distributions over the models and the parameters of the models, and given a symmetric loss function with respect to the choice of an incorrect model, Geisel shows that the procedure of choosing a model with the highest posterior model probability, \( P'(M_1 | y, D) \), is equivalent to the procedure of selecting the model with the lowest \( s^2 \) or highest \( R^2 \).\(^1\) This result is very similar to a result derived by Thornber.\(^2\) Thornber, however, uses as priors on the parameters of the models those suggested by Jeffreys' invariance theory,\(^3\) whereas Geisel's priors on the parameters of the models take the form of multinormal and inverted gamma-2 distributions. These results will be discussed in more detail in Chapter III.

Another important Geisel result that will be drawn upon is his proof that given, say, \( M_1 \) is the true model in the set of \( N \) competing models, as sample evidence accumulates (i.e., \( n \to \infty \)) \( P'(M_1 | y, D) \to 1 \) and the BMMP \( \to P(y_1| M_1, y, D_1) \).\(^4\) Thus, if the decision maker could wait long enough, the data he would observe would tell him with near certainty which of the \( N \) models was generating the data. This result will be discussed in more detail in Chapter III.

\(^1\)Ibid., pp. 24-37.


\(^4\)Geisel, p. 23.
In the next chapter some of the consequences of forecasting with and without the use of the Bayesian Model Comparison procedure are explored. Particular attention is paid to the comparison of the Bayesian Model Comparison procedure and the Bayesian Model Selection procedure.
CHAPTER III
FORECASTING WITH AND WITHOUT REGARD FOR MODEL SPECIFICATION UNCERTAINTY

If a decision maker is uncertain as to which one of N random processes is generating future values of a random variable upon which the effectiveness of his current decision depends, Geisel contends that the decision maker should use the Bayesian Mixed Model Predictive (BMMP) distribution of the Bayesian Model Comparison (BMC) procedure to reflect the information he has concerning the process of interest.¹ His justification for this approach rests primarily on the following statement:²

Note again that this procedure does not select one model as "true" or "best" and eliminate the rest. If the probabilistic weight of a model is positive it contributes to our knowledge of the future observations and there is no reason to neglect this contribution. Thus, any decision theoretic procedure which is designed to eliminate some of the models is viewed as an approximation which is used for reasons of simplicity of view or to reduce the cost of computation.

This chapter explores some of the consequences of forecasting with and without the use of the Bayesian Model Comparison procedure and, in so doing, attempts to more rigorously justify advocation of the BMC procedure for use in decision-making problems in which model specification uncertainty is present. The chapter attempts to explain why

¹Geisel, Chapter II.
²Ibid., p. 19.
it would frequently be worth the extra cost to use the BMC approach rather than approaches which, though perhaps simpler and less costly, fail to fully reflect model specification uncertainty and the totality of information the decision maker has concerning the process of interest.

In this chapter, forecasting via the Bayesian Model Comparison procedure will be compared to forecasting via the Bayesian Model Selection procedure and the maximize-$R^2$ rule. It is shown that when model specification uncertainty exists, of these three procedures only the BMC procedure optimally handles the information the decision maker has concerning the data-generating process whose future values he wants to forecast. More specifically, if a decision maker forecasts via the BMS procedure, it is shown that the risk he takes in predicting future values of the random process of interest is misspecified. It is also shown that the decision-maker's posterior expected loss from using a BMC forecast is less than his posterior expected loss from using a BMS forecast. The last two sections of this chapter compare the effectiveness of point and interval forecasts generated via the BMC procedure with those generated via the BMS procedure. It is shown that BMS point estimates are typically misplaced and that the reliability of BMS credible intervals may be misspecified.

The following section introduces notation which will be used in the remainder of the chapter and examines the relationship between the predictive variance of $y$ as defined by a BMMP distribution, and the predictive variance of $y$ as defined by the model selected by the BMS procedure.
III.1 A Comparison of the Predictive Variances Generated by the Bayesian Mixed Model Distribution and the Bayesian Model Selection Procedure

Much of the analysis in Section III.2 draws on the relative sizes of the predictive variance of $y$ as defined by a Bayesian Mixed Model Predictive distribution, $V(\text{BMMP})$, and the predictive variance of $y$ as defined by a Bayesian Model Selection Predictive distribution, $V(\text{BMSP})$. Accordingly, to avoid awkward digressions in Section III.2, this section will be devoted to a comparison of $V(\text{BMMP})$ and $V(\text{BMSP})$.

It was shown in equations (2.14) and (2.15) that the BMMP is a weighted average of the predictive densities of $y$ for each of $N$ alternative models. Equation (2.15) is repeated here:

$$f(y_F|y,D,D_F) = \sum_{i=1}^{N} P^*(M_i|y,D)[\int f(y_F|\theta_i,M_i,D_{F_i})g^*(\theta_i|M_i,y,D_i) d\theta_i]$$

$$= \sum_{i=1}^{N} P^*(M_i|y,D)f(y_F|M_i,y,D_i,D_{F_i}). \quad (3.1)$$

The function $f(y_F|M_i,y,D_i,D_{F_i})$ will be referred to as a "model predictive." Recalling equation (2.11), a model predictive is a distribution of realizations from the data-generating process conditioned on

1) $M_i$ being the correct model of the process; 2) previous observations of $y$, the dependent variable of interest, and $D_i$, the decision variable; and 3) $D_{F_i}$, the value of the decision variable with which the next $y$ to be observed, $y_F$, is associated. Thus, if the Bayesian Model Selection procedure chooses, say, $M_i$, it is $M_i$'s predictive distribution, $f(y_F|M_i,y,D_i,D_{F_i})$, that is being chosen to characterize future observations of $y$, $y_F$. It is the variance of this predictive

$^1V(\text{BMMP})$ and $V(\text{BMSP})$ are formally defined below.
distribution that is referred to as $V(BMSP)$. In general, the mean and variance of the predictive distribution generated by $M_i$ will be denoted by $\mu_i$ and $\sigma_i^2$, respectively. The mean and variance of a BMMP will be denoted by $\mu$ and $\sigma^2$ (or $V(BMMP)$), respectively.

It is shown below that

$$\mu = P''(M_1|y,D)\mu_1 + \ldots + P''(M_N|y,D)\mu_N$$  \hspace{1cm} (3.2)

and

$$\sigma^2 = P''(M_1|y,D)[\sigma_1^2 + (\mu_1 - \mu)^2] + \ldots$$

$$+ P''(M_N|y,D)[\sigma_N^2 + (\mu_N - \mu)^2].$$  \hspace{1cm} (3.3)

To demonstrate, first note that $\mu$ can be obtained by definition as

$$\mu = \int_{-\infty}^{\infty} y_F f(y_F|y,D,D_F)dy_F.$$  \hspace{1cm} (3.4)

Substituting (3.1) for $f(y_F|y,D,D_F)$ in (3.4) yields

$$\mu = \int_{-\infty}^{\infty} y_F \left[ \sum_{i=1}^{N} P''(M_i|y,D)f(y_F|M_i,y,D_i,D_{F_i}) \right]dy_F.$$  \hspace{1cm} (3.5)

With the expansion of the sum in equation (3.5), (3.2) is obtained:

$$\mu = \int_{-\infty}^{\infty} y_F \left[ P''(M_1|y,D)f(y_F|M_1,y,D_1,D_{F_1}) \right.$$

$$+ \ldots + P''(M_N|y,D)f(y_F|M_N,y,D_N,D_{F_N}) \right]dy_F$$

$$= P''(M_1|y,D) \int_{-\infty}^{\infty} y_F f(y_F|M_1,y,D_1,D_{F_1})dy_F$$

$$+ \ldots + P''(M_N|y,D) \int_{-\infty}^{\infty} y_F f(y_F|M_N,y,D_N,D_{F_N})dy_F$$

$$= P''(M_1|y,D)\mu_1 + \ldots + P''(M_N|y,D)\mu_N.$$  \hspace{1cm} (3.6)
To obtain an expression for the predictive variance of the BMMP, note that by definition

$$V(\text{BMMP}) = \sigma^2 = \int_{-\infty}^{\infty} (y_F - \mu)^2 f(y_F | y, D, D_F) dy_F.$$  \hspace{1cm} (3.7)

Substituting (3.1) for $f(y_F | y, D, D_F)$ in equation (3.7) yields

$$\sigma^2 = \int_{-\infty}^{\infty} (y_F - \mu)^2 \sum_{i=1}^{N} p''(M_i | y, D_i) f(y_F | M_i, y, D_i, D_{F_i}) dy_F.$$ \hspace{1cm} (3.8)

The following is obtained by expanding the sum in equation (3.8):

$$\sigma^2 = \sum_{i=1}^{N} p''(M_i | y, D) \int_{-\infty}^{\infty} (y_F - \mu)^2 f(y_F | M_i, y, D_i, D_{F_i}) dy_F$$

$$= \sum_{i=1}^{N} p''(M_i | y, D) \int_{-\infty}^{\infty} (y_F^2 - 2y_F \mu + \mu^2) f(y_F | M_i, y, D_i, D_{F_i}) dy_F.$$  

Working with the $i$th term of this sum, the following is obtained:

$$p''(M_i | y, D) \{ E^2_i(y_F) - 2 \mu E_i(y_F) + \mu^2 \}.$$  \hspace{1cm} (3.9)

Noting that $E^2_i(y_F) = \sigma_i^2 + [E_i(y_F)]^2$ and $E_i(y_F) = \mu_i$, (3.9) becomes

$$p''(M_i | y, D) \{ \sigma_i^2 + \mu_i^2 - 2 \mu_i \mu + \mu^2 \}.$$  \hspace{1cm} (3.10)

The three right-hand terms inside the brackets of (3.10) may be factored yielding:

$$p''(M_i | y, D) \{ \sigma_i^2 + (\mu_i - \mu)^2 \}.$$  

Thus, $\sigma^2$ may be written as follows:

$$\sigma^2 = \sum_{i=1}^{N} p''(M_i | y, D) \{ \sigma_i^2 + (\mu_i - \mu)^2 \}.$$ \hspace{1cm} (3.11)
This is the same as equation (3.3). Defining $P''(M_1) = P''(M_1|y,D)$, (3.11) can be rewritten as follows:

$$\sigma^2 = \sum_{i=1}^{N} P''(M_i)\sigma^2_i + \sum_{i=1}^{N} P''(M_i)(\mu_i - \mu)^2.$$ (3.12)

Having defined $V(BMMP)$ and $V(BMSP)$, it is now possible to compare their magnitudes. Assuming, as will be done for the remainder of this dissertation unless otherwise noted, that the decision-maker's model space contains only two models, $M_1$ and $M_2$, the relative magnitudes of $V(BMMP)$ and $V(BMSP)$ will be examined for each of the following cases:

**CASE I**: $\sigma_1^2 = \sigma_2^2$.

**CASE II**: $\sigma_1^2 < \sigma_2^2$ and BMS chooses $M_1$.

**CASE III**: $\sigma_1^2 < \sigma_2^2$ and BMS chooses $M_2$.

For convenience, $P''(M_1)$ will be used in place of $P''(M_1|y,D)$ in the discussion and proofs of these cases and the lemmas that follow.

**THEOREM 1**: If $\sigma_1^2 = \sigma_2^2$, then $V(BMMP) \geq V(BMSP)$.

**PROOF**: When $N = 2$,

$$\sigma^2 = P''(M_1)\sigma_1^2 + P''(M_2)\sigma_2^2 + P''(M_1)(\mu_1 - \mu)^2 + P''(M_2)(\mu_2 - \mu)^2,$$

and when $\sigma_1^2 = \sigma_2^2$, $V(BMSP) = \sigma_1^2 = \sigma_2^2$.

---

1This assumption is made in order to simplify the analysis which follows. For a more precise explanation of this assumption, see Section III.2.4.
Since by definition $0 \leq P'(M_1), P'(M_2) \leq 1$, it follows trivially that when $\sigma_1^2 = \sigma_2^2$,

$$P'(M_1)\sigma_1^2 + P'(M_2)\sigma_2^2 = \sigma_1^2 = \sigma_2^2.$$ 

Thus, if

$$P'(M_1)(\mu_1 - \mu)^2 + P'(M_2)(\mu_2 - \mu)^2 \geq 0,$$

then $V(BMMP) \geq V(BMSP)$. Since $(\mu_1 - \mu)^2$ and $(\mu_2 - \mu)^2$ are nonnegative,

$$P'(M_1)(\mu_1 - \mu)^2 + P'(M_2)(\mu_2 - \mu)^2 \geq 0$$

and $V(BMMP) \geq V(BMSP)$. Unless $\mu_1 = \mu_2$, in which case $\mu_1 = \mu_2 = \mu$, $P'(M_1)(\mu_1 - \mu)^2 + P'(M_2)(\mu_2 - \mu)^2$ is strictly greater than zero and $V(BMMP)$ is strictly greater than $V(BMSP)$.

**Theorem 2:** If $\sigma_1^2 < \sigma_2^2$ and BiS chooses $M_1$, then $V(BMMP) \geq V(BMSP)$.

**Proof:** Refer to the proof of Theorem 1. Since $\sigma_1^2 < \sigma_2^2$,

$$P'(M_1)\sigma_1^2 + P'(M_2)\sigma_2^2 \geq \sigma_1^2.$$ 

From the proof of Case I,

$$P'(M_1)(\mu_1 - \mu)^2 + P'(M_2)(\mu_2 - \mu)^2 \geq 0.$$ 

Thus, it follows that

$$P'(M_1)\sigma_1^2 + P'(M_2)\sigma_2^2 + P'(M_1)(\mu_1 - \mu)^2 + P'(M_2)(\mu_2 - \mu)^2 \geq \sigma_1^2,$$

and

1 This dissertation is not concerned with special cases in which $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$. 


i.e., $V(BMMP) \geq V(BMSP)$. However, $V(BMMP)$ equals $V(BMSP)$ only if $P''(M_1) = 1$. But, if $P''(M_1) = 1$, there exists no model specification uncertainty. Thus, when model specification uncertainty exists, $V(BMMP)$ is strictly greater than $V(BMSP)$.

**THEOREM 3:** If $\sigma_1^2 < \sigma_2^2$ and BMS chooses $M_2$, then $V(BMMP) \leq V(BMSP)$.

**PROOF:** Refer to Theorem 1. Whenever $P''(M_2) \neq 1$,

$$P''(M_1)\sigma_1^2 + P''(M_2)\sigma_2^2 < \sigma_2^2.$$  

Therefore,

$$\sigma^2 = P''(M_1)\sigma_1^2 + P''(M_2)\sigma_2^2 + P''(M_1)(\mu_1 - \mu)^2 + P''(M_2)(\mu_2 - \mu)^2 \leq \sigma_2^2$$

depending on the size of $P''(M_1)(\mu_1 - \mu)^2 + P''(M_2)(\mu_2 - \mu)^2$.

Perhaps the most important thing that Theorems 1, 2, and 3 reveal is that if model specification uncertainty exists, $V(BMMP) \neq V(BMSP)$, except for uninteresting cases. This fact will be referred to repeatedly throughout Chapters III and IV. As will be seen in Section III.2, the inability to order $V(BMMP)$ and $V(BMSP)$ in Case III poses no problem with respect to comparing the relative merits of the BMC and BMS procedures as aids to forecasting. It does, however, make identification of whether the measure of forecast-risk provided the decision maker by the BMS procedure (defined in Section III.2.6 to be $V(BMSP)$) understates or overstates the actual forecast-risk faced by the decision maker.

This problem is discussed in Section III.2.6. In Sections III.2.6 and IV.3.2, it is shown that Case III may never arise, since situations exist in which only Case I applies.

The following three lemmas and the discussion that follows them are useful for helping to order $V(BMMP)$ and $V(BMSP)$ in situations
in which Case III applies. The first provides a necessary and sufficient condition for \( V_{BMMP} \) to be greater than \( V_{BMSP} \).

**Lemma 1:** If \( \sigma_1^2 < \sigma_2^2, \mu_1 \neq \mu_2 \), and BMS chooses \( M_2 \), then

\[
V_{BMMP} > V_{BMSP} \text{ if and only if } \quad \mathbb{P}(M_2 | y,D) > \frac{1}{(\mu_1 - \mu_2)^2} \cdot \frac{\sigma_2^2 - \sigma_1^2}{(\mu_1 - \mu_2)^2}.
\]

**Proof:** 1. If \( V_{BMMP} > V_{BMSP} \), it must be shown that

\[
\mathbb{P}(M_2) > \frac{1}{(\mu_1 - \mu_2)^2} \cdot \frac{\sigma_2^2 - \sigma_1^2}{(\mu_1 - \mu_2)^2}.
\]

Since \( V_{BMMP} = \sigma_1^2 = \mathbb{P}(M_1)\sigma_1^2 + \mathbb{P}(M_2)\sigma_2^2 + \mathbb{P}(M_1)(\mu_1 - \mu)^2 + \mathbb{P}(M_2)(\mu_2 - \mu)^2 \)

and \( V_{BMSP} = \sigma_2^2 \), \( V_{BMMP} > V_{BMSP} \) is the same as

\[
\mathbb{P}(M_1)\sigma_1^2 + \mathbb{P}(M_2)\sigma_2^2 + \mathbb{P}(M_1)(\mu_1 - \mu)^2 + \mathbb{P}(M_2)(\mu_2 - \mu)^2 > \sigma_2^2. \quad (3.13)
\]

Subtracting \( \mathbb{P}(M_1)\sigma_1^2 + \mathbb{P}(M_2)\sigma_2^2 \) from both sides of (3.13) yields

\[
\mathbb{P}(M_1)(\mu_1 - \mu)^2 + \mathbb{P}(M_2)(\mu_2 - \mu)^2 > \sigma_2^2 - [\mathbb{P}(M_1)\sigma_1^2 + \mathbb{P}(M_2)\sigma_2^2]. \quad (3.14)
\]

From (3.2) it is known that \( \mu = \mathbb{P}(M_1)\mu_1 + \mathbb{P}(M_2)\mu_2 \). Let the rhs of (3.14) equal \( R \), and define \( P_1 = \mathbb{P}(M_1) \) and \( P_2 = \mathbb{P}(M_2) \). Then substituting for \( \mu \) in (3.14) yields

\[
P_1(\mu_1 - P_1\mu_1 - P_2\mu_2)^2 + P_2(\mu_2 - P_1\mu_1 - P_2\mu_2)^2 > R. \quad (3.15)
\]

Noting that \( P_2 = 1 - P_1 \), (3.15) can be written

\[
P_1(\mu_1 P_2 - \mu_2 P_2)^2 + P_2(\mu_2 P_1 - \mu_1 P_1)^2 > R. \quad (3.16)
\]
Factoring $P_2$ out of the first term on the lhs of (3.16) and $P_1$ out of the second term on the lhs yields

$$P_2P_1P_2(\mu_1 - \mu_2)^2 + P_1P_1P_2(\mu_2 - \mu_1)^2 > R. \quad (3.17)$$

Noting that $P_1 + P_2 = 1$, and that $P_1P_2(\mu_1 - \mu_2)^2 = P_1P_2(\mu_2 - \mu_1)^2$, (3.17) becomes

$$P_1P_2(\mu_1 - \mu_2)^2 > \sigma_2^2 - (P_1\sigma_1^2 + P_2\sigma_2^2) = P_1(\sigma_2^2 - \sigma_1^2). \quad (3.18)$$

Dividing both sides of this inequality by $P_1(\sigma_2^2 - \sigma_1^2)$ yields the desired result

$$P_2 > \frac{(\sigma_2^2 - \sigma_1^2)}{(\mu_1 - \mu_2)^2}.$$ 

2. If $P_2 > \frac{(\sigma_2^2 - \sigma_1^2)}{(\mu_1 - \mu_2)^2}$, then $V(BMMP) > V(BMSP)$, i.e.,

$$P_1\sigma_1^2 + P_2\sigma_2^2 + P_1(\mu_1 - \mu)^2 + P_2(\mu_2 - \mu)^2 > \sigma_2^2.$$

A reversal of the steps in the first half of the proof leads immediately to this result.

Lemma 1 can be combined with Theorem 1 to form a necessary and sufficient condition for $V(BMMP)$ to be greater than $V(BMSP)$ when, say, $\sigma_1^2 < \sigma_2^2$, regardless of which model BMS selects.

**Lemma 2:** If $\sigma_1^2 < \sigma_2^2$, the $V(BMMP) > V(BMSP)$ if and only if

a) Model 1 is selected by BMS,

or

b) Model 2 is selected by BMS, $\mu_1 \neq \mu_2$, and
\[ P^*(M_2 | Y, D) > \frac{(\sigma_2^2 - \sigma_1^2)}{(\mu_1 - \mu_2)^2}. \]

**PROOF:** Lemma 2 results from combining Theorem 1 and Lemma 1, and its proof follows directly from their proofs.

It is clear that \( V(BMMP) > V(BMSP) \) whenever condition a or b of Lemma 2 is satisfied. Thus, upon examining condition b the following can be said:

1. Other things equal, the greater the distance between the means, \( \mu_1 \) and \( \mu_2 \), of the predictive distributions of the two models in question, the smaller is the rhs of the inequality of condition b, and the more likely it is that \( V(BMMP) > V(BMSP) \).

2. Other things equal, the closer in size are the predictive variances, \( \sigma_1^2 \) and \( \sigma_2^2 \), the smaller is the rhs of the inequality of condition b, and the more likely it is that condition b holds, i.e., the more likely it is that \( V(BMMP) > V(BMSP) \).

Both these statements apply irregardless of which model is chosen by BMS, i.e., whether it be the model with the lower or higher predictive variance.

As an example of how statements one and two might help determine the relationship between \( V(BMMP) \) and \( V(BMSP) \), the following is offered. Suppose the decision-maker's prior information about \( y \) leads him to believe that the predictive variances of both models are roughly equal, but that their predictive means differ
significantly. By Theorems 1 and 2 and statements one and two above, the decision maker should consider it more likely that $V(BMMP)$ exceeds $V(BMSP)$ than if he believed, say, that $\mu_1$ and $\mu_2$ were about the same size. This follows since a) if $\sigma_1^2$ in fact equals $\sigma_2^2$, then by Theorem 1 $V(BMMP) > V(BMSP)$; b) if, say, $\sigma_1^2 < \sigma_2^2$ and the BMS procedure chooses $M_1$, then Theorem 2 applies and $V(BMMP) > V(BMSP)$; and c) if $\sigma_1^2 < \sigma_2^2$ and the BMS procedure chooses $M_2$, then Theorem 3 applies and the decision-maker's prior information about $\sigma_1^2$, $\sigma_2^2$, $\mu_1$ and $\mu_2$ in concert with statements one and two above indicate that it is more likely that $V(BMMP)$ exceeds $V(BMSP)$ than if, say, the decision maker thought $\mu_1$ and $\mu_2$ were about the same size.

The next section utilizes the results of this section in comparing the effectiveness of the BMC, BMS, and maximize-$R^2$ approaches to forecasting.

