MONASH UNIVERSITY
THESIS ACCEPTED IN SATISFACTION OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
ON..... 1 March 2002 .....
.....[redacted]..
for Sec. Research Graduate School Committee

ERRATA

- P 99, Section 4.4.1, line 7 from bottom; replace "1986" with "1985".
- P 111, Section 4.5.2.8, paragraph 2, line 1; replace "δ" with "ϕ".
- P 242, line 10; replace "extend" with "extent".
- P 252, insert missing reference:

Engle, R.F., D.F. Hendry, and D. Trumble, 1985, Small sample properties of ARCH estimators and tests, *Canadian Journal of Economics* 18, 66-93.

Modelling and Forecasting in the Presence of Structural Change in the Linear Regression Model

A thesis submitted for the degree of
Doctor of Philosophy

by

Mohammad Nurul Azam
BSc.(Hons.), MSc.(Statistics) Jahangirnagar University
Dhaka, Bangladesh,
Master of Applied Statistics (MAS), Memorial University of
Newfoundland, Canada

Department of Econometrics and Business Statistics
Faculty of Business and Economics
Monash University
Australia
July 2001

Contents

Chapter 4 The Use of Model Selection for Detecting Unknown Changepoints

List of Tables

List of Figures

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except where reference is made in the text of the thesis.

[redacted]

Mohammad Nurul Azam

Acknowledgments

In the name of Allah, Most Gracious, Most Merciful.

This thesis would not have been completed without the guidance of my supervisor, Professor Maxwell L. King, to whom I express my deep debt of gratitude. Professor King suggested the topic and I was fortunate to have benefited greatly from the profound insight that Professor King so generously imparted. His constructive criticisms, continuous encouragement and affection at every stage of my work are highly appreciated. I also acknowledge his editorial assistance in the writing of this thesis. I would also like to express my sincere gratitude to Associate Professor Brett Inder for supervising me for the first six months of my candidature.

Many thanks are due to Mrs. Mary Englefield, Mrs. Philippa Geurens, Mrs. Julie Larsen, Mrs. Pauline Froggatt, Mr. Myles Tooher, Ms. Inge Meldgaard and Ms. Deanna Orchard for their administrative support in various ways that have helped me successfully complete this thesis.

I want to acknowledge with gratitude my debt to Monash University for awarding me a Monash Graduate Scholarship, Faculty Fee Paying Scholarship, and to the Department of Econometrics and Business Statistics for offering me a Departmental Scholarship and a fractional assistant lecturer's position.

Last, but not least, my main indebtedness and gratitude are to my wife, Mahbuba Yeasmin and my son Mim Mahbubul Azam, for all their love, sacrifice, patience, support and encouragement and for managing to survive on less than their reasonable share of attention during the entire course of my study.

Finally, I dedicate this thesis to the memory of my mother who died when I was in grade 5 and my father who died last year.

Abstract

The importance of testing for and modelling structural change of possible unknown timing in econometric relationships is well recognized. There is a well-developed literature on testing and modelling when a structural change may be present and a lesser literature on forecasting. Unfortunately, almost no work has been reported on model selection procedures for detecting the presence of a structural change with changepoint of unknown timing. In this thesis we investigate the use of model selection and develop a new model selection procedure which involves of maximizing the average mean probability of correct selection (AMPCS). New results are also presented for testing and forecasting with possible structural change in mind.

The first contribution of this thesis concerns the use of the likelihood ratio (LR) test statistic to test for the presence of structural change when there is a possible unknown changepoint in the data. Since this test does not have a known distribution for finite sample sizes, we calculate exact critical values for the test by simulation for different sample sizes, numbers of regressors and types of regressors. We find that the critical value clearly depends on sample size, number of regressors and to a lesser extend on the type of explanatory variables. We develop formulae for critical values using a response surface approach and check the accuracy of these formulae by Monte Carlo simulations. Overall the actual sizes of the test using our formulae to calculate appropriate critical values are quite satisfactory.

As a second contribution, this thesis demonstrates that model selection procedures can be used to detect a possible changepoint in the data. It cautions against the use of one particular information criteria (IC) procedure in order to detect

the presence of a structural change because none of the IC procedures stand out as a clear best method. In order to assess different selection strategies, we use our criterion AMPCS that summarizes the quality of different IC procedures. The results of the Monte Carlo experiment show that in terms of our AMPCS criterion, Hocking's S_p criterion (HSPC) is the best IC procedures for small samples and Schwarz's Bayesian information criterion (BIC) for large samples. Findings also show that BIC outperformed all existing IC procedures considered when there is no structural change, and Theil's R-squared criterion (TRSC) performed best overall when a changepoint is present.

The third contribution is to outline methods for finding optimal penalties for different changepoint models in such a way that no one model is favoured unwillingly. We propose four new methods which are the complete grid search algorithm (CGSA), a block grid search algorithm (BGSA), polynomials of degree four algorithm (PDFA) based on a grid search and a relatively new global optimization algorithm called the simulated annealing algorithm (SAA). Our simulation results show that the CGSA is the best, BGSA second, SAA third and PDFA fourth best as measured by maximum AMPCS. We have found that all of our four suggested procedures dominate the existing IC procedures considered in terms of having higher AMPCS.

Finally, we investigate random changes in the coefficients of linear regression models and their effect on prediction models. We derive the distributional pattern especially the mean, variance and covariance structure of different linear regression models for stochastic changes in either slope or intercept parameters in turn by a fixed amount with a very low probability. We find that this results in a linear regression with a nonscalar variance-covariance matrix, which allows standard approaches to estimation and prediction to be used.

Glossary

AIC	Akaike's information criterion.
AICc	Bias-corrected AIC.
AMPCS	Average mean probability of correct selection.
APC	Amemiya's prediction criterion.
ARCH	Autoregressive conditional heteroscdastic.
BA	Boltzmann annealing.
BGSA	Block grid search algorithm.
BIC	Schwarz's Bayesian information criterion.
CGSA	Complete grid search algorithm.
CL	Conditional likelihood.
CIC	Controlled information criterion.
CUSUM	Cumulated sums of recursive residuals.
CUSUMSQ	CUSUM square.
DW	Durbin-Watson.
FIC	IC based on the F distribution.
FPE	Final prediction error.
GARCH	Generalized autoregressive conditional heteroscdastic.
GCVC	Generalized cross-validation criterion.
GMM	Generalized method of moments.
GSA	Grid search algorithm.
HQC	Hannan and Quinn's criterion.
HSPC	Hocking's S_p criterion.
IC	Information criteria.
ICL	Improved conditional likelihood.
KL	Kullback-Leibler.
LM	Lagrange multiplier.
LR	Likelihood ratio.

MAE	Mean absolute error.
MCPC	Mallow's Cp criterion.
MDL	Minimum description length.
ME	Mean error.
ML	Maximum likelihood.
MML	Maximum marginal likelihood.
MOSUM	Moving sums of recursive residuals.
MPCS	Mean probability of correct selection.
MSE	Mean squared error.
OLS	Ordinary least squares.
OSAIC	One-sided AIC.
PDFA	Polynomial of degree four algorithm.
QFIC	IC based on the quasi F distribution.
RMSE	Root mean square error.
SAA	Simulated annealing algorithm.
SSE	Sum of squared residuals.
SSR	Regression sum of squares
TRSC	Theil's R-squared criterion.
W	Wald.

CHAPTER 1

Introduction

1.1 Background

Econometricians, economists, statisticians and researchers of a number of other disciplines use the term model to mean a simplifying approximation of the real data, which captures the relevant features of a particular phenomenon. Grassa (1989, p1) defined an econometric model as "an analytical characterisation of the joint probability distribution of some random variables of interest which yields some information on how the actual economy works". On the other hand, a mathematical economic model describes the behaviour of an economy within the framework of a set of assumptions. In other words, an economic model typically involves some degree of abstraction from reality whereas an econometric model reflects this abstraction in practice (see for example, Zarembka (1974)).

Bergstrom (1993) indicated that economic models are usually less precise than other models used for statistical fitting and testing. Econometric models are usually more precise in the sense that parameters of such models can be estimated and tested through statistical techniques using available data. In other words, we are able to see how well the model fits the data. Much of the literature in econometrics

is mainly concerned with the problems of estimation and inference from a sample of data. The properties of estimation techniques, and the quality of inferences, are heavily dependent on the correct specification of the model under consideration. There is the problem, that for one modelling situation, there might be many different specifications that constitute different alternative models. Thus an obvious question arises as to which model provides the best characterisation from the viewpoint of the data. The term model selection emerges from this simple idea.

Econometricians usually expect that economic theory will help them to find causal links and formulate appropriate models. But unfortunately, existing economic theory often fails to suggest an adequate functional form of such relationships. Because of this weakness in economic theory, econometricians often use their own methods for deciding on the functional forms of models. In doing so, they typically propose a range of alternative models to reflect the relationship between dependent and independent variables. The question arises; how should one model be selected from a number of alternative possible models using the available data? This is typically known as the model selection problem in the econometrics literature.

During the 1950's and 1960's, regression analysis became the principal tool for economic data analysis. It was not long until there was a concern about the assumption that parameters in the regression are constant over the entire sample period. It is quite common to observe occasional changes in economic systems, which alter the underlying relationships between variables of interest. As a result, these changes must be considered while forming the model. This realization has led to a large literature on the possibility, detection and modelling of structural change in econometric models.

Modern econometric practice advocates that while modelling any econometric problem; researchers should test their models for misspecification,

especially for structural change. The problem of testing for structural change basically involves testing for the consistency of regression coefficients in two or more separate subsamples. In the case of time series data, the subsamples may be different economic situations, such as particular government policy periods or particular foreign exchange regimes. In the case of cross sectional data, the subsamples may correspond to different groups of observations such as large and small firms, developed and underdeveloped countries or men and women. Many researchers have shown that proper consideration of possible structural change is needed while analysing data, otherwise poor estimates, inferences and forecasts can result.

Over the last few decades, there has been considerable interest in the problem of testing for structural change. Some of the literature covers the case when the timing of the changepoint is assumed known. This is not always a realistic assumption because for example in economics, some economic changes can take place when the timing of the changepoint is unknown. The difficulty with the problem of testing for a structural change with an unknown changepoint is that it does not fit into the traditional testing framework. The reason is that the timing of the changepoint appears only under the alternative hypothesis but not under the null hypothesis. As a result, standard tests such as the Wald (W), Likelihood ratio (LR) and Lagrange multiplier (LM) tests do not possess their usual large sample distributions under the null hypothesis. One of the main aims of this thesis is to provide a small sample based LR test procedure for structural change in the linear regression model when the changepoint is unknown.

In the literature there have been a number of model selection procedures suggested by researchers. These procedures can mainly be classified into four categories: (1) procedures based on hypothesis testing, (2) procedures based on

minimum sum of squared residuals, (3) procedures based on Bayesian criteria and (4) procedures based on information criteria (IC). However, if we look at the development of the model selection literature, we see there is a wide range of research related to IC based procedures. These particular procedures are probably the most workable, popular and widely used methods for model selection in econometrics.

Clayton et al. (1986) showed that IC based procedures can be regarded as a more substantial approach to model selection than any other procedures. In addition, Granger et al. (1995) noted that IC based procedures involve fewer limitations than hypothesis test based procedures, and hence have become more popular with practitioners. Thus, the focus in this thesis will be on IC based procedures which can be defined as choosing the model with the largest maximized log-likelihood function minus a penalty term, where the penalty is a function of the number of parameters included in the model and possibly also sample size.

One can easily design a new IC procedure by slightly changing the value of the penalty function. As such, interest in introducing various IC based procedures for different types of models continues to grow and that make the users confused as to which IC procedure to use for a particular problem in hand. Therefore, an IC based procedure that would work well for any kind of model selection problem is a current disparity in the IC literature.

King, Forbes and Morgan (1995) and Forbes, King and Morgan (1995) proposed a new approach for estimating penalties through simulation. This method is known as the controlled probability approach. Hossain and King (1998) applied this approach to Box-Cox transformation models and found that it produces high selection rates in picking the true (data generating) model. Kwek and King (1997a, 1997b, 1998), and King and Bose (2000) considered model selection problems in

conditional heteroscedastic models, and linear regression models, respectively, in the context of maximizing the average mean probability of correct selection (AMPCS). The AMPCS is calculated by averaging the mean probabilities of correct selection for all models in the plausible group. All of these applications produced, on average, a high probability of selecting the true model. For the sample size and plausible models under consideration, this new model selection approach maximizes the AMPCS through the estimation of penalty values numerically.

This has motivated us to develop a model selection approach for the detection of the possible presence of structural change in the linear regression model. In this thesis we discuss in detail the idea of model selection for this problem both using existing IC procedures and the new approach of maximizing the AMPCS.

Finding penalty values that maximize the AMPCS is a difficult numerical problem. The AMPCS is a step function, and hence, it may not be easy to maximize using standard methods. The grid search algorithm (GSA) could be one way of estimating penalty values so that the AMPCS is maximized. Because we are dealing with so many models based on the position of the changepoint, to ease the computational burden, we investigate the use of a block grid search algorithm (BGSA) and a polynomial of degree four algorithm (PDFA) based on grid search. We also use a relatively new global optimization algorithm called the simulated annealing algorithm (SAA) to maximize this AMPCS. The SAA works well as an optimization algorithm, even when optimizing very complicated functions such as functions with a large number of local maxima (see Corana et al. (1987), Kirkpatrick et al. (1983) and Goffe et al. (1994)). A contribution of this thesis is to investigate the use of the BGSA, PDFA and SAA to find optimal penalties for a small or large number of models when the data contain a possible structural change.

In making policy decisions, forecasting is becoming an increasingly important issue both in the regulation of developed economies as well as for the planning of the economic development of the underdeveloped countries. Goldfeld (1976) and Zellner (1979) showed that forecasting without taking possible structural change into account may provide misleading or poor forecasts. Also, a test procedure that does not account for the possible presence of structural change may reject some well-established econometric theories (e.g., the Lucas hypothesis), see Ilmakunnas and Tsurumi (1985). In this thesis, we also discuss in detail the idea of forecasting in presence of possible stochastic change of unknown timing in the parameters with a low probability.

In summary, the main aims of this thesis are to study the problems of testing for a structural change in the linear regression model with an unknown changepoint, model selection in the presence of possible structural change where there is a large number of models each based on the position of the changepoint in the data, finding optimal penalties for such model selection problems and forecasting in the presence of structural change.

The specific objectives of this thesis are to:

(i) Develop the ability to calculate critical values for the LR test for structural change of unknown timing by developing formulae using the response surface method. We also check the usefulness of the proposed formulae by conducting a small Monte Carlo experiment.

(ii) Investigate the use of IC model selection procedures to detect a structural change when there is large number of models each with a different timing for the changepoint. We also examine which criterion among existing IC has the best ability to detect a

changepoint in the context of a linear regression model when the timing of the changepoint is unknown.

(iii) Consider the use of AMPCS as a method of obtaining optimal penalties. We also examine how such optimal penalties might be calculated in practice, particularly when there are a large number of alternative models involved.

(iv) Investigate how to incorporate possible future structural change as a stochastic element in linear models with possible stochastic changes in their parameters and to compare predictions from different strategies for such models.

1.2 Outline of the Thesis

In Chapter 2, we briefly review the testing for structural change literature from an econometric and statistical viewpoint covering likelihood ratio tests, Bayesian procedures, nonparametric approaches, CUSUM and CUSUM of squares tests, the Chow test, the sup F test for time series and linear regression models.

We briefly survey the model selection literature in econometrics beginning with a short discussion on some obvious demerits involved with model selection through hypothesis testing. The survey reveals that there is a large body of literature on model selection ranging from stepwise hypothesis testing to IC based model selection procedures. However, the discussion on IC based procedures mainly focuses on the penalty term that is one of the main ingredients of such procedures. We also briefly survey the literature on the simulated annealing algorithm and the varying coefficient model.

Chapter 3 develops a small-sample test procedure that allows the use of the LR test statistic based on maximization of the likelihood function of the linear

regression model in the presence of structural change. Under the null hypothesis, the regression parameters are constant across all periods. Under the alternative, a particular regression parameter changes at an unknown changepoint. In our case, the critical values of the LR test statistic depend on the number of regressors, types of regressors and the sample size. Unfortunately, the LR test statistic does not have a known finite sample distribution, although the critical values for the test can be calculated by Monte Carlo estimation. We develop formulae for critical values of the LR test for different sample sizes, different significance levels, number of regressors in the model and types of regressors using a response surface approach applied to estimated critical values obtained via simulation.

In Chapter 4, we argue that the problem of detecting a changepoint of unknown timing can be viewed as a model selection problem. In particular, we examine which criterion has the best ability to detect a changepoint in the context of a linear regression model when the timing of the changepoint is unknown. We use the average mean probability of correct selection (AMPCS) criterion as a measure of accuracy in detecting a changepoint.

In Chapter 5 we develop algorithms that compute optimal penalties in such a way that the AMPCS is maximized for different models involving structural change. We use grid search, polynomial of degree four combined with grid search, and simulated annealing optimization algorithms that estimate optimal penalties for different models. We give a working version of each of the algorithms and evaluate them by discussing their advantages and disadvantages. We perform a small Monte Carlo experiment to calculate the penalties for different models using these algorithms. We look for the algorithm that gives the optimal penalties in a sense that these penalties will provide maximum AMPCS with minimum computational cost and effort.

In Chapter 6, we investigate how to incorporate possible future structural change as a stochastic element in different linear models with stochastic changes in parameters. We consider the case in which there are two possibilities of a changing parameter. One is a change in the slope parameter by a fixed amount with a very low probability, keeping the intercept constant and another is a change in the intercept parameter by a fixed amount with a very low probability, keeping the slope unchanged. We use linear regression models with single or multiple time changing coefficients but with low probabilities of a change at any point in time. We find that this results in a linear regression with a nonscalar variance-covariance matrix, which allows standard approaches to estimation and prediction to be used. We perform a Monte Carlo experiment to investigate whether our forecast procedures are likely to be useful.

We conclude this thesis with Chapter 7. It summarises the results, conclusions and contributions of this thesis, and gives some suggestions for future research.

CHAPTER 2

Literature Review

2.1 Introduction

There are four main themes in this thesis. These are testing for structural change when the timing of the changepoint is unknown, model selection as a method of detecting a possible changepoint, the use of the special optimization procedures to find optimal penalties for model selection and prediction in the linear regression model in presence of possible random change of unknown timing in the parameters. The purpose of this chapter is to review the literature relevant to these four topics with particular emphasis on the problem of structural change when the changepoint is unknown.

The structural change problem can be considered to be one of the central problems of statistical inference, linking together the theory of estimation and testing using classical and Bayesian approaches, and fixed sample and sequential procedures. We survey the literature on testing for a structural change when the changepoint is unknown and discuss the developments of related topics in the structural change literature.

In applied work, model selection is a frequently occurring problem of great importance, as inferences, forecasts, interpretations, and policy decisions etc. can depend critically on the particular model selected from the range of models examined. Most often, model selection is done by mechanical application of one or several of the criteria that have been developed for this purpose. Among these, we review some of the commonly used IC procedures.

The maximised likelihood and the penalty function are the two main ingredients of IC based model selection procedures which involve choosing the model with the largest maximised log-likelihood function minus a penalty term, that is, the largest penalized maximum likelihood. The penalty term is a function of the number of parameters included in the model under consideration and typically also the sample size. Unfortunately, there is little agreement about the best form of this penalty function. To find the penalties for different models that maximize the average mean probability of correct selection, a global optimization method needs to be used. One relatively new optimization method that we will find useful in later chapters, is a method called the simulated annealing algorithm (SAA). We also briefly review this topic in this chapter.

We begin the survey in Section 2.2 with a brief discussion of the structural change literature in econometrics and statistics with an emphasis on hypothesis testing. In Subsection 2.2.2 we survey the likelihood ratio approach and diagnostic tests for structural change are outlined in Subsection 2.2.3. Subsection 2.2.4 summarizes the Chow test and Subsection 2.2.5 the Sup F test. Subsection 2.2.6 discusses some other approaches.

In Section 2.3 we survey the literature on information criterion based model selection. In Subsection 2.3.2 we discuss the consequences of model selection through testing. In Subsection 2.3.3 we survey the historical development of some

existing IC based model selection procedures namely, Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (BIC), generalized cross validation criterion (GCVC), Hannan and Quinn's criterion (HQC), Theil's adjusted R-square criterion (TRSC), Mallow's C_p criterion (MCPC), Hocking's S_p criterion (HSPC), and Amemiya's prediction criterion (APC). We end this section discussing the work of King and others. In Section 2.4 we review simulated annealing algorithms. Finally, some concluding remarks are made in the Section 2.5.

2.2 Brief Review of Testing for Structural Change

2.2.1 Introduction

The problem of possible structural change in economic relationships has a long history in econometrics. Parameter constancy is especially important when one wants to use a model for forecasting and policy implementation. Generally, it is essential in policy analysis that the parameters of the model be invariant with respect to the possible policy intervention if the effect of such a policy change is to be predictable.

The aim of this section is to provide a literature survey of various test statistics for structural change. Since the literature in this area is vast and growing rapidly we will in particular focus on those tests which are widely used in practice.

2.2.2 Likelihood Ratio Approach

The likelihood ratio test has practical and theoretical importance, because it provides a unified approach to the problem of testing for structural change. In this survey we review some tests which can be applied to hypotheses about regression

coefficients containing possible structural change of a known or unknown changepoint.

Consider the standard linear regression model

$$y_t = x_t'\beta_t + u_t, \qquad t = 1,\cdots,n, \tag{2.1}$$

where y_t is the dependent variable at time t, x_t is a $k \times 1$ vector of observations of the independent variables at time t and β_t is a $k \times 1$ vector of unknown parameters of the model that may change over time. The error term u_t is assumed to be independently normally distributed with constant variance, i.e., $u_t \sim IN(0,\sigma^2)$.

The null hypothesis of interest is

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{n_1} = \beta_{n_1+1} = \cdots = \beta_n \tag{2.2}$$

where n_1 is the time of the changepoint.

2.2.2.1 Testing when the Alternative is Specified

One particularly important alternative hypothesis is the one time discrete change, which can be expressed as

$$H_1 : \beta_1 = \beta_2 = \cdots = \beta_{n_1} \neq \beta_{n_1+1} = \cdots = \beta_n \tag{2.3}$$

where n_1 is the time of the changepoint.

Quandt (1958) derived a method of estimating the changepoint of a linear regression model where there are two regimes and when it is known that time period under consideration contains one changepoint. He considered the problem under the assumptions that the data are free of error, error terms are independent of each other and the errors are independent of explanatory variables. He suggested a LR test of H_0 against a more general alternative, where the variance is also allowed to change. When the sample size is finite, n_1 is discrete and takes values from $n_1 = (k+1)/n$ to

$n_1 = (n-k)/n$. For such n_1 values, suppose the observations in the time segment $(1,\cdots,n_1,n_1+1,\cdots,n)$ come from two different regressions, then the likelihood ratio is

$$LR = 2\log\frac{\text{max likelihood under } H_1}{\text{max likelihood under } H_0} = 2\log\frac{\hat{\sigma}_1^{n_1}\hat{\sigma}_2^{n-n_1}}{\hat{\sigma}^n} \tag{2.4}$$

where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are the standard errors of estimate of the regressions before and after the changepoint respectively, and $\hat{\sigma}$ is the standard error of estimate of the overall regression based on all n observation. For this reason, expression (2.4) is often called a Quandt ratio, and the resulting test is referred to as the Quandt test. This test requires re-estimating the model for each subsample. Therefore, a natural answer to such a testing problem is to calculate the LR test statistic at every possible changepoint, then examine the largest test statistic. This may be quite time consuming or difficult in the case of non-linear regression models.

The estimated n_1 corresponds to the value of n_1 at which LR attains its maximum. Unfortunately, the likelihood function here is not differentiable with respect to the parameter n_1 and the distribution of maximum LR is unknown. In other words, the problem of testing for a one-time structural change with unknown changepoint does not fit into the standard testing framework, see Davies (1977, 1987). The reason is that the parameter n_1 only appears under the alternative hypothesis and not under the null hypothesis.

Quandt (1960) tested the conjecture that the distribution of maximum LR is asymptotically chi-squared. Based on Monte Carlo simulation results, he rejected it and concluded that the proposed approximation to the null distribution of the LR test is very poor, thus the LR procedure suffers from a lack of knowledge of the appropriate distributional theory. Such a problem is only partly overcome by either

using an approximation method or constraining to some special data generating process.

Farley and Hinch (1970) and Farley, Hinch and McGuire (1975), partitioned the unknown parameter vector of (2.1) $\beta_t' = (\beta_1', \beta_0')$, where β_1 is a $(k_1 \times 1)$ vector with $k_1 \leq k$, and β_0 is a $((k-k_1)\times 1)$ vector, they approximated the discrete change in the coefficient at an unknown point by a linear continuous change

$$\beta_1 = \beta_0 + \delta t \tag{2.5}$$

where δ is a scalar. Then under the alternative hypothesis, we have an augmented model, which is still linear in the parameters

$$Y = X\beta + Z\delta + u \tag{2.6}$$

where Z is an $n \times k_1$ matrix. Their procedure rejects H_0 of no structural change whenever

$$R = \frac{(SSE0 - SSE)/k}{SSE0/(n-2k)} \tag{2.7}$$

is significantly different from zero. Under the null hypothesis, the statistic $\left((n-2k)/k\right)R$ has a conventional F distribution with $(k_1, n-2k_1)$ degrees of freedom. Here $SSE0$ is the sum of squared residuals in (2.6) and SSE is the sum of squared residuals in (2.6) subject to the constraint $\delta = 0$. They also proved that their test is asymptotically more powerful than the mid-point Chow test if the structural change occurs in the intervals $0 < \frac{n_1}{n} < 0.42$ or $0.58 < \frac{n_1}{n} < 1$.

Hinkley (1969) derived the asymptotic distribution of the maximum likelihood estimate of the changepoint $\hat{n}_1$. He discussed computational aspects of the asymptotic distribution of the likelihood ratio statistic and compared the asymptotic results with some empirical finite sample results. He concluded that the asymptotic distributions are poor approximations in small samples.

Hinkley and Hinkley (1970) considered models of the form (2.1) and showed that the maximum likelihood estimate of the changepoint converges in distribution based on an infinite sample. They considered a sequence of random variables at which the probability distribution changes and derived the asymptotic distribution of the likelihood ratio statistic based on random walk techniques for testing hypotheses about the changepoint and compared it with finite sample empirical distributions. They used the maximum likelihood method to estimate n_1 and derived exact and asymptotic distributions of $\hat{n}_1$. They used the LR test to make inferences about the changepoint and concluded that the asymptotic distribution of the test statistic is not consistent.

Hawkins (1977) studied the likelihood ratio test for the alternative of a location change, found its distribution under the null hypothesis, and gave tables of standard percentiles along with asymptotic results. Worsley (1979) used likelihood ratio test statistics for location of the changepoint of normal population. Brown et al. (1975) suggested a way to test whether regression coefficients changed without specifying a changepoint.

Worsley (1983) gave an iterative procedure to determine the exact null and alternative distributions of the likelihood ratio statistics in the context of a change in a binomial probability. He provided approximate upper bounds for the p-value of the LR test and his numerical results indicate that such bounds are reasonably good in small samples.

Srivastava and Worsley (1986) used the LR statistic to test the null hypothesis H_0 of no change against the alternative there is a change. The unknown changepoint n_1 is estimated by the maximum likelihood method. They showed that the LR test statistic is a maximum Hotelling T^2 statistic, and the estimate of the change point is the point at which this is a maximum. Their main result is a conservative

approximation for Hotelling T^2 statistic's null distribution. They used the binary segmentation procedure by Vostrikova (1981) for detecting more than one changepoint. The disadvantage of this procedure is that it relies on the assumption of equal variance throughout the whole data period and also depends on the order in which data were split.

James, James and Siegmund (1987) considered likelihood ratio tests to detect a changepoint. Instead of directly addressing the problem under general linear regression model, they considered the simple linear regression

$$y_t = \begin{cases} \alpha_0 + \beta_0 x_t + u_t & \text{for } t \le n_1 \\ \alpha_1 + \beta_1 x_t + u_t & \text{for } t > n_1 \end{cases} \tag{2.8}$$

to test $H_0: \beta_0 = \beta_1$, and $\alpha_0 = \alpha_1$, against one of the alternatives, $H_1: \beta_0 = \beta_1 = \beta$, and there exists an n_1 $(1 \le n_1 \le n)$ such that $\alpha_0 \ne \alpha_1$ or there exists an unknown changepoint n_1 $(1 \le n_1 \le n)$ such that $H_2: \beta_0 \ne \beta_1$, or $\alpha_0 \ne \alpha_1$. They investigated some tests, including the LR test, for a sequence of independent normal random variables with constant, known or unknown variance for no change versus the alternative of a single changepoint.

Andrews and Fair (1988) extended Chow's (1960) classical test for structural change in linear regression models to a variety of nonlinear models, estimated by a variety of different procedures. They introduced Wald, Lagrange multiplier-like, and likelihood ratio-like test statistics, and provided a compact presentation of general unifying results for estimation and testing in nonlinear parametric econometric models.

Kim and Siegmund (1989) derived analytical likelihood ratio tests of constancy of a regression model over time and obtained an approximate p-value under reasonably general assumptions about the empirical distribution of the independent variable. A difficulty associated with this problem is that under the null

hypothesis, the distribution (even the asymptotic distribution) of the test statistic depends on the values of the independent variable. They estimated approximations to the p-values of the likelihood ratio tests of $H_0:\beta_0 = \beta_1$ against $H_1:\beta_0 = \beta_1 = \beta$ or $H_2:\beta_0 \neq \beta_1$.

Nyblom (1989) proposed tests for detecting possible changes in parameters when the observations are obtained sequentially in time. He mentioned when the starting point is unknown, an efficient estimate is substituted for it. In addition, he established the corresponding limiting distribution. The proposed tests turn out to be based on cumulative sums of the score function (the derivative of the log-likelihood).

Henderson (1990) considered testing the null hypothesis of no change, against the alternative of change at an unknown changepoint. Under the null hypothesis H_0, the variables are identically distributed, but under the alternative hypothesis H_1 there is a change in distribution at some unknown point n_1 in the sequence $(1 \leq n_1 \leq n)$. That is, the first n_1 observations are drawn from one distribution and the remaining $(n - n_1)$ are drawn from a different distribution. Thus H_1, contains a family of alternatives indexed by a parameter n_1 that disappears under the null hypothesis. He considered a likelihood ratio test and mentioned that additional information in the maximum likelihood estimate of the changepoint can seriously affect the interpretation of test results. He considered some modifications, derived exact percentage points and performed Monte Carlo power and mean squared error comparisons. He concluded that his results were encouraging.

Loader (1992) considered a changepoint model of the form (2.1) and used likelihood ratio tests for testing for the presence of a changepoint, for which standard asymptotic theory is not valid. He developed the log-likelihood ratio statistic for testing H_0 and showed that under the null hypothesis, the log-likelihood ratio

statistic is distributed approximately $\chi_1^2/2$. He mentioned that for small δ this approximation is not very good without giving a proof. Applying large deviation methods, he approximated the p-values, and gave power approximations. He also derived confidence regions for the changepoint and illustrated the methodology using a British coal mining accident data set.

Muller (1992) proposed estimators for location and size of a changepoint in smooth regression model. The assumptions he made are much weaker than those made in parametric models. His estimators apply to the detection of changepoints of slope and of higher order curvature based on a comparison of left and right one-sided kernel smoothers. He illustrated the methods by means of the well-known data on the annual flow volume of the Nile River between 1871 and 1970.

Andrews (1993) considered tests for parameter instability and structural change with an unknown changepoint for a wide class of parametric models estimated by generalized method of moments (GMM). He considered test statistics of the form

$$\sup_{n_1 \in \Pi} \text{Wald}(n_1),\ \sup_{n_1 \in \Pi} \text{LM}(n_1) \text{ and } \sup_{n_1 \in \Pi} \text{LR}(n_1) \tag{2.9}$$

where $\text{Wald}(n_1)$, $\text{LM}(n_1)$ and $\text{LR}(n_1)$ are the Wald, Lagrange multiplier, and likelihood ratio test statistics for testing H_0 versus H_1 and Π is a subset of the integers $\{1,\cdots,n\}$. The LR test statistic for the case of a specified parameter n_1 within the parameter space Π. Andrews justified the test statistics of the form (2.9) as the test statistics $\sup_{n_1 \in \Pi} \text{Wald}(n_1)$, $\sup_{n_1 \in \Pi} \text{LM}(n_1)$ and $\sup_{n_1 \in \Pi} \text{LR}(n_1)$ correspond to the tests derived from Roy's type I principle, see Roy (1953) and Roy, Gnanadesikan and Srivastava (1971, pp 36-46).

Andrews under some regularity conditions, proved that under H_0

$$\sup_{n_1 \in \Pi} K(n_1) \stackrel{d}{\Rightarrow} \sup_{n_1 \in \Pi} B(n_1)' B(n_1) / \left(n_1(n-n_1)\right) \tag{2.10}$$

is same as Wald, LM or LR, where K is any of these test statistic, $\stackrel{d}{\Rightarrow}$ means convergence in distribution, $B(n_1) = W(n_1) - n_1 W(1)$ is a vector tied-down Brownian motion on (0,1). In other words, the $\sup_{n_1 \in \Pi} \text{Wald}(n_1)$, $\sup_{n_1 \in \Pi} \text{LM}(n_1)$ and $\sup_{n_1 \in \Pi} \text{LR}(n_1)$ tests have the same asymptotic distribution which is free of nuisance parameters. In particular, in the case of a full structural change,

$$\sup F = \sup_{n_1 \in \Pi} F(n_1) \stackrel{d}{\Rightarrow} \sup_{n_1 \in \Pi} B(n_1)' B(n_1) / \left(n_1(n-n_1)\right) \tag{2.11}$$

where $F(n_1)$ is the conventional F test statistic which we discuss further in Section 2.2.5.

He noted that the asymptotic distributions of the test statistics are nonstandard because the changepoint parameter only appears under the alternative hypothesis and not under the null. He showed that the asymptotic null distributions of his test statistics are the supremum of the square of a standardized tied-down Bessel-process. This allowed him to provide tables of critical values based on this asymptotic null distribution. He found the tests performed quite well in a Monte Carlo experiment.

Kim and Cai (1993) examined the distributional robustness of the likelihood ratio test for a changepoint in a simple linear regression of type (2.8). They checked whether the level and power of the test remain unchanged when the underlying error distribution is nonnormal. They summarized the normal theory of the likelihood ratio tests for no change in the regression coefficients versus the alternatives with a change in the intercept alone and with a change in the intercept and slope, then discussed the robustness of these tests. Using the convergence theory of stochastic

processes, they showed that the test statistics converge to the same limiting distributions regardless of the underlying error distribution. They performed simulations to assess the distributional insensitivity of the test statistics to a Weibull, a lognormal, and a contaminated normal distribution in two different cases: fixed and random independent variables. Their numerical examples showed that the test has correct size and retains its power when the distribution is nonnormal. They performed simulation experiments and observed that the LR tests achieved almost the same level, and power regardless of the distributional assumption. They also found that the approximations for the p-values of the test achieved almost the same accuracy for each of the selected nonnormal errors.

Other related work includes Miller and Siegmund (1982) who considered a special case of maximally selected chi-square statistics. Vostrikova (1983) established the weak convergence of the likelihood ratio statistics in a very general case. Bhattacharya (1987) studied the problem of estimation of a changepoint in a general multiparameter case. Haccou, Meelis and van de Geer (1988) obtained the limit distribution of the maximally selected likelihood ratio in the exponential case and they showed that the test is optimal. Gombay and Horvath (1990) studied the changepoint problem when the observations are from a one-parameter exponential family. Yao (1993a,b) obtained approximations for a modified likelihood ratio test in the normal case. The result in Horvath and Serbinowska (1995) covers the case when the maximum is taken with respect to all possible changepoints. Rukhin (1994) and Hu and Rukhin (1995) computed the asymptotic minimaxity of the LR test in the context of changepoint problem. Baron and Rukhin (1997) constructed confidence regions for the estimated changepoint.

There have also been extensive studies on the analysis of structural change in a variety of models such as dynamic, nonlinear, simultaneous and time-series

models. For example, see Bacon and Watts (1971), Feder (1975), Deshaves and Picard (1982), Lo and Newey (1985), Sen (1985), Huskova (1988a,b), Kramer, Ploberger and Alt (1988), Lutkepohl (1988, 1989), Miao (1988), Antoch and Huskova (1989), Ploberger, Kramer and Kontrus (1989), Huskova (1990a,b, 1991), Mills (1992), Antoch and Huskova (1993), Huskova (1994a,b), Chu, Hornik and Kuan (1995), Watson (1995), Huskova (1996) and Horvath, Huskova and Serbinowska (1997) for details.

2.2.2.2 Testing when Alternative is Random Coefficient

One of the important specifications of β_t in context of one discrete changepoint of (2.1) under the alternative is to treat its variation as random. In this situation a test for parameter constancy reduces to a zero restriction on the variance of the innovations moving the random parameters. The reason behind this specification is that if the regression coefficients are to be regarded as the true partial derivatives of the dependent variable with respect to the independent variables, then it is unlikely that these partial derivatives will be identical for two different observations (Rosenberg (1973)). A number of varying parameter models have been proposed in the context of time series models. Among them, three major types of varying parameter models are of particular interest. They are, respectively, the random coefficient models of Hildreth-Houck (1968), the random walk models and the return to normalcy models of Rosenberg (1973). A detailed survey of testing varying coefficient regression models can be found in Brooks and King (1994).

The linear regression model (2.1) can be written under a specification of one random coefficient as

$$y_t = x_t\beta_t + z_t'\alpha + \varepsilon_t, \tag{2.12}$$

in which y_t is the dependent variable, x_t is the non-stochastic regressor with the single varying coefficient β_t, z_t is a $k \times 1$ vector of non-stochastic explanatory variables with fixed coefficient vector α, $\varepsilon_t \sim IN(0,\sigma^2)$ and $t = 1, 2, ..., n$. The null hypothesis is (2.2).

The Hildreth and Houck (1968) random coefficient model states that the single varying coefficient β_t from (2.12) follows the process,

$$\beta_t = \bar{\beta} + u_t, \tag{2.13}$$

in which $\bar{\beta}$ is a constant parameter, $u_t \sim IN(0,\lambda_0\sigma^2)$ and is independent of ε_t. Testing for regression coefficient stability is equivalent to testing

$$H_0: \lambda_0 = 0 \text{ against } H_1: \lambda_0 > 0.$$

If β_t follows this process then by substitution, model (2.12) becomes,

$$y_t = x_t\bar{\beta} + z_t'\alpha + v_t, \tag{2.14}$$

in which $v_t = \varepsilon_t + x_t u_t$. The properties of v_t are that it is normally distributed with,

$$E(v_t) = 0,\ Var(v_t) = \sigma^2(1 + \lambda_0 x_t^2) \text{ and } Cov(v_t, v_s) = 0, \quad t \neq s.$$

Therefore, the effect of introducing random coefficient variation is to give the dependent variable a different variance at each observation, the testing problem can thus be considered as testing for heteroscedasticity in the standard linear regression model. A large variety of test statistics have been developed in the literature. Among them, the LM test of Breusch and Pagan (1979) and the point optimal test of Evans and King (1985, 1988) appear to be most appropriate.

Rosenberg's (1973) return to normalcy random coefficient model assumes the coefficient following a first-order stationary AR process,

$$\beta_t = \phi\,\beta_{t-1} + (1-\phi)\bar{\beta} + a_t, \tag{2.15}$$

in which $a_t \sim IN(0,\lambda_1\sigma^2)$ and is independent of ε_t. For (2.15) to be a stationary process, it is required that $|\phi|<1$. It incorporates some of the best features of the random walk and random coefficient models and is therefore of great importance.

Watson and Engle (1985) pointed out that the standard tests, (e.g. LR, Wald or LM tests) cannot be used in the regular approach since the transition parameter ϕ is only identified under the alternative hypothesis but not under the null hypothesis. This is because under the null hypothesis the information matrix will be singular. To overcome this difficulty they applied the testing approach as suggested by Davies (1977).

Nyblom (1989) introduced the martingale formulation for the general problem of testing for structural change. Under the martingale specification, a test for constant coefficients reduces to a zero restriction on the variance of the innovations disturbing the random parameters. Such an approach possesses substantial flexibility. It allows, for example, β_t to be a random walk or to have a specified or unknown number of discrete jumps during the observation period. Based on the martingale specification, it is therefore possible to develop a test, which is sensitive to different types of nonconstancy of parameters (see Nyblom (1989) for details). Using a Taylor's series expansion, Nyblom first formed the approximation of the joint density function of the observations under the alternative. Then, by evaluating the ratio of the joint density functions respectively under null and alternative hypotheses, he achieved a locally most powerful test statistic.

2.2.3 Diagnostic Tests for Structural Change

In the previous section we discuss the test procedures which are intended to test against a specified form of structural change, there is also a need for an explanatory procedure aimed at being sensitive to a wide variety of non-stability patterns without specifying any particular alternative. It is possible to detect possible structural change by defining test statistics on the basis of various types of residuals and inspecting their probability distributions. Belsley, Kuh and Welsch (1980) called this approach an analysis of residuals and Box and Jenkins (1970) called it diagnostic checking.

Brown, Durbin and Evans (1975) introduced CUSUM and CUSUM of squares tests based on recursive residuals to test the stability of regression coefficients over time, they have become the standard diagnostic procedures when the timing and type of structural change are unknown. They considered the simple linear regression model of type (2.1). The null hypothesis of interest is $H_0:\beta_1=\cdots=\beta_n=\beta$. Tests are constructed on the basis of recursive residuals. The recursive residuals, are defined as

$$W_{n_1}=(y_{n_1}-x_{n_1}^{'}\hat{\beta}_{n_1-1})/\sqrt{1+x_{n_1}(x_{n_1-1}^{\prime}x_{n_1-1})^{-1}x_{n_1}^{\prime}},\ n_1=k+1,\cdots,n \qquad (2.16)$$

where $\hat{\beta}_{n_1-1}$ is the OLS estimate of β from the first n_1-1 observations and the CUSUM test statistic is defined as

$$CUSUM=\frac{1}{\hat{\sigma}}\max_{k+1\le n_1\le n}\left|\frac{W_{n_1}}{\sqrt{n-k}}\right|/\left(1+2\frac{n_1-k}{n-k}\right), \qquad (2.17)$$

where $\hat{\sigma}^2=\frac{1}{n-k}\sum_{i=1}^{n}\left(y_i-x_i^{'}\hat{\beta}\right)^2$ is a consistent estimate for σ^2. The CUSUM of squares test statistic is of the form

$$CUSUMSQ = \max_{k+1 \le n_1 \le n} \left| V_{n_1} \right| - \frac{n_1 - k}{n - k}, \tag{2.18}$$

where $V_{n_1} = \dfrac{\sum_{s=k+1}^{n_1} \tilde{w}_s^2}{\sum_{s=k+1}^{n} \tilde{w}_s^2}$.

The advantage of this test is that under the null hypothesis the recursive residuals are independent under the normality assumption. The major disadvantages of the CUSUM test is the requirement that all the regressors be independent of the disturbances which excludes the lagged dependent variable from being included in the model. The CUSUMSQ test is recommended if there is instability of a random fashion rather than of a systematic nature and the sample size is small.

Hsu (1977) investigated two tests for variance change in a sequence of independent normal random variables, when the initial level of the variance is unknown. He investigated two methods, namely the locally most powerful test and the test based upon CUSUM of squares values. He approximated the distribution functions of the two test statistics through the use of Edgeworth expansions and/or the beta distribution by matching the first few moments. He gave critical points of both test statistics for various sample sizes and also compared the powers of the two tests using a Monte Carlo experiment. His results showed that both tests behave almost the same. The power of the tests is relatively high when the changepoint located in the middle position of the data rather than at the end or beginning of the data.

Hackl (1980) constructed a test called the MOSUM test which is based on the moving sums rather than cumulated sums of recursive residuals. The MOSUM test statistic is based on: $M_{n_1} = \dfrac{1}{\hat{\sigma}} \sum_{j=n_1-G+1}^{n} w_j$ for any fixed constant G. Obviously, the

null hypothesis of no structural change is $H_0: E(\mathrm{M}_{n_1}) = 0$. Similar to the CUSUM of squares test, Hackl also constructed the MOSUM-SQ test, which is based on

$$MQ_{n_1} = \frac{\sum_{j=n_1-G+1}^{n} w_j^2}{\sum_{j=k+1}^{n} w_j^2}.$$

McCabe and Harrison (1980) developed a test using the cumulative sum of squares of OLS residuals. Their test has the advantage of being computationally simple and comparable in power with Brown et al.'s (1975) CUSUM of squares test when the degree of instability is high and/or the sample size is large. A major drawback of their procedure is that the critical value of their test depends on the particular set of observations. They thus derive bounds for their test following the approach of Durbin and Watson (1951). Because of the existence of an inconclusive region, their test has not been widely accepted in empirical work.

Pettitt (1980) considered a simple CUSUM type statistic for the changepoint in the case of zero-one observations. He introduced a conditional test of the null hypothesis of no change against the alternative there is change and compared his test with the likelihood ratio test. He also considered the estimation of the changepoint using a simple statistic and showed that the method is asymptotically equivalent to the maximum likelihood estimator in certain circumstances and almost equivalent in others. To investigate its small sample behaviour, he carried out simulation experiments and showed that the new estimator is generally superior to the maximum likelihood estimator.

Dufour (1982) pointed out that a major drawback of the CUSUM test is the requirement that all regressors be independent of the disturbances. In particular, this excludes the lagged dependent variable from being included in the equation. In this

case, he suggested replacing the coefficients of lagged dependent variables by their consistent estimates from the full sample, and hoping that the resulting recursive residuals and any tests based on them will have approximately the same properties as those based on the true coefficients of lagged dependent variables.

Westlund (1985) conducted a Monte Carlo experiment and showed that the MOSUM test has some advantages over the CUSUM test, the CUSUM of square test and the MOSUM-SQ test. However, the choice of G is somewhat arbitrary, the null distribution of the MOSUM test is very complicated and depends on the number of observations. Such a shortcoming prevents the MOSUM test from widespread use.

Kramer et al. (1988) investigated the CUSUM test in the context of structural change when there are lagged dependent variables among the regressors in a linear model. They showed that both a modified CUSUM test, suggested by Dufour (1982), and the straightforward CUSUM test retain their asymptotic significance levels in dynamic models. They showed that, asymptotically, one can disregard the dynamic character of the regression and proceed with the CUSUM test as in the static model.

McCabe (1988) attempted to justify the use of CUSUM type procedures based on OLS residuals rather than recursive residuals using the analysis of multiple decision theory. Ploberger, Kramer and Kontrous (1989) considered the fluctuation test that is based on successive OLS parameter estimates. They derived the limiting null distribution of the test statistic, and showed that it compared favourably to both the CUSUM and CUSUM of squares tests. Ploberger and Kramer (1992) extended the CUSUM procedure to the case of OLS residuals. They found the power of the CUSUM test to be similar to that of their test, except for changes late in the sample when their test has more power.

Ploberger, Kramer and Alt (1989) showed that under some regularity conditions, Dufour's (1982) approach is asymptotically valid if the coefficients of lagged dependent variables are estimated under the null hypothesis of no structural change. Any choice between Dufour's approach and that of Kramer et al.'s is a matter of power and of the accuracy of the nominal size of the test. Their Monte Carlo results also showed that the dynamic CUSUM test performs much better than Dufour's approach.

Some related works are Johnson and Bagshaw (1974) who obtained the limit processes for partial sums of observations from ARMA processes and explored the effect of ARMA noise on CUSUM statistics. Bagshaw and Johnson (1975) examined the effect of ARMA noise on the run length distribution for CUSUM statistics. Tang and MacNeill (1993) give theoretical results and report simulations on the effect of correlation. Bai (1994b), Antoch, Huskova and Pragkova (1997), Horvath (1997) and Lombard and Hart (1994) used the least squares method to construct consistent, asymptotically normal and efficient estimators of the error spectral density function and covariances. Boldin (1982) and Bai (1994a) obtained the weak convergence of empirical processes of residuals, in stationary ARMA processes, while Koul and Levental (1989) studied the explosive case of autoregression.

2.2.4 The Chow test

A very widely used test in the literature for detecting structural change when the changepoint is known, is the Chow test named after Chow (1960). The statistic can be explained using the regression model (2.1). Under the alternative hypothesis

the regression coefficient vector β_1 changes to β_2 after the n_1th observation. Under this model, the statistic for the Chow test can be given as

$$F_{(k,n-2k)} = \frac{(SSE - (SSE_{n_1} + SSE_{n-n_1}))/k}{(SSE_{n_1} + SSE_{n-n_1})/(n-2k)} \tag{2.19}$$

where $SSE = (y - X\hat{\beta})'(y - X\hat{\beta})$, $\hat{\beta} = (X'X)^{-1}X'y$, $\hat{\beta}_1 = (X_{n_1}'X_{n_1})^{-1}X_{n_1}'y_{n_1}$,

$SSE_{n_1} = (y_{n_1} - X_{n_1}\hat{\beta}_1)'(y_{n_1} - X_{n_1}\hat{\beta}_1)$, $SSE_{n-n_1} = (y_{n-n_1} - X_{n-n_1}\hat{\beta}_2)'(y_{n-n_1} - X_{n-n_1}\hat{\beta}_2)$,

$\hat{\beta}_2 = (X_{n-n_1}'X_{n-n_1})^{-1}X_{n-n_1}'y_{n-n_1}$, y_{n_1} and X_{n_1} are the parts of y, and X up to the change point n_1 respectively and, y_{n-n_1} and X_{n-n_1} are the parts of y, and X after the change point n_1 respectively.

This statistic has an F distribution with k and $n-2k$ degrees of freedom under H_0. According to the Chow test, if the calculated value of the test statistic (2.19) is greater than critical value we reject the null hypothesis and we conclude that structural change has occurred.

Chow (1960) developed this test to test the equality of two sets of coefficients in the linear regression model. Despite its simplicity and hence its widespread popularity, the test suffers a serious limitation. It is, in general, valid only under the rather strong assumption that the disturbance variances in the two regressions are equal. There has been a lot of research conducted on investigating the non-robustness of the Chow test, or on proposing new competitive tests that do not require the assumption of equal variances.

Another limitation of the Chow test is that the timing of the break is assumed known. This is not always a realistic assumption because, for example in economics, it could take some unknown period of time before the influence of international events, be they political or environmental, are felt. This is especially

true for countries such as Australia, which is geographically isolated from the rest of the world. Unanticipated decisions may also happen when agents have rational expectations or when policy announcements are partially predicted. Also, in the study of impacts of treatments (say, of a drug treatment or an advertising campaign), the point when the treatment might take effect is usually unknown. The difficulty with the problem of testing for a structural break with an unknown break point is that it does not fit into the "regular" testing framework. The reason is that the break point only appears under the alternative hypothesis, and not under the null.

The Chow test suffers from serious distortion in size even with moderate heteroscedasticity, for example see Toyoda (1974), Schmidt and Sickles (1977) and Ohtani and Toyoda (1985) for details. Attempts have been made to approximate the null distribution of the Chow test by Toyoda (1974), Ali and Silver (1985) and Conerly and Mansfield (1989), among others.

2.2.5 The Sup F Test

The model for this test is the same as the one given in (2.1) where n_1 is an unknown. The test statistic can be written as follows

$$SupF = \underset{k \le n_1 \le n-k}{Sup} F_{n_1} \tag{2.20}$$

where $F_{n_1} = \dfrac{(SSE - (SSE_{n_1} + SSE_{n-n_1}))/k}{(SSE_{n_1} + SSE_{n-n_1})/(n-2k)}$.

An alternative test, which is equivalent to the Sup F test depends on the idea of maximizing the likelihood, function with respect to n_1. For fixed n_1, and given the model for the unknown changepoint, the likelihood function can be written as follows

$$l(\beta_1,\beta_2,\sigma_1,n_1)=(2\pi)^{-\frac{n}{2}}(\sigma_1^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma_1^2}\left[\left(y_{n_1}-X_{n_1}\beta_1\right)'\left(y_{n_1}-X_{n_1}\beta_1\right)\right.\right.$$
$$\left.\left.+\left(y_{n-n_1}-X_{n-n_1}\beta_2\right)'\left(y_{n-n_1}-X_{n-n_1}\beta_2\right)\right]\right\}. \tag{2.21}$$

For fixed n_1 the maximum likelihood estimator for the variance is

$$\hat{\sigma}_1^2=\frac{SSE_{n_1}+SSE_{n-n_1}}{n}.$$

The maximized loglikelihood function can be written as

$$l(n_1)=-\frac{n}{2}-\frac{n}{2}\log 2\pi-\frac{n}{2}\log\hat{\sigma}_1^2. \tag{2.22}$$

The maximum likelihood estimate of the changepoint can be found by maximizing the above loglikelihood over n_1. Then, for that n_1 the F statistic can be found.

The problem with the Sup F test and also the other tests considering the changepoint endogenously is that the changepoint appears only under the alternative hypothesis but not under the null hypothesis as a parameter. Asymptotic analysis of such problems can be found in Davies (1977, 1987), Andrews and Ploberger (1991), Hansen (1991) and King and Shively (1993). They show that the asymptotic distributions differ from the standard ones.

Andrews (1990) determined the asymptotic distributions of the W, LM and LR test statistics under the null hypothesis of parameter stability and under the alternative hypothesis of parameter instability including one time structural change. Since the W, LM and LR test statistics are extensions of the Sup F test statistic, the same asymptotic distribution applies for the Sup F test. Moreover, he compared this test with tests such as the CUSUM and the CUSUM of squares of Brown, Durbin and Evans (1975) and the fluctuation test of Sen (1980) and Ploberger, Kramer and Kontrus (1989) in terms of power and concluded that Sup F test is a more powerful test.

Andrews and Ploberger (1992) derived a class of test statistics which has an optimality property based on weighted local asymptotic power. They considered non-linear models with non-trending observations, and showed that the asymptotic null distribution is an exponential average of the square of a standardized tied-down Bessel process of appropriate order.

Andrews and Ploberger (1994) derived asymptotically optimal tests for testing problems in which a nuisance parameter exists under the alternative hypothesis but not under the null. They mentioned that the testing problem is nonstandard and the classical asymptotic optimality results for the Lagrange multiplier, Wald, and likelihood ratio tests do not apply and used a weighted average power criterion to generate optimal tests. This criterion is similar to that used by Wald (1943) to obtain the classical asymptotic optimality properties of Wald tests in "regular" testing problems. In fact, the optimal tests they introduced reduce to the standard LM, Wald, and LR tests when standard regularity conditions hold. They gave a new optimal test in the nonstandard cases and found that LR test is not an optimal test

Andrews, Lee and Ploberger (1996) derived a class of finite sample optimal tests for one or more changepoints at unknown times in a multiple linear regression model. Their tests can be used to test the null hypothesis of parameter consistency against the alternative of multiple parameter change at unknown times. They considered a weighted average power criterion function and obtained a class of test statistics indexed by a scalar measure c of the magnitude of the parameter changes. They checked the sensitivity of their optimal tests to the scalar measure c and the relative power of the tests to other tests such as likelihood ratio test, the midpoint F test, the CUSUM test of Brown et al. (1975) and a test introduced by Nyblom (1989) for martingale parameter changes. They concluded that their test statistics are not

very sensitive to the choice of c and their asymptotic F statistic is preferable to the likelihood ratio statistic.

2.2.6 Other Approaches

In this section we discuss Bayesian and nonparametric approaches to structural change in the literature. There were many Bayesian contributions to the changepoint literature during the mid 1970's and early 1980's, particularly for univariate and multivariate linear models.

Sen and Srivastava (1975) considered a procedure for testing whether the means of each variable in a sequence of independent random variables are the same, against alternatives that a change might have occurred after some point. They provided Bayesian test statistics as well as some statistics depending on estimates of the changepoint. They derived the exact and asymptotic distribution functions for some of the Bayesian statistics. They compared the relative powers of the Bayesian procedure and the classical LR test using Monte Carlo simulations and showed that the latter has superior power when the change is close to 1 or to n, and the former has more power when the change is near the middle of the sample period.

Broemeling and Choy (1980) studied the linear regression of type (2.1) with $1 \le n_1 \le n-1$ as the unknown changepoint. They used Bayesian analysis based on the marginal posterior distribution of the changepoint, n_1. In the discussion of their simulation study, they mentioned that the posterior probability is sensitive to the corresponding prior probability.