III.2 Forecasting: Bayesian Model Comparison Versus Bayesian Model Selection and the Maximize-$R^2$ Rule

Most forecasting procedures handle model specification uncertainty suboptimally. Typically, a forecaster proposes a number of alternative statistical models as possible candidates to represent the data-generating process whose future value he is interested in predicting and then, via some model screening procedure, eliminates all but one model.\(^1\)

In this section, forecasting as accomplished via two model-screening procedures, Bayesian Model Selection (BMS) and the classical maximize-\( R^2 \) rule approach (max-\( R^2 \)), is compared to forecasting as handled by a procedure that optimally considers model specification uncertainty, the Bayesian Model Comparison approach (BMC). Before proceeding with the comparison a brief review of BMS, max-\( R^2 \), and BMC is in order.

III.2.1 The Bayesian Model Selection Procedure (BMS)

Bayesian Model Selection was discussed in some detail in Chapter II. Briefly, it requires the following:

1. The specification of a set of \( N \) alternative statistical models each of which purports to represent the data-generating process of interest to the forecaster.
2. The assessment of a prior probability mass function over the set of \( N \) models, \( P'(M_i) \), \( i=1,2,\ldots,N \).
3. The assessment of prior probability density functions over the parameters of each model, \( g'(\theta_i|M_i) \), \( i=1,2,\ldots,N \).
4. The specification of a likelihood function for each model, \( f(y|\theta_i,M_i,D_i) \), \( i=1,2,\ldots,N \).
5. The computation of posterior probabilities for the models, \( P''(M_i|y,D) \), \( i=1,2,\ldots,N \).

The posterior model probabilities are often used to select one model from among the set of \( N \) models to represent the data-generating

\(^1\)When thought of as a function of \( y \) with \( \theta_i \), \( M_i \), and \( D_i \) given, \( f(y|\theta_i,M_i,D_i) \) is model \( i \).
process of interest to the forecaster. The usual procedure is to select the model with the highest posterior model probability. In the event that the forecaster can estimate the loss that results from choosing an inappropriate model and can do this for each of the $N$ models,\footnote{Actually the decision-maker must be able to determine the loss from choosing model $i$ when model $j$ is the true model, $i \neq j$. There are $N(N - 1)$ such losses.} he can compute his expected loss from choosing each model and select the model which yields the lowest expected loss.

It should be noted that BMS may also be used for reasons other than for the selection of a single model from among a set of $N$ models. For example, if $N$ is large, BMS can be used to reduce the number of models in the model space to a number that can be more easily and inexpensively dealt with by a procedure such as BMC. This can be accomplished by eliminating all models from consideration whose posterior model probability is, say, less than some $\alpha$, $0 < \alpha < 1$. In this dissertation, however, BMS will be regarded as a procedure for selecting a single model from among $N$ alternative models.

The forecaster who uses BMS essentially handles his forecasting problem in a two-step sequence: first, a single model is chosen to represent the data-generating process; second, under the assumption that the chosen model is in fact a "true" reflection of the data-generating process, the forecaster addresses his prediction problem.
III.2.2 The Maximize-$R^2$ Rule

The maximize-$R^2$ rule is frequently used to choose one from among a set of alternative competing linear statistical models whose explanatory variables are nonrandom.\(^1\) The usual procedure is to estimate the parameters of each of the alternative models, compute each model's coefficient of determination, $R^2$, and then select as being the best representation of the data-generating process the model with the highest $R^2$. Forecasting is then carried out utilizing the chosen model as if it were in fact the true model.

It is important to reiterate the well-known fact that $R^2$ is inversely related to $S^2$, the estimate of the dependent variable's residual variance. A maximize-$R^2$ rule is therefore equivalent to a minimize-$S^2$ rule. In other words, the model with the maximum $R^2$ is also the model with the minimum $S^2$.

Geisel\(^2\) and Thornber\(^3\) have shown that under certain conditions model selection as accomplished via the max-$R^2$ rule is equivalent to the Bayesian Model Selection procedure. The conditions are the following:

1. The loss structure with respect to the selection of an incorrect model is symmetric. That is, if the loss from

---

\(^1\)For a more detailed discussion of the max-$R^2$ rule see Gaver and Geisel, pp. 52-53.

\(^2\)Geisel, pp. 24-37.

\(^3\)Thornber, Chapter 2.
choosing $M_i$ when $M_j$ is true is represented by $L_{ij}$, then $L_{ij} = L_{k\ell}$, for all $i,j,k,\ell = 1,2,\ldots,N$, with $i \neq j$ and $k \neq \ell$, and $L_{ij}, L_{k\ell} > 0$.

2. $P'(M_1) = P'(M_2) = \ldots = P'(M_N)$, i.e., the prior model probabilities are equal.

3. The statistical models in question are normal regression models each of which has the same number of parameters. The parameters of each are its regression coefficients, usually denoted by $\beta$'s, and its residual variance, $\sigma^2_{e_i}$. That each model has the same number of coefficients implies that each model has the same number of independent (explanatory) variables.

4. The prior density function for the parameters, $\beta_i$ and $\sigma_{e_i}$, is diffuse.

Geisel and Thornber used different forms for the diffuse prior density function for the parameters $\beta_i$ and $\sigma_{e_i}$, but both showed that selection of the model with the highest posterior probability is equivalent to selection of the model with the lowest $S^2$. Since the model with the lowest $S^2$ also has the highest $R^2$, Geisel and Thornber have shown that selection of a model via the BMS procedure is equivalent to selection via the maximize-$R^2$ rule.

Since a model's $R^2$ can be increased simply by adding more "explanatory" variables to the model, a maximize-$R^2$ rule is frequently used in place of the maximize-$R^2$ rule. $R^2$ is defined as follows:\footnote{See Gaver and Geisel, pp. 52-54.}

\footnote{See Gaver and Geisel, pp. 52-54.}
\[
\tilde{R}^2 = R^2 - \frac{(k - 1)}{(n - k)} (1 - R^2)
\]

where \( n \) is the sample size and \( k \) is the number of explanatory variables. The addition of variables will increase the model's \( \tilde{R}^2 \), adjusted coefficient of determination, if and only if the F statistic for the hypothesis that the added variables' coefficients are all zero is greater than one.\(^1\) Geisel showed that in the two-model case, model selection via the BMS procedure can be made equivalent to selection via the maximize-\( \tilde{R}^2 \) rule if the relationships between the parameters of \( M_1 \) and \( M_2 \) are appropriately specified. The required parameter relationships are, unfortunately, somewhat nonsensical. There are no known intuitively meaningful sets of assumptions under which the BMS procedure and the maximize-\( \tilde{R}^2 \) rule are equivalent.\(^2\)

In the remainder of this chapter the four conditions listed above apply, unless noted otherwise. Thus, to avoid redundancy, the maximize-\( R^2 \) rule will not be discussed directly in what follows but will be addressed indirectly through comments about the equivalent selection procedure, BMS. Since the BMS and maximize-\( R^2 \) procedures are equivalent only in that they select the same model, only comments concerning the fact that the BMS procedure actually chooses a model, or comments about which model it chooses, also apply to the maximize-\( R^2 \) procedure.

---


\(^2\) Geisel, pp. 41-45.
III.2.3 The Bayesian Model Comparison Procedure (BMC)

The Bayesian Model Comparison procedure was discussed in detail in Chapter II. Briefly, it requires the following:

1. The specification of a set of N alternative statistical models, each of which purports to represent the data-generating process of interest to the forecaster.
2. The assessment of a prior probability mass function over the set of N models, \( P'(M_i) \), \( i=1,...,N \).
3. The assessment of prior probability density functions over the parameters of each model, \( g'(\theta_i|M_i) \), \( i=1,...,N \).
4. The specification of a likelihood function for each model, \( f(y|\theta_i,M_i,D_i) \), \( i=1,...,N \).
5. The computation of posterior probabilities for the models (referred to as model probabilities), \( P''(M_i|y,D) \), \( i=1,2,...,N \).
6. The computation of the marginal distribution of future values of the data-generating process. (This distribution, as noted earlier, is a predictive distribution. It will be referred to herein as the Bayesian Mixed Model Predictive (BMMP).)

The first five requirements are the same as the five requirements of the Bayesian Model Selection procedure. It is the sixth requirement that distinguishes the Bayesian Model Comparison procedure from the Bayesian Model Selection procedure. Instead of choosing one of the N models, as does the BMS procedure, BMC models the data-generating process of interest with the BMMP distribution.
Recalling (2.14), the BMMP distribution is defined as follows:

\[
f(y|D) = \sum_{i=1}^{N} P'(M_i)[\int f(y|\theta_i,M_i,D_i)g'(\theta_i|M_i)d\theta_i]
\]

\[
= \sum_{i=1}^{N} P'(M_i)f(y|M_i,D_i).
\]

(3.19)  
(3.20)

All the terms denoted in (3.19) and (3.20) were defined in Chapter II, and the distributions denoted in (3.19) and (3.20) were redefined in the six requirements above.

After observing realizations of the data-generating process in question, the BMMP takes the form presented in (2.15):

\[
f(y_F|y,D,D_F) = \sum_{i=1}^{N} P''(M_i|y,D)[\int f(y_F|\theta_i,M_i,D_{F_i})g''(\theta_i|M_i,y,D_i)d\theta_i]
\]

\[
= \sum_{i=1}^{N} P''(M_i|y,D)f(y_F|M_i,y,D_i,D_{F_i}).
\]

(3.21)  
(3.22)

Recall that \(D = (D_1,D_2,\ldots,D_N)'\), where \(D_i\) is a vector containing the values of model \(i\)'s explanatory variables that correspond to the most recently observed \(y\) value. \(D_F = (D_{F1},D_{F2},\ldots,D_{FN})'\), where \(D_{Fi}\) is a vector containing the values of model \(i\)'s explanatory variables at the time the next \(y\) value is to be generated. From (3.20) or (3.22), it can be seen that a BMMP distribution is a weighted average -- or mixture -- of each model's predictive density of \(y_F\), \(f(y_F|M_i,y,D_i,D_{Fi})\).
The implications of parameter and residual uncertainty for prediction and decision making have been given considerable attention. See, for example, any of the following: Theil, Fisher, Brainard, Leland, Basu, Zellner, Barry and Horowitz, and/or Waud.1 As noted in Chapter II, the BMC procedure considers residual, parameter, and model specification uncertainty. Accordingly, if each model in the set of N competing models is viewed as a possible "parameter value" for the process of interest, the BMC procedure may be thought of as a means for extending the parametric analysis of prediction and decision-making problems to include consideration of the possibly widely differing predictive and decision-making implications of the competing models. Thus, just as a Bayesian can extend predictive analysis by explicitly allowing for parameter uncertainty instead of just using parameter estimates, the BMC procedure extends parametric analysis by explicitly considering model specification uncertainty.

A forecaster using the BMC procedure rather than, say, the BMS procedure, does not have to unnaturally divide the forecasting problem into two parts. He does not have to first select a model from the set

of N competing models and then, assuming the chosen model to be the
correct model of the process, proceed with his forecasting. He
computes the BMMP distribution for his set of models and uses it
directly to determine, say, point or interval predictions for future
values of y. The forecaster's BMMP distribution reflects his residual,
parameter, and model specification uncertainty, and any predictions
that he makes using his BMMP are made in light of all three types of
uncertainty and with the use of information bearing on any and all of
them. This point will be discussed in greater detail in Section
III.2.5.

The next section sets forth the specific assumptions under which
the BMC and BMS procedures will be compared in the remainder of the
chapter.

III.2.4 Model Space and Assumptions

The comparison of the BMC and BMS procedures (and indirectly
max-$R^2$) that follows will be based on the following assumptions:

1. The decision maker (forecaster) behaves as if he believes
that one or the other of the following two models is an
accurate representation of the random process of interest,
but he is unsure which model is appropriate:

\[ M_1: \quad y = \beta_1 X + \epsilon; \]
\[ M_2: \quad y = \beta_2 Z + \delta. \]

y is the variable whose future value the forecaster is
interested in predicting. X and Z are two different
explanatory variables. X and Z are random, but their
values associated with the next $y$ to be generated are known prior to $y$'s observation. $\beta_1$ and $\beta_2$ are unknown parameters. $\varepsilon$ and $\delta$ are the usual normally distributed error terms, each with mean zero and unknown variance, $\sigma^2_\varepsilon$ and $\sigma^2_\delta$, respectively. It is also assumed that $\text{cov}(\beta_1, \varepsilon) = \text{cov}(\beta_2, \delta) = \text{cov}(\varepsilon, \delta) = 0$. Thus, $M_1$ and $M_2$ are normal univariate regression models which, to keep the number of each model's unknown parameters at two, have been forced through the origin. Since the values of the explanatory and dependent variables can always be scaled so that $M_1$ and $M_2$ pass through the origin, no generality is lost by using models without intercept terms. Care must be taken, however, to interpret results in the appropriate units.

2. The random process of interest to the forecaster is stationary.

3. $X$ and $Z$ are uncorrelated and only the explanatory variable in the true model affects $y$. Thus, if $M_1$ were the true model, $\beta_2$ would be zero. If neither $M_1$ nor $M_2$ were the true model, it may be that $\beta_1 = \beta_2 = 0$.

4. In comparing the BMC and BMS procedures, it will be assumed that the forecaster may have prior information about the parameters of $M_1$ and $M_2$. Since model selection via the BMS procedure and the maximize-$R^2$ rule are equivalent only if the forecaster has no prior information about the parameters
of the models, any comments made about the BMS procedure under this assumption do not apply to the maximize-\(R^2\) rule.\(^1\)

Note that in assumption one above the residual variance of each model is assumed to be unknown. It would be unrealistic to assume the residual variance to be known when the correct model of the process is not known. Further if \(\sigma^2_x\) and \(\sigma^2_o\) were known, or were assumed to be known, and the correct model was known to be either \(M_1\) or \(M_2\), the correct model could be selected by the forecaster with probability one and there would be no need for procedures such as BMC or BMS.

To illustrate, consider the following argument. For a given \(X\) value the conditional variance of \(y\), \(\sigma^2_y|X\), is \(\sigma^2\). For a given value of \(Z\) the conditional variance of \(y\), \(\sigma^2_y|Z\), is \(\sigma^2_o\). The marginal variance of \(y\) (i.e., \(y\)'s variance unconditioned on \(X\)), \(\sigma^2_y\), as described by \(M_1\) is \(\beta^2_x \sigma^2_x + \sigma^2_o\) and the marginal variance of \(y\) as described by \(M_2\) is \(\beta^2_2 \sigma^2_Z + \sigma^2_o\). If \(M_1\) were in fact the true model, then

\[
\sigma^2_y = \beta^2_x \sigma^2_x + \sigma^2_o,
\]

\[
\sigma^2_y|X = \sigma^2\,
\]

and

\[\beta_2 = 0.\]

Since \(\beta_2 = 0\), the marginal variance of \(y\) as described by \(M_2\) is simply \(\sigma^2_o\). Thus, since the marginal variance of \(y\) is now known to be \(\beta^2_x \sigma^2_x + \sigma^2_o\), it follows that \(\sigma^2_o = \beta^2_2 \sigma^2_Z + \sigma^2_o\). This says that when \(M_1\) is the true model \(\sigma^2_o < \sigma^2_o\). Consequently, if it is assumed that \(\sigma^2_x\) and \(\sigma^2_o\)

---

\(^1\)The specific conditions under which the BMS and the max-\(R^2\) approaches to model selection are equivalent were listed in Section III.2. Only assumption four of this section affects their equivalency.
are known, the model with the lower residual variance can be identified with probability one as being the true model.

In the next section the BMS and BMC procedures are compared with respect to how well each accounts for a forecaster's model specification uncertainty.

III.2.5 The Treatment of Model Specification Uncertainty

Assuming that the random process of interest is stationary and that one of a proposed set of alternative models is a true representation of the process, Geisel has shown that in the limit the BMMP and BMS predictive distributions are the same. Thus, in the limit, the BMC and BMS approaches to forecasting are equivalent. This result is demonstrated below.

Recalling (2.15), a BMMP can be written as a weighted average of model predictives:

\[ f(y_F|y,D,D_F) = \sum_{i=1}^{N} P^u(M_i|y,D)f(y_F|M_i,y,D_i,D_F). \]  

(3.23)

Each of the individual model predictives, \( f(y_F|M_i,y,D_i,D_F) \), is the distribution that would be used to characterize the random process in question if the BMS procedure chose \( M_i \).

Geisel has shown that if \( M_i \) is in fact the true model, then as sample evidence accumulates (i.e., as \( n \to \infty \)) \( P^u(M_i|y,D) \) approaches one. It follows trivially that as \( n \) approaches \( \infty \), \( f(y_F|y,D,D_F) \) approaches

---

1 Geisel, pp. 22-23.

2 Ibid.
Thus, since the distribution yielded by the BMC procedure to forecast future values of y is \( f(y_F|y,D,D_F) \), and that yielded by the BMS procedure for forecasting purposes is \( f(y_F|y,D_1, D_{F_1}) \), in the limit the BMC and BMS procedures are equivalent forecasting procedures. This unsurprising result says that in the limit, under the assumed conditions, truth is obtained, i.e., the accumulated data would indicate with certainty the model that had been generating the data. If such were the case, everybody would ultimately use the same—correct—model to predict future values of y.

In both the BMS and BMC procedures the forecaster or decision maker proposes a set of N models each of which he believes might correctly represent the random process whose future values he is interested in predicting. Theoretically, if he assesses a nonzero probability for a particular model, that model should be included in his model space. In both the BMS and BMC procedure the forecaster assesses a prior probability mass function over the N models in his model space. By so doing the forecaster is formally acknowledging the fact that he is uncertain as to the correct model. He is thus faced with a forecasting problem in which model specification uncertainty is present and must be dealt with.
By selecting one of the N models and assuming it to be true, the BMS approach to forecasting yields predictions that do not appropriately reflect the forecaster's model specification uncertainty. The BMMP of the BMC procedure, however, by utilizing all N model predictions and their associated model probabilities acknowledges the forecaster's model specification uncertainty and yields predictions that do reflect this uncertainty. Forecasting via the BMS procedure should therefore be regarded as an approximation to the "optimal" approach to forecasting offered by the BMC procedure.

In the next section of this chapter the risk involved in forecasting via the BMC procedure is compared to that involved in forecasting via the BMS procedure. These risks are measured by \( V(\text{BMMP}) \) and \( V(\text{BMSP}) \), respectively.

III.2.6 Risk Specification

Forecasts are frequently used as inputs to decision-making problems. For example, predicted new-car demand might be used by an auto manufacturer in determining the rate and timing of automobile production, as well as the size of his labor force. Much of the risk taken by a decision maker in making a decision that utilizes a forecast stems from the possibility of forecasting error. If, for example, the forecasted new-car demand errs on the high side, both the manufacturer and many of his distributors might be burdened with an excess stock of cars, leading to unnecessarily high inventory costs. The risk passed on to a decision maker by a forecaster, called here forecast-risk, will be assumed to be adequately measured in terms of
the variance of the forecaster's predictive distribution. Such an assumption would be appropriate, for example, if losses associated with forecast errors were proportional to the squared error of the forecast.

A forecaster that utilizes the BMS or BMC procedure is admitting that he is uncertain of the specification of the process whose future values he wishes to predict. It has been noted above that this uncertainty is fully reflected in a BMMP distribution but not in a Bayesian Model Selection Predictive (BMSP) distribution. Thus, unless \( V(\text{BMMP}) = V(\text{BMSP}) \), or if no model specification uncertainty exists, \( V(\text{BMSP}) \) is an inappropriate measure of forecast-risk, either under or overstating it as \( V(\text{BMMP}) > V(\text{BMSP}) \) or \( V(\text{BMMP}) < V(\text{BMSP}) \).

Thus, the decisions that utilize a prediction arrived at via the BMS procedure will have been made under the assumption that the risk involved is either less than or greater than it is in reality. The BMS procedure, therefore, has the potential to provide the decision maker with information that may lead him to generate inappropriate and excessively costly decisions.

As seen in Cases I, II, and III of Section III.1, \( V(\text{BMMP}) \) may be greater than or less than \( V(\text{BMSP}) \). In certain situations it is more likely that \( V(\text{BMMP}) \) is greater than \( V(\text{BMSP}) \), and in others it is always the case that \( V(\text{BMMP}) \) is greater than \( V(\text{BMSP}) \). Such situations will be discussed below.

It was noted in Section III.2.2 that a model's posterior probability is inversely related to its estimated residual variance, \( S_i^2 \).
and, therefore, directly related to its coefficient of determination, \( R^2 \). Thus, if \( M_1 \)'s posterior probability is high relative to \( M_2 \)'s posterior probability, then \( S_1^2 \) is low relative to \( S_2^2 \), and \( R_1^2 \) is high relative to \( R_2^2 \). If such were the case, it could be said that the accumulated evidence supports \( M_1 \) rather than \( M_2 \) as being the more likely data-generating source. Accordingly, a forecaster might be tempted to invoke the BMS procedure or the maximize-\( R^2 \) rule and choose \( M_1 \) and its predictive distribution with which to forecast \( y_F \). But in such cases it is more likely that \( V(\textrm{BMMP}) > V(\textrm{BMSP}) \) than it would be if the evidence did not so clearly support one model or the other.\(^1\) This is explained below.

CLAIM: If \( \frac{|\sigma_2^2 - \sigma_1^2|}{(u_2 - u_1)^2} \) remains constant, the larger the difference in \( P''(M_1) \) and \( P''(M_2) \), the more likely that \( V(\textrm{BMMP}) > V(\textrm{BMSP}) \).

DISCUSSION: Zellner has shown that for a normal regression model (see the assumptions of Section III.2.4) with diffuse prior information on the parameters of the model, \( V(\textrm{BMSP}) \), also denoted \( \sigma_1^2 \), is defined as follows:\(^2\)

\[
\sigma_1^2 = \frac{(n - 1)S_1^2}{(n - 3)} \left[ \frac{D_{F1}^2}{N} + 1 \right] \sum_{i=1}^{j} D_{ji}^2
\]

\((3.24)\)

\(^1\)From (3.12) it can be seen that when, say, \( P''(M_1) \) is close to one, the difference between \( V(\textrm{BMMP}) \) and \( V(\textrm{BMSP}) \) is of no practical significance. Under such circumstances a comparison of \( V(\textrm{BMMP}) \) and \( V(\textrm{BMSP}) \) serves little purpose.

\(^2\)Zellner, pp. 72-74.
where \( n \) is the sample size, i.e., the number of \( y \) values observed to date; the \( D_{ji} \)'s are the values of model \( i \)'s independent (explanatory) variable, \( D_i \), observed to date; \( D_{F_i} \) is the value of \( D_i \) that corresponds to the next \( y \) value generated by the process in question; \( S_i^2 \) is the estimated residual variance of model \( i \). It can be seen from (3.24) that \( \sigma_i^2 \) is proportional to \( S_i^2 \).

It is known from Geisel's work that \( P''(M_i) \) is inversely proportional to \( S_i^2 \). Thus, the larger \( |P''(M_1) - P''(M_2)| \), the larger is \( |S_1^2 - S_2^2| \).

Conditions 2a and 2b of Section III.1 provide necessary and sufficient conditions for \( V(BMMP) > V(BMSP) \). The conditions are that if, say, \( \sigma_1^2 < \sigma_2^2 \), then \( P''(M_1) \) must be greater than .5 or \( P''(M_2) \) must be greater than \( \frac{\sigma_2^2 - \sigma_1^2}{(\mu_2 - \mu_1)^2} \). Thus, other things equal, if \( P''(M_1) < .5 \), then the larger is \( [P''(M_2) - P''(M_1)] \), the more likely it is that \( P''(M_2) \) satisfies either condition 2a or 2b, i.e., the more likely it is that \( V(BMMP) > V(BMSP) \). Of course if \( P''(M_1) > .5 \), then \( V(BMMP) \) is greater than \( V(BMSP) \) regardless of how large \( [P''(M_1) - P''(M_2)] \) is.

The phrase "other things equal" used above refers specifically to the ratio of \( |\sigma_2^2 - \sigma_1^2| \) to \( (\mu_2 - \mu_1)^2 \). What is being said is that given two model selection situations in which the absolute value of the ratio of \( (\sigma_2^2 - \sigma_1^2) \) to \( (\mu_2 - \mu_1)^2 \) is the same in both, but that in the first situation \( |P''(M_1) - P''(M_2)| \) is larger than it is in the

---

1See Section III.2.2.

2The "other things" are clarified in the next paragraph.
second, then it is more likely in the first situation that \( V(\text{BMMP}) > V(\text{BMSP}) \).

This claim can be supported from another angle. Since \( \sigma_1^2 \) is proportional to \( S_i^2 \), it can be said that the smaller, say \( S_1^2 \) is in relation to \( S_2^2 \), the more likely it is that \( \sigma_1^2 < \sigma_2^2 \). By the Geisel result discussed in Section III.2.2, the smaller is \( S_1^2 \) in relation to \( S_2^2 \), the larger is \( P'(M_1) \) in relation to \( P'(M_2) \). Thus, the smaller \( S_1^2 \) is in relation to \( S_2^2 \), the more likely it is that the model with the lower predictive variance will be chosen by the BMS procedure. Therefore, by Theorem 1 of Section III.1, the more likely it is that \( V(\text{BMMP}) \) is greater than \( V(\text{BMSP}) \).

There is a special forecasting case worth noting in which \( V(\text{BMMP}) \) is greater than \( V(\text{BMSP}) \) no matter which model the BMS procedure chooses. It is a result of the following lemma.

**LEMMA 3:** If \( \frac{X_F^2}{\sum_{j=1}^{n} x_j^2} = \frac{Z_F^2}{\sum_{j=1}^{n} z_j^2} \), then the model with the lower estimated residual variance \( S_i^2 \), also has the lower predictive variance, \( \sigma_i^2 \).

**PROOF:** Proof of this lemma follows directly from the definition of \( \sigma_i^2 \). Recalling (3.24) and the model space assumptions of Section III.2.4, \( \sigma_1^2 \) and \( \sigma_2^2 \) are defined as follows:

\[
\sigma_1^2 = \frac{(n - 1)S_1^2}{(n - 3)} \left( \frac{X_F^2}{\sum_{j=1}^{n} x_j^2} + 1 \right)
\]

(3.25)
Thus, since \( n \), the sample size, is a constant, and \( \frac{X_F^2}{\sum Z_j^2} \) is assumed equal to \( \frac{Z_F^2}{\sum Z_j^2} \), \( \sigma_1^2 \) and \( \sigma_2^2 \) are proportional to \( S_1^2 \) and \( S_2^2 \), respectively.

Since the model chosen by BMS has the smaller estimated residual variance, by Lemma 3 it also has the lower predictive variance. Thus if Lemma 3 holds, by Theorem 2 of Section III.1 \( V(BMMP) > V(BMSP) \). In this special case, a decision maker using a forecast obtained via the BMS procedure would be making a decision that fails to recognize the full extent of the uncertainty involved in the outcome of his decision.

Under the assumptions of Section III.2.4, Zellner has shown that the posterior expected value of the residual variance of, say, Model 1 is

\[
E''(\sigma_e^2) = \frac{(n - 1)S_1^2}{(n - 3)} ,
\]

and Raiffa and Schlaifer have shown that the posterior variance of, say, \( \beta_1 \) is

\[
V''(\beta_1) = \frac{(n - 1)S_1^2}{(n - 3) \sum x_j^2} .
\]

Thus, recalling (3.25), the predictive variance of model 1 may be written

\[\sigma_2^2 = \frac{(n - 1)S_2^2}{(n - 3)} \left( \frac{Z_F^2}{n \sum Z_j^2} + 1 \right) \]

\[
(3.26)
\]

\[1\] Zellner, p. 62.

\[ \sigma_1^2 = V''(\beta_1)X_F^2 + E''(\sigma_1^2). \]  

(3.29)

The following lemma, based on the above facts, is offered to further explain the relationship between \( V(BMMP) \) and \( V(BMSP) \):

**Lemma 4:** If \( E''(\alpha_1^2) < E''(\alpha_2^2) \), \( \sum_{j=1}^{n} X_j^2 > \sum_{j=1}^{n} Z_j^2 \), and \( X_F \leq Z_F \), then \( V''(\beta_1) < V''(\beta_2) \) and \( \sigma_1^2 < \sigma_2^2 \).

**Proof:** From equation (3.27) it can be seen that \( E''(\alpha_1^2) \) and \( E''(\alpha_2^2) \) are proportional to \( S_1^2 \) and \( S_2^2 \), respectively. Thus, \( E''(\alpha_1^2) < E''(\alpha_2^2) \) means that \( S_1^2 < S_2^2 \). From (3.28) it can be seen that \( V''(\beta_1) \) and \( V''(\beta_2) \) are inversely related to \( \sum_{j=1}^{n} X_j^2 \) and \( \sum_{j=1}^{n} Z_j^2 \), respectively. Consequently, if \( S_1^2 < S_2^2 \) and \( \sum_{j=1}^{n} X_j^2 > \sum_{j=1}^{n} Z_j^2 \), it can be seen from (3.28) that \( V''(\beta_1) < V''(\beta_2) \). Thus, since \( V''(\beta_1) < V''(\beta_2) \), \( E''(\alpha_1^2) < E''(\alpha_2^2) \), and \( X_F \leq Z_F \), it follows from equation (3.29) that \( \sigma_1^2 < \sigma_2^2 \).

If the conditions of Lemma 4 are fulfilled, the model selected by the BMS procedure will have the lower predictive variance and by Theorem 2 of Section III.1, \( V(BMMP) > V(BMSP) \). Thus, as is the case when Lemma 3 holds, a decision maker using a forecast obtained via the BMS procedure would be making a decision which fails to recognize the full extent of the uncertainty involved in the outcome of his decision.

---

\( S_1^2 \) and \( S_2^2 \) could, of course be substituted for \( E''(\alpha_1^2) \) and \( E''(\alpha_2^2) \), respectively, but one of the goals of this lemma is to explain the relationship of \( V(BMMP) \) and \( V(BMSP) \) via the, perhaps, more easily interpretable definition of \( \sigma_1^2 \): \( \sigma_1^2 = V''(\beta_1)X_F^2 + E''(\sigma_1^2). \)
The next section examines the decision-maker's posterior expected losses from utilizing BMS and BMC-generated predictions of $y_F$.

III.2.7 A Comparison of Expected Losses

Given a loss function, sample $y$ values, and a predictive distribution of $y$, a forecaster can find an optimal point estimate for $y$ by minimizing the decision-maker's posterior expected loss:

$$
\min_{\hat{y}} \int_{-\infty}^{\infty} L(y, \hat{y}) f(y|y,D,D_F) dy
$$

(3.30)

It is well known that if a quadratic loss function is used in (3.30), the solution to the minimization problem is the mean of $f(y_F|y,D,D_F)$. If the forecaster chooses to forecast via the BMS procedure he would utilize a model predictive, $f(y_F|M_1,y,D_i,D_{F1})$, to solve (3.30). The solution to (3.30) and his point estimate for $y_F$ would therefore be the mean of his model predictive, $\mu_1$. If he chooses to forecast via the BMC procedure, he would use a BMMP, $f(y_F|y,D,D_F)$, to solve (3.30) and his solution and point estimate would be the mean of the BMMP, $\mu$. As has been mentioned several times earlier in this chapter, however, a forecaster who opts for forecasting via BMS is not making use of all the available information about $y_F$. The Bayesian Mixed Model Predictive (BMMP) of the BMC procedure reflects all the available information, whereas a BMSP is merely an approximation to the BMMP. Therefore, the appropriate predictive distribution to use in (3.30) is a BMMP. Consequently, the optimal solution to (3.30) is $\mu$, the mean of the BMMP, i.e., $\hat{y} = \mu$. Only if the forecaster and/or decision
maker assess a probability of one for a particular model being the true model of \( y_F \)'s process would a single model predictive provide full information to the forecaster and/or decision maker and, hence, an optimal solution to (3.30).\(^1\)

Since the appropriate distribution to use in solving (3.30) is a BMMP, the decision-maker's posterior expected loss using a BMS forecast, \( \mu_i \), is greater than his posterior expected loss using a BMC forecast, \( \mu \):

\[
\text{EL}(\mu_i) = \int_{-\infty}^{\infty} L(y_F, \mu_i)f(y_F | y, D, D_F)dy_F > \text{EL}(\mu)
\]

\[
= \int_{-\infty}^{\infty} L(y_F, \mu)f(y_F | y, D_i, D_{F_i})dy_F.
\] (3.31)

This follows from the fact that it is \( \mu \), and not \( \mu_i \), that minimizes

\[
\int_{-\infty}^{\infty} L(y_F, \hat{y})f(y_F | y, D, D_F)dy_F.
\] (3.32)

When \( P(M_i) > 0, i=1,2 \), then only if \( \mu_1 = \mu_2 \) would, say, \( \mu_1 \), minimize (3.32) since then \( \mu_1 = \mu_2 = P'(M_i)\mu_1 + P'(M_2)\mu_2 = \mu \). Of course if for some \( i \) \( P(M_i) = 1 \), then \( \mu_i = \mu \) also. But in the context of this dissertation, this case is of no interest.

Let \( C(\text{BMC}) \) and \( C(\text{BMS}) \) stand for the costs required to forecast with BMC and BMS, respectively.\(^2\) Then, assuming that the decision maker's loss function and the cost functions \( C(\text{BMC}) \) and \( C(\text{BMS}) \) can be

\(^1\)Note that when \( P(M_i) = 1 \), the BMSP and BMMP distribution are the same.

\(^2\)In general \( C(\text{BMC}) \) and \( C(\text{BMS}) \) cannot be computed without going through the actual computations required by the BMC and BMS procedures.
meaningfully compared, if experimentation with the BMC and BMS procedures shows that in general

$$EL(\mu_i) - EL(\mu) > C(BMC) - C(BMS),$$

it is materially as well as theoretically advantageous for the forecaster to use the BMC procedure rather than the BMS procedure.

Future values of a random variable are typically predicted using point or interval estimates. The implications of making point and interval estimates via the BMS procedure as opposed to the BMC procedure are discussed in the next two sections.

III.2.8 Implications for Point Estimation

The point estimate of a future value of some random process will be denoted by $\hat{y}_F$. The use of loss functions to determine optimal point estimates was discussed in the preceding section of this chapter. If a loss function can be specified by the forecaster and/or decision maker, it should be used to determine $\hat{y}_F$. Frequently, however, loss functions are too costly to develop and predictions must be made without the information that a loss function provides. In such cases forecasters usually examine $y_F$'s predictive distribution and choose a measure of its central tendency as their estimate of $y_F$. Their logic is that central tendency measures are usually in the high density region of the distribution and will not err significantly even if the actual $y_F$ falls in a tail of $y_F$'s predictive distribution. Further, it is
well known that commonly-used loss functions often result in mean, median, or modal estimates of parameters.

In the preceding section, it was noted that if the BMS procedure and a quadratic loss function are utilized for forecasting, \( \hat{y}_F = \mu_1 \). However, even if a BMS forecaster does not have a loss function with which to work, he might again choose the mean of the chosen model predictive, \( \mu_1 \), as his point estimate of \( y_F \). In either of these cases, if the BMS procedure chooses \( M_1 \) and \( \mu_1 \neq \mu_2 \),\(^1\) then, for reasons explained below, it can be said that the forecaster's point estimate is inappropriately high or low with probability one. For example, if \( \mu_1 < \mu \), and \( \mu_1 \) is used by a BMS forecaster to predict \( y_F \), \( \mu_1 \) is said to be an inappropriately low prediction of \( y_F \).

Suppose it is the next \( y \) value that the forecaster would like to predict. By assessing nonzero model probabilities for \( M_1 \) and \( M_2 \), as is done in both the BMS and BMC procedures, the forecaster/decision maker is acknowledging that he believes the next observation could be generated by either \( M_1 \) or \( M_2 \). A prediction of the next \( y_F \) value should acknowledge this uncertainty. But forecasting procedures that utilize the BMS procedure do not optimally account for this sort of uncertainty (model specification uncertainty) because they do not appropriately reflect the possibility that a rejected model may be the true model. Thus, in the example of the preceding paragraph, \( \mu_1 \) is said to be an inappropriately low forecast because it does not appropriately reflect the fact that \( y_F \) may be generated by \( M_2 \).

\(^1\)Since \( \mu = P''(M_1)\mu_1 + P''(M_2)\mu_2 \) and \( P''(M_1), P''(M_2) > 0, \mu_1 \neq \mu_2 \) means \( \mu_1 \neq \mu_2 \neq \mu \).
Forecasts made utilizing the BMC procedure do reflect model specification uncertainty. \( \mu \), the mean of the BMMP distribution is an example of a BMC-generated prediction. As can be seen by examining its definition, \( \mu \) reflects the belief that \( y_F \) may be generated by either \( M_1 \) or \( M_2 \):

\[
\mu = P^n(M_1)\mu_1 + P^n(M_2)\mu_2.
\]

Since the decision maker's predictive distribution is a mixture of the model predictions, his optimal estimator will arise from the mixture as well, and in this case will be \( \mu \). It is just as appropriate to use \( \mu \) when model specification uncertainty exists as it is to use, say, \( \mu_1 \) when it is known that \( y_F \) will be generated by \( M_1 \).

If a forecaster's loss function is asymmetric, the mean of \( y_F \)'s predictive distribution would not be appropriate for forecasting \( y_F \). Suppose his losses are best represented by an asymmetric linear loss function and model specification uncertainty exists. Then his optimal point estimate for \( y_F \) would be a fractile of \( y_F \)'s BMMP distribution.\(^1\)

A BMS forecaster utilizing an asymmetric linear loss function would use a fractile of the BMSP distribution. If the asymmetric linear loss function describes losses from underestimating \( y_F \) as being greater than losses from overestimating \( y_F \), the BMS forecaster's point estimate would be a fractile of the BMSP distribution which is greater than the

\(1\) If the linear loss function were symmetric, the optimal point estimate would be the median of the BMMP.
mean. In such cases the BMS forecaster may seriously underestimate \( y_F \) and incur a large loss while thinking is is protecting against such an occurrence. Suppose \( \mu_1 < \mu \), and the BMS procedure selects \( M_1 \). Then, if the forecaster chooses to estimate \( y_F \) with a fractile of \( M_1 \)'s BMSP distribution which is less than \( \mu \), say, the .7 fractile, the BMSP reflects his probability of underestimating \( y_F \) as being only .3. But, if the .7 fractile of the BMSP distribution is less than \( \mu \), the BMMP distribution reflects his probability of underestimating \( y_F \) as being greater than .5. Thus, a BMS forecaster may believe he is protecting against underestimating \( y_F \) when in fact he has a higher probability of an underestimate than an overestimate.

The results of this section were generated via a comparison of BMC and BMS forecasts. It should be noted, however, that point estimates determined by any procedure which utilizes a single model that has been selected from a set of viable models will typically be misplaced. This is due to the fact that use of a single model, however selected, has the effect of ignoring information provided by those remaining models which have positive posterior probability.

III.2.9 Implications for Interval Estimation

The procedure of predicting that a future value of a random process will take on a value between two specified real numbers with \( ^1 \)Raiffa and Schlaifer, p. 345, have shown that the predictive distributions for \( y_F \) yielded by \( M_1 \) and \( M_2 \) are Student. Since the Student distribution is unimodal and symmetric, its mean and median are equal.
some positive probability is referred to as Bayesian interval estimation. The interval represented by the two given numbers is called a credible interval. Often, a Bayesian will choose as his credible interval a Highest Posterior Density (HPD) region.\footnote{Bayesian methods for optimal interval estimates exist when, as in the case of point estimation, appropriate loss functions may be specified. See R. L. Winkler, "Decision-Theoretic Approach to Interval Estimation," Journal of the American Statistical Association, 67 (1972), 187-191.} Denoting \( y_F \)'s predictive distribution as \( f(y_F|y) \), an interval \( I \) in the domain of \( y_F \) is called a HPD region of content \( 1 - \alpha \) if

\[
\begin{align*}
& a) \quad P(y_F \in I) = 1 - \alpha \\
& b) \quad y_{F1} \in I \text{ and } y_{F2} \notin I \text{ implies } f(y_{F1}|y) \geq f(y_{F2}|y). \quad 2
\end{align*}
\]

BMS interval forecasts of \( y_F \) are determined from the predictive distribution of \( y_F \) generated by the model chosen by the BMS procedure, i.e., a Bayesian Model Selection Predictive (BMSp). BMC interval forecasts of \( y_F \) are determined from the appropriate Bayesian Mixed Model Predictive (BMMP).

Recall that under the assumptions of Section III.2.4, \( M_1 \) and \( M_2 \) define unimodal, symmetric distributions (Student distributions). Accordingly, a HPD credible interval determined from \( M_1 \)'s BMSp will be centered at \( \mu_1 \). Thus, when model specification uncertainty exists and \( \mu_1 \neq \mu_2 \), the midpoint of a BMS credible interval is inappropriately high or low in the same sense as BMS point estimates were in the

\footnote{George E. P. Box and George C. Tiao, Bayesian Inference in Statistical Analysis (Reading, MA: Addison-Wesley, 1973), p. 123.}
preceding section. The discussion of this phenomenon with respect to point estimates in the preceding section applies equally well here.

Under the assumptions that \( M_1 \) and \( M_2 \) are normal regression models, \( \mu_1 \neq \mu_2 \), and \( P'(M_1), P'(M_2) > 0 \) (see Section III.2.4), the BMMP distribution is bimodal. Accordingly, an HPD BMC credible region will frequently consist of two intervals; one with midpoint \( \mu_1 \), the other with midpoint \( \mu_2 \). Interval forecasts that are comprised of more than one interval will be referred to as split-interval forecasts or split credible intervals. An HPD split credible interval serves to warn a decision maker that it is highly probable that \( y_F \) will take on a value in one of two or more noncontiguous regions.

The following two lemmas demonstrate how a credible interval formed using a BMSP can be misleading when model specification uncertainty exists. In Lemma 5, the intersection of the BMSP's of \( M_1 \) and \( M_2 \) between their modes is referred to as the inter-modal intersection. The \( y_F \) value that corresponds to the inter-modal intersection will be denoted \( y_F^I \).

**Lemma 5:** Let \( \mu_1 \neq \mu_2 \) and suppose BMS chooses model i. If the length of a credible interval formed using the BMSP is less than or equal to \( 2|\mu_1 - y_F^I| \), then the BMS credible interval overstates the probability that it will cover \( y_F \).

**Proof:** Recall that the BMMP is a mixture of predictive distributions generated by \( M_1 \) and \( M_2 \):
\[ f(y_F | y, D, D_F) = P''(M_1 | y, D)f(y_F | M_1, y, D_1, D_{F_1}) \\
+ P''(M_2 | y, D)f(y_F | M_2, y, D_2, D_{F_2}) \]

Thus,
\[ f(y_F | M_1, y, D_1, D_{F_1}) \geq f(y_F | y, D, D_F). \]

When \( P''(M_1 | y, D) \neq 1 \), then
\[ f(y_F | M_1, y, D_1, D_{F_1}) > f(y_F | y, D, D_F). \]

If, say, \( u_1 < u_2 \) and \( y_F < \bar{I} \), then
\[ f(y_F | M_1, y, D_1, D_{F_1}) > f(y_F | y, D, D_F). \]

Thus, the probability of an interval centered on \( u_i \) of length less than \( 2|u_1 - \bar{I}| \) containing \( y_F \) is greater when the probability is evaluated via \( f(y_F | M_1, y, D, D_{F_1}) \), rather than \( f(y_F | y, D, D_F) \).

If the conditions of Lemma 5 are fulfilled, the probability of a BMS credible interval covering \( y_F \) is actually smaller than claimed by the forecaster using the BMS credible interval. Thus, the BMS credible interval overstates the probability of \( y_F \) being covered and therefore understates the risk involved in using the interval forecast for decision-making purposes. Notice that since \( f(y_F | y, D, D_F) > f(y_F | M_1, y, D_1, D_F) \) when \( y_F > \bar{I} \), it is unclear whether a BMS credible interval of length greater than \( 2|u_1 - \bar{I}| \) understates or overstates the probability that it will cover \( y_F \).

**Lemma 6:** If \( u_1 = u_2 \) and the BMS procedure chooses the model with the higher (lower) predictive variance, then a BMS credible interval under-
states (overstates) the probability that it will cover \( y_F \).

**PROOF:** Theorem 3 of Section III.1 showed that if the BMS procedure chooses the model with the higher predictive variance, then \( V(\text{BMMP}) \) may be less than, greater than, or equal to \( V(\text{BMSP}) \). Recall that

\[
V(\text{BMMP}) = \sigma^2 = p''(M_1 | y, D)\sigma_1^2 + p''(M_2 | y, D)\sigma_2^2 + p''(M_1 | y, D)(\mu_1 - \mu)^2
\]

\[+ p''(M_2 | y, D)(\mu_2 - \mu)^2\]

and

\[
E(\text{BMMP}) = \mu = p''(M_1 | y, D)\mu_1 + p''(M_2 | y, D)\mu_2.
\]

Thus, \( \mu_1 = \mu_2 \) implies that \( \mu_1 = \mu_2 = \mu \) and

\[
\sigma^2 = p''(M_1 | y, D)\sigma_1^2 + p''(M_2 | y, D)\sigma_2^2.
\]

Therefore, if the BMS procedure chooses the model with the higher predictive variance, \( V(\text{BMMP}) < V(\text{BMSP}) \). Under the assumptions of this chapter, a BMSP distribution is Student and, therefore, unimodal and symmetric. Accordingly a 95 percent credible interval, say, formed using the BMSP distribution will be wider than a 95 percent credible interval formed using the BMMP distribution. It follows that the probability of \( y_F \) being covered by a BMMP (i.e., BMC) credible interval of the same size as a 95 percent BMS credible interval is greater than .95. Thus, it may be said that when the conditions of Lemma 6 are fulfilled, a BMS credible interval understates its probability of covering \( y_F \).

In this chapter, it has been shown that when model specification uncertainty is present, the appropriate distribution with which to
characterize a data-generating process is the BMMP of the BMC procedure. Failure to use the BMMP when model specification uncertainty exists results in two interesting and seemingly contradictory effects. First, in using a single model, however selected, information provided by the remaining models which have positive posterior probability is ignored. Second, in ignoring available information about model specification uncertainty, the forecaster behaves in many cases as if he is facing a lesser degree of uncertainty than is actually the case. Thus, the forecaster simultaneously discards relevant information and behaves as if he possesses more information than is actually possessed.¹ This phenomenon was noted in both point and interval forecasting situations.

In the next chapter, the BMC procedure is applied to single-period economic control problems.