Feder (1975) discussed the asymptotic distribution theory of least squares estimators in regression models having different analytical forms in different regions of the domain of the independent variable. He showed that the unrestricted least

squares estimator is consistent under suitable assumptions. He assigned a uniform prior distribution to the unknown regression parameters and assigned three different prior distributions to n_1. He used the Monte Carlo method to compare the mean squared error (MSE) and the mean biases of the Bayesian estimates of the changepoint corresponding to the three different prior distributions, with that of the maximum likelihood estimate. He found the MSE of each of the three Bayesian estimates are smaller than those of the maximum likelihood estimates.

Smith (1975) considered a Bayesian approach to the problem of making inferences about the unknown changepoint n_1. He derived the posterior probabilities of the occurrence of changepoint being at various possible points $1 \le n_1 \le n$, used these probabilities to calculate Bayesian estimates, and derived hypothesis tests using posterior odds. His inferences are based on the posterior probabilities of the possible changepoints and he gave a detailed analysis for the cases in which the distributions are binomial and normal with some numerical illustrations.

Nonparametric methods play an active role in the estimation and testing of changepoints. With respect to change in location, following Page (1954, 1955), Blum (1987), and Bhattacharya and Johnson (1968), Pettitt (1979) obtained the first theoretical results on changepoint detection using nonparametetric methods. He described how to use the Mann-Whitney statistic to detect a changepoint and derived approximate significance probabilities for testing no change against change. He gave exact and approximate results for testing the null hypothesis of no change. The methods he gave were illustrated by the analysis of three sets of data for zero-one observations, binomial observations and continuous observations. He made some comparisons with other methods based on differences of means. He concluded that his techniques for continuous data are highly efficient when normal or near normal data is used.

Some related changepoint problems solved using the non-parametric approach can found in Darkhovsky (1976), Csorgo and Horvath (1987), Lombard (1983), Csorgo and Horvath (1988), Praagman (1988), Yao (1990), Gombay (1994), Gombay and Horvath (1995), Gombay and Huskova (1996), and Horvath and Shao (1996).

Lombard (1988) adapted the Fourier expansion of Brownian bridges to handle problems of changepoint analysis. Ferger (1994a,b) constructed changepoint estimators when small disorders occur and studied the power of some nonparametric changepoint tests. Boukai (1993), Ferger (1994c), and Antoch, Huskova and Veraverbeke (1995) discussed the applications of bootstrap to the estimation of the time of change. Stute (1996) and Horvath (1998) investigated the properties of U-statistics.

2.3 Brief Review of Information Criteria (IC) Based Model Selection

2.3.1 Introduction

Model selection plays an important role in econometric and statistical modelling. In the literature, many methods of model selection have been suggested over the last few decades. Indeed, the area of model selection is now quite vast in its scope. A full treatment is really beyond the scope of this section. Thus we will confine our attention mainly to a particular type of model selection technique which is called IC based model selection.

The organisation of this section is as follows. Subsection 2.3.2 discusses the consequences of model selection through testing. Subsection 2.3.3 presents the

historical development of some IC based model selection procedures. Some concluding remarks are made in the final subsection.

2.3.2 Consequences of Model Selection through Testing

Over the last few decades, various testing mechanisms have been proposed for model selection in the literature. A partial list includes Gaver and Geisel (1974), Atkinson and Fedorov (1975), Leamer (1978), White (1982a, 1982b, 1983, 1990), MacKinnon (1983), Davidson and MacKinnon (1984), Bunke and Droge (1985), Linhart and Zuchini (1986), McAleer (1987), Grassa (1989), Brownstone (1990), Pötscher (1991), Maddala (1992), Hurvich and Tsai (1993) and Wess and Indurkhya (1996). Despite the fact that hypothesis tests have been widely used in the model building process, there are many opportunities to make mistakes when selecting the best possible model by using such testing mechanisms. We discuss this problem in the remainder of this subsection.

In the early days of econometric model building, models were formulated taking into account highly parsimonious relationships in accordance with the contemporary sophisticated economic theories, then using statistical procedures, these models were estimated and tested for model adequacy. If the models were still found to be inadequate, further terms were added and the process was repeated until an adequate model was found. This model building procedure is often known as specific-to-general methodology. In contrast to this idea, general-to-specific methodology, also known as the Hendry methodology (Hendry, 1989) has became popular. This procedure involves the formulation of a general model and then sequentially testing the model for various parameters under some specified criteria until a desirable model is found. The disadvantage of this procedure is that there are

a number of problems in such a long sequence of testing procedures. Pretesting is one such problem.

If a type I error occurs as a consequence of the test, the estimated parameters of the model will be less efficient. On the other hand, if a type II error occurs, then the parameter estimates suffer from an omitted variable bias. The second problem involves the assumption of the error distribution of the model. For diagnostic testing, it is frequently assumed that the error term of the model is normally distributed. But in many situations, this assumption is violated. As a result, the test may have incorrect size and may also lack power. One more problem with such a sequential testing mechanism is that the size of the overall procedure is often difficult to control.

Granger et al. (1995) noted that model selection through hypothesis testing has a number of limitations. For example, two investigators working on the same set of data could easily end up with different models just because they performed their tests in different orders or used different levels of significance. In their view, model selection decisions should be based on a well-thought-out model selection procedure rather than a series of classical pairwise hypothesis tests. They noted a number of advantages of this approach. These are that no one model is favoured due to the choice of null hypothesis, it does not matter in what order the calculations have been done, using an information criterion is equivalent to testing each model against all other models by the likelihood ratio (LR) test and selecting the model which is accepted against all other models, no pretesting problem arises if the model selection criterion is consistent as the sample size n tends to infinity, the judgement of significance level is no longer needed although there is a big issue of what penalty function is appropriate.

2.3.3 Historical Development of Some Existing IC Based Model Selection Procedures

During the last few decades, many researchers have worked on IC based model selection procedures in the literature. A partial list includes Akaike (1973, 1974, 1981), Tong (1975), Hocking (1976), Bhansali and Downham (1977), Thompson (1978), Schwarz (1978), Leamer (1979), Hannan and Quinn (1979), Amemiya (1980), Rissanen (1986, 1987, 1988), Quinn (1988), Nishii (1988), Franses (1989), Sin and White (1992, 1996), Mills and Prasad (1992), Hurvich and Tsai (1993), Grose and King (1994), Fox (1995), King *et al.* (1995) and Granger *et al.* (1995). These articles cover both Bayesian and non-Bayesian approaches for IC based model selection procedures. Very recently, Hughes (1997), Hossain (1998), Kwek (1999) and Billah (2001) provided reviews of IC based model selection procedures in their Ph.D. dissertations.

In this section, we firstly discuss the usual definitions and limitations of some of the important IC based model selection procedures. Then we review the IC based model selection work so far reported in the literature during the last few decades. In fact, the range of work on the development of model selection procedures in econometrics is very wide today. It ranges from stepwise hypothesis testing to IC based model selection procedures. Our purpose is to highlight some of the important points.

The maximised likelihood function and the penalty function are the two main ingredients of IC based model selection procedures which involve choosing the model with the largest maximised log of the likelihood function minus a penalty term. The penalty term is a function of the number of parameters included in the

model under consideration and typically the sample size. But unfortunately, there is little agreement about the best form of this penalty function.

We review in the following section, commonly used important IC procedures, namely, Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (BIC), generalized cross validation criterion (GCVC), Hannan and Quinn's criterion (HQC), Theil's R-squared criterion (TRSC), Mallow's C_p criterion (MCPC), Hocking's S_p criterion (HSPC), and Amemiya's prediction criterion (APC).

2.3.3.1 Akaike's Information Criterion (AIC)

Akaike (1973) proposed a simple and very useful criterion called Akaike's information criterion for selecting the best-fitting model among alternative models. AIC was developed incorporating Kullback-Leibler (KL) information with the use of maximum likelihood principles and negative entropy. The form of AIC varies form author to author. In this study we use the penalized maximized log-likelihood form given by Fox (1995) as

$$AIC = L(\hat{\theta}) - k, \tag{2.23}$$

where $L(\hat{\theta})$ is the maximized log-likelihood of the model, $\hat{\theta}$ is the estimated parameter vector and k is the penalty term which is the number of free parameters included in the model under consideration. Another maximized log-likelihood form of AIC in the case of a linear regression model (2.1) under H_0 is

$$n\log(\hat{\sigma}^2) + 2k \tag{2.24}$$

where $\hat{\sigma}^2$ is the maximum likelihood estimate of the residual variance for k parameters. We select the model with maximum AIC among alternative models.

Akaike's (1969) paper was the basis of the development of AIC; for selecting the best order of an AR process using the minimum final prediction error (FPE) criterion. Papers on applications of AIC to particular model selection problems include Akaike (1977) on factor analysis and polynomial fitting and Akaike (1978) on choosing AR processes.

Ozaki (1975) proposed an effective algorithm for the fitting of nonstationary autoregressive models. Ozaki (1977) applied the AIC procedure to ARMA model selection from series A-F of Box and Jenkins (1970), and found that when only the models with orders of AR greater than 0 and MA greater than 2 are considered, the procedure nearly always selects the models identified by the graphical method recommended by Box and Jenkins (1970).

Tong (1975) developed a procedure for determining the order of an AR signal process from noisy data by employing AIC. The procedure was illustrated through some numerical examples using both artificially generated and real data. His proposed procedure gives the asymptotic properties of the maximum LR statistics and KL information for discriminating between two distributions.

Shibata (1976) analysed the statistical properties of Akaike's (1973, 1974) proposed method. In particular, the author examined the property of consistency of AIC and pointed out that AIC does not provide consistent model order selections for an autoregressive model of finite order.

Soderstrom (1977) investigated two criteria, AIC and FPE, by using the *F*-test to choose between two models where the smaller model is nested within the larger one and concluded that these two procedures are asymptotically equivalent.

Hurvich and Tsai (1989) found that as the dimension of the candidate model increases with the size of sample, AIC tends to provide a negatively biased estimate

of the KL information. They argued that this under estimation is due to the fact that AIC tends to overfit in very small samples.

Kozin and Nakajima (1980) showed that the AIC procedure is applicable to a class of time-varying non-stationary AR processes. A similar issue has also been considered by several other authors, see for example, Tong (1978), Stone (1979), Shibata (1980, 1981), Sakai (1981), Woodroofe (1982), Breiman and Freedman (1983), Ronchetti (1985), Shibata (1986), Kabaila (1995) and Pötscher and Novak (1996).

2.3.3.2 Schwarz's Bayesian Information Criterion (BIC)

The Schwarz's (1978) Bayesian information criterion (BIC) provides a simple reference method for choosing between competing models. The problem with AIC is inconsistency in the sense that it does not always select the model having maximum information with probability tending to one as the sample size tends to infinity. This problem seems to be overcome by BIC, which is usually recommended for large sample cases. BIC is a widely used criterion in econometrics today and is given by penalized maximized log-likelihood form

$$BIC = L(\hat{\theta}) - \frac{k \log(n)}{2} \tag{2.25}$$

where $L(\hat{\theta})$ is the maximized log-likelihood of the model and $\hat{\theta}$ is the estimated parameter vector. When the number of observations is large, BIC penalizes additional parameters much more than AIC, leading to more parsimonious models being chosen.

Akaike (1981) indicated that in many practical situations, the use of the BIC procedure is problematic if there is no clearly defined proper prior distribution of the

parameters. However, this difficulty has been solved by Rissanen (1986, 1987, 1988) who derived a criterion based on stochastic complexity and the associated minimum description length (MDL) principle which is similar to BIC but with more general applicability.

Stone (1979) remarked that Schwarz's (1978) work was a special case of earlier criteria for model discrimination by Jeffreys (1967). Kohn (1983) showed that BIC consistently chooses a minimal dimension in a large class of models.

2.3.3.3 Hannan and Quinn's Criterion (HQC)

The criterion due to Hannan and Quinn (1979), which is less commonly used, can be written in penalized maximized log likelihood form as

$$HQC = L(\hat{\theta}) - k\log(\log(n)). \tag{2.26}$$

This procedure is particularly employed to choose the overall lag length in a vector autoregressive model. Hannan (1981) extended the results of Hannan (1980) to multivariate ARMA processes.

Fox (1995) noted that HQC shares a common property with AIC and BIC, i.e., the marginal penalties for these three criteria are constant as the number of parameters increases for a fixed n. For the general form of a regression model, HQC is equivalent to:

$$\log(\sigma_k^2) + \frac{2k\log(\log n)}{n} \tag{2.27}$$

where $\hat{\sigma}_k^2$ is the maximum likelihood estimate of the error variance of the model with k parameters.

2.3.3.4 Theil's R-Squared Criterion (TRSC)

In the history of econometrics, perhaps the first model selection procedure in the case of the linear regression model was the coefficient of multiple determination called *R*-squared which is given as

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{2.28}$$

where *RSS* is the residual sum of squares, *TSS* is the total sum of squares of the model of interest and $\bar{y}$ is the mean of the y_is. This procedure involves choosing the model with the largest R^2. But the problem is that the addition of an extra regressor in a model usually increases (and never decreases) the value of R^2.

Theil (1961) proposed the adjusted R^2 denoted, as $\bar{R}^2$ that takes into account the number of estimated parameters for model comparison yielding a criterion, which is sensitive to the number of remaining degrees of freedom. The $\bar{R}^2$ criterion is given by

$$\bar{R}^2 = 1 - R^2\left(\frac{n}{n-k+1}\right), \tag{2.29}$$

where n is the sample size and k is the total number of parameters included in the model.

Theil (1971) showed that a decision rule which favours the model with the largest $\bar{R}^2$ will result 'on average' in the correct choice of model. It takes into account the goodness of fit of the model, as well as its parsimony. But a difficulty arises when the model that fitted the data well does not seem to have a good predictive performance. Schmidt (1973, 1975) has shown that the $\bar{R}^2$ criterion does

not help us in selecting the true model when the regression contains both the variables of the true model plus some extra irrelevant independent variables. Moreover, it has been suggested that $\overline{R}^2$ does not penalize the loss of degrees of freedom heavily enough in practice (see for example, Amemiya, (1985)).

Fox (1995) showed an equivalent procedure to maximizing $\overline{R}^2$ is to maximize

$$\text{TRSC} = L(\hat{\theta}) + \frac{n}{2}\log(n-k+1) \tag{2.30}$$

where $L(\hat{\theta})$ is the maximized log-likelihood.

2.2.3.5 Mallows' C_p Criterion (MCPC)

Mallows (1964) suggested a variable selection criterion that has been used in economics, econometrics and many other social sciences. Common references to this criterion include Gorman and Toman (1966) and Mallows (1973).

Fox (1995) expressed Mallows' statistic in the penalized maximized log-likelihood form as,

$$\text{MCPC} = L\left(\hat{\theta}\right) - \frac{1}{2}n\ln\left(1+\frac{2k}{\left(n-k^{*}\right)}\right), \tag{2.31}$$

where k^* is the number of free parameters in the smallest model which nests all models under consideration. The model with highest MCPC in chosen.

AIC can also be viewed as an extension of MCPC in the linear context (Atkinson, 1980). However, a closer link is through FPE (Akaike, 1969) as it is based on the mean square error. Although the MCPC and the PRESS (prediction sum of squares) criterion of Allen (1974) have been around much longer, they have

not been as popular as AIC. Also see Akaike (1974) for a discussion on MCPC and Young (1982) for further discussion on the generalized MCPC.

2.3.3.6 Amemiya's Prediction Criterion (APC)

Amemiya's (1972, 1980) PC was derived as an alternative method to estimate the variance in Mallows' criterion within a hypothesis-testing framework. Expressing APC as the average prediction variance based on regression models, we choose the model that minimizes:

$$\hat{\sigma}^2\big((n+k)/(n-k)\big). \tag{2.32}$$

It is interesting to note that both FPE and APC evaluate the mean squared prediction error of the predictor derived from each model. Then using Fox's generalization, we can express APC as the penalized maximized log-likelihood of the form:

$$\text{APC} = L(\hat{\theta}) - \frac{1}{2}n(\log(n+k)) + \frac{1}{2}n(\log(n-k)). \tag{2.33}$$

This criterion was also suggested by Rothman (1968) and Akaike (1969) (Rothman called it J_p and Akaike called it FPE). It is interesting to note that Mallows' criterion is the same as APC as both have an identical penalty term.

Chan *et al.* (1974) proposed a criterion that resembles the FPE statistic for order estimation of ARMAX systems. Its interpretation as an FPE type criterion may, however, be questioned because the calculations in the paper contain some flaws (see, e.g., Soderstrom (1977)).

Nevertheless, the proposed criterion appears to be a compromise between small residual variance and accurate parameter estimates (like, for example, the criteria of Rissanen (1976), Maklad and Nichols, (1980), etc.) and it was reported to

behave well in a number of Monte Carlo simulations. Comparisons with the F-test and Akaike's FPE criterion are included. Chan et al. (1975) gave a discussion of some order selection procedures of ARMAX systems with emphasis on the F-test and Akaike's FPE criterion.

2.3.3.7 Hocking's S_P Criterion (HSPC)

Hocking (1976) suggested a criterion that was reviewed by Thompson (1978) and can be expressed in penalized maximized log likelihood form as

$$\text{HSPC} = L(\hat{\theta}) + \frac{n}{2}\log(n-k) + \frac{n}{2}\log(n-k-1). \tag{2.34}$$

Thompson (1978) examined methods for variable selection according to whether the regressors included in the models are fixed or random. One major limitation of HSPC is it is not applicable for nonstochastic regressors.

2.3.3.8 Generalized Cross-Validation Criterion (GCVC)

Schmidt (1971) suggested cross-validation (CV) which involves splitting the sample into roughly two equal parts. The first part is used for fitting a model and the second part is reserved for assessing the predictive ability of the model (often-called model validation).

In fact, the CV procedure is used to determine the loss of efficiency in parameter estimation of the model by providing a measure of future prediction error. Nevertheless, one of the main problems is that, like AIC and MCPC, this criterion is inconsistent. Another problem is that the calculation of CV is really cumbersome. Moreover, this criterion is rather different in nature from the other criteria mentioned

above. However, a good approximation to CV, called the generalized cross-validation criterion (GCVC) has been derived by Golub et al. (1979). When applied to regression models it has the form,

$$\left(1-\frac{k}{n}\right)^{-2}\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}. \tag{2.35}$$

The model with the minimum GCVC is chosen. Fox (1995) derived the penalized maximized log-likelihood form that is given as

$$GCVC = L(\hat{\theta}) + n\log\left(1-\frac{k}{n}\right). \tag{2.36}$$

Other studies related to CV are Schmidt (1974, 1975) and Allen (1971a, 1974). Stone (1974) re-introduced and systematized CV procedures, and Geisser (1975) discussed predictive sample re-use methods. Schmidt called the CV score, the sum of squared predictive errors (SSPE) and Allen called it the PRESS. Stone (1977a) showed the asymptotic equivalence between AIC and a cross-validation criterion. Stone (1977b) analyzed the asymptotic properties (consistency and efficiency) of the one-item-out cross-validatory assessment scheme of Stone (1974) mainly in the context of some particular applications. Other forms of CV are found in Craven and Wahba (1979).

2.3.3.9 The Works of King and Others

King et al. (1995) developed general model selection procedures in which one can calculate the penalties by controlling the probabilities of correct selection so that no one model is unnecessarily favoured. In doing so they proposed two methods; one called the common model approach and the other the representative fixed points approach. They also provided an algorithm for calculating the penalties

based on fixing probabilities of correct selection according to both of these approaches.

Forbes et al. (1995), presented three IC based model selection procedures of which the first two (one is called FIC (IC based on the F distribution) and the other is called QFIC (IC based on the quasi F distribution)) are explicitly designed for regressor selection and the third one is for general model selection. For the FIC procedure, they derived a new penalty function on the basis of sums of critical values from particular F distributions whilst for QFIC, they calculated penalties from particular chi-squared distributions. Both penalties come from a desire to control the probability of incorrectly choosing additional regressors.

The consistency property of one-sided AIC (OSAIC) and a bias corrected OSAIC (OSAICc) for small sample regression problems involving knowledge of signs of parameter values is discussed by Hughes (1997) (also see Hughes and King (1994)). He conducted a Monte Carlo study to investigate OSAIC and showed that there is no reason for practitioners to uniformly favour a consistent criterion, such as BIC. Examples of bias-corrected versions of IC are AICc by Hurvich and Tsai (1989) and AIC_1 by Hurvich, Shumway and Tsai (1990).

Rahman and King (1997) and Rahman, Bose and King (1998) developed other forms of penalty functions where the functions consist of composite variables of n and k. Hossain (1998) proposed simulation based information criterion called controlled information criterion (CIC) for selecting between Box-Cox transformation models and compared the performance of AIC, BIC and CIC and found that CIC performs better than the existing IC procedures considered in his study.

Kwek and King (1997a) explored a set of IC-based model selection procedures and compared their small sample performance by their relative performance as measured by the AMPCS in the context of choosing between

autoregressive conditional heteroscdastic (ARCH) and generalized autoregressive conditional heteroscdastic (GARCH) models. In terms of AMPCS, they found that TRSC has the best penalty function for small sample performance but not always for very large samples. AIC did not perform as poorly as BIC for small samples. In large samples, AIC performed relatively better than TRSC in terms of AMPCS.

Kwek and King (1997b) introduced the conditional heteroscedastic IC (CHIC) as the criterion with an optimal penalty function for ARCH and GARCH models and concluded that this criterion has a better performance than IC procedures for finite sample model selection problems.

Kwek (1998) considered model selection in the context of ARCH and GARCH models and derived optimal penalties based on maximizing the AMPCS. She claimed that her results provides a good way to evaluate different procedures and is more efficient compared to other approaches which give higher average mean probabilities of correct selection.

Kwek (1999) compared the performances of penalized log-likelihood based IC procedures (AIC, BIC, HQC, MCPC, GCVC, TRSC, HSPC and APC) and found that in terms of AMPCS, TRSC is the best criterion and BIC is the worst. On the other hand, when the sample size is large TRSC losses its efficiency to AIC and the latter become comparatively a better criterion. The performance of the optimal AMPCS procedure was found to be clearly better than all the other existing IC procedures.

King and Bose (2000) considered model selection problems in linear regression models using optimal penalties in the sense of choosing the model with largest penalised maximized log-likelihood function. They use simulation methods to estimate probabilities of correct selection and suggested choosing penalties that optimize the average of these probabilities. Results from their Monte Carlo

experiment showed that AMPCS gives better average probabilities than other IC procedures considered.

Billah and King (1998) considered optimal AMPCS penalties for choosing between different time-series processes for linear regression disturbances. Their Monte Carlo results show that optimal AMPCS penalties provide the best criterion for both model selections for shorter and longer forecasting horizons.

Billah and King (2000) studied the application of AMPCS to time series model selection. They claim from their Monte Carlo results that the optimal AMPCS penalties consistently dominate all existing IC procedures.

Billah (2001) investigated several important issues concerning IC based small sample model selection for exponential smoothing models as well as regression models with ARMA error processes. He introduced conditional likelihood (CL) based IC procedures for selecting between exponential smoothing models and improved conditional likelihood (ICL) based IC procedures. He found that optimal AMPCS penalties provide the best procedure compared to existing IC procedures in the sense that it gives higher average probabilities of correct selection.

In this thesis, we aim to develop a simulation based model selection procedure in the presence of structural change when the possible changepoint is unknown and compare our method with existing IC procedures discussed above. Clearly AMPCS penalties are worth investigating. Unfortunately the studies to date have only been for circumstances where there are only 3 or 4 competing models. We are interested in seeing how it works when there is a very large number of competing models.

2.4 Brief Review of the Simulated Annealing Algorithm (SAA)

In this section, we introduce a global optimization algorithm called SAA, which works well to maximize complex functions. More specifically, we wish to investigate whether this algorithm can be used to estimate penalty values when dealing with many models.

To locate the real global minimum with certainty, a global optimization method has to be used. In this section, we will introduce a global optimization called the simulated annealing algorithm (SAA) which is capable of even maximizing very complex functions. More specifically, we wish to investigate whether this algorithm can be used to estimate penalty values of the model selection procedures.

The algorithm is based upon that of Metropolis et al. (1953), which was originally proposed as a means of finding the equilibrium configuration of a collection of atoms at a given temperature. The minimization of the objective function corresponds to the energy state of the solid. Therefore, the name of the algorithm is drawn from an analogy between solving an optimization problem and simulating the annealing of a solid. In econometric literature methods used to estimate parameters of a model, for example, the generalized method of moments the maximum likelihood method and nonlinear least squares, depend upon optimization algorithm, such as Newton-Raphson, to estimate parameters in the model. However, almost all-conventional algorithms occasionally fail to estimate the optimum value of parameters. Popular statistical and econometric packages use these algorithms to solve optimization problems. Reviews on these packages can be found in Judge et al. (1985) and Press et al. (1986).

Generally traditional optimization algorithms assume approximately quadratic nature of the function to be optimized. Unfortunately, some functions frequently do not follow this assumption. A common problem to the classical algorithms is that although these algorithms converge; yet they may converge to local maxima instead of the global maxima. In this situation, researchers generally try to solve these problems by using different approaches, for example, trying different starting values (see Cramer (1986) and Finch et al. (1989)). Fortunately, the SAA, assumes very little about the function, can tackle the optimization problem very efficiently (see Corana et al. (1987) and Goffe et al. (1994)). The advantage of this algorithm is that it is explicitly designed for functions with multiple maxima and also works well for complex functions. The SAA discovers the function's complete surface and while moving both up hill and downhill tries to optimize the function. Therefore, the SAA is much more user friendly than traditional algorithms found in econometric literature.

The connection between this algorithm and mathematical optimization was first noted by Pincus (1970) and Kirkpatrick et al. (1983) who proposed that the SAA form the basis of an optimization technique for combinatorial (and other) problems. Extensions of simulated annealing to the case of functions defined on continuous sets have also been introduced in the literature (e.g., Geman and Hwang (1986), Gidas, (1985), Holley, Kusuoka and Stroock (1989), Jeng and Woods, (1990), Kushner (1985), Cerny (1984), Fox (1988a,b), Hajek (1988) and Otten (1989)).

Geman et al. (1984) first gave a necessary and sufficient condition for the convergence of the annealing method to the global minimum. Their method is usually called either Boltzmann annealing (BA) or classical simulated annealing. Szu et al. (1987) proposed the fast annealing method, which is a semi local search and consists of occasional long jumps. They made some improvements to the

Boltzmann form. Ingber et al. (1989) presented the very fast-simulated re-annealing method. They argued that their algorithm permits a fast exponential cooling schedule, while fast annealing has only an inverse cooling schedule, and Boltzmann annealing has only an inverse logarithmic cooling schedule.

In a comprehensive study of the SAA, Johnson et al. (1990, 1991, 1992) discussed the performance of the SAA on four problems: the travelling salesman problem, graph partitioning problem, graph colouring problem and number partitioning problem. In general, the performance of the SAA was mixed: in some problems, it outperformed the best known heuristics for these problems, and, in other cases, specialized heuristics performed better.

Many researchers have considered the SAA as a tool in the development of optimal experimental designs. Some examples include Van Laarhoven (1987) and Meyer and Nachtsheim (1988). Variants of SAA based on Bayesian ideas have been proposed by Laud, Berliner and Goel (1989), Van Laarhoven et al (1989) and Aarts et al. (1989).

In the initial stage SAA was known as the combinatorial SAA because it was introduced in combinatorial optimization problems. This SAA has been successfully used in image processing (Carnevali et al. (1985), reconstruction of pollycrystalline structures (Telly et al. (1987)), pollution control (Derwent (1988)), neural networks (Wasserman and Schwartz, (1988)), and computer and circuit design (Wong et al. (1988))). Other SAAs proposed in the optimization literature are as follows: adaptive random search (Pronzato et al. (1984)), fast SAA (Szu and Hartly (1987)), down hill simplex with annealing (Vetterling et al. (1994)) and direct search SAA (Ali et al. (1997)). Corana et al. (1987) derived a new SAA for optimization of functions of continuous variables from the SAA introduced in combinatorial optimization. This new SAA has been found to be more reliable, being nearly

always able to find the optimum, or at least a point very close to it. However, this algorithm appears to be the best with respect to the combination of ease of use and robustness.

For estimation of econometric models Goffe et al. (1994) compared four algorithms introduced by Corana et al. (1987) with that of the SAA. Compared to the three traditional algorithms, the SAA was found to have several advantages. The most important advantage is that it can maximize functions with which traditional algorithms have extreme difficulty or simply cannot maximize at all. This algorithm can also be used as a diagnostic tool to understand how conventional algorithms fail.

The SAA has a number of other advantages over existing traditional optimization procedures. For any function, if there are more than one maxima then the SAA can escape from a local maxima by moving both up hill and down hill to find the global maxima. In classical optimization, one of the conditions is that the function to be optimized should be approximately quadratic and it needs to be differentiable, but in case of SAA these conditions are not necessary (see Corana et al. (1987)). Second, the SAA can handle a very complex function. Another advantage of this algorithm is that it provides valuable information about the function through the step length vector. The most important advantage of the SAA is that it can properly optimize functions that are very complex and nearly impossible to optimize (see Goffe et al. (1994)). The only drawback of the SAA is that the required very high powered computer. On the other hand, because of the availability of high-powered computers now a day, this problem seems not to be a major problem. As a consequence, the SAA is an attractive optimization algorithm for difficult functions. In this thesis, we implement the SAA to estimate penalty functions, (by optimizing what is a step function) for a small and large numbers of alternative models. Overall, the SAA is a generally applicable and easy-to-

implement probabilistic approximation algorithm that is able to produce good solutions for an optimization problem, even if we do not understand the structure of the problem well.

2.5 Conclusions

This chapter is devoted to three areas of the literature; testing for structural change, IC model selection procedures and the simulated annealing optimization algorithm.

In Section 2.2, some developments of related areas in the literature on testing for structural change were briefly discussed in context of econometrics and statistics. The survey revealed that there is a large body of literature on changepoint testing problems. From our review, we see that LR tests are widely used in practice when data may possess a possible changepoint of unknown timing. Unfortunately, the finite sample distribution of the LR test statistic is unknown, although the critical values for the test can be calculated by simulation. We can develop formulae for critical values of the LR test for different sample sizes, different significance levels, number of regressors in the model and types of regressors. We can apply a response surface approach to estimate formulae for critical values. When there is a possible structural change with an unknown changepoint in the data, we consider the use of the LR test statistic in Chapter 3.

In Section 2.3, we reviewed of some existing IC procedures for model selection. Due to the enormous literature on model selection contributed to by mathematicians, statisticians and econometricians, we are not able to review all the statistical properties that come with each criterion. We therefore highlighted only the salient points for some leading criteria. In this section we also discussed some of

their asymptotic and finite sample properties. Most of these statistical properties were derived out of model selection problems in linear regression and time series models. This section begins with a brief discussion on some obvious demerits involved with model selection through hypothesis testing. The survey reveals that there is a large body of literature on model selection ranging from hypothesis testing to IC based model selection procedures.

The evidence so far supports the contention that model selection decisions should be based on some well-thought-out model selection criteria rather than classical hypothesis testing. However, the discussion on IC based procedures mainly focuses on the penalty term that is one of the main ingredients of such procedures. We also reviewed model selection based on optimal penalties. This penalty has the ability to select a model from a group of alternative models by estimating mean probabilities of correct selection of a model and choosing the penalty that maximizes the average mean probability of correct selection (AMPCS). Unfortunately, it appears that almost no work has been reported on the use of IC based model selection procedures for detecting the presence of structural change. Chapter 4 of this thesis aims to concentrate on this particular issue.

If the differences between the penalty values of an IC procedure are large, keeping other things the same, then smaller models are favoured. On the other hand, for a small difference in penalty values, the larger models are favoured. Thus, it is difficult for a given problem to assess which IC is best. Regardless of these limitations, improvements in existing IC procedure may be possible by estimating penalty values numerically. We briefly reviewed the simulated annealing optimization algorithm in Section 2.4 and will use it to calculate optimal penalties for model selection in Chapter 5.

CHAPTER 3

Testing for Structural Change when the Changepoint is Unknown

3.1 Introduction

In this chapter we consider the problem of testing for structural change in the presence of an unknown changepoint. The presence of a structural change in data that is not detected is a hazard for applied economists, econometricians and statisticians, with serious consequences for model performance and forecasting. If the model selected is misspecified, that is, if the data possess a structural change at some point then the model chosen may not perform well in the sense that it will not provide good forecasts. For this reason, it is important to test from the beginning whether the data possesses a significant structural change or not. If the changepoint is known (such as World War II, 1973 oil shock etc.), one can use Chow's (1960) F statistic to test for possible structural change in the linear regression model.

Since economic conditions are constantly changing, it is not always possible to know with certainty, which of the changes, and with what timing, affect the performance of a linear regression model applied to economic time series data.

It is helpful to have a test for structural change that does not require the knowledge of the changepoint in advance. In other words, we need to be able to test for structural change when the changepoint is unknown. In response to this need, many tests for structural change have been developed. A number of these tests were discussed in Chapter 2. Among these, the CUSUM and the CUSUM of squares test of Brown, Durbin and Evans (1975), and the fluctuation test of Sen (1980) and Ploberger, Kramer and Kontrus (1989) are well known. In the former two tests, recursive residuals are used and in the latter, recursive estimates of parameters are used.

As mentioned in Chapter 2, Andrews (1990) compared the likelihood ratio (LR) test with tests such as the CUSUM and CUSUM of squares tests and the fluctuation test of Sen (1980) and Ploberger, Kramer and Kontrus (1989) in terms of power. He concluded that the LR test is more powerful than these other tests. Andrews (1993) determined the asymptotic distributions of the LR test statistics under the null hypothesis of parameter stability and for the alternative hypothesis of parameter instability including one time structural change. He mentioned that the Wald and Lagrange multiplier test statistics are generally asymptotically equivalent to LR test statistic under the null and local alternatives.

Seber and Wild (1989) showed that under the null hypothesis of parameter stability, the finite sample null distribution of the LR statistic does not depend on the parameters and error variance of the model although it does depend on the explanatory variables in finite samples (but not asymptotically). Therefore, it is possible to simulate the null distribution of the LR test for any particular data set and arbitrary values of the parameters including the error variance of the model. A critical value for any significance level can be found from this simulation. The null hypothesis is rejected if the value of the test statistic is greater than this value.

King and Shively (1993) worked on hypothesis testing problems in which the nuisance parameter is present only under the alternative hypothesis. They reparameterized the testing problem in such a way that the reparameterized testing problem involves testing a greater number of parameters. They started by considering testing $H_0: \theta = \theta_0$, where θ_0 is a known $p \times 1$ vector, against $H_1: \theta \neq \theta_0$. They reparameterized $\overline{\theta} = \theta - \theta_0$ so that the testing problem becomes one of the testing $H_0: \overline{\theta} = 0$ against $H_1: \overline{\theta} \neq 0$. Then they reparameterized $\overline{\theta}$ into polar coordinates $\overline{\theta}_1 = r\cos\gamma_1$, $\overline{\theta}_j = r\left(\prod_{k=1}^{j-1}\sin\gamma_k\right)\cos\gamma_1$, $j = 2,\cdots,p-1$, $\overline{\theta}_p = r\prod_{k=1}^{p-1}\sin\gamma_k$ with $r = \sqrt{\overline{\theta}'\overline{\theta}}$ in such a way that the null hypothesis $H_0: \overline{\theta} = 0$ is now equivalent to $H_0: r = 0$ and $H_1: \overline{\theta} \neq 0$ is equivalent to $H_1: r \neq 0$. Under the null hypothesis $\gamma_1,\cdots,\gamma_{p-1}$ are not defined, so we now have a test of $r = 0$ in which $\gamma_1,\cdots,\gamma_{p-1}$ are nuisance parameters and present only under the alternative.

King and Shively's main contribution was to observe that for testing problems in which nuisance parameters are present only under the alternative hypothesis, we may be able to reparameterize to a higher dimensional testing problem by the reverse of the above transformations. For the LR test, this will have consequences with respect to which critical value should be used.

Tan and King (1994) considered King and Shively's approach for testing structural change when the changepoint is unknown. They conjectured that for the case of a change only in one parameter under standard regularity conditions, the LR test statistic follows an asymptotic distribution under the null hypothesis of no structural change that is a probability mixture of χ_1^2 and χ_2^2 and suggested expressions for the respective probability weights. Through simulation

experiments, their critical values were found to be misleading for both one-sided and two-sided tests when the range of possible changepoints is wide compared to the total period. They also found that the small-sample distribution of the LR test statistic might be quite sensitive to the regressors making a useful asymptotic solution difficult to find.

Andrews (1993) gave asymptotic critical values based on the asymptotic null distribution that covers 1%, 5% and 10% levels of significance, different numbers of regressors and changepoint or point of structural change for any value of the changepoint expressed as a proportion of the sample between 0.05 and 0.50.

From the above discussion and related literature survey in Chapter 2 we can conclude that the finite sample distribution of the LR test statistics is unknown, and sensitive to the number and kind of regressors used in the model. The evidence appears to be that it is more powerful than other tests for testing for structural change of unknown timing. Our aim is to develop a small-sample test procedure that allows the LR test to be applied with confidence in finite samples. Under the null hypothesis, the regression parameters are constant across all the periods. Under the alternative, a particular regression parameter changes at an unknown changepoint. In our case, critical values of the LR test statistic depend on the number of regressors, types of regressors and the sample size.

Because the LR test statistic does not have a known finite sample distribution, the critical values for the test must be found by Monte Carlo estimation. An applied econometrician using the available data can estimate the critical value specific to that setting. A few decades ago it was relatively expensive in terms of computer time and effort to do this, but today, due to the availability of high-powered computers, it is easier to find critical values via simulation. We develop formulae for critical values of the LR test for different sample sizes,

different significance levels, different numbers of regressors in the model and types of regressors. We will check whether estimated critical values of the LR statistic depends on these factors in the context of our framework of analysis. We will apply a response surface approach to estimate formulae for critical values. To check the accuracy of these response surface formulae, we will conduct a small Monte Carlo experiment.

The remainder of this chapter is organised as follows. In Section 3.2, we discuss the model and construct our test statistics based on maximization of the likelihood function for different changepoints of the linear regression model. In Section 3.3, we calculate critical values of the test statistics for different sample sizes, significance levels, numbers of regressors and types of regressors. Estimation of the response surfaces formulae for critical values of the test statistic from simulated critical values is outlined in Section 3.4. The accuracy of the response surface formulae is checked through a Monte Carlo study which is outlined in Section 3.5. Section 3.6 contains a brief discussion of results of the Monte Carlo study. Section 3.7 presents some concluding remarks.

3.2 The Model and the Test Statistic

3.2.1 The Model

We consider the linear regression model for $t = 1, \cdots, n$, with a possible change of unknown timing in one coefficient,

$$y_t = \begin{cases} x_t'\beta_0 + w_t\gamma + u_t & \text{for } t < n_1; \\ x_t'\beta_0 + w_t(\gamma + \delta) + u_t & \text{for } t \geq n_1, \end{cases} \tag{3.1}$$

where y_t is the dependent variable at time t, x_t is a $k \times 1$ vector of regressors at time t, w_t is a scalar variable that is of interest, β_0 is a $k \times 1$ vector of regression

coefficients, and γ and δ are unknown scalar parameters. The error term is assumed independent, identically distributed $N(0,\sigma^2)$. We have chosen to work with this simple form with a possible change in one coefficient in the hope of making progress which can be applied to other more complicated models at a later time. Model (3.1) can be written jointly as

$$y_t = x_t'\beta_0 + w_t\gamma + z_t\delta + u_t, \qquad \text{for} \quad t = 1, \ldots, n, \tag{3.2}$$

where z_t is a dummy variable defined as

$$z_t = \begin{cases} 0, & t \le n_1, \\ w_t, & t > n_1. \end{cases} \tag{3.3}$$

Denoting

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

$$X_1 = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} & w_1 \\ \vdots & & \cdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} & w_n \end{bmatrix},$$

$$X_2 = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} & w_1 & z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} & w_n & z_n \end{bmatrix},$$

$$u = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}',$$

$$\beta_0 = \begin{bmatrix} \alpha_0 & \beta_1 & \cdots & \beta_{k-1} \end{bmatrix}',$$

$$\theta_1 = \begin{bmatrix} \beta_0 & \gamma \end{bmatrix}',$$

$$\theta_2 = \begin{bmatrix} \beta_0 & \gamma & \delta \end{bmatrix}',$$

model (3.2) can be rewritten in matrix form as

$$y = \begin{cases} X_1\theta_1 + u, & \text{under } H_0, \\ X_2\theta_2 + u, & \text{under } H_1, \end{cases} \tag{3.4}$$

where $u \sim N(0, \sigma^2 I_n)$.

3.2.2 The Test Statistics

The null hypothesis of interest is

$$H_0 : \ \delta = 0,$$

and the alternative hypothesis is

$$H_1 : \ \delta \neq 0.$$

The likelihood ratio test of a structural change of unknown timing can be described as follows. The loglikelihood function of the sample under the alternative hypothesis that there is a changepoint in the data after period n_1 is

$$l_1 = -\frac{n}{2}\log(2\pi\sigma_2^2) - \frac{1}{2\sigma_2^2}(y - X_2\theta_2)'(y - X_2\theta_2). \tag{3.5}$$

The loglikelihood function under the null hypothesis of no changepoint in the data is

$$l_0 = -\frac{n}{2}\log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2}(y - X_1\theta_1)'(y - X_1\theta_1). \tag{3.6}$$

Differentiating (3.5) with respect to the parameters β_0, γ, δ and σ_2^2, and equating the resultant equations to zero, we obtain the conditional maximum likelihood (ML) estimates of β_0, γ, δ, and σ_2^2 under the alternative hypothesis that there is a changepoint in the data after period n_1 as

$$\begin{aligned} \hat{\theta}_2 &= (X_2'X_2)^{-1}X_2'y, \\ \hat{\sigma}_2^2 &= (y - X_2\hat{\theta}_2)'(y - X_2\hat{\theta}_2)/n. \end{aligned} \tag{3.7}$$

Differentiating (3.6) with respect to the parameters β_0, γ and σ_1^2, and equating the resultant equations to zero, we obtain the ML estimates of β_0, γ and σ_1^2 under the null hypothesis that there is no changepoint in the data as

$$\begin{aligned} \hat{\theta}_1 &= (X_1'X_1)^{-1}X_1'y, \\ \hat{\sigma}_1^2 &= (y - X_1\hat{\theta}_1)'(y - X_1\hat{\theta}_1)/n. \end{aligned} \tag{3.8}$$

Substituting the estimates of (3.7) into (3.5), we obtain the concentrated log-likelihood function under the alternative hypothesis of the sample given a changepoint in the data after period n_1 as

$$\hat{l}_1 = -\frac{n}{2}\log(2\pi\hat{\sigma}_2^2) - \frac{n}{2}. \tag{3.9}$$

Substituting the estimated values from (3.8) into (3.6), we obtain the maximized log-likelihood function under the null hypothesis of no changepoint

$$\hat{l}_0 = -\frac{n}{2}\log(2\pi\hat{\sigma}_1^2) - \frac{n}{2}. \tag{3.10}$$

When the changepoint n_1 is unknown, a naturally, intuitive approach would be to estimate n_1 and then apply the LR test at that estimate of n_1. Given n_1, (3.9) is maximized by substituting in the estimated value of σ_2^2 from (3.7). Maximizing (3.9) with respect to n_1 is equivalent to finding the n_1 for which $\hat{\sigma}_2^2$ is minimum. The LR test statistic can be obtained by substituting minimum values of $\hat{\sigma}_2^2$ in (3.9) which we will denote by $\hat{\hat{l}}_1$ and then taking twice the difference between it and the log-likelihood of (3.10) that is, $LR = 2(\hat{\hat{l}}_1 - \hat{l}_0)$.

3.3 Calculation of Critical Values via Simulation

Analytically it is difficult, if not impossible, to find the distributional form of the test statistic; therefore, in this section we calculate critical values for the LR test statistic via simulation. It is possible to simulate the null distribution of the test statistic LR for any particular regressor set. Because critical values are invariant to a change of parameters and error variance of the model, any arbitrary values of the parameters and error variance of the model can be used in the simulation. We conducted a number of simulation experiments to find the critical values of the LR test statistic in a range of circumstances. These critical values will allow us to get a formula that can be applied to the LR test for structural change of unknown changepoint with some degree of confidence.

3.3.1 Design of Simulation Experiments

Our main purpose is to calculate the critical values by simulation. Under the null hypothesis, values of the dependent variable y_t were generated from the following equation

$$y_t = x_t' \beta_0 + w_t \gamma + u_t, \qquad \text{for} \quad t = 1, \ldots, n,$$

where $u_t \sim N(0,1)$, w_t is a scalar variable, β_0 is a $k \times 1$ vector of regression coefficients, and γ is a constant coefficient. The $k \times 1$ independent variables are generated following Engle et al.'s (1985) Monte Carlo experimental design; that is explanatory variables (excluding the constant term) were generated from the first order autoregressive process $x_{it} = \phi x_{it-1} + e_{it}$, with $e_{it} \sim IN(0,1)$ for $t = 1, \ldots, n$, where ϕ takes values 0, 0.7, 1.0 and 1.02 which covers white noise, autoregressive,

random walk and explosive processes, respectively. The number of regressors k was allowed to range from 1 to 15 in turn with ϕ being the same for each regressor and w_t was generated from the uniform distribution with range from 0 to 1.

We set initial values of the parameters as one throughout because under null hypothesis, the distribution of the LR statistic does not depend on β_0 and γ. Four different sample sizes of 25, 50, 75 and 100 were used. We generated 10000 LR test statistics for the linear regression model for each set of ϕ's, k and n. We then ordered the calculated LR from the lowest to the highest values, and obtained the 90th, 95th, 97.5th, and 99th percentiles, which are the required critical values for the 10%, 5%, 2.5% and 1% level of significance respectively. Throughout, we use the GAUSS 3.2.12 software (GAUSS is a mathematical and statistical programming language, produced by Aptech Systems, Inc., Kent, Washington) to estimate the parameters of the model by the method of ML estimation. In the model, error terms were simulated using pseudo random numbers from the GAUSS function RNDNS that generates standard normal variates for regression errors. The seed for the random number generator for each experiment was 1786.

3.3.2 Results of the Simulation

Tables 3.1 to 3.4 report the critical value calculation results of the Monte Carlo simulations. We will now discuss the results. We discuss the overall trends in the critical values in four stages. The first stage involves the patterns or trends with respect to sample size variation, the second involves patterns as the number of regressors in the model changes, the third considers changes in the type of autoregressive regressors and the fourth discusses some general patterns with regard to the significance level.

A noticeable feature is that the simulated critical values of the LR test increase as the sample size increases from 25 to 75 for small $k = 1$ and 2, and it decreases as the sample size increases from 75 to 100 at the 1% level of significance for different values of ϕ considered. At the 2.5%, 5% and 10% levels of significance, critical values of the LR test increase as the sample size increases from 25 to 50 and they decrease as the sample size increases from 50 to 75. They decrease as n increases from 75 to 100 for the different values of ϕ considered. The critical values of the LR test increase as the sample size increases for $k = 3$ or more at different levels of significance for different values of ϕ.

The largest calculated critical value of the test occurs at the 1% level of significance when $\phi = 1.02$ and $n = 25$ and takes the value 17.453 whereas when $n = 100$ it takes the value 6.633. The largest critical values of the test at the 2.5%, 5% and 10% levels of significance occur when $\phi = 1.02$ and $n = 25$ and are 13.903, 11.386 and 8.995, respectively, whereas when $n = 100$, the critical values are respectively, 5.615, 4.662 and 3.838. The minimum critical values of the test statistic at the 1%, 2.5%, 5% and 10% levels of significance occur when $\phi = 0$ and $n = 25$ and are respectively, 4.779, 3.814, 3.136 and 2.446, whereas when $n = 100$ these critical values are respectively, 4.906, 4.059, 3.497 and 2.818.

The critical values of the test almost always increase with an increase in the number of regressors k. The largest increases occur for small n and for small ϕ. The smallest increases occur for $n = 100$ and for large levels of significance. For a large number of regressors in the model when $\phi = 0$ and $k = 15$ at the 1%, 2.5%, 5% and 10% levels of significance, the maximum value of the critical values are 16.196, 12.477, 10.209 and 7.698, respectively.

The simulated critical values of the LR type test statistic appear to be practically unchanged as the type of autoregressive regressors change with everything else held constant. The critical values are typically the same for $\phi = 0$ and for $\phi = 0.70$ at different levels of significance and are also roughly the same for $\phi = 1.0$ and $\phi = 1.02$. The latter are almost always slightly bigger than the former.

Obviously the critical values decrease as the level of significance increases. This decrease is largest for large k. This variation in critical values with n, k, ϕ, and α suggests the need for formulae for critical values of the test statistic, which will be developed in the next section.

3.4 Estimation of Critical Values via Response Surface Approach

The simulated critical values calculated in the previous section show that they vary reasonably systematically with the number of explanatory variables k, the autoregressive parameter ϕ of the regressors and the sample size n. For large samples, it can be very time consuming to calculate the critical values via simulation even on fast computers. For example, it took a Pentium-III personal computer several days to perform relevant computations for $n = 400$ and 10,000 replications for the one-regressor case. To reduce the computational load, we use the critical values of Tables 3.1 to 3.4 as data to estimate formulae for critical values at different levels of significance by using a response surface approach.

The advantages of response surfaces are (i) they reduce computational costs and effort using specific factors to calculate critical values in a simple regression; (ii) they allow easy calculation of critical values for sample sizes not included in the experimental design; and (iii) response surfaces for commonly used

significance levels are easily programmed so as to provide estimated finite sample critical values directly.

In our case, the response surface approach involves developing formulae to estimate critical values using various functions of k, ϕ and n. Normally, researchers have tabulated critical values at different levels of significance for different sample sizes. Such tabulations recognize the dependence of the critical values on the sample size, significance level and other factors. That dependence can be approximated by regressing the Monte Carlo estimates of the critical values on functions for different factors such as k, ϕ and n.

The response surface approach determines the combination of levels of different factors, which will produce the best-fitting model. The regressors are an intercept, four inverse powers of the sample size n, four-power functions of the number of regressors k and types of autoregressive regressor ϕ. We also considered other plausible interactive terms between n, k and ϕ. We considered a total of 80 regressors including intercept for the response surfaces given below:

k_1, k_2, k_3, k_4, n_1, n_2, n_3, n_4, ϕ, k_1k_2, k_1k_3, k_1k_4, k_2k_3, k_2k_4, k_3k_4, n_1n_2, n_1n_3, n_1n_4, n_2n_3, n_2n_4, n_3n_4, k_1n_1, k_1n_2, k_1n_3, k_1n_4, k_2n_1, k_2n_2, k_2n_3, k_2n_4, k_3n_1, k_3n_2, k_3n_3, k_3n_4, k_4n_1, k_4n_2, k_4n_3, k_4n_4, $k_1\phi$, $k_2\phi$, $k_3\phi$, $k_4\phi$, $n_1\phi$, $n_2\phi$, $n_3\phi$, $n_4\phi$, $k_1k_2k_3$, $k_1k_2k_4$, $k_2k_3k_4$, $n_1n_2n_3$, $n_1n_2n_4$, $n_2n_3n_4$, $k_1k_2\phi$, $k_1k_3\phi$, $k_1k_4\phi$, $k_2k_3\phi$, $k_2k_4\phi$, $k_3k_4\phi$, $n_1n_2\phi$, $n_1n_3\phi$, $n_1n_4\phi$, $n_2n_3\phi$, $n_2n_4\phi$, $n_3n_4\phi$, $k_1n_1\phi$, $k_1n_2\phi$, $k_1n_3\phi$, $k_1n_4\phi$, $k_2n_1\phi$, $k_2n_2\phi$, $k_2n_3\phi$, $k_2n_4\phi$, $k_3n_1\phi$, $k_3n_2\phi$, $k_3n_3\phi$, $k_3n_4\phi$, $k_4n_1\phi$, $k_4n_2\phi$, $k_4n_3\phi$, $k_4n_4\phi$, where $k_1 = k$, $k_2 = k^2$, $k_3 = k^3$, $k_4 = k^4$, $n_1 = n$, $n_2 = 1/n$, $n_3 = n_1^2$, and $n_4 = n_1^3$.

The estimation process involved estimating many equations of penalty responses using the above factors. The coefficients that are clearly insignificantly different from zero were eliminated using smallest absolute t-statistic (well below 1 in absolute value) or highest p-values of the coefficients. After obvious insignificant regressors were dealt with, we then used criteria such as AIC, BIC, the Durbin-Watson (DW) test and $\overline{R}^2$ to help make the decision of choosing a final formulae for critical values at a certain level of significance. In the following subsection we will discuss the estimated response surface formula for critical value at different levels of significance.

3.4.1 Formulae for Critical Values

We fitted models using all regressors and by applying the criteria mentioned in the above section. Our finally selected estimated models for 1%, 2.5%, 5% and 10% significance levels respectively are as follows:

$$\begin{aligned} CV_{1\%} = {} & 0.2699\phi + 0.4508k_1 - 0.029k_2 + 62.630n_1 - 5.930n_1\phi - 47.859n_1k_1 \\ & \quad (0.034) \quad (0.099) \quad (0.007) \quad (19.790) \quad (1.919) \quad (10.962) \\ & + 1.256n_1k_1\phi + 2.376n_1k_2 + 0.007n_1k_4 - 1405.684n_2 + 1730.065n_2k_2 \\ & \quad (0.161) \quad (0.508) \quad (0.002) \quad (353.493) \quad (361.239) \\ & - 0.708n_2k_4 - 18060.632n_3k_1 - 136.468n_3k_3 + 20.841n_3k_4 + 4.360 \\ & \quad (0.127) \quad (4688.339) \quad (21.206) \quad (2.833) \quad (0.208) \end{aligned} \tag{3.13}$$

$\overline{R}^2 = 0.997$, $s^2 = 0.125$, $SSR = 3.519$, $DW = 1.830$, $AIC = -1.251$, $BIC = -1.019$.

$$
\begin{aligned}
CV_{2.5\%} = {} & 0.709\phi + 0.153k_1 - 0.044k_1\phi + 0.061k_2 + 0.002k_2\phi - 0.007k_3 \\
& (0.123) \quad (0.048) \quad (0.016) \quad (0.024) \quad (0.001) \quad (0.003) \\
& + 0.00021k_4 + 62.332n_1 - 37.173n_1\phi + 2.429n_1k_1\phi - 11.712n_1k_2 \\
& (0.00043) \quad (14.294) \quad (10.139) \quad (0.302) \quad (2.851) \\
& + 1.159n_1k_3 - 0.029n_1k_4 + 465.345n_2\phi + 527.363n_2k_2 - 43.750n_2k_3 \\
& (0.297) \quad (0.009) \quad (189.372) \quad (124.038) \quad (10.245) \\
& + 0.723n_2k_4 - 6367.748n_3k_2 - 1346.942n_2 + 431.051n_3k_3 + 3.383 \\
& (0.178) \quad (1725.869) \quad (274.105) \quad (131.999) \quad (0.159)
\end{aligned}
\tag{3.14}
$$

$\overline{R}^2 = 0.996$, $s^2 = 0.115$, $SSR = 2.892$, $DW = 2.025$, $AIC = -1.406$, $BIC = -1.101$.

$$
\begin{aligned}
CV_{5\%} = {} & 0.446\phi + 0.346k_1 - 0.017k_2 + 0.0004k_3 + 44.285n_1 \\
& (0.074) \quad (0.073) \quad (0.008) \quad (0.00082) \quad (14.436) \\
& - 26.59n_1\phi - 41.23n_1k_1 + 1.64n_1k_1\phi + 1.02n_1k_2 - 1130.29n_2 \\
& (0.728) \quad (6.667) \quad (0.108) \quad (0.395) \quad (255.864) \\
& + 431.363n_2\phi + 1885.226n_2k_1 - 0.157n_2k_4 - 25012.300n_3k_1 \\
& (139.141) \quad (200.446) \quad (0.045) \quad (2635.103) \\
& - 52.549n_3k_3 + 7.003n_3k_4 + 3.044 \\
& (16.174) \quad (1.497) \quad (0.158)
\end{aligned}
\tag{3.15}
$$

$\overline{R}^2 = 0.996$, $s^2 = 0.084$, $SSR = 1.590$, $DW = 1.752$, $AIC = -2.037$, $BIC = -1.791$.

$$
\begin{aligned}
CV_{10\%} = {} & 0.201k_1 + 0.041k_1\phi + 27.95n_1\phi - 30.45n_1k_1 - 1198.418n_2\phi \\
& (0.018) \quad (0.009) \quad (6.655) \quad (3.042) \quad (343.590) \\
& + 2152.19n_2k_1 - 130.46n_2k_2 + 13.69n_2k_3 - 0.468n_2k_4 - 6941.37n_3 \\
& (211.473) \quad (41.808) \quad (4.151) \quad (0.135) \quad (1284.9) \\
& - 37411.53n_3n_1 + 3795.375n_3k_2 - 407.256n_3k_3 + 15.177n_3k_4 \\
& (4422.9) \quad (1088.38) \quad (107.74) \quad (3.493) \\
& + 372674.501n_4\phi + 9.706n_3k_3\phi - 9.041n_4k_4\phi - 0.001k_2\phi + 2.752 \\
& (121597.8) \quad (2.481) \quad (4.113) \quad (0.001) \quad (0.030)
\end{aligned}
\tag{3.16}
$$

$\overline{R}^2 = 0.995$, $s^2 = 0.071$, $SSR = 1.129$, $DW = 1.402$, $AIC = -2.363$, $BIC = -2.087$.
Figures in parenthesis indicate standard errors, SSR is the regression sum of squares and s^2 is the estimated error variance of the model.