CHAPTER IV
MODEL SPECIFICATION UNCERTAINTY IN SINGLE-PERIOD ECONOMIC CONTROL PROBLEMS

In Chapter III the consequences of forecasting with and without considering model specification uncertainty were examined. Given the existence of model specification uncertainty, it was concluded that the BMC procedure was an appropriate procedure to utilize in predicting future values of a random process. In this chapter, the BMC procedure is applied to single-period economic control problems. In particular, the BMC procedure will be used to find both certainty-equivalent and optimal analytic solutions to single-period control problems. In both cases, control solutions will be derived which take into consideration costs that may be incurred by a controller as a result of his employing a particular instrument (controllable variable) to help control a random process.

By using the BMC procedure to solve economic control problems, control solutions need not be artificially conditioned on the assumption that a particular econometric model is in fact an accurate characterization of the process whose control is desired. Instead, a controller's model specification uncertainty is reflected in his control solutions, i.e., in his decisions concerning the levels or rates at which to set his controllable variables. By explicitly recognizing model specification uncertainty and including it through the BMC
procedure as part of the economic control problem, the controller is appropriately specifying the risk that control entails.

In the following section, the economic control problem is defined, references to previous work in this area are cited, and the integration of the BMC procedure and the economic control problem is discussed.

IV.1 The Economic Control Problem

The problem of effecting the outcome of some economic data-generating process such as the GNP, rate of inflation, or unemployment rate, is referred to as an economic control problem. More specifically, given an econometric model of the data-generating process of interest, a single-period economic control problem involves determining settings for the model's instruments -- controllable variables -- in one time period such that in the next time period the model's dependent variable -- a desideratum or policy objective -- is close to a specified target value or within a specified target interval. Controlling values for the model's instruments are determined by optimizing an objective or criterion function that is typically a function of the difference between the target and realized values of the dependent variable.

An economic control problem may be expressed as:

\[
\min_{X} \int_{-\infty}^{\infty} L(y - y^*) f(y | X) dy
\]

(4.1)

\(^1\)When a control of a dependent variable is desired over more than one time period, the problem is referred to as a multiperiod control problem. For a discussion of multiperiod control, see Zellner, 336-354.
where \( y \) is the dependent variable whose control is desired, \( y^* \) is the value or target the controller would like \( y \) to attain next period, and \( X \) is the model's vector-valued instrument. \( L(y - y^*) \) is a loss function that describes the losses incurred by the controller as a result of \( y \) not equalling the target, \( y^* \), or not falling in the target interval.\(^1\) The loss values may be viewed as opportunity losses or "social costs". The function \( f(y|X) \) is the predictive distribution of future values of \( y \) as determined by the econometric model used to characterize \( y \). Thus, in this case control of \( y \) is effected by setting \( X \) in the current time period so as to minimize next period's expected loss.

If sample information about the process is available, the control problem is still solved by minimizing expected loss, but the expectation is taken with respect to a predictive distribution that reflects the sample information. Letting \( y \) and \( X \) now refer to observed data points and \( y_F \) and \( X_F \) refer to the control-period values of the target and control variable, respectively, the problem becomes

\[
\min_{X_F} \int_{-\infty}^{\infty} L(y_F - y^*_F) f(y_F|y,X,X_F) dy_F. \tag{4.2}
\]

In the single-period control problem described above, it is assumed that the controller knows the correct econometric model or random process he wishes to control. Accordingly, in solving his control problem, the controller has only to contend with parameter and residual uncertainty, not model specification uncertainty. Much work has been done on such problems by, for example, Fisher, Brainard, Leland, Basu, and Zellner. The more complicated multiperiod control problem, in which it is assumed that the controller knows the correct econometric model of the process he desires to control, has also received attention. See, for example, Aoki, Prescott, Zellner, Taylor, and Chow. The approaches to single and multiperiod control of the

1For a more complete discussion of control problems, see Zellner, 319-359.


above-mentioned authors are theoretically appropriate only if the controller can assert with probability one that the model he has chosen to represent the process whose control is desired is in fact the correct representation of the process. If the controller can make such a statement, then in solving his control problem he only has to contend with the model's parameter and residual uncertainty. If, however, he specifies the chosen model's appropriateness with a model probability less than one, he is acknowledging the existence of model specification uncertainty. Theoretically, if model specification uncertainty exists, it should be dealt with in control problems. It should not be ignored or assumed away via some model selection procedure such as Bayesian Model Selection.\(^1\) Control procedures that fail to consider model specification when it exists are not optimal procedures. Such procedures, in the sense of Chapter III, misspecify the uncertainty involved in controlling \(y\), and therefore, the risk faced by the controller in using them to set the rate or level of his instruments.

Model specification uncertainty has not been explicitly considered in the control literature. Since it may have an impact upon optimal control solutions, it merits consideration. That the consideration of model specification uncertainty in control contexts is important and warrants

---

\(^1\)For a discussion of several model selection procedures that are frequently used to establish econometric models of processes whose control is desired, see Gaver and Geisel, "Discriminating Among Alternative Models: Bayesian and Non-Bayesian Methods," pp. 49-77.
a great deal of attention has been expressed by Pierce:

Another area of uncertainty has to do with our models. I want to stress this because users of control theory often tend to take models as given and work out solutions without seriously questioning the reasonableness of the models. This tendency is not very harmful when one is working on technique. However, there is a real danger of giving more credence to model results than they deserve, especially if a particular policy trajectory is highly influenced by the choice of a model.¹

He goes on to say:

The problem lies not with uncertainty concerning the true value of the model parameters, but also with the structure of the models themselves.²

By utilizing the Bayesian Model Comparison procedure to develop a Bayesian Mixed Model Predictive distribution for the process whose single-period control is desired, a controller can determine settings for his instruments in light of residual, parameter, and model specification uncertainty.³ When single-period control is desired, the solution to the following minimization problem provides optimal settings for the controller’s instruments, D₉:

\[
\min_{D_9} \int \int L(y_9, y_9^*) f(y_9 | y, D, D_9) dy_9.
\]

²Ibid.
³That a BMMP in fact reflects model specification uncertainty was discussed in Chapter III.
Recall that $D = (X, Z)'$ and $D_F = (X_F, Z_F)'$. The function $f(y_F|y, D, D_F)$ is the controller's BMMP for the data-generating process. All other terms in (4.3) are as previously defined. The only difference between (4.3) and (4.1) or (4.2) is the use of a Bayesian Mixed Model Predictive in (4.3) rather than a predictive distribution determined from a single model.\footnote{In what follows, control problems that deal with a single model will be expressed as in (4.1) or (4.2).} Since all relevant major forms of uncertainty, residual, parameter, and model specification uncertainty are reflected in (4.3) and, therefore, influence its solution, it is said that (4.3) provides optimal settings for $D_F$.

In the next section of this chapter, assumptions are presented under which various single-period control solutions are obtained using the BMC procedure in the remainder of the chapter.

IV.2 Model Space and Assumptions

In the remainder of this chapter, solutions will be derived for single-period control problems based on the following assumptions:

1. The decision maker (controller) believes that one or the other of the following two models is an accurate representation of a data-generating process to be controlled, but he is unsure which one is correct:

   \begin{align*}
   M_1: \quad & y = \beta_1 X + \epsilon; \\
   M_2: \quad & y = \beta_2 Z + \delta.
   \end{align*}

   (4.4)

   $y$ is the target variable, and $X$ and $Z$ are two different nonrandom explanatory variables, instruments...
over which the controller has complete control. $\beta_1$ and $\beta_2$ are unknown parameters. $\varepsilon$ and $\delta$ are the usual normally distributed error terms, each with zero mean and unknown variance, $\sigma^2_\varepsilon$ and $\sigma^2_\delta$, respectively. It is also assumed that $\text{cov}(\beta_1, \varepsilon) = \text{cov}(\beta_2, \delta) = \text{cov}(\varepsilon, \delta) = 0$. Thus, $M_1$ and $M_2$ are normal univariate regression models which, to keep the number of each model's unknown parameters at two, have been forced through the origin.

2. The data-generating process over which control is desired is stationary.

3. $X$ and $Z$ are uncorrelated, and only the controllable variable in the true model affects $y$. Thus, if $M_1$ were the true model, $\beta_2$ would be zero. If neither $M_1$ nor $M_2$ were the true model, it may be that $\beta_1 = \beta_2 = 0$.

4. The controller's loss function is a quadratic loss function of the form

$$L(y_F, y_{F*}) = K(y_F - y_{F*})^2$$

where $K$ is a constant. In what follows, $K$ is set equal to one without loss of generality.

Aside from the change in emphasis from forecasting to control, and the assumption that $X$ and $Z$ are controllable variables, the above assumptions are similar to those under which the Bayesian Model Comparison and Bayesian Model Selection procedures were compared in Chapter III (see Section III.2.4).

In the next section, certainty-equivalent solutions to single
period control problems will be derived under the above assumptions with the use of the BMC procedure.

IV.3 Single-Period Certainty-Equivalent Control

If in attempting to control y's value next period, the controller behaves this period as if \( E(y) \) is the value of \( y \) that will occur with certainty next period, then \( E(y) \) is said to be a "certainty equivalent" for \( y \).\(^1\) When the process which generates \( y \) is known or assumed to be known, the single-period control problem under parameter and residual uncertainty is reduced to a deterministic problem. If the process which generates \( y \) is not known, but is believed to be best represented by one of \( N \) alternative models, the single-period control problem under model specification, parameter, and residual uncertainty reduces to one of control under model specification uncertainty alone. In this section, single-period control solutions are derived for the controller who admits to model specification uncertainty and behaves as if \( E(y) \) will occur with certainty next period.

The use of the certainty-equivalent \( E(y) \) for \( y \) reduces models 1 and 2 of Section IV.2 to the following:

\[
M_1: \quad E_y|x(y) = E_{\beta_1^*} (\beta_1)X + E_\epsilon (\epsilon) = b_1^*X;
\]

\[
M_2: \quad E_y|x(y) = E_{\beta_2^*} (\beta_2)Z + E_\delta (\delta) = b_2^*Z.
\]

In (4.5) the parameter and residual uncertainty of M\(_1\) and M\(_2\) are treated as if they do not exist. Thus, if neither M\(_1\) nor M\(_2\) is known or assumed to be the true model, it is only necessary to deal with model specification uncertainty.

From (4.2), assuming that M\(_1\) is the true model, the single-period control problem is solved by determining:

\[
\min_{X_F} \int_{-\infty}^{\infty} L(y_F,y^*,y\mid y,X,F)dy_F = \min_{X_F} \mathbb{E}_{y,F\mid y,X,F} L(y_F,y^*). \quad (4.6)
\]

The solution to (4.6) yields the controller's minimum expected loss under M\(_1\)'s predictive distribution of y\(_F\). If the loss function is quadratic, as is assumed for the remainder of this chapter, (4.6) becomes:

\[
\min_{X_F} \mathbb{E}_{y,F\mid y,X,F} (y_F - y^*)^2. \quad (4.7)
\]

The use by the controller of \(\mathbb{E}_{y,F\mid y,X,F} (y_F)\) as a certainty equivalent for y\(_F\) reduces (4.7) to the following:

\[
\min_{X_F} [\mathbb{E}_{y,F\mid y,X,F} (y_F) - y^*]^2. \quad (4.8)
\]

Note that (4.8) contains no random terms. Thus, (4.8) is minimized by the value of X\(_F\) that sets \(\mathbb{E}_{y,F\mid y,X,F} (y_F)\) equal to y\(_F^*\). From (4.5) it can be seen that \(\mathbb{E}_{y,F\mid y,X,F} (y_F) = \mathbb{E}_{\beta_1\mid y} (\beta_1)X_F + \mathbb{E}_{\epsilon\mid y} (\epsilon) = b_1X_F \). Thus the appropriate setting for X\(_F\) is one such that \(b_1X_F = y^*\). Accordingly, X\(_F\) should be set equal to \(\frac{y^*}{b_1}\). This is the single-period certainty-equivalent solution when it is assumed that model 1 generates
Similarly, when model 2 is assumed to generate \( y_F \), the single-period certainty-equivalent solution is to set \( Z \) equal to \( \frac{y^*}{b^*_2} \).\(^2\)

Single-period certainty-equivalent control solutions in which a particular model is assumed to generate \( y_F \) are derived assuming the mean of \( y_F \)'s predictive distribution is the value of \( y_F \) that will occur with certainty next period. This is equivalent to assuming that \( b_1 = b^-_1 \) and \( \varepsilon = 0 \). \( y_F \)'s predictive variance is ignored in the certainty-equivalent solution. Consequently, such solutions are not optimal but are only approximations to optimal solutions, as explained by Zellner.\(^3\) In general, since certainty-equivalent control problems ignore \( y_F \)'s predictive variance and, therefore, parameter and residual uncertainty, their solutions are much easier and less costly to obtain than are optimal control solutions. Consequently, certainty-equivalent control may at times provide the controller with an attractive alternative to full-scale optimal control.

**IV.3.1 Certainty-Equivalent Control Using the BMMP Distribution**

By using the Bayesian Model Comparison procedure's Bayesian Mixed Model Predictive as \( y_F \)'s predictive distribution, single-period certainty-equivalent control solutions can be derived which reflect the controller's model specification uncertainty concerning \( M_1 \) and \( M_2 \) of

\(^1\)This solution can also be found in Zellner, pp. 320-322.

\(^2\)Notice that these certainty-equivalent solutions make the control target, \( y^* \), the mean of \( y_F \)'s predictive distribution.

\(^3\)Zellner, pp. 322-324.
the previous section. This approach to certainty-equivalent control also does not explicitly consider parameter and residual uncertainty and is, therefore, also suboptimal. But by enabling the controller to solve his control problems in light of any model specification uncertainty, this approach may improve the effectiveness of certainty-equivalent control solutions. As will be seen below, single-period BMC certainty-equivalent control, as it will be called, requires little more computational effort than the certainty-equivalent control solutions derived above in which specification uncertainty was not treated.

The BMC certainty-equivalent control solution can be obtained from the full-scale BMC control problem of (4.3). (4.3) is repeated here and the BMC certainty-equivalent control solution is derived below:

$$\min_{D_F} \int_{-\infty}^{\infty} L(y_F|y,F,D,F)dy_F. \quad (4.9)$$

Recall from (3.1) that for the two-model case the BMMP distribution would be expressed as

$$f(y_F|y,D,D_F) = P^n(M_1|y,X)f(y_F|M_1,y,X,X_F)$$
$$+ P^n(M_2|y,Z)f(y_F|M_2,y,Z,Z_F). \quad (4.10)$$

In this case, $D$ and $D_F$ are vectors of control variables: $D = (X,Z)'$ and $D_F = (X_F,Z_F)'$. Accordingly, (4.9) may be written

$$\min_{D_F} \left[ P^n(M_1|y,X) \int_{-\infty}^{\infty} L(y_F,y_F^*)f(y_F|M_1,y,X,X_F)dy_F \right.$$
$$+ P^n(M_2|y,Z) \int_{-\infty}^{\infty} L(y_F,y_F^*)f(y_F|M_2,y,Z,Z_F)dy_F]. \quad (4.11)$$
Under the assumption that the loss function is quadratic, (4.11) may be rewritten

\[
\min_{D_F} \left[ P'(M_1|y,X)E_{y_F|M_1,y,X}(y_F - y_F^*)^2 + P'(M_2|y,Z)E_{y_F|M_2,y,Z}(y_F - y_F^*)^2 \right].
\]  

(4.12)

The use by the controller of \( E_{y_F|M_1,y,X}(y_F) = C_1 \) and \( E_{y_F|M_2,y,Z}(y_F) = C_2 \) as certainty equivalents for \( y_F \) in \( M_1 \) and \( M_2 \), respectively, means that \( y_F \) is no longer treated as being random. Consequently, (4.12) reduces to

\[
\min_{D_F} \left[ P'(M_1|y,X)(C_1 - y_F^*)^2 + P'(M_2|y,Z)(C_2 - y_F^*)^2 \right].
\]  

(4.13)

Because the right-hand term inside the brackets of (4.13) is not a function of \( X_F \), and the left-hand term is not a function of \( Z_F \), the vector optimizing (4.13), \( D_F^* \), may be found by minimizing each of the terms within the brackets separately. Thus, in order to find \( D_F^* \), the single-period BMC certainty-equivalent control solution, the following two problems must be solved:

\[
\min_{X_F} P'(M_1|y,X)(C_1 - y_F^*)^2 \]  

(4.14)

and

\[
\min_{Z_F} P'(M_2|y,Z)(C_2 - y_F^*)^2. \]  

(4.15)

Noting that \( P'(M_1|y,X) \) is not a function of \( X_F \), and that \( P'(M_2|y,Z) \) is not a function of \( Z_F \), (4.14) and (4.15) reduce to the following:
\[
\begin{align*}
\min_{X_F} (C_1 - y_F^*)^2 ; \\
\min_{Z_F} (C_2 - y_F^*)^2 .
\end{align*}
\]  

(4.16)  

(4.17)

Notice that (4.16) is the same as (4.8). Thus, for example, in order to solve (4.16) \(X_F\) should be set equal to \(\frac{y^*}{b_2} \) . Thus, \(D_F^* = (\frac{y^*}{b_1}, \frac{y^*}{b_2}) \). In words, the BMC certainty-equivalent control solution is to set \(X_F\) as if \(M_1\) were in fact the model generating \(y_F\) and to set \(Z_F\) as if \(M_2\) were the true model.

The rationale behind the BMC control solution is that since the controller is unsure of which of the two control instruments, \(X_F\) or \(Z_F\), affects \(y_F\), and since he believes that only one of them actually affects \(y\), he should use them both fully in attempting to attain \(y_F^*\). Due to the restrictive assumptions under which it was derived, this solution is somewhat unrealistic. A more realistic solution would account for the possibility that (1) costs might be incurred for the use of an instrument,\(^1\) especially for the use of an inappropriate instrument; (2) both instruments might affect \(y_F\); (3) the instruments interact in some manner; and/or (4) the process generating \(y\) may be nonstationary. The first of these more realistic cases will be discussed with respect to optimal BMC control in Section IV.4. At

---

\(^1\) For a solution to how to account for the cost of changing the setting of an instrument in the optimal single-period control problem in which a particular model is assumed to generate \(y_F\), see Zellner, pp. 324-325.
that time, the appropriate optimal BMC and BMC certainty-equivalent control solutions for various cases in which instrument use costs are involved will be derived. Case (2) above is discussed in Section IV.4.5, and an approach to case (4) is discussed in Chapter V.

IV.3.2 Risk Specification in Certainty-Equivalent Control

Even though a controller may behave as though the expected value of \( y_F \) is certain to occur next period, he should not ignore the risk involved in his choosing to do so. This risk may be represented by \( y_F \)'s predictive variance. The larger \( y_F \)'s predictive variance, the more likely that \( L(y_F, y_F^*) = (y_F - y_F^*)^2 \) will be large. Thus, the controller can use \( y_F \)'s predictive variance as a measure of the risk involved in his attempt to attain \( y_F^* \). If the risk appears too great, the controller may choose a different control method, perhaps optimal single-period control (discussed in Section IV.4), since it considers the size of \( y_F \)'s predictive variance in determining settings for the controller's policy instruments.

If the controller knows that a particular model, say \( M_1 \), will generate \( y_F \), recalling (3.24), \( y_F \)'s predictive variance and a measure of the risk being taken by the controller is

\[
\sigma^2_1 = \frac{(n - 1)S^2}{(n - 3)} \left[ \frac{X_F^2}{\sum_{i=1}^{N} x_i^2} + 1 \right]
\]

(4.18)

where \( x_i \) is the \( i \)th sample observation of \( X \). Notice that \( \sigma^2_1 \) is a function of the controller's instrument setting, \( X_F \). Consequently, since the control method chosen affects \( X_F \), it also influences the size of \( \sigma^2_1 \).

In the case of certainty-equivalent
control, $X_F = \frac{y_F^*}{b_1}$, and the predictive variance is

$$\sigma^2 = \frac{(n - 1)s^2}{(n - 3)} \left[ \left( \frac{y_F^*}{b_1} \right)^2 \left( \frac{1}{\sum_{i=1}^{N} x_i} \right) + 1 \right]. \quad (4.19)$$

If the controller acknowledges model specification uncertainty and chooses to control via BMC certainty-equivalent control, then, recalling (3.12), $y_F$'s predictive variance is

$$\sigma^2 = p_1(M_1 | y, X)\sigma_1^2 + p_2(M_2 | y, Z)\sigma_2^2 + p_1(M_1 | y, X)(\mu_1 - \mu)^2$$

$$+ p_2(M_2 | y, Z)(\mu_2 - \mu)^2. \quad (4.20)$$

$\mu_i$ is the mean of $y_F$'s predictive distribution as characterized by model $i$, and $\mu$ is the mean of the BMMP distribution for $y_F$.

Equation (4.18) provides an appropriate risk measure only if the controller is certain that a particular model will generate $y_F$. If he utilizes the BMS procedure to choose a model for $y_F$, he is acknowledging that he is uncertain of the form of the process generating $y_F$. Consequently, (4.18) is not an appropriate measure of his risk. If the BMS procedure chooses, say, $M_1$, $\sigma_1^2$ understates the risk involved in his attempt to attain $y_F^*$. The following lemma is needed to prove this statement.

**Lemma 7.** Let $n > 3$. When BMS and the max-$R^2$ rule provide equivalent methods for choosing between $M_1$ and $M_2$ (see Section III.2.2), and single-period certainty-equivalent control is applied to the model
chosen by BMS, say, model 1, then $V(BMSP) = \sigma_1^2 \leq V(BMMP) = \sigma^2$.

**PROOF:** Suppose the BMS procedure chooses $M_1$, and $M_1$ is used to control $y_F$. The certainty-equivalent control solution is $X_F^* = \frac{y_F^*}{b_1^{**}}$. Accordingly, $\sigma_1^2$ is as shown in (4.19). Raiffa and Schlaifer show that

$$b_1^{**} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2},$$

where $x_i$ and $y_i$ may be the $i$th sample observation of $X$ and $y$, or reflect prior information about $\beta_i$ in a form equivalent to sample observations.\(^1\) Substituting for $b_1^{**}$ in (4.19) yields

$$\sigma_1^2 = \frac{(n - 1)S_1^2}{(n - 3)} \left[ \frac{\left( \sum_{i=1}^{n} x_i^2 \right) y_F^{*2}}{\left( \sum_{i=1}^{n} x_i y_i \right)^2} + 1 \right].$$  

(4.21)

$M_1$'s estimated residual variance, $S_1^2$, is, by definition,

$$S_1^2 = \frac{\sum_{i=1}^{n} (y_i - x_i b_1^{**})^2}{n - 1} = \frac{\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} x_i y_i)^2}{n - 1}$$

(4.22)

Thus, a necessary and sufficient condition for $S_1^2 = S_2^2$ is

---

\[
\frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\sum_{i=1}^{n} x_i^2} = \frac{\left(\sum_{i=1}^{N} z_i y_i\right)^2}{\sum_{i=1}^{n} z_i^2},
\]

(4.23)

where \(z_i\) is the \(i^{th}\) sample observation of \(Z\). Accordingly, if \(S_1^2 = S_2^2\), then \(\frac{(\sum xy)^2}{\sum x^2} = \frac{(\sum zy)^2}{\sum z^2}\) and, noting \(\sigma_1^2\)'s definition in (4.21), \(\sigma_1^2 = \sigma_2^2\).

If it can be shown that \(\frac{\partial \sigma_1^2}{\partial S_1^2} > 0\), then it can also be said that \(\sigma_1^2 > \sigma_2^2\) when \(S_1^2 > S_2^2\). That \(\frac{\partial \sigma_1^2}{\partial S_1^2} > 0\) is demonstrated in the next paragraph.

Noting that

\[
\frac{(\sum xy)^2}{\sum x^2} = \left[\sum y^2 - (n - 1)S_1^2\right] > 0,
\]

(4.24)

(4.21) can be rewritten

\[
\sigma_1^2 = \frac{(n - 1)S_1^2}{(n - 3)} \left[\frac{y_F^*_2}{\left[\sum y^2 - (n - 1)S_1^2\right]} + 1\right],
\]

(4.25)

Taking the partial derivative of (4.25) with respect to \(S_1^2\) yields

\[
\frac{\partial \sigma_1^2}{\partial S_1^2} = \frac{(n - 1)y_F^*_2}{(n - 3)[\sum y^2 - (n - 1)S_1^2]} + \frac{(n - 1)^2S_1^2y_F^*_2}{(n - 3)[\sum y^2 - (n - 1)S_1^2]^2} + \frac{(n - 1)}{(n - 3)}. \]

(4.26)
By (4.24), the denominator and, therefore, the entire first term on the rhs of (4.26) is positive when \( n > 3 \). The second and third terms on the rhs of (4.26) are also obviously both positive if \( n > 3 \). Consequently, \( \frac{\partial \sigma_1^2}{\partial S_1} > 0 \).

Under the conditions of this lemma, the BMS procedure selects the model with the lower \( S^2 \) (higher \( R^2 \)). Thus, since \( \sigma_1^2 = \sigma_2^2 \) when \( S_1^2 = S_2^2 \), and \( \sigma_1^2 > \sigma_2^2 \) when \( S_1^2 > S_2^2 \), the BMS procedure also selects the model with the lower predictive variance. Recall Theorem 2 of Chapter III in which it was shown that if the BMS procedure chooses the model with the lower predictive variance, then \( V(\text{BMMP}) \geq V(\text{BMSP}) \). Accordingly, by Theorem 2, the desired result is obtained.¹

If model specification uncertainty exists, the BMMP distribution of the BMC procedure is the appropriate distribution with which to characterize \( y_F \); any other procedure for determining the predictive distribution will fail to include relevant information. Accordingly, when model specification uncertainty exists, the appropriate measure of the controller's risk is \( \sigma^2 \), not \( \sigma_1^2 \). As shown in Lemma 7, \( \sigma_1^2 \) is less than \( \sigma^2 \) and therefore understates the controller's risk.

In this section, certainty-equivalent solutions have been considered, but, certainty-equivalent solutions are not fully optimal in

¹Recall that Theorem 2 showed that \( V(\text{BMMP}) \geq V(\text{BMSP}) \). It was noted, however, that \( V(\text{BMMP}) = V(\text{BMSP}) \) only when one or the other of \( P'(M_1) \) and \( P'(M_2) \) equalled one. But neither of these cases involve model specification uncertainty and, therefore, are not of interest in this dissertation. Therefore, under the conditions of Lemma 7, \( V(\text{BMMP}) > V(\text{BMSP}) \) in cases of interest.
general. In the next section, optimal single-period BMC control solutions will be derived.

**IV.4 Optimal Single-Period Control**

A control procedure will be referred to as providing an optimal solution to a control problem if it explicitly recognizes all existing major forms of uncertainty and utilizes the information provided by them in its solution to the control problem. Thus, for example, for a control procedure and its solution to be called optimal when the controller knows the form of the model generating y, but does not know the parameter values of the model, the procedure need only consider residual and parameter uncertainty. However, should specification uncertainty concerning the model be present as well, the procedure would have to consider residual uncertainty, parameter uncertainty, and model specification uncertainty. As discussed in Section IV.3, the certainty-equivalent approach to economic control problems treats residual and parameter uncertainty suboptimally and, unless BMC certainty-equivalent control procedures are used, also treats model specification uncertainty suboptimally. In this section, optimal control solutions, i.e., solutions that appropriately treat residual, parameter and model specification uncertainty, will be derived using the BMC procedure. These solutions will be referred to as "optimal BMC control solutions."

Before proceeding with the derivation of optimal BMC control solutions, mention should be made of the optimal control solution for the case in which the controller knows the form of the model generating
y, but not its parameters. Assuming $M_1$ is the true model, and employing a quadratic loss function, Zellner shows that the optimal solution to (4.2) is

$$X_F = \frac{y_F^*}{(n - 1)S_1^2 + \sum_{i=1}^{n} x_i y_i} \cdot \frac{n}{(n - 3) \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} x_i^2}$$

Equation (4.27) may be rewritten so that its relationship to the certainty-equivalent solution to this problem may be examined:

$$X_F = \frac{y_F^*}{b_1^n} \left[ \frac{1}{(n - 1)S_1^2 + \sum_{i=1}^{n} x_i y_i} \right]$$

Recall from Section IV.3 that the certainty-equivalent solution is

$$X_F = \frac{y_F^*}{b_1^n}.$$

Thus, as Zellner has noted, the certainty-equivalent solution is just the first term on the rhs of (4.28). Zellner has shown that as the precision of the estimation of $\beta_1$ improves (i.e., as the posterior variance of $\beta_1$ decreases),

\[\text{Zellner, pp. 320-322.}\]

\[\text{Ibid.}\]
\begin{align*}
\frac{(n - 1)s_1^2}{(n - 3)b_1 n^2 \sum_{i=1}^{n} x_i^2} & \to 0.
\end{align*}

and, accordingly, the second term on the rhs of (4.28) approaches 1.\(^1\)

Thus, if \(b_1"\) is a very precise estimate of \(\beta_1\), (4.29) is approximately (4.28). Zellner has also demonstrated that the use of the certainty-equivalent solution (4.29) leads to higher expected losses than the use of (4.28).\(^2\)

**IV.4.1 Optimal BMC Control**

The optimal BMC control solution is obtained by minimizing expected loss over \(X_F\) and \(Z_F\) using the BMMP distribution of \(y_F\). This problem was stated in (4.3) and is repeated here for convenience:

\[
\min_{D_F} \lim_{\lambda \to \infty} \int L(y_F, y_F^*) f(y_F|y, D, D_F) dy_F. \tag{4.29}
\]

(4.29) is solved below for the two-model case (see the assumptions of Section IV.2) under study in this dissertation.

Substituting (4.10) for \(f(y_F|y, D, D_F)\) in (4.29), the minimization problem becomes

\[
\min_{D_F} \left[ P''(M_1|y, X) \lim_{\lambda \to \infty} \int L(y_F, y_F^*) f(y_F|M_1, y, X, X_F) dy_F \right.
\]

\[
\left. + P''(M_2|y, Z) \lim_{\lambda \to \infty} \int L(y_F, y_F^*) f(y_F|M_2, y, Z, Z_F) dy_F \right]. \tag{4.30}
\]

\(^1\)Ibid.

\(^2\)Zellner, pp. 322-324.
Recalling that $\beta_1$ and $\varepsilon$, and $\beta_2$ and $\delta$ are assumed to be independent, the following transformation of variables can be made in the first and second terms of (4.30), respectively, so that (4.30) may be written in a more convenient form:

$$y_F = \beta_1 x_F + \varepsilon$$

$$y_F = \beta_2 z_F + \delta.$$

Thus, utilizing a quadratic loss function for $L(y_F, y_F^*)$, (4.30) may be written

$$\min_{D_F} \left( p'(M_1|y,x) \epsilon_{\beta_1, \varepsilon, \sigma_\varepsilon^2} y, x, x_F [y_F^* - (\beta_1 x_F + \varepsilon)]^2 ight)$$

$$+ p'(M_2|y,z) \epsilon_{\beta_2, \delta, \sigma_\delta^2} y, z, z_F [y_F^* - (\beta_2 z_F + \delta)]^2 \right)$$

(4.31)

It can be seen that, as in the case of the BMC certainty-equivalent control problem, (4.31) separates into two minimization problems,

$$\min_{X_F} E_{\beta_1, \varepsilon, \sigma_\varepsilon^2} [y_F^* - (\beta_1 x_F + \varepsilon)]^2$$

(4.32)

and

$$\min_{Z_F} E_{\beta_2, \delta, \sigma_\delta^2} [y_F^* - (\beta_2 z_F + \delta)]^2.$$ 

(4.33)

Recall that (4.2) is the mathematical statement of the control problem when it is known that $M_1$ will generate $y_F$. After the transformation of variables noted above, (4.2) and (4.32) are the same. Thus, the solution to (4.32) will be the same as that derived by Zellner for (4.2). Except that it is $M_2$ that is known to be generating $y_F$ in
(4.33), the solution to (4.33) will also be of the same form as Zellner's solution noted in (4.27).

Even though the solutions to (4.32) and (4.33) are already known, (4.32) will be solved below for later use as a reference in solving optimal BMC control problems when instrument use costs are considered. Squaring the term in brackets in (4.32) yields

$$\min_{x_F} E_{\beta_1, \epsilon, \sigma_\epsilon^2}[y, x, x_F \left[ y_F^2 - 2y_F \beta_1^1 x_F - 2y_F \epsilon + \beta_1^2 x_F^2 + 2\beta_1^1 x_F \epsilon + \epsilon^2 \right]],$$

(4.34)

which can be expressed as

$$\min_{x_F} \left[ y_F^2 - 2y_F \beta_1^1 x_F + 2x_F \beta_1^2 + 2x_F \epsilon \beta_1^1 + E_{\beta_1, \epsilon} \left[ y, x, x_F \right] \right].$$

(4.35)

Recall that $E(\beta_1) = b_1^2$, $E(\epsilon) = 0$, $E(\epsilon^2) = [E(\epsilon)]^2 = E(\epsilon^2)$, $V(\beta_1) = E(\beta_1^2) - [E(\beta_1)]^2$, and $\text{cov}(\beta_1, \epsilon) = E(\beta_1 \epsilon) - E(\beta_1)E(\epsilon) = 0$. Thus, (4.35) may be written as

$$\min_{x_F} \left[ y_F^2 - 2y_F x_F b_1^1 + x_F^2 \left[ V(\beta_1) + b_1^2 \epsilon^2 \right] + 2x_F \text{cov}(\beta_1, \epsilon) + E(\beta_1)E(\epsilon) \right].$$

(4.36)

Further simplification reduces (4.36) to

$$\min_{x_F} \left[ y_F^2 - 2y_F x_F b_1^1 + x_F^2 V(\beta_1) + x_F^2 b_1^2 \epsilon^2 + \sigma_\epsilon^2 \right].$$

(4.37)

(4.37) can be solved by taking the partial derivative of the term in
brackets with respect to $X_F$, setting the resultant derivative equal to zero, and solving for $X_F$. Before proceeding to solve (4.37), however, recall that

$$b_1'' = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

and note that the variance of $\beta_1$'s marginal distribution is

$$V(\beta_1) = \frac{(n - 1)S_1^2}{(n - 3)\sum_{i=1}^{n} x_i^2}.$$  \hspace{1cm} (4.38)

Thus, the partial derivatives of $b_1''$ and $V(\beta_1)$ with respect to $X_F$ are both zero. Calling the bracketed term in (4.37) $A$,

$$\frac{\partial A}{\partial X_F} = -2y_F^* b_1'' + 2V(\beta_1)X_F + 2b_1''x_F^2.$$  \hspace{1cm} (4.39)

Setting (4.39) equal to zero, solving for $X_F$, and substituting (4.38) for $V(\beta_1)$ yields

$$X_F = \frac{b_1''y_F^*}{(n - 1)S_1^2 + b_1''x_F^2}.$$  \hspace{1cm} (4.40)

Recalling the definition of $b_1''$ above, $X_F$ may be expressed as

---

1Raiffa and Schlaifer, pp. 344-345.
Similarly, the solution to (4.33) is

\[
X_F = \frac{b_1 y_F^*}{(n - 1)S_1^2 + \sum_{i=1}^{n} x_i^2 y_i} .
\]  

Comparing (4.41) and (4.42) with (4.27), it can be seen that these solutions are in fact the same as Zellner's. Thus, the optimal BMC control solution is to set \( X_F \) as if it were known that \( M_1 \) had generated \( y \) and would be generating \( y_F \), and to set \( Z_F \) as if it were known that \( M_2 \) had generated \( y \) and would be generating \( y_F \).

As with the BMC certainty-equivalent control solution, the rationale behind the optimal BMC control solution is that since the controller is unsure of which of the two control instruments, \( X_F \) or \( Z_F \), affects \( y_F \), he should use them both fully in attempting to attain \( y_F^* \). Due to the restrictive assumptions under which it was derived, this solution suffers the same lack of realism as did the BMC certainty-equivalent solution. A more realistic solution would account for the possibility that (1) costs might be incurred for use of an
instrument, especially for the use of an inappropriate instrument; (2) both instruments might affect $y_F$; (3) the instruments interact in some manner; and/or (4) the process generating $y_F$ may be nonstationary. Situation (1) is treated in the next section, situation (2) is discussed in Section IV.4.5, and an approach to situation (4) is presented in Chapter V.

IV.4.2 Optimal BMC Control When Instrument Use Costs Are Considered

The optimal BMC control solution derived in Section IV.1 is appropriate if the assumptions of Section IV.2 hold and the use of the instruments, $X_F$ and $Z_F$, involves no cost to the controller. Optimal BMC control solutions will now be derived for the following cases in which instrument use costs are involved:

Case 1: The instrument-use cost varies with the level of instrument usage. Let $X_F$ and $Z_F$ cost the controller $C_1$ and $C_2$ per squared unit of $X_F$ and $Z_F$ used, respectively.

Case 2: A fixed as well as a variable instrument-use cost is incurred by the controller. Let the controller incur a fixed instrument usage cost of $e_1$ for use of $X_F$ and $e_2$ for use of $Z_F$. Let each unit of $X_F$ and $Z_F$ used by the controller cost him $d_1$ and $d_2$, respectively. $e_1$, $e_2$, $d_1$ and $d_2$ are constants known to the controller.

Case 3: There is no cost incurred by the controller as a result of his using the appropriate instrument (the instrument that affects $y_F$), but there is a cost for use of the inappropriate instrument.

Thus, if $X_F$ is used needlessly, let a cost of $C_1$ per squared unit of $X_F$ used be incurred. If $Z_F$ is used needlessly, let a cost of $C_2$ per
squared unit of $Z_F$ used be incurred.

It is assumed that instrument costs are expressed in the same measurement units as the losses described by the controller's loss function. Accordingly, each of the above cost considerations may be included in the solution to (4.29) by explicitly introducing each into (4.31).

For Case 1, if $X_F^o$ and $Z_F^o$ units of $X_F$ and $Z_F$ are used by the controller, he will, with probability one, incur costs of $C_1$ per squared unit of $X_F$ used and $C_2$ per squared unit of $Z_F$ used, no matter which model is in fact appropriate. Accordingly, with the inclusion of Case 1 costs, (4.31) becomes

$$\min \{P''(M_1|y,X)E_{\beta_1,\varepsilon,\sigma^2}|y,X,X_F [y_F^* - (\beta_1 X_F + \varepsilon)]^2$$

$$+ P''(M_2|y,Z)E_{\beta_2,\delta,\sigma^2}|y,Z,Z_F [y_F^* - (\beta_2 Z_F + \delta)]^2 + C_1 X_F^2 + C_2 Z_F^2\}.$$  \hspace{1cm} (4.43)

As with (4.31), and for the same reason, (4.43) separates into two minimization problems,

$$\min \{P''(M_1|y,X)E_{\beta_1,\varepsilon,\sigma^2}|y,X,X_F [y_F^* - (\beta_1 X_F + \varepsilon)]^2 + C_1 X_F^2\}.$$ \hspace{1cm} (4.44)

and

$$\min \{P''(M_2|y,Z)E_{\beta_2,\delta,\sigma^2}|y,Z,Z_F [y_F^* - (\beta_2 Z_F + \delta)]^2 + C_2 Z_F^2\}.$$ \hspace{1cm} (4.45)

(4.44) and (4.45) may be solved in the same manner as (4.32) and (4.33). Accordingly, prior to taking the partial derivative of (4.44) with respect to $X_F$, (4.44) may be written in the same form as (4.37):
\[
\min_{X_F} \{ P''(M_1 | y, X)[y_F^*]^2 - 2y_F^*X_Fb_1'' + X_F^2V(\beta_1) + X_F^2b_1''^2 + V(\epsilon)] + C_1X_F^2 \}. \quad (4.46)
\]

Call everything inside the braces of (4.46) A, and denote \( P''(M_1 | y, X) \) by \( P_1 \). Then, the partial derivative of \( A \) with respect to \( X_F \) is

\[
\frac{\partial A}{\partial X_F} = -2y_F^*b_1''p_1 + 2X_FV(\beta_1)p_1 + 2X_Fb_1''^2p_1 + 2C_1X_F. \quad (4.47)
\]

Setting (4.47) equal to zero and solving for \( X_F \), the optimal setting for \( X_F \) is obtained (second order conditions are easily shown to hold):

\[
X_F = \frac{y_F^*b_1''p_1}{P_1[V(\beta_1) + b_1''^2] + C_1} = \frac{y_F^*}{V(\beta_1)b_1'' + C_1P_1b_1''}. \quad (4.48)
\]

The optimal setting for \( Z_F \) may be obtained by solving (4.45) in a similar fashion:

\[
Z_F = \frac{y_F^*}{V(\beta_2)b_2'' + b_2'' + C_2P_2b_2''}. \quad (4.49)
\]
Thus, the optimal single-period BMC control solution for Case 1 is

\[
D_F = \left( \frac{y_F^*}{V(\beta_1) + \frac{C_1}{b_1'' + \frac{p_1 b_1''}{b_1''}}}, \frac{y_F^*}{V(\beta_2) + \frac{C_2}{b_2'' + \frac{p_2 b_2''}{b_2''}}} \right). \tag{4.50}
\]

The implications of the solutions to Cases 1, 2, and 3 will be discussed together after all three solutions have been derived.

For Case 2, a fixed cost of \( e_1 \) is incurred by the controller if he uses \( X_F \) (i.e., if he sets \( X_F \) at any value other than zero), and a fixed cost of \( e_2 \) is incurred if he uses \( Z_F \). Also, the controller will incur costs of \( d_1 \) per unit of \( X_F \) used and \( d_2 \) per unit of \( Z_F \) used no matter which model is appropriate. Thus, since the costs of Case 2, like Case 1, are incurred irrespective of which model is appropriate, \( e_1 + d_1 X_F \) and \( e_2 + d_2 Z_F \) should be added to the expected loss expression in (4.31). Accordingly, with the inclusion of Case 2 costs, (4.31) becomes
\[
\min_{D_F} \{ P''(M_1 \mid y, X) E_{\beta_1, \varepsilon, \sigma^2} [y, X, Z_F \mid y, X, X_F] [y_F^* - (\beta_1 X_F + \varepsilon)]^2 +
\]

\[
P''(M_2 \mid y, Z) E_{\beta_2, \delta, \sigma^2} [y, Z, Z_F \mid y, Z, Z_F] [y_F^* - (\beta_2 Z_F + \delta)]^2 +
\]

\[e_1 + e_2 + d_1 X_F + d_2 Z_F \}.
\] (4.51)

As with (4.31) and (4.43), (4.51) separates into two minimization problems, one whose solution is the optimal setting for \(X_F\), and one whose solution is the optimal setting for \(Z_F\). Solving the two problems in the same manner that (4.32) and (4.44) were solved, the optimal setting for \(X_F\) is found to be

\[
X_F = \begin{cases} 
\frac{y_F^* - \frac{d_1}{2P_F b_1}}{V(\beta_1) b_1} + b_1 & \text{if by using } X_F \text{ the controller's total expected loss is reduced by at least } X_F' \text{'s fixed use-cost} \\
0 & \text{otherwise},
\end{cases}
\] (4.52)

where, again, \(P''(M_1 \mid y, X)\) is represented by \(P_1\).
The optimal setting for $Z_F$ turns out to be

$$Z_F = \begin{cases} 
\frac{y_F^* - d_2}{2P_2b_2''} & \text{if by using } Z_F \text{ the controller's} \\
\frac{V(\beta_2)}{b_2''} + b_2'' & \text{expected loss is reduced by at least } Z_F 's \text{ fixed use cost} \\
0 & \text{otherwise,} 
\end{cases} \quad (4.53)$$

where $P''(M_2|y,Z)$ is represented by $P_2$. Thus, the optimal single-period BMC control solution for Case 2 is $D_F = (X_F, Z_F)'$, where $X_F$ and $Z_F$ are as given in (4.52) and (4.53), respectively.

For Case 3, if $X_F$ is utilized by the controller when in fact it is $Z_F$ that affects $y_F$, a cost of $c_1$ per squared unit of $X_F$ used is incurred by the controller. Similarly, if $Z_F$ is used to control $y_F$ when it is $X_F$ that affects $y_F$, a cost of $c_2$ per squared unit of $Z_F$ used is incurred. A cost structure of this nature would be appropriate if, say, use of the inappropriate instrument caused some undesirable side effects. For example, suppose a firm were interested in increasing the productivity of its assembly line workers and believed that an annual Christmas pay bonus was an appropriate means of so doing. It may turn out, however, that the bonus does not affect the assembly line workers' productivity, but does hurt the morale and, therefore, the productivity of the firm's employees that did not receive a pay bonus. If such were the case, the pay bonus would be an inappropriate
incentive that would have costly side effects.

Since in Case 3, the controller is uncertain as to which instrument affects $y_F$, he is uncertain which instrument is inappropriate, and, therefore, is uncertain of the cost he would incur by utilizing $X_F$ and/or $Z_F$. Thus, unlike Cases 1 and 2, it is the expected cost of Case 3 that the controller should consider. Noting that the probability of incurring a cost for use of, say, $X_F$ is the probability that $M_1$ is inappropriate for use in control of $y_F$, i.e., $P''(M_2|y,z)$, the controller can consider Case 3 costs by including $c_1X_F^2$ and $c_2Z_F^2$ in (4.31) as follows:

\[
\min_{D_F} \{ P'(M_1|y,x)E_{\beta_1,\varepsilon,\sigma^2}\{y,x,x_F[y_F^* - (\beta_1 X_F + \varepsilon)]^2 + P''(M_2|y,z)E_{\beta_2,\delta,\sigma^2}\{y,z,z_F[y_F^* - (\beta_2 Z_F + \delta)]^2 + P''(M_1|y,x)C_2Z_F^2 + P''(M_2|y,z)C_1X_F^2 \}. \quad (4.54)
\]

As in the previous cases considered, the minimization problem of (4.54) separates into two minimization problems, one of which is

\[
\min_{X_F} \{ P'(M_1|y,x)E_{\beta_1,\varepsilon,\sigma^2}\{y,x,x_F[y_F^* - (\beta_1 X_F + \varepsilon)]^2 + P''(M_2|y,z)C_1X_F^2 \}. \quad (4.55)
\]

The other is
\[
\min_{Z_F} \{P''(M_2|y,z)E_{\beta_2,\delta,\sigma_0^2|y,z,Z_F}[y_F^* - (\beta_2 Z_F + \delta)]^2 + P''(M_1|y,x)C_2 Z_F^2 \}.
\]

(4.56)

The solutions to (4.55) and (4.56) may be obtained via the same procedure used to solve (4.32) and (4.44). Accordingly, the optimal setting for \(X_F\) is found to be

\[
X_F = \frac{y_F^* b_1''}{V(\beta_1) + b_1''^2 + \frac{p_2}{p_1} C_1}.
\]

(4.57)

and the optimal setting for \(Z_F\) is found to be

\[
Z_F = \frac{y_F^* b_2''}{V(\beta_2) + b_2''^2 + \frac{p_1}{p_2} C_2}.
\]

(4.58)

In both (4.57) and (4.58), \(P''(M_1|y,x) = P_1\) and \(P''(M_2|y,z) = P_2\).