The response surface formula for the 1% level of significance is (3.13) which consists of 16 factors including an intercept of which there are five individual factors, 9 two-factor interactions where most of the interactions are functions of n and k and one three-factor interaction. The response surface formula for the 2.5% level of significance is (3.14) which consists of 21 factors including the intercept of which there are 7 individual factors, 12 two-factor interactions and one three-factor interaction.

The response surface formula for the 5% level of significance is given by (3.15) which consists of 17 factors including the intercept of which there are six individual factors, 9 two-factor interactions where most of the interactions are functions of n and k and one three-factor interaction. The response surface formula for the 10% level of significance is (3.16) which consists of 19 factors including an intercept of which there are two individual factors, 14 two-factor interactions and two three-factor interactions. Overall, out of the three main factors n, k and ϕ, the first two factors are found to feature much more than the last factor.

3.5 How to Use the Estimated Response Surface Formulae

The response surface formulae (3.13) to (3.16) given above quantify the straightforward dependencies of the critical values of the test on k, ϕ and n and they therefore offer a simple way to estimate critical values when Tables 3.1-3.4 are

not available or do not apply. For any particular values of k, ϕ and n and using formulae (3.13) to (3.16), desired critical values can be found. The choice of k, and n are straightforward but the decision of what to use for ϕ may be a little more problematic. We suggest using the average of the estimated first order autocorrelation coefficients of each of the nonconstant regressors. For example, if x_{it} $i = 1, \cdots, k$ are k regressors and x_{1t} is the intercept, then fit AR(1) models to each nonconstant regressors excluding the intercept. Suppose $\hat{\phi}_2, \cdots, \hat{\phi}_k$ are the estimates of the coefficients in the AR(1) models. Then replace ϕ with the mean $\overline{\phi} = (\hat{\phi}_2 + \cdots + \hat{\phi}_k)/(k-1)$. As we will see, the formulae are not particularly sensitive to the value of ϕ.

3.5.1 Monte Carlo Experiment

In order to check whether there are problems caused by our choice of variables and to evaluate the practical usefulness of our proposed response surface formulae for critical values, we conducted a Monte Carlo experiment. We considered testing the null hypothesis

$$H_0: \quad \delta = 0,$$

against the alternative hypothesis

$$H_0: \quad \delta \neq 0.$$

Under the null hypothesis, the n observations of the dependent variable y_t were generated from the following equation

$$y_t = x_t' \beta_0 + w_t \gamma + z_t \delta + u_t, \qquad \text{for} \quad t = 1, \ldots, n, \text{ with } \delta = 0,$$

where y_t is the dependent variable at time t, x_t is a $k \times 1$ vector of regressors at time t, w_t is a scalar variable that is of interest, z_t is a dummy variable defined as in (3.3), β_0 is a $k \times 1$ vector of regression coefficients, and γ and δ are unknown scalar parameters. The error term is assumed independent, identically distributed $N(0, \sigma^2)$.

Under the null hypothesis of parameter stability, the distribution of the LR statistic does not depend on the parameters β_0 and γ. Therefore it is possible to simulate the distribution of LR for any particular data set and arbitrary β_0 and γ values. We set all elements of the β_0 vector and γ as unity throughout. The error term u_t is assumed independent, identically distributed $N(0,\sigma^2)$ and we set $\sigma^2 = 1$. The following design matrices were used in this experiment:

X_1 A constant, monthly US seasonally adjusted total volume of real retail sales on domestic trade (in billion 1992 USD) and lagged one month commencing 1960(1).

X_2 A constant, monthly US interest rate, the same interest rate lagged one month, real personal income (in billion 1995 USD), and the same variable lagged one month commencing 1960(1).

These design matrices were chosen to reflect a variety of economic and statistical phenomena. In this case, X_1 and X_2 show some long term fluctuations. We used as w_t monthly US seasonally adjusted total volume of real retail sales on domestic trade (in billion 1992 USD) in the case of X_1 and the monthly US interest rate in the case of X_2. After the disturbances were generated, and given the appropriate design matrix, the y's were generated. The LR test statistics were calculated for each of the sample sizes 30, 60, 120 and 240 with 2000 replications.

Note that sample sizes 120 and 240 are outside the range of our sample sizes used to fit our critical value formulae. One of them, $n = 120$, is close to the sample range while $n = 240$ is quite distant. It will therefore be interesting to see how well the formulae work for these two values. When the null hypothesis is true, the proportion of replications in which the test statistic LR rejects the null gives an estimate of the size of the test.

We calculated critical values at the 2.5%, 5% and 10% levels of significance using the response surface formulae given in the previous section. Three parameters n, k and ϕ are central to the test statistic's distributional properties. The first two are known from the data and the problem is how to choose the value of the last one namely ϕ. We calculated ϕ as $\overline{\phi}$ discussed in Section 3.5. We also checked the sensitivity of ϕ on critical values at different levels of significance using response surface formulae (3.13) to (3.16). We used $\phi = 0, .02, .05, .1, .2, .3, \ldots, 1.02$ keeping the sample size n as 60 and the number of regressors k as 6 and found that the critical values generally are not sensitive to changes in ϕ. All computer programs were written in GAUSS (see Aptech, 1997, version 3.2.17) and computations were carried out on a Pentium III with a 933 Mhz CPU.

3.6 Discussion of Results

In this section we report the Monte Carlo results for sample sizes of 30, 60, 120 and 240 for design matrices X_1 and X_2 using response surface equations given by equations (3.14) to (3.16) respectively. We calculated the estimated critical values using the formulae at the 2.5%, 5% and 10% significance levels and present the results in Table 3.5.

We observe that large sample sizes give smaller critical values and as the number of regressor increases, the critical values also increase. Obviously critical values decrease as levels of significance increase.

Using the critical values from Table 3.5, we estimated the sizes of the test. The results are given in Table 3.6. The estimated sizes for the design matrix X_1 are closer to the nominal significance level than those for the design matrix X_2. One possible explanation for this behavior is that the design matrix we used here only has one variable and its lag as regressors. Thus, the results seem to indicate that when regressors involved lagged independent variables, increasing the sample size increases the reliability of the test.

Tables 3.7 and 3.8 give lower and upper bounds of the 95% confidence interval of the estimated sizes of Table 3.6 for design matrices X_1 and X_2 using different sample sizes. We find that all estimated sizes of Table 3.6 are within the bounds indicating the estimated sizes are not significantly different from the nominal sizes. There is a clear sign of improvement as the sample size increases from 30 to 240. Thus, from this Monte Carlo study, we found that the critical values calculated from our response surface formulae have very acceptable sizes, at least for the design matrices we used in the study. Overall, we have seen that

increasing the number of observations improves the reliability of the critical value formulae. We note that in both design matrices the regressors include lagged independent variables which may be helpful with respect to reliability.

3.7 Conclusion

While there are so many ways to develop a test statistic to test for the presence of structural change when there is a possible unknown changepoint in the data, we recommend the use of the LR test. Since this test does not have a known distribution for finite sample sizes, we calculated exact critical values for the test by simulation using 10000 replications for different sample sizes, numbers of regressors and types of regressors. We found that the critical values clearly depend on sample size, the number of regressors and to a less extend on the type of explanatory variables. A portion of this finding supports King and Tan's (1994) finding that the LR test statistic is sensitive to the number of regressors used in the model.

We found that the calculation of critical values via simulation is very time consuming and the computational cost is very high, particularly for very large samples. To overcome this difficulty, we developed formulae for critical values using a response surface approach, which helps to estimate critical values of the test statistic directly avoiding the use of a table at a desired level of significance when the sample size and the number of regressors are known. Response surfaces provide complementary summaries of the vast array of results from the Monte Carlo study undertaken. The response surfaces highlight some simple dependencies of the critical values on the number of regressors, the degree of autocorrelation in the regressors and the sample size.

The problem of how to choose ϕ is the main concern in using our formula. We suggest the use of the average of estimated coefficients from fitted AR(1) models to each of the nonconstant regressors. The reported response surfaces provide a computationally convenient way of finding finite sample critical values at the 1%, 2.5%, 5%, and 10% levels.

We checked the accuracy of the critical value formulae by performing a small Monte Carlo experiment. We calculated the estimated sizes of the test using response surface formulae for critical values at the different nominal levels. We found that the estimated sizes are not significantly different from the nominal size regardless of the sample sizes. Overall the actual sizes of the test are quite satisfactory. We recommend using the LR test statistic for testing structural change of unknown timing with our critical value formulae.

One question we have not answered here is how to make inference about the changepoint n_1 when parameters of the models have to be estimated. This problem is the subject of the next two chapters.

Table 3.1 Empirical Critical Values of the LR Test for Different Numbers of Regressors and ϕ when $n = 25$

α	1%				2.5%			
k	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02
1	4.779	4.766	4.759	4.837	3.814	3.840	3.829	4.022
2	5.113	5.162	5.403	5.500	4.137	4.111	4.266	4.252
3	5.402	5.395	5.724	5.765	4.356	4.353	4.531	4.534
4	5.761	5.754	6.093	6.172	4.679	4.673	4.803	4.845
5	6.082	6.058	6.441	6.732	4.778	4.811	5.217	5.416
6	7.002	6.899	6.952	6.810	5.429	5.432	5.588	5.650
7	7.069	7.126	7.402	7.261	5.462	5.436	5.860	5.856
8	7.509	7.323	7.454	7.469	6.005	6.033	6.098	6.203
9	8.291	8.254	8.686	8.702	6.573	6.622	7.144	7.279
10	8.710	8.652	9.454	9.129	7.006	7.004	7.651	7.590
11	9.766	9.993	10.354	10.278	8.123	8.036	8.412	8.423
12	10.633	10.749	11.513	11.302	8.783	8.830	9.309	9.549
13	11.808	11.912	12.569	12.570	9.136	9.107	10.578	10.725
14	13.871	14.053	14.695	14.852	11.128	11.153	11.955	12.163
15	16.196	16.169	17.106	17.453	12.477	12.567	14.090	13.903

Table 3.1 (cont'd) Empirical Critical Values of the LR Test for Different Numbers of Regressors and ϕ when $n = 25$

α	5%				10%			
k	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02
1	3.136	3.139	3.179	3.356	2.446	2.455	2.463	2.636
2	3.338	3.373	3.524	3.494	2.480	2.480	2.590	2.685
3	3.445	3.462	3.741	3.863	2.590	2.587	2.884	2.832
4	3.720	3.737	3.946	3.968	2.823	2.806	3.039	3.080
5	3.938	3.962	4.298	4.435	3.031	3.016	3.331	3.514
6	4.338	4.347	4.622	4.784	3.220	3.198	3.632	3.698
7	4.358	4.388	4.843	4.822	3.230	3.204	3.785	3.825
8	4.763	4.756	5.066	5.170	3.432	3.448	3.985	4.067
9	5.182	5.111	5.754	5.993	3.717	3.727	4.456	4.691
10	5.478	5.485	6.285	6.299	4.026	4.020	4.715	4.877
11	6.298	6.346	6.973	7.114	4.705	4.737	5.339	5.588
12	6.952	6.986	7.744	7.838	5.165	5.204	5.961	6.109
13	7.363	7.478	8.502	8.577	5.609	5.654	6.577	6.914
14	8.801	8.822	9.651	9.998	6.479	6.585	7.509	7.712
15	10.209	10.182	11.249	11.386	7.698	7.735	9.033	8.995

Table 3.2 Empirical Critical Values of the LR Test for Different Numbers of Regressors and ϕ when $n = 50$

α	1%				2.5%			
k	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02
1	5.044	5.046	5.153	5.259	4.237	4.243	4.328	4.504
2	5.117	5.133	5.364	5.526	4.247	4.272	4.391	4.566
3	5.420	5.352	5.605	5.641	4.555	4.547	4.596	4.760
4	5.636	5.430	5.758	5.795	4.786	4.583	4.862	4.852
5	5.350	5.689	5.775	5.881	4.603	4.677	4.932	4.993
6	5.731	5.755	5.822	5.953	4.643	4.770	4.945	5.108
7	5.923	5.944	6.060	6.192	4.898	4.930	5.223	5.191
8	6.222	6.255	6.440	6.362	4.995	5.054	5.534	5.390
9	6.351	6.398	6.737	6.563	5.428	5.448	5.587	5.507
10	6.438	6.489	6.785	6.674	5.440	5.512	5.701	5.673
11	6.708	6.638	7.150	7.161	5.515	5.588	6.066	6.113
12	6.773	6.924	7.227	7.271	5.583	5.613	6.139	6.123
13	7.158	6.957	7.267	7.320	5.957	5.804	6.236	6.166
14	6.955	7.135	7.640	7.532	5.799	5.978	6.483	6.370
15	7.232	7.263	7.643	7.787	6.166	6.140	6.602	6.568

Table 3.2 (cont'd) Empirical Critical Values of the LR Test for Different Numbers of Regressors and ϕ when $n = 50$

α	5%				10%			
k	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02
1	3.570	3.511	3.665	3.791	2.820	2.829	2.980	3.129
2	3.497	3.567	3.678	3.796	2.900	2.917	3.107	3.177
3	3.805	3.792	3.914	4.036	3.065	3.082	3.264	3.305
4	3.985	3.839	4.172	4.225	3.167	3.130	3.363	3.511
5	3.824	3.995	4.212	4.267	3.112	3.152	3.367	3.516
6	3.970	4.025	4.242	4.321	3.199	3.214	3.438	3.531
7	4.103	4.140	4.495	4.465	3.255	3.241	3.696	3.659
8	4.286	4.320	4.634	4.608	3.435	3.437	3.808	3.782
9	4.468	4.510	4.790	4.668	3.585	3.608	3.951	3.894
10	4.697	4.682	4.875	4.813	3.724	3.741	4.048	4.034
11	4.667	4.694	5.215	5.146	3.763	3.768	4.231	4.198
12	4.664	4.709	5.239	5.212	3.743	3.797	4.291	4.248
13	5.076	4.945	5.399	5.340	4.074	4.044	4.494	4.495
14	4.967	5.084	5.557	5.566	4.051	4.087	4.649	4.620
15	5.280	5.296	5.621	5.587	4.300	4.355	4.678	4.689

Table 3.3 Empirical Critical Values of the LR Test for Different Numbers of Regressors and ϕ when $n = 75$

α	1%				2.5%			
k	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02
1	5.153	5.140	5.228	5.293	4.168	4.161	4.313	4.464
2	5.229	5.233	5.304	5.330	4.261	4.263	4.503	4.574
3	5.283	5.274	5.535	5.448	4.307	4.300	4.601	4.629
4	5.342	5.297	5.624	5.554	4.310	4.323	4.665	4.684
5	5.353	5.344	5.660	5.700	4.458	4.439	4.686	4.760
6	5.424	5.401	5.690	5.835	4.524	4.516	4.778	4.888
7	5.482	5.472	5.757	6.111	4.570	4.551	4.833	4.966
8	5.590	5.565	6.125	6.131	4.600	4.610	4.993	5.085
9	5.660	5.731	6.142	6.147	4.648	4.625	5.127	5.173
10	5.764	5.803	6.158	6.266	4.757	4.734	5.134	5.250
11	5.846	5.812	6.159	6.331	4.981	4.988	5.228	5.369
12	6.177	6.127	6.222	6.402	5.051	5.032	5.275	5.464
13	6.178	6.230	6.504	6.586	5.062	5.056	5.512	5.558
14	6.405	6.453	6.550	6.742	5.072	5.106	5.539	5.565
15	6.457	6.468	6.569	6.911	5.212	5.180	5.597	5.741

Table 3.3 (cont'd) Empirical Critical Values of the LR Test for Different Numbers of Regressors and ϕ when $n = 75$

α	5%				10%			
k	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02
1	3.486	3.484	3.609	3.780	2.783	2.787	2.962	3.062
2	3.527	3.514	3.826	3.913	2.836	2.842	3.146	3.210
3	3.569	3.571	3.859	3.958	2.859	2.862	3.148	3.258
4	3.653	3.671	3.866	4.014	2.956	2.953	3.165	3.281
5	3.683	3.674	3.933	4.077	2.968	2.968	3.186	3.310
6	3.701	3.702	4.081	4.164	2.969	2.982	3.342	3.418
7	3.809	3.818	4.145	4.238	3.019	3.027	3.426	3.527
8	3.840	3.863	4.299	4.369	3.076	3.080	3.452	3.564
9	3.982	3.963	4.348	4.378	3.174	3.184	3.515	3.613
10	4.051	4.043	4.389	4.465	3.210	3.189	3.582	3.665
11	4.076	4.089	4.476	4.562	3.230	3.239	3.599	3.711
12	4.151	4.150	4.553	4.579	3.348	3.332	3.687	3.808
13	4.243	4.243	4.619	4.751	3.397	3.405	3.791	3.898
14	4.249	4.258	4.645	4.769	3.422	3.417	3.800	3.946
15	4.304	4.284	4.697	4.872	3.450	3.468	3.863	3.957

Table 3.4 Empirical Critical Values of the LR Test for Different Numbers of Regressors and ϕ when $n = 100$

α	1%				2.5%			
k	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02
1	4.906	4.917	4.829	5.302	4.059	4.060	4.105	4.451
2	5.055	5.098	5.310	5.372	4.199	4.218	4.335	4.543
3	5.227	5.219	5.313	5.492	4.362	4.379	4.497	4.615
4	5.257	5.272	5.600	5.492	4.385	4.392	4.581	4.784
5	5.284	5.281	5.658	5.659	4.444	4.424	4.656	4.824
6	5.366	5.358	5.717	5.749	4.492	4.509	4.782	4.864
7	5.376	5.407	5.744	5.800	4.578	4.564	4.833	4.972
8	5.567	5.593	5.830	5.964	4.583	4.564	4.945	5.107
9	5.665	5.633	5.894	6.068	4.656	4.646	4.961	5.154
10	5.678	5.744	5.981	6.071	4.696	4.715	5.068	5.171
11	5.691	5.760	6.068	6.134	4.764	4.792	5.075	5.227
12	5.753	5.781	6.168	6.151	4.803	4.801	5.125	5.259
13	5.761	5.786	6.184	6.311	4.895	4.927	5.251	5.297
14	5.983	6.064	6.192	6.375	4.945	4.965	5.300	5.358
15	6.017	6.072	6.610	6.633	4.989	5.029	5.543	5.615

Table 3.4 (cont'd) Empirical Critical Values of the LR Test for Different Numbers of Regressors and ϕ when $n = 100$

α	5%				10%			
k	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02	ϕ=0	ϕ=0.7	ϕ=1.0	ϕ=1.02
1	3.497	3.511	3.508	3.800	2.818	2.826	2.863	3.129
2	3.594	3.607	3.675	3.897	2.924	2.936	2.938	3.202
3	3.643	3.630	3.824	3.934	2.943	2.939	3.176	3.225
4	3.753	3.762	3.921	4.009	3.012	3.009	3.257	3.337
5	3.757	3.777	3.972	4.046	3.016	3.016	3.310	3.340
6	3.842	3.850	4.081	4.123	3.041	3.066	3.333	3.377
7	3.868	3.856	4.139	4.190	3.108	3.111	3.337	3.432
8	3.876	3.872	4.176	4.319	3.161	3.162	3.435	3.523
9	3.924	3.927	4.245	4.335	3.173	3.172	3.500	3.624
10	3.994	4.004	4.304	4.403	3.237	3.249	3.516	3.639
11	4.008	4.018	4.412	4.435	3.250	3.258	3.603	3.643
12	4.096	4.120	4.443	4.504	3.328	3.339	3.697	3.690
13	4.115	4.125	4.502	4.622	3.351	3.363	3.761	3.800
14	4.211	4.193	4.616	4.626	3.367	3.369	3.790	3.834
15	4.219	4.232	4.644	4.662	3.448	3.450	3.820	3.838

Table 3.5 Estimated Critical Values from Response Surface Formulae at Different Levels of Significance for Design Matrices X_1 and X_2 for Different Sample Sizes

	X_1			X_2		
n	2.5%	5%	10%	2.5%	5%	10%
30	4.588	3.846	2.927	5.227	4.221	3.542
60	4.456	3.735	3.102	4.897	4.037	3.423
120	4.416	3.617	3.013	4.609	4.009	3.372
240	4.296	3.575	3.187	4.488	3.962	3.338

Table 3.6 Estimated Sizes of the LR Test Based on Critical Values from Response Surface Formulae at Different Levels of Significance for Design Matrices X_1 and X_5 for Different Sample Sizes

	X_1			X_2		
n	**2.5%**	**5%**	**10%**	**2.5%**	**5%**	**10%**
30	0.0313	0.0527	0.1190	0.0340	0.0614	0.1201
60	0.0270	0.0518	0.1110	0.0327	0.0582	0.1098
120	0.0259	0.0511	0.0986	0.0279	0.0542	0.1020
240	0.0235	0.0508	0.0949	0.0270	0.0510	0.0998

Table 3.7 The 95% Lower and Upper Bounds of the Estimated Sizes of the LR Test Based on Exact Critical Values at Different Level of Significance Using Design Matrix X_1 for Different Sample Sizes

Nominal Size	2.5%		5%		10%	
Sample size	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound
30	0.0196	0.0430	0.0377	0.0677	0.0919	0.1461
60	0.0161	0.0378	0.0369	0.0667	0.0899	0.1320
120	0.0153	0.0366	0.0363	0.0659	0.0786	0.1186
240	0.0133	0.0337	0.0361	0.0655	0.0753	0.1146

Table 3.8 The 95% Lower and Upper Bounds of the Estimated Sizes of the LR Test Based on Exact Critical Values at Different Level of Significance Using Design Matrix X_2 for Different Sample Sizes

Nominal Size	2.5%		5%		10%	
Sample size	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound
30	0.0219	0.0462	0.0453	0.0775	0.0983	0.1419
60	0.0208	0.0447	0.0425	0.0739	0.0888	0.1308
120	0.0168	0.0389	0.0390	0.0693	0.0817	0.1223
240	0.0161	0.0378	0.0362	0.0658	0.0797	0.1199

CHAPTER 4

The Use of Model Selection for Detecting Unknown Changepoints[1]

4.1 Introduction

The conventional approach to linear regression analysis involves the formulation of a model with constant coefficients across the entire time domain. As we have seen in Chapter 2, the appropriateness of this framework is highly questionable in many economic applications. Because of changes that often occur in the structure of the economy and important institutions in the economy, econometricians need to be mindful of the possibility of a structural change although there can be considerable uncertainty as to the timing of the changepoint. It is therefore desirable to be able to detect a changepoint when the timing of the change is unknown. The use of hypothesis testing to do this has gained a great deal of attention from econometricians in recent years. A difficulty with this approach, as we have seen in Chapter 2, is that the timing of the change is a nuisance parameter, which is present only under the alternative hypothesis.

[1] The preliminary findings of this chapter were presented at the Third Annual Doctoral Research Conference, Faculty of Business and Economics, Monash University. See Azam and King (1997).

In this chapter we argue that the problem of detecting a changepoint of unknown timing can be viewed as a model selection problem. There is a history of the use of hypothesis testing to make decisions about model specification in the econometric literature (for more details, see Section 2.3.2). In this regard, Granger et al. (1995) argued model selection decisions should be based on some well-thought-out model selection procedure rather than a series of classical pairwise hypothesis tests.

The purpose of this chapter is to investigate the use of IC model selection procedures to detect a structural change when the changepoint is unknown. To the best of our knowledge, this has not yet been investigated in the literature. A disadvantage of this approach is that we have many different models that have to be estimated, one for each different timing of the possible changepoint, and we are not sure about the quality of inferences from model selection procedures in cases where a high number of models are involved. In particular, our aim is to find which criteria among existing IC has the best ability to detect a changepoint in the context of a linear regression model when the timing of the changepoint is unknown. The first and foremost aim of this chapter is to see if model selection can be successfully used in this case.

A comprehensive investigation of the application of all possible IC procedures to the problem of detecting structural change with unknown timing is not feasible because of the size of the task. Therefore, we have had to narrow the number of criterion functions. Under these constraints, we use a unified strategy to find a best choice of IC model selection; that is, we choose a number of prominent IC (among AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC) and apply them to simulated DGPs. We use as our measure of the ability of a criterion to detect a

changepoint the average mean probability of correct selection (AMPCS) when the model is being selected from a group of alternative models.

The plan of this chapter is as follows. Section 4.2 looks at the use of model selection procedures to detect structural change and discusses our measure of AMPCS as a method of assessing different procedures. In Section 4.3 the design of our Monte Carlo experiment is provided. Section 4.4 discusses the results of this experiment. Concluding remarks are given in Section 4.5.

4.2 Model Selection Procedures to Detect Structural Change

Consider the following multiple linear regression model

$$y_t = x_t'\beta + z_t\gamma_1 + \varepsilon_t \tag{4.1}$$

where y_t is an observation on the dependent variable at time t, x_t is a $k \times 1$ vector of regressors at time t, z_t is a scalar variable, β is a $k \times 1$ vector of regressor coefficients, and γ_1 is an unknown scalar parameter. The residuals ε_t of the n observations are assumed to be independent and identically normally distributed with mean zero and a constant variance σ^2. Our interest is in detecting whether γ_1 has changed in value at some point in time. This means we are interested in selecting between the following n different models:

$$\begin{aligned}
&\text{Model}-1 \quad && y_t = x_t'\beta + z_t\gamma_1 + \varepsilon_t && \text{for} \quad t = 1, 2, \ldots, n, \\
&\text{Model}-2 \quad && y_t = x_t'\beta + z_t\gamma_1 + z_{D1t}\gamma_2 + \varepsilon_t && \text{for} \quad t = 1, 2, \ldots, n, \\
&\qquad\vdots \qquad && \qquad\vdots && \qquad\qquad (4.2) \\
&\text{Model}-n \quad && y_t = x_t'\beta + z_t\gamma_1 + z_{D(n-1)t}\gamma_2 + \varepsilon_t && \text{for} \quad t = 1, 2, \ldots, n,
\end{aligned}$$

where $z_{D\#}$, $i = 1, \ldots, n$, denotes zeros up to the ith observation and z_i afterwards.

In obvious matrix notation, model-i can be rewritten as

$$y = X_i\Gamma + \varepsilon \tag{4.3}$$

where X_1 represents the $n \times (k+1)$ matrix of observations on the regressors in model-1 in the case of no changepoint; X_i $i = 2, \ldots, n$, represents the $n \times (k+2)$ matrix of observations on the regressors in model-i when there is a changepoint in the observations. In this case, X_i consists of k columns made up of the rows of x'_t, $t = 1, \ldots, n$, a column of z_i values and an additional column consisting of zeros up to the ith element and z_i afterwards. Also Γ is a $(k+2) \times 1$ [$(k+1) \times 1$ in the case of model-1] vector of coefficients.

The log-likelihood function of the model is

$$L_i(\theta) = L(\Gamma, \sigma^2, y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - X_i\Gamma)'(y - X_i\Gamma) \tag{4.4}$$

and is maximised for $\hat{\Gamma} = (X_i'X_i)^{-1}X_i'y$, the ordinary least squares (OLS) estimator of Γ, and $\hat{\sigma}_i^2 = (y - X_i\hat{\Gamma})'(y - X_i\hat{\Gamma})/n$, $i = 1, \ldots, n$. Therefore, under the null hypothesis of no changepoint in the model (i.e., model-1), the maximised log-likelihood is

$$L_1(\hat{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\hat{\sigma}_1^2 - \frac{n}{2} \tag{4.5}$$

where $\hat{\sigma}_1^2 = (y - X_1\hat{\Gamma})'(y - X_1\hat{\Gamma})/n$ is the estimated error variance of the model without a changepoint. On the other hand, also under the alternative hypothesis of a changepoint in the model (i.e., model-i), the maximized log-likelihood is

$$L_i(\hat{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\hat{\sigma}_i^2 - \frac{n}{2} \tag{4.6}$$

where $\hat{\sigma}_i^2 = (y - X_i\hat{\Gamma})'(y - X_i\hat{\Gamma})/n$; $i = 2, \ldots, n$, is the estimated error variance calculated after modelling the changepoint.

In the simulation study reported below, we included only $m = (n-4)$ changepoint models because it is extremely hard to detect models when there is a changepoint right at the beginning or right at the end of the data period. Therefore, we dropped off models 2 and 3 that have changepoints at the beginning and also models $n-1$ and n which have their changepoints at the end of the data period.

Let $L_1(\hat{\theta}), \cdots, L_m(\hat{\theta})$ correspond to the maximised log-likelihoods of the models $M_1, \cdots, M_m$, respectively, with $\hat{\theta}$ being the maximum likelihood estimate of the parameter vector θ. Here M_1 corresponds to the model without structural change, M_2 corresponds to the model with a changepoint at the third time period and so on, with M_m corresponding to the model with a changepoint at the $(n-2)$ th period.

In Chapter 2 we discussed various IC based model selection procedures. The usual form of almost all IC based model selection procedures is to select the model with the largest penalised maximised log-likelihood function, namely

$$IC_i = L_i\left(\hat{\theta}\right) - p_i \tag{4.7}$$

where p_i is the penalty function for the i th model, M_i, for $i = 1, \cdots, m$, dependent on the number of parameters, among other things. For example, p_i in case of AIC takes the value k_i, BIC takes the value $k_i \log(n)/2$, HQC takes the value $-k_i \log(\log(n))/n$, RSC takes the value $-n \log((n-k_i))/2$, GCVC takes the value $-n \log((n-k_i)/n)$, HSPC takes the value $-n \log((n-k_i)(n-k_i-1))/2$ and MCPC takes the value $n \log(1+2k_i/(n-k^*))/2$ where k^* is the number of free parameters in the smallest model that includes all models under consideration as special cases.

For each model under consideration and for a given choice of penalty function, we estimate the probability of correctly selecting this model when it is indeed the true model. Under repeated sampling we count how many times these models come up as the true model and then divide the total count by the number of repetitions which gives us an estimate of the mean probability of correct selection (MPCS). For the same penalty set, we then average these probabilities across the different models to find the average mean probability of correct selection (AMPCS). In order to understand more closely what is involved, let $\Pr(CSM_j|M_j,\theta_j, p_1, \ldots, p_m)$ denote the probability of correctly selecting Model-j when it is true with parameter vector θ_j and using penalties $p_1, \ldots, p_m$. This probability can be given by

$$\begin{aligned}&\Pr(CSM_j|M_j,\theta_j, p_1, \ldots, p_m)\\ &\quad = \Pr(L_j(\hat{\theta}_j) - p_j > L_i(\hat{\theta}_i) - p_i, i \neq j, i = 1,\cdots,m \,|\, M_j,\theta_j).\end{aligned} \tag{4.8}$$

The problem with this probability is that it is not fixed but varies as θ_j. An alternative way to overcome this difficulty is to work with the MPCS for the j th model which is given by

$$MPCS_j(p_1, \ldots, p_m) = \int \Pr(CSM_j|M_j,\ \theta_j; p_1, \ldots, p_m) f(\theta_j)\, d\theta_j \tag{4.9}$$

where $f(\theta_j)$ is a weighting density function rather like a prior density function used in Bayesian methods. Its purpose is to weight different parameter vector values when calculating MPCS. Further, we can take the average of MPCS to obtain the overall AMPCS over m models and for a given set of penalty values $p_1,\cdots,p_m$, namely,

$$AMPCS = \frac{1}{m}\sum_{j=1}^{m}\int \Pr(CSM_j|M_j,\ \theta_j, p_1, \ldots, p_m) f(\theta_j)\, d\theta_j. \tag{4.10}$$

Unfortunately, it is very difficult to calculate (4.9) and (4.10) analytically. To make our approach operational, we need a method of estimating (4.9). Given that $f(\theta_j)$ is a joint density function, the Monte Carlo estimate of (4.9) can be found by taking a large sample of drawings from the distribution represented by $f(\theta_j)$, denoted by $\theta(i)$, $i = 1, \ldots, R$, where R is the number of drawings and then calculating

$$\frac{1}{R}\sum_{i=1}^{R} \Pr(CSM_j | M_j, \theta(i), p_1, \ldots, p_m). \tag{4.11}$$

This then requires estimating $\Pr(CSM_j | M_j, \theta(i), p_1, \ldots, p_m)$ for given M_j, $\theta(i)$ and $p_1, \ldots, p_m$ by a Monte Carlo simulation. Some recent research, for example, King and Bose (2000), in this context has confirmed that for a fixed total number of replications, good results are achieved by using only one replication but the maximum number of drawings of $\theta(i)$ from $f(\theta_j)$. In the following section we discuss a procedure for estimation of MPCS and consequently the AMPCS.

4.3 Procedure for Estimation of AMPCS

The selection of a best information criterion among existing IC can be based on the relative performance of the criteria or the power to pick the correct model. Obviously, the best choice of IC may differ from model to model, with the dimensions of the models or the values of the parameters, sample sizes and the timing of changepoints in the data. The evaluation of these IC procedures is based on the AMPCS discussed in the previous section. Therefore, we propose to take the average of the mean probability of correctly choosing models for a given set of penalty values $p_1, \ldots, p_m$ as our measure of accuracy of the resultant IC procedure.

Our Monte Carlo study involves simulating models in the presence of structural change. For data simulated from a particular DGP, maximizing the log-likelihood function fits the different models. Estimation is by maximum likelihood (ML), under the assumption that $\varepsilon_i \sim IN(0,\sigma^2)$ so that the likelihood is easily specified. Table 4.1 gives all the possible outcomes of the fitted maximized log-likelihood values, given the true models. The maximized log-likelihood function is denoted by $L_{ij}(\hat{\theta})$, where i denotes "fitted" for model $i = 1, \ldots, m$ and j denotes M_j is the true model for $j = 1, \ldots, m$. For example, a typical element in the first row and second column of the matrix of all possible outcomes, $L_{12}(\hat{\theta})$, denotes that M_2 has been fitted when the actual true model is M_1. The diagonal elements of the matrix would give the outcomes of the correctly fitted models. The upper and lower triangles of the off-diagonals are outcomes of wrongly fitted models.

Table 4.1 Experimental Design: Values of the Maximized Log-likelihood

True	Fitted Model		
Model	M_1	M_2	$\ldots M_m$
M_1	$L_{11}(\hat{\theta})$	$L_{12}(\hat{\theta})$	$\ldots L_{1m}(\hat{\theta})$
M_2	$L_{21}(\hat{\theta})$	$L_{22}(\hat{\theta})$	$\ldots L_{2m}(\hat{\theta})$
	⋮	⋮	⋮
M_m	$L_{m1}(\hat{\theta})$	$L_{m2}(\hat{\theta})$	$\ldots L_{mm}(\hat{\theta})$

We then penalize each of these estimated maximized log-likelihoods with a penalty term from a particular IC and compute the number of times the true model is selected. Evaluating (4.9) requires random drawings of θ_j from $f(\theta_j)$, for model M_j in order to compute this MPCS. For each of the m models, R random

drawings of θ_j are obtained from $f(\theta_j)$ and then for each drawing the model (4.3) is used to generate a dependent variable y. Then the log-likelihood functions for each of the models is maximized and the maximized values are stored. This is repeated for each model so that a file of Rm^2 maximized likelihoods is generated. This can then be used to estimate (4.11) for different values of $p_1, \ldots, p_m$. The selection process can be represented by an indicator function:

$$\begin{aligned} I_j(p_1, \ldots, p_m; \theta(i)) &= 1 \quad \text{(when model } j \text{ is chosen)} \\ &= 0 \quad \text{(when model } j \text{ is not chosen).} \end{aligned} \tag{4.12}$$

Given this first set of parameter values for true model M_j, $j = 1, \ldots, m$, we can keep a count of the number of times each of the true models is selected. As we are only interested in correct selection, we aggregate the I_js for correct selection of the true models by summing all the ones. Because there are a total of m competing models, and a total number of R replications, we thus can obtain the MPCS over m models and R replications for a given set of penalty values $p_1, \ldots, p_m$. That is, using indicator functions, (4.12) can be estimated by

$$MPCS(M_j|p_1, \ldots, p_m) = \frac{\sum_{i=1}^{R} I_j(p_1, \ldots, p_m; \theta(i))}{R}. \tag{4.13}$$

and (4.10) can be estimated as

$$AMPCS = \frac{\sum_{j=1}^{m} MPCS(M_j|p_1, \ldots, p_m)}{m}. \tag{4.14}$$

4.4 The Monte Carlo Experiment

In order to find the best model selection procedure for detecting linear regression models with a structural change of unknown timing, we conducted a Monte Carlo experiment. The experiment aimed to evaluate the relative performance of some existing IC model selection procedures (namely, AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC) in terms of AMPCS.

4.4.1 Experimental Design

Our aim is to estimate the probabilities of correct selection for each of the models as the true DGP and for a range of different procedures. We are able to compare the results for different IC procedures and select the best procedure that gives the largest average mean probability of correct selection. We performed the following simulation:

The y's are generated from the following equation

$$y_t = \beta_1 + \beta_2 x_t + \gamma_1 z_t + \gamma_2 z_{Dt} + \varepsilon_t \tag{4.14}$$

where $\varepsilon_t \sim IN(0,\sigma^2)$. The regressors chosen here are influenced by Engle et al.'s (1986) Monte Carlo study; that is, x_t is generated from the AR(1) process $x_t = \phi x_{t-1} + u_t$, where $u_t \sim IN(0,\sigma^2)$.

We set $\phi = 0, 0.7, 1.0$ and 1.2 which covers white noise, autoregressive, random walk and explosive processes, respectively. In each case, x_t is generated artificially and held fixed from replication to replication. z_t is generated from the uniform distribution ranging from 0 to 1. z_{Dt} takes the value zero up to and including the changepoint and z_t afterwards. Five different samples sizes of

15, 25, 50, 75 and 100 were used for comparison purposes. For the estimation of AMPCS, we considered 11 models when $n = 15$, 21 models when $n = 25$, 46 models when $n = 50$, 71 models for $n = 75$ and 96 models for $n = 100$.

For each model M_j, note that all MPCS are invariant to the values taken by β_1, β_2, γ_1 and depend only on γ_2 / σ. Consequently we set $\beta_1 = 1$, $\beta_2 = 1$, $\sigma = 1$, $\gamma_1 = 1$. The $f(\theta_j)$ prior distribution was only required for $\theta = \gamma_2 / \sigma$. As we have seen, the MPCS is very sensitive to the choice of values for θ. Therefore, we would like to draw parameter values from a distribution, which allows each parameter to take a range of realistic values. Based on this argument, we decided to use a uniform distribution so that values are drawn from a uniform spread of small, medium and large values. The uniform distribution seems to be a natural choice but there is the question of how to choose the limits of the distribution. We have seen that if the bound of the distribution is very large, the AMPCS tends to one, and if the bound is very small the AMPCS tends to zero. Therefore, we choose parameter values that give some randomness in the selection, but provide a good coverage of admissible parameter values. This approach resulted in θ being generating uniformly from the interval -10 to 10.

Throughout, when ML estimation was needed it was conducted using the GAUSS 3.2.12 software. All simulations were carried out using $R = 2{,}000$ replications for five different samples sizes mentioned earlier. All the models were simulated using pseudo random numbers from the GAUSS function RNDNS that generates standard normal variates for regression errors. The seed for generating random numbers for each experiment was 1786. To evaluate our model selection criterion, we averaged our estimated MPCS to three decimals.

4.5 Results of the Monte Carlo Data Analysis

Tables 4.2 – 4.26 contain the Monte Carlo results for the selection probabilities of AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for different sample sizes (i.e., $n = 15, 25, 50, 75$ and 100), and different values of the autoregressive coefficient for the exogenous regressor ($\phi = 0, 0.7, 1.0$ and 1.2). We summarize the probabilities of correct model selection by each of the IC methods along with their average mean probability of correct selection when there is no changepoint and when there is a changepoint. In terms of overall probabilities, we calculate the average of MPCS for no changepoint models and changepoint models for different sample sizes and different values of ϕ.

The performance of the IC procedures in the context of changepoint and no changepoint models will be discussed in this section. Tables 4.2 to 4.26 contain calculated MPCS and AMPCS for changepoint and no changepoint models for different values of ϕ for the autoregressive exogenous variable. In order to discuss this massive set of results, we have presented the results for individual MPCS for each model for different values of ϕ, and different sample sizes for each changepoint model. We then computed AMPCS along with their standard deviation for different models for different IC procedures. These AMPCS, standard deviations and ranking of different IC procedures were also tabulated for different sample sizes and different ϕ's. From these general results we will proceed to take a closer look at the performance of the IC procedures and will highlight any peculiar cases. But first, some general results.

4.5.1 Results for Models with Changepoints

4.5.1.1 For $n=15$

For models with changepoints and for all values of ϕ, the MPCS of the seven IC procedures considered here show a general downward trend as the changepoint moves forward in time. Also as ϕ increases from 0 to 1.02, there is a slight tendency for the AMPCS to decrease, ceteris paribus.

A closer examination of the results shows that the AMPCS of all the seven IC procedures are very similar with the difference of AMPCS between any pair of IC procedures being no more than 0.0500, 0.0429, 0.0728 and 0.0630, respectively, for ϕ values of 0, 0.7, 1.0 and 1.02. Overall, the difference of AMPCS between the best and worst IC procedure is 0.0527 irrespective of the effect of different ϕ values.

Overall, RSC is statistically the best choice among the IC procedure for all ϕ values considered, because it gives the largest AMPCS. The ranking of the other procedures are AIC ranked second, MCPC ranked third, HQC ranked fourth, GCVC ranked fifth, HSPC ranked sixth and BIC ranked last. We can group these IC procedures into three major groups based on similar performances, these are RSC and AIC as the best group, MCPC, HQC and GCVC as the second best group, and, HSPC and BIC as the worst group.

4.5.1.2 For $n=25$

In the presence of structural change, the MPCS of the various IC procedures indicate a general downward tendency as the changepoint moves forward in time, with a few exceptions. A closer look at the pattern reveals very low MPCS when the changepoint is at the beginning or the end of the time period. The MPCS is highest when the changepoint is situated in the middle of the data period.

When the ϕ value is 0, 0.7, 1.0 and 1.02, then the AMPCS of all the seven IC procedures are very similar with the difference of AMPCS between any two IC procedures being with no more than 0.0700, 0.0526, 0.0620 and 0.0834, respectively. The difference of the overall mean of the AMPCS between the best and worst IC procedure is 0.0670 irrespective of the effect of different ϕ values.

The overall ranking of the IC procedures are RSC ranked first, AIC second, MCPC third, HQC fourth, GCVC fifth, HSPC sixth and BIC seventh. The three major groupings in terms of performance are the same as for $n=15$, namely RSC and AIC as the best group, MCPC, HQC and GCVC as the second best group, and, HSPC and BIC as the worst group.

4.5.1.3 For $n=50$

For models with changepoints and for all values of ϕ, the MPCS of the seven IC procedures considered here show almost a bi-modal pattern as the changepoint moves forward in time. When the ϕ value is 0, the difference of AMPCS between any two IC procedures is no more than 0.0595. Also when the ϕ value is 0.7, 1.0 and 1.02 then the AMPCS of all the seven IC procedures are similar with the difference in AMPCS between any two IC procedures being no more than 0.1003, 0.1079 and 0.1196, respectively. The difference of the overall mean of the AMPCS between the best and worst IC procedure is 0.0968 irrespective of the effect of different ϕ values.

The ranking of IC procedures is the same as for sample size 25. We are able to group these IC procedures into four major groups based on their performances, these being RSC as the best group, AIC, MCPC and HQC as the second best group, GCVC and HSPC as the third best group and BIC as the worst group.

4.5.1.4 For $n=75$

In the presence of a structural change, the MPCS of the various IC procedures indicate a general upward tendency as the changepoint moves forward in time, with a few exceptions. When the ϕ value is 0, 0.7, 1.0 and 1.02, then the difference in AMPCS between any two IC procedures is no more than 0.0501, 0.0696, 0.1161 and 0.1073, respectively. The difference of the overall mean of the AMPCS between the best and worst IC procedure is 0.0858 irrespective of the effect of different ϕ values.

The ranking of IC procedures is the same as for sample size 50. We again can group these IC procedures into four major groups based on their performances, these being RSC and AIC as the best group, MCPC and HQC as the second best group, GCVC and HSPC as the third best group and BIC as the worst group.

4.5.1.5 For $n=100$

In the presence of structural change, for different ϕ values, the MPCS of the seven IC procedures considered here indicate a general downward trend with greater MPCS values observed at the beginning of the data period with a gradual decrease as the changepoint moves forward in time. A closer examination of the results indicates that when the ϕ value is 0, 0.7, 1.0 and 1.02, then the difference in AMPCS between any two IC procedures is no more than 0.0930, 0.0819, 0.1215 and 0.1219, respectively. The difference of the overall mean of the AMPCS between the best and worst IC procedure is 0.1046 irrespective of the effect of different ϕ values.

The ranking of IC procedures is the same as for sample size 50. We can group these IC procedures into three major groups based on their performances, these

being RSC, AIC and MCPC as the best group, HQC, GCVC and HSPC as the second best group and BIC making up the worst group.

4.5.1.6 For $\phi = 0$

In the presence of a structural change for an autoregressive exogenous regressor with a coefficient of $\phi = 0$, the performance of different IC procedures in terms of AMPCS is consistently increasing as the sample size increases from 15 to 25, 50, 75 and 100. That is, the AMPCS generally increases with an increase in sample size. When the n value is 15, 25, 50, 75 and 100, then the difference in AMPCS between any two IC procedures is no more than 0.0500, 0.0700, 0.0595, 0.0501 and 0.0930, respectively. The difference of the overall mean of the AMPCS between the best and worst IC procedure is 0.0645 irrespective of the effect of different n values.

Overall, RSC is statistically the best choice among the IC procedures for all n values considered, because it gives the largest AMPCS. The rankings of the other procedures are AIC ranked second, MCPC third, HQC fourth, GCVC fifth, HSPC sixth and BIC seventh. We can group these IC procedures into three major groups based on their performances, these being RSC, AIC and MCPC as the best group, HQC, GCVC and HSPC as the second best group, and BIC as the worst group.

4.5.1.7 For $\phi = 0.70$

In the presence of a structural change for an autoregressive exogenous regressor with coefficient $\phi = 0.7$, the performance of different IC procedures in terms of AMPCS is consistently increasing as the sample size increases from 15 to 100. When the n value is 15, 25, 50, 75 and 100 then the difference in AMPCS between any two IC procedures is no more than 0.0429, 0.0526, 0.1003, 0.0696 and

0.0819, respectively. The difference of the overall mean of the AMPCS between the best and worst IC procedure is 0.0695 irrespective of the effect of different n values.

The ranking of IC procedures is the same as for $\phi = 0$. We can group these IC procedures into four major groups based on their performances, these being RSC as the best group, AIC, MCPC and HQC, as the second best group, GCVC and HSPC as the third best group and BIC as the worst group.

4.5.1.8 For $\phi = 1.0$

In the presence of a structural change for an autoregressive exogenous regressor with coefficient $\phi = 1.0$, the performance of different IC procedures in terms of AMPCS is consistently increasing as the sample sizes increase from 15 to 100. When the n value is 15, 25, 50, 75 and 100 then the difference in AMPCS between any two IC procedures is no more than 0.0728, 0.0620, 0.1079, 0.1161 and 0.1215, respectively. The difference of the overall mean of the AMPCS between the best and worst IC procedure is 0.0961 irrespective of the effect of different n values.

The ranking of IC procedures is the same as for $\phi = 0.7$. We can group these IC procedures into four major groups based on their performances, these being RSC as the best group, AIC, MCPC and HQC, as the second best group, GCVC and HSPC as the third best group and BIC as the worst group.

4.5.1.9 For $\phi = 1.02$

In the presence of a structural change for an autoregressive exogenous regressor with coefficient $\phi = 1.02$, the performance of the different IC procedures in terms of AMPCS is consistently increasing as the sample size increases from 15 to 100. When the n value is 15, 25, 50, 75 and 100 then the difference in AMPCS

between any two IC procedures is no more than 0.0630, 0.0834, 0.1196, 0.1073 and 0.1219 respectively. The difference of the overall mean of the AMPCS between the best and worst IC procedure is 0.0990 irrespective of the effect of different n values.

The ranking of IC procedures is the same as for $\phi = 1.0$. We can group these IC procedures into four major groups based on their performances, these being RSC as the best group, AIC, MCPC and HQC, as the second best group, GCVC and HSPC as the third best group and BIC as the worst group.

4.5.2 Results for Models without Changepoints

4.5.2.1 For $n=15$

For models without changepoints and for all values of ϕ, the MPCS of the seven IC procedures considered here show an increasing trend as the ϕ value increases from 0 to 1.02, except for BIC, which has a decreasing trend, ceteris paribus. A closer examination of the results shows that the difference of MPCS between any pair of IC procedures is no more than 0.1230, 0.0831, 0.0755 and 0.0825 respectively, for ϕ values of 0, 0.7, 1.0 and 1.02. The difference of the overall mean of MPCS between the best and worst IC procedure is 0.0910 irrespective of the effect of different ϕ values.

Overall, BIC is statistically the best choice among the IC procedures for all ϕ values considered, because it gives the largest MPCS when there is no changepoint. The ranking of the other procedures are HSPC second, GCVC third, HQC fourth, MCPC fifth, AIC sixth and RSC seventh. We can group these IC procedures into three major groups based on their performances, these being BIC, HSPC and GCVC as the best group, HQC and MCPC as the second best group, and AIC and RSC as the worst group.

4.5.2.2 For $n=25$

When no changepoints are present and for all values of ϕ, the MPCS of the seven IC procedures considered here show an increasing trend as the ϕ value increases from 0 to 1.02, except for BIC, which exhibits a decreasing trend, ceteris paribus. A closer examination of the results shows that the difference in MPCS between any pair of IC procedures is no more than 0.1385, 0.0770, 0.0846 and 0.0795 for ϕ values of 0, 0.7, 1.0 and 1.02, respectively. The difference of the overall mean of MPCS between the best and worst IC procedure is 0.0949 irrespective of the effect of different ϕ values. The ranking of IC procedures is the same as for $n=15$. We can group these IC procedures into three major groups based on their performances, these being BIC and HSPC as the best group, GCVC, HQC and MCPC as the second best group, and AIC and RSC as the worst group.

4.5.2.3 For $n=50$

For models without changepoints and for all values of ϕ, the MPCS of the seven IC procedures considered here show the same pattern as for $n=25$. A closer examination of the results shows that the difference in MPCS between any pair of IC procedures is no more than 0.1277, 0.0821, 0.0811 and 0.0813 for ϕ values of 0, 0.7, 1.0 and 1.02, respectively. The difference in the overall mean of MPCS between the best and worst IC procedure is 0.0931 irrespective of the effect of different ϕ values.

The ranking of IC procedures is the same as for $n=15$. We can group these IC procedures into four major groups based on their performances; these groups are BIC as the best group, HSPC, GCVC and HQC as the second best group, MCPC and AIC as the third best group, and RSC as the worst group.

4.5.2.4 For $n=75$

In the presence of no structural change and for all values of ϕ, the MPCS of the seven IC procedures considered here show the same pattern as for $n=50$. A closer examination of the results shows that the difference in MPCS between any pair of IC procedures is no more than 0.1328, 0.0870, 0.0861 and 0.0778 for ϕ values of 0, 0.7, 1.0 and 1.02, respectively. The difference of the overall mean of MPCS between the best and worst IC procedure is 0.0959 irrespective of the effect of different ϕ values.

The ranking of IC procedures is the same as for $n=15$. We can group these IC procedures into three major groups based on their performances, these being BIC and HSPC as the best group, GCVC, HQC and MCPC as the second best group, and AIC and RSC as the worst group.

4.5.2.5 For $n=100$

For models without changepoints and for all values of ϕ, the MPCS of the seven IC procedures considered here show the same pattern as for $n=75$. A closer examination of the results shows that the difference in MPCS between any pair of IC procedures is no more than 0.1083, 0.0804, 0.0749 and 0.0832 for ϕ values of 0, 0.7, 1.0 and 1.02, respectively. The difference of the overall mean of MPCS between the best and worst IC procedure is 0.0867 irrespective of the effect of different ϕ values.

The ranking of IC procedures is the same as for $n=75$. We can group these IC procedures into four major groups based on their performances, these being BIC as the best group, HSPC, GCVC and HQC as the second best group, MCPC and AIC as the third best group, and RSC as the worst group.

4.5.2.6 For $\phi = 0$

When there is no structural change for an autoregressive exogenous regressor with coefficient $\phi = 0$, the performance of different IC procedures in terms of MPCS are consistently increasing as the sample sizes increase from 15 to 25, 50, 75 and 100. That is, the MPCS generally increases with an increase in sample size. When the n value is 15, 25, 50, 75 and 100 then the difference in MPCS between any two IC procedures is no more than 0.1230, 0.1385, 0.1277, 0.1328 and 0.1083 respectively. The difference in the overall mean of the MPCS between the best and worst IC procedure is 0.1261 irrespective of the effect of different n values.

Overall, BIC is statistically the best choice among the IC procedures for all n values considered, since it gives the largest MPCS when there is no structural change. The ranking of the other procedures are HSPC second, GCVC third, HQC fourth, MCPC fifth, AIC sixth and RSC seventh. We can group these IC procedures into four major groups based on their performances, these being BIC as the best group, HSPC GCVC and HQC as the second best group, MCPC and AIC as the third best group and RSC as the worst group.

4.5.2.7 For $\phi = 0.70$

In the presence of no structural change for an autoregressive exogenous regressor with coefficient $\phi = 0.7$, the performance of different IC procedures in terms of MPCS is the same as for $\phi = 0$ when the sample size increases from 15 to 25, 50, 75 and 100. When the n value is 15, 25, 50, 75 and 100 then the difference in MPCS between any two IC procedures is no more than 0.0831, 0.0770, 0.0821, 0.0870 and 0.0804, respectively. The difference of the overall mean of MPCS

between the best and worst IC procedure is 0.0819 irrespective of the effect of different n values.

The ranking of IC procedures is the same as for $\phi = 0$. We can group these IC procedures into three major groups based on their performances; these groups are BIC, HSPC and GCVC as the best group, HQC, MCPC and AIC as the second best group and RSC as the worst group.

4.5.2.8 For $\phi = 1.0$

In the presence of no structural change for an autoregressive exogenous regressor with coefficient $\phi = 1.0$, the performance of different IC procedures in terms of MPCS is the same as for $\phi = 0.7$ when the sample size increases from 15 to 25, 50, 75 and 100. When the n value is 15, 25, 50, 75 and 100, then the difference in MPCS between any two IC procedures is no more than 0.0755, 0.0846, 0.0811, 0.0861 and 0.0749, respectively. The difference of the overall mean of MPCS between the best and worst IC procedure is 0.0804 irrespective of the effect of different n values.

The ranking of IC procedures is the same as for $\delta = 0.7$. We can group these IC procedures into three major groups based on their performances; these groups are BIC, HSPC and GCVC as the best group, HQC, MCPC and AIC as the second best group and RSC as the worst group.

4.5.2.9 For $\phi = 1.02$

In the presence of no structural change for an autoregressive exogenous regressor with coefficient $\phi = 1.02$, the performance of different IC procedures in terms of MPCS is the same as for $\phi = 1.0$ when the sample size increases from 15 to

25, 50, 75 and 100. When the n value is 15, 25, 50, 75 and 100 then the MPCS between any two IC procedures is no more than 0.0825, 0.0795, 0.0813, 0.0778 and 0.0832, respectively. The difference of the overall mean of MPCS between the best and worst IC procedure is 0.0809 irrespective of the effect of different n values.

The ranking of IC procedures is the same as for $\phi = 1.0$. We can group these IC procedures into three major groups based on their performances, these being BIC, HSPC and GCVC as the best group, HQC, MCPC and AIC as the second best group and RSC as the worst group.

4.5.2.10 Results Based on AMPCS

When we do not have any knowledge of the presence or absence of a structural change in the data, for small samples (15 and 25) the overall performance of different IC procedures in terms of AMPCS, HSPC is statistically the best choice among the IC procedures. The ranking of the other procedures are GCVC second, BIC third, MCPC fourth, AIC fifth, HQC sixth and RSC seventh. On the other hand, when the sample size is large (more than 50), the overall performance of different IC procedures in terms of AMPCS is that BIC is statistically the best choice among the IC procedures. The ranking of the other procedures are HQC second, AIC third, HSPC fourth, MCPC fifth, GCVC sixth and RSC seventh.

4.6 Concluding Remarks

This study investigated the relative performance of IC model selection procedures when detecting the possible presence of a structural change. It was limited in scope, which suggests that one has to take care in making too many generalized conclusions from the results. In particular, because it was based on a Monte Carlo study, the results are specific to the choice of models and, within that set of models, the choice of the autoregressive exogenous regressor coefficient ϕ, sample size n etc. used. However, subject to these limitations, there are a number of conclusions that may be reached. Overall, the important conclusions are:

i. BIC outperformed all existing IC procedures considered when there is no structural change but its performance is the worst of all procedures in the presence of structural change.

ii. RSC's performance is the worst of all existing IC procedures in the presence of no structural change however it outperformed all other IC procedures considered when there is structural change.

iii. Based on AMPCS, HSPC is the best IC procedures for small samples and BIC for large samples.

iv. In the presence of no structural chance, the ranking of the relative performance of the other IC procedures is HSPC > GCVC > HQC > MCPC > AIC.

v. In the presence of structural change, the ranking of the relative performance of the other IC procedures is AIC > MCPC > HQC > GCVC > HSPC.

vi. The AMPCS of all IC procedures decreases as the ϕ value increases, ceteris paribus.

vii. The AMPCS of all IC procedures increases as the sample size increases, ceteris paribus.

viii. The performance of two groups of IC procedures, that is, BIC, HSPC and GCVC, and HQC, MCPC and AIC are more or less the same.

ix. None of the IC procedures considered stands out as a clear best method for modelling involving structural change.

x. As we use more sample observations, we deal with a greater number of models. As we deal with more models, the average probability of choosing the true model tends to increase.

xi. Models with changepoints at the beginning or at the end of data points give comparatively high MPCS.

This chapter demonstrates that model selection procedures can be applied to detect a possible changepoint in the data. As a policy lesson, this chapter cautions against the use of one particular IC procedure in order to detect the presence of a structural change because none of the IC procedures stand out as a clear best method. Findings showed that BIC outperformed all existing IC procedures considered when there was no structural change, and RSC performed best overall when a changepoint was present. When there is no structural change, the performance of AIC is generally the worst of all procedures. The AMPCS criterion summarizes the quality of different IC procedures and suggests HSPC is the best IC procedures for small samples and BIC for large samples. Clearly we could use AMPCS to determine an optimal penalty function that will maximize the probability of correctly selecting the true model on average. We will discuss this topic in detail in the next chapter.

Table 4.2 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting from 11 Models, $\phi = 0$, for $n = 15$[1].

Model[2]	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1130	0.2430	0.1115	0.1040	0.1505	0.2900	0.2675
2	0.7155	0.7090	0.7155	0.7190	0.7130	0.7015	0.7055
3	0.5750	0.5690	0.5755	0.5755	0.5745	0.5665	0.5655
4	0.5435	0.5400	0.5435	0.5445	0.5430	0.5365	0.5390
5	0.6430	0.6390	0.6430	0.6450	0.6430	0.6355	0.6375
6	0.6195	0.6155	0.6195	0.6260	0.6180	0.6110	0.6140
7	0.6320	0.6225	0.6320	0.6370	0.6300	0.6200	0.6210
8	0.6580	0.6505	0.6580	0.6640	0.6570	0.6460	0.6485
9	0.3825	0.3700	0.3825	0.3890	0.3805	0.3640	0.3675
10	0.4270	0.4170	0.4270	0.4335	0.4255	0.4125	0.4145
11	0.4475	0.4260	0.4475	0.4575	0.4455	0.4200	0.4220

Table 4.3 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 11 Models, $\phi = 0.70$, for $n = 15$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1630	0.3020	0.1620	0.1725	0.1985	0.3555	0.3275
2	0.7170	0.7105	0.7175	0.7220	0.7160	0.7080	0.7095
3	0.5665	0.5630	0.5665	0.5670	0.5655	0.5600	0.5620
4	0.5590	0.5560	0.5590	0.5600	0.5580	0.5530	0.5545
5	0.6445	0.6395	0.6445	0.6480	0.6435	0.6370	0.6395
6	0.6330	0.6300	0.6330	0.6360	0.6330	0.6290	0.6295
7	0.6575	0.6510	0.6575	0.6600	0.6560	0.6460	0.6490
8	0.6830	0.6745	0.6830	0.6885	0.6815	0.6725	0.6745
9	0.4235	0.4140	0.4235	0.4295	0.4225	0.4110	0.4130
10	0.4610	0.4530	0.4610	0.4655	0.4610	0.4485	0.4515
11	0.5220	0.5035	0.5220	0.5320	0.5195	0.4965	0.5010

1 Model 1 is the only model without changepoint

[2] Model 2 is for changepoint at 3rd observation, Model 3 changepoint at 4th observation, last model changepoint at $(n-2)^{th}$ observation and so on.

Table 4.4 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting from 11 Models, $\phi = 1.00$, for $n = 15$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1695	0.3080	0.1670	0.1076	0.2080	0.3565	0.3310
2	0.7025	0.6960	0.7025	0.7035	0.7005	0.6875	0.6915
3	0.5805	0.5755	0.5805	0.5815	0.5800	0.5740	0.5750
4	0.5625	0.5550	0.5625	0.5635	0.5605	0.5545	0.5550
5	0.6530	0.6495	0.6530	0.6535	0.6520	0.6475	0.6490
6	0.5905	0.5855	0.5905	0.5945	0.5890	0.5835	0.5845
7	0.6370	0.6340	0.6370	0.6395	0.6365	0.6305	0.6315
8	0.6800	0.6745	0.6800	0.6845	0.6790	0.6715	0.6730
9	0.4500	0.4440	0.4505	0.4555	0.4490	0.4405	0.4415
10	0.4740	0.4695	0.4740	0.4780	0.4720	0.4655	0.4690
11	0.5215	0.5065	0.5225	0.5320	0.5185	0.4995	0.5020

Table 4.5 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 11 Models, $\phi = 1.20$, for $n = 15$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.2050	0.3440	0.2025	0.1005	0.2405	0.3950	0.3685
2	0.7310	0.7235	0.7315	0.7330	0.7295	0.7185	0.7205
3	0.5780	0.5750	0.5780	0.5800	0.5770	0.5730	0.5740
4	0.5705	0.5650	0.5705	0.5715	0.5690	0.5640	0.5640
5	0.6585	0.6540	0.6585	0.6600	0.6580	0.6530	0.6540
6	0.6095	0.6065	0.6095	0.6110	0.6095	0.6045	0.6045
7	0.6415	0.6345	0.6415	0.6440	0.6385	0.6320	0.6330
8	0.6865	0.6810	0.6865	0.6890	0.6850	0.6760	0.6785
9	0.4210	0.4140	0.4215	0.4275	0.4195	0.4080	0.4105
10	0.4735	0.4680	0.4735	0.4780	0.4725	0.4630	0.4670
11	0.5265	0.5110	0.5265	0.5385	0.5225	0.5030	0.5080

Table 4.6 Average Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Different ϕ when $n = 15$.

ϕ		AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
0				With and without structural change				
	Average	0.5547	0.5615	0.5545	0.5485	0.5565	0.5627	0.5620
	Rank	5	3	6	7	4	1	2
	Stdev	0.1477	0.1170	0.1483	0.1735	0.1392	0.1084	0.1125
				With structural change				
	Average	0.5897	0.5833	0.5898	0.5933	0.5881	0.5795	0.5814
	Rank	3	5	2	1	4	7	6
	Stdev	0.0964	0.0972	0.0964	0.0945	0.0965	0.0981	0.0973
				Without structural change				
	Average	0.2050	0.3440	0.2025	0.1005	0.2405	0.3950	0.3685
	Rank	6	3	6	7	4	1	2
0.70				With and without structural change				
	Average	0.5475	0.5544	0.5473	0.5420	0.5495	0.5555	0.5548
	Rank	6	3	6	7	4	1	2
	Stdev	0.1491	0.1156	0.1497	0.1738	0.1393	0.1050	0.1104
				With structural change				
	Average	0.5853	0.5790	0.5853	0.5886	0.5837	0.5755	0.5772
	Rank	3	6	2	1	4	7	6
	Stdev	0.0850	0.0862	0.0850	0.0837	0.0854	0.0860	0.0861
				Without structural change				
	Average	0.1630	0.3020	0.1620	0.0725	0.1985	0.3555	0.3275
	Rank	6	3	6	7	4	1	2
1.0				With and without structural change				
	Average	0.5475	0.5544	0.5473	0.5420	0.5495	0.5555	0.5548
	Rank	6	3	6	7	4	1	2
	Stdev	0.1491	0.1156	0.1497	0.1738	0.1393	0.1050	0.1104
				With structural change				
	Average	0.5853	0.5790	0.5853	0.5886	0.5837	0.5755	0.5772
	Rank	3	6	2	1	4	7	6
	Stdev	0.0850	0.0862	0.0850	0.0837	0.0854	0.0860	0.0861
				Without structural change				
	Average	0.1695	0.3080	0.1670	0.0760	0.2080	0.3565	0.3310
	Rank	6	3	6	7	4	1	2
1.02				With and without structural change				
	Average	0.5233	0.5274	0.5232	0.5205	0.5255	0.5277	0.5277
	Rank	6	3	6	7	4	1	1
	Stdev	0.1723	0.1440	0.1726	0.1905	0.1632	0.1346	0.1392
				With structural change				
	Average	0.5644	0.5559	0.5644	0.5686	0.5630	0.5515	0.5537
	Rank	3	6	2	1	4	7	6
	Stdev	0.1113	0.1147	0.1113	0.1097	0.1115	0.1150	0.1151
				Without structural change				
	Average	0.1130	0.2430	0.1115	0.0395	0.1505	0.2900	0.2675
	Rank	6	3	6	7	4	1	2

Table 4.7 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 21 Models, $\phi = 0$, for $n = 25$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1270	0.3655	0.1915	0.2028	0.1435	0.2205	0.2090
2	0.3465	0.3195	0.3410	0.3520	0.3450	0.3375	0.3390
3	0.3735	0.3535	0.3695	0.3815	0.3720	0.3665	0.3675
4	0.6255	0.6095	0.6225	0.6305	0.6255	0.6195	0.6210
5	0.7030	0.6950	0.7015	0.7060	0.7020	0.7000	0.7000
6	0.6315	0.6265	0.6300	0.6330	0.6315	0.6290	0.6290
7	0.5265	0.5245	0.5265	0.5270	0.5265	0.5260	0.5265
8	0.5560	0.5535	0.5555	0.5580	0.5560	0.5555	0.5555
9	0.3325	0.3310	0.3325	0.3325	0.3325	0.3320	0.3320
10	0.3565	0.3550	0.3555	0.3570	0.3560	0.3550	0.3550
11	0.6460	0.6440	0.6460	0.6475	0.6460	0.6460	0.6460
12	0.4835	0.4825	0.4830	0.4835	0.4830	0.4830	0.4830
13	0.4885	0.4870	0.4880	0.4890	0.4885	0.4880	0.4880
14	0.7420	0.7395	0.7420	0.7440	0.7420	0.7420	0.7420
15	0.5330	0.5295	0.5325	0.5330	0.5330	0.5320	0.5320
16	0.5185	0.5150	0.5180	0.5200	0.5185	0.5170	0.5170
17	0.7100	0.7010	0.7085	0.7115	0.7100	0.7080	0.7080
18	0.5170	0.5105	0.5155	0.5205	0.5170	0.5155	0.5155
19	0.2695	0.2630	0.2685	0.2730	0.2695	0.2675	0.2675
20	0.2845	0.2810	0.2835	0.2855	0.2845	0.2830	0.2830
21	0.4410	0.4135	0.4350	0.4500	0.4395	0.4335	0.4340

Table 4.8 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 21 Models, $\phi = 0.70$, for $n = 25$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1255	0.3655	0.1930	0.1255	0.1365	0.2260	0.2130
2	0.3490	0.3260	0.3420	0.3550	0.3480	0.3385	0.3400
3	0.3555	0.3330	0.3535	0.3615	0.3555	0.3505	0.3515
4	0.6050	0.5915	0.6015	0.6140	0.6045	0.6005	0.6010
5	0.6990	0.6940	0.6985	0.7035	0.6990	0.6980	0.6985
6	0.6220	0.6175	0.6215	0.6245	0.6220	0.6210	0.6210
7	0.5470	0.5450	0.5465	0.5475	0.5470	0.5460	0.5465
8	0.5665	0.5655	0.5655	0.5685	0.5660	0.5655	0.5655
9	0.3345	0.3325	0.3340	0.3350	0.3345	0.3340	0.3340
10	0.3525	0.3515	0.3520	0.3535	0.3525	0.3515	0.3520
11	0.6505	0.6485	0.6500	0.6520	0.6505	0.6500	0.6500
12	0.4750	0.4740	0.4750	0.4750	0.4750	0.4750	0.4750
13	0.4885	0.4860	0.4885	0.4890	0.4885	0.4880	0.4885
14	0.7350	0.7315	0.7340	0.7360	0.7340	0.7340	0.7340
15	0.5160	0.5120	0.5160	0.5170	0.5160	0.5155	0.5155
16	0.4925	0.4905	0.4920	0.4930	0.4925	0.4915	0.4920
17	0.6890	0.6795	0.6875	0.6910	0.6885	0.6860	0.6865
18	0.5125	0.5050	0.5110	0.5150	0.5120	0.5011	0.5105
19	0.2680	0.2615	0.2670	0.2690	0.2675	0.2655	0.2660
20	0.2835	0.2775	0.2825	0.2840	0.2835	0.2810	0.2815
21	0.4445	0.4170	0.4400	0.4525	0.4440	0.4355	0.4380

Table 4.9 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 21 Models, $\phi = 1.0$ for $n = 25$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1299	0.3065	0.1420	0.1019	0.1060	0.1725	0.1615
2	0.3500	0.3345	0.3575	0.3530	0.3495	0.3455	0.3465
3	0.3775	0.3560	0.3735	0.3810	0.3775	0.3710	0.3715
4	0.6195	0.6055	0.6170	0.6240	0.6190	0.6155	0.6155
5	0.7105	0.7040	0.7095	0.7135	0.7100	0.7090	0.7090
6	0.6290	0.6255	0.6285	0.6305	0.6290	0.6275	0.6280
7	0.5555	0.5530	0.5550	0.5560	0.5555	0.5545	0.5545
8	0.5640	0.5635	0.5635	0.5645	0.5640	0.5635	0.5635
9	0.3310	0.3310	0.3310	0.3325	0.3310	0.3310	0.3310
10	0.3520	0.3515	0.3520	0.3530	0.3520	0.3515	0.3515
11	0.6465	0.6440	0.6455	0.6470	0.6460	0.6455	0.6455
12	0.4835	0.4825	0.4835	0.4840	0.4835	0.4835	0.4835
13	0.5005	0.5000	0.5000	0.5015	0.5005	0.5000	0.5000
14	0.7465	0.7420	0.7465	0.7480	0.7465	0.7465	0.7465
15	0.5220	0.5190	0.5220	0.5230	0.5220	0.5220	0.5220
16	0.5175	0.5135	0.5170	0.5175	0.5170	0.5170	0.5170
17	0.6995	0.6940	0.6985	0.7005	0.6990	0.6980	0.6980
18	0.4980	0.4920	0.4970	0.5020	0.4980	0.4965	0.4970
19	0.2670	0.2590	0.2650	0.2695	0.2665	0.2630	0.2645
20	0.2745	0.2680	0.2735	0.2760	0.2745	0.2725	0.2735
21	0.4235	0.3980	0.4175	0.4335	0.4225	0.4155	0.4160

Table 4.10 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 21 Models, $\phi = 1.02$ for $n = 25$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.2765	0.3289	0.3131	0.3016	0.3186	0.3155	0.3147
2	0.3550	0.3420	0.3520	0.3585	0.3540	0.3505	0.3510
3	0.3695	0.3560	0.3670	0.3735	0.3690	0.3640	0.3660
4	0.6160	0.5980	0.6135	0.6180	0.6160	0.6115	0.6125
5	0.7100	0.7050	0.7095	0.7110	0.7100	0.7080	0.7090
6	0.6295	0.6260	0.6280	0.6295	0.6295	0.6270	0.6275
7	0.5465	0.5450	0.5465	0.5470	0.5450	0.5460	0.5460
8	0.5640	0.5605	0.5640	0.5640	0.5640	0.5640	0.5640
9	0.3300	0.3290	0.3295	0.3305	0.3300	0.3295	0.3295
10	0.3555	0.3530	0.3555	0.3565	0.3555	0.3550	0.3555
11	0.6520	0.6485	0.6520	0.6525	0.6520	0.6510	0.6515
12	0.4800	0.4780	0.4800	0.4800	0.4800	0.4795	0.4800
13	0.4995	0.4965	0.4990	0.4995	0.4990	0.4980	0.4990
14	0.7440	0.7340	0.7435	0.7455	0.7440	0.7430	0.7430
15	0.5360	0.5325	0.5360	0.5385	0.5360	0.5355	0.5355
16	0.5275	0.5230	0.5270	0.5285	0.5270	0.5265	0.5265
17	0.7070	0.7000	0.7060	0.7085	0.7070	0.7040	0.7055
18	0.5015	0.4875	0.4990	0.5045	0.5015	0.4965	0.4970
19	0.2610	0.2520	0.2580	0.2625	0.2610	0.2575	0.2575
20	0.2615	0.2570	0.2605	0.2620	0.2610	0.2595	0.2600
21	0.3770	0.3510	0.3700	0.3850	0.3770	0.3675	0.3680

Table 4.11 Average Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Different ϕ when $n = 25$.

ϕ		AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
0				With and Without structural change				
	Average	0.4863	0.4905	0.4879	0.4840	0.4868	0.4884	0.4881
	Rank	6	1	4	7	5	2	3
	Stdev	0.1629	0.1451	0.1565	0.1748	0.1612	0.1540	0.1549
				With structural change				
	Average	0.5043	0.4967	0.5028	0.5068	0.5039	0.5018	0.5021
	Rank	2	7	4	1	3	6	5
	Stdev	0.1442	0.1459	0.1446	0.1438	0.1444	0.1449	0.1448
				Without structural change				
	Average	0.1270	0.3655	0.1915	0.0280	0.1435	0.2205	0.2090
	Rank	6	1	4	7	5	2	3
0.70				With and without structural change				
	Average	0.4815	0.4860	0.4834	0.4791	0.4818	0.4836	0.4838
	Rank	6	1	4	7	5	3	2
	Stdev	0.1607	0.1433	0.1540	0.1733	0.1594	0.1513	0.1524
				With structural change				
	Average	0.4993	0.4920	0.4979	0.5018	0.4991	0.4964	0.4974
	Rank	2	7	4	1	3	6	5
	Stdev	0.1421	0.1443	0.1424	0.1423	0.1420	0.1429	0.1428
				Without structural change				
	Average	0.1255	0.3655	0.1930	0.0255	0.1365	0.2260	0.2130
	Rank	6	1	4	7	5	2	3
1.0				With and without structural change				
	Average	0.4799	0.4878	0.4855	0.4824	0.4843	0.4858	0.4855
	Rank	7	1	4	6	5	2	3
	Stdev	0.1784	0.1498	0.1624	0.1772	0.1665	0.1599	0.1608
				With structural change				
	Average	0.5034	0.4968	0.5027	0.5055	0.5032	0.5015	0.5017
	Rank	2	7	4	1	3	6	5
	Stdev	0.1459	0.1476	0.1457	0.1455	0.1458	0.1467	0.1464
				Without structural change				
	Average	0.0099	0.3065	0.1420	0.0190	0.1060	0.1725	0.1615
	Rank	7	1	4	6	5	2	3
1.02				With and without structural change				
	Average	0.4809	0.4840	0.4823	0.4796	0.4812	0.4823	0.4825
	Rank	6	1	4	7	5	3	2
	Stdev	0.1728	0.1534	0.1671	0.1801	0.1717	0.1645	0.1655
				With structural change				
	Average	0.5012	0.4937	0.4998	0.5028	0.5009	0.4987	0.4992
	Rank	2	7	4	1	3	6	5
	Stdev	0.1496	0.1506	0.1503	0.1492	0.1497	0.1503	0.1503
				Without structural change				
	Average	0.0765	0.2885	0.1310	0.0160	0.0860	0.1550	0.1470
	Rank	6	1	4	7	5	2	3

Table 4.12 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 46 Models, $\phi = 0$ for $n = 50$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1550	0.5270	0.3080	0.2215	0.1585	0.2040	0.1980
2	0.8055	0.7850	0.8015	0.8100	0.8055	0.8025	0.8035
3	0.5415	0.5395	0.5405	0.5425	0.5415	0.5415	0.5415
4	0.5535	0.5485	0.5530	0.5540	0.5535	0.5535	0.5535
5	0.6475	0.6470	0.6475	0.6485	0.6475	0.6475	0.6475
6	0.4625	0.4605	0.4615	0.4630	0.4625	0.4620	0.4625
7	0.2640	0.2630	0.2635	0.2640	0.2640	0.2640	0.2640
8	0.2745	0.2735	0.2745	0.2745	0.2745	0.2745	0.2745
9	0.4675	0.4670	0.4675	0.4680	0.4675	0.4675	0.4675
10	0.4155	0.4155	0.4155	0.4155	0.4155	0.4155	0.4155
11	0.5105	0.5095	0.5105	0.5105	0.5105	0.5105	0.5105
12	0.7315	0.7290	0.7310	0.7330	0.7315	0.7315	0.7315
13	0.6980	0.6955	0.6955	0.6980	0.6975	0.6970	0.6970
14	0.6940	0.6925	0.6935	0.6940	0.6940	0.6940	0.6940
15	0.7500	0.7470	0.7485	0.7505	0.7495	0.7495	0.7495
16	0.7430	0.7410	0.7430	0.7430	0.7430	0.7430	0.7430
17	0.7990	0.7960	0.7990	0.7995	0.7990	0.7990	0.7990
18	0.7400	0.7385	0.7400	0.7405	0.7400	0.7400	0.7400
19	0.7380	0.7360	0.7375	0.7380	0.7380	0.7380	0.7380
20	0.7185	0.7170	0.7185	0.7185	0.7185	0.7185	0.7185
21	0.6960	0.6955	0.6960	0.6960	0.6960	0.6960	0.6960
22	0.6840	0.6830	0.6835	0.6840	0.6840	0.6835	0.6835
23	0.7320	0.7310	0.7315	0.7330	0.7320	0.7320	0.7320
24	0.8470	0.8440	0.8460	0.8475	0.8470	0.8470	0.8470
25	0.8510	0.8480	0.8505	0.8510	0.8510	0.8510	0.8510
26	0.5385	0.5370	0.5385	0.5390	0.5385	0.5385	0.5385
27	0.5140	0.5110	0.5135	0.5145	0.5140	0.5135	0.5135
28	0.8190	0.8170	0.8190	0.8195	0.8190	0.8190	0.8190
29	0.6405	0.6385	0.6400	0.6405	0.6405	0.6405	0.6405
30	0.6130	0.6125	0.6130	0.6130	0.6130	0.6130	0.6130
31	0.7025	0.7015	0.7015	0.7025	0.7025	0.7025	0.7025
32	0.7855	0.7850	0.7855	0.7855	0.7855	0.7855	0.7855
33	0.7385	0.7385	0.7385	0.7385	0.7385	0.7385	0.7385
34	0.7730	0.7720	0.7725	0.7730	0.7730	0.7730	0.7730
35	0.8295	0.8265	0.8280	0.8300	0.8295	0.8290	0.8290
36	0.8290	0.8250	0.8280	0.8305	0.8290	0.8290	0.8290
37	0.7330	0.7315	0.7325	0.7330	0.7330	0.7325	0.7325
38	0.5130	0.5125	0.5130	0.5130	0.5130	0.5130	0.5130
39	0.5080	0.5070	0.5080	0.5080	0.5080	0.5080	0.5080
40	0.6450	0.6445	0.6445	0.6450	0.6450	0.6450	0.6450
41	0.5230	0.5210	0.5225	0.5240	0.5230	0.5230	0.5230
42	0.5175	0.5165	0.5175	0.5180	0.5175	0.5175	0.5175
43	0.8180	0.8140	0.8190	0.8190	0.8180	0.8140	0.8170
44	0.8330	0.8285	0.8380	0.8350	0.8330	0.8225	0.8215
45	0.8355	0.8305	0.8345	0.8370	0.8345	0.8350	0.8350
46	0.8340	0.8265	0.8345	0.8355	0.8245	0.8340	0.8340

Table 4.13 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 46 Models, $\phi = 0.7$ for $n = 50$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1550	0.5270	0.3080	0.1215	0.1585	0.2040	0.1980
2	0.9055	0.7850	0.8015	0.8100	0.8055	0.8250	0.8035
3	0.5415	0.5395	0.5405	0.5425	0.5415	0.5415	0.5415
4	0.5535	0.5485	0.5530	0.5540	0.5535	0.5535	0.5535
5	0.6475	0.6470	0.6475	0.6485	0.6475	0.6475	0.6475
6	0.4625	0.4605	0.4615	0.4630	0.4625	0.4620	0.4625
7	0.2640	0.2630	0.2635	0.2640	0.2640	0.2640	0.2640
8	0.2745	0.2735	0.2745	0.2745	0.2745	0.2745	0.2745
9	0.4675	0.4670	0.4675	0.4680	0.4675	0.4675	0.4675
10	0.4155	0.4150	0.4155	0.4155	0.4155	0.4155	0.4155
11	0.5105	0.5095	0.5105	0.5105	0.5105	0.5105	0.5105
12	0.7315	0.7290	0.7310	0.7330	0.7315	0.7315	0.7315
13	0.6980	0.6955	0.6965	0.6980	0.6975	0.6970	0.6970
14	0.6940	0.6925	0.6935	0.6940	0.6940	0.6940	0.6940
15	0.7500	0.7470	0.7485	0.7505	0.7495	0.7495	0.7495
16	0.7430	0.7410	0.7430	0.7430	0.7430	0.7430	0.7430
17	0.7990	0.7960	0.7990	0.7995	0.7990	0.7990	0.7990
18	0.7400	0.7385	0.7400	0.7405	0.7400	0.7400	0.7400
19	0.7380	0.7360	0.7375	0.7380	0.7380	0.7380	0.7380
20	0.7185	0.7170	0.7185	0.7185	0.7185	0.7185	0.7185
21	0.6960	0.6955	0.6960	0.6960	0.6960	0.6960	0.6960
22	0.6840	0.6830	0.6835	0.6840	0.6840	0.6835	0.6835
23	0.7320	0.7310	0.7315	0.7330	0.7320	0.7320	0.7320
24	0.8470	0.8440	0.8460	0.8475	0.8470	0.8470	0.8470
25	0.8510	0.8480	0.8505	0.8510	0.8510	0.8510	0.8510
26	0.5385	0.5370	0.5385	0.5390	0.5385	0.5385	0.5385
27	0.5140	0.5110	0.5135	0.5140	0.5140	0.5135	0.5135
28	0.8190	0.8170	0.8190	0.8195	0.8190	0.8190	0.8190
29	0.6405	0.6385	0.6400	0.6405	0.6405	0.6405	0.6405
30	0.6130	0.6125	0.6130	0.6130	0.6130	0.6130	0.6130
31	0.7025	0.7015	0.7015	0.7085	0.7025	0.7025	0.7025
32	0.7855	0.7850	0.7855	0.7895	0.7855	0.7855	0.7855
33	0.7385	0.7385	0.7385	0.7385	0.7385	0.7385	0.7385
34	0.7730	0.7720	0.7725	0.7730	0.7730	0.7730	0.7730
35	0.8295	0.8265	0.8280	0.8300	0.8295	0.8290	0.8290
36	0.8290	0.8250	0.8280	0.8305	0.8290	0.8290	0.8290
37	0.7330	0.7315	0.7325	0.7330	0.7330	0.7325	0.7325
38	0.5130	0.5125	0.5130	0.5180	0.5130	0.5130	0.5130
39	0.5080	0.5070	0.5080	0.5080	0.5080	0.5080	0.5080
40	0.6450	0.6445	0.6445	0.6450	0.6450	0.6450	0.6450
41	0.5230	0.5210	0.5225	0.5240	0.5230	0.5230	0.5230
42	0.5175	0.5165	0.5175	0.5180	0.5175	0.5175	0.5175
43	0.8180	0.8140	0.8180	0.8190	0.8180	0.8180	0.8180
44	0.8130	0.8285	0.8310	0.8390	0.8330	0.8325	0.8325
45	0.8155	0.8305	0.8345	0.8395	0.8355	0.8350	0.8350
46	0.8145	0.8265	0.8325	0.8355	0.8345	0.8340	0.8340

Table 4.14 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 46 Models, $\phi = 1.0$ for $n = 50$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1815	0.5580	0.3430	0.1445	0.1835	0.2340	0.2275
2	0.8065	0.7870	0.8025	0.8095	0.8065	0.8045	0.8045
3	0.5420	0.5385	0.5410	0.5435	0.5420	0.5415	0.5415
4	0.5480	0.5455	0.5470	0.5490	0.5480	0.5480	0.5480
5	0.6415	0.6415	0.6415	0.6420	0.6415	0.6415	0.6415
6	0.4665	0.4645	0.4655	0.4670	0.4665	0.4665	0.4665
7	0.2630	0.2625	0.2630	0.2635	0.2630	0.2630	0.2630
8	0.2730	0.2720	0.2730	0.2730	0.2730	0.2730	0.2730
9	0.4665	0.4665	0.4660	0.4670	0.4665	0.4665	0.4665
10	0.4160	0.4155	0.4160	0.4165	0.4160	0.4160	0.4160
11	0.4910	0.4895	0.4905	0.4910	0.4910	0.4910	0.4910
12	0.7195	0.7190	0.7195	0.7200	0.7195	0.7195	0.7195
13	0.6940	0.6930	0.6935	0.6940	0.6940	0.6940	0.6940
14	0.6905	0.6905	0.6905	0.6910	0.6905	0.6905	0.6905
15	0.7440	0.7425	0.7435	0.7445	0.7440	0.7440	0.7440
16	0.7430	0.7380	0.7400	0.7400	0.7400	0.7400	0.7400
17	0.7890	0.7885	0.7880	0.7895	0.7890	0.7885	0.7885
18	0.7345	0.7315	0.7330	0.7350	0.7345	0.7340	0.7340
19	0.7330	0.7315	0.7325	0.7330	0.7330	0.7325	0.7330
20	0.7080	0.7070	0.7080	0.7080	0.7080	0.7080	0.7080
21	0.6940	0.6935	0.6940	0.6940	0.6940	0.6940	0.6940
22	0.6745	0.6735	0.6745	0.6750	0.6745	0.6745	0.6745
23	0.7150	0.7145	0.7150	0.7150	0.7150	0.7150	0.7150
24	0.8400	0.8395	0.8395	0.8400	0.8400	0.8400	0.8400
25	0.8515	0.8505	0.8515	0.8515	0.8515	0.8515	0.8515
26	0.5235	0.5225	0.5235	0.5235	0.5235	0.5235	0.5235
27	0.5070	0.5055	0.5065	0.5070	0.5070	0.5070	0.5070
28	0.8135	0.8125	0.8130	0.8135	0.8135	0.8135	0.8135
29	0.6295	0.6285	0.6295	0.6295	0.6295	0.6295	0.6295
30	0.5990	0.5980	0.5990	0.5990	0.5990	0.5990	0.5990
31	0.6835	0.6835	0.6835	0.6840	0.6835	0.6835	0.6835
32	0.7775	0.7770	0.7775	0.7775	0.7775	0.7775	0.7775
33	0.7345	0.7345	0.7345	0.7345	0.7345	0.7345	0.7345
34	0.7790	0.7780	0.7785	0.7790	0.7790	0.7785	0.7785
35	0.8255	0.8240	0.8255	0.8260	0.8255	0.8255	0.8255
36	0.8220	0.8205	0.8215	0.8225	0.8220	0.8220	0.8220
37	0.7395	0.7375	0.7385	0.7400	0.7395	0.7395	0.7395
38	0.5165	0.5160	0.5160	0.5160	0.5160	0.5160	0.5160
39	0.5105	0.5100	0.5105	0.5115	0.5105	0.5105	0.5105
40	0.6510	0.6505	0.6510	0.6515	0.6510	0.6510	0.6510
41	0.5300	0.5280	0.5295	0.5300	0.5300	0.5300	0.5300
42	0.5100	0.5095	0.5100	0.5105	0.5100	0.5100	0.5100
43	0.8125	0.8075	0.8120	0.8125	0.8125	0.8125	0.8125
44	0.8255	0.8225	0.8240	0.8265	0.8255	0.8255	0.8255
45	0.8325	0.8255	0.8310	0.8335	0.8325	0.8310	0.8310
46	0.8355	0.8280	0.8335	0.8375	0.8355	0.8345	0.8345

Table 4.15 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 46 Models, $\phi = 1.02$ for $n = 50$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.3086	0.4360	0.3208	0.3072	0.3088	0.3175	0.3113
2	0.8015	0.7870	0.7980	0.8030	0.8015	0.8010	0.8010
3	0.5390	0.5380	0.5380	0.5390	0.5390	0.5390	0.5390
4	0.5500	0.5485	0.5380	0.5390	0.5390	0.5390	0.5390
5	0.6480	0.6475	0.6480	0.6480	0.6480	0.6480	0.6480
6	0.4625	0.4615	0.4620	0.4625	0.4625	0.4625	0.4625
7	0.2640	0.2640	0.2640	0.2640	0.2640	0.2640	0.2640
8	0.2745	0.2740	0.2740	0.2745	0.2745	0.2740	0.2745
9	0.4780	0.4770	0.4780	0.4780	0.4780	0.4780	0.4780
10	0.4190	0.4180	0.4190	0.4195	0.4190	0.4190	0.4190
11	0.5120	0.5120	0.5120	0.5120	0.5120	0.5120	0.5120
12	0.7270	0.7260	0.7270	0.7275	0.7270	0.7270	0.7270
13	0.6955	0.6945	0.6955	0.6955	0.6955	0.6955	0.6955
14	0.6970	0.6960	0.6970	0.6970	0.6970	0.6970	0.6970
15	0.7480	0.7475	0.7475	0.7480	0.7480	0.7475	0.7475
16	0.7395	0.7385	0.7395	0.7395	0.7395	0.7395	0.7395
17	0.7935	0.7920	0.7935	0.7935	0.7935	0.7935	0.7935
18	0.7385	0.7395	0.7400	0.7400	0.7400	0.7395	0.7395
19	0.7405	0.7400	0.7405	0.7405	0.7405	0.7405	0.7405
20	0.7340	0.7320	0.7340	0.7345	0.7340	0.7340	0.7340
21	0.6940	0.6940	0.6940	0.6940	0.6940	0.6940	0.6940
22	0.6905	0.6895	0.6900	0.6910	0.6905	0.6905	0.6905
23	0.7350	0.7350	0.7350	0.7355	0.7350	0.7350	0.7350
24	0.8455	0.8435	0.8455	0.8465	0.8455	0.8455	0.8455
25	0.8490	0.8455	0.8485	0.8490	0.8490	0.8485	0.8485
26	0.5385	0.5385	0.5385	0.5385	0.5385	0.5385	0.5385
27	0.5100	0.5080	0.5095	0.5100	0.5095	0.5095	0.5095
28	0.8180	0.8145	0.8170	0.8195	0.8180	0.8175	0.8175
29	0.6320	0.6310	0.6315	0.6325	0.6320	0.6320	0.6320
30	0.6015	0.6005	0.6015	0.6015	0.6015	0.6015	0.6015
31	0.6990	0.6960	0.6985	0.6990	0.6990	0.6990	0.6990
32	0.7855	0.7830	0.7845	0.7855	0.7855	0.7855	0.7855
33	0.7365	0.7345	0.7355	0.7370	0.7365	0.7365	0.7365
34	0.7760	0.7740	0.7745	0.7765	0.7760	0.7760	0.7760
35	0.8230	0.8195	0.8230	0.8245	0.8230	0.8230	0.8230
36	0.8220	0.8160	0.8200	0.8225	0.8220	0.8210	0.8210
37	0.7250	0.7190	0.7225	0.7265	0.7250	0.7250	0.7250
38	0.5055	0.5015	0.5040	0.5060	0.5055	0.5050	0.5050
39	0.4960	0.4945	0.4960	0.4960	0.4960	0.4960	0.4960
40	0.6420	0.6385	0.6405	0.6420	0.6420	0.6420	0.6420
41	0.5315	0.5280	0.5300	0.5330	0.5315	0.5310	0.5310
42	0.5065	0.5030	0.5050	0.5075	0.5065	0.5055	0.5055
43	0.8085	0.8015	0.8065	0.8105	0.8085	0.8085	0.8085
44	0.8130	0.8000	0.8100	0.8145	0.8130	0.8120	0.8120
45	0.8125	0.8000	0.8090	0.8150	0.8125	0.8115	0.8120
46	0.7955	0.7860	0.7935	0.7970	0.7955	0.7950	0.7950

Table 4.16 Average Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Different ϕ when $n = 50$.

ϕ		AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
0				With and without structural change				
	Average	0.6535	0.6593	0.6562	0.6511	0.6536	0.6544	0.6543
	Rank	6	1	2	4	5	3	4
	Stdev	0.1675	0.1500	0.1583	0.1775	0.1673	0.1643	0.1647
				With structural change				
	Average	0.6646	0.6622	0.6639	0.6651	0.6646	0.6644	0.6645
	Rank	2	7	6	1	3	5	5
	Stdev	0.1514	0.1504	0.1511	0.1517	0.1514	0.1513	0.1513
				Without structural change				
	Average	0.1550	0.5270	0.3080	0.0215	0.1585	0.2040	0.1980
	Rank	6	1	2	7	5	3	4
0.70				With and without structural change				
	Average	0.6557	0.6593	0.6563	0.6511	0.6536	0.6549	0.6543
	Rank	3	1	2	7	5	4	5
	Stdev	0.1702	0.1500	0.1584	0.1775	0.1673	0.1648	0.1647
				With structural change				
	Average	0.6668	0.6622	0.6641	0.6651	0.6646	0.6649	0.6645
	Rank	2	7	6	1	3	5	5
	Stdev	0.1543	0.1504	0.1512	0.1517	0.1514	0.1518	0.1513
				Without structural change				
	Average	0.1550	0.5270	0.3080	0.0215	0.1585	0.2040	0.1980
	Rank	6	1	2	7	5	3	4
1.0				With and without structural change				
	Average	0.6497	0.6559	0.6526	0.6466	0.6496	0.6506	0.6504
	Rank	5	1	2	7	5	3	4
	Stdev	0.1649	0.1489	0.1559	0.1761	0.1647	0.1616	0.1619
				With structural change				
	Average	0.6601	0.6581	0.6595	0.6604	0.6600	0.6598	0.6598
	Rank	2	7	6	1	3	5	5
	Stdev	0.1507	0.1498	0.1505	0.1508	0.1507	0.1506	0.1506
				Without structural change				
	Average	0.1815	0.5580	0.3430	0.0245	0.1835	0.2340	0.2275
	Rank	6	1	2	7	5	3	4
1.02				With and without structural change				
	Average	0.6487	0.6536	0.6504	0.6473	0.6485	0.6490	0.6489
	Rank	5	1	2	7	5	3	4
	Stdev	0.1701	0.1496	0.1617	0.1766	0.1701	0.1679	0.1682
				With structural change				
	Average	0.6612	0.6584	0.6602	0.6615	0.6610	0.6608	0.6608
	Rank	2	7	6	1	3	5	5
	Stdev	0.1491	0.1476	0.1490	0.1496	0.1493	0.1492	0.1492
				Without structural change				
	Average	0.0860	0.4360	0.2080	0.0070	0.0875	0.1175	0.1130
	Rank	6	1	2	7	5	3	4

Table 4.17 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 71 Models, $\phi = 0$ for $n = 75$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1190	0.5735	0.3080	0.2065	0.1200	0.1460	0.1420
2	0.8365	0.8395	0.8345	0.8365	0.8365	0.8365	0.8365
3	0.6510	0.6475	0.6500	0.6515	0.6510	0.6510	0.6510
4	0.6650	0.6635	0.6645	0.6655	0.6650	0.6650	0.6650
5	0.8955	0.8940	0.8955	0.8960	0.8955	0.8955	0.8955
6	0.4945	0.4945	0.4945	0.4945	0.4945	0.4945	0.4945
7	0.4875	0.4860	0.4875	0.4875	0.4875	0.4875	0.4875
8	0.8850	0.8825	0.8850	0.8850	0.8850	0.8850	0.8850
9	0.8970	0.8960	0.8970	0.8970	0.8970	0.8970	0.8970
10	0.7160	0.7160	0.7160	0.7160	0.7160	0.7160	0.7160
11	0.4670	0.4665	0.4670	0.4670	0.4670	0.4670	0.4670
12	0.4995	0.4995	0.4995	0.4995	0.4995	0.4995	0.4995
13	0.8825	0.8820	0.8825	0.8825	0.8825	0.8825	0.8825
14	0.8330	0.8330	0.8330	0.8330	0.8330	0.8330	0.8330
15	0.8295	0.8285	0.8295	0.8295	0.8295	0.8295	0.8295
16	0.8755	0.8750	0.8755	0.8755	0.8755	0.8755	0.8755
17	0.8780	0.8770	0.8780	0.8780	0.8780	0.8780	0.8780
18	0.8865	0.8860	0.8865	0.8865	0.8865	0.8865	0.8865
19	0.6195	0.6185	0.6190	0.6195	0.6195	0.6195	0.6195
20	0.6290	0.6280	0.6290	0.6290	0.6290	0.6290	0.6290
21	0.8850	0.8845	0.8850	0.8850	0.8850	0.8850	0.8850
22	0.6750	0.6750	0.6750	0.6750	0.6750	0.6750	0.6750
23	0.6125	0.6125	0.6125	0.6125	0.6125	0.6125	0.6125
24	0.7160	0.7160	0.7160	0.7160	0.7160	0.7160	0.7160
25	0.8810	0.8810	0.8810	0.8810	0.8810	0.8810	0.8810
26	0.8680	0.8675	0.8680	0.8680	0.8680	0.8680	0.8680
27	0.8560	0.8560	0.8560	0.8560	0.8560	0.8560	0.8560
28	0.5990	0.5990	0.5990	0.5990	0.5990	0.5990	0.5990
29	0.5735	0.5735	0.5735	0.5734	0.5735	0.5735	0.5735
30	0.7875	0.7870	0.7870	0.7875	0.7875	0.7875	0.7875
31	0.8830	0.8830	0.8830	0.8830	0.8830	0.8830	0.8830
32	0.8935	0.8915	0.8925	0.8935	0.8935	0.8935	0.8935
33	0.8175	0.8175	0.8175	0.8175	0.8175	0.8175	0.8175
34	0.8045	0.8045	0.8045	0.8045	0.8045	0.8045	0.8045
35	0.8745	0.8745	0.8745	0.8745	0.8745	0.8745	0.8745
36	0.8115	0.8110	0.8115	0.8115	0.8115	0.8115	0.8115
37	0.6635	0.6635	0.6635	0.6635	0.6635	0.6635	0.6635
38	0.5235	0.5235	0.5235	0.5235	0.5235	0.5235	0.5235
39	0.5915	0.5915	0.5915	0.5915	0.5915	0.5915	0.5915
40	0.8665	0.8660	0.8665	0.8665	0.8665	0.8665	0.8665
41	0.8575	0.8570	0.8575	0.8575	0.8575	0.8575	0.8575
42	0.9050	0.9045	0.9050	0.9050	0.9050	0.9050	0.9050
43	0.8130	0.8130	0.8130	0.8130	0.8130	0.8130	0.8130
44	0.7735	0.7735	0.7735	0.7735	0.7735	0.7735	0.7735
45	0.8270	0.8270	0.8270	0.8270	0.8270	0.8270	0.8270

Table 4.17 (cont'd)

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
46	0.8685	0.8685	0.8685	0.8685	0.8685	0.8685	0.8685
47	0.8755	0.8755	0.8755	0.8755	0.8755	0.8755	0.8755
48	0.8760	0.8760	0.8760	0.8760	0.8760	0.8760	0.8760
49	0.8110	0.8110	0.8110	0.8110	0.8110	0.8110	0.8110
50	0.7850	0.7850	0.7850	0.7850	0.7850	0.7850	0.7850
51	0.8340	0.8340	0.8340	0.8340	0.8340	0.8340	0.8340
52	0.8555	0.8555	0.8555	0.8555	0.8555	0.8555	0.8555
53	0.7270	0.7270	0.7270	0.7270	0.7270	0.7270	0.7270
54	0.7310	0.7310	0.7310	0.7310	0.7310	0.7310	0.7310
55	0.8715	0.8705	0.8715	0.8715	0.8715	0.8715	0.8715
56	0.8570	0.8565	0.8570	0.8570	0.8570	0.8570	0.8570
57	0.8465	0.8465	0.8465	0.8465	0.8465	0.8460	0.8465
58	0.8715	0.8715	0.8715	0.8715	0.8715	0.8715	0.8715
59	0.8630	0.8630	0.8630	0.8630	0.8630	0.8630	0.8630
60	0.8045	0.8030	0.8040	0.8050	0.8045	0.8045	0.8045
61	0.8225	0.8215	0.8220	0.8225	0.8225	0.8225	0.8225
62	0.8920	0.8910	0.8920	0.8920	0.8920	0.8920	0.8920
63	0.8270	0.8270	0.8270	0.8270	0.8270	0.8270	0.8270
64	0.4505	0.4500	0.4500	0.4505	0.4505	0.4505	0.4505
65	0.4095	0.4095	0.4095	0.4095	0.4095	0.4095	0.4095
66	0.8430	0.8400	0.8410	0.8430	0.8430	0.8425	0.8425
67	0.8245	0.8235	0.8245	0.8250	0.8245	0.8245	0.8245
68	0.8605	0.8585	0.8600	0.8610	0.8605	0.8605	0.8605
69	0.8025	0.7995	0.8025	0.8025	0.8025	0.8025	0.8025
70	0.7995	0.7975	0.7985	0.7995	0.7995	0.7995	0.7995
71	0.8670	0.8610	0.8655	0.8685	0.8670	0.8670	0.8670

Table 4.18 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 71 Models, $\phi = 0.7$ for $n = 75$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1095	0.5780	0.3040	0.1085	0.1115	0.1315	0.1315
2	0.8350	0.8295	0.8335	0.8350	0.8350	0.8350	0.8350
3	0.6470	0.6435	0.6460	0.6475	0.6470	0.6465	0.6465
4	0.6595	0.6585	0.6895	0.6600	0.6595	0.6595	0.6595
5	0.8965	0.8930	0.8955	0.8965	0.8965	0.8960	0.8965
6	0.4915	0.4915	0.4915	0.4915	0.4915	0.4915	0.4915
7	0.4920	0.4910	0.4920	0.4920	0.4920	0.4920	0.4920
8	0.8845	0.8825	0.8840	0.8845	0.8845	0.8845	0.8845
9	0.8980	0.8980	0.8980	0.8980	0.8980	0.8980	0.8980
10	0.7020	0.7020	0.7020	0.7020	0.7020	0.7020	0.7020
11	0.4640	0.4635	0.4640	0.4640	0.4640	0.4640	0.4640
12	0.4920	0.4905	0.4915	0.4920	0.4920	0.4920	0.4920
13	0.8795	0.8790	0.8795	0.8795	0.8795	0.8795	0.8795
14	0.8340	0.8340	0.8340	0.8340	0.8340	0.8340	0.8340
15	0.8280	0.8275	0.8280	0.8285	0.8280	0.8280	0.8280
16	0.8725	0.8720	0.8720	0.8725	0.8725	0.8725	0.8725
17	0.8775	0.8765	0.8770	0.8775	0.8775	0.8775	0.8775
18	0.8865	0.8865	0.8865	0.8865	0.8865	0.8865	0.8865
19	0.6240	0.6235	0.6240	0.6245	0.6240	0.6240	0.6240
20	0.6255	0.6245	0.6250	0.6255	0.6255	0.6255	0.6255
21	0.8850	0.8850	0.8850	0.8850	0.8850	0.8850	0.8850
22	0.6775	0.6770	0.6770	0.6775	0.6775	0.6770	0.6770
23	0.6120	0.6120	0.6120	0.6120	0.6120	0.6120	0.6120
24	0.7175	0.7175	0.7175	0.7175	0.7175	0.7175	0.7175
25	0.8840	0.8840	0.8840	0.8840	0.8840	0.8840	0.8840
26	0.8670	0.8665	0.8670	0.8670	0.8670	0.8670	0.8670
27	0.8555	0.8555	0.8555	0.8555	0.8555	0.8555	0.8555
28	0.6010	0.6005	0.6005	0.6010	0.6010	0.6010	0.6010
29	0.5710	0.5705	0.5710	0.5710	0.5710	0.5710	0.5710
30	0.7830	0.7830	0.7830	0.7830	0.7830	0.7830	0.7830
31	0.8830	0.8830	0.8830	0.8830	0.8830	0.8830	0.8830
32	0.8945	0.8945	0.8945	0.8945	0.8945	0.8945	0.8945
33	0.8150	0.8150	0.8150	0.8150	0.8150	0.8150	0.8150
34	0.8050	0.8050	0.8050	0.8050	0.8050	0.8050	0.8050
35	0.8740	0.8740	0.8740	0.8740	0.8740	0.8740	0.8740
36	0.8110	0.8105	0.8105	0.8110	0.8110	0.8110	0.8110
37	0.6615	0.6615	0.6615	0.6615	0.6615	0.6615	0.5615
38	0.5305	0.5305	0.5305	0.5305	0.5305	0.5305	0.5305
39	0.5955	0.5955	0.5955	0.5955	0.5955	0.5955	0.5955
40	0.8615	0.8615	0.8615	0.8615	0.8615	0.8615	0.8615
41	0.8550	0.8550	0.8550	0.8550	0.8550	0.8550	0.8550
42	0.9040	0.9035	0.9040	0.9040	0.9040	0.9040	0.9040
43	0.8115	0.8110	0.8110	0.8115	0.8115	0.8115	0.8110
44	0.7750	0.7745	0.7750	0.7750	0.7750	0.7750	0.7750
45	0.8215	0.8215	0.8215	0.8220	0.8215	0.8215	0.8215

Table 4.18 (cont'd)

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
46	0.8600	0.8600	0.8600	0.8600	0.8600	0.8600	0.8600
47	0.8720	0.8715	0.8720	0.8720	0.8720	0.8720	0.8720
48	0.8720	0.8720	0.8720	0.8720	0.8720	0.8720	0.8720
49	0.8120	0.8120	0.8120	0.8120	0.8120	0.8120	0.8120
50	0.7870	0.7865	0.7870	0.7870	0.7870	0.7870	0.7870
51	0.8355	0.8355	0.8355	0.8355	0.8355	0.8355	0.8355
52	0.8555	0.8555	0.8555	0.8555	0.8555	0.8555	0.8555
53	0.7290	0.7285	0.7290	0.7290	0.7290	0.7290	0.7290
54	0.7325	0.7325	0.7325	0.7325	0.7325	0.7325	0.7325
55	0.8685	0.8675	0.8685	0.8685	0.8685	0.8685	0.8685
56	0.8570	0.8570	0.8570	0.8570	0.8570	0.8570	0.8570
57	0.8465	0.8465	0.8465	0.8465	0.8465	0.8465	0.8465
58	0.8725	0.8725	0.8725	0.8725	0.8725	0.8725	0.8725
59	0.8610	0.8605	0.8610	0.8610	0.8610	0.8610	0.8610
60	0.8075	0.8065	0.8075	0.8075	0.8075	0.8075	0.8075
61	0.8265	0.8265	0.8265	0.8265	0.8265	0.8265	0.8265
62	0.8955	0.8940	0.8950	0.8955	0.8955	0.8955	0.8955
63	0.8270	0.8260	0.8270	0.8270	0.8270	0.8270	0.8270
64	0.4465	0.4455	0.4460	0.4465	0.4465	0.4460	0.4465
65	0.4165	0.4165	0.4165	0.4165	0.4165	0.4165	0.4165
66	0.8450	0.8425	0.8430	0.8455	0.8450	0.8445	0.8445
67	0.8235	0.8230	0.8230	0.8240	0.8235	0.8235	0.8235
68	0.8600	0.8570	0.8590	0.8600	0.8600	0.8600	0.8600
69	0.8030	0.8030	0.8030	0.8030	0.8030	0.8030	0.8030
70	0.7995	0.7980	0.7995	0.7995	0.7995	0.7995	0.7995
71	0.8660	0.8600	0.8640	0.8675	0.8660	0.8655	0.8655

Table 4.19 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 71 Models, $\phi = 1.0$ for $n = 75$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.2155	0.4815	0.2010	0.2025	0.2555	0.2755	0.2745
2	0.8340	0.8290	0.8330	0.8340	0.8340	0.8340	0.8340
3	0.6460	0.6440	0.6455	0.6460	0.6460	0.0646	0.6460
4	0.6590	0.6590	0.6590	0.6595	0.6590	0.6590	0.6590
5	0.8905	0.8865	0.8900	0.8905	0.8905	0.8905	0.8905
6	0.4900	0.4895	0.4900	0.4900	0.4900	0.4900	0.4900
7	0.4840	0.4820	0.4835	0.4840	0.4840	0.4840	0.4840
8	0.8870	0.8835	0.8860	0.8875	0.8870	0.8870	0.8870
9	0.8955	0.8945	0.8955	0.8955	0.8955	0.8955	0.8955
10	0.7070	0.7065	0.7065	0.7065	0.7070	0.7070	0.7070
11	0.4755	0.4750	0.4750	0.4755	0.4755	0.4755	0.4755
12	0.4835	0.4825	0.4830	0.4835	0.4835	0.4835	0.4835
13	0.8805	0.8795	0.8800	0.8805	0.8805	0.8805	0.8805
14	0.8320	0.8305	0.8320	0.8320	0.8320	0.8320	0.8320
15	0.8280	0.8280	0.8280	0.8280	0.8280	0.8280	0.8280
16	0.8740	0.8730	0.8740	0.8740	0.8740	0.8740	0.8740
17	0.8760	0.8755	0.8760	0.8760	0.8760	0.8760	0.8760
18	0.8795	0.8795	0.8795	0.8795	0.8795	0.8795	0.8795
19	0.6125	0.6115	0.6125	0.6125	0.6125	0.6125	0.6125
20	0.6215	0.6205	0.6215	0.6215	0.6215	0.6215	0.6215
21	0.8810	0.8805	0.8810	0.8810	0.8810	0.8810	0.8810
22	0.6710	0.6695	0.6705	0.6710	0.6710	0.6710	0.6710
23	0.6140	0.6140	0.6140	0.6140	0.6140	0.6140	0.6140
24	0.7155	0.7155	0.7155	0.7155	0.7155	0.7155	0.7155
25	0.8800	0.8795	0.8800	0.8800	0.8800	0.8800	0.8800
26	0.8640	0.8630	0.8640	0.8640	0.8640	0.8640	0.8640
27	0.8520	0.8520	0.8520	0.8520	0.8520	0.8520	0.8520
28	0.6000	0.5995	0.5995	0.6000	0.6000	0.6000	0.6000
29	0.5720	0.5715	0.5715	0.5720	0.5720	0.5720	0.5720
30	0.7865	0.7855	0.7865	0.7865	0.7865	0.7865	0.7865
31	0.8865	0.8865	0.8865	0.8865	0.8865	0.8865	0.8865
32	0.8915	0.8910	0.8910	0.8915	0.8915	0.8910	0.8910
33	0.8155	0.8145	0.8150	0.8155	0.8155	0.8155	0.8155
34	0.8120	0.8115	0.8115	0.8120	0.8120	0.8120	0.8120
35	0.8745	0.8740	0.8740	0.8745	0.8745	0.8745	0.8745
36	0.8070	0.8070	0.8070	0.8070	0.8070	0.8070	0.8070
37	0.6615	0.6615	0.6615	0.6615	0.6615	0.6615	0.6615
38	0.5290	0.5280	0.5290	0.5290	0.5290	0.5290	0.5290
39	0.5920	0.5915	0.5915	0.5920	0.5920	0.5920	0.5920
40	0.8575	0.8560	0.8575	0.8575	0.8575	0.8575	0.8575
41	0.8520	0.8505	0.8520	0.8520	0.8520	0.8520	0.8520
42	0.8970	0.8945	0.8970	0.8970	0.8970	0.8970	0.8970
43	0.8075	0.8060	0.8075	0.8075	0.8075	0.8075	0.8075
44	0.7730	0.7725	0.7730	0.7730	0.7730	0.7730	0.7730
45	0.8255	0.8245	0.8255	0.8255	0.8255	0.8255	0.8255