Thus, the optimal BMC control solution for Case 3 is \(D_F = (X_F,Z_F)'\), where \(X_F\) and \(Z_F\) are as given in (4.57) and (4.58), respectively.

The following statements can be made about the instrument settings established by the optimal BMC control procedure:

1. In each of Cases 1, 2, and 3, as well as the cost-free case of Section IV.4.1, as the precision of the information known about an instrument's response parameter, \(\beta_1\),
increases, the instrument's optimal setting increases; i.e., as, say, $V(\beta_1)$ decreases, $X_F$ increases.

(2) In each of Cases 1, 2, and 3, as the cost per unit (or cost per squared unit) of an instrument used increases, the instrument's optimal control setting decreases.

(3) In each of Cases 1, 2, and 3, as the probability of a particular model being true increases, the optimal setting of the model's instrument increases, while at the same time the optimal setting of the other model's instrument decreases. In Case 3, this is a result of the expected cost per squared unit of, say, $X_F$ decreasing as $P'(M_1|y,x)$ increases.

Assuredly, Cases 1, 2, and 3 are not the only instrument-use cost situations that the controller may face. Consider the following:

(1) Zellner has examined the problem of how to account for the cost of changing the setting of an instrument in optimal single-period control problems in which a particular model is assumed to generate $y_F$.\(^1\)

(2) The forms of the cost structures used in Cases 1, 2, and 3 are certainly not exhaustive.

(3) Some combination of Cases 1, 2, and 3 may be appropriate.

(4) Some combination of Cases 1, 2, and/or 3 which also involves a cost for changing the level or rate of existing instrument settings may be appropriate.

\(^1\)Zellner, pp. 324-325.
As an example of (2), suppose the cost of using the inappropriate variable is better expressed by a linear function than by the quadratic function used in Case 3. Then the optimal BMC control settings for $X_F$ and $Z_F$ would be

$$X_F = \frac{y_F b_1'' - \frac{P_2 d_1}{2P_1}}{V(\beta_1) + b_1''^2}$$  \hspace{1cm} (4.59)$$

and

$$Z_F = \frac{y_F b_2'' - \frac{P_1 d_2}{2P_2}}{V(\beta_2) + b_2''^2}.$$  \hspace{1cm} (4.60)$$

Concerning (3), suppose Cases 1 and 3 are both relevant. Let Case 1 costs be as described above. For Case 3, let a cost of $g_1$ per squared unit of $X_F$ used be incurred if $X_F$ is the inappropriate instrument, or $g_2$ per squared unit of $Z_F$ used if $Z_F$ is the inappropriate instrument. Then, the optimal BMC control settings for $X_F$ and $Z_F$ would be

$$X_F = \frac{y_F b_1''}{V(\beta_1) + b_1''^2 + \frac{P_2}{P_1} g_1 + \frac{C_1}{P_1}}$$  \hspace{1cm} (4.61)$$

and
\[Z_F = \frac{y_F^* b_1''}{V(\beta_2) + b_2''^2 + \frac{p_1}{p_2} g_2 + \frac{c_2}{p_2}}. \quad (4.62)\]

In the next section, the certainty-equivalent BMC control solutions for Cases 1, 2, and 3 of this section are presented.

**IV.4.3 Certainty-Equivalent BMC Control Solutions When Instrument Use Costs Are Considered**

The certainty-equivalent BMC control solutions for Cases 1, 2, and 3 of the previous section are derived from the minimization problem of (4.13):

\[
\min_{D_F} \{ P''(M_1 | y, X)[E_{y_F | M_1, y, X} (y_F) - y_F^*] + P''(M_2 | y, Z)[E_{y_F | M_2, y, Z} (y_F) - y_F^*]^2 \}. \quad (4.63)
\]

The solutions are derived in the same manner as the optimal BMC control solutions for Cases 1, 2, and 3.

The following are the certainty-equivalent BMC control solutions for Cases 1, 2, and 3:

**Case 1:**

\[x_F = \frac{y_F^*}{b_1'' + \frac{c_1}{b_1'' p_1}}. \quad (4.64)\]
\[ Z_F = \frac{y_F^*}{b_2'' + \frac{c_2}{b_2''P_2}} \]  \hspace{1cm} (4.65)

CASE 2:

\[ X_F = \frac{y_F^* - \frac{d_1}{2P_1}}{b_1''} \]  \hspace{1cm} (4.66)

\[ Z_F = \frac{y_F^* - \frac{d_2}{2P_2}}{b_2''} \]  \hspace{1cm} (4.67)

CASE 3:

\[ X_F = \frac{y_F^*}{b_1'' + \frac{p_2c_1}{b_2'' + \frac{p_1c_2}{b_1''}}} \]  \hspace{1cm} (4.68)

\[ Z_F = \frac{y_F^*}{b_2'' + \frac{p_1c_2}{b_2''}} \]  \hspace{1cm} (4.69)

The notation used in (4.64) through (4.69) is the same as that of the previous section. Notice that the only difference between these solutions and those of the previous section is in the denominators. In particular, the \( V(\beta_1) \) term does not appear in the denominators of (4.64) through (4.69). As a result, the certainty-equivalent BMC control settings for \( X_F \) and \( Z_F \) are at least as high in all cases as...
the optimal BMC control settings.\(^1\) The absence of \(V(\beta_i)\) in the denominator, and the resulting higher settings of \(X_F\) and \(Z_F\), is due to the failure of the certainty-equivalent approach to consider the uncertainty of \(y_F\).

By not considering \(y_F\)'s uncertainty, the certainty-equivalent approach ignores some of the risk involved in choosing settings for \(X_F\) and \(Z_F\). The optimal BMC control problem reflects this risk; i.e., it recognizes that the higher are the settings for \(X_F\) and \(Z_F\), the higher will be \(y_F\)'s predictive variance, and, therefore, the greater is the probability of \(|y_F - y_F^*|\) being very large. The optimal BMC control procedure sets \(X_F\) and \(Z_F\) in such a manner that the probability of the controller incurring a large loss as a result of \(y_F\)'s predictive variance being large is smaller than under the certainty-equivalent approach. Accordingly, if the controller's loss function is symmetric\(^2\) (as with the quadratic loss function used in this chapter), optimal BMC control settings for \(X_F\) and \(Z_F\) will be lower than those of certainty-equivalent BMC control.

---

\(1\) The certainty-equivalent BMC control settings will be higher in all cases than the optimal BMC control settings if \(V(\beta_i) > 0, i = 1, 2\), i.e., if \(\beta_1\) and \(\beta_2\) are not known.

The risk taken by the controller in setting his instruments is considered from the point of view of optimal BMC control in the next section.

IV.4.4 Risk Specification in Optimal BMC Control

As discussed extensively in Chapter III and in Sections IV.3.2 and IV.4, if residual, parameter, and model specification uncertainty exist for a controller and/or forecaster, the BMMP is the appropriate distribution to use to characterize the process generating $y_F$. Thus, if the predictive variance of $y_F$ is viewed as a measure of the risk the controller takes in setting his instruments so as to attain $y_F^*$, the variance of the BMMP distribution is the appropriate measure of his risk. The optimal BMC control procedure utilizes a BMMP to determine control settings for $X_F$ and $Z_F$. Accordingly, by using optimal BMC control, the controller can determine settings for his instruments which appropriately reflect his uncertainty concerning the process generating $y_F$ and the risk he will be taking in attempting to attain $y_F^*$. As previously noted, this is why it is said that the BMC control procedure, introduced in Section IV.4, yields optimal instrument settings for single-period control problems. Control procedures that ignore existing model specification uncertainty ignore available information concerning the risk involved with controlling $y_F$, and, by definition, cannot yield optimal instrument settings.
IV.4.5 BMC Control When More Complicated Models Are Included in the Model Space

Two reasons why the model space considered in the previous sections of this chapter might be employed by a controller are the following:

(1) The controller believes that the target variable is affected exclusively by one or the other of the two instruments.

(2) It is more convenient and/or computationally efficient for the controller to specify his beliefs about the nature of the data-generating process via two simple models.

In either case, both the certainty-equivalent and optimal BMC control solutions for such model spaces, when instrument use costs are not considered, would have the controller set each instrument as if it were the only instrument that affected the target variable. Such solutions, and therefore such simplistic model spaces, could in reality lead to significant over-shooting or under-shooting of the established target. For example, suppose the target variable were the rate of inflation in period t and the two instruments were government expenditures and the rate of growth of the money supply in period t-1. The above-mentioned control solutions would set government expenditures as if the growth rate of the money supply had no effect on the inflation rate, and would set the growth rate of the money supply as if government expenditures had no effect on the inflation rate. Should both government expenditures and the growth rate of the money supply in period t-1 be positively related to the inflation rate in period t,
such a control policy could result in overspending, an inappropriately high money supply level, and, therefore, a rate of inflation above the targeted level.

If the controller believes that both instruments might affect the target variable he can help avoid over-shooting or under-shooting his target by expanding the model space to reflect his belief. The following are examples of two possible expansions:

(1) $M_1$: \[ y = \beta_1 X + \epsilon \]
$M_2$: \[ y = \beta_2 Z + \delta \]
$M_3$: \[ y = \beta_3 X + \beta_4 Z + \gamma \]

(2) $M_1$: \[ y = \beta_1 X + \epsilon \]
$M_2$: \[ y = \beta_2 Z + \delta \]
$M_3$: \[ y = \beta_3 X + \beta_4 Z + \beta_5 XZ + \gamma \].

In both cases, the inclusion of model three forces the control solutions yielded by the BMC procedures to reflect the possible dependence of $y_F$ on both $X_F$ and $Z_F$, as well as the possibility that only one or the other of $X_F$ and $Z_F$ affect $y_F$. Model three of example one describes the dependence of $y_F$ on $X_F$ and $Z_F$ as being additive. Model three of example two describes $y_F$ as being dependent on the interaction between $X_F$ and $Z_F$, as well as $X_F$ and $Z_F$ main effects. Thus, depending on the magnitudes of the $\beta$ coefficients, the $X_F$ and $Z_F$ settings yielded by the BMC certainty-equivalent and optimal control procedures may be lower or higher than in the two-model cases perviously considered.
Analytic BMC certainty-equivalent and optimal control solutions for more complex model spaces, like the two above, can be found by the straightforward application of the methods used to solve the two model cases earlier in this chapter. One complicating difference arises, however, when the model set reflects the target variable as being possibly dependent on more than one instrument. In such cases, the solution to the control problem, i.e., the minimization problem of (4.3), does not separate into a series of independent minimization problems, one for each instrument, as was the case in (4.13) and (4.31). Instead, the simultaneous solution of two or more equations is required to obtain control settings for the instruments. For example, obtaining the optimal BMC control settings for $X_F$ and $Z_F$ utilizing the model space of example one above requires the simultaneous solution of two equations in two unknowns. This presents no particular problem unless the equations to be solved simultaneously are of degree higher than one. In that case, the determination of instrument settings may involve the cumbersome task of finding the roots of high degree polynomials. This is precisely what happens when the model space of example two above is utilized to find optimal BMC control settings for $X_F$ and $Z_F$.

This chapter discussed the application of the BMC procedure to single-period economic control problems. It was noted that, unlike the traditional control solutions which are artificially conditioned on the assumption that a particular econometric model is the true model of the data-generating process, BMC control solutions appropriately reflect the controller's model specification uncertainty. Analytic
solutions were found for simple single-period control problems using both the BMC certainty-equivalent and BMC optimal approaches to control. Solutions were obtained assuming control was cost free, as well as for cases in which instrument-use cost functions were assumed to be known. A discussion of several important questions regarding the relation of BMC control to other control methods is deferred until Chapter VI.

The next chapter introduces a Bayesian procedure for making inferences about certain types of nonstationary data-generating processes.
In an approach to modeling the uncertainty involved in
decision making situations referred to here as the Bayesian
Model Switching (BMSW) approach, it is assumed that the random
variable upon which a decision hinges is generated by different
statistical models in different time periods with the switch
between models described by some random process. In this chapter
the methodology of the BMSW approach is developed and applied to
the case where it is known that the decision variable is generated
by two different "switching" normal distributions. As a practical
approach to handling model nonstationarity the BMSW methodology
is shown to be computationally unwieldy even when the number of
time periods considered is very small. It is hoped, however,
that the BMSW approach yields useful new insights into the
problem of modeling nonstationary processes.

V.1 Bayesian Model Switching Methodology

The BMSW approach was suggested by anomalies observed in
sequences of posterior model probabilities of sets of competing
models. For example, in comparing five alternative aggregate
consumption functions via the Bayesian Model Comparison
procedure, Wiginton\(^1\) generated the following sequence and others like it:\(^2\)

<table>
<thead>
<tr>
<th>Year</th>
<th>(P(M_1))</th>
<th>(P(M_2))</th>
<th>(P(M_3))</th>
<th>(P(M_4))</th>
<th>(P(M_5))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1911</td>
<td>.008</td>
<td>.781</td>
<td>.200</td>
<td>.011</td>
<td>*</td>
</tr>
<tr>
<td>1912</td>
<td>.005</td>
<td>.796</td>
<td>.193</td>
<td>.006</td>
<td>*</td>
</tr>
<tr>
<td>1913</td>
<td>.004</td>
<td>.794</td>
<td>.199</td>
<td>.003</td>
<td>*</td>
</tr>
<tr>
<td>1914</td>
<td>.001</td>
<td>.600</td>
<td>.395</td>
<td>.003</td>
<td>*</td>
</tr>
<tr>
<td>1915</td>
<td>*</td>
<td>.015</td>
<td>.982</td>
<td>.002</td>
<td>*</td>
</tr>
<tr>
<td>1916</td>
<td>*</td>
<td>.010</td>
<td>.990</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

where * signifies posterior probability < \(10^{-3}\). Notice the dramatic change that occurs in the posterior probabilities of model 2 and model 3 between the years 1914 and 1915. With only one additional bit of data (i.e., 1915's aggregate consumption), model 2's posterior probability drops from .6 to .015 and model 3's increases from .395 to .982. Given that during any one particular period one of the five alternative representations of the aggregate consumption process is appropriate, the significant changes in posterior model probabilities suggest that the data being observed may have been generated by different models in different time periods. In the above example, it appears that model 2 may have been the more accurate representation of the aggregate consumption process before 1915 and model 3 the more accurate during and after 1915. Thus, in situations where the


\(^2\)Martin S. Geisel, "Bayesian Comparison of Simple Macroeconomic Models," *Journal of Money, Credit, and Banking*, 5 (October, 1973), 759-762.
posterior model probabilities tend to vacillate rather than converge over time, it may be better to assume that the process generating the observations of interest is nonstationary rather than assuming stationarity as is done in the applications of the Bayesian Model Comparison and Bayesian Model Selection procedures considered in this dissertation. Certainly environmental change is the rule rather than the exception, and could explain movements in posterior model probabilities such as those witnessed above.

Given a set of $N$ alternative switching models the BMSW approach assumes that one of the $N$ models is "in control" each period (i.e., is an appropriate representation of the process generating the random variable of interest each period) and that all $N$ models have a positive probability of "taking control" in any period. It is not known, however, which model is in control in any particular period. Utilizing a multinomial process to describe the switches among models from period to period, the random variable of interest may be characterized as follows:

\[ f(Y|\theta_1, \theta_2, \ldots, \theta_N) = \pi_1 f(Y|M_1, \theta_1) + \]
\[ \quad \pi_2 f(Y|M_2, \theta_2) + \ldots + \pi_N f(Y|M_N, \theta_N). \tag{5.1} \]

$Y$ is the variable of interest, $M_i$ stands for "model $i$", $\theta_i$ is the unknown parameter of model $i$ (possibly a vector), and $\pi_i$ is the probability that model $i$ is in control during any particular period. For simplicity it will be assumed that the $\pi_i$'s are
known \( \sum_{i=1}^{N} \pi_i = 1 \). The function \( f(Y|M_i, \theta_i) \) is the probability distribution of \( Y \) as characterized by model \( i \). Thus the process generating \( Y \) is modeled as a convex linear combination of the \( N \) alternative switching models.

The characterization of a random process as a linear combination (or mixture) of two or more other processes can arise as a result of "richness" considerations as well as from nonstationarity. Statistical models are, after all, only approximations to reality, and a mixture of models may, for reasons of computational efficacy, intuitive appeal, and/or predictability, prove to be a better approximation than any one model by itself.

Quandt (and later Goldfeld and Quandt)\(^1\) uses a mixture of processes from the same family to model the behavior of his "switching regression regimes." He was concerned with discontinuous shifts in the parameters (specifically the regression coefficients) of a single linear regression model. The methodology of this chapter, however, can be applied equally as well to shifts in the parameters of the controlling model, and/or to switches in the mathematical form of the controlling model, and/or to switches in the independent variables of the controlling model over time. Swamy and Mehta note that the likelihood

---

function yielded by Quandt's classical approach to estimating the parameters of his switching regimes generates likelihoods that are "very high along unreasonable paths of the parameter space. . . ." ¹ They also point out that the parameter vector containing the parameters of Quandt's regression regimes is not identifiable without restrictions. Drawing on Lindley's work, Swamy and Mehta suggest that the problems that arise in Quandt's approach can be avoided if proper prior distributions for the parameters in his switching regimes are employed. ² The approach taken in this chapter is essentially that suggested by Swamy and Mehta, the difference being that the methodology of this chapter explicitly permits the model of the changing data-generating process to reflect changes in the process's mathematical form and/or controlling variables as well as changes in its parameters.

As it stands, with its parameters unknown, (5.1) is not very useful to the decision maker. What is needed is a predictive distribution, a distribution of as-yet-to-be-seen \( Y \) values, \( f(Y) \), which is not conditioned on the parameters. However, the development of such a distribution, as will be seen below, is extremely difficult even in the special cases in which \( N = 2 \) and \( M_1 \) and \( M_2 \)


² Ibid.
are both normal models with known variances and unknown means. A proxy, therefore, is needed for the desired predictive. One proxy is the distribution arrived at by substituting estimates of \( \theta_1, \theta_2, \ldots, \theta_N \) obtained via a Bayesian estimation procedure into \( f(Y|\theta_1, \theta_2, \ldots, \theta_N) \).

In order to be able to obtain estimates for the unknown parameters and be able to revise them as new observations on \( Y \) are observed, thus revising \( f(Y) \), it is necessary to establish a scheme for revising the decision-maker's prior distribution on \( \theta_i \). It is assumed that initially the decision maker knows the set of \( N \) switching models and knows \( \pi_i, i = 1, 2, \ldots, N \). All other relevant information possessed by the decision maker is summarized in his prior distributions of the parameters of the switching models, \( g'(\theta_i), i = 1, 2, \ldots, N \). When one or more new \( Y \) values are observed (i.e., when the process being modeled is sampled), \( g'(\theta_i) \) is revised via Bayes Rule as follows:

\[
\text{g}''(\theta_i|Y) = g'(\theta_i)\mathcal{L}(Y|\theta_i)\int_{\theta_i} g'(\theta_i)\mathcal{L}(Y|\theta_i)\,d\theta_i.
\]

The function \( g'(\theta_i) \) is the decision-maker's prior distribution of \( \theta_i \), \( g''(\theta_i|Y) \) is the decision-maker's posterior distribution of \( \theta_i \), and \( \mathcal{L}(Y|\theta_i) \) reflects the likelihood of a particular value of \( \theta_i \) having "generated" the observed \( Y \). The likelihood function \( \mathcal{L}(Y|\theta_i) \), is structurally quite unusual. Like \( f(Y) \), it is a mixture, but instead of being a mixture of models, it is a mixture of likelihood functions:
\[ l(Y|\theta_i) = \pi_1 f(Y|M_1, \theta_i) + \pi_2 f(Y|M_2) + \ldots + \pi_i f(Y|M_i, \theta_i) + \ldots + \pi_N f(Y|M_N). \]  

(5.3)

\( \pi_i \), as before, is the probability that model \( i \) is in control during any particular period. The function \( f(Y|M_i) \), given \( Y \), is a model likelihood. It reflects the likelihood of observing \( Y \) given that \( Y \) was generated by model \( i \) (i.e., given that model \( i \) was in control). The function \( f(Y|M_i, \theta_i) \) is also a likelihood function, but it reflects the likelihood of observing \( Y \) given that \( Y \) was generated by model \( i \) and that model \( i \)'s parameter is \( \theta_i \). Thus, the probability revision of (5.2) is carried out without assuming that a particular model has generated the observed data, \( Y \). This is unlike the revision processes in the BMC and BMS approaches where in updating either \( f'(\theta_i) \) or \( P'(M_i) \) it is always assumed that the observed data have been generated by the model being revised.

Equation (5.3) is developed as follows:

\[ l(Y|\theta_i) = \pi_1 f(Y|M_1, \theta_i) + \pi_2 f(Y|M_2, \theta_i) + \ldots + \pi_i f(Y|M_i, \theta_i) + \ldots + \pi_N f(Y|M_N, \theta_i) \]  

(5.4)

where

\[ f(Y|M_i, \theta_i) = f(Y|M_i), \ i \neq j, \text{ and} \]  

(5.5)

\[ f(Y|M_i) = \int_{\Theta_i} f(Y|M_i, \theta_i)g'(\theta_i)d\theta_i. \]  

(5.6)

Thus, upon substituting (5.5) and (5.6) into (5.4), (5.3) is
obtained. Equation (5.3) is repeated here:

\[
\xi(Y|\theta_i) = \pi_1 f(Y|M_1) + \pi_2 f(Y|M_2) + ... + \\
\pi_i f(Y|M_i, \theta_i) + ... + \pi_N f(Y|M_N).
\]  

(5.7)

In (5.2) \(\theta_i\) is revised without assuming that the observed data was generated by model i. Thus, the likelihood function being used in the revision process should reflect the likelihood of the data having been generated by each of the N switching models. Accordingly, a likelihood function that is a mixture of model likelihoods, where the mixing weights are the probabilities that particular models generated \(Y\), is appropriate.