Table 4.19 (cont'd)

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
46	0.8620	0.8605	0.8620	0.8620	0.8620	0.8620	0.8620
47	0.8715	0.8705	0.8715	0.8715	0.8715	0.8715	0.8715
48	0.8705	0.8700	0.8705	0.8705	0.8705	0.8705	0.8705
49	0.8075	0.8075	0.8075	0.8075	0.8075	0.8075	0.8075
50	0.7805	0.7805	0.7805	0.7805	0.7805	0.7805	0.7805
51	0.8350	0.8350	0.8350	0.8350	0.8350	0.8350	0.8350
52	0.8515	0.8515	0.8515	0.8515	0.8515	0.8515	0.8515
53	0.7290	0.7285	0.7290	0.7290	0.7290	0.7290	0.7290
54	0.7335	0.7335	0.7335	0.7335	0.7335	0.7335	0.7335
55	0.8670	0.8665	0.8670	0.8670	0.8670	0.8670	0.8670
56	0.8570	0.8565	0.8570	0.8570	0.8570	0.8570	0.8570
57	0.8485	0.8480	0.8485	0.8485	0.8485	0.8485	0.8485
58	0.8705	0.8705	0.8705	0.8705	0.8705	0.8705	0.8705
59	0.8600	0.8590	0.9600	0.8600	0.8600	0.8600	0.8600
60	0.8140	0.8140	0.8140	0.8140	0.8140	0.8140	0.8140
61	0.8260	0.8260	0.8260	0.8260	0.8260	0.8260	0.8260
62	0.8920	0.8895	0.8920	0.8920	0.8920	0.8920	0.8920
63	0.8225	0.8210	0.8220	0.8225	0.8225	0.8225	0.8225
64	0.4410	0.4400	0.4410	0.4410	0.4410	0.4410	0.4410
65	0.4155	0.4155	0.4155	0.4155	0.4155	0.4155	0.4155
66	0.8415	0.8385	0.8410	0.8415	0.8415	0.8415	0.8415
67	0.8220	0.8215	0.8220	0.8220	0.8220	0.8220	0.8220
68	0.8590	0.9580	0.8590	0.8590	0.8590	0.8590	0.8590
69	0.7975	0.7950	0.7970	0.7975	0.7975	0.7975	0.7975
70	0.7970	0.7955	0.7965	0.7970	0.7970	0.7970	0.7970
71	0.8625	0.7590	0.8620	0.8635	0.8625	0.8625	0.8625

Table 4.20 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 71 Models, $\phi = 1.02$ for $n = 75$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.3065	0.4980	0.3215	0.3055	0.3166	0.3785	0.3760
2	0.8405	0.8330	0.8385	0.8405	0.8405	0.8405	0.8405
3	0.6510	0.6475	0.6495	0.6510	0.6510	0.6500	0.6500
4	0.6665	0.6645	0.6660	0.6665	0.6665	0.6660	0.6660
5	0.8955	0.8945	0.8955	0.8955	0.8955	0.8955	0.8955
6	0.4950	0.4950	0.4950	0.4950	0.4950	0.4950	0.4950
7	0.4810	0.4810	0.4810	0.4810	0.4810	0.4810	0.4810
8	0.8845	0.8830	0.8835	0.8845	0.8845	0.8845	0.8845
9	0.8960	0.8945	0.8960	0.8960	0.8960	0.8960	0.8960
10	0.7150	0.7150	0.7150	0.7150	0.7150	0.7150	0.7150
11	0.4700	0.4695	0.4700	0.4700	0.4700	0.4700	0.4700
12	0.4940	0.4940	0.4940	0.4940	0.4940	0.4940	0.4940
13	0.8825	0.8825	0.8825	0.8825	0.8825	0.8825	0.8825
14	0.8350	0.8350	0.8350	0.8350	0.8350	0.8350	0.8350
15	0.8290	0.8290	0.8290	0.8290	0.8290	0.8290	0.8290
16	0.8740	0.8740	0.8740	0.8740	0.8740	0.8740	0.8740
17	0.8775	0.8770	0.8775	0.8775	0.8775	0.8775	0.8775
18	0.8825	0.8810	0.8815	0.8815	0.8825	0.8825	0.8825
19	0.6195	0.6190	0.6195	0.6195	0.6195	0.6195	0.6195
20	0.6250	0.6235	0.6240	0.6250	0.6250	0.6250	0.6250
21	0.8840	0.8835	0.8835	0.8840	0.8840	0.8840	0.8840
22	0.6780	0.6780	0.6780	0.6780	0.6780	0.6780	0.6780
23	0.6090	0.6085	0.6090	0.6090	0.6090	0.6090	0.6090
24	0.7205	0.7205	0.7205	0.7205	0.7205	0.7205	0.7205
25	0.8835	0.8835	0.8835	0.8835	0.8835	0.8835	0.8835
26	0.8665	0.8660	0.8665	0.8665	0.8665	0.8665	0.8665
27	0.8575	0.8575	0.8575	0.8575	0.8575	0.8575	0.8575
28	0.6020	0.6015	0.6020	0.6020	0.6020	0.6020	0.6020
29	0.5725	0.5720	0.5720	0.5725	0.5725	0.5725	0.5725
30	0.7860	0.7855	0.7860	0.7860	0.7860	0.7860	0.7860
31	0.8860	0.8855	0.8855	0.8860	0.8860	0.8860	0.8860
32	0.8935	0.8935	0.8935	0.8935	0.8935	0.8935	0.8935
33	0.8170	0.8170	0.8170	0.8170	0.8170	0.8170	0.8170
34	0.8140	0.8140	0.8140	0.8140	0.8140	0.8140	0.8140
35	0.8765	0.8765	0.8765	0.8765	0.8765	0.8765	0.8765
36	0.8095	0.8085	0.8095	0.8095	0.8095	0.8095	0.8095
37	0.6635	0.6630	0.6635	0.6635	0.6635	0.6635	0.6635
38	0.5325	0.5325	0.5325	0.5325	0.5325	0.5325	0.5325
39	0.5960	0.5960	0.5960	0.5960	0.5960	0.5960	0.5960
40	0.8630	0.8625	0.8630	0.8630	0.8630	0.8630	0.8630
41	0.8575	0.8575	0.8575	0.8575	0.8575	0.8575	0.8575
42	0.9040	0.9035	0.9035	0.9040	0.9040	0.9040	0.9040
43	0.8205	0.8200	0.8205	0.8205	0.8205	0.8205	0.8205
44	0.7755	0.7755	0.7755	0.7755	0.7755	0.7755	0.7755
45	0.8270	0.8270	0.8270	0.8270	0.8270	0.8270	0.8270
46	0.8700	0.8695	0.8700	0.8700	0.8700	0.8700	0.8700

Table 4.20 (cont'd)

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
47	0.8715	0.8710	0.8710	0.8715	0.8715	0.8715	0.8715
48	0.8715	0.8705	0.8705	0.8715	0.8715	0.8710	0.8710
49	0.8095	0.8095	0.8095	0.8095	0.8095	0.8095	0.8095
50	0.7860	0.7855	0.7855	0.7860	0.7860	0.7860	0.7860
51	0.8320	0.8320	0.8320	0.8320	0.8320	0.8320	0.8320
52	0.8520	0.8515	0.8520	0.8520	0.8520	0.8520	0.8520
53	0.7260	0.7255	0.7260	0.7260	0.7260	0.7260	0.7260
54	0.7290	0.7290	0.7290	0.7290	0.7290	0.7290	0.7290
55	0.8665	0.8660	0.8660	0.8665	0.8665	0.8665	0.8665
56	0.8585	0.8580	0.8585	0.8585	0.8585	0.8585	0.8585
57	0.8470	0.8465	0.8470	0.8470	0.8470	0.8470	0.8470
58	0.8675	0.8675	0.8675	0.8675	0.8675	0.8675	0.8675
59	0.8600	0.8585	0.8600	0.8600	0.8600	0.8600	0.8600
60	0.8085	0.8070	0.8080	0.8090	0.8085	0.8085	0.8085
61	0.8275	0.8245	0.8275	0.8295	0.8275	0.8275	0.8275
62	0.8900	0.8845	0.8885	0.8900	0.8900	0.8900	0.8900
63	0.8185	0.8155	0.8175	0.8185	0.8185	0.8185	0.8185
64	0.4400	0.4360	0.4390	0.4400	0.4400	0.4400	0.4400
65	0.4120	0.4115	0.4120	0.4120	0.4120	0.4120	0.4120
66	0.8330	0.8300	0.8325	0.8335	0.8330	0.8330	0.8330
67	0.8220	0.8190	0.8205	0.8220	0.8220	0.8220	0.8220
68	0.8545	0.8500	0.8530	0.8555	0.8545	0.8540	0.8540
69	0.7635	0.7565	0.7610	0.7640	0.7635	0.7635	0.7635
70	0.7870	0.7820	0.7870	0.7875	0.7870	0.7870	0.7870
71	0.7545	0.7465	0.7530	0.7560	0.7545	0.7545	0.7545

Table 4.21 Average Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Different ϕ when n=75.

ϕ		AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
0				With and without structural change				
	Average	0.7645	0.7703	0.7670	0.7630	0.7645	0.7649	0.7649
	Rank	6	1	2	7	5	3	4
	Stdev	0.1525	0.1332	0.1424	0.1598	0.1525	0.1509	0.1511
				With structural change				
	Average	0.7738	0.7732	0.7736	0.7738	0.7738	0.7737	0.7738
	Rank	2	7	6	1	2	5	5
	Stdev	0.1322	0.1321	0.1322	0.1322	0.1322	0.1322	0.1322
				Without structural change				
	Average	0.1190	0.5735	0.3080	0.0065	0.1200	0.1460	0.1420
	Rank	6	1	2	7	5	3	4
0.70				With and without structural change				
	Average	0.7638	0.7697	0.7667	0.7624	0.7638	0.7640	0.7641
	Rank	6	1	2	7	5	4	3
	Stdev	0.1529	0.1330	0.1420	0.1594	0.1527	0.1516	0.1515
				With structural change				
	Average	0.7731	0.7725	0.7734	0.7732	0.7731	0.7731	0.7731
	Rank	2	7	6	1	2	5	5
	Stdev	0.1320	0.1319	0.1316	0.1320	0.1320	0.1320	0.1320
				Without structural change				
	Average	0.1095	0.5780	0.3040	0.0085	0.1115	0.1315	0.1315
	Rank	6	1	2	7	5	3	4
1.0				With and without structural change				
	Average	0.7615	0.7665	0.7648	0.7607	0.7615	0.7536	0.7617
	Rank	5	1	2	6	4	7	3
	Stdev	0.1563	0.1364	0.1490	0.1598	0.1562	0.1752	0.1550
				With structural change				
	Average	0.7600	0.7666	0.7634	0.7593	0.7600	0.7520	0.7603
	Rank	2	7	6	1	2	5	5
	Stdev	0.1569	0.1374	0.1496	0.1604	0.1569	0.1760	0.1557
				Without structural change				
	Average	0.0550	0.4815	0.2010	0.0025	0.0555	0.0755	0.0745
	Rank	6	1	2	7	5	3	4
1.02				With and without structural change				
	Average	0.7610	0.7660	0.7629	0.7603	0.7610	0.7612	0.7611
	Rank	6	1	2	7	5	3	4
	Stdev	0.1551	0.1344	0.1459	0.1591	0.1551	0.1543	0.1544
				With structural change				
	Average	0.7710	0.7698	0.7706	0.7710	0.7710	0.7709	0.7709
	Rank	2	7	6	1	2	5	5
	Stdev	0.1315	0.1314	0.1315	0.1315	0.1315	0.1315	0.1315
				Without structural change				
	Average	0.0650	0.4980	0.2215	0.0055	0.0660	0.0785	0.0760
	Rank	6	1	2	7	5	3	4

Table 4.22 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 96 Models, $\phi = 0$ for $n = 100$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.3575	0.5450	0.3415	0.3015	0.1258	0.1740	0.2715
2	0.9015	0.8960	0.9005	0.9020	0.9015	0.9015	0.9015
3	0.9065	0.9065	0.9065	0.9065	0.9065	0.9065	0.9065
4	0.9030	0.9005	0.9020	0.9030	0.9030	0.9030	0.9030
5	0.8945	0.8930	0.8945	0.8945	0.8945	0.8945	0.8945
6	0.9005	0.9005	0.9005	0.9005	0.9005	0.9005	0.9005
7	0.8770	0.8765	0.8770	0.8770	0.8770	0.8770	0.8770
8	0.8655	0.8645	0.8655	0.8655	0.8655	0.8655	0.8655
9	0.5790	0.5785	0.5790	0.5790	0.5790	0.5790	0.5790
10	0.5590	0.5585	0.5590	0.5590	0.5590	0.5590	0.5590
11	0.8975	0.8970	0.8975	0.8975	0.8975	0.8975	0.8975
12	0.9165	0.9160	0.9165	0.9165	0.9165	0.9165	0.9165
13	0.8760	0.8760	0.8760	0.8760	0.8760	0.8760	0.8760
14	0.8790	0.8785	0.8790	0.8790	0.8790	0.8790	0.8790
15	0.9105	0.9105	0.9105	0.9105	0.9105	0.9105	0.9105
16	0.8575	0.8575	0.8575	0.8575	0.8575	0.8575	0.8575
17	0.7705	0.7700	0.7705	0.7705	0.7705	0.7705	0.7705
18	0.7795	0.7795	0.7795	0.7795	0.7795	0.7795	0.7795
19	0.7210	0.7295	0.7200	0.7210	0.7210	0.7210	0.7210
20	0.7115	0.7110	0.7115	0.7115	0.7115	0.7115	0.7115
21	0.9145	0.9145	0.9145	0.9145	0.9145	0.9145	0.9145
22	0.9245	0.9240	0.9245	0.9245	0.9245	0.9245	0.9245
23	0.9305	0.9295	0.9300	0.9305	0.9305	0.9305	0.9305
24	0.9215	0.9215	0.9215	0.9215	0.9215	0.9215	0.9215
25	0.8915	0.8915	0.8915	0.8915	0.8915	0.8915	0.8915
26	0.8390	0.8390	0.8390	0.8390	0.8390	0.8390	0.8390
27	0.8415	0.8415	0.8415	0.8415	0.8415	0.8415	0.8415
28	0.8880	0.8880	0.8880	0.8880	0.8880	0.8880	0.8880
29	0.8985	0.8980	0.8980	0.8985	0.8985	0.8985	0.8985
30	0.9110	0.9110	0.9110	0.9110	0.9110	0.9110	0.9110
31	0.8915	0.8915	0.8915	0.8920	0.8915	0.8915	0.8915
32	0.5325	0.5325	0.5325	0.5325	0.5325	0.5325	0.5325
33	0.5370	0.5370	0.5370	0.5370	0.5370	0.5370	0.5370
34	0.8810	0.8810	0.8810	0.8810	0.8810	0.8810	0.8810
35	0.8405	0.8405	0.8405	0.8405	0.8405	0.8405	0.8405
36	0.7775	0.7775	0.7775	0.7775	0.7775	0.7775	0.7775
37	0.7700	0.7700	0.7700	0.7700	0.7700	0.7700	0.7700
38	0.8685	0.8685	0.8685	0.8685	0.8685	0.8685	0.8685
39	0.9120	0.9115	0.9120	0.9120	0.9120	0.9120	0.9120
40	0.8880	0.8870	0.8875	0.8880	0.8880	0.8880	0.8880
41	0.9030	0.9025	0.9025	0.9030	0.9030	0.9030	0.9030
42	0.9225	0.9220	0.9225	0.9225	0.9225	0.9225	0.9225
43	0.8970	0.8960	0.8965	0.8970	0.8970	0.8970	0.8970
44	0.9115	0.9105	0.9115	0.9115	0.9115	0.9115	0.9115
45	0.9255	0.9250	0.9255	0.9255	0.9255	0.9255	0.9255
46	0.9295	0.9285	0.9290	0.9295	0.9295	0.9295	0.9295
47	0.9210	0.9200	0.9205	0.9210	0.9210	0.9210	0.9210

Table 4.22 (cont'd)

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
48	0.9255	0.9250	0.9255	0.9255	0.9255	0.9255	0.9255
49	0.8445	0.8435	0.8440	0.8445	0.8445	0.8440	0.8440
50	0.8440	0.8440	0.8440	0.8440	0.8440	0.8440	0.8440
51	0.8715	0.8715	0.8715	0.8715	0.8715	0.8715	0.8715
52	0.6330	0.6325	0.6330	0.6330	0.6330	0.6330	0.6330
53	0.6125	0.6125	0.6125	0.6125	0.6125	0.6125	0.6125
54	0.9265	0.9265	0.9265	0.9265	0.9265	0.9265	0.9265
55	0.5225	0.5225	0.5225	0.5225	0.5225	0.5225	0.5225
56	0.4985	0.4985	0.4985	0.4985	0.4985	0.4985	0.4985
57	0.9300	0.9295	0.9300	0.9300	0.9300	0.9300	0.9300
58	0.9175	0.9165	0.9170	0.9175	0.9175	0.9175	0.9175
59	0.8540	0.8530	0.8540	0.8540	0.8540	0.8540	0.8540
60	0.8520	0.8520	0.8520	0.8520	0.8520	0.8520	0.8520
61	0.9135	0.9135	0.9135	0.9135	0.9135	0.9135	0.9135
62	0.6455	0.6750	0.6450	0.6455	0.6455	0.6455	0.6455
63	0.6385	0.6380	0.8380	0.6385	0.6385	0.6385	0.6385
64	0.8930	0.8920	0.8930	0.8930	0.8930	0.8930	0.8930
65	0.7720	0.7720	0.7720	0.7720	0.7720	0.7720	0.7720
66	0.7285	0.7285	0.7285	0.7285	0.7285	0.7285	0.7285
67	0.8000	0.7995	0.8000	0.8000	0.8000	0.8000	0.8000
68	0.9155	0.9155	0.9155	0.9155	0.9155	0.9155	0.9155
69	0.8765	0.8765	0.8765	0.8765	0.8765	0.8765	0.8765
70	0.8830	0.8830	0.8830	0.8830	0.8830	0.8830	0.8830
71	0.9360	0.9345	0.9355	0.9360	0.9360	0.9360	0.9360
72	0.7720	0.7715	0.7715	0.7720	0.7720	0.7720	0.7720
73	0.5730	0.5725	0.5730	0.5730	0.5730	0.5730	0.5730
74	0.2940	0.2935	0.2940	0.2940	0.2940	0.2940	0.2940
75	0.2990	0.2985	0.2990	0.2990	0.2990	0.2990	0.2990
76	0.9155	0.9155	0.9155	0.9155	0.9155	0.9155	0.9155
77	0.5560	0.5560	0.5560	0.5560	0.5560	0.5560	0.5560
78	0.5680	0.5670	0.5675	0.5680	0.5680	0.5680	0.5680
79	0.7570	0.7565	0.7570	0.7570	0.7570	0.7570	0.7570
80	0.7750	0.7740	0.7740	0.7750	0.7750	0.7750	0.7750
81	0.9100	0.9090	0.9095	0.9100	0.9100	0.9100	0.9100
82	0.8950	0.8935	0.8950	0.8950	0.8950	0.8950	0.8950
83	0.8895	0.8895	0.8895	0.8895	0.8895	0.8895	0.8895
84	0.9080	0.9060	0.9075	0.9080	0.9080	0.9080	0.9080
85	0.8950	0.8920	0.8940	0.8950	0.8950	0.8950	0.8950
86	0.8975	0.8940	0.8965	0.8980	0.8975	0.8975	0.8975
87	0.9105	0.9080	0.9100	0.9115	0.9105	0.9105	0.9105
88	0.9035	0.8985	0.9030	0.9035	0.9035	0.9035	0.9035
89	0.5190	0.5165	0.5190	0.5190	0.5190	0.5190	0.5190
90	0.5185	0.5160	0.5180	0.5180	0.5190	0.5185	0.5185
91	0.6615	0.6655	0.6605	0.6620	0.6615	0.6615	0.6615
92	0.6770	0.6730	0.6760	0.6770	0.6770	0.6765	0.6765
93	0.8235	0.8170	0.8230	0.8235	0.8235	0.8235	0.8235
94	0.8090	0.8050	0.8085	0.8095	0.8090	0.8090	0.8090
95	0.4800	0.4755	0.4795	0.4805	0.4800	0.4800	0.4800
96	0.4270	0.4225	0.4265	0.4275	0.4270	0.4270	0.4270

Table 4.23 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 96 Models, $\phi = 0.7$ for $n = 100$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1965	0.5790	0.3000	0.1004	0.1980	0.1115	0.1100
2	0.8995	0.8970	0.8995	0.8995	0.8995	0.8995	0.8995
3	0.9060	0.9010	0.9055	0.9070	0.9060	0.9060	0.9060
4	0.9015	0.9010	0.9015	0.9015	0.9015	0.9015	0.9015
5	0.8925	0.8910	0.8925	0.8925	0.8925	0.8925	0.8925
6	0.8990	0.8985	0.8990	0.8990	0.8990	0.8990	0.8990
7	0.8795	0.8790	0.8790	0.8795	0.8795	0.8795	0.8795
8	0.8670	0.8655	0.8665	0.8675	0.8670	0.8670	0.8670
9	0.5645	0.5640	0.5645	0.5645	0.5645	0.5645	0.5645
10	0.5565	0.5565	0.5565	0.5565	0.5565	0.5565	0.5565
11	0.8955	0.8955	0.8955	0.8955	0.8955	0.8955	0.8955
12	0.9150	0.9145	0.9145	0.9150	0.9150	0.9150	0.9150
13	0.8740	0.8740	0.8740	0.8740	0.8740	0.8740	0.8740
14	0.8790	0.8790	0.8790	0.8790	0.8790	0.8790	0.8790
15	0.9095	0.9090	0.9090	0.9095	0.9095	0.9095	0.9095
16	0.8565	0.8565	0.8565	0.8565	0.8565	0.8565	0.8565
17	0.7690	0.7685	0.7690	0.7690	0.7690	0.7690	0.7690
18	0.7755	0.7755	0.7755	0.7755	0.7755	0.7755	0.7755
19	0.7205	0.7200	0.7205	0.7210	0.7205	0.7205	0.7205
20	0.7125	0.7120	0.7125	0.7125	0.7125	0.7125	0.7125
21	0.9150	0.9145	0.9150	0.9150	0.9150	0.9150	0.9150
22	0.9250	0.9240	0.9250	0.9250	0.9250	0.9250	0.9250
23	0.9285	0.9285	0.9285	0.9285	0.9285	0.9285	0.9285
24	0.9205	0.9205	0.9205	0.9205	0.9205	0.9205	0.9205
25	0.8915	0.8915	0.8915	0.8915	0.8915	0.8915	0.8915
26	0.8395	0.8395	0.8395	0.8395	0.8395	0.8395	0.8395
27	0.8450	0.8450	0.8450	0.8450	0.8450	0.8450	0.8450
28	0.8850	0.8850	0.8850	0.8850	0.8850	0.8850	0.8850
29	0.8975	0.8975	0.8975	0.8975	0.8975	0.8975	0.8975
30	0.9105	0.9100	0.9105	0.9105	0.9105	0.9105	0.9105
31	0.8875	0.8875	0.8875	0.8875	0.8880	0.8875	0.8875
32	0.5355	0.5355	0.5355	0.5360	0.5355	0.5355	0.5355
33	0.5280	0.5280	0.5280	0.5280	0.5280	0.5280	0.5280
34	0.8750	0.8745	0.8750	0.8750	0.8750	0.8750	0.8750
35	0.8385	0.8385	0.8385	0.8385	0.8385	0.8385	0.8385
36	0.7770	0.7770	0.7770	0.7770	0.7770	0.7770	0.7770
37	0.7650	0.7650	0.7650	0.7650	0.7650	0.7650	0.7650
38	0.8705	0.8705	0.8705	0.8705	0.8705	0.8705	0.8705
39	0.9070	0.9070	0.9070	0.9070	0.9070	0.9070	0.9070
40	0.8900	0.8890	0.8890	0.8900	0.8900	0.8900	0.8900
41	0.9015	0.9015	0.9015	0.9015	0.9015	0.9015	0.9015
42	0.9235	0.9235	0.9235	0.9235	0.9235	0.9235	0.9235
43	0.8980	0.8970	0.8970	0.8970	0.8980	0.8980	0.8980
44	0.9100	0.9095	0.9100	0.9100	0.9100	0.9100	0.9100
45	0.9265	0.9260	0.9265	0.9265	0.9265	0.9265	0.9265
46	0.9340	0.9335	0.9335	0.9340	0.9340	0.9340	0.9340
47	0.9225	0.9215	0.9220	0.9225	0.9225	0.9225	0.9225

Table 4.23 (cont'd)

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
48	0.9260	0.9255	0.9260	0.9260	0.9260	0.9260	0.9260
49	0.8455	0.8450	0.8450	0.8455	0.8455	0.8455	0.8455
50	0.8455	0.8455	0.8455	0.8455	0.8455	0.8455	0.8455
51	0.8705	0.8705	0.8705	0.8705	0.8705	0.8705	0.8705
52	0.6305	0.6300	0.9305	0.6305	0.6305	0.6305	0.6305
53	0.6130	0.6130	0.6130	0.6130	0.6130	0.6130	0.6130
54	0.9270	0.9270	0.9270	0.9270	0.9270	0.9270	0.9270
55	0.5215	0.5210	0.5210	0.5215	0.5215	0.5215	0.5215
56	0.4990	0.4990	0.4990	0.4990	0.4990	0.4990	0.4990
57	0.9305	0.9305	0.9305	0.9305	0.9305	0.9305	0.9305
58	0.9170	0.9155	0.9165	0.9170	0.9170	0.9170	0.9170
59	0.8525	0.8515	0.8520	0.8525	0.8525	0.8525	0.8525
60	0.8515	0.8515	0.8515	0.8515	0.8515	0.8515	0.8515
61	0.9135	0.9135	0.9135	0.9135	0.9135	0.9135	0.9135
62	0.6450	0.6445	0.6450	0.6450	0.6450	0.6450	0.6450
63	0.6325	0.6320	0.6325	0.6325	0.6325	0.6325	0.6325
64	0.8925	0.8915	0.8920	0.8925	0.8925	0.8925	0.8925
65	0.7660	0.7660	0.7660	0.7660	0.7660	0.7660	0.7660
66	0.7285	0.7285	0.7285	0.7285	0.7285	0.7285	0.7285
67	0.7950	0.7945	0.7945	0.7950	0.7950	0.7950	0.7950
68	0.9140	0.9140	0.9140	0.9140	0.9140	0.9140	0.9140
69	0.8810	0.8810	0.8810	0.8810	0.8810	0.8810	0.8810
70	0.8835	0.8825	0.8830	0.8835	0.8835	0.8835	0.8835
71	0.9360	0.9355	0.9360	0.9360	0.9360	0.9360	0.9360
72	0.7675	0.7665	0.7665	0.7675	0.7675	0.7675	0.7675
73	0.5680	0.5680	0.5680	0.5680	0.5680	0.5680	0.5680
74	0.2885	0.2885	0.2885	0.2885	0.2885	0.2885	0.2885
75	0.2920	0.2920	0.2920	0.2920	0.2920	0.2920	0.2920
76	0.9145	0.9145	0.9145	0.9145	0.9145	0.9145	0.9145
77	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495	0.5495
78	0.5665	0.5665	0.5665	0.5665	0.5665	0.5665	0.5665
79	0.7570	0.7570	0.7570	0.7570	0.7570	0.7570	0.7570
80	0.7770	0.7765	0.7765	0.7770	0.7770	0.7770	0.7770
81	0.9125	0.9125	0.9125	0.9125	0.9125	0.9125	0.9125
82	0.8965	0.8960	0.8965	0.8970	0.8965	0.8965	0.8965
83	0.8910	0.8900	0.8905	0.8910	0.8910	0.8910	0.8910
84	0.9090	0.9090	0.9090	0.9090	0.9090	0.9090	0.9090
85	0.8935	0.8920	0.8935	0.8935	0.8935	0.8935	0.8935
86	0.8965	0.8955	0.8960	0.8970	0.8965	0.8965	0.8965
87	0.9100	0.9085	0.9095	0.9095	0.9095	0.9100	0.9100
88	0.9050	0.9025	0.9045	0.9050	0.9050	0.9050	0.9050
89	0.5235	0.5225	0.5230	0.5240	0.5235	0.5235	0.5235
90	0.5290	0.5280	0.5285	0.5285	0.5290	0.5290	0.5290
91	0.6680	0.6665	0.6675	0.6685	0.6685	0.6680	0.6680
92	0.6900	0.6870	0.6890	0.6890	0.6900	0.6900	0.6900
93	0.8350	0.8315	0.8345	0.8360	0.8350	0.8350	0.8350
94	0.8185	0.8140	0.8180	0.8195	0.8185	0.8185	0.8185
95	0.4975	0.4945	0.4965	0.4980	0.4975	0.4975	0.4975
96	0.4380	0.4360	0.4380	0.4380	0.4380	0.4380	0.4380

Table 4.24 Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 96 Models, $\phi = 1.0$ for $n = 100$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.3047	0.5135	0.1965	0.2205	0.2475	0.1575	0.1550
2	0.9000	0.8970	0.9000	0.9000	0.9000	0.9000	0.9000
3	0.9055	0.9035	0.9055	0.9060	0.9055	0.9055	0.9055
4	0.9020	0.9005	0.9020	0.9020	0.9020	0.9020	0.9020
5	0.8900	0.8885	0.8895	0.8900	0.8900	0.8900	0.8900
6	0.8990	0.8990	0.8990	0.8990	0.8990	0.8990	0.8990
7	0.8775	0.8765	0.8770	0.8775	0.8775	0.8775	0.8775
8	0.8655	0.8640	0.8650	0.8655	0.8655	0.8655	0.8655
9	0.5770	0.5765	0.5770	0.5770	0.5770	0.5770	0.5770
10	0.5565	0.5560	0.5565	0.5565	0.5565	0.5565	0.5565
11	0.8945	0.8945	0.8945	0.8945	0.8945	0.8945	0.8945
12	0.9150	0.9150	0.9150	0.9150	0.9150	0.9150	0.9150
13	0.8775	0.8765	0.8775	0.8775	0.8775	0.8775	0.8775
14	0.8820	0.8815	0.8820	0.8820	0.8820	0.8820	0.8820
15	0.9040	0.9040	0.9040	0.9040	0.9040	0.9040	0.9040
16	0.8525	0.8525	0.8525	0.8525	0.8525	0.8525	0.8525
17	0.7610	0.7610	0.7610	0.7610	0.7610	0.7610	0.7610
18	0.7730	0.7730	0.7730	0.7730	0.7730	0.7730	0.7730
19	0.7230	0.7225	0.7230	0.7230	0.7230	0.7230	0.7230
20	0.7025	0.7015	0.7025	0.7025	0.7025	0.7025	0.7025
21	0.9105	0.9100	0.9105	0.9105	0.9105	0.9105	0.9105
22	0.9225	0.9210	0.9225	0.9225	0.9225	0.9225	0.9225
23	0.9280	0.9270	0.9280	0.9280	0.9280	0.9280	0.9280
24	0.9195	0.9185	0.9195	0.9195	0.9195	0.9195	0.9195
25	0.8890	0.8890	0.8890	0.8890	0.8890	0.8890	0.8890
26	0.8420	0.8410	0.8420	0.8420	0.8420	0.8420	0.8420
27	0.8460	0.8450	0.8460	0.8460	0.8460	0.8460	0.8460
28	0.8830	0.8825	0.8830	0.8830	0.8830	0.8830	0.8830
29	0.8975	0.8960	0.8975	0.8975	0.8975	0.8975	0.8975
30	0.9085	0.9085	0.9085	0.9085	0.9085	0.9085	0.9085
31	0.8875	0.8875	0.8875	0.8875	0.8875	0.8875	0 8875
32	0.5280	0.5275	0.5280	0.5280	0.5280	0.5280	0.5280
33	0.5310	0.5305	0.5305	0.5310	0.5310	0.5310	0.5310
34	0.8790	0.8790	0.8790	0.8790	0.8790	0.8790	0.8790
35	0.8395	0.8390	0.8390	0.8395	0.8395	0.8395	0.8395
36	0.7720	0.7720	0.7720	0.7720	0.7720	0.7720	0.7720
37	0.7690	0.7685	0.7690	0.7690	0.7690	0.7690	0.7690
38	0.8655	0.8655	0.8655	0.8655	0.8655	0.8655	0.8655
39	0.9080	0.9075	0.9080	0.9080	0.9080	0.9080	0.9080
40	0.8865	0.8865	0.8865	0.8865	0.8865	0.8865	0.8865
41	0.9010	0.9005	0.9010	0.9010	0.9010	0.9010	0.9010
42	0.9215	0.9210	0.9215	0.9215	0.9215	0.9215	0.9215
43	0.8945	0.8945	0.8945	0.8945	0.8945	0.8945	0.8945
44	0.9070	0.9070	0.9070	0.9070	0.9070	0.9070	0.9070
45	0.9265	0.9265	0.9265	0.9265	0.9265	0.9265	0.9265
46	0.9280	0.9275	0.9280	0.9280	0.9280	0.9280	0.9280
47	0.9220	0.9215	0.9220	0.9220	0.9220	0.9220	0.9220

Table 4.24 (cont'd)

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
48	0.9270	0.9265	0.9270	0.9270	0.9270	0.9270	0.9270
49	0.8455	0.8450	0.8455	0.8455	0.8455	0.8455	0.8455
50	0.8430	0.8425	0.8425	0.8430	0.8430	0.8430	0.8430
51	0.8740	0.8740	0.8740	0.8740	0.8740	0.8740	0.8740
52	0.6355	0.6355	0.6355	0.6355	0.6355	0.6355	0.6355
53	0.6130	0.6130	0.6130	0.6130	0.6130	0.6130	0.6130
54	0.9280	0.9280	0.9280	0.9280	0.9280	0.9280	0.9280
55	0.5155	0.5155	0.5155	0.5155	0.5155	0.5155	0.5155
56	0.4985	0.4985	0.4985	0.4985	0.4985	0.4985	0.4985
57	0.9275	0.9270	0.9275	0.9275	0.9275	0.9275	0.9275
58	0.9160	0.9155	0.9160	0.9160	0.9160	0.9160	0.9160
59	0.8520	0.8515	0.8520	0.8520	0.8520	0.8520	0.8520
60	0.8535	0.8535	0.8535	0.8535	0.8535	0.8535	0.8535
61	0.9125	0.9120	0.9120	0.9125	0.9125	0.9125	0.9125
62	0.6380	0.6375	0.6380	0.6380	0.6380	0.6380	0.6380
63	0.6360	0.6355	0.6360	0.6360	0.6360	0.6360	0.6360
64	0.8920	0.8915	0.8920	0.8920	0.8920	0.8920	0.8920
65	0.7670	0.7670	0.7670	0.7670	0.7670	0.7670	0.7670
66	0.7310	0.7310	0.7310	0.7310	0.7310	0.7310	0.7310
67	0.7980	0.7975	0.7980	0.7980	0.7980	0.7980	0.7980
68	0.9145	0.9145	0.9145	0.9145	0.9145	0.9145	0.9145
69	0.8830	0.8825	0.8825	0.8830	0.8830	0.8830	0.8830
70	0.8830	0.8830	0.8830	0.8830	0.8830	0.8830	0.8830
71	0.9335	0.9330	0.9335	0.9335	0.9335	0.9335	0.9335
72	0.7725	0.7715	0.7720	0.7725	0.7725	0.7725	0.7725
73	0.5715	0.5715	0.5715	0.5715	0.5715	0.5715	0.5715
74	0.2925	0.2925	0.2925	0.2925	0.2925	0.2925	0.2925
75	0.3015	0.3015	0.3015	0.3015	0.3015	0.3015	0.3015
76	0.9130	0.9130	0.9130	0.9130	0.9130	0.9130	0.9130
77	0.5555	0.5555	0.5555	0.5555	0.5555	0.5555	0.5555
78	0.5680	0.5675	0.5680	0.5680	0.5680	0.5680	0.5680
79	0.7615	0.7615	0.7615	0.7615	0.7615	0.7615	0.7615
80	0.7770	0.7770	0.7770	0.7770	0.7770	0.7770	0.7770
81	0.9140	0.9140	0.9140	0.9140	0.9140	0.9140	0.9140
82	0.8960	0.8955	0.8960	0.8960	0.8960	0.8960	0.8960
83	0.8895	0.8890	0.8890	0.8895	0.8895	0.8895	0.8895
84	0.9075	0.9075	0.9075	0.9075	0.9075	0.9075	0.9075
85	0.9005	0.8975	0.8995	0.9005	0.9005	0.9005	0.9005
86	0.8980	0.8965	0.8980	0.8980	0.8980	0.8980	0.8980
87	0.9125	0.9100	0.9115	0.9130	0.9125	0.9125	0.9125
88	0.9045	0.9030	0.9045	0.9045	0.9045	0.9045	0.9045
89	0.5245	0.5235	0.5245	0.5250	0.5245	0.5245	0.4525
90	0.5250	0.5230	0.5245	0.5250	0.5250	0.5250	0.5250
91	0.6740	0.6715	0.6730	0.6740	0.6740	0.6740	0.6740
92	0.6845	0.6815	0.6845	0.6850	0.6845	0.6845	0.6845
93	0.8340	0.8295	0.8330	0.8340	0.8340	0.8340	0.8340
94	0.8205	0.8180	0.8600	0.8220	0.8205	0.8205	0.8205
95	0.4935	0.4895	0.4930	0.4940	0.4935	0.4935	0.4935
96	0.4355	0.4340	0.4355	0.4355	0.4355	0.4355	0.4355

Table 4.25 Mean Probabilities of Correct ..lection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Selecting From 96 Models, $\phi = 1.02$ for $n = 100$.

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
1	0.1965	0.5750	0.2970	0.2765	0.1972	0.1090	0.1085
2	0.9025	0.8985	0.9020	0.9025	0.9025	0.9025	0.9025
3	0.9075	0.9030	0.9065	0.9085	0.9075	0.9075	0.9075
4	0.9035	0.9025	0.9035	0.9040	0.9035	0.9035	0.9035
5	0.8930	0.8915	0.8920	0.8930	0.8930	0.8925	0.8925
6	0.8995	0.8990	0.8995	0.8995	0.8995	0.8995	0.8995
7	0.8795	0.8795	0.8795	0.8795	0.8795	0.8795	0.8795
8	0.8670	0.8650	0.8665	0.8670	0.8670	0.8670	0.8670
9	0.5685	0.5680	0.5685	0.5685	0.5685	0.5685	0.5685
10	0.5585	0.5580	0.5585	0.5585	0.5585	0.5585	0.5585
11	0.8960	0.8955	0.8960	0.8960	0.8960	0.8960	0.8960
12	0.9160	0.9135	0.9155	0.9160	0.9160	0.9160	0.9160
13	0.8765	0.8765	0.8765	0.8765	0.8765	0.8765	0.8765
14	0.8785	0.8775	0.8785	0.8785	0.8785	0.8785	0.8785
15	0.9100	0.9095	0.9100	0.9100	0.9100	0.9100	0.9100
16	0.8585	0.8585	0.8585	0.8585	0.8585	0.8585	0.8585
17	0.7695	0.7685	0.7695	0.7695	0.7695	0.7695	0.7695
18	0.7740	0.7740	0.7740	0.7740	0.7740	0.7740	0.7740
19	0.7180	0.7175	0.7175	0.7180	0.7180	0.7180	0.7180
20	0.7105	0.7100	0.7105	0.7105	0.7105	0.7105	0.7105
21	0.9145	0.9135	0.9145	0.9145	0.9145	0.9145	0.9145
22	0.9240	0.9235	0.9240	0.9245	0.9240	0.9240	0.9240
23	0.9300	0.9295	0.9300	0.9300	0.9300	0.9300	0.9300
24	0.9230	0.9230	0.9230	0.9230	0.9230	0.9230	0.9230
25	0.8915	0.8915	0.8915	0.8915	0.8915	0.8915	0.8915
26	0.8415	0.8415	0.8415	0.8415	0.8415	0.8415	0.8415
27	0.8445	0.8445	0.8445	0.8445	0.8445	0.8445	0.8445
28	0.8860	0.8860	0.8860	0.8860	0.8860	0.8860	0.8860
29	0.8975	0.8975	0.8975	0.8975	0.8975	0.8975	0.8975
30	0.9105	0.9100	0.9105	0.9105	0.9105	0.9105	0.9105
31	0.8905	0.8900	0.8900	0.8905	0.8905	0.8900	0.8900
32	0.5385	0.5375	0.5380	0.5385	0.5385	0.5380	0.5380
33	0.5355	0.5350	0.5355	0.5355	0.5355	0.5355	0.5355
34	0.8790	0.8785	0.8790	0.8790	0.8790	0.8790	0.8790
35	0.8375	0.8370	0.8375	0.8375	0.8375	0.8375	0.8375
36	0.7705	0.7705	0.7705	0.7705	0.7705	0.7705	0.7705
37	0.7685	0.7685	0.7685	0.7685	0.7685	0.7685	0.7685
38	0.8670	0.7670	0.7670	0.7670	0.7670	0.7670	0.7670
39	0.9075	0.9075	0.9075	0.9075	0.9075	0.9075	0.9075
40	0.8890	0.8880	0.8880	0.8890	0.8890	0.8890	0.8890
41	0.9025	0.9020	0.9020	0.9025	0.9025	0.9025	0.9025
42	0.9230	0.9230	0.9230	0.9230	0.9230	0.9230	0.9230
43	0.8995	0.8985	0.8985	0.8995	0.8995	0.8995	0.8995
44	0.9115	0.9110	0.9115	0.9115	0.9115	0.9115	0.9115
45	0.9265	0.9260	0.9265	0.9265	0.9265	0.9265	0.9265
46	0.9290	0.9285	0.9285	0.9290	0.9290	0.9290	0.9290
47	0.9215	0.9205	0.9205	0.9205	0.9215	0.9210	0.9210

Table 4.25 (cont'd)

Model	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
48	0.9255	0.9250	0.9255	0.9255	0.9255	0.9255	0.9255
49	0.8465	0.8465	0.8465	0.8465	0.8465	0.8465	0.8465
50	0.8435	0.8435	0.8435	0.8435	0.8435	0.8435	0.8435
51	0.8720	0.8720	0.8720	0.8720	0.8720	0.8720	0.8720
52	0.6325	0.6325	0.6325	0.6325	0.6325	0.6325	0.6325
53	0.6174	0.6140	0.6140	0.6140	0.6140	0.6140	0.6140
54	0.9275	0.9275	0.9275	0.9275	0.9275	0.9275	0.9275
55	0.5230	0.5225	0.5230	0.5230	0.5230	0.5230	0.5230
56	0.5020	0.5020	0.5020	0.5020	0.5020	0.5020	0.5020
57	0.9280	0.9275	0.9280	0.9280	0.9280	0.9280	0.9080
58	0.9165	0.9150	0.9165	0.9165	0.9165	0.9165	0.9165
59	0.8525	0.8525	0.8525	0.8525	0.8525	0.8525	0.8525
60	0.8510	0.8510	0.8510	0.8510	0.8510	0.8510	0.8510
61	0.9085	0.9075	0.9085	0.9085	0.9085	0.9085	0.9085
62	0.6415	0.6415	0.6415	0.6415	0.6415	0.6415	0.6415
63	0.6380	0.6375	0.6380	0.6380	0.6380	0.6380	0.6380
64	0.8930	0.8920	0.8925	0.8930	0.8930	0.8930	0.8930
65	0.7700	0.7700	0.7700	0.7700	0.7700	0.7700	0.7700
66	0.7290	0.7290	0.7290	0.7290	0.7290	0.7290	0.7290
67	0.7985	0.7985	0.7985	0.7985	0.7985	0.7985	0.7985
68	0.9175	0.9175	0.9175	0.9175	0.9175	0.9175	0.9175
69	0.8820	0.8820	0.8820	0.8820	0.8820	0.8820	0.8820
70	0.8840	0.8835	0.8840	0.8840	0.8840	0.8840	0.8840
71	0.9365	0.9355	0.9365	0.9365	0.9365	0.9365	0.9365
72	0.7720	0.7710	0.7720	0.7720	0.7720	0.7720	0.7720
73	0.5675	0.5670	0.5675	0.5675	0.5675	0.5675	0.5675
74	0.2920	0.2920	0.2920	0.2920	0.2920	0.2920	0.2920
75	0.2910	0.2910	0.2910	0.2910	0.2910	0.2910	0.2910
76	0.9130	0.9130	0.9130	0.9130	0.9130	0.9130	0.9130
77	0.5545	0.5545	0.5545	0.5545	0.5545	0.5545	0.5545
78	0.5675	0.5675	0.5675	0.5675	0.5675	0.5675	0.5675
79	0.7570	0.7565	0.7570	0.7570	0.7570	0.7570	0.7570
80	0.7780	0.7775	0.7775	0.7780	0.7780	0.7780	0.7780
81	0.9130	0.9120	0.9125	0.9135	0.9130	0.9130	0.9130
82	0.8955	0.8950	0.8955	0.8955	0.8955	0.8955	0.8955
83	0.8900	0.8890	0.8895	0.8900	0.8900	0.8900	0.8900
84	0.9095	0.9090	0.9095	0.9085	0.9095	0.9095	0.9095
85	0.8945	0.8930	0.8945	0.8945	0.8945	0.8945	0.8945
86	0.8990	0.8970	0.8975	0.8990	0.8990	0.8985	0.8985
87	0.9115	0.9095	0.9100	0.9125	0.9115	0.9115	0.9115
88	0.9020	0.8990	0.9015	0.9020	0.9020	0.9015	0.9015
89	0.5320	0.5305	0.5310	0.5320	0.5320	0.5320	0.5320
90	0.5245	0.5230	0.5240	0.5250	0.5245	0.5245	0.5245
91	0.6715	0.6695	0.6715	0.6720	0.6715	0.6715	0.6715
92	0.6875	0.6850	0.6860	0.6875	0.6875	0.6875	0.6875
93	0.8335	0.8305	0.8330	0.8340	0.8335	0.8335	0.8335
94	0.8165	0.8140	0.8155	0.8170	0.8165	0.8165	0.8165
95	0.4950	0.4900	0.4930	0.4950	0.4950	0.4950	0.4950
96	0.4305	0.4290	0.4305	0.4305	0.4305	0.4305	0.4305

Table 4.26 Average Mean Probabilities of Correct Selection for AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC for Different ϕ when $n = 100$.

ϕ		AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC
0				With and without structural change				
	Average	0.7914	0.7946	0.7922	0.7895	0.7904	0.7905	0.7903
	Rank	3	1	2	7	5	4	6
	Stdev	0.1694	0.1550	0.1617	0.1733	0.1692	0.1687	0.1686
				With structural change				
	Average	0.7988	0.7969	0.7974	0.7977	0.7977	0.7976	0.7974
	Rank	1	7	5	2	3	5	6
	Stdev	0.1543	0.1542	0.1542	0.1542	0.1542	0.1542	0.1540
				Without structural change				
	Average	0.0965	0.5750	0.2970	0.0065	0.0970	0.1090	0.1085
	Rank	6	1	2	7	5	3	4
0.70				With and without structural change				
	Average	0.7909	0.7953	0.7960	0.7900	0.7909	0.7911	0.7911
	Rank	6	2	1	7	5	3	4
	Stdev	0.1697	0.1555	0.1618	0.1739	0.1697	0.1691	0.1692
				With structural change				
	Average	0.7982	0.7976	0.8012	0.7983	0.7982	0.7982	0.7982
	Rank	1	7	5	2	3	5	6
	Stdev	0.1547	0.1547	0.1543	0.1547	0.1547	0.1547	0.1547
				Without structural change				
	Average	0.0965	0.5790	0.3000	0.0040	0.0980	0.1115	0.1100
	Rank	6	1	2	7	5	3	4
1.0				With and without structural change				
	Average	0.7902	0.7944	0.7921	0.7898	0.7902	0.7903	0.7896
	Rank	5	1	2	6	4	3	7
	Stdev	0.1713	0.1559	0.1652	0.1735	0.1713	0.1708	0.1723
				With structural change				
	Average	0.7981	0.7974	0.7984	0.7981	0.7981	0.7981	0.7973
	Rank	1	7	5	2	3	5	6
	Stdev	0.1540	0.1540	0.1541	0.1540	0.1540	0.1540	0.1555
				Without structural change				
	Average	0.0470	0.5135	0.1965	0.0005	0.047	0.0575	0.0550
	Rank	6	1	2	7	5	3	4
1.02				With and without structural change				
	Average	0.7906	0.7953	0.7944	0.7901	0.7906	0.7908	0.7908
	Rank	6	1	2	7	5	3	4
	Stdev	0.1715	0.1558	0.1633	0.1741	0.1714	0.1707	0.1709
				With structural change				
	Average	0.7983	0.7979	0.8002	0.7984	0.7984	0.7983	0.7983
	Rank	1	7	5	2	3	5	6
	Stdev	0.1547	0.1544	0.1539	0.1547	0.1547	0.1547	0.1547
				Without structural change				
	Average	0.0575	0.5450	0.2415	0.0000	0.0580	0.0740	0.0710
	Rank	6	1	2	7	5	3	4

CHAPTER 5

An Optimal Method for Finding Penalties for the Problem of Detecting Structural Change[1]

5.1 Introduction

In Chapter 4, we investigated the use of several IC model selection procedures for detecting simple structural change in the linear regression model. Our simulation results showed that BIC outperformed all existing IC procedures considered when there was no structural change but was the worst performer in the presence of structural change, and RSC performed best overall when a changepoint was present. When there is no structural change, the performance of AIC is generally the worst of all procedures. The results also revealed that none of the IC procedures considered stand out as a clear best method for this model selection problem.

As is well known, the penalty function suggested for different IC procedures depends very much on the number of parameters (k) of the model and generally also

[1] Some of the preliminary findings of this chapter were presented at the Australasian Meeting of the Econometric Society, Australian National University, Canberra. See Azam and King (1998).

on the sample size (n) of the data. An obvious question then is can we use simulation methods to find the optimal penalties for model selection of possible structural change in a linear regression model? In Chapter 4, we used the AMPCS criteria to assess the quality of different IC procedures. Clearly, we could use this to determine an optimal penalty function. By an optimal penalty function, we mean one that maximizes the average mean probability of correctly selecting the true model. It can then offer an appropriate solution to a wide class of selection problems for structural change models. In particular, this penalty function would balance penalizing additional parameters while being sensitive to give a small enough penalty to select bigger models when they are true. The optimal penalties have a special property in that they are constructed in such a way that no one model is favoured unknowingly. Kwek (1999), King and Bose (2000), and Billah and King (2000a) have employed this approach with some success in the cases of selecting between ARCH, linear regression and time series models respectively. Unfortunately, none of these studies involved choosing between a larger number of different models, such as 96 models in the previous chapter when $n = 100$.

Existing IC procedures penalise changepoint models equally although the changepoint models vary from one to another by the position of the changepoint in time. This is a weakness of these procedures. It may be that different changepoint models should be penalised differently because changes in some cases are easier to detect. With this view in mind, our aim is to develop an algorithm that would compute optimal penalties for different models involving structural change. We investigate the use of grid search, polynomial of degree four combined with grid search, and simulated annealing optimization algorithms that will estimate optimal penalties for different models in such a way that the AMPCS is maximized. The latter is a difficult maximization problem that can be very time consuming to solve.

We performed a small Monte Carlo experiment to calculate the penalties for different models using these algorithms. We look for the algorithm that gives the optimal penalties in a sense that these penalties will provide maximum AMPCS with minimum computational cost and effort.

An outline of this chapter is as follows. In Section 5.2, we discuss the issues in finding optimal penalties. Section 5.3 describes the algorithm for finding optimal penalties for different models that includes the grid search method, polynomial of degree four based on grid search and simulated annealing. Section 5.4 gives a working version of each of the algorithms and evaluates them by discussing their advantages and disadvantages. Section 5.5 gives the details of the simulation experiments. The results of these experiments are reported in Section 5.6 and the final section contains some concluding remarks.

5.2 Derivation of the Procedure for Finding Optimal Penalties

In the literature on IC based model selection procedures, there is disagreement about the proper form of the penalty function. This is because, from the definition of IC, obviously, one can easily suggest a new criterion by slightly changing the value of the penalty function. For this reason, researchers introduced various IC based procedures for different types of models. As a result, there has been a huge growth in the literature that may make the users confused as to which IC procedure to use for a particular problem in hand. Further, the small sample performance of these new IC procedures may not be satisfactory. Therefore, an IC based procedure that would perform well for any kind of model selection problem is of interest.

As explained in Chapter 4, the AMPCS is calculated by averaging the mean of the probabilities of correct selection (MPCS) for all models in the plausible group. All of these applications produced, on average, a high probability of selecting the true model and can be used to find penalties in such a way that the probabilities of correctly choosing the right model at each of these points are the same. In other words, optimal penalties are constructed in such a way that no one model is being favoured unknowingly. In our proposed procedure, we use this philosophy with a modification that will result in the optimal penalties that maximize the AMPCS for the set of models under consideration. This allows us to find a data driven penalty that depends on the nature of the data and the sample size. This approach can also be used for other model selection problems with a large number of alternative models. For the sample size and plausible models under consideration, this model selection approach will maximize the AMPCS through the estimation of penalty values numerically. The AMPCS is a step function, and hence, it may not be easy to maximize it using standard methods.

Our interest is in finding a suitable algorithm for estimating appropriate optimal penalties that will save computing time and at the same time, giving penalties, which outperform all existing IC model selection procedures when looking for changepoints of unknown timing. Our approach involves an optimization principle with a simple probabilistic algorithm, i.e., finding penalties that maximize the AMPCS discussed in Chapter 4.

5.3 Optimization Methods for Optimal Penalties

Analytically we are not able to outline how to find optimal penalties $p_1, \ldots, p_m$ for changepoint models. In this section, we outline how to use the computer to maximize AMPCS so that optimal penalties are obtained. Note that because the IC procedure involves looking at differences in penalties, one of these penalties can be fixed. We set $p_1 = 0$ without loss of generality. The derivative methods of the optimization, such as the Newton-Raphson type procedure, are not entirely satisfactory for finding global maxima because they might sometimes end up with local maxima, so other methods need to be used. We use the derivative-free grid search, polynomial equation based on grid search and simulated annealing methods for finding optimal penalties, and these are discussed in the following subsections.

5.3.1 Complete Grid Search Algorithm (CGSA)

This section discusses the development of the CGSA for finding optimal penalties. We can maximize AMPCS with respect to the penalties by a trial and error method. The disadvantage is that it takes a long time to find the global optimum and might end up giving a local optimum instead of the global maximum. A CGSA can be used to find the global optimization point. It evaluates the function at grid points that cover the entire range of possible penalty values. The CGSA inspects the results and repeats the process with a finer grid over a selected zone, which is centred at the penalty values for the largest calculated value of AMPCS. In this section, we outline and discuss the grid search algorithm.

We set $p_1 = 0$ as mentioned earlier so that the AMPCS is maximized with respect to the remaining penalties, i.e., $p_2, \ldots, p_m$. Thus, the optimal penalties are found when the AMPCS is at its global maximum. In order to understand more closely what is involved in the grid search algorithm, we consider a set of penalty values (i.e., p_j, $j = 2, \cdots, m$). Let l_j and u_j be preselected lower and upper limits (l_j and u_j could be the same or different for different values of j) for the penalties. Beginning from the lower value l_j, a sequence of penalties is generated such that the difference between any two successive elements of the sequence is equal to ε, where ε is a small number.

We evaluate the AMPCS at grid points that covers the entire area or a likely area, inspect the result, and repeat the process over a selected zone with a finer grid. For each set of penalties, the AMPCS is calculated and then recorded. We then check for which penalty set the AMPCS is a maximum. Let p_j^* be the selected penalty set. We then generate a new sequence $p_j^* - \varepsilon^*$ to $p_j^* + \varepsilon^*$ and for each j where ε^* is a new value other than ε and such that $\varepsilon^* < \varepsilon$. Using the same procedure discussed earlier, we select the maximum AMPCS. We change ε^* to ε^{**} such that $\varepsilon^{**} < \varepsilon^*$ and repeat the above process. We continue the grid search procedure until changing the ε values does not change the maximum AMPCS with in a prescribed level of tolerance. Ultimately the search converges to the global maximum giving optimal penalties. The algorithm we applied in the experiments described below is as follows:

Step 1. Set $p_1 = 0$. Set upper and lower limits of the grid points for the penalty of each model. The spacing and number of grid points depend on the choice of the researcher. For computational convenience, we recommend equal

spacing. The larger the number of grid points for each penalty, the more computational time needed. If the number of grid points for each penalty is n_g then for m models, we need n_g^{m-1} penalty combinations and therefore n_g^{m-1} evaluations of AMPCS.

Step 2. Calculate maximised log-likelihoods for different models assuming each of the models is true in turn. These maximized log-likelihoods are stored. Subtract each set of penalties considered in step 1 from these maximised log-likelihoods for different models. The one that is the maximum is our chosen model. Perform this process for R replications, calculating Rm sets of penalised maximised log-likelihoods for each model as the true model in turn and count how many times the true model gives the largest value. Calculate MPCS by dividing these counts by R. This is done with each model being the true model in turn.

Step 3. For each set of penalties, calculate AMPCS using (4.10) and check for which penalty set the AMPCS is a maximum. The penalty values corresponding to the highest calculated value of AMPCS is recorded, then a new, but much finer, grid of penalty values is calculated and the process is repeated. The whole process is repeated several times. The CGSA ends by comparing the last maximum AMPCS with the most recent maximum AMPCS. Check the difference; if the difference is relatively small, stop the algorithm.

Step 4. From step 3, check for which penalty set the AMPCS is a maximum. This is the required optimal penalty set.

5.3.1.1 Block Grid Search Algorithm (BGSA)

The CGSA takes a huge amount of computational time even with a high-speed computer. If the number of grid points per penalty is small then the total number of penalty sets will be small and the cost of computation will be less. Literally, 2 grid points per penalty set is the minimum required points to start with a grid search, but in practice more grid points per penalty set will help to locate the global maximum. The disadvantage of taking more grid points per penalty is a relatively high computational cost. A naturally arising question is how to choose the minimum number of grid points per penalty that will provide minimum computational cost with little sacrifice of AMPCS.

If we have $m-1$ penalties with n_g grid points for each penalty set and if m and n_g are large, therefore we have n_g^{m-1} penalty combinations that will give a huge number of computations and take an enormous amount of computational time. However, we might not need all n_g^{m-1} penalty combinations. In this section, we discuss a modification to the CGSA and develop BGSA that gives almost the same solution as CGSA but has a significant improvement in speed of computational time over CGSA. In this algorithm, we use five grid points and five model groups giving 5^5 penalty combinations. The algorithm involves the following steps:

Step 1. Without any loss of generality, we set $p_1 = 0$.

Step 2. Divide the rest of the models into 5 groups such that each group consists of $(n-4)/5$ models. Figures 5.1 to 5.8 shows plot of estimated optimal penalties for $n = 15$ and 25 for different values of ϕ by CGSA. These plots give us an indication that penalties for different models follow more or less

a symmetric pattern at the beginning and end period of the data duration, therefore we set penalties for different models as illustrated below:

For $n = 15$, or $m = 12$:

$$p_1 = 0;\ p_2 = p_{12};\qquad p_3 = p_{11};$$

$$p_4 = p_{10};\qquad p_5 = p_9;\qquad p_6 = p_7 = p_8.$$

For $n = 25$, or $m = 22$:

$$p_1 = 0;\qquad p_2 = p_3 = p_{21} = p_{22};$$

$$p_4 = p_5 = p_{19} = p_{20};\qquad p_6 = p_7 = p_{17} = p_{18};$$

$$p_8 = p_9 = p_{15} = p_{16};\qquad p_{10} = p_{11} = p_{12} = p_{13} = p_{14}.$$

For $n = 50$, or $m = 47$:

$$p_1 = 0;$$

$$p_2 = p_3 = p_4 = p_5 = p_6 = p_{43} = p_{44} = p_{45} = p_{46} = p_{47};$$

$$p_7 = p_8 = p_9 = p_{10} = p_{11} = p_{38} = p_{39} = p_{40} = p_{41} = p_{42};$$

$$p_{12} = p_{13} = p_{14} = p_{15} = p_{16} = p_{33} = p_{34} = p_{35} = p_{36} = p_{37};$$

$$p_{17} = p_{18} = p_{19} = p_{20} = p_{21} = p_{28} = p_{29} = p_{30} = p_{31} = p_{32};$$

$$p_{22} = p_{23} = p_{24} = p_{25} = p_{26} = p_{27}.$$

For $n = 75$, or $m = 72$:

$$p_1 = 0;$$

$$p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_{66} = p_{67} = p_{68} = p_{69} = p_{70} = p_{71} = p_{72};$$

$$p_9 = p_{10} = p_{11} = p_{12} = p_{13} = p_{14} = p_{15} = p_{59} = p_{60} = p_{61} = p_{62} = p_{63} = p_{64} = p_{65};$$

$$p_{16} = p_{17} = p_{18} = p_{19} = p_{20} = p_{21} = p_{22} = p_{52} = p_{53} = p_{54} = p_{55} = p_{56} = p_{57} = p_{58};$$

$$p_{23} = p_{24} = p_{25} = p_{26} = p_{27} = p_{28} = p_{29} = p_{45} = p_{46} = p_{47} = p_{48} = p_{49} = p_{50} = p_{51};$$

$$p_{30} = p_{31} = p_{32} = p_{33} = p_{34} = p_{35} = p_{36} = p_{37} = p_{38} = p_{39} = p_{40} = p_{41} = p_{42} = p_{43} = p_{44};$$

For $n = 100$, or $m = 97$:

$$p_1 = 0;$$

$$p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9 = p_{10} = p_{11} = p_{88} = p_{89} = p_{90} = p_{91} = p_{92} = p_{93} = p_{94} = p_{95} = p_{96} = p_{97};$$

$$p_{12} = p_{13} = p_{14} = p_{15} = p_{16} = p_{17} = p_{18} = p_{19} = p_{20} = p_{21} = p_{78} = p_{79} = p_{80} = p_{81} = p_{82} = p_{83} = p_{84} = p_{85} = p_{86} = p_{87};$$

$$p_{22} = p_{23} = p_{24} = p_{25} = p_{26} = p_{27} = p_{28} = p_{29} = p_{30} = p_{31} = p_{68} = p_{69} = p_{70} = p_{71} = p_{72} = p_{73} = p_{74} = p_{75} = p_{76} = p_{77};$$

$$p_{32} = p_{33} = p_{34} = p_{35} = p_{36} = p_{37} = p_{38} = p_{39} = p_{40} = p_{41} = p_{58} = p_{59} = p_{60} = p_{61} = p_{62} = p_{63} = p_{64} = p_{65} = p_{66} = p_{67};$$

$$p_{42} = p_{43} = p_{44} = p_{45} = p_{46} = p_{47} = p_{48} = p_{49} = p_{50} = p_{51} = p_{52} = p_{53} = p_{54} = p_{55} = p_{56} = p_{57}.$$

Step 3. Calculate maximised log-likelihoods for different models assuming each of the models is true in turn. These maximized log-likelihoods are stored. Subtract each set of penalties considered in step 2 from these maximised log-likelihoods for different models. These penalized maximized log-likelihoods are compared. The one that is the maximum indicates our chosen model. Perform this process for R replications, calculate Rm sets of penalised maximised log-likelihoods for each model as the true model in turn and count how many times the true model gives the largest value. Calculate MPCS by dividing these counts by R. This is done with each model being the true model in turn.

Step 4. For each set of penalties, calculate the AMPCS and check for which penalty set the AMPCS is a maximum. The penalty values corresponding to the highest calculated value of AMPCS is recorded, then a new, but much finer, grid of penalty values is calculated and the process is repeated. The whole process is repeated several times. The BGSA ends by comparing the last

maximum AMPCS with the most recent maximum AMPCS. Check the difference; if the difference is relatively small, stop the algorithm.

5.3.1.2 Polynomial of Degree Four Algorithm (PDFA)

We have many different models, because for each timing of the possible changepoint we have one model. The optimal penalty for neighbouring models might be more or less the same. As the timing of the changepoint moves through the sample, we might except the optimal penalty to change in a smooth manner, which might be well approximated by a polynomial of some degree, which is not known. Based on this idea, consider a polynomial of degree four to estimate optimal penalties, known as PDFA.

PDFA is similar to that of BGSA, but assumes penalties for different models follow a polynomial function because we are assuming optimal penalties change slowly as the timing of changepoints change. A polynomial allows gradual but also substantial change if required. We ask the data to speak in the sense of finding the best penalties constrained by a polynomial of degree four. The algorithm used for this purpose in now given.

Step 1. Without any loss of generality, we set $p_1 = 0$.

Step 2. We consider the following penalty function

$$p_{i+1} = a_0 + a_1(i) + a_2(i)^2 + a_3(i)^3 + a_4(i)^4 \qquad (5.1)$$

where $i = 1, 2, \ldots, m-1$, denotes the timing of the changepoint we considered for different changepoint models. Our aim is to find a_0, a_1, a_2, a_3, a_4 so that p_{i+1} from (5.1) maximize AMPCS.

Step 3. Choose initial values of p_i's from AIC penalties (or from any other IC penalties). Solve the polynomial (5.1) for the a_i's with the initial values of

p_i's. Because the number of equations is more than the number of unknown coefficients, we only consider five equations and solve for the five unknown coefficients, a_i. We used penalties for the equations $i = 0, \frac{m}{4}, \frac{m}{2}, \frac{3m}{4}$ as the timing of the changepoints and then solved for a_i, $i = 0,\ldots,4$.

Step 4. Perform a grid search on the estimated $\hat{a}_i$'s found in step 3. Consider a set of different values for a_i, namely $a_i = a_{i1}, a_{i2}, \cdots, a_{in_g} = a_u$ for each model so that a range of a_i values from lower limit a_l to upper limit is a_u are considered. Changes of the grid and the grid search are done using same procedure discussed in Section 5.3.1. If we consider 5 grid points for 5 sets of a_i's, then we have 5^5 sets of a_i values. Substitute each set of a_i values in equation (5.1) and calculate different sets of penalties for different models.

Step 5. Calculate maximised log-likelihoods for the different models assuming each of the models is true. Subtract each set of penalties from these maximised log-likelihoods for different models and choose that model which gives the largest penalised maximised log-likelihood. Perform this process using R replications, and count how many times the true model gives the largest value. Calculate MPCS by dividing these counts by R. This is done for each model as the true model in turn and for each set of $\hat{a}_i$ values (which determine the p_i values) in the grid. Average the m MPCS values to obtain the AMPCS for all models. Check for that set of $\hat{a}_i$ which gives the maximum AMPCS. The $\hat{a}_i$ set estimated here are the required polynomial coefficients for use in (5.1) to calculate the required penalties.

5.3.2 Simulated Annealing Algorithm

In previous sections, we introduced CGSA, BGSA, and PDFA to estimate optimal penalties for different models that give maximum AMPCS. We have noted that the CGSA requires a huge amount of computational time. The other two algorithms save computational time by imposing some restrictions on the penalties. We now focus our attention on reducing the computational time for finding optimal penalties using stochastic optimization techniques. Simulated annealing algorithms (SAAs) have recently gained a great deal of the attention in the optimization literature. A discussion of the SAA was given in Chapter 2 and we investigate the use of this class of algorithms to estimate optimal penalties in this chapter.

SAA's major advantage over other methods is its ability to avoid becoming trapped at local optima. The algorithm employs a random search, which not only accepts changes that increase the objective function, but also some changes that decrease it. As its name implies, simulated annealing exploits an analogy between the way that a metal cools and freezes into a minimum energy crystalline structure (the annealing process) and the search for an optimum of a general function.

5.3.2.1 Estimation of Penalties by Simulated Annealing

We are looking for the penalties that maximize AMPCS. The algorithm can then be formulated as follows. Let p be the vector of penalties, with $p=(p_2,\cdots,p_m)'$ and let $f(p)$ represent penalties which need to be optimized. Let the initial penalty set be p_s. The algorithm moves both up and downhill as the optimization process is carried out and looks for the area where the optimum is. It randomly selects a penalty from the neighbourhood of the current penalty and then

calculates the corresponding change in $f(p)$. Let $l_i \le p_i \le u_i$, $i = 2, \ldots, m$, where the values of l_i and u_i are lower and upper bounds of the penalty values chosen by the researcher.

At the beginning, using the initial p_s, the algorithm randomly chooses a new point p' within the step length determined by V_{m-1}, a step length vector of order $m-1$ selected by the researcher, in the neighbourhood of p_s. The function is evaluated at this new point and its value is compared to that of the initial point p_s. If the change in $f(p)$ is positive, the transition is unconditionally accepted; if the $f(p)$ decreases, the transition is accepted with a probability based upon the distribution

$$\Pr = \exp\left(-\frac{\Delta f(p)}{kT}\right) \tag{5.2}$$

where k is a constant selected by the user, the temperature T is a control parameter and $\Delta f(p)$ is the change in $f(p)$.

Each element of the step length vector V_{m-1} is adjusted periodically so that about half of all points are accepted. The standard implementation of the algorithm requires the specification of a cooling schedule. The initial temperature should be high enough to ensure that there is little chance of the algorithm moving very quickly towards a local maximum in the early stages. A fall in temperature is imposed upon the system with a temperature reduction factor r_T ranging from 0 to 1. Finally, a stopping criterion is imposed to terminate the algorithm.

5.3.3 Different Factors of the SAA

The basic principle of estimation of penalties using simulated annealing was discussed in subsection 5.3.2.1. The algorithm randomly chooses a new penalty and calculates AMPCS, which is compared with the previous estimated value. Out of these two penalty sets, we check which gives the maximum AMPCS. The choice of penalties depends on a few factors; a detailed description of how to implement the algorithm for optimum penalty estimation is given below:

5.3.3.1 Initial Temperature

Initial temperature is an important factor for running the SAA. Kirkpatrick et al. (1983) suggested that a suitable initial temperature should be chosen so that about 80% of all positive transitions (i.e., transitions that increase the AMPCS) are accepted.

5.3.3.2 Temperature Reduction Factor (r_T)

The temperature is decreased by multiplication by a constant factor. There is actually a trade-off between temperature reduction between stages and the number of iterations per stage. The most common temperature reduction rule is $r_T + 1 = \alpha r_T$, where α is a constant close to, but smaller than, 1. Kirkpatrick et al. (1982) used it with $\alpha = 0.95$. The suggested value for temperature reduction by Corana et al. (1987) is 0.85.

5.3.3.3 Boundaries for Penalties

Let the lower bound for the allowable penalties be l_i and the upper bound for the penalties be u_i. Unless the user wishes to concentrate the search to a particular region, l_i and u_i should be set to cover a very large range.

5.3.3.4 Number of Cycles (n_s)

n_s is the number of cycles used before adjusting the step length vector. When the SAA starts its operation, it evaluates n_s cycles of $m-1$ functions and then each element of the step length vector (v_m) is adjusted in such a way that approximately half of all functions evaluated are accepted.

5.3.3.5 Number of Iterations before Temperature Reduction (n_T)

Frequently used criteria for n_T are a constant number of iterations, or iterating until a constant number of transitions is accepted. Experiments show that better results are achieved by considering the physical background of simulated annealing and the concept of thermal quasi-equilibrium. This means keeping the temperature constant until the AMPCS has reached a constant value. After $n_T \times n_s \times (m-1)$ function evaluations, temperature (T) is changed by the factor r_T. The value n_T suggested by Corana et al. (1987) is $\max(100, 5n)$.

5.3.3.6 Termination Criterion (n_ε)

Since the emphasis of our research is on the quality of the estimated penalties, we have developed a workable stopping criterion. It is clear that when the optimal values of the AMPCS for successive stages are constant themselves, the iterative process can be stopped. Let f_T be the most recent function. For every T_{j+1} if

$$\left|f_{T_j} - f_{T_{j+1}}\right| \leq \xi, \tag{5.3}$$

stop the search where ξ is very small.

5.4 Working Version of Different Algorithms

In the previous sections, we discussed theoretical aspects of different algorithms and the steps involved in using them. The algorithms we discussed are different in nature although some are developed on the basis of almost the same principles with a slight modification and/or alternation. For example, the CGSA, BGSA and PDFA are based on grid search but the SAA is completely different. In the following subsection we discuss and evaluate the working performance of these algorithms. We use the same experimental design as reported in Chapter 4 for this purpose.

5.4.1 CGSA

We performed a pilot experiment to evaluate the performance of the CGSA in the sense of investigating how long it takes to find optimal penalties. The number of models and number of grid points for each penalty set are two main ingredients of the CGSA. If the number of these ingredients increases, the computational cost of the CGSA increases.

Table 5.1 shows the CPU time required per penalty for different models, total number of penalty sets searched, total time required for different models with different grid points for each group. Note that the experimental design is the same as the experimental design used in Chapter 4 in the case of random walk regressors ($\phi = 1.0$). All computer programs were written in GAUSS (see Aptech, 1997, version 3.2.17) and computations were carried out on a Pentium III with a 933 Mhz CPU. It is clear from the table that if the number of grid points for each group and the number of models increases, then the computational time increases in such a way that it will be impossible to complete the task. Therefore, in practice, it is almost impossible to apply the CGSA even using a very high-powered computer.

In addition, the disadvantage of the CGSA is that it works well for a small number of models and a small number of grid points but is more difficult to apply as the number of models and grid points grows. In our case, we have a large number of penalties to be estimated. We need a method that will work well for a large number of models and grid points.

5.4.2 BGSA

We estimated the total time required for calculating optimal penalties for different models using five grid points for each penalty. Table 5.2 shows the estimated total time (in hours) required for different models by the BGSA with 5 grid points for each group comprised of $5^5 = 3125$ penalty sets. We found that for 12 models, the computation time required by the BGSA is only 0.5843 hours whereas the CGSA takes more than a year to complete the task. We also checked the loss of AMPCS when the BGSA is used over the CGSA. Table 5.3 gives the comparison between the computational time for the CGSA and the BGSA, loss of AMPCS and time saved for using the BGSA over the CGSA. From the table, it is clear that the

BGSA saves a lot of computational time for a small sacrifice of AMPCS. The disadvantage of this algorithm is that it considers all penalty points as five groups instead of considering each point separately.

5.4.3 PDFA

We first estimated coefficients of the polynomial using (5.1) and then used the CGSA on these coefficients. After performing a fine search for the coefficients of the polynomial, the estimated formulae for penalties for different values of ϕ and n are given below:

$$p_{i+1,0.15} = -0.0042(i)^4 + 0.117(i)^3 - 1.1403(i)^2 + 4.4185(i) - 3.21$$

$$p_{i+1,0.25} = -0.0001(i)^4 + 0.0052(i)^3 - 0.1089(i)^2 + 0.9666(i) - 1.0324$$

$$p_{i+1,0.50} = -0.000003(i)^4 + 0.0004(i)^3 - 0.0148(i)^2 + 0.2308(i) + 0.1198$$

$$p_{i+1,0.75} = -0.0000000006(i)^4 + 0.000008(i)^3 - 0.0008(i)^2 + 0.0263(i) + 0.9395$$

$$p_{i+1,0.100} = -0.0000002(i)^4 + 0.00003(i)^3 - 0.0015(i)^2 + 0.0113(i) + 3.1645$$

$$p_{i+1,0.7,15} = -0.0007(i)^4 + 0.0297(i)^3 - 0.4363(i)^2 + 2.3942(i) - 1.7925$$

$$p_{i+1,0.7,25} = -0.0001(i)^4 + 0.0062(i)^3 - 0.1267(i)^2 + 1.0389(i) - 1.1535$$

$$p_{i+1,0.7,50} = -0.000002(i)^4 + 0.0002(i)^3 - 0.0072(i)^2 + 0.1344(i) + 0.4278$$

$$p_{i+1,0.7,75} = -0.0000004(i)^4 + 0.00006(i)^3 - 0.0028(i)^2 + 0.0508(i) + 2.8948$$

$$p_{i+1,0.7,100} = -0.0000004(i)^4 + 0.00007(i)^3 - 0.0038(i)^2 + 0.0524(i) + 2.0675$$

$$p_{i+1,1.0,15} = -0.0028(i)^4 + 0.0773(i)^3 - 0.7771(i)^2 + 3.3041(i) - 2.4422$$

$$p_{i+1,1.0,25} = -0.000008(i)^4 + 0.0004(i)^3 - 0.0102(i)^2 + 0.136(i) + 0.0102$$

$$p_{i+1,1.0,50} = -0.000002(i)^4 + 0.0002(i)^3 - 0.0101(i)^2 + 0.2289(i) + 0.3095$$

$$p_{i+1,1.0,75} = -0.0000004(i)^4 + 0.00007(i)^3 - 0.0038(i)^2 + 0.0846(i) + 0.6394$$

$$p_{i+1,1.0,100} = -0.0000003(i)^4 + 0.00005(i)^3 - 0.0026(i)^2 + 0.0433(i) + 1.0529$$

$$p_{i+1,1.02,15} = -0.0026(i)^4 + 0.0743(i)^3 - 0.7473(i)^2 + 3.0459(i) - 2.228$$

$$p_{i+1,1.02,25} = -0.00009(i)^4 + 0.0052(i)^3 - 0.1080(i)^2 + 0.9359(i) - 0.9503$$

$$p_{i+1,1.02,50} = -0.000003(i)^4 + 0.0003(i)^3 - 0.0131(i)^2 + 0.2598(i) + 0.3158$$

$$p_{i+1,1.02,75} = -0.0000008(i)^4 + 0.0001(i)^3 - 0.0063(i)^2 + 0.1956(i) + 1.2134$$

$$p_{i+1,1.02,100} = -0.0000002(i)^4 + 0.000025(i)^3 - 0.0025(i)^2 - 0.0348(i) + 1.9843.$$

Here the first suffix indicates $i = 1,\cdots,m-1$, the second suffix indicates the values of ϕ and the third suffix indicates the sample size n. These polynomials can be used to find penalties for different changepoint models for different sample sizes and types of autoregressive regressors.

5.4.4 SAA

We performed a pilot experiment to find the effect of the choice of starting points of penalties on estimating optimal penalties. We used different IC penalties such as AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC as initial penalty values for the SAA to optimize AMPCS for different sample sizes. We found that the choice of different starting penalties does not affect the maximum AMPCS. In our case, we used AIC penalties as starting values of the penalties. Considering estimated AMPCS and total computational time, we used values of different factors as follows:

Initial temperature $T_s = 5$.

Temperature reduction factor $r_T = 0.65$.

Lower boundary for penalties $l_i = 0$ and the upper boundary $u_i = 3$.

Number of cycles n_s =15.

Number of iterations before temperature reduction $n_T = 10$.

Termination criterion $n_\varepsilon = 4$.

5.5 Computer Simulation

The performance of the various IC methods and the capabilities of the CGSA, BGSA, PODFA and SAA, were examined via extensive computer simulations to evaluate the performance of our methods discussed in the previous section. Our aim was to estimate the optimal penalties for each of the models as the true DGP. For our proposed procedure, we empirically estimated penalties for different models and compared the results of our methods with those of existing methods. In order to do this, we used the same simulation design as in Chapter 4.

5.6 Results of Simulation

We compared the gain of our procedures over existing IC procedures considered in Chapter 4. In other words, we evaluated the performance according to the criterion 'what percent of times one approach outperforms other approaches' by considering the magnitude of gain or losses of the use of different algorithms. The comparisons are based on AMPCS for different ϕ or n values in the presence of structural change or no structural change.

5.6.1 Comparison between IC and CGSA

It was mentioned earlier that the CGSA is very time consuming and difficult to apply for a large sample, that is, for a big number of models. Because of this, we applied this approach only for sample size 15 when $\phi = 0$. The estimated AMPCS and the percentage gain of the CGSA over existing IC procedures namely, AIC, BIC,

HQC, RSC, MCPC, HSPC, and GCVC in the presence of structural change and no structural change are given in Tables 5.4 and 5.5.

The form of the IC output is similar to that of Chapter 4 with the exception that the CGSA outperforms all existing IC procedures. In the presence of structural change, the percentage gain of the CGSA over existing IC procedures is highest for BIC (23.55 percent) and lowest for RSC (11.03 percent). When there is no structural change, the percentage gain of the CGSA over all IC is highest for RSC (124.13 percent) and lowest for BIC (11.15 percent). Overall, the CGSA is preferable to the existing IC procedures considered in this study.

5.6.2 Comparison between IC and BGSA

The estimated AMPCS and the percentage gain of the BGSA compared to existing procedures namely, AIC, BIC, HQC, RSC, MCPC, HSPC, and GCVC in the presence of structural change and no structural change are given in Tables 5.6 and 5.7. We would like to observe how the results change with changes in n and ϕ. The results show that when there is no structural change, the percentage gain of the BGSA is highest for RSC (123.38 percent) for $\phi = 0$ and $n = 25$, and the percentage gain is lowest for BIC (4.22 percent) for $\phi = 0$ and $n = 15$. The results reveal that as the sample size n or ϕ increases, the gain over RSC decreases, on the other hand with respect to BIC, it increases. In the presence of structural change, the percentage gain of the BGSA is highest for BIC (59.37 percent) for $\phi = 0$ and $n = 75$ and lowest for RSC (2.37 percent) for $\phi = 1.02$ and $n = 15$. Overall, the BGSA is clearly preferable to existing IC procedures in terms of always having a better AMPCS.

5.6.3 Comparison between IC and PDFA

The estimated AMPCS and the percentage gain of the PDFA compared to existing IC procedures in the presence of structural change and no structural change are given in Tables 5.8 and 5.9. The results show that when there is no structural change, the percentage gain of the PDFA is highest for RSC (113.31 percent) when $\phi = 0$ and $n = 25$. The percentage gain is lowest for BIC (1.98 percent) for $\phi = 0$ and $n = 50$. The results reveal that as the sample size n or ϕ increases, the gain over RSC decreases while, on the other hand, with respect to BIC it does not follow any particular pattern. In the presence of structural change, the percentage gain of the PDFA is highest for BIC (28.02 percent) for $\phi = 1.02$ and $n = 50$ and lowest for RSC (0.83 percent) for $\phi = 1.02$ and $n = 100$. Overall, the PDFA is clearly preferable to existing IC procedures in terms of always having a better AMPCS.

5.6.4 Comparison between IC and SAA

The estimated AMPCS and the percentage gain of the SAA compared to existing IC procedures in the presence of structural change and no structural change are given in Tables 5.10 and 5.11. The results show that when there is no structural change, the percentage gain of the SAA is highest for RSC (114.80 percent) for $\phi = 0$ and $n = 25$. The percentage gain is lowest for BIC (2.58 percent) for $\phi = 0$ and $n = 100$. The results reveal that as the sample size n or ϕ increases, the gain over RSC decreases while, on the other hand, with respect to BIC it does not follow any particular pattern. In the presence of structural change, the percentage gain of the SAA is highest for BIC (28.53 percent) for $\phi = 1.02$ and $n = 50$ and lowest for RSC (1.16 percent) for $\phi = 1.02$ and $n = 100$. Overall, the SAA is undoubtedly

preferable to all existing IC procedures in terms of always having a better AMPCS and preferable to the CGSA because of its small computational time.

5.6.5 Comparison between BGSA and PDFA

The percentage gains of the BGSA over the PDFA in the presence of structural change and no structural change are given in Table 5.12. The results reveal that as the sample size n or ϕ increases, the gain for the BGSA increases. When there is no structural change, the percentage gain of the BGSA over the PDFA is largest (about 20.06 percent) for sample size $n=100$ and $\phi=0.7$, and smallest (about 2.04 percent) for sample size $n=15$ and $\phi=0$. In the presence of structural change, the percentage gain is a maximum (about 10.14 percent) for sample size $n=100$ and $\phi=1.0$, and minimum (about 1.26 percent) when the sample size is $n=15$ and $\phi=1.02$. Overall, the PDFA has a smaller computational time (approximately 14 minutes) compared to the BGSA but the BGSA has a slightly larger AMPCS than the PDFA.

5.6.6 Comparison between BGSA and SAA

The percentage gains of the BGSA over the SAA in the presence of structural change and no structural change are given in Table 5.12. The results show that as the sample size n or ϕ increases, the gain for the BGSA increases. When there is no structural change, the percentage gain is a maximum (about 19.67 percent) for sample size $n=100$ and $\phi=0.7$, and a minimum (about 1.23 percent) for sample size $n=15$ and $\phi=0$. In the presence of structural change, the gain is largest (about 9.87 percent) for sample size $n=100$ and $\phi=1.0$, and smallest (about 0.84 percent) when the sample size is $n=15$ and $\phi=1.02$. Overall, the SAA has a

smaller computational time compared to the BGSA but the BGSA has slightly larger AMPCS than the SAA.

5.6.7 Comparison between PDFA and SAA

The percentage gains of the SAA over the PDFA in the presence of structural change and no structural change are given in Table 5.12. The results reveal that as the sample size n or ϕ increases, the gain of the SAA over the PDFA decreases. When there is no structural change, the gain is a maximum (about 0.98 percent) for sample size $n=15$ and $\phi=1.0$, and a minimum (about 0.46 percent) when $n=100$ and $\phi=0.7$. In the presence of structural change, the gain is largest (about 0.46 percent) for sample size $n=50$ and $\phi=0.7$, and smallest (about 0.27 percent) when the sample size is $n=50$ and $\phi=1.0$. Overall, the PDFA has a smaller computational time compared to the SAA but the PDFA has a slightly larger AMPCS than the SAA.

5.7. Concluding Remarks

In this chapter, we examined a new method for finding penalties for the problem of detecting possible structural change through model selection procedures. Our method includes a family of procedures, based on grid search algorithms such as the CGSA, BGSA and PDFA, and the SAA. These procedures do not require conditions such as regularity or existence of derivatives. We are interested in optimizing AMPCS. One of the basic aims of this chapter was to develop an algorithm for finding penalties that optimize AMPCS, that is completely stable and does not use too much computational time. It means that we can always find the penalties that give maximum AMPCS in a reasonable time.

Grid search is one of the appealing ways to maximize a function, which is difficult to maximize using other algorithms. The disadvantage of using this procedure is that it is very time consuming when there is a reasonably large number of grid points for the penalty vector. The computational time of grid search algorithms will increase dramatically with an increase in the number of grid points and so can be exceptionally excessive. To overcome the computational limits imposed by grid search algorithms, we estimate the optimum penalties that give maximum AMPCS using the SAA whose performance is similar to that of grid search algorithms while its computational time is much lower.

The simulation results show that our procedure outperforms existing IC procedures, including AIC, BIC, HQC, RSC, MCPC, HSPC and GCVC in small samples as well as in moderately sized samples. Out of these procedures, the CGSA involves heavy computation giving the highest percentage gain over all IC procedures while for the rest of our procedures the highest gain is about 59.37%.

We performed some simulation experiments using the CGSA for $n = 15$ and 25 when ϕ takes the value 1.0. The computational time for optimizing AMPCS by this algorithm is enormous. On the other hand, from the plot of penalties versus different changepoint models, we found that for a particular changepoint model and nearby changepoint model, the estimated penalties are almost the same and these penalties follow an approximately symmetric pattern.

We also calculated maximum AMPCS for experiments with penalties corresponding to each model and the same penalties for a model and its neighbouring models. The former is called a complete list of penalties and the latter is a partial list of penalties. We found that a partial list of the penalties is enough to get reasonably optimal penalties without a major sacrifice of AMPCS. These important considerations motivated us to investigate two alternative approaches to the CGSA.

The two alternatives namely, the BGSA and the PDFA, are straight forward and save computational time. The former gives higher AMPCS than the latter. However, the computational cost is lower for the latter. The major advantage is that it does not require detailed analytic knowledge of the function to be optimized.

It was revealed from a comparison of the performance of our procedures that the CGSA is the best, the BGSA second, the SAA third and the PDFA fourth best as measured by maximum AMPCS. In the context of computational time for maximizing AMPCS, the ranking performance of our procedures are the SAA first, the PDFA second, the BGSA third and the CGSA last.

We have found that all of our four suggested procedures dominate the existing IC procedures considered in terms of maximizing AMPCS. We therefore suggest using any one of the procedures when we need to choose optimal penalties depending on the needs and wishes of the user. In particular, we suggest for very small samples, the use of the CGSA, for moderately sized samples, the use of the BGSA or the PDFA, and for large samples, the use of the SAA.

Figure 5.1 Plot of Penalties for Different Models when $n = 15$ and $\phi = 0$

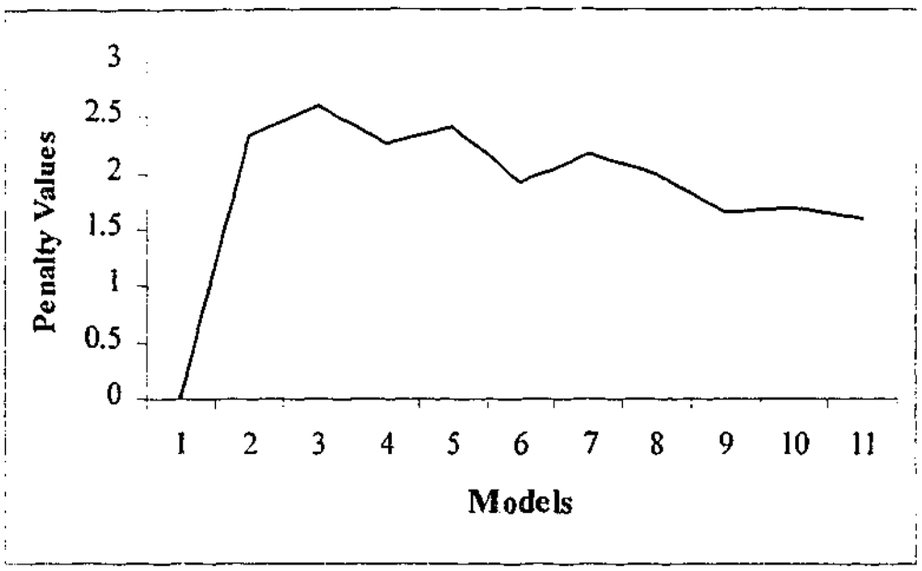

Figure 5.2 Plot of Penalties for Different Models when $n = 15$ and $\phi = 0.7$

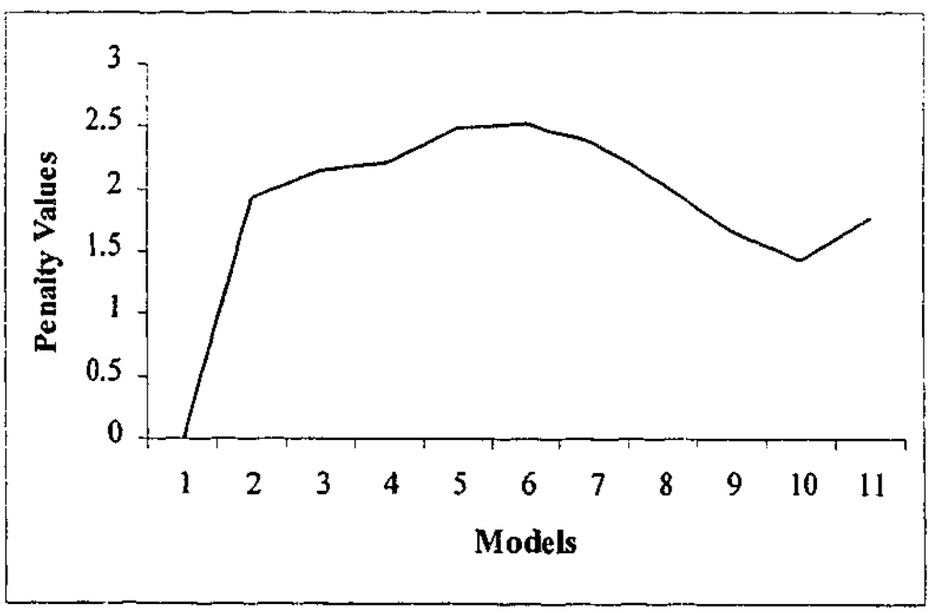

Figure 5.3 Plot of Penalties for Different Models when $n = 15$ and $\phi = 1.0$

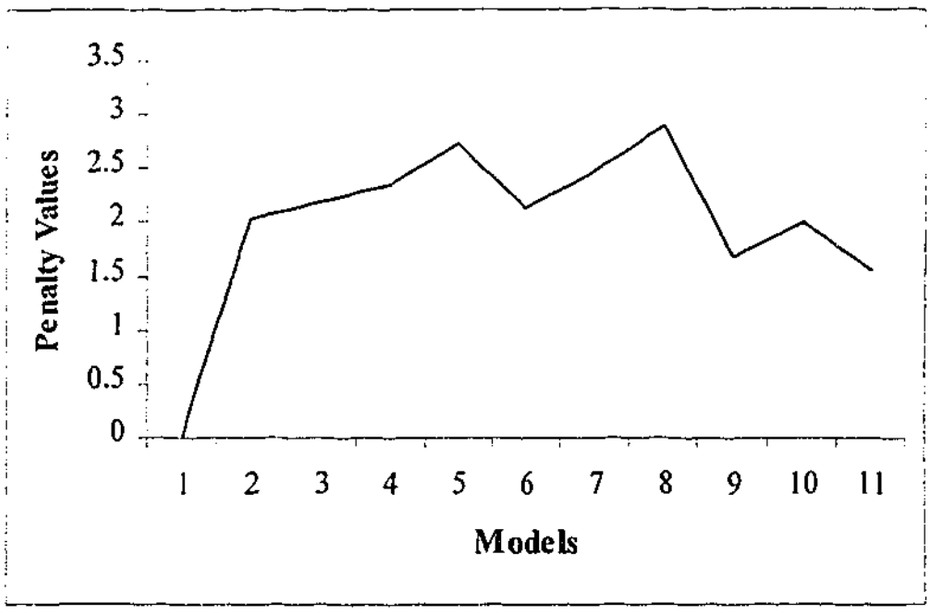

Figure 5.4 Plot of Penalties for Different Models when $n = 15$ and $\phi = 1.02$

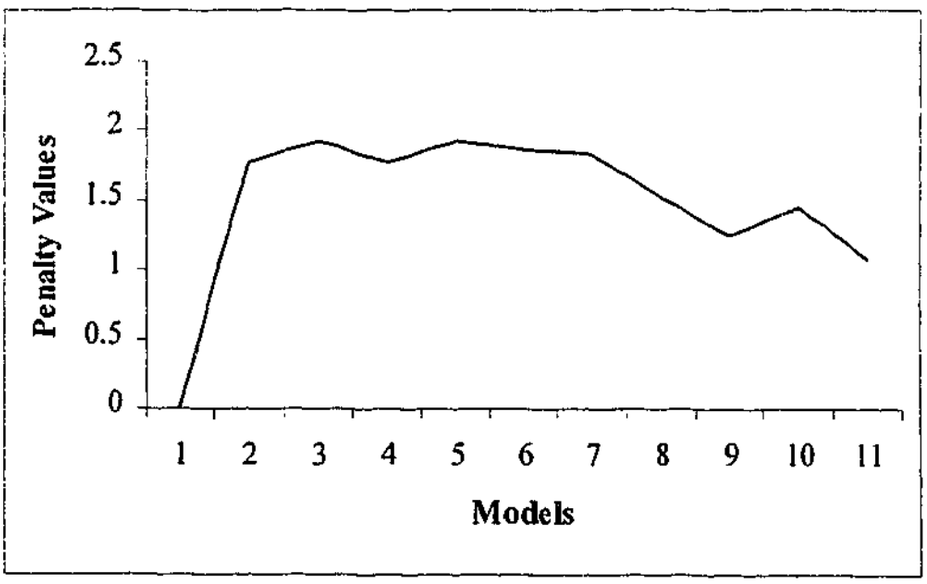

Figure 5.5 Plot of Penalties for Different Models when $n = 25$ and $\phi = 0$

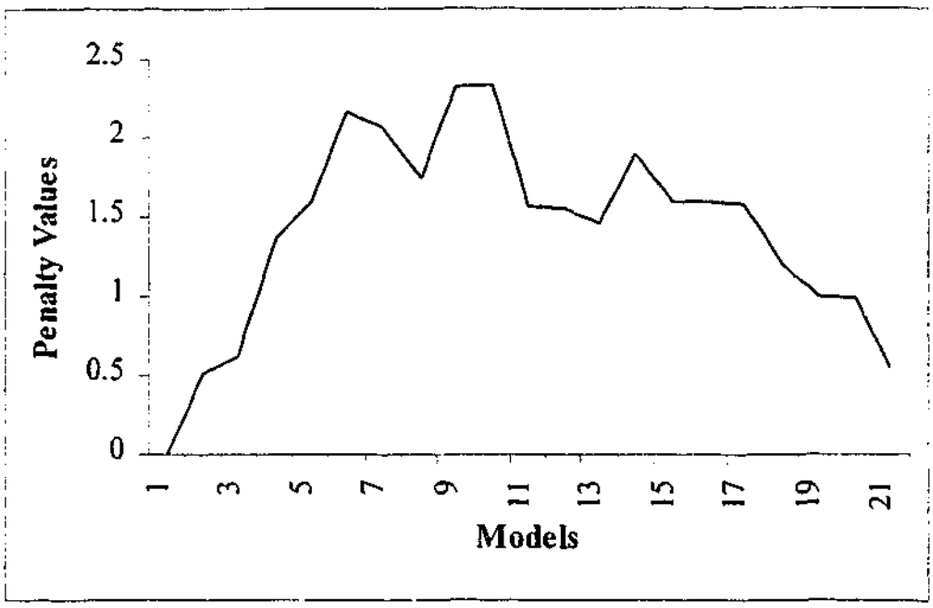

Figure 5.6 Plot of Penalties for Different Models when $n = 25$ and $\phi = 0.7$

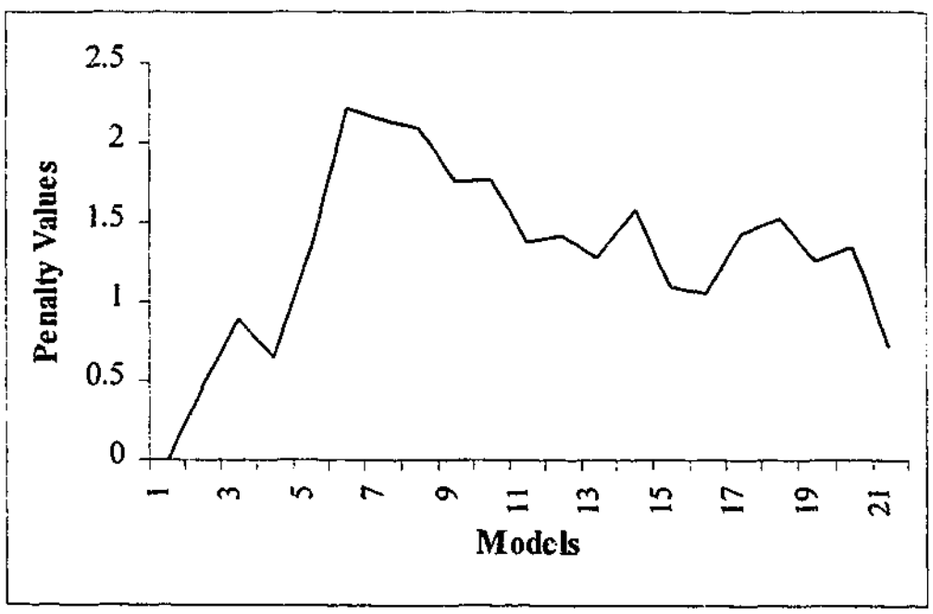

Figure 5.7 Plot of Penalties for Different Models when $n = 25$ and $\phi = 1.0$

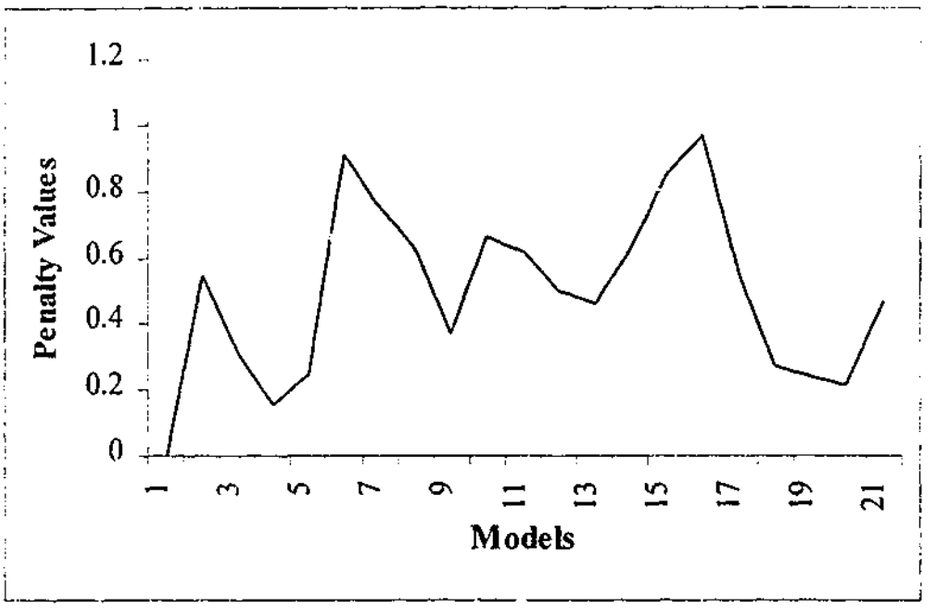

Figure 5.8 Plot of Penalties for Different Models when $n = 25$ and $\phi = 1.02$

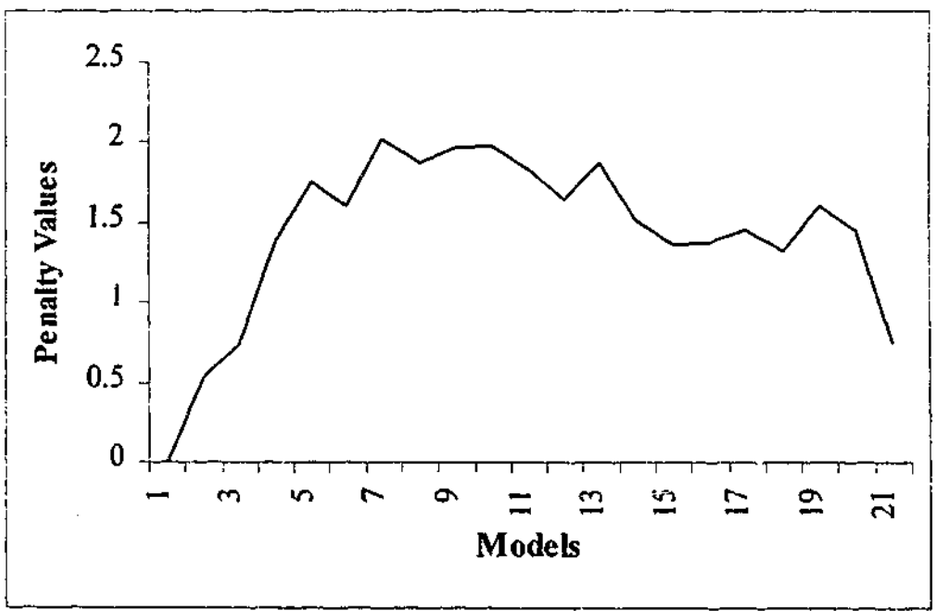

Table 5.1 Estimated Timing and Number of Penalty Sets for CGSA and BGSA with 5^5 Penalty Sets.

Number of Models	11	21	46	71	98
		CGSA			
Time per penalty (Seconds)	0.6714	1.4502	3.3061	3.7581	5.0803
		$n_g = 5$			
Penalty sets	4.88×10^{7}	4.77×10^{14}	1.42×10^{32}	4.24×10^{49}	1.26×10^{67}
Time (years)	1.0385	2.19×10^{6}	1.49×10^{25}	5.04×10^{42}	2.05×10^{60}
		$n_g = 3$			
Penalty sets	177147	1.05×10^{9}	8.86×10^{21}	7.50×10^{33}	6.36×10^{45}
Time (years)	0.0038	480.959	9.20×10^{14}	8.95×10^{26}	1.02×10^{39}
		$n_g = 2$			
Penalty sets	2048	2.10×10^{6}	7.04×10^{13}	2.36×10^{21}	7.92×10^{28}
Time (years)	4.36×10^{-5}	0.096	7.38×10^{6}	2.81×10^{14}	1.28×10^{22}
		BGSA			
Time (hours)	0.5843	1.1474	2.8702	4.3383	6.0853

Table 5.2 Comparison of CGSA and BGSA on Computational Time and Loss of AMPCS and Time Saved for Sample Size 15.