Now continuing with the development of the revision process, substitute (5.7) into (5.2):

\[
g''(\theta_i | Y) = g'(\theta_i) [\pi_1 f(Y|M_1) + ... + \pi_i f(Y|M_i, \theta_i) + ... + \\
\pi_N f(Y|M_N)] / \int_{\theta_i} g'(\theta_i) [\pi_1 f(Y|M_1) + ... + \\
\pi_i f(Y|M_i, \theta_i) + ... + \pi_N f(Y|M_N)] d\theta_i
\]

(5.8)

\[
g'(\theta_i) \left[ \frac{\text{same as above}}{\pi_1 f(Y|M_1) + ... + \pi_i f(Y|M_i) + ... + \pi_N f(Y|M_N)} \right].
\]  

(5.9)

The denominator of (5.9) is in a form similar to the predictive distribution of interest. \(Y\)'s predictive distribution prior to sampling is:
\[ f(Y) = \pi_1 f(Y|M_1) + \ldots + \pi_i f(Y|M_i) + \ldots + \pi_N f(Y|M_N). \] (5.10)

Unlike the model in equation (5.1), \( f(Y) \) is not conditioned on the parameters of the switching models. The decision-maker's uncertainty about the parameters is reflected in (5.10), making (5.10) generally more useful to the decision maker than (5.1).

After \( Y \) is observed and \( g'(\theta_i), i = 1, 2, \ldots, N \) is revised, the following predictive distribution is obtained:

\[
f(Y_F|Y) = \int g''(\theta_i)[\pi_1 f(Y_F|M_1, Y) + \ldots +
\pi_i f(Y_F|M_i, \theta_i) + \ldots + \pi_N f(Y_F|M_N, Y)]
= \pi_1 f(Y_F|M_1, Y) + \ldots + \pi_i f(Y_F|M_i, Y) + \ldots +
\pi_N f(Y_F|M_N, Y),
\] (5.11)

where

\[
f(Y_F|M_i, Y) = \int g''(\theta_i|Y)f(Y_F|M_i, \theta_i)d\theta_i
\]

and \( Y_F \) denotes a future value of \( Y \). It is this predictive distribution, \( f(Y_F|Y) \), that the decision maker should use to model the process generating \( Y \). But, it is also this distribution that, as mentioned earlier, and will be demonstrated below, is extremely difficult to derive.

Before applying the preceding methodology to a specific problem, it would be helpful to understand an expansion of (5.9) that arises in the application. (5.9) may be expanded as follows:
\[ g'(\Theta_i | \gamma) = \left[ \pi_i g'(\Theta_i) f(Y|M_i, \Theta_i) / \pi_i f(Y|M_i) \right] + \ldots + \pi_i f(Y|M_i) + \ldots + \pi_N f(Y|M_N) \]

\[ \left\{ g'(\Theta_i) \left[ \pi_1 f(Y|M_1) + \ldots + \pi_{i-1} f(Y|M_{i-1}) + \pi_i f(Y|M_i) + \ldots + \pi_N f(Y|M_N) \right] \right\} / \left[ \pi_1 f(Y|M_1) + \ldots + \pi_i f(Y|M_i) + \ldots + \pi_N f(Y|M_N) \right]. \]

(5.12)

Call the first term after the equal sign on the right-hand side of (5.12) \( A \) and the second term on the right \( B \). It is \( B \) that is of interest.

\[ B = g'(\Theta_i) \left[ \pi_i^n + \ldots + \pi_{i+1}^n + \ldots + \pi_N^n \right] \]

(5.13)

where

\[ \pi_j^n = \pi_j f(Y|M_j) / \left[ \pi_1 f(Y|M_1) + \ldots + \pi_i f(Y|M_i) + \ldots + \pi_j f(Y|M_j) + \ldots + \pi_N f(Y|M_N) \right]. \]

(5.14)

Since \( \pi_j \) equals the probability that model \( j \) is in control during any particular period, and \( f(Y|M_j) \) represents the likelihood of \( Y \) being generated by model \( j \), \( \pi_j^n \) can be interpreted as the posterior probability (i.e., posterior to \( Y \)) that model \( j \) was in control during the period in which \( Y \) was observed. Accordingly, from now on \( \pi_j = \pi_j' \) and will be referred to as the prior probability that model \( j \) will be in effect during any particular
time period.

V.2 Special Case: Two Normal Models

The revision methodology developed in Section V.1 is applied below to the problem established by the following assumptions:

(1) The decision maker knows that the data-generating process in which he is interested is nonstationary.

(2) Further, he knows that the nonstationarity is caused by the fact that in different time periods one or the other of two different processes is generating the data.

(3) He cannot identify when a particular process will be in control, but he does know \( \pi_i \), \( i = 1, 2 \), the fixed probability that process \( i \) will be in control during any particular period.

(4) The decision maker knows that both processes may be best represented by normal models (i.e., normal distributions) with known variance \( \sigma^2 \), and unknown means, \( \mu_1 \) and \( \mu_2 \):

\[
M_1: f_N(Y|\mu_1, \sigma^2);
M_2: f_N(Y|\mu_2, \sigma^2). \tag{5.15}
\]

The decision maker should model the nonstationary process as follows:1

1Since \( Y \) is generated each period by either \( M_1 \) or \( M_2 \) the data-generating process is nonstationary relative to \( M_1 \) and \( M_2 \). Relative to (5.16), however, the process is stationary since the data are assumed to be generated by (5.16) in each period.
\[
f(Y|\pi_1, \pi_2, \mu_1, \mu_2, \sigma^2) = \pi_1 f_N(Y|\mu_1, \sigma^2) + \\
\pi_2 f_N(Y|\mu_2, \sigma^2).
\] (5.16)

For example, suppose the decision maker is a security analyst who views the returns on shares as depending on some event. Conditioned on the event occurring, he might assess the distribution of returns to be \(f_N(Y|\mu_1, \sigma^2)\). However, if he does not know a priori whether the event will occur, he may assess the marginal distribution of returns to be a bimodal distribution similar to (5.16).¹

As data are observed (5.2) is used to obtain the posterior distributions of \(\mu_1\):

\[
g^*(\mu_1|Y) = g'(\mu_1)\mathcal{L}(Y|\mu_1) / \int_{-\infty}^{\infty} g'(\mu_1)\mathcal{L}(Y|\mu_1) d\mu_1, \quad (5.17)
\]

\(i = 1, 2\). The functions \(g^*(\mu_1|Y)\), \(g'(\mu_1)\), and \(\mathcal{L}(Y|\mu_1)\) were all defined above. For convenience, the revision process will be demonstrated via revision of \(\mu_1\). Utilizing (5.9), (5.17) becomes

\[
g^*(\mu_1|Y) = g'(\mu_1)[\pi_1 f(Y|M_1, \mu_1) + \pi_2 f(Y|M_2)] / \\
[\pi_1 f(Y|M_1) + \pi_2 f(Y|M_2)]. \quad (5.18)
\]

Before proceeding with the specifics of the revision process, it is necessary to know the form of the decision-maker's prior

distributions on \( \mu_1 \) and \( \mu_2 \). It will be assumed that

\[
g'(\mu_1) = N(m_1, \frac{\sigma_1^2}{n_1})
\]

\[
g'(\mu_2) = N(m_2, \frac{\sigma_2^2}{n_2}),
\]

(5.19)

where \( n_i > 0, i = 1, 2 \). Continuing with the development of

(5.18), \( f(Y|M_1, u_1) = f_N(Y|\mu_1, \sigma^2) \) from (5.15). Or, in terms

of the sufficient statistic

\[
m = \frac{1}{n} \sum_{j=1}^{n} Y_j,
\]

as derived by factorization in Raiffa and Schlaifer,\(^1\)

\[
f(Y|M_1, u_1) = R_1(Y)f_N(m|M_1, \mu_1, \frac{\sigma^2}{n}),
\]

(5.20)

where \( n \) is the sample size and \( R_1(Y) \) is a residual which is a function of \( Y \) alone. Similarly,

\[
f(Y|M_2, u_2) = R_2(Y)f_N(m|M_2, \mu_2, \frac{\sigma^2}{n}).
\]

(5.21)

Then, using another Raiffa and Schlaifer result:\(^2\)

\[
f(Y|M_2) = \int_{-\infty}^{\infty} f(Y|M_2, \mu_2)g'(\mu_2)d\mu_2
\]

(5.22)

\[
= \int_{-\infty}^{\infty} R_2(Y)f_N(m|M_2, \mu_2, \frac{\sigma^2}{n})g'(\mu_2)d\mu_2
\]

\[
= R_2(Y)f_N(m|m_2', \frac{n_2' + n}{n_2'n} \sigma^2).
\]

---

\(^1\)Raiffa and Schlaifer, p. 294.

\(^2\)Ibid., p. 296.
All of the terms in (5.18) have now been identified. However, before actually substituting these results into (5.18), it will be shown that \( R_1(Y) = R_2(Y) \), thus permitting all further analysis of \( g''(\theta | Y) \) to be done in terms of the sufficient statistic \( m \) instead of \( Y \). From (5.20) and (5.21),

\[
f(Y | \mu_1, M_1) = R_1(Y) f_N(m | \mu_1, \frac{\sigma^2}{n}, M_1)
\]

and

\[
f(Y | \mu_2, M_2) = R_2(Y) f_N(m | \mu_2, \frac{\sigma^2}{n}, M_2).
\]

More specifically,

\[
f(Y | \mu_1, M_1) = \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp\left[ -\frac{1}{2\sigma^2} \left( \sum (Y_j - \mu_1)^2 \right) \right] \quad (5.24)
\]

\[
= \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp\left[ -\frac{1}{2\sigma^2} \left( \sum (Y_j - m)^2 \right) \right] \exp\left[ -\frac{2}{2\sigma^2} (m - \mu_1)^2 \right]
\]

\[
= \frac{1}{n} \left( \frac{1}{\sqrt{2\pi\sigma}} \right)^{n-1} \exp\left[ -\frac{1}{2\sigma^2} (Y_j - m)^2 \right] \frac{\sqrt{n}}{\sqrt{2\pi\sigma}} \exp\left[ -\frac{n}{2\sigma^2} (m - \mu_1)^2 \right] \quad (5.25)
\]

\[
= R_1(Y) f_N(m | \mu_1, \frac{\sigma^2}{n}, M_1)
\]

as in (5.20). Notice that \( R_1(Y) \) is only a function of the observed data. Thus, if \( f(Y | \mu_2, M_2) \) were broken down in an analogous fashion, utilizing the same data, \( R_2(Y) \) would obviously equal \( R_1(Y) \). To see this, substitute \( \mu_2 \) for \( \mu_1 \) in (5.25).
Returning with the above results to the revision process, (5.19), (5.20) and (5.23) are substituted into (5.18) yielding

\[ g''(\mu_1 | Y) = \left\{ \frac{\pi_1 R_1(Y)f_\text{N}(m|m_1^1, \frac{n_1^1 + n}{n_1^1 n}, \sigma^2)}{\pi_1 f_\text{N}(m|m_1^1, \frac{n_1^1 + n}{n_1^1 n}, \sigma^2) + \pi_2 f(m|m_2^1, \frac{n_2^1 + n}{n_2^1 n}, \sigma^2)} \right\} \]  

(5.26)

The residuals cancel leaving the following:

\[ g''(\mu_1 | Y) = \frac{g_N'(\mu_1 | m_1^1, \frac{\sigma^2}{n_1^1})\pi_1 f_\text{N}(m|\mu_1, \frac{\sigma^2}{n}, M)}{\pi_1 f_\text{N}(m|m_1^1, \frac{n_1^1 + n}{n_1^1 n}, \sigma^2) + \pi_2 f(m|m_2^1, \frac{n_2^1 + n}{n_2^1 n}, \sigma^2)} + \frac{2}{\text{same denominator as above}}. \]  

(5.27)

Recalling (5.14), (5.27) reduces further to

\[ g''(\mu_1 | Y) = \text{(same first term)} + g_N'(\mu_1 | m_1^1, \frac{\sigma^2}{n_1^1})\pi_2'. \]

Now, calling the denominator of (5.27) D and multiplying the numerator and denominator of the first term on the rhs of (5.27) by

\[ \int_{-\infty}^{\infty} f_N(\mu_1 | m_1^1, \frac{\sigma^2}{n_1^1})f_N(m|\mu_1, \frac{\sigma^2}{n})d\mu_1, \]
the following is obtained:

\[
g''(\mu_1 | Y) = \frac{1}{D} \left[ \frac{g^i_N(\mu_1 | m_1^i, \sigma^2/n_1^i) f_N(m_1^i | \mu_1, \sigma^2/n_1^i)}{\int f_N(\mu_1 | m_1^i, \sigma^2/n_1^i) f_N(m | \mu_1, \sigma^2/n_1^i) d\mu_1} \right] + \\
\int g^i_N(\mu_1 | m_1^i, \sigma^2/n_1^i) f_N(m | \mu_1, \sigma^2/n_1^i) d\mu_1 \\
g^i_N(\mu_1 | m_1^i, \sigma^2/n_1^i)_{\pi_2}^{\pi_2}
\]

Substituting for D in (5.28) and utilizing (5.14), (5.28) reduces to the following:

\[
g''(\mu_1 | Y) = \frac{g^i_N(\mu_1 | m_1^i, \frac{\sigma^2}{n_1^i + n})_{\pi_1} f_N(m | m_1^i, \frac{n_1^i + n}{n_1^i n} \sigma^2)}{\pi_1 f_N(m | m_1^i, \frac{n_1^i + n}{n_1^i n} \sigma^2) + \pi_2 f_N(m | m_2^i, \frac{n_2^i + n}{n_2^i n} \sigma^2)} + \\
g^i_N(\mu_1 | m_1^i, \frac{\sigma^2}{n_1^i})_{\pi_2}^{\pi_2}
\]

\[
g''(\mu_1 | Y) = g''(\mu_1 | m_1^i, \frac{\sigma^2}{n_1^i + n})_{\pi_1}^{\pi_1} + g^i_N(\mu_1 | m_1^i, \frac{\sigma^2}{n_1^i})_{\pi_2}^{\pi_2}.
\]  

Given that the observed data were generated by model i, \(m_1^i\) represents the posterior mean of \(\mu_i\). Equation (5.29) can be readily interpreted. It says that the posterior distribution of \(\mu_1\) is
a mixture of the posterior distribution of \( \mu_1 \) that would have resulted had it been known that the model 1 generated the data and the distribution of \( \mu_1 \) that would apply if it were known the data were generated by model 2 (i.e., the prior distribution of \( \mu_1 \)). The mixing weights \( \pi_1^m \) and \( \pi_2^m \), are, respectively, the probability that model 1 was in control during the period of observation, and the probability that model 2 was in control during the period of observation.

The form of the posterior mean of \( \mu_1 \) also has intuitive appeal and can be derived using (5.29) as follows:

\[
E^m(\mu_1) = \int_{-\infty}^{\infty} \mu_1 g^m(\mu_1 | Y) d\mu_1
\]

(5.30)

\[
= \int_{-\infty}^{\infty} \mu_1 \pi_1^m(\mu_1 | \mu_1^{m1}, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n}) d\mu_1 + \int_{-\infty}^{\infty} \mu_1 \pi_2^m g_N^m(\mu_1 | \mu_1^{m1}, \frac{\sigma^2}{n_1}) d\mu_1
\]

\[
= \pi_1^m \int_{-\infty}^{\infty} \mu_1 g_N^m(\mu_1 | \mu_1^{m1}, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n}) d\mu_1 + \pi_2^m \int_{-\infty}^{\infty} \mu_1 g_N^m(\mu_1 | \mu_1^{m1}, \frac{\sigma^2}{n_1}) d\mu_1
\]

\[
= \pi_1^m \mu_1^{m1} + \pi_2^m \mu_1^{m1}.
\]

(5.31)

The posterior mean of \( \mu_1 \) is thus a mixture of the posterior mean of \( \mu_1 \) that would have been obtained had it been known that \( Y \) was generated by model 1, and the posterior mean of \( \mu_1 \) (really the prior mean) that would apply if it were known that \( Y \) was generated by model 2. The mixing weights are the same as in (5.29).

\[\footnote{If the decision maker wished to revise his prior distribution on the parameters of \( M_1 \), and knew that \( Y \) had been generated by \( M_2 \), he obviously would not use \( Y \) to revise \( M_1 \)'s parameters.}\]
Revision of $\nu_2$'s distribution leads to results analogous to (5.29) and (5.31):

$$g''(\mu_2 | Y) = \pi_1 g''_N(\mu_2 | m_2, \frac{2}{n_2}) + \pi_2 g''_N(\mu_2 | m_2^*, \frac{2}{n_2^* + n})$$  (5.32)

and

$$E''(\mu_2) = \pi_1 m_2 + \pi_2 m_2^*.$$  (5.33)

Throughout all of the above analysis $Y$ could be either a scalar or a vector. If it were a vector, then it would seem that it is possible for it to contain observations generated by both model 1 and model 2, in which case the above analysis would not make sense. But, it was assumed that the decision maker is able to specify the time periods (called here control periods) during which one or the other of the models is in control. During these control periods the process of interest may generate more than one bit of data. Thus, $Y$ can be a scalar or a vector, but will always reflect only the data from one of these periods at a time, which is to say that $Y$ will never reflect data generated by more than one model. This assumption is not restrictive. For example, if the decision maker cannot say with certainty that the next two observations will be generated by the same model, then he must reduce his control period to that interval of time during which just one $Y$ value is generated.

The revision procedure developed above utilizes the fact that the next $Y$ observed, be it scalar or vector, contains all
the data generated in the next control period. Thus, this revision procedure is an end-of-period revision scheme. Accordingly, revision results (5.29), (5.31), (5.32), and (5.33) apply only at the end of the first control period for which data were available.

The next step in developing the BMSW revision process for this normal-normal case is to revise (5.29) and (5.30) using the data observed during the second control period. Thus, (5.18) must be revised. Equation (5.18) is repeated here with $Y_j$ representing the data from the $j^{th}$ time period:

$$g''(\mu | Y_2) = g'(\mu)[\pi_1 f(Y_2 | M_1, \mu) + \pi_2 f(Y_2 | M_2)]/$$

$$[\pi_1 f(Y_2 | M_1) + \pi_2 f(Y_2 | M_2)].$$

(5.34)

Functions $g'(\mu_i), i = 1, 2$, are the posteriors of period one, (5.28) and (5.32), here referred to as being "prior" to the data observed in period two. $\pi_i, i = 1, 2$, is, as before, the probability that model $i$ is in control during any particular period, in this case period two. $f(Y_2 | M_i, \mu)$ has exactly the same form as (5.25), leaving only $f(Y_2 | M_i)$ to be derived. Unlike the other terms in (5.34), $f(Y_j | M_i)$, being conditioned on all available information about $\mu_i$, changes from period to period as new information is incorporated into the probability distributions of model $i$'s parameters. This can be seen by examining the way in
which \( f(Y_j | M_1) \) is defined:

\[
f(Y_j | M_1) = \int_{-\infty}^{\infty} f(Y_j | M_1, \mu_i) g''(\mu_i | Y_{j-1}) d\mu_i. \tag{5.35}
\]

The function \( f(Y_2 | M_2) \) is derived here. By (5.25), \( f(Y_2 | M_2, \mu_2) = R_2(Y_2) f_N(m|\mu_2, \frac{\sigma^2}{n}, M_2) \). The function \( g''(\theta_2 | Y_1) \) is the mixture of normals displayed in (5.32). Thus,

\[
f(Y_2 | M_2) = \int R_2(Y_2) f_H(m|\mu_2, \frac{\sigma^2}{n}, M_2) [\pi'' f_N(\mu_2 | m_2^1, \frac{\sigma^2}{n_1^2}) + \\
\quad \pi_2'' f''(\mu_2 | m_2^2, \frac{\sigma^2}{n_2^2 + n})] d\mu_2
\]

\[
= R_2(Y_2) \left[ \pi'' \int_{-\infty}^{\infty} f_H(m|\mu_2, \frac{\sigma^2}{n}, M_2) f_N(\mu_2 | m_2^1, \frac{\sigma^2}{n_1^2}) d\mu_2 + \\
\quad \pi_2'' \int_{-\infty}^{\infty} f_N(m|\mu_2, \frac{\sigma^2}{n}, M_2) f''(\mu_2 | m_2^2, \frac{\sigma^2}{n_2^2 + n}) d\mu_2 \right]
\]

\[
= R_2(Y_2) \pi'' f_N(m|m_2^1, \frac{n + n_2^1}{n_2}, \sigma^2) + \\
\quad \pi_2'' f''(m|m_2^2, \frac{2n + n_2^1}{n(n_2 + n)}, \sigma^2). \tag{5.36}
\]

Similarly,

\[
f(Y_1 | M_1) = R_1(Y_2) \pi'' f_N(m|m_1^1, \frac{2n + n_2^1}{n_2}, \sigma^2) + \\
\quad \pi_2'' f''(m|m_1^1, \frac{n + n_1^1}{n_1}, \sigma^2). \tag{5.37}
\]

---

Even though \( g''(\mu_i | Y_{j-1}) \) is written as being conditioned on only \( Y_{j-1} \), it is, of course, conditioned on all the observed data of periods one through \( j-1 \).
Before (5.37) is substituted into (5.34), it will be helpful to introduce additional notation.

1) Let the prior distribution of \( u_i \), \( f^\prime_N(u_i|m_i, \frac{\sigma^2}{n_i}) \), be denoted by \( f^\prime_{\mu_i} \).

2) Let the posterior distribution of \( u_i \) after one period, where the revision was performed assuming \( Y_1 \) was generated by model \( i \), \( f^\prime_N(u_i|m_i, \frac{\sigma^2}{n'_i + n}) \), be denoted by \( f^\prime_{\mu_i} \).

3) Let the likelihood function of the sufficient statistic \( m \), when it is assumed that \( Y_j \) is generated by model \( i \) with mean \( \mu_i \), \( f_N(m|\mu_i, \frac{\sigma^2}{n}, M_i) \), be denoted by \( f^\prime_{M_i} \).

4) Let the model likelihood for model \( i \) in the first control period in terms of the sufficient statistic \( M \),
\[ f_N(m|m_i, \frac{n + n'_i}{nn'_i} \sigma^2) \], be denoted by \( f_{M_i} \).

5) Let \( f_N(m|m_i, \frac{2n + n'_i}{n(n'_i + n)} \sigma^2) \), the model likelihood for model \( i \) in the second control period conditioned on model \( i \) having generated \( Y_i \) (in terms of the sufficient statistic \( M \)), be denoted by \( f^2_{M_i} \).

6) Let the posterior distribution of \( u_i \) after period two be denoted by \( f^\prime_{\mu_i} \), \( f^\prime_{\mu_i} \), or \( f^\prime_{\mu_i} \) as it is assumed that \( Y_1 \) was generated by model 1 and \( Y_2 \) by model 1, or \( Y_1 \) by
model 1 and \( Y_2 \) by model 2, or \( Y_1 \) by model 2 and \( Y_2 \) by model 1, or \( Y_1 \) by model 2 and \( Y_2 \) by model 2, respectively.