Method	AMPCS	Time taken	Loss of AMPCS when BGSA used.	Total time saved when BGSA used
CGSA	8.3498	1.0385(yrs)		
BGSA	8.0653	0.5843(hours)	3.53%	9096.68(hours)

Table 5.3 Estimated AMPCS and Percentage Gain of CGSA over IC when there is no Structural Change for $n = 15$ and $\phi = 0$

	AIC	**BIC**	**HQC**	**RSC**	**MCPC**	**HSPC**	**GCVC**	**CGSA**
AMPCS	0.1325	0.2440	0.1805	0.1210	0.1505	0.2195	0.2069	0.2712
% gain	104.68	11.15	50.25	124.13	80.20	23.55	31.11	

Table 5.4 Estimated AMPCS and Percentage Gain of CGSA over IC in Presence of Structural Change for $n = 25$ and $\phi = 0$

	AIC	**BIC**	**HQC**	**RSC**	**MCPC**	**HSPC**	**GCVC**	**CGSA**
AMPCS	0.4904	0.4433	0.4765	0.4933	0.4881	0.4695	0.4740	0.5476
% gain	11.67	23.55	14.93	11.03	12.20	16.64	15.53	

Table 5.5 Percentage Gain of BGSA over IC in Presence of no Structural Change for Different Sample Sizes and Types of Autoregressive Regressors

	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC	BGSA
				$\phi = 0$				
$n = 15$	91.92	4.22	40.89	110.17	68.97	15.85	22.94	0.25
$n = 25$	97.70	6.86	48.15	123.39	77.00	28.66	35.74	0.28
$n = 50$	91.99	7.64	44.35	96.32	67.66	32.07	38.44	0.30
$n = 75$	96.23	14.85	51.74	109.08	63.92	36.16	50.98	0.34
$n = 100$	77.89	15.11	52.72	78.52	62.03	40.26	51.81	0.35
				$\phi = 0.7$				
$n = 15$	44.62	6.60	23.40	68.77	30.07	9.42	15.85	0.24
$n = 25$	43.64	8.87	26.31	61.72	32.40	12.16	20.37	0.26
$n = 50$	59.67	20.60	35.14	81.68	41.30	23.06	28.93	0.29
$n = 75$	61.60	25.06	40.54	88.70	47.14	32.59	37.48	0.32
$n = 100$	64.23	28.09	46.95	82.75	53.16	37.09	42.80	0.34
				$\phi = 1.0$				
$n = 15$	54.22	13.01	26.48	67.13	33.74	17.95	25.35	0.26
$n = 25$	51.93	12.86	30.92	71.83	32.54	20.17	24.09	0.28
$n = 50$	49.38	15.93	29.82	69.85	37.11	19.72	26.93	0.30
$n = 75$	54.36	18.29	34.69	74.26	40.67	22.12	32.11	0.32
$n = 100$	51.02	20.10	31.68	65.85	38.27	22.75	27.06	0.33
				$\phi = 1.02$				
$n = 15$	40.58	6.42	21.55	61.12	27.39	11.03	16.75	0.26
$n = 25$	55.61	14.31	31.69	65.01	35.95	20.53	27.92	0.29
$n = 50$	52.20	15.56	31.05	66.86	38.32	21.35	26.20	0.31
$n = 75$	45.20	15.47	26.37	61.99	34.25	16.82	24.41	0.31
$n = 100$	52.38	19.68	32.41	69.19	37.58	25.30	27.94	0.34

Table 5.6 Percentage Gain of BGSA over IC in Presence of Structural Change for Different Sample Sizes and Types of Autoregressive Regressors

	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC	BGSA
				$\phi = 0$				
$n = 15$	8.59	20.14	11.75	7.96	9.10	13.42	12.34	0.53
$n = 25$	11.99	26.09	16.87	9.33	14.28	22.48	19.46	0.58
$n = 50$	11.73	23.00	14.72	9.57	13.86	17.04	16.10	0.60
$n = 75$	46.47	59.37	50.78	45.55	49.87	56.37	54.56	0.64
$n = 100$	14.13	29.65	22.66	10.74	16.21	27.91	24.46	0.71
				$\phi = 0.7$				
$n = 15$	4.74	13.52	8.77	3.60	6.91	12.81	11.07	0.51
$n = 25$	4.77	16.61	9.30	4.38	9.01	13.96	12.07	0.52
$n = 50$	9.51	30.80	16.36	7.64	11.59	20.77	18.03	0.61
$n = 75$	10.54	22.61	14.57	7.83	11.61	15.43	15.08	0.62
$n = 100$	15.37	24.18	19.59	8.05	17.68	21.90	19.76	0.68
				$\phi = 1.0$				
$n = 15$	5.43	20.52	9.28	2.56	6.99	15.00	11.92	0.50
$n = 25$	12.34	26.02	13.48	10.56	15.67	23.30	21.37	0.56
$n = 50$	20.97	38.26	26.81	11.90	26.63	29.64	32.29	0.63
$n = 75$	13.33	40.28	22.73	12.34	18.16	37.75	34.72	0.65
$n = 100$	13.90	40.43	20.29	12.37	18.20	25.47	24.11	0.68
				$\phi = 1.02$				
$n = 15$	5.10	18.27	9.91	2.37	7.48	15.45	12.73	0.48
$n = 25$	5.65	24.71	11.28	4.02	7.83	18.90	16.24	0.52
$n = 50$	9.60	31.69	13.84	3.88	11.68	26.18	19.75	0.59
$n = 75$	11.14	28.29	14.63	7.96	14.05	20.29	19.65	0.73
$n = 100$	9.20	29.03	15.35	7.48	10.74	21.05	17.43	0.78

Table 5.7 Percentage Gain of PDFA over IC in Presence of no Structural Change for Different Sample Sizes and Types of Autoregressive Regressors

	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC	PDFA
				$\phi = 0$				
$n = 15$	88.00	2.09	38.01	105.87	65.52	13.49	20.43	0.25
$n = 25$	88.78	2.03	41.46	113.31	69.02	22.86	29.62	0.27
$n = 50$	81.89	1.98	36.76	86.00	58.84	25.13	31.16	0.29
$n = 75$	74.32	2.02	34.79	85.73	45.62	20.96	34.12	0.30
$n = 100$	57.66	2.02	35.35	58.22	43.60	24.31	34.54	0.31
				$\phi = 0.7$				
$n = 15$	38.79	2.31	18.42	61.96	24.82	5.00	11.18	0.23
$n = 25$	35.01	2.34	18.72	52.00	24.45	5.42	13.14	0.24
$n = 50$	35.45	2.30	14.63	54.12	19.86	4.39	9.37	0.25
$n = 75$	34.99	4.46	17.39	57.62	22.90	10.75	14.84	0.26
$n = 100$	31.31	2.42	17.50	46.13	22.46	9.62	14.19	0.28
				$\phi = 1.0$				
$n = 15$	39.40	2.15	14.33	51.08	20.89	6.62	13.31	0.24
$n = 25$	37.52	2.15	18.49	55.53	19.96	8.77	12.31	0.25
$n = 50$	31.74	2.24	14.49	49.79	20.91	5.58	11.94	0.26
$n = 75$	33.38	2.22	16.39	50.58	21.55	5.53	14.16	0.27
$n = 100$	28.64	2.30	12.16	41.27	17.78	4.56	8.23	0.28
				$\phi = 1.02$				
$n = 15$	35.20	2.35	16.90	54.95	22.51	6.78	12.28	0.25
$n = 25$	39.15	2.22	17.76	47.56	21.57	7.78	14.39	0.26
$n = 50$	34.67	2.25	15.95	47.63	22.39	7.37	11.66	0.27
$n = 75$	28.49	2.18	11.83	43.35	18.81	3.38	10.10	0.28
$n = 100$	30.01	2.11	12.97	44.34	17.37	6.90	9.15	0.29

Table 5.8 Percentage Gain of PDFA over IC in Presence of Structural Change for Different Sample Sizes and Types of Autoregressive Regressors

	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC	PDFA
				$\phi = 0$				
$n = 15$	1.65	12.46	4.62	1.06	2.13	6.18	5.16	0.50
$n = 25$	3.49	16.53	8.01	1.03	5.61	13.19	10.40	0.53
$n = 50$	2.99	13.38	5.75	1.00	4.96	7.89	7.02	0.55
$n = 75$	1.61	10.56	4.60	0.97	3.97	8.48	7.22	0.58
$n = 100$	4.02	18.16	11.79	0.92	5.91	16.57	13.42	0.64
				$\phi = 0.7$				
$n = 15$	2.14	10.70	6.07	1.03	4.26	10.01	8.31	0.50
$n = 25$	1.41	12.87	5.79	1.03	5.52	10.30	8.48	0.51
$n = 50$	2.67	22.62	9.09	0.92	4.62	13.22	10.66	0.57
$n = 75$	3.46	14.75	7.22	0.91	4.46	8.03	7.70	0.58
$n = 100$	7.79	16.03	11.74	0.95	9.95	13.89	11.89	0.64
				$\phi = 1.0$				
$n = 15$	3.93	18.81	7.74	1.11	5.47	13.37	10.34	0.49
$n = 25$	2.73	15.24	3.78	1.10	5.78	12.75	10.99	0.51
$n = 50$	9.19	24.79	14.45	1.00	14.29	17.01	19.40	0.57
$n = 75$	1.90	26.14	10.36	1.01	6.25	23.86	21.14	0.59
$n = 100$	2.37	26.21	8.10	0.99	6.23	12.76	11.54	0.61
				$\phi = 1.02$				
$n = 15$	3.76	16.77	8.51	1.07	6.12	13.98	11.29	0.47
$n = 25$	2.66	21.18	8.13	1.08	4.78	15.53	12.95	0.51
$n = 50$	6.54	28.02	10.67	0.98	8.56	22.66	16.41	0.57
$n = 75$	3.81	19.83	7.07	0.84	6.53	12.36	11.76	0.68
$n = 100$	2.44	21.05	8.22	0.83	3.89	13.57	10.17	0.74

Table 5.9 Percentage Gain of SAA over IC in Presence of no Structural Change for Different Sample Sizes and Types of Autoregressive Regressors

	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC	PDFA
				$\phi = 0$				
$n = 15$	89.58	2.95	39.17	107.60	66.91	14.44	21.44	0.25
$n = 25$	90.10	2.75	42.45	114.80	70.20	23.72	30.53	0.27
$n = 50$	82.97	2.58	37.57	87.10	59.78	25.87	31.94	0.29
$n = 75$	75.42	2.67	35.65	86.91	46.54	21.72	34.97	0.30
$n = 100$	58.52	2.58	36.09	59.08	44.39	24.99	35.27	0.31
				$\phi = 0.7$				
$n = 15$	39.93	3.15	19.39	63.30	25.85	5.87	12.09	0.23
$n = 25$	35.97	3.06	19.56	53.07	25.33	6.17	13.94	0.24
$n = 50$	36.42	3.04	15.46	55.23	20.72	5.14	10.15	0.25
$n = 75$	33.03	2.95	15.69	55.34	21.12	9.15	13.18	0.27
$n = 100$	31.89	2.87	18.01	46.76	23.00	10.09	14.68	0.28
				$\phi = 1.0$				
$n = 15$	40.81	3.18	15.48	52.60	22.11	7.70	14.46	0.24
$n = 25$	38.72	3.04	19.53	56.89	21.01	9.72	13.29	0.25
$n = 50$	32.69	2.98	15.32	50.87	21.79	6.34	12.75	0.26
$n = 75$	34.26	2.89	17.16	51.57	22.35	6.22	14.91	0.28
$n = 100$	29.33	2.86	12.77	42.04	18.41	5.12	8.81	0.28
				$\phi = 1.02$				
$n = 15$	36.02	2.96	17.60	55.89	23.25	7.43	12.96	0.25
$n = 25$	40.15	2.96	18.60	48.62	22.44	8.55	15.21	0.27
$n = 50$	35.56	2.93	16.72	48.62	23.20	8.08	12.40	0.27
$n = 75$	29.42	2.92	12.64	44.38	19.67	4.13	10.89	0.28
$n = 100$	30.90	2.81	13.75	45.34	18.18	7.64	9.90	0.29

Table 5.10 Percentage Gain of SAA over IC in Presence of Structural Change for Different Sample Sizes and Types of Autoregressive Regressors

	AIC	BIC	HQC	RSC	MCPC	HSPC	GCVC	PDFA
				$\phi = 0$				
$n = 15$	2.06	12.92	5.04	1.47	2.54	6.60	5.59	0.50
$n = 25$	3.86	16.94	8.39	1.40	5.99	13.59	10.79	0.53
$n = 50$	3.35	13.77	6.12	1.35	5.32	8.26	7.39	0.55
$n = 75$	2.01	11.00	5.01	1.37	4.38	8.91	7.64	0.59
$n = 100$	4.37	18.56	12.17	1.27	6.28	16.97	13.81	0.65
				$\phi = 0.7$				
$n = 15$	2.57	11.17	6.52	1.46	4.70	10.48	8.77	0.50
$n = 25$	1.83	13.34	6.23	1.45	5.95	10.76	8.93	0.51
$n = 50$	3.13	23.18	9.59	1.37	5.09	13.74	11.16	0.57
$n = 75$	3.92	15.26	7.70	1.36	4.93	8.51	8.18	0.59
$n = 100$	8.17	16.43	12.12	1.30	10.33	14.29	12.28	0.64
				$\phi = 1.0$				
$n = 15$	4.29	19.22	8.11	1.45	5.84	13.76	10.72	0.50
$n = 25$	3.06	15.60	4.10	1.42	6.11	13.10	11.34	0.51
$n = 50$	9.47	25.12	14.75	1.26	14.59	17.31	19.72	0.57
$n = 75$	2.20	26.50	10.67	1.30	6.55	24.22	21.49	0.59
$n = 100$	2.67	26.58	8.42	1.29	6.54	13.09	11.86	0.62
				$\phi = 1.02$				
$n = 15$	4.22	17.29	9.00	1.52	6.59	14.48	11.79	0.48
$n = 25$	3.09	21.68	8.58	1.50	5.21	16.01	13.42	0.51
$n = 50$	6.97	28.53	11.11	1.39	9.00	23.16	16.88	0.57
$n = 75$	4.21	20.29	7.48	1.23	6.93	12.78	12.19	0.69
$n = 100$	2.78	21.45	8.57	1.16	4.23	13.94	10.53	0.74

Table 5.11 Percentage Gain of BGSA over SAA (GOGOS), BGSA over PDFA (GOGOP) and SSA over PDFA (GOSOP) for Different Sample Sizes and Types of Autoregressive Regressors

	Without Structural Change			With Structural Change		
	GOGOS	GOGOP	GOSOP	GOGOS	GOGOP	GOSOP
			$\phi = 0$			
$n = 15$	1.23	2.04	0.81	6.02	6.39	0.39
$n = 25$	3.85	4.50	0.68	7.24	7.58	0.37
$n = 50$	4.72	5.25	0.56	7.50	7.82	0.35
$n = 75$	10.60	11.18	0.64	8.60	8.95	0.38
$n = 100$	10.89	11.37	0.55	8.56	8.87	0.35
			$\phi = 0.7$			
$n = 15$	3.26	4.03	0.79	2.05	2.47	0.42
$n = 25$	5.34	5.97	0.67	2.81	3.22	0.43
$n = 50$	14.57	15.19	0.72	5.82	6.25	0.46
$n = 75$	17.67	18.14	0.57	6.00	6.41	0.44
$n = 100$	19.69	20.06	0.46	6.25	6.58	0.35
			$\phi = 1.0$			
$n = 15$	8.71	9.61	0.99	1.09	1.41	0.33
$n = 25$	8.69	9.49	0.87	8.26	8.54	0.31
$n = 50$	11.18	11.82	0.72	9.49	9.74	0.27
$n = 75$	13.02	13.60	0.66	9.82	10.07	0.28
$n = 100$	14.36	14.83	0.53	9.87	10.14	0.30
			$\phi = 1.02$			
$n = 15$	3.26	3.83	0.59	0.84	1.26	0.42
$n = 25$	6.78	7.44	0.70	2.43	2.83	0.42
$n = 50$	10.94	11.54	0.68	2.40	2.78	0.39
$n = 75$	10.86	11.49	0.71	6.25	6.59	0.37
$n = 100$	14.10	14.69	0.69	5.88	6.19	0.33

CHAPTER 6

Prediction with the Linear Regression Model in the Presence of Random Structural Change

6.1 Introduction

People have always wanted to predict the future to reduce their fear and anxiety about the unknown and an uncertain tomorrow. This desire has been since the dawn of civilization. Today, the need to predict the future is fulfilled in a wide range of ways, from horoscopes to econometric services. Predictions are simply extrapolations (or interpolations) of established past patterns and/or existing relationships. Prediction techniques play an important role in the fields of economics, business administration, engineering and meteorological sciences, among others. The main purpose is to predict at time t, the future value of a variable. Like others, economists are interested to know the possible future values of economic time series variables.

In formulating policy decisions, it is essential to be able to forecast the values of important economic variables. Such forecasts will enable the policy-maker to judge whether it is necessary to take any measures in order to influence the relevant economic variables. For example, suppose that the government wants to pursue its employment policy. It is necessary to know what the current level of employment is, as well as what the level of employment will be, say in five years' time, if the government takes no new measures. With econometric techniques, we are able to obtain such an estimate of the level of employment. Forecasting is becoming increasingly important both for the regulation of developed economies as well as for the planning of the economic development of underdeveloped countries.

Wallis (1989) pointed out three main motivations for forecasting. The first is that in order to make policy decisions when the current situation is uncertain, some kind of forward-looking prediction is essential, particularly when the decisions cannot be reversed. The second is to anticipate events for private gains and the third is to put hypotheses about the behaviour of the world to test. Zellner (1988) emphasized the importance of prediction in evaluating hypotheses and models in econometrics. In his opinion, econometricians are overly concerned with estimating parameters and fail to appreciate the important role of forecasting. We share this view.

One of the objectives of applied econometric research is to obtain good numerical estimates of the coefficients of economic relationships and to use them for the prediction of the values of economic variables. Before using an estimated model for forecasting the value of the dependent variable, there must be an assessment in some way of the predictive power of the model. It is conceivably possible that the

model is economically meaningful, and statistically and econometrically correct for the sample period for which the model has been estimated, yet it may very well not be suitable for forecasting due to a change in the parameters of the relationship in the real world.

In Chapter 4, we discussed the use of model selection to detect an unknown changepoint when there is a possible structural change in the data. We also suggested a procedure that outperformed existing procedures in identifying the model that best fits the data in the sense that the probability of the model being selected is unity. It is natural to ask the question "Does the best fitted model also produce the best forecasts?" In certain situations, the answer may be "yes" but we should bear in mind that fitting and forecasting are two different issues. In practice a particular model might produce a very good fit, but because of the mathematical properties of the model involved, may produce ridiculous forecasts (see Bryant (1960)). In addition, Makridakis (1986) and Mills and Prasad (1992) observe that a model having the best fit for a given series does not necessarily mean it is the best forecasting model.

Forecasting with a regression model assumes there is no change in any of the parameters over the forecast period. Unfortunately, it is quite possible that one or more of the parameters might change at some unknown point in the future. The aim of this chapter is to investigate how we might incorporate possible future structural change as a stochastic element of our model. With this view in mind, we consider different models for stochastic changes in parameters, to compare predictions from different strategies. We shall begin with the case in which there are two possibilities of a changing parameter. The first is a change in the slope parameter by a fixed

amount with a very low probability, keeping the intercept constant. The second is a change in the intercept parameter by a fixed amount with a very low probability, keeping the slope unchanged. As usual, these estimators depend on nuisance parameters whose values are unknown. To obtain an operational procedure we replace these unknown parameters by sample estimates.

We adopt linear regression models with single or multiple time changing coefficients but with low probabilities of a change at any point in time. The time changing coefficient approach considered here will provide a simple procedure for handling possible changepoints in the data. The traditional fixed coefficient linear regression models can be treated as a special case of the time changing coefficient linear regression model.

We look at the distributional pattern of the model, derive the distribution of the changing parameter model, develop the theoretical variance-covariance matrices for three special situations, and construct 'out of sample' forecast procedures. In order to investigate whether our forecast procedures are likely to be useful, we perform a Monte Carlo study of forecast performance. We use OLS, maximum likelihood (ML) and maximum marginal likelihood (MML) methods for the estimation of parameters.

Once the forecasts are made, they can be evaluated by computing the prediction error (PE) and these errors can tell us a lot about the quality of the forecasting model. We use three forecast evaluation statistics: (i) mean error (ME), (ii) mean absolute error (MAE) and (iii) root mean squared error (RMSE), prominent in the forecasting literature.

The rest of the chapter is organised as follows. Section 6.2 contains a theoretical discussion that includes the distribution of time series data for changing the intercept and slope parameters keeping one fixed while the other may change in a simple linear regression model. In addition, the variance-covariance matrix for our model in these two cases is derived for three special situations. In Section 6.3, we discuss the estimation of parameters and prediction errors in the case of changepoints and of no changepoints in the model. Section 6.4 presents three different schemes for estimating the parameters of the model to predict consecutive periods with new data becoming available. Section 6.5 contains a discussion of the Monte Carlo experiment and includes a description of the experimental design and the models used. Section 6.6 discusses the prediction accuracy and some statistical measures for it. Discussion of the results of the Monte Carlo experiments is presented in Section 6.7. We furnish the concluding remarks of our study in the final section.

6.2 Theoretical Discussion

Consider the simple linear regression model

$$y_t = \alpha_t + \beta_t x_t + \varepsilon_t , \qquad t = 1, \ldots, n, \tag{6.1}$$

where y_t is the dependent variable at time t, x_t is the value of the independent variable at time t, α_t and β_t are parameters of the model that may change over time. We will consider two possibilities of changes. The first is where the intercept α_t is a constant and β_t may change at each point t by a fixed amount δ but with a very low probability p. The second is where the slope β_t is a constant and α_t may

change at each point t by a fixed amount δ but with a very low probability p. In other words in the first case, $\alpha_t = \alpha_0$ for all t and

$$\Pr(\beta_{t+1} = \beta_t + \delta) = p,$$

$$\Pr(\beta_{t+1} = \beta_t) = 1 - p,$$

where p is a small and unknown probability and δ is a fixed but unknown value while in the second case, $\beta_t = \beta_0$ for all t and

$$\Pr(\alpha_{t+1} = \alpha_t + \delta) = p,$$

$$\Pr(\alpha_{t+1} = \alpha_t) = 1 - p.$$

We are interested in finding the distribution of y_t for a changing slope and for a changing intercept. The simple linear regression model (6.1) in these two cases can be written respectively as

$$y_t = \alpha_0 + (\beta_0 + \delta\,\tau_t)x_t + u_t, \qquad t = 1, 2, \cdots, n, \tag{6.2}$$

$$y_t = \alpha_0 + \delta\,\tau_t + \beta_0\,x_t + u_t, \qquad t = 1, 2, \cdots, n, \tag{6.3}$$

where $\tau_t \sim$ Binomial (t, p). τ_t can be approximated by a normal random variable for large t with mean $t\,p$ and variance $t\,p(1-p)$. The error term u_t is assumed to be independently normally distributed with constant variance, i.e., $u_t \sim IN(0, \sigma^2)$ and we assume u_t is independent of τ_t. We introduce a sequence of independent random variables z_i, $i = 1, 2, \cdots, n$, which are allowed to have only two different values, 0 and 1, with corresponding probability $(1-p)$ and p so that $\tau_t = \sum_{i=1}^{t} z_i$. Then models (6.2) and (6.3) become

$$y_t = \alpha_0 + (\beta_0 + \delta \sum_{i=1}^{t} z_i) x_t + u_t, \qquad t = 1, 2, \cdots, n, \tag{6.4}$$

$$y_t = \alpha_0 + \delta \sum_{i=1}^{t} z_i + \beta_0 x_t + u_t, \qquad t = 1, 2, \cdots, n, \tag{6.5}$$

respectively. Models (6.4) and (6.5) are characterized as changing slope and intercept models, respectively.

6.2.1 Distributional Pattern of y_t for Changing Slope Parameters

Consider the slope changing model (6.4) and assume δ is fixed. We are interested in finding the distribution of y_t. We note the following:

Period	Distribution of y_t	Probability	Values of z_i
1	$y_1 \sim N(\alpha_0 + \beta_0 x_1, \sigma^2)$	$1-p$	$z_1 = 0$
	$y_1 \sim N(\alpha_0 + (\beta_0 + \delta) x_1, \sigma^2)$	p	$z_1 = 1$
2	$y_2 \sim N(\alpha_0 + \beta_0 x_2, \sigma^2)$	$(1-p)^2$	all z_i's are 0
	$y_2 \sim N(\alpha_0 + (\beta_0 + \delta) x_2, \sigma^2)$	$2p(1-p)$	one z_i is 1
	$y_2 \sim N(\alpha_0 + (\beta_0 + 2\delta) x_2, \sigma^2)$	p^2	all z_i's are 1
3	$y_3 \sim N(\alpha_0 + \beta_0 x_3, \sigma^2)$	$(1-p)^3$	all z_i's are 0
	$y_3 \sim N(\alpha_0 + (\beta_0 + \delta) x_3, \sigma^2)$	$3p(1-p)^2$	one z_i is 1
	$y_3 \sim N(\alpha_0 + (\beta_0 + 2\delta) x_3, \sigma^2)$	$3p^2(1-p)$	two z_i's are 1
	$y_3 \sim N(\alpha_0 + (\beta_0 + 3\delta) x_3, \sigma^2)$	p^3	all z_i's are 1
4	$y_4 \sim N(\alpha_0 + \beta_0 x_4, \sigma^2)$	$(1-p)^4$	all z_i's are 0

Period	Distribution of y_t	Probability	Values of z_i
	$y_4 \sim N(\alpha_0 + (\beta_0 + \delta)\, x_4, \sigma^2)$	$4p(1-p)^3$	one z_i is 1
	$y_4 \sim N(\alpha_0 + (\beta_0 + 2\delta)\, x_4, \sigma^2)$	$6p^2(1-p)^2$	two z_i's are 1
	$y_4 \sim N(\alpha_0 + (\beta_0 + 3\delta)\, x_4, \sigma^2)$	$4p^3(1-p)$	3 z_i's are 1
	$y_4 \sim N(\alpha_0 + (\beta_0 + 4\delta)\, x_4, \sigma^2)$	p^4	all z_i's are 1.
5	$y_5 \sim N(\alpha_0 + \beta_0\, x_5, \sigma^2)$	$(1-p)^5$	all z_i's are 0
	$y_5 \sim N(\alpha_0 + (\beta_0 + \delta)\, x_5, \sigma^2)$	$5p(1-p)^4$	one z_i is 1
	$y_5 \sim N(\alpha_0 + (\beta_0 + 2\delta)\, x_5, \sigma^2)$	$10p^2(1-p)^3$	two z_i's are 1
	$y_5 \sim N(\alpha_0 + (\beta_0 + 3\delta)\, x_5, \sigma^2)$	$10p^3(1-p)^2$	3 z_i's are 1
	$y_5 \sim N(\alpha_0 + (\beta_0 + 4\delta)\, x_5, \sigma^2)$	$5p^4(1-p)$	4 z_i's are 1
	$y_5 \sim N(\alpha_0 + (\beta_0 + 5\delta)\, x_5, \sigma^2)$	p^5	all z_i's are 1

and so on.

6.2.1.1 Derivation of Covariance Matrix (Ω_{IS}) when Slope is Changing in One Direction

Under the same assumptions as stated in the previous section regarding the probability distribution of z_i, the expectation, variance and covariance of z_i are

$$E(z_i) = p,$$

$$\text{var}(z_i) = p(1-p), \text{and}$$

$$Cov(z_i z_j) = 0, \text{ for } i \neq j.$$

Under these assumptions the y_t given by (6.4) have expected value

$$E(y_t) = \alpha_0 + \beta_0 x_t + x_t \delta\, E(\sum_{i=1}^{t} z_i) = \alpha_0 + (\beta_0 + t\ p\delta)\, x_t .$$

The variance of y_t under the assumptions that

$$E(u_i) = 0,\ E(u_i u_j) = 0, \text{ for } i \neq j,\ E(u_i^2) = \sigma^2$$

and that the x_t are fixed is

$$\begin{aligned} \text{var}(y_t) &= \text{var}(\alpha_0 + (\beta_0 + \delta \sum_{i=1}^{t} z_i)\, x_t) + \text{var}(u_t) \\ &= t\, p(1-p)\delta^2 x_t^2 + \sigma^2 . \end{aligned}$$

Also the covariances of the y_t's are

$$Cov(y_i, y_j) = \min(i, j) p(1-p)\, \delta^2 x_i x_j, \text{ for } i \neq j .$$

From the above results, model (6.4) can be rewritten as

$$y_t = \alpha_0 + (\beta_0 + t\ p\delta) x_t + \omega_t ,$$

where $\omega_t = u_t + v_t$; $v_t = (\tau_t - pt) x_t \delta$;

$$v_t \sim N(0, t\, p(1-p)\delta^2 x_t^2) \text{ and}$$

$$\omega_t \sim N(0, \sigma^2 + t\, p(1-p)\delta^2 x_t^2) .$$

The variance-covariance matrix of y is

$$\Omega_{1S} = \begin{bmatrix} p(1-p)\delta^2 x_1^2 + \sigma^2 & p(1-p)\delta^2 x_1 x_2 & \cdots & p(1-p)\delta^2 x_1 x_n \\ p(1-p)\delta^2 x_2 x_1 & 2p(1-p)\delta^2 x_2^2 + \sigma^2 & \cdots & 2p(1-p)\delta^2 x_2 x_n \\ \vdots & \vdots & \vdots & \\ p(1-p)\delta^2 x_n x_1 & 2p(1-p)\delta^2 x_n x_2 & \cdots & np(1-p)\delta^2 x_n^2 + \sigma^2 \end{bmatrix} . \quad (6.6)$$

6.2.1.2 Derivation of Covariance Matrix (Ω_{2S}) when Slope is Changing in Two Directions

Consider a sequence of independent random variables z_i, $i = 1, 2, \cdots, n$, which can take three values –1, 0 and 1 with corresponding probabilities $q, 1-p-q$ and p respectively. In other words z_i are defined by

$$z_i = \begin{cases} -1 & \text{with probability } q, \\ 0 & \text{with probability } 1-p-q, \\ 1 & \text{with probability } p, \end{cases}$$

where p and q are small and unknown probabilities. The expectation, variance and covariance of z_i are now

$$E(z_i) = (p-q),$$

$$\text{var}(z_i) = ((p+q)-(p-q)^2) \text{ and}$$

$$Cov(z_i z_j) = 0, \text{ for } i \neq j.$$

Under these assumptions, the y_t given by (6.4) have expected value

$$E(y_t) = \alpha_0 + \beta_0 x_t + x_t \delta E(\sum_{i=1}^{t} z_i) = \alpha_0 + (\beta_0 + t(p-q)\delta) x_t.$$

The variance of y_t under the assumptions that

$$E(u_i) = 0, \; E(u_i u_j) = 0, \text{ for } i \neq j, \; E(u_i^2) = \sigma^2$$

and that the x_t are fixed is

$$\text{var}(y_t) = \text{var}(\alpha_0 + (\beta_0 + \delta \sum_{i=1}^{t} z_i) x_t) + \text{var}(u_t)$$
$$= t((p+q)-(p-q))\delta^2 x_t^2 + \sigma^2.$$

Also the covariances of the y_t's are

$$Cov(y_i, y_j) = \min(i, j)((p+q)-(p-q)^2)\delta^2 x_i x_j, \text{ for } i \neq j.$$

From the above results, model (6.4) can be rewritten as

$$y_t = \alpha_0 + (\beta_0 + t\,(p-q)\delta)x_t + \omega_t,$$

where $\omega_t = u_t + v_t$; $v_t = (\tau_t - (p-q)t)x_t\delta$;

$$v_t \sim N(0, t\,((p+q)-(p-q))\delta^2 x_t^2)$$

$$\omega_t \sim N(0, \sigma^2 + t((p+q)-(p-q))\delta^2 x_t^2).$$

The variance-covariance matrix of y is

$$\Omega_{2S} = \begin{bmatrix} c_{2s}x_1^2+\sigma^2 & c_{2s}x_1x_2 & \cdots & c_{2s}x_1x_n \\ c_{2s}x_2x_1 & 2c_{2s}x_2^2+\sigma^2 & \cdots & 2c_{2s}x_2x_n \\ \vdots & \vdots & & \vdots \\ c_{2s}x_nx_1 & 2c_{2s}x_nx_2 & \cdots & nc_{2s}x_n^2+\sigma^2 \end{bmatrix} \tag{6.7}$$

where $c_{2s} = ((p+q)-(p-q)^2)\delta^2$.

6.2.1.3 Derivation of Covariance Matrix (Ω_{3S}) when Slope is Changing in Two Directions by Different Amounts

Consider a sequence of independent random variables z_i, $i = 1, 2, \cdots, n$ which can take three values $-\delta_1, 0$ and δ_2 with corresponding probabilities $q, 1-p-q$ and p respectively. In other words, the z_i are defined by

$$z_i = \begin{cases} -\delta_1 & \text{with probability } q, \\ 0 & \text{with probability } 1-p-q, \\ \delta_2 & \text{with probability } p, \end{cases}$$

where p and q are small and unknown probabilities. The expectation, variance and covariance of z_i are now

$$E(z_i) = \delta_2 p - \delta_1 q,$$

$$\text{var}(z_i) = \delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2 \text{ and}$$

$$Cov(z_i z_j) = 0, \text{ for } i \neq j.$$

Under these assumptions, the y_t given by (6.4) have expected value

$$E(y_t) = \alpha_0 + \beta_0 x_t + x_t \; E(\sum_{i=1}^{t} z_i) = \alpha_0 + (\beta_0 + t(\delta_2 p - \delta_1 q))x_t.$$

The variance of y_t under the assumptions that

$$E(u_i) = 0, \; E(u_i u_j) = 0, \text{ for } i \neq j, \; E(u_i^2) = \sigma^2$$

and that the x_t are fixed is

$$\begin{aligned} \text{var}(y_t) &= \text{var}(\alpha_0 + (\beta_0 + \sum_{i=1}^{t} z_i)x_t) + \text{var}(u_t) \\ &= (\delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2)\, t\, x_t^2 + \sigma^2. \end{aligned}$$

Also the covariances of the y_t's are

$$\begin{aligned} Cov(y_i, y_j) &= \min(i, j)\, \text{var}(z_i) x_i x_j \\ &= \min(i, j)(\delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2) x_i x_j, \quad \text{for} \quad i \neq j. \end{aligned}$$

From the above results, model (6.4) can be rewritten as

$$y_t = \alpha_0 + (\beta_0 + t\,(\delta_2 p - \delta_1 q))x_t + \omega_t,$$

where $\omega_t = u_t + v_t$; $v_t = (\tau_t - (\delta_2 p - \delta_1 q)t)x_t \delta$;

$$v_t \sim N(0, (\delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2)\, t\, x_t^2)$$

$$\omega_t \sim N(0, \sigma^2 + (\delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2)\, t\, x_t^2).$$

The variance-covariance matrix of y is

$$\Omega_{3S} = \begin{bmatrix} c_{3s}x_1^2 + \sigma^2 & c_{3s}x_1x_2 & \cdots & c_{3s}x_1x_n \\ c_{3s}x_2x_1 & 2c_{3s}x_2^2 + \sigma^2 & \cdots & 2c_{3s}x_2x_n \\ \vdots & \vdots & & \vdots \\ c_{3s}x_nx_1 & 2c_{3s}x_nx_2 & \cdots & nc_{3s}x_n^2 + \sigma^2 \end{bmatrix} \tag{6.8}$$

where $c_{3s} = \delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2$.

6.2.2 Distributional Pattern of y_t for Changing Intercept

Consider the slope changing model (6.5) and assume δ is fixed. We are interested to find the distribution of y_t. We note the following:

Period	Distribution of y_t	Probability	Values of z_i
1	$y_1 \sim N(\alpha_0 + \beta_0 x_1, \sigma^2)$	$1-p$	$z_1 = 0$
	$y_1 \sim N((\alpha_0 + \delta) + \beta_0 x_1, \sigma^2)$	p	$z_1 = 1$.
2	$y_2 \sim N(\alpha_0 + \beta_0 x_2, \sigma^2)$	$(1-p)^2$	all z_i's are 0
	$y_2 \sim N((\alpha_0 + \delta) + \beta_0 x_2, \sigma^2)$	$2p(1-p)$	one z_i is 1
	$y_2 \sim N((\alpha_0 + 2\delta) + \beta_0 x_2, \sigma^2)$	p^2	all z_i's are 1.
3	$y_3 \sim N(\alpha_0 + \beta_0 x_3, \sigma^2)$	$(1-p)^3$	all z_i's are 0
	$y_3 \sim N((\alpha_0 + \delta) + \beta_0 x_3, \sigma^2)$	$3p(1-p)^2$	one z_i is 1
	$y_3 \sim N((\alpha_0 + 2\delta) + \beta_0 x_3, \sigma^2)$	$3p^2(1-p)$	two z_i's are 1
	$y_3 \sim N((\alpha_0 + 3\delta) + \beta_0 x_3, \sigma^2)$	p^3	all z_i's are 1.
4	$y_4 \sim N(\alpha_0 + \beta_0 x_4, \sigma^2)$	$(1-p)^4$	all z_i's are 0

Period	Distribution of y_t	Probability	Values of z_i
	$y_4 \sim N((\alpha_0+\delta)+\beta_0 x_4, \sigma^2)$	$4p(1-p)^3$	one z_i is 1
	$y_4 \sim N((\alpha_0+2\delta)+\beta_0 x_4, \sigma^2)$	$6p^2(1-p)^2$	two z_i's are 1
	$y_4 \sim N((\alpha_0+3\delta)+\beta_0 x_4, \sigma^2)$	$4p^3(1-p)$	3 z_i's are 1
	$y_4 \sim N((\alpha_0+4\delta)+\beta_0 x_4, \sigma^2)$	p^4	all z_i's are 1.
5	$y_5 \sim N(\alpha_0+\beta_0 x_5, \sigma^2)$	$(1-p)^5$	all z_i's are 0
	$y_5 \sim N((\alpha_0+\delta)+\beta_0 x_5, \sigma^2)$	$5p(1-p)^4$	one z_i is 1
	$y_5 \sim N((\alpha_0+2\delta)+\beta_0 x_5, \sigma^2)$	$10p^2(1-p)^3$	two z_i's are 1
	$y_5 \sim N((\alpha_0+3\delta)+\beta_0 x_5, \sigma^2)$	$10p^3(1-p)^2$	3 z_i's are 1
	$y_5 \sim N((\alpha_0+4\delta)+\beta_0 x_5, \sigma^2)$	$5p^4(1-p)$	4 z_i's are 1
	$y_5 \sim N((\alpha_0+5\delta)+\beta_0 x_5, \sigma^2)$	p^5	all z_i's are 1

and so on.

6.2.2.1 Derivation of Covariance Matrix (Ω_{IC}) when Intercept is Changing in One Direction

Consider a sequence of independent random variables z_i, $i=1,2,\cdots,n$, which can take two values 0 and 1 with corresponding probabilities $1-p$ and p respectively. In other words z_i, are defined by

$$z_i = \begin{cases} 0 & \text{with probability } 1-p \\ 1 & \text{with probability } p \end{cases}$$

where p is a small and unknown probability. The expectation, variance and covariance of z_i are now

$$E(z_i) = p,$$

$$\text{var}(z_i) = p(1-p) \text{ and}$$

$$Cov(z_i z_j) = 0, \text{ for } i \neq j.$$

Under these assumptions, the y_t given by (6.5) have expected value

$$E(y_t) = \alpha_0 + \delta\, E(\sum_{i=1}^{t} z_i) + \beta_0 x_t = (\alpha_0 + t\,p\delta) + \beta_0 x_t.$$

and the variance of y_t is

$$\text{var}(y_t) = \text{var}(\alpha_0 + \delta \sum_{i=1}^{t} z_i) + \text{var}(u_t) = t\,p(1-p)\delta^2 + \sigma^2.$$

Also the covariances of the y_t's are

$$Cov(y_i, y_j) = \min(i, j)\,p(1-p)\delta^2 \text{ for } i \neq j.$$

From the above results, model (6.5) can be rewritten as

$$y_t = \alpha_0 + t\,p\delta + \beta_0 x_t + \omega_t,$$

where $\omega_t = u_t + v_t$; $v_t = (\tau_t - pt)\delta$;

$$v_t \sim N(0, t\,p(1-p)\delta^2),\ \omega_t \sim N(0, t\,p(1-p)\delta^2 + \sigma^2).$$

The variance-covariance matrix of y is

$$\Omega_{IC} = \begin{bmatrix} p(1-p)\delta^2 + \sigma^2 & p(1-p)\delta^2 & \cdots & p(1-p)\delta^2 \\ p(1-p)\delta^2 & 2p(1-p)\delta^2 + \sigma^2 & \cdots & 2p(1-p)\delta^2 \\ \vdots & \vdots & & \vdots \\ p(1-p)\delta^2 & 2p(1-p)\delta^2 & \cdots & np(1-p)\delta^2 + \sigma^2 \end{bmatrix} \quad (6.9)$$

6.2.2.2 Derivation of Covariance Matrix (Ω_{2C}) when Intercept is Changing in Two Directions

Consider now a sequence of independent random variables z_i, $i = 1, 2, \cdots, n$, which can take three values –1, 0 and 1 with corresponding probabilities $q, 1-p-q$ and p respectively. In other words, z_i are defined by

$$z_i = \begin{cases} -1 & \text{with probability } q, \\ 0 & \text{with probability } 1-p-q, \\ 1 & \text{with probability } p, \end{cases}$$

where p and q are small and unknown probabilities. The expectation, variance and covariance of z_i are now

$$E(z_i) = (p-q),$$

$$\text{var}(z_i) = ((p+q)-(p-q)^2) \text{ and}$$

$$Cov(z_i z_j) = 0, \text{ for } i \neq j.$$

Under these assumptions, the y_t given by (6.4) have expected value

$$E(y_t) = \alpha_0 + \delta\, E(\sum_{i=1}^{t} z_i) + \beta_0 x_t = (\alpha_0 + t\,(p-q)\delta) + \beta_0 x_t.$$

The variance of y_t is

$$\begin{aligned} \text{var}(y_t) &= \text{var}(\alpha_0 + \delta \sum_{i=1}^{t} z_i) + \text{var}(u_t) \\ &= t((p+q)-(p-q)^2)\delta^2 + \sigma^2. \end{aligned}$$

Also the covariances of the y_t's are

$$Cov(y_i, y_j) = \min(i, j)((p+q)-(p-q)^2)\delta^2, \text{ for } i \neq j.$$

From the above results, model (6.5) can be rewritten as

$$y_t = \alpha_0 + t(p-q)\delta + \beta_0 x_t + \omega_t,$$

where $\omega_t = u_t + v_t$; $v_t = (\tau_t - (p-q)t)\delta$;

$$v_t \sim N(0, t((p+q)-(p-q)^2))\delta^2),$$

$$\omega_t \sim N(0, \sigma^2 + t((p+q)-(p-q)^2)\delta^2).$$

The variance-covariance matrix of y is

$$\Omega_{2C} = \begin{bmatrix} c_{2c}+\sigma^2 & c_{2c} & \cdots & c_{2c} \\ c_{2c} & 2c_{2c}+\sigma^2 & \cdots & 2c_{2c} \\ \vdots & \vdots & & \vdots \\ c_{2c} & 2c_{2c} & \cdots & nc_{2c}+\sigma^2 \end{bmatrix} \tag{6.10}$$

where $c_{2c} = ((p+q)-(p-q)^2)\delta^2$.

6.2.2.3 Derivation of Covariance Matrix (Ω_{3C}) when Intercept is Changing in Two Directions by Different Amounts

Consider a sequence of independent random variables z_i, $i = 1, 2, \cdots, n$ which can take three values $-\delta_1, 0$ and δ_2 with corresponding probabilities $q, 1-p-q$ and p respectively. In other words, the z_i are defined by

$$z_i = \begin{cases} -\delta_1 & \text{with probability } q, \\ 0 & \text{with probability } 1-p-q, \\ \delta_2 & \text{with probability } p, \end{cases}$$

where p and q are small and unknown probabilities. The expectation, variance and covariance of z_i are now

$$E(z_i) = \delta_2 p - \delta_1 q,$$

$$\text{var}(z_i) = \delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2 \text{ and}$$

$$Cov(z_i z_j) = 0, \text{ for } i \neq j.$$

Under these assumptions, the y_t given by (6.5) have expected value

$$E(y_t) = \alpha_0 + E(\sum_{i=1}^{t} z_i) + \beta_0 x_t = (\alpha_0 + t(\delta_2 p - \delta_1 q)) + \beta_0 x_t.$$

The variance of y_t is

$$\begin{aligned}\text{var}(y_t) &= \text{var}(\alpha_0 + \sum_{i=1}^{t} z_i) + \text{var}(u_t) \\ &= t(\delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2) + \sigma^2.\end{aligned}$$

Also the covariances of the y_t's are

$$Cov(y_i, y_j) = \min(i, j)(\delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2), \text{ for } i \neq j.$$

From the above results, model (6.5) can be rewritten as

$$y_t = \alpha_0 + t(\delta_2 p - \delta_1 q) + \beta_0 x_t + \omega_t,$$

where $\omega_t = u_t + v_t$; $v_t = (\tau_t - (\delta_2 p - \delta_1 q)t)\delta$;

$$v_t \sim N(0, t(\delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2))$$

$$\omega_t \sim N(0,\ \sigma^2 + t(\delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2)).$$

The variance-covariance matrix of y is

$$\Omega_{3c} = \begin{bmatrix} c_{3c} + \sigma^2 & c_{3c} & \cdots & c_{3c} \\ c_{3c} & 2c_{3c} + \sigma^2 & \cdots & 2c_{3c} \\ \vdots & \vdots & & \vdots \\ c_{3c} & 2c_{3c} & \cdots & nc_{3c} + \sigma^2 \end{bmatrix} \tag{6.11}$$

where $c_{3c} = \delta_1^2 q + \delta_2^2 p - (\delta_1 q + \delta_2 p)^2$.

The Ω_{ij} $i=1,2,3$ and $j=s,c$ matrices help in data generation as well as in calculating ML and MML estimates with a minimum of computational effort and time. They are also needed for prediction.

6.3 Estimation and Prediction Procedures

After making a decision about the form of the model in the presence of possible changepoints in the intercept or slope, the parameters of the model have to be determined in the best possible manner from the available data. In other words, the parameters of the changing parameters of the model are unknown and we need to estimate them from the data available at hand. In the following section, we discuss some estimation and prediction procedures.

In the previous section, we outlined changepoint models with unknown parameters α, β, δ_1, δ_2, p, q, and σ to be estimated. We discussed three special situations and developed three variance-covariance matrices Ω_{iS}; $i=1,2,3$ for changing slope and three variance-covariance matrices Ω_{iC}; $i=1,2,3$ for changing intercept. For simplicity, we denote $\gamma=[\alpha,\beta]'$, $\psi=[\delta_1,\delta_2,p,q,\sigma]'$ and Ω for the variance-covariance matrix and use these variance-covariance matrices to estimate parameters. We now discuss some prediction procedures below.

We consider the linear regression model

$$y=X\gamma+u, \qquad u\sim N(0,\sigma^2\Omega(\psi)) \tag{6.12}$$

where y is $n\times 1$, X is an $n\times k$ nonstochatic matrix of rank $k<n$, γ is a $k\times 1$ vector of unknown parameters, u is an $n\times 1$ disturbance vector assumed to be

normally distributed with expected value 0 and covariance matrix $\sigma^2\Omega(\psi)$ which is an $n \times n$ known positive-definite matrix. If the disturbances are independent and homoscedastic then $\Omega(\psi) = I$ and so $u \sim N(0, \sigma^2 I)$. The main problem in this context is the estimation of the parameter ψ. Therefore our initial emphasis is on estimating ψ and in that case γ and σ^2 are regarded as unknown nuisance parameters. Once ψ is estimated, γ and σ^2 can be easily estimated as a consequence.

The OLS predictor of y, $\hat{y}$, is defined as

$$\hat{y} = X\hat{\gamma}_{OLS}, \tag{6.13}$$

where $\hat{\gamma}_{OLS} = (X'X)^{-1}X'y$ is the OLS estimator of γ. When OLS is applied, we assume that the errors are independent of each other and do not suffer from heteroscedasticity.

We are interested in estimating the parameter ψ and this can be done by maximizing the log-likelihood. The process can be described as follows. The log-likelihood function is

$$L(\gamma, \psi) = \text{const.} - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\log|\Omega(\psi)| - \frac{(y - X\gamma)'\Omega^{-1}(\psi)(y - X\gamma)}{2\sigma^2} \tag{6.14}$$

which is a function of γ, ψ and σ^2. Its first derivatives with respect to γ and σ^2 when set to zero, provide the estimated values of γ and σ^2 as $\hat{\gamma}$ and $\hat{\sigma}^2$ respectively

$$\hat{\gamma} = \left(X'\Omega^{-1}(\psi)X\right)^{-1}X'\Omega^{-1}(\psi)y, \tag{6.15}$$

$$\hat{\sigma}^2 = \frac{(y - X\hat{\gamma})' \Omega^{-1}(\psi)(y - X\hat{\gamma})}{n}, \tag{6.16}$$

assuming ψ fixed. Substituting the estimated values of $\hat{\gamma}$ and $\hat{\sigma}^2$ in the above log-likelihood function, the maximum log-likelihood (ML) value can be found from

$$ML = \text{const.} - \frac{n}{2}\log\hat{\sigma}^2 - \frac{1}{2}\log|\Omega(\psi)| - \frac{n}{2}, \tag{6.17}$$

which is a function of ψ. We need to estimate ψ by maximizing this function. This can be done by using a suitable computing algorithm namely Newton-Raphson, Fisher's scoring method, simulated annealing (Krikpatrick et. al (1983)) or the method of Brendt, Hall, Hall and Hausman (1974) that maximizes the likelihood for different values of ψ by an iterative procedure.

There is a literature which suggests that maximizing the marginal likelihood gives better results than the maximum likelihood method. Fraser (1967), and Kalbfleisch and Sprott (1970) introduced maximum marginal likelihood (MML). Ara and King (1993, 1995), and Rahman and King (1997) used the marginal likelihood to construct different tests and observed a significant improvement in small sample properties over those of traditional tests. The main principle is to transform the dependent variable to another random vector, and a subvector of which has a likelihood (marginal likelihood) that only contains the parameters of interest and the remainder of which contains no information about those parameters. From Tunnicliffe Wilson (1989), our marginal likelihood is

$$MGL = |\Omega(\psi)|^{-\frac{1}{2}} \left| X' \Omega^{-1}(\psi) X \right|^{-\frac{1}{2}} \hat{s}^{-\frac{n-k}{2}} \tag{6.18}$$

where $\hat{s} = (y - X\hat{\gamma})'\Omega^{-1}(\psi)(y - X\hat{\gamma})$ and $\hat{\gamma}$ as given in (6.15). The MML estimate of the parameter can be obtained by maximizing the marginal likelihood function with respect to ψ. To obtain a final estimate of γ, we need to estimate the value of ψ by maximizing either equation (6.17) or (6.18) and then replace ψ by the estimated value in equation (6.15). Suppose we wish to find the one-step-ahead forecasts of y in model (6.12), which can be written as

$$y_{n+1} = x'_{n+1}\gamma + u_{n+1} \tag{6.19}$$

where y_{n+1} is the next value of y, x_{n+1} is the $k \times 1$ vector of observations on the regressors at time $n+1$ and u_{n+1} is its associated disturbance term. The predicted value of y_{n+1} can be written as

$$\hat{y}^j_{n+1} = x'_{n+1}\hat{\gamma}^j + \upsilon(\hat{\psi}^j)'\Omega^{-1}(\hat{\psi}^j)(y - X\hat{\gamma}^j) \tag{6.20}$$

where $\hat{\psi}^j = \left[\hat{\delta}^j_1, \hat{\delta}^j_2, \hat{p}^j, \hat{q}^j, \hat{\sigma}^j\right]'$, $j = 1, 2$, in which $j = 1$ indicates the estimated value of γ comes from the ML method by maximising (6.17) and $j = 2$ indicates the estimated value comes from the MML method by maximising (6.18) and following Goldberger (1962) and Toyooka (1982), $\upsilon(\psi) = \mathrm{E}(u_{n+1}u)$ is the top $n \times 1$ vector from the final column of variance-covariance matrices Ω_{iS}; $i = 1, 2, 3$, for changing slope and Ω_{iC}; $i = 1, 2, 3$, for changing intercept, in the case of a sample size of $n+1$. For example, in the case of a changing slope in one direction

$$\upsilon(\psi) = \begin{bmatrix} p(1-p)\delta^2 x_1 x_{n+1} \\ 2p(1-p)\delta^2 x_2 x_{n+1} \\ \vdots \\ np(1-p)\delta^2 x_n x_{n+1} \end{bmatrix},$$

in the case of a changing slope in two directions

$$\upsilon(\psi) = \begin{bmatrix} c_{2s}x_1x_{n+1} \\ 2c_{2s}x_2x_{n+1} \\ \vdots \\ nc_{2s}x_nx_{n+1} \end{bmatrix},$$

in the case of a changing slope in two directions by different amounts

$$\upsilon(\psi) = \begin{bmatrix} c_{3s}x_1x_{n+1} \\ 2c_{3s}x_2x_{n+1} \\ \vdots \\ nc_{3s}x_nx_{n+1} \end{bmatrix},$$

in the case of a changing intercept in one direction

$$\upsilon(\psi) = \begin{bmatrix} p(1-p)\delta^2 \\ 2p(1-p)\delta^2 \\ \vdots \\ (n+1)p(1-p)\delta^2 \end{bmatrix},$$

in the case of a changing intercept in two directions

$$\upsilon(\psi) = \begin{bmatrix} c_{2c} \\ 2c_{2c} \\ \vdots \\ (n+1)c_{2c} \end{bmatrix},$$

and in the case of a changing intercept in two directions by different amounts

$$\upsilon(\psi) = \begin{bmatrix} c_{3c} \\ 2c_{3c} \\ \vdots \\ (n+1)c_{3c} \end{bmatrix}.$$

6.4. Prediction Schemes

For prediction, which relies on the estimates of unknown parameters, in this section we discuss three different schemes for estimating the parameters using the available data. We consider these schemes explicitly and discuss them below.

6.4.1 Scheme-1: Fixed

The first scheme, which we call fixed, and which was used by, for example Pagan and Schwert (1990), is as follows. Let $y_1,\cdots,y_n$ be generated data from a model of interest. We divide the total generated data into two parts, i.e., $n = n_1 + n_2$. The first n_1 data points are used to estimate the parameters and then predict for n_2 observations using information from the fitted model. This scheme estimates β, just once, using data from 1 to n_1, and uses the one estimate in forming all n_2 predictions. In this case, $\hat{\beta}$ is the same for all n, and depends only on n_1.

6.4.2 Scheme-2: Recursive

The second scheme, which we call recursive, was used by, for example, Fair and Shiller (1990). This scheme uses all available data, estimating β first with data from 1 to n_1, next with data from 1 to n_1+1, so on, and finally with data from 1 to n. In the model $y_t = x_t'\beta + u_t$, for example, β is estimated using data from 1 to $n_1 + r$ in the recursive scheme, $\hat{\beta}_{1+r} = (\sum_{s=1}^{n_1+r} x_s'x_s)^{-1}\sum_{s=1}^{n_1+r} x_s'y_s,\ r = 0,\cdots,n_2$.

6.4.3 Scheme-3: Rolling

The third scheme, which we call rolling, was used by, for example, Akgiray (1989). This scheme fixes the sample size, say at n_1, and drops distant observations as recent ones are added. Thus, β is estimated first with data from 1 to n_1, next with data from 2 to n_1+1, so on, and finally with data from $n-n_1+1$ to n. In the least squares model $y_t = x_t'\beta + u_t$, for example, β is estimated using data from 1 to n_1, 2 to n_1+1, and $n-n_1+1$ to n in the rolling scheme, $\hat{\beta} = (\sum_{s=n-n_1+1}^{n} x_s'x_s)^{-1} \sum_{s=n-n_1+1}^{n} x_s'y_s$. In the case of a rolling scheme, different regression estimates are used for each sample size.

6.5 The Monte Carlo Experiment

In order to compare average prediction errors for different situations in linear regression models with structural changes of random timing, we conducted a Monte Carlo experiment. The experiment aimed to evaluate the predictive performance of different models, different schemes and different forecasting methods.

6.5.1 Experimental Design

The following design matrices were used in the experiment:

X_1 A constant dummy plus two independent stationary AR(1) regressors generated as, $x_{it} = 0.5x_{it-1} + \eta_{it}$ where $\eta_{it} \sim IN(0,1)$, $t = 1, \ldots, n$, $i = 1, 2$.

X_2 A constant dummy plus two independent trending regressors generated as $x_{it} = 0.25t + w_{it}$ where w_{it}, $t = 1, \ldots, n$, $i = 2, 3$, is an AR(1) time series generated as for X_1 above.

X_3 A constant, monthly US seasonally adjusted total volume of real retail sales on domestic trade (in billion 1992 USD) and lagged one month commencing 1960(1).

X_4 A constant, monthly US seasonally adjusted personal income (in billion 1995 USD) and lagged one month commencing 1960(1).

X_5 A constant, monthly US interest rate, the same interest rate lagged one month, real personal income (in billion 1995 USD), and the same variable lagged one month commencing 1960(1).

These design matrices were chosen to reflect a variety of economic and statistical phenomena. X_1 is comprised of stationary regressors, X_2 and X_4 have trending regressors while X_3 and X_5 show some long term fluctuations. After the disturbances were generated, and given the appropriate design matrix, the y's are generated from the following equations:

$$y_t = \alpha_0 + x_t'\beta_0 + u_t, \qquad t = 1, 2, \cdots, n \qquad \text{(Model 1)}$$

where $u_t \sim IN(0, \sigma^2)$.

$$y_t = \begin{cases} \alpha_0 + x_t'\beta_0 + u_t & \text{for} \quad t = 1, \cdots, t_1 \\ \alpha_0 + x_t'(\beta_0 + \delta) + u_t & \text{for} \quad t = t_1 + 1, \cdots, n \end{cases} \qquad \text{(Model 2)}$$

where $u_t \sim IN(0, \sigma^2)$.

$$y_t = \alpha_0 + x_t'(\beta_0 + \delta\sum_{i=1}^{t} z_i) + u_t, \qquad t = 1, 2, \cdots, n \qquad \text{(Model 3)}$$

where the z_i are independently distributed

$$z_i = \begin{cases} 1 & \text{with probability} \quad p, \\ 0 & \text{with probability} \quad 1 - p, \end{cases}$$

and $u_t \sim IN(0, \sigma^2)$.

$$y_t = \alpha_0 + x_t'(\beta_0 + \delta\sum_{i=1}^{t} z_i) + u_t, \qquad t = 1, 2, \cdots, n \qquad \text{(Model 4)}$$

where the z_i are independently distributed

$$z_i = \begin{cases} -1 & \text{with probability } q, \\ 0 & \text{with probability } 1 - p - q, \\ 1 & \text{with probability } p, \end{cases}$$

and $u_t \sim IN(0, \sigma^2)$.

$$y_t = \alpha_0 + x_t'(\beta_0 + \sum_{i=1}^{t} z_i) + u_t, \qquad t = 1, 2, \cdots, n \qquad \text{(Model 5)}$$

where the z_i are independently distributed

$$z_i = \begin{cases} -\delta_1 & \text{with probability } q, \\ 0 & \text{with probability } 1 - p - q, \\ \delta_2 & \text{with probability } p, \end{cases}$$

and $u_t \sim IN(0, \sigma^2)$.

We performed a pilot experiment where different combinations of α_0 and β_0 values were used and it was found that the prediction error values did not change for the changes in α_0 and β_0 values. In other words, prediction errors are largely invariant to changes in α_0 and β_0. We set α_0, β_0 and σ to unity. In addition, for each model M_j, we set $p = 0.005$ and $q = 0.005$. The values of δ_1 and δ_2 were chosen in such a way that $\frac{\text{var}(\delta_i x_t)}{\text{var}(u_t)} \approx 1$. In the end we set δ_1 and δ_2 to 0.15 and 0.35 respectively.

After the y's were generated according to each of the above true models, the parameters of the model were estimated by using ML, MML and OLS methods then one-step-ahead predictions were produced by the recursive, rolling and fixed schemes in turn. Throughout, when ML, MML estimation was needed, it was conducted using the GAUSS 3.2.12 software.

The overall aim is to assess the performance of each of the above methods and prediction schemes. The experiment we present involved 2000 replications. In the case of the fixed prediction method, for each replication, we generated 480 data points. We throw away some initial observations and then split the remaining into regression (n_1) and prediction (n_2) samples. The prediction (n_2) samples were taken to be 200 in size and the regression (n_1) samples of size 25, 50, 100 and 200 were taken discarding 255, 230, 180 and 80 observations respectively from the generated data. In the case of Model 2, we use t_1 in such a way that for each regression period

n_1, the prediction period n_2, contains the changepoint. Based on this principle we choo. t_1 at the 220th position of the data period.

6.6 Prediction Accuracy

In Section 6.3 we discussed different methods of prediction. Once the predictions are obtained we need to evaluate their accuracy. The prediction error that is the difference between the predicted value and the actual value, can be used as the tool for prediction accuracy evaluation. If y_{n_1+1} is the observed value of y, for the time period n_1+1, and $\hat{y}_{n_1+1}$ is the predicted value of y given by (6.20), for the same period, then the prediction error is defined as $\hat{e}_{n_1+1} = \left[y_{n_1+1} - \hat{y}_{n_1+1}\right]$. These prediction errors were then compared in terms of the following statistical measures to evaluate the prediction performance.

6.6.1 Root Mean Square Error (RMSE)

The root mean square error is the square root of the sum of the squared prediction errors for each observation divided by the number of observations and defined as

$$RMSE = \sqrt{\frac{1}{n_2}\sum_{i=1}^{n_2}\hat{e}_{n_1+i}^2}$$

RMSE is popular among econometricians and statisticians. It is the most widely used prediction evaluation criterion in applied and theoretical research. RMSE is mathematically more amenable than other methods and it has a relationship

to the least squares criterion. The strategy that has minimum RMSE is often thought to be the best strategy in terms of prediction performance.

6.6.2 Mean Error (ME)

The ME is the mean or average of the forecasting errors and is defined as

$$ME = \frac{1}{n_2}\sum_{i=1}^{n_2}\hat{e}_{n_1+i}.$$

In general, MEs have a tendency towards zero in most forecasting situations. Positive ME indicates that on average the bias in prediction is downwards and negative ME indicates upward bias.

6.6.3 Mean Absolute Error (MAE)

The MAE involves calculating the average of the absolute value of the prediction errors, namely

$$MAE = \frac{1}{n_2}\sum_{i=1}^{n_2}\left|\hat{e}_{n_1+i}\right|$$

It is an acceptable measure when the loss of making a prediction error is proportional to the absolute size of the prediction error.

6.7 Discussion of Results

In this section we analyse the overall picture of our results separately. We took averages over the 2000 iterations of prediction errors (root mean square error, mean absolute error and mean error) using different sample sizes n (25, 50, 100 and 200), schemes (recursive, rolling and fixed), methods (ML, MML and OLS) and models.

In Tables 6.1 to 6.4 we report the average of ME, MAE and RMSE over 2000 iterations of model-1 considering different design matrices, schemes and sample sizes. The results show that as the sample size n increases, the average RMSEs for all predictors decreases. The comparative performances of different forecasting methods show that the OLS method outperforms the other methods in the sense that it has the minimum average RMSE. The performance of the ML based method was the second best and MML performed the worst. The ME and the MAE also support the overall good performance of the OLS method. Over the different design matrices, OLS has the lowest average RMSE for X_5 and the highest for X_1.