Now, substituting (5.37) into (5.34), which is repeated here, the following is obtained:

\[
g''(\mu_1 | Y_2) = \frac{g''(\mu_1 | Y_1)[\pi_1 f(Y_2 | M_1, \mu_1) + \pi_2 f(Y_2 | M_2)]}{\pi_1 f(Y_2 | M_1) + \pi_2 f(Y_2 | M_2)} \tag{5.38}
\]

\[
= \frac{[\pi_1 f'' + \pi_2 f''][\pi_1 f_{\mu_1}^\phi + \pi_2 \{\pi_1 f_{M_1}^\phi + \pi_2 f_{M_2}^\phi\}]}{\int [\text{numerator}] d\mu_1} \tag{5.39}
\]

where the residuals, \( R_1(Y_2) \) and \( R_2(Y_2) \), have been excluded since they cancel. Expanding (5.39),

\[
g''(\mu_1 | Y_2) = \left\{ \pi_1 f'' f_{\mu_1}^\phi + [\pi_1 \pi_2 f_{M_1}^\phi + \pi_1 \pi_2 f_{M_2}^\phi] f'' \right\} + \pi_2 f'' f_{\mu_1}^\phi + [\pi_2 \pi_1 f_{M_1}^\phi + \pi_2 \pi_2 f_{M_2}^\phi] f'' \right\} / \int [\text{numerator}] d\mu_1 \tag{5.40}
\]
\[
\frac{\pi^m}{2} \int_{\mu_1} f^1_{\mu_1} f^2_{M_1} d\mu_1 + \left[ \pi^m \pi^m f^1_{M_2} + \pi^2 \pi^2 f^2_{M_2} \right] \} . 
\tag{5.41}
\]

Let \( I = \int_{-\infty}^{\infty} f^1_{\mu_1} f^2_{M_1} d\mu_1 \) and \( I_2 = \int_{-\infty}^{\infty} f^2_{\mu_1} f^2_{M_1} d\mu_1 \). Noticing that

\[
\frac{f^1_{\mu_1} f^2_{M_1}}{I_2} = f^1_{\mu_1}, \quad \frac{f^2_{\mu_1} f^2_{M_1}}{I_1} = f^2_{\mu_1}, \quad \frac{f^2_{\mu_1}}{I_1} = f^2_{\mu_1}, \quad \frac{f^1_{\mu_1}}{I_1} = f^1_{\mu_1},
\]

multiply and divide the first and third terms in the numerator of (5.41) by \( I_2 \) and \( I_1 \), respectively. Then,

\[
g^m(\mu_1 | \gamma_2) = \left\{ \left[ I_2 \pi^m f^1_{\mu_1} \right] \frac{f^1_{\mu_1}}{I_2} + \left[ \pi^m \pi^m f^2_{M_2} + \pi^m \pi^2 f^2_{M_2} \right] \right\} \frac{f^1_{\mu_1}}{I_1} + \\
\left\{ \left[ I_1 \pi^m f^2_{M_1} \right] \frac{f^2_{\mu_1}}{I_1} + \left[ \pi^m \pi^m f^2_{M_2} + \pi^m \pi^m f^2_{M_2} \right] \right\} \frac{f^2_{\mu_1}}{I_1}
\]

\[
\left\{ \left[ I_2 \pi^m f^1_{\mu_1} \right] + \left[ \pi^m \pi^m f^2_{M_2} \right] + \left[ I_1 \pi^m f^2_{M_1} \right] + \left[ \pi^m \pi^m f^2_{M_2} \right] \right\} . \tag{5.42}
\]

Using a by now familiar Raiffa and Schlaifer result,

\[ I_1 = \int_{-\infty}^{\infty} f^1_{\mu_1} f^2_{M_1} d\mu_1 = f_N(m|m_1', \frac{n + n_1'}{nn_1} \sigma^2) = f_{M_1} \]

and

\[ I_2 = \int_{-\infty}^{\infty} f^2_{\mu_1} f^2_{M_1} d\mu_1 = f_N(m|m_1, \frac{2n + n_1'}{n(n_1' + n)} \sigma^2) = f_{M_1}^2 . \]

Substituting these results in (5.42) and factoring \( \pi^m \) and \( \pi^m \) out of the second and fourth terms in the denominator yields:
\[ g''(\mu_1 | Y_2) = \left\{ \pi_1^{\pi_1 f_{M_1}^2} \right\}^{1,1}_{\mu_1} + \left\{ \pi_2^{\pi_2 f_{M_2}^2} \right\}^{1,2}_{\mu_1} + \left\{ \pi_1^{\pi_1 f_{M_1}^2} + \pi_2^{\pi_2 f_{M_2}^2} \right\}^{1,2}_{\mu_1} + \]

\[ \left\{ \pi_1^{\pi_1 f_{M_1}^2} \right\}^{2,1}_{\mu_1} + \left\{ \pi_2^{\pi_2 f_{M_2}^2} \right\}^{2,1}_{\mu_1} + \left\{ \pi_2^{\pi_2 f_{M_2}^2} \right\}^{2,2}_{\mu_1} \}

\[ \pi_1^{\pi_1 f_{M_1}^2} + \pi_2^{\pi_2 f_{M_2}^2} \}

\[ \left\{ \pi_1^{\pi_1 f_{M_1}^2} \right\} + \pi_2^{\pi_2 f_{M_2}^2} \}

(5.43)

Noticing that \( \mu_1 + \mu_2 = 1 \), the denominator of (5.43) simplifies:

\[ g''(\mu_1 | Y_2) = \left\{ \pi_1^{\pi_1 f_{M_1}^2} \right\}^{1,1}_{\mu_1} + \left\{ \pi_2^{\pi_2 f_{M_2}^2} \right\}^{1,2}_{\mu_1} + \]

\[ \pi_1^{\pi_1 f_{M_1}^2} + \pi_2^{\pi_2 f_{M_2}^2} \}

\[ \left\{ \pi_1^{\pi_1 f_{M_1}^2} \right\}^{2,1}_{\mu_1} + \left\{ \pi_2^{\pi_2 f_{M_2}^2} \right\}^{2,1}_{\mu_1} + \left\{ \pi_2^{\pi_2 f_{M_2}^2} \right\}^{2,2}_{\mu_1} \}

\[ \pi_1^{\pi_1 f_{M_1}^2} + \pi_2^{\pi_2 f_{M_2}^2} \}

\[ \left\{ \pi_1^{\pi_1 f_{M_1}^2} \right\} + \pi_2^{\pi_2 f_{M_2}^2} \}

\[ \left\{ \pi_2^{\pi_2 f_{M_2}^2} \right\} \}

(5.44)
\[
\pi_{2,1}^{2,1} f_{\mu_1}^{2,1} + \pi_{1,2}^{2,2} f_{\mu_2}^{2,2} + \pi_{2,2}^{2,2} f_{\mu_1}^{2,2} = \pi_{1,1}^{1,1} f_{\mu_1}^{1,1} + \pi_1 \left( \pi_{1,2}^{1,2} f_{\mu_1}^{1,2} + \pi_{2,2}^{2,2} f_{\mu_1}^{2,2} \right) + \pi_2 \left[ \pi_{1,2}^{1,2} f_{\mu_1}^{1,2} + \pi_{2,2}^{2,2} f_{\mu_1}^{2,2} \right].
\] (5.45)

(5.46)

where, for example,

\[
\pi_{1,1}^{1,1} = \frac{\pi_1 f_{M_1}^2}{\left[ \pi_1 f_{M_1}^2 \right] + \left[ \pi_2 f_{M_2}^2 \right] + \left[ \pi_2 f_{M_2}^2 \right] + \left[ \pi_2 f_{M_2}^2 \right]}.
\] (5.47)

Looking at the numerator of (5.47), \(\pi_1^{1,1}\) is the posterior probability that model 1 was in control during period one; \(\pi_1^{1,2}\) is the prior probability that model 1 is in control during period two; \(f_{M_2}^2\) is proportional to the likelihood of observing \(Y_2\) given that model 1 was in control during period one and that model 1 is in control during period two. Thus, \(\pi_{1,1}^{1,1}\) may be interpreted as the posterior probability that model 1 was in control during period one and that model 1 was in control during period two. \(\pi_{1,2}^{1,2}\) is the posterior form of \(\pi_1^{1,2}\).

To summarize, after the second control period we have the following results:

\[
g''(\mu_1 | Y_2) = \pi_{1,1}^{1,1} f_{\mu_1}^{1,1} + \pi_1 \left[ \pi_{1,2}^{1,2} f_{\mu_1}^{1,2} + \pi_{2,2}^{2,2} f_{\mu_1}^{2,2} \right].
\]
\[ g''(\mu_2 | Y_2) = \pi_2,2 f''_{\mu_2} + \pi_2,1 \pi_1,1 f''_{\mu_2} + \pi_1,2 f''_{\mu_2} + \pi_1,1 \pi_2,1 f''_{\mu_2} \]  

\[ (5.49) \]

Since \( f''_{\mu_1} \) and \( f''_{\mu_2} \), \( i = 1, 2 \), were obtained by the application of Bayes' rule to normal prior distributions and likelihood functions whose form is normal (see equations (5.41) and (5.42)), \( f''_{\mu_1} \) and \( f''_{\mu_2} \), \( i = 1, 2 \), are normal distributions. Thus, (5.48) and (5.49) are mixtures of normal distributions.

The form of the posterior distributions after period 1 (see equation (5.29)) and after period 2 (see equation (5.48)) indicate that \( g''(\mu_i | Y_n) \) will be a mixture of \( 2^n \) normal distributions. This is a consequence of the special nature of the likelihood function used in the BMSW revision process. Because of the likelihood function's dependence on the revision results of previous periods, it expands as \( n \) gets large. It can be seen by examining equations (5.25), (5.34) and (5.37) that when used for revising \( g''(\mu_i | Y_{n-1}) \), the form of the likelihood function is that of a mixture of \( 2^n + 1 \) normal distributions. With the likelihood function, and therefore the posterior distribution of \( \mu_i \), expanding as \( n \) increases, the revision procedure described above
becomes computationally unwieldy. Even when \( n = 2 \) the procedure is quite messy. Swamy and Mehta ran into a similar problem in their work. They noted that "...any reasonable prior distribution on \( \theta \) that we may have leads to integrals which cannot all be expressed in closed form and, as a result, the Bayesian argument is numerically most complex to execute."

While BMSW is certainly a feasible procedure for handling many model nonstationarity problems, its computational complexities make it extremely costly to implement. Thus, unless \( n \) is quite small, or the rewards from being able to utilize both subjective and sample information in analyzing "switching" data-generating processes are quite large, the BMSW approach is probably too costly to be of much practical value. BMSW does, however, serve to demonstrate the difficulties that can arise in modeling nonstationary processes via mixtures of posterior distributions.

---

Swamy and Mehta, pp. 594-595.
CHAPTER VI
CONCLUDING COMMENTS AND SUGGESTIONS FOR FURTHER RESEARCH

This dissertation has dealt with the treatment of uncertainty concerning the specification of data-generating processes in statistical decision problems. The point of view taken has been that a decision maker's model specification uncertainty is in itself information about the data-generating process and should be appropriately reflected in the decision-maker's decision analysis. Thus, in modelling data-generating processes, the emphasis should not be on methods, such as model selection procedures, which discard relevant information and mask model specification uncertainty, but rather on methods, such as the Bayesian Model Comparison procedure, which formally collect and display information about the data-generating process. The BMC procedure was shown to be an appropriate and natural way to formally extend the parametric analysis of statistical decision problems to include consideration of the possibly widely differing predictive and decision-making implications of alternative specifications of the data-generating process. Accordingly, this dissertation has advocated the use of the BMC procedure in decision-making problems in which model specification uncertainty is present.

In Chapter III, forecasting with and without formal regard for model specification uncertainty was examined via comparisons of BMC
forecasting with BMS forecasting and, indirectly, forecasts obtained using the max-$R^2$ model selection procedure. It was shown that as a result of not appropriately treating model specification uncertainty, BMS point forecasts may be misplaced, BMS credible interval forecasts may mis specify their reliability, and that the predictive distribution yielded by the BMS procedure misspecifies forecast-risk.

In Chapter IV, the BMC procedure was applied to single-period economic control problems. In particular, certainty-equivalent and optimal analytic solutions were found for simple two-model control problems in which control is cost-free and in which instrument-use cost functions are known.

In Chapter V, an approach to handling nonstationary data-generating processes was introduced. Called Bayesian Model Switching, the approach models the random variable upon which a decision hinges as being generated by different statistical models in different time periods with the switch between models described by a multinomial process. The BMSW methodology can be used to make inferential statements about nonstationary processes that have been identified as fitting the specifications of the BMSW model. Unfortunately, however, the procedure proves to be computationally unwieldy. Even for the straightforward case of two switching normal models, the probability revision involved becomes computationally burdensome after only two time periods.

The remainder of this chapter will be devoted to discussions of (1) problems encountered in the course of the research for this dissertation, (2) the shortcomings of the BMC procedure, and (3) possible
extensions to the work of this dissertation and suggestions for further research in the area of model specification uncertainty.

VI.1 Research Difficulties Encountered

The most significant difficulties encountered in the research for this dissertation arose in the application of BMC control to complex model spaces, and in the probability revision scheme of the Bayesian Model Switching methodology. These difficulties are discussed in this section.

As noted in Chapter IV, analytic BMC control solutions, whether certainty-equivalent or optimal, may not be easy to obtain. Difficulties arise when model spaces are utilized that include one or more models in which the target variable is characterized as being a function of two or more instruments. In such cases, BMC control solutions can only be obtained by solving simultaneously a set of possibly very complex equations. The reason for this can be seen by examining the general form of the BMC control problem for the following model space:

\[ M_1: y = \beta_1 X + \varepsilon \]  \hspace{1cm} (6.1)

\[ M_2: y = \beta_2 X + \beta_3 Z + \delta \]  \hspace{1cm} (6.2)

The optimal BMC control settings for \( X_F \) and \( Z_F \) would be obtained by solving:
\begin{equation}
\min_{D_F} \left[ P^\prime(M_1 \mid y, X) \int_{-\infty}^{\infty} L(y_F, y_F^\ast) f(y_F \mid M_1, y, X, X_F) dy_F \right. \\
+ \left. P^\prime(M_2 \mid y, X, Z) \int_{-\infty}^{\infty} L(y_F, y_F^\ast) f(y_F \mid M_2, y, X, X_F, Z, Z_F) dy_F \right].
\end{equation}

Refer to Chapter IV for definitions of the terms in (6.1), (6.2), and (6.3). Both terms in the sum of (6.3) depend on \( X_F \). Accordingly, the minimization problem cannot be separated into two independent minimization problems as was done repeatedly in Chapter IV, but must be solved by taking partial derivatives of the entire expression in brackets, setting the resulting derivatives equal to zero, and solving the resulting equations simultaneously for \( X_F \) and \( Z_F \). For the model space given above, this solution is not difficult to obtain. However, when interaction or other higher order terms are included in one or more of the models, the solution may be computationally quite complex, perhaps involving the finding of the roots of a higher order polynomial. Even when solutions to more complex control problems can be efficiently obtained via the BMC procedure, such solutions frequently result in awkward analytical expressions for the instruments, making their policy implications difficult to ascertain.

In Chapter V, it was explained that the computational complexities of the BMSW procedure are the result of the special nature of the likelihood function used in the BMSW probability revision process. The likelihood function is dependent on the previous period's revision results and, consequently, expands with each succeeding revision. This dependence occurs because the BMSW likelihood function is designed to
reflect the likelihood of all n periods of data observed to date
having been generated by any possible sequence of the switching models.
The need for such a complex likelihood function arises because of the
decision-maker's inability to identify which of the switching models
has generated the observed realizations of the data-generating process.
The computational inefficiencies of the BMSW procedure are, therefore,
a result of the way in which the BMSW methodology treats the identifi-
cation problem.

The next section discusses reasons for the limited applicability
of the BMC procedure and, in so doing, points out areas in which
further research is needed.

VI.2 Shortcomings of the BMC Procedure and Suggestions
for Further Research
Two of the major difficulties encountered in the application of
the BMC procedure are technical limitations which are common to Bayesian
inference in general. The first has to do with the assessment of prior
probability distributions; the second with the computational methods
required for performing probability revision.

If a decision maker has substantial prior information in the form
of his own judgments or experience, it is important that this informa-
tion be reflected in his prior probability distribution over the models,
P'(M_i) i=1,..., N, and his prior probability distribution over the
parameters, g'(θ_i|M_i) i=1,..., N. The assessment of judgmental priors,
however, for problems involving many parameters is a considerable task.
Even in situations in which the decision maker does not have prior
information, as Gaver and Geisel have pointed out, the choice of an
informationless prior may be quite difficult and the results may be
biased in favor of the model with the most parameters.\(^1\) Probability
assessment for problems of this nature appears to be an area deserving
substantial research.

In applying the BMC procedure, probability revision must be per-
formed many times. With each successive set of sample observations
obtained, revision is performed on each model's prior parameter distri-
bution and on each prior model probability. Currently, however,
computationally efficient methods of performing this revision exist
for only a relatively few classes of statistical models. Thus, in
order for the BMC procedure to become more generally applicable, work
is needed in expanding the classes of models which can be conveniently
handled by Bayesian inferential procedures.

The form of the predictive distribution yielded by the BMC pro-
cedure can also make the application of the procedure computationally
inconvenient. Recall from equation (2.15) that the BMMP is a mixture
-- a linear combination -- of single-model predictive distributions.
Accordingly, the BMMP is quite difficult to display graphically and
numerical methods may be required to find the probability of \(y_F\) taking
on a value in a particular region; i.e., to find \(P[a \leq y_F \leq b]\), where
a and b are constants. Thus, a significant amount of effort may be
required to find credible intervals and, depending on the decision-
maker's loss function, point estimates for \(y_F\).

\(^1\)Gaver and Geisel, pp. 62-72.
The BMC procedure is also limited in its applicability by the requirement that the decision maker must assign prior model probabilities as if he believes the correct model of the data-generating process is represented in his model space. No formal provision exists within the BMC procedure for the decision maker to express doubt concerning whether his model space contains the correct model. This then is a weakness in both the BMC and BMS procedures, for certainly most decision makers would admit to some doubt concerning the existence of the correct model in a particular model space no matter how extensive the model space. What is needed is a procedure for making inferences about a set of N models comprised of N-1 well-defined statistical models and one ill-defined dummy model representing the infinitely many models not explicitly included in the set of N-1 models. Until such a procedure is available, the meaningfulness of the prior model probabilities and, therefore, the results of the BMC procedure are dependent on the decision maker choosing his model space so as to make virtually insignificant his doubt concerning whether he has in fact captured the correct model of the data-generating process in his model space.

The next section offers suggestions for further research in the areas of economic control and model nonstationarity.

VI.3 Suggestions for Further Research in the Areas of Economic Control and Model Nonstationarity

Several important questions concerning BMC control solutions remain unanswered. These questions, along with suggestions for
further research in the area of model nonstationarity, are outlined in this section.

The details of the BMC approach to handling model specification uncertainty in economic control problems were presented in Chapter IV. Now what is needed are answers to the following questions:

(1) Is the BMC approach to control equivalent to the usual Bayesian approach in which all $N$ alternative models are nested in a single model?

(2) A popular classical approach to control involves including all instruments thought to be related to the target variable in a single model, estimating the parameters of that model, and testing hypotheses about the model parameters in order to establish a single model of the process whose control is desired. The question is: How does the BMC approach to control perform relative to this classical approach?

(3) Is there a problem in logically assessing prior model probabilities for, say, three models when the third model is a nest of the first two models?

If BMC control procedures compare favorably with other Bayesian and classical approaches, solutions should be derived and their policy-making implications explored for a variety of model spaces, i.e., for model spaces of differing dimensions containing models of differing functional form.

Concerning suggestions for further study of model nonstationarity
via the BMSW approach, the next step should involve either the development of models with simpler likelihood functions or a search for asymptotic properties of BMSW probability revision that would at least permit BMSW inferential statements to be approximated. Assuming the computational inefficiencies of the BMSW procedure can be overcome, other specific forms of nonstationarity (recall that in Chapter V, the switching models were controlled by a multinomial process) should be considered in the context of BMSW. For example, a Markov process might be utilized to control the switches between models, or perhaps the model switches might be controlled by a Bernoulli process with an exponential process dictating the time the switch is to occur. The development of BMSW methodology for Bernoulli and Markov switching control processes in which the transition probabilities are unknown would help add even more realism to the modeling of nonstationary processes. Whether or not the computational inefficiencies of the BMSW methodology can be overcome, it would seem sensible to approach the study of nonstationary data-generating processes by initially examining highly specific forms of nonstationarity as was done in Chapter V and suggested above.
BIBLIOGRAPHY

Books and Dissertations


Articles


Geisel, Martin S. "Bayesian Comparison of Simple Macroeconomic Models." Journal of Money, Credit, and Banking, 5 (October, 1973): 751-772.


BIOGRAPHICAL SKETCH

Paul George Benson was born in Montour Township, Pennsylvania on June 3, 1946 to Paul and Anna Louise Benson. He was graduated from Lewisburg Joint High School, Lewisburg, Pennsylvania in 1964 and from Bucknell University with a Bachelor of Science in Mathematics in June, 1968. While at Bucknell he was a member of Sigma Chi, president of Phi Eta Sigma, and upon graduation received the Dorothy Walls McCormack citizenship-leadership prize. He won a Rotary International Fellowship for a year of graduate study abroad which he was unable to accept due to the military draft. He is currently the President of the Class of 1968, having been originally elected in June, 1968, and re-elected in June, 1974.

From June, 1968, to June, 1969, Mr. Benson was employed by the U.S. Army Security Agency, Arlington Hall Station, Virginia as a management analyst in the area of manpower control. He was part of a manpower survey team responsible for determining the agency's manpower needs worldwide. From June, 1969, to September, 1971, Mr. Benson was employed as a computer program design engineer by Bell Telephone Laboratories, Whippany, New Jersey. He was involved with helping to maintain and develop the Laboratories' Data Management System. From June, 1970, to September, 1971, Mr. Benson was a Member of the Laboratories' Technical Staff. While at the Laboratories he was a full-time student for two semesters in the Department of Operations Research of New York University.
From September, 1971, through December, 1976, Mr. Benson was a full-time student in the Department of Management in the College of Business Administration at the University of Florida. He minored in statistics and economics. While at Florida he was President of Alpha Iota Delta and held an NDEA Title IV Fellowship and a fellowship from the Public Utilities Research Center at the University of Florida.

Since December, 1976, Mr. Benson has been an assistant professor in the Department of Management Sciences in the College of Business Administration at the University of Minnesota.
I certify that I have read this study and that in my opinion it
conforms to acceptable standards of scholarly presentation and is fully
adequate, in scope and quality, as a dissertation for the degree of
Doctor of Philosophy.

Christophe B. Barry, Chairman
Associate Professor of Management and Finance

I certify that I have read this study and that in my opinion it
conforms to acceptable standards of scholarly presentation and is fully
adequate, in scope and quality, as a dissertation for the degree of
Doctor of Philosophy.

Ira Horowitz
Professor of Management

I certify that I have read this study and that in my opinion it
conforms to acceptable standards of scholarly presentation and is fully
adequate, in scope and quality, as a dissertation for the degree of
Doctor of Philosophy.

Roger D. Blair
Associate Professor of Economics

I certify that I have read this study and that in my opinion it
conforms to acceptable standards of scholarly presentation and is fully
adequate, in scope and quality, as a dissertation for the degree of
Doctor of Philosophy.

H. Russell Fogler
Associate Professor of Management
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James T. McClave
Associate Professor of Statistics

This dissertation was submitted to the Graduate Faculty of the Department of Management in the College of Business Administration and to the Graduate Council, and was accepted as partial fulfillment for the requirements for the degree of Doctor of Philosophy.

August 1977

Dean, Graduate School