From Tables 6.5 to 6.8, we observe that the forecasting performance of the OLS method is the best compared to other methods for model-2 considering the average of ME, MAE and RMSE for different design matrices, schemes, and sample size. A trait of model-2 is that the MML method has the second best performance and the worst performance comes from ML. The difference between the best and worst performances is 0.0218, 0.0235, 0.0001 and 0.0559 in terms of RMSE respectively when sample sizes are 25, 50, 100 and 200. The above analysis means that when the sample size increases, the distance between best and worst becomes

narrow and then wide. The recursive scheme has the best performance and the rolling scheme has the second best while the fixed scheme has the worst performance. The difference between the best and worst performances of recursive scheme is 0.0787 in terms of RMSE when the sample size is 50. We have found the difference is not statistically significant.

The estimated average of ME, MAE and RMSE for different methods without considering different models, design matrices and sample sizes shows that OLS performs best, MML is the second best and ML is the worst performed method. The difference between the best and worst performances is 0.0554 in the context of RMSE. The recursive scheme gives the smallest prediction errors on average, and therefore is the best performed scheme, the rolling scheme performed second best and the fixed scheme works relatively poorly. The difference between the best and worst performance is 0.0119 in the context of RMSE when the sample size is 25 and for design matrix X_5.

In Tables 6.9 to 6.12 we report the average of ME, MAE and RMSE for model-3 considering different design matrices, schemes and sample sizes. The results show that as the sample size n increases, the average RMSE for all predictors decrease. The comparative performances of different forecasting methods show that the MML estimator outperforms the other methods in the sense that it has the minimum average RMSE. The performance of the ML estimator was the second best and the OLS estimator performed worst. The ME and MAE results also support the overall good performance of the MML method. The performance of the recursive scheme is the best, the rolling scheme is the second best and the fixed scheme is the worst.

The estimated averages of MAE and RMSE for model-4 for different design matrices and sample sizes are reported in Tables 6.13 to 6.16. A characteristic of model-4 is that the MML method performs best in the sense that it gives minimum average RMSEs in the case of all schemes and sample sizes. The ML method has the second best performance. The worst performance is from OLS. A clear feature of the results is that when the sample size increases, the gap between best and worst becomes wider. The results also show that as the sample size n increases, the average RMSEs for all predictors decrease. In other words, the performance of predictors changes with an increase in the sample size. The performance of the recursive scheme is the best, the rolling scheme is the second best and the fixed scheme is the worst.

From Tables 6.17 to 6.20, we observe that the forecasting performance of the MML method is the best compared to other methods for model-5 based the average of MAE and RMSE for different design matrices, schemes, and sample sizes. A trait of model-5 is that the ML method has the second best performance. The OLS method gives the worst performance. The difference between the best and worst performances is 0.0157, 0.0166, 0.0186 and 0.0459 in terms of RMSE respectively when sample sizes are 25, 50, 100 and 200. We have found that these differences are not statistically significant. The above analysis means that when the sample size increases, the distance between best and worst becomes broader. The recursive scheme has the best performance and the rolling scheme has the second best performance while the fixed scheme has the worst performance. The estimated average prediction errors for different methods without considering different models, design matrices and sample sizes shows that the MML performs best and ML is the

second best performed method. The worst performance is from OLS (in terms of MAE and RMSE). The difference between the best and worst performances is 0.01460 in context of RMSE. The recursive scheme gives minimum RMSE, and so is the best performed scheme, the rolling scheme performed second best and the fixed scheme works very poorly. The difference between the best and worst performance is 0.0139 in the context of RMSE, which is not statistically significant. The estimated predicted errors generally decrease as the sample size increases.

The estimated average prediction errors for different methods without considering different models, schemes, design matrices and sample sizes shows that the MML method performs best, ML is the second best and OLS has the worst performance.

Overall, in the case of models 1 and 2 OLS, gives minimum average RMSEs, on the other hand, in case of models 3 to 5, MML methods gives minimum average RMSEs. The estimated predicted average RMSE and MAE generally decreases as the sample size increases. Overall, one would have to recommend the use of MML method for models 3 to 5 and the OLS method for models 1 and 2 because their use can result an improvement in prediction in the sense of the minimum RMSE and MAE.

6.8 Conclusions

In this chapter, we investigated the forecasting performance of the linear regression model with a random change in coefficients. We derived the distributional patterns and especially the mean, variance and covariance structure of different linear regression models for stochastic changes in either the slope or intercept parameters in turn by a fixed amount with a very low probability. We found that this resulted in a linear regression with a nonscalar variance-covariance matrix, which allows standard approaches to estimation and prediction to be used.

We compared the predictive performances for three methods of estimation of parameters, three different schemes, four different sample sizes and five different models in turn and evaluated the forecasting performance of the estimators using RMSE, MAE and ME. The simulation results convey that the MML estimator is clearly better than the other estimator in terms of small sample properties for models 3 to 5, on the other hand the OLS estimator is better than others in the case of models 1 and 2. In addition, the MML estimator is quite promising in terms of average ME, MAE and RMSE. Estimation of covariance matrix parameters in the models demonstrate that if estimation is based on maximizing the marginal likelihood rather than the classical likelihood, then the average of ME, MAE and RMSE can be reduced for model-3 to model-5. In contrast for model-1 and model-2 predictions from OLS are quite different from those of MML and ML methods.

The results based on different prediction schemes show that recursive forecasts perform better than other forecasts. From the discussion of the results, we found that overall, the adaptation of the recursive scheme is always the optimal

choice. We, therefore, recommend the use of the OLS method and recursive scheme for models 1 and 2, and the MML method and recursive scheme when we deal with a situation in which there are possible random changes in either the intercept or slope of the model.

This study has some limitations that suggest that one has to take extreme care in making too many generalized conclusions from the results based on the Monte Carlo experiments. The results achieved by Monte Carlo experiments are often specific to the design of the experiment.

There is thus perhaps scope for further studies regarding the factors responsible for coefficient changes, testing the significance and constructing the confidence intervals of the changing parameters, as well as construction of prediction intervals. It would also be interesting to see how our predictors behave for different forecast lead times.

Table 6.1 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-1 when $n=25$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0560	1.0908	1.3517	0.0675	1.1750	1.4712	0.0869	1.1921	1.5084
MML	0.0558	1.0912	1.3514	0.0669	1.1744	1.4711	0.0864	1.1920	1.5094
OLS	0.0553	1.0901	1.3505	0.0668	1.1737	1.4701	0.0857	1.1918	1.5081
				X2					
ML	0.0464	1.0573	1.3321	0.0493	1.0621	1.3328	0.0490	1.0746	1.3345
MML	0.0468	1.0567	1.3321	0.0492	1.0614	1.3329	0.0485	1.0750	1.3346
OLS	0.0459	1.0562	1.3319	0.0482	1.0613	1.3323	0.0481	1.0737	1.3339
				X3					
ML	0.0463	1.0587	1.3373	0.0491	1.0654	1.3373	0.0490	1.0764	1.3412
MML	0.0473	1.0590	1.3372	0.0497	1.0644	1.3370	0.0490	1.0770	1.3411
OLS	0.0461	1.0585	1.3371	0.0485	1.0642	1.3368	0.0480	1.0757	1.3408
				X4					
ML	0.0479	1.0603	1.3344	0.0501	1.0619	1.3348	0.0518	1.0629	1.3354
MML	0.0483	1.0596	1.3341	0.0502	1.0621	1.3351	0.0515	1.0632	1.3350
OLS	0.0472	1.0591	1.3343	0.0493	1.0616	1.3347	0.0509	1.0622	1.3342
				X5					
ML	0.0481	1.0612	1.3361	0.0491	1.0615	1.3354	0.0526	1.0601	1.3147
MML	0.0479	1.0605	1.3362	0.0490	1.0616	1.3345	0.0529	1.0606	1.3156
OLS	0.0476	1.0602	1.3355	0.0487	1.0604	1.3342	0.0522	1.0593	1.3145

Table 6.2 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-1 when $n = 50$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0736	0.9812	1.2254	0.0738	0.9991	1.2575	0.0749	1.0105	1.2667
MML	0.0735	0.9823	1.2261	0.0733	0.9987	1.2580	0.0756	1.0112	1.2661
OLS	0.0726	0.9810	1.2249	0.0731	0.9985	1.2568	0.0744	1.0101	1.2660
				X2					
ML	0.0687	0.9769	1.2026	0.0744	0.9791	1.2030	0.0705	0.9803	1.2063
MML	0.0685	0.9765	1.2026	0.0744	0.9784	1.2037	0.0703	0.9799	1.2057
OLS	0.0676	0.9761	1.2021	0.0739	0.9781	1.2028	0.0693	0.9794	1.2056
				X3					
ML	0.0715	0.9757	1.1981	0.0695	0.9777	1.2002	0.0736	0.9792	1.1964
MML	0.0717	0.9757	1.1983	0.0695	0.9785	1.1990	0.0730	0.9792	1.1960
OLS	0.0706	0.9748	1.1972	0.0692	0.9773	1.1989	0.0725	0.9786	1.1954
				X4					
ML	0.0732	0.9762	1.1979	0.0729	0.9768	1.1991	0.0743	0.9780	1.1973
MML	0.0724	0.9754	1.1985	0.0734	0.9770	1.2000	0.0744	0.9779	1.1967
OLS	0.0720	0.9751	1.1973	0.0727	0.9765	1.1988	0.0735	0.9772	1.1964
				X5					
ML	0.0728	0.9751	1.1991	0.0746	0.9788	1.1989	0.0748	0.9792	1.1996
MML	0.0720	0.9752	1.1985	0.0757	0.9793	1.1985	0.0742	0.9793	1.1993
OLS	0.0717	0.9744	1.1981	0.0744	0.9784	1.1979	0.0737	0.9787	1.1988

Table 6.3 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-1 when $n = 100$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0610	0.9243	1.0861	0.0468	0.9292	1.1543	0.0387	0.9194	1.1777
MML	0.0603	0.9249	1.0866	0.0473	0.9293	1.1538	0.0383	0.9185	1.1776
OLS	0.0599	0.9241	1.0858	0.0466	0.9284	1.1536	0.0381	0.9184	1.1772
				X2					
ML	0.0629	0.9151	1.1008	0.0631	0.9271	1.1072	0.0701	0.9326	1.1303
MML	0.0630	0.9152	1.1009	0.0637	0.9269	1.1071	0.0706	0.9329	1.1305
OLS	0.0618	0.9149	1.1001	0.0630	0.9267	1.1064	0.0697	0.9317	1.1298
				X3					
ML	0.0715	0.9127	1.0931	0.0702	0.9162	1.0974	0.0693	0.9202	1.1051
MML	0.0713	0.9130	1.0931	0.0703	0.9155	1.0978	0.0691	0.9207	1.1049
OLS	0.0707	0.9123	1.0921	0.0697	0.9154	1.0967	0.0681	0.9200	1.1040
				X4					
ML	0.0646	0.9089	1.0891	0.0634	0.9104	1.0916	0.0723	0.9133	1.1440
MML	0.0654	0.9093	1.0900	0.0632	0.9100	1.0923	0.0719	0.9139	1.1442
OLS	0.0644	0.9086	1.0887	0.0631	0.9099	1.0911	0.0713	0.9132	1.1429
				X5					
ML	0.0654	0.9029	1.0853	0.0621	0.9067	1.0901	0.0707	0.9085	1.0800
MML	0.0657	0.9038	1.0847	0.0625	0.9058	1.0905	0.0707	0.9088	1.0810
OLS	0.0649	0.9026	1.0844	0.0613	0.9056	1.0895	0.0695	0.9077	1.0799

Table 6.4 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-1 when $n=200$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
					X1				
ML	0.0483	0.8708	1.0869	0.0505	0.8726	1.0863	0.0489	0.8720	1.0859
MML	0.0489	0.8713	1.0873	0.0497	0.8728	1.0870	0.0487	0.8718	1.0862
OLS	0.0480	0.8706	1.0860	0.0492	0.8719	1.0859	0.0482	0.8708	1.0855
					X2				
ML	0.0495	0.8907	1.1242	0.0479	0.9581	1.1951	0.0458	0.9665	1.2053
MML	0.0495	0.8912	1.1246	0.0477	0.9578	1.1939	0.0461	0.9661	1.2057
OLS	0.0487	0.8902	1.1237	0.0468	0.9574	1.1938	0.0456	0.9656	1.2047
					X3				
ML	0.0450	0.9012	1.1262	0.0406	0.9012	1.1262	0.0451	0.8999	1.1222
MML	0.0450	0.9016	1.1251	0.0411	0.9012	1.1268	0.0443	0.9006	1.1216
OLS	0.0444	0.9005	1.1249	0.0403	0.9004	1.1258	0.0439	0.8995	1.1212
					X4				
ML	0.0704	1.0935	1.6074	0.0734	1.0938	1.6073	0.0739	1.0937	1.6091
MML	0.0713	1.0924	1.6074	0.0737	1.0932	1.6083	0.0732	1.0939	1.6089
OLS	0.0700	1.0922	1.6062	0.0726	1.0929	1.6070	0.0727	1.0929	1.6085
					X5				
ML	0.0484	0.9097	1.1440	0.0515	0.9126	1.1424	0.0531	0.9152	1.1311
MML	0.0485	0.9101	1.1443	0.0518	0.9134	1.1421	0.0523	0.9155	1.1313
OLS	0.0483	0.9092	1.1431	0.0511	0.9124	1.1414	0.0521	0.9144	1.1304

Table 6.5 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-2 when $n=25$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0479	1.0399	1.2968	0.0564	1.1266	1.4096	0.0758	1.1376	1.4447
MML	0.0479	1.0396	1.2962	0.0565	1.1267	1.4101	0.0757	1.1367	1.4452
OLS	0.0474	1.0389	1.2957	0.0555	1.1260	1.4090	0.0745	1.1364	1.4440
				X2					
ML	0.0359	1.0120	1.2721	0.0374	1.0128	1.2718	0.0364	1.0234	1.2754
MML	0.0367	1.0124	1.2721	0.0381	1.0130	1.2723	0.0370	1.0233	1.2751
OLS	0.0354	1.0118	1.2712	0.0371	1.0121	1.2712	0.0362	1.0225	1.2743
				X3					
ML	0.0372	1.0119	1.2761	0.0380	1.0145	1.2766	0.0397	1.0259	1.2804
MML	0.0374	1.0121	1.2770	0.0382	1.0147	1.2756	0.0395	1.0260	1.2797
OLS	0.0363	1.0110	1.2760	0.0373	1.0139	1.2755	0.0393	1.0248	1.2792
				X4					
ML	0.0366	1.0093	1.2731	0.0373	1.0140	1.2740	0.0394	1.0158	1.2754
MML	0.0367	1.0098	1.2726	0.0382	1.0134	1.2749	0.0396	1.0152	1.2748
OLS	0.0357	1.0091	1.2724	0.0372	1.0130	1.2738	0.0391	1.0146	1.2745
				X5					
ML	0.0363	1.0106	1.2746	0.0374	1.0103	1.2744	0.0414	1.0135	1.2548
MML	0.0362	1.0106	1.2749	0.0373	1.0113	1.2752	0.0412	1.0133	1.2553
OLS	0.0356	1.0096	1.2739	0.0370	1.0100	1.2742	0.0401	1.0129	1.2544

Table 6.6 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-2 when $n = 50$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0612	0.9346	1.1674	0.0639	0.9540	1.2052	0.0638	0.9655	1.2096
MML	0.0610	0.9350	1.1671	0.0633	0.9534	1.2051	0.0633	0.9654	1.2106
OLS	0.0605	0.9339	1.1662	0.0632	0.9527	1.2041	0.0626	0.9652	1.2093
				X2					
ML	0.0612	0.9329	1.1475	0.0630	0.9343	1.1482	0.0604	0.9347	1.1524
MML	0.0616	0.9323	1.1475	0.0629	0.9336	1.1483	0.0599	0.9351	1.1525
OLS	0.0607	0.9318	1.1473	0.0619	0.9335	1.1477	0.0595	0.9338	1.1518
				X3					
ML	0.0599	0.9295	1.1434	0.0589	0.9326	1.1446	0.0631	0.9325	1.1441
MML	0.0609	0.9298	1.1426	0.0595	0.9316	1.1443	0.0631	0.9331	1.1440
OLS	0.0597	0.9293	1.1424	0.0583	0.9314	1.1441	0.0621	0.9318	1.1437
				X4					
ML	0.0592	0.9303	1.1430	0.0599	0.9294	1.1429	0.0609	0.9324	1.1428
MML	0.0596	0.9296	1.1427	0.0600	0.9296	1.1432	0.0606	0.9327	1.1424
OLS	0.0585	0.9291	1.1419	0.0591	0.9291	1.1428	0.0600	0.9317	1.1416
				X5					
ML	0.0608	0.9321	1.1433	0.0623	0.9332	1.1459	0.0617	0.9321	1.1454
MML	0.0606	0.9314	1.1434	0.0622	0.9333	1.1450	0.0620	0.9326	1.1463
OLS	0.0603	0.9311	1.1427	0.0619	0.9321	1.1447	0.0613	0.9313	1.1452

Table 6.7 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-2 when $n = 100$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0507	0.8800	1.0364	0.0361	0.8841	1.1062	0.0289	0.8759	1.1234
MML	0.0505	0.8800	1.0358	0.0371	0.8846	1.1054	0.0280	0.8758	1.1243
OLS	0.0498	0.8787	1.0352	0.0359	0.8839	1.1050	0.0277	0.8748	1.1233
				X2					
ML	0.0507	0.8714	1.0482	0.0526	0.8873	1.0574	0.0606	0.8877	1.0811
MML	0.0500	0.8719	1.0483	0.0527	0.8877	1.0578	0.0607	0.8875	1.0821
OLS	0.0498	0.8710	1.0477	0.0521	0.8870	1.0565	0.0596	0.8872	1.0810
				X3					
ML	0.0601	0.8699	1.0430	0.0584	0.8739	1.0490	0.0570	0.8794	1.0534
MML	0.0594	0.8694	1.0428	0.0594	0.8731	1.0495	0.0577	0.8798	1.0531
OLS	0.0590	0.8690	1.0419	0.0581	0.8727	1.0489	0.0565	0.8788	1.0528
				X4					
ML	0.0535	0.8648	1.0385	0.0558	0.8672	1.0420	0.0600	0.8722	1.0917
MML	0.0540	0.8652	1.0376	0.0556	0.8677	1.0418	0.0603	0.8722	1.0926
OLS	0.0531	0.8645	1.0373	0.0551	0.8667	1.0408	0.0591	0.8715	1.0914
				X5					
ML	0.0539	0.8603	1.0357	0.0549	0.8660	1.0384	0.0597	0.8684	1.0333
MML	0.0544	0.8599	1.0355	0.0544	0.8653	1.0388	0.0594	0.8687	1.0337
OLS	0.0536	0.8595	1.0353	0.0536	0.8651	1.0378	0.0592	0.8678	1.0326

Table 6.8 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-2 when $n = 200$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0384	0.8294	1.0354	0.0386	0.8305	1.0355	0.0375	0.8277	1.0350
MML	0.0386	0.8290	1.0356	0.0382	0.8308	1.0363	0.0378	0.8288	1.0349
OLS	0.0380	0.8286	1.0346	0.0379	0.8296	1.0353	0.0373	0.8276	1.0346
				X2					
ML	0.0386	0.8507	1.0711	0.0358	0.9163	1.1394	0.0351	0.9204	1.1497
MML	0.0380	0.8512	1.0706	0.0360	0.9173	1.1397	0.0344	0.9197	1.1499
OLS	0.0374	0.8504	1.0703	0.0351	0.9161	1.1387	0.0342	0.9192	1.1490
				X3					
ML	0.0334	0.8568	1.0740	0.0330	0.8565	1.0739	0.0341	0.8562	1.0754
MML	0.0334	0.8565	1.0743	0.0330	0.8577	1.0734	0.0341	0.8558	1.0754
OLS	0.0328	0.8561	1.0730	0.0324	0.8564	1.0728	0.0331	0.8555	1.0745
				X4					
ML	0.0612	1.0416	1.5384	0.0614	1.0426	1.5381	0.0624	1.0420	1.5396
MML	0.0612	1.0421	1.5380	0.0620	1.0430	1.5372	0.0630	1.0426	1.5388
OLS	0.0602	1.0409	1.5371	0.0613	1.0425	1.5369	0.0618	1.0413	1.5383
				X5					
ML	0.0389	0.8684	1.0908	0.0418	0.8593	1.0915	0.0420	0.8721	1.0789
MML	0.0388	0.8681	1.0914	0.0418	0.8704	1.0917	0.0416	0.8716	1.0779
OLS	0.0381	0.8672	1.0905	0.0406	0.8691	1.0905	0.0408	0.8710	1.0778

Table 6.9 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-3 when $n=25$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0335	1.0063	1.2669	0.0474	1.0990	1.3791	0.0636	1.1128	1.4138
MML	0.0327	0.9970	1.2566	0.0450	1.0981	1.3648	0.0632	1.1071	1.4105
OLS	0.0418	1.0218	1.2779	0.0499	1.1083	1.3868	0.0687	1.1203	1.4222
				X2					
ML	0.0308	0.9937	1.2508	0.0316	0.9947	1.2517	0.0323	1.0045	1.2533
MML	0.0295	0.9924	1.2495	0.0314	0.9946	1.2515	0.0321	1.0043	1.2532
OLS	0.0317	0.9946	1.2517	0.0323	0.9954	1.2524	0.0331	1.0053	1.2541
				X3					
ML	0.0315	0.9946	1.2546	0.0322	0.9957	1.2558	0.0338	1.0069	1.2583
MML	0.0308	0.9939	1.2538	0.0319	0.9954	1.2555	0.0337	1.0069	1.2583
OLS	0.0319	0.9950	1.2550	0.0326	0.9961	1.2562	0.0343	1.0074	1.2588
				X4					
ML	0.0312	0.9921	1.2523	0.0321	0.9951	1.2531	0.0333	0.9964	1.2539
MML	0.0304	0.9914	1.2516	0.0318	0.9948	1.2528	0.0333	0.9964	1.2531
OLS	0.0316	0.9925	1.2527	0.0325	0.9955	1.2535	0.0338	0.9969	1.2545
				X5					
ML	0.0312	0.9923	1.2524	0.0320	0.9932	1.2532	0.0339	0.9946	1.2336
MML	0.0305	0.9922	1.2517	0.0317	0.9929	1.2529	0.0338	0.9946	1.2332
OLS	0.0316	0.9926	1.2528	0.0325	0.9936	1.2536	0.0344	0.9952	1.2339

Table 6.10 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-3 when $n = 50$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
					X1				
ML	0.0541	0.9169	1.1433	0.0576	0.9352	1.1812	0.0586	0.9474	1.1887
MML	0.0540	0.9165	1.1398	0.0565	0.9350	1.1804	0.0582	0.9465	1.1876
OLS	0.0544	0.9183	1.1472	0.0590	0.9369	1.1892	0.0592	0.9488	1.1906
					X2				
ML	0.0538	0.9143	1.1242	0.0534	0.9152	1.1269	0.0534	0.9163	1.1306
MML	0.0531	0.9136	1.1216	0.0528	0.9141	1.1233	0.0529	0.9151	1.1301
OLS	0.0552	0.9157	1.1292	0.0557	0.9173	1.1292	0.0544	0.9190	1.1350
					X3				
ML	0.0536	0.9131	1.1229	0.0543	0.9141	1.1243	0.0546	0.9139	1.1251
MML	0.0529	0.9113	1.1222	0.0542	0.9138	1.1230	0.0545	0.9131	1.1241
OLS	0.0542	0.9137	1.1235	0.0558	0.9154	1.1256	0.0564	0.9160	1.1271
					X4				
ML	0.0536	0.9135	1.1234	0.0547	0.9143	1.1247	0.0558	0.9153	1.1214
MML	0.0540	0.9139	1.1238	0.0554	0.9150	1.1254	0.0568	0.9164	1.1207
OLS	0.0529	0.9127	1.1227	0.0533	0.9129	1.1233	0.0537	0.9132	1.1228
					X5				
ML	0.0538	0.9134	1.1239	0.0553	0.9141	1.1239	0.0547	0.9147	1.1246
MML	0.0532	0.9132	1.1234	0.0547	0.9134	1.1235	0.0546	0.9143	1.1242
OLS	0.0545	0.9145	1.1244	0.0561	0.9154	1.1252	0.0567	0.9168	1.1267

Table 6.11 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-3 when $n = 100$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0439	0.8624	1.0171	0.0320	0.8676	1.0863	0.0222	0.8594	1.1041
MML	0.0431	0.8618	1.0168	0.0317	0.8673	1.0859	0.0219	0.8591	1.1038
OLS	0.0446	0.8629	1.0174	0.0323	0.8679	1.0866	0.0225	0.8596	1.1043
				X2					
ML	0.0436	0.8546	1.0291	0.0452	0.8675	1.0362	0.0526	0.8686	1.0526
MML	0.0429	0.8541	1.0282	0.0442	0.8631	1.0329	0.0482	0.8638	1.0413
OLS	0.0442	0.8550	1.0299	0.0462	0.8719	1.0394	0.0569	0.8735	1.0639
				X3					
ML	0.0490	0.8504	1.0216	0.0492	0.8555	1.0275	0.0472	0.8603	1.0303
MML	0.0451	0.8470	1.0181	0.0450	0.8511	1.0235	0.0429	0.8553	1.0251
OLS	0.0529	0.8538	1.0251	0.0534	0.8599	1.0315	0.0514	0.8652	1.0346
				X4					
ML	0.0453	0.8475	1.0185	0.0478	0.8501	1.0213	0.0508	0.8533	1.0697
MML	0.0434	0.8456	1.0166	0.0448	0.8471	1.0183	0.0465	0.8490	1.0655
OLS	0.0472	0.8493	1.0203	0.0508	0.8531	1.0243	0.0550	0.8576	1.0740
				X5					
ML	0.0462	0.8441	1.0170	0.0466	0.8469	1.0197	0.0493	0.8499	1.0125
MML	0.0436	0.8428	1.0157	0.0441	0.8443	1.0171	0.0455	0.8461	1.0087
OLS	0.0488	0.8455	1.0184	0.0492	0.8494	1.0223	0.0531	0.8536	1.0162

Table 6.12 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-3 when $n = 200$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
					X1				
ML	0.0327	0.8138	1.0169	0.0322	0.8143	1.0173	0.0321	0.8139	1.0171
MML	0.0321	0.8135	1.0166	0.0319	0.8138	1.0168	0.0319	0.8136	1.0168
OLS	0.0332	0.8142	1.0172	0.0324	0.8149	1.0177	0.0324	0.8141	1.0173
					X2				
ML	0.0322	0.8346	1.0526	0.0293	0.8990	1.1215	0.0285	0.9023	1.1293
MML	0.0321	0.8343	1.0523	0.0287	0.8985	1.1208	0.0282	0.9021	1.1290
OLS	0.0323	0.8349	1.0529	0.0298	0.8995	1.1223	0.0288	0.9026	1.1295
					X3				
ML	0.0276	0.8400	1.0549	0.0281	0.8403	1.0553	0.0282	0.8412	1.0557
MML	0.0273	0.8395	1.0545	0.0278	0.8400	1.0551	0.0279	0.8406	1.0555
OLS	0.0279	0.8404	1.0554	0.0285	0.8407	1.0556	0.0286	0.8418	1.0560
					X4				
ML	0.0544	1.0236	1.5132	0.0548	1.0241	1.5138	0.0556	1.0245	1.5141
MML	0.0543	1.0235	1.5132	0.0546	1.0238	1.5136	0.0550	1.0241	1.5138
OLS	0.0544	1.0237	1.5132	0.0551	1.0243	1.5141	0.0562	1.0250	1.5143
					X5				
ML	0.0336	0.8524	1.0715	0.0345	0.8536	1.0727	0.0347	0.8546	1.0612
MML	0.0333	0.8522	1.0713	0.0343	0.8532	1.0723	0.0343	0.8541	1.0608
OLS	0.0338	0.8526	1.0717	0.0355	0.8540	1.0731	0.0352	0.8555	1.0618

Table 6.13 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-4 when $n=25$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0390	1.0109	1.2717	0.0548	1.1067	1.3859	0.0723	1.1195	1.4201
MML	0.0363	0.9985	1.2664	0.0515	1.1049	1.3848	0.0696	1.1174	1.4142
OLS	0.0518	1.0304	1.2816	0.0598	1.1160	1.3965	0.0785	1.1245	1.4303
				X2					
ML	0.0358	1.0039	1.2534	0.0342	0.9999	1.2565	0.0351	1.0134	1.2594
MML	0.0349	1.0012	1.2531	0.0333	0.9994	1.2562	0.0345	1.0083	1.2574
OLS	0.0383	1.0041	1.2595	0.0410	1.0039	1.2588	0.0384	1.0146	1.2632
				X3					
ML	0.0391	0.9983	1.2586	0.0380	1.0035	1.2605	0.0414	1.0165	1.2627
MML	0.0387	0.9965	1.2585	0.0353	1.0012	1.2594	0.0411	1.0144	1.2611
OLS	0.0399	1.0022	1.2656	0.0411	1.0067	1.2634	0.0434	1.0170	1.2681
				X4					
ML	0.0349	0.9995	1.2575	0.0341	1.0023	1.2582	0.0394	1.0045	1.2577
MML	0.0331	0.9968	1.2568	0.0336	1.0001	1.2552	0.0368	1.0022	1.2570
OLS	0.0390	1.0009	1.2608	0.0410	1.0055	1.2628	0.0435	1.0074	1.2632
				X5					
ML	0.0372	0.9976	1.2591	0.0354	0.9969	1.2621	0.0410	1.0003	1.2384
MML	0.0369	0.9936	1.2562	0.0348	0.9954	1.2615	0.0389	0.9999	1.2376
OLS	0.0387	1.0018	1.2637	0.0406	1.0016	1.2634	0.0448	1.0056	1.2441

Table 6.14 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-4 when $n=50$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
					X1				
ML	0.0605	0.9231	1.1447	0.0652	0.9426	1.1836	0.0633	0.9544	1.1964
MML	0.0565	0.9228	1.1443	0.0636	0.9394	1.1829	0.0617	0.9523	1.1911
OLS	0.0651	0.9266	1.1565	0.0658	0.9451	1.1893	0.0644	0.9579	1.1983
					X2				
ML	0.0624	0.9205	1.1355	0.0602	0.9230	1.1351	0.0616	0.9234	1.1342
MML	0.0588	0.9202	1.1332	0.0593	0.9207	1.1328	0.0606	0.9221	1.1338
OLS	0.0648	0.9249	1.1372	0.0665	0.9268	1.1380	0.0632	0.9257	1.1403
					X3				
ML	0.0592	0.9163	1.1311	0.0592	0.9198	1.1336	0.0620	0.9231	1.1284
MML	0.0556	0.9162	1.1303	0.0578	0.9160	1.1280	0.0604	0.9199	1.1281
OLS	0.0637	0.9221	1.1332	0.0594	0.9244	1.1344	0.0663	0.9247	1.1321
					X4				
ML	0.0607	0.9186	1.1308	0.0607	0.9201	1.1311	0.0623	0.9221	1.1311
MML	0.0575	0.9184	1.1301	0.0596	0.9169	1.1273	0.0610	0.9212	1.1311
OLS	0.0626	0.9226	1.1330	0.0634	0.9225	1.1341	0.0648	0.9273	1.1323
					X5				
ML	0.0616	0.9236	1.1305	0.0621	0.9186	1.1313	0.0578	0.9213	1.1318
MML	0.0611	0.9219	1.1294	0.0577	0.9181	1.1297	0.0573	0.9207	1.1299
OLS	0.0646	0.9247	1.1329	0.0662	0.9259	1.1360	0.0643	0.9229	1.1355

Table 6.15 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-4 when $n = 100$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0518	0.8693	1.0266	0.0383	0.8747	1.0931	0.0290	0.8647	1.1130
MML	0.0490	0.8673	1.0236	0.0373	0.8736	1.0911	0.0251	0.8628	1.1087
OLS	0.0538	0.8728	1.0275	0.0387	0.8781	1.0963	0.0323	0.8684	1.1147
				X2					
ML	0.0512	0.8639	1.0363	0.0521	0.8726	1.0426	0.0587	0.8739	1.0621
MML	0.0460	0.8630	1.0345	0.0482	0.8693	1.0399	0.0564	0.8733	1.0511
OLS	0.0541	0.8656	1.0397	0.0567	0.8802	1.0476	0.0609	0.8790	1.0715
				X3					
ML	0.0530	0.8574	1.0292	0.0510	0.8630	1.0352	0.0539	0.8658	1.0354
MML	0.0518	0.8537	1.0211	0.0505	0.8525	1.0265	0.0486	0.8646	1.0315
OLS	0.0636	0.8628	1.0330	0.0613	0.8640	1.0405	0.0602	0.8707	1.0451
				X4					
ML	0.0511	0.8565	1.0240	0.0539	0.8571	1.0241	0.0564	0.8622	1.0757
MML	0.0500	0.8534	1.0222	0.0512	0.8551	1.0225	0.0502	0.8527	1.0686
OLS	0.0576	0.8585	1.0288	0.0581	0.8589	1.0316	0.0618	0.8639	1.0819
				X5					
ML	0.0522	0.8494	1.0186	0.0502	0.8543	1.0244	0.0538	0.8535	1.0171
MML	0.0512	0.8455	1.0184	0.0492	0.8541	1.0213	0.0521	0.8526	1.0150
OLS	0.0571	0.8524	1.0267	0.0567	0.8596	1.0277	0.0638	0.8606	1.0235

Table 6.16 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-4 when $n=200$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0385	0.8192	1.0232	0.0374	0.8225	1.0257	0.0388	0.8173	1.0237
MML	0.0365	0.8160	1.0215	0.0367	0.8209	1.0235	0.0388	0.8165	1.0234
OLS	0.0418	0.8226	1.0265	0.0425	0.8239	1.0274	0.0412	0.8207	1.0264
				X2					
ML	0.0364	0.8428	1.0586	0.0355	0.9064	1.1244	0.0334	0.9105	1.1337
MML	0.0362	0.8422	1.0577	0.0333	0.9052	1.1243	0.0328	0.9074	1.1324
OLS	0.0416	0.8449	1.0614	0.0396	0.9102	1.1270	0.0387	0.9132	1.1403
				X3					
ML	0.0306	0.8455	1.0591	0.0333	0.8426	1.0580	0.0332	0.8451	1.0640
MML	0.0303	0.8443	1.0589	0.0319	0.8412	1.0579	0.0321	0.8438	1.0632
OLS	0.0368	0.8508	1.0643	0.0355	0.8511	1.0637	0.0367	0.8482	1.0665
				X4					
ML	0.0640	1.0265	1.5202	0.0599	1.0323	1.5198	0.0597	1.0299	1.5227
MML	0.0615	1.0262	1.5171	0.0596	1.0276	1.5192	0.0585	1.0275	1.5170
OLS	0.0645	1.0324	1.5231	0.0660	1.0351	1.5218	0.0658	1.0320	1.5244
				X5					
ML	0.0386	0.8567	1.0759	0.0408	0.8579	1.0794	0.0375	0.8597	1.0637
MML	0.0370	0.8556	1.0729	0.0364	0.8572	1.0765	0.0368	0.8585	1.0637
OLS	0.0415	0.8604	1.0825	0.0448	0.8629	1.0811	0.0453	0.8651	1.0673

Table 6.17 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-5 when $n=25$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0493	1.0272	1.2808	0.0616	1.1111	1.3923	0.0776	1.1235	1.4236
MML	0.0379	1.0063	1.2727	0.0594	1.1083	1.3850	0.0689	1.1173	1.4205
OLS	0.0533	1.0408	1.2844	0.0672	1.1141	1.3967	0.0844	1.1399	1.4351
				X2					
ML	0.0394	0.9995	1.2613	0.0453	1.0082	1.2695	0.0518	1.0224	1.2665
MML	0.0381	0.9986	1.2570	0.0375	1.0069	1.2653	0.0462	1.0164	1.2659
OLS	0.0489	1.0027	1.2711	0.0505	1.0126	1.2726	0.0529	1.0255	1.2713
				X3					
ML	0.0404	1.0054	1.2632	0.0405	1.0026	1.2643	0.0435	1.0227	1.2691
MML	0.0361	1.0054	1.2588	0.0372	1.0024	1.2614	0.0404	1.0162	1.2657
OLS	0.0476	1.0090	1.2747	0.0511	1.0152	1.2765	0.0478	1.0268	1.2794
				X4					
ML	0.0455	1.0028	1.2623	0.0515	1.0072	1.2706	0.0444	1.0064	1.2675
MML	0.0432	1.0001	1.2605	0.0512	1.0056	1.2629	0.0428	1.0051	1.2617
OLS	0.0506	1.0114	1.2725	0.0526	1.0115	1.2731	0.0532	1.0107	1.2718
				X5					
ML	0.0480	1.0098	1.2665	0.0422	1.0055	1.2693	0.0487	1.0020	1.2520
MML	0.0471	1.0011	1.2659	0.0406	1.0007	1.2608	0.0404	1.0008	1.2437
OLS	0.0517	1.0126	1.2737	0.0519	1.0132	1.2716	0.0543	1.0071	1.2534

Table 6.18 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-5 when $n=50$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0665	0.9348	1.1571	0.0711	0.9426	1.1893	0.0733	0.9568	1.2043
MML	0.0644	0.9323	1.1449	0.0669	0.9420	1.1878	0.0731	0.9534	1.1981
OLS	0.0729	0.9373	1.1708	0.0730	0.9520	1.1986	0.0770	0.9613	1.2071
				X2					
ML	0.0611	0.9250	1.1431	0.0666	0.9272	1.1442	0.0663	0.9339	1.1439
MML	0.0600	0.9220	1.1345	0.0641	0.9226	1.1372	0.0624	0.9302	1.1355
OLS	0.0637	0.9296	1.1467	0.0740	0.9316	1.1473	0.0685	0.9351	1.1503
				X3					
ML	0.0630	0.9206	1.1352	0.0671	0.9271	1.1410	0.0632	0.9299	1.1309
MML	0.0626	0.9201	1.1283	0.0608	0.9239	1.1367	0.0623	0.9245	1.1308
OLS	0.0704	0.9301	1.1414	0.0720	0.9324	1.1435	0.0714	0.9347	1.1391
				X4					
ML	0.0691	0.9263	1.1362	0.0677	0.9302	1.1343	0.0619	0.9273	1.1337
MML	0.0612	0.9211	1.1336	0.0663	0.9291	1.1336	0.0611	0.9266	1.1302
OLS	0.0742	0.9300	1.1419	0.0748	0.9329	1.1437	0.0748	0.9294	1.1408
				X5					
ML	0.0580	0.9223	1.1310	0.0614	0.9288	1.1339	0.0665	0.9261	1.1392
MML	0.0578	0.9207	1.1299	0.0612	0.9230	1.1298	0.0617	0.9240	1.1330
OLS	0.0716	0.9266	1.1435	0.0752	0.9330	1.1400	0.0758	0.9366	1.1422

Table 6.19 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-5 when $n = 100$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0584	0.8827	1.0265	0.0470	0.8775	1.0906	0.0288	0.8705	1.1192
MML	0.0554	0.8738	1.0248	0.0433	0.8734	1.0899	0.0268	0.8696	1.1176
OLS	0.0601	0.8829	1.0356	0.0499	0.8859	1.0955	0.0400	0.8766	1.1219
				X2					
ML	0.0555	0.8690	1.0434	0.0531	0.8788	1.0486	0.0642	0.8825	1.0670
MML	0.0481	0.8664	1.0434	0.0525	0.8772	1.0427	0.0601	0.8776	1.0557
OLS	0.0633	0.8727	1.0504	0.0630	0.8806	1.0545	0.0715	0.8913	1.0752
				X3					
ML	0.0596	0.8601	1.0377	0.0644	0.8682	1.0399	0.0572	0.8718	1.0403
MML	0.0540	0.8569	1.0336	0.0535	0.8650	1.0335	0.0479	0.8692	1.0369
OLS	0.0708	0.8706	1.0420	0.0712	0.8752	1.0432	0.0691	0.8773	1.0525
				X4					
ML	0.0536	0.8660	1.0343	0.0599	0.8624	1.0363	0.0689	0.8651	1.0811
MML	0.0535	0.8638	1.0339	0.0588	0.8597	1.0300	0.0663	0.8625	1.0751
OLS	0.0647	0.8679	1.0397	0.0617	0.8699	1.0414	0.0737	0.8712	1.0897
				X5					
ML	0.0583	0.8596	1.0295	0.0576	0.8584	1.0372	0.0593	0.8586	1.0220
MML	0.0525	0.8520	1.0219	0.0513	0.8565	1.0339	0.0568	0.8528	1.0217
OLS	0.0663	0.8625	1.0336	0.0597	0.8610	1.0423	0.0683	0.8640	1.0283

Table 6.20 The Average of ME, MAE and RMSE Using Different Schemes, Different Methods for Model-5 when $n = 200$

Method	Scheme								
	Recursive			Rolling			Fixed		
	ME	MAE	RMSE	ME	MAE	RMSE	ME	MAE	RMSE
				X1					
ML	0.0447	0.8303	1.0307	0.0396	0.8257	1.0327	0.0470	0.8254	1.0333
MML	0.0420	0.8298	1.0266	0.0382	0.8237	1.0306	0.0469	0.8210	1.0263
OLS	0.0494	0.8315	1.0369	0.0510	0.8328	1.0358	0.0504	0.8339	1.0360
				X2					
ML	0.0446	0.8453	1.0595	0.0372	0.9068	1.1341	0.0448	0.9148	1.0445
MML	0.0425	0.8405	1.0579	0.0363	0.9049	1.1321	0.0419	0.9078	1.0427
OLS	0.0510	0.8464	1.0736	0.0493	0.9089	1.1413	0.0480	0.9214	1.0486
				X3					
ML	0.0388	0.8542	1.0667	0.0356	0.8527	1.0694	0.0389	0.8589	1.0598
MML	0.0373	0.8527	1.0655	0.0334	0.8514	1.0693	0.0370	0.8557	1.0594
OLS	0.0475	0.8602	1.0731	0.0410	0.8597	1.0753	0.0468	0.8609	1.0637
				X4					
ML	0.0640	1.0363	1.0209	0.0681	1.0306	1.0263	0.0709	1.0324	1.0292
MML	0.0631	1.0314	1.0196	0.0612	1.0284	1.0213	0.0664	1.0317	1.0271
OLS	0.0685	1.0428	1.0286	0.0719	1.0414	1.0315	0.0724	1.0445	1.0317
				X5					
ML	0.0469	0.8604	1.0854	0.0416	0.8614	1.0849	0.0516	0.8607	1.0782
MML	0.0456	0.8579	1.0790	0.0375	0.8611	1.0840	0.0421	0.8598	1.0767
OLS	0.0503	0.8671	1.0894	0.0522	0.8707	1.0875	0.0536	0.8722	1.0804

CHAPTER 7

Summary and Concluding Remarks

7.1 Introduction

This thesis has investigated four important issues. The first was to develop a small-sample procedure for calculating critical values of the LR test statistic for testing a changepoint of unknown timing in the linear regression model. The second was to investigate the use of IC model selection criteria for detecting a changepoint and to find which criteria among some existing IC has the best ability to detect a changepoint in the context of a linear regression model when the timing of the changepoint is unknown. We used the average mean probability of correct selection (AMPCS) criterion as a measure of the ability to detect a changepoint. The third was to find a suitable algorithm for estimating appropriate optimal penalties that saves computing time and at the same time, giving penalties which outperform all existing IC model selection procedures when looking for a changepoint of unknown timing. This approach involves finding penalties that maximize the AMPCS.

Finally, this thesis investigated the distributional pattern of a linear model with random changes in the parameter with low probability, derived the distribution

of the changing parameter model, developed the variance-covariance matrices for three special situations, and constructed 'out of sample' forecast procedures.

In the section that follows, we give a detailed discussion of the results and contributions of this thesis. We conclude this chapter by giving aspects related to, but not covered in this thesis, with some suggestions for potential topics for future research.

7. 2 Summary of the Thesis

Chapter 2 reviewed three topics with particular emphasis on the problem of a structural change. The survey began with a brief historical review of the structural change literature particularly on hypothesis testing from a statistical and econometric viewpoint. The survey revealed that there is a large body of literature on this problem, with procedures ranging from non-parametric to classical and to Bayesian methods.

The literature survey conducted in Chapter 2 highlighted the fact that hypothesis tests are not always the most advantageous way to choose the best-specified model. Rather, it was argued following Granger et al. (1995) that, model selection decisions should be based on some well-thought-out model selection criteria rather than such classical mechanisms. This provides a justification for working with information criteria based model selection procedures when faced with a choice of model to be decided using the available data.

Chapter 2 showed that there are not many comprehensive studies that have evaluated the relatively small sample performance of various existing IC model selection procedures. The majority of the research in this area has been related to asymptotic properties and Monte Carlo studies have mainly been used to illustrate

the asymptotic results. This led to the identification of the need to develop model selection procedures for choosing between different possible models with structural change of different timing.

In Chapter 3 we examined the use of the LR test to test for the presence of structural change when there is a possible unknown changepoint in the data. Since this test does not have a known distribution for finite sample sizes, we calculated exact critical values for the test by simulation using 10000 replications for different sample sizes, numbers of regressors and types of regressors. We found that the critical value clearly depends on sample size, number of regressors and to a lesser extend on the type of explanatory variables. We found that the calculation of critical values via simulation can be very time consuming. To overcome this difficulty, we developed formulae for critical values using a response surface approach. This avoids the use of a table at a desired level of significance when the sample size and the number of regressors are known. We checked the accuracy of the critical value formulae by using the Monte Carlo method and found that the estimated sizes based on the new formulae are not significantly different from nominal size regardless of the sample size. Overall the actual sizes of the test using our formulae to calculate appropriate critical values are quite satisfactory.

In Chapter 4, we investigated the relative performance of IC model selection procedures when detecting the possible presence of a structural change. We found BIC outperformed all existing IC procedures considered when there is no structural change but its performance is the worst of all procedures in the presence of structural change. RSC's performance is the worst of all existing IC procedures in the presence of no structural change however it outperformed all other IC procedures considered when there is structural change. In presence of no structural chance, the ranking of the relative performance of the other IC procedures is HSPC > GCVC > HQC >

MCPC > AIC. On the other hand, in the presence of structural change, the ranking of the relative performance of the other IC procedures is AIC > MCPC > HQC > GCVC > HSPC. None of the IC procedures considered stands out as a clear best method for modelling involving structural change. We conclude from the results that model selection can be useful when there are a large number of models involved.

In Chapter 5, we proposed new methods for finding optimal penalties for different models while detecting possible structural change through model selection procedures by maximizing AMPCS. Our method includes a family of procedures, based on grid search algorithms such as the CGSA, BGSA and PDFA, and the SAA. These procedures do not require conditions such as regularity or existence of derivatives. The grid search algorithm is one of the appealing ways to maximize a function. The disadvantage of this procedure is that it is very time consuming when there is a reasonably large number of grid points for the penalty vector. The computational time of grid search algorithms will increase dramatically with an increase in the number of penalties and so can be exceptionally excessive. To overcome the computational limits imposed by grid search algorithms, we estimated the optimum penalties using SAA whose performance is similar to that of grid search algorithms while its computational time is much lower. The simulation results show that the CGSA involves heavy computation giving the highest percentage gain over all IC procedures.

We investigated the use of two alternative approaches to CGSA, namely BGSA and PDFA. These are straight forward and save computational time. The former gives higher AMPCS than the latter two. However, the computational cost is lower for the latter two. We have found that CGSA is the best, BGSA second, SAA third and PDFA fourth best as measured by maximum AMPCS. In the context of computational time, the rankings are SAA first, PDFA second, BGSA third and

CGSA last. We have found that all of our four suggested procedures dominate the existing IC procedures considered in terms of maximizing AMPCS.

In Chapter 6, we investigated random changes in the coefficients of linear regression models and their effect on predictions. We derived the distributional pattern especially the mean, variance and covariance structure of different linear regression models for stochastic changes in either slope or intercept parameters in turn by a fixed amount with a very low probability. We find that this results in a linear regression with a nonscalar variance-covariance matrix, which allows standard approaches to estimation and prediction to be used.

We compared the predictive performances for three methods of estimation of parameters, three different schemes, four different sample sizes and five different models in turn and evaluated the forecasting performance of the estimators using RMSE, MAE and ME. The simulation results suggest that the MML estimator is clearly better than the ML estimation in terms of small sample properties. In addition, the MML estimator is quite promising in terms of ME, MAE and RMSE.

This study has some limitations in that one has to take care in making too many generalized conclusions from the results based on Monte Carlo experiments. The results achieved by Monte Carlo experiments are often specific to the design of the experiment.

7. 3 Future Work

Given the encouraging results of our research on LR test statistics, model selection procedures and penalty estimation, they are indeed worthy of further attention and work. A few suggestions for future research are outlined below.

In Chapter 3 we examined the use of the LR test to test for the presence of structural change when there is a possible unknown changepoint in the data and gave critical values of the LR test empirically. An obvious extension is to testing changepoints in more than one coefficient and using the same approach to find critical values. Comparison of empirical sizes and powers of the test can be done with other tests.

In this thesis, testing for structural change is confined to the regression coefficients. We could relax the assumption that the variances of two the subsamples are the same and let the disturbance variances vary among the different subsamples and then investigate robustness of a structural change test to the presence of heteroscedasticity or to develop same new tests that account for this situation. Also further work could be done to develop the corresponding tests for structural change in dynamic and simultaneous equation models.

Relaxing the assumption that the structural change occurs only once, other possible extensions to the test include the possibility of multiple unknown changepoints, a two-sided testing problem, or testing for stability of a vector (as opposed to a scalar) of coefficients.

Our proposed small sample model selection procedures can be applied to a number of other model selection problems which have not been considered in this thesis such as heteroscedasticity and error component regression error models. The error distribution of the models in this thesis was assumed to be normal. Further

research should be conducted on the robustness of these methods in the presence of nonnormal errors.

A mathematical derivation of the penalty function for different models may be of great interest in order to cut computational time for the application of the model. There is perhaps some scope for further studies regarding the factors responsible for coefficient changes, testing the significance and constructing the confidence intervals of the changing parameters, as well as construction of prediction intervals. It would also be interesting to investigate the forecast performance for different lead times (other than one-period-ahead) for different models. As one can see, there are many directions for further research.

References

Aarts, E.H.L. and J.H.M. Korst, 1989, *Simulated Annealing and Boltzmann Machines*, Wiley (Interscience), New York.

Akaike, H., 1969, Fitting autoregressive models for prediction, *Annals of the Institute of Statistical Mathematics* 21, 243-247.

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle in B.N. Petrov and F. Csaki eds, *Second International Symposium on Information Theory, Akademiai Kiado*, Budapest, 267-281, Also in S. Kotz. and N.L. Johnson eds., Breakthroughs in Statistics Volume 1, *Foundations and Basic Theory*, Springer-Verlag, New York, 601-609.

Akaike, H., 1974, A new look at the statistical model identification, *IEEE, Transactions on Automatic Control* 19, 716-723.

Akaike, H., 1978, A Bayesian analysis of the minimum AIC procedure, *Annals of the Institute of Statistical Mathematics* 30, 9-14.

Akaike, H., 1981, Likelihood of a model and information criteria, *Journal of Econometrics* 16, 3-14.

Akgiray, V., 1989, Conditional heteroskedasticity in time series of stock return: Evidence and forecasts, *Journal of Business* 62, 55-80.

Ali, M.M. and J.L. Silver, 1985, Tests for equality between sets of coefficients in two linear regressions under heteroscedasticity, *Journal of the American Statistical Association* 80, 730-735.

Ali, M., A. Torn and S. Viitanen, 1997, A direct search simulated annealing algorithm for optimization involving continuous variables, TUCS Technical Report 97, Turku Centre for Computer Science Finland.

Allen, D.M., 1971, The prediction sum of squares as a criterion for selecting predictor variables, Technical Report 23, Department of Statistics University of Kentucky, Lexington, KY.

Allen, D.M., 1974, The relationship between variable selection and data augmentation and a method for prediction, *Technometrics* 16, 125-127.

Amemiya, T., 1972, Lecture notes on econometrics, Department of Economics, Stanford University.

Amemiya, T., 1980, Selection of regressors, *International Economic Review* 21, 331-354.

Amemiya, T., 1985, *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.

Andrews, D.W.K., 1990, Testing for parameter instability and structural change with unknown change point, Cowles Foundation Discussion Paper 943, Yale University.

Andrews, D.W.K., 1991, Heteroscedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817-858.

Andrews, D.W.K., 1993, Tests for parameter instability and structural change with unknown change point, *Econometrica* 61, 821-856.

Andrews, D.W.K. and R.C. Fair, 1988, Inference in nonlinear econometric models with structural change, *The Review of Economic Studies* 55, 615-639.

Andrews, D.W.K., I. Lee and W. Ploberger, 1996, Optimal changepoint tests for normal regression, *Journal of Econometrics* 70, 9-38.

Andrews, D.W.K. and W. Ploberger, 1994, Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* 62, 1383-1414.

Antoch, J. and M. Huskova, 1989, Some M-tests for detection of a change in linear models, In: *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics* Charles University, Prague, 123-136.

Antoch, J. and M. Huskova, 1993, Changepoint problem, In: *Computational Aspects of Model Choice,* Physica-Verlag, Heidelberg, 11-37.

Antoch, J., M. Huskova and Z. Praskova, 1997, Effect of dependency on statistics for determination of change, *Journal of Statistical Planning and Inference* 60, 291-310.

Antoch, J., M. Huskova and N. Veraverbeke, 1995, Changepoint problem and bootstrap, *Journal of Nonparametric Statistics* 5, 123-144.

Ara, I. and M.L. King, 1993, Marginal likelihood based tests of regression disturbances, mimeo, Department of Econometrics and Business Statistics, Monash University, Australia.

Ara, I., 1995, Marginal likelihood based tests of regression disturbances, unpublished Ph.D. thesis, Department of Econometrics and Business Statistics Monash University, Australia.

Atkinson, A.C., 1980, A note on the generalized information criterion for choice of a model, *Biometrika* 67, 413-418.

Atkinson, A.C. and V.V. Fedorov, 1975, The design of experiments for discriminating between two rival models, *Biometrika* 62, 289-304.

Azam, M.N. and M.L. King, 1997, Model selection in presence of structural change when the changepoint is unknown, *Proceedings of the Third Annual Doctoral Research Conference,* Faculty of Business and Economics, Monash University, Australia.

Azam, M.N. and M.L. King, 1998, Model selection with data driven penalty when there is an unknown changepoint in the data, in C.L. Skeels ed., *Proceedings of the Econometric Society Australasian Meeting,* CD Rom, Australian National University, Canberra, Australia.

Bacon, D.W. and D.G. Watts, 1971, Estimating the transition between two intersecting straight lines, *Biometrika* 58, 525-534.

Bagshaw, M. and R.A. Johnson, 1975, The effect of serial correlation on the performance of CUSUM tests, *Technometrics* 17, 73-80.

Bai, J., 1994a, Weak convergence of the sequential empirical processes of residuals in ARMA models, *Annals of Statistics* 22, 2051-2061.

Bai, J., 1994b, Least squares estimation of a shift in linear processes, *Time Series Analysis* 15, 453-472.

Baron, M. and A.L. Rukhin, 1997, Asymptotic behavior of confidence regions in the changepoint problem, *Journal of Statistical Planning and Inference* 58, 263-282.

Belsley, D.A., E. Kuh and R.E. Welsch, 1980, *Regression Diagnostics, Identifying Influential Data and Sources of Collinearity,* Wiley, New York.

Bergstrom, A.R., 1993, What is econometrics? In P.C.B. Phillips ed., *Models, Methods, and Applications of Econometrics,* Basil Blackwell, Cambridge, Massachusetts, 9-11.

Bhansali, R.J. and D.Y. Downham, 1977, Some properties of the order of an autoregressive model selected by a generalisation of Akaike's FPE criterion, *Biometrika* 64, 547-551.

Bhattacharya, P.K., 1987, Maximum likelihood estimation of a changepoint in the distribution of independent random variables, general multiparameter case, *Journal of Multivariate Analysis* 23, 183-208.

Bhattacharyya, G.K. and A. Johnson, 1968, Nonparametric tests for shift at an unknown time point, *Annals of Mathematical Statistics* 39, 1731-1743.

Billah, M.B., 2001, Model selection for time series forecasting models, Unpublished Ph.D. Thesis, Department of Econometrics and Business Statistics Monash University, Australia.

Billah, M.B. and M.L. King, 1998, Model selection and optimal penalty estimation for time series models, in C.L. Skeels (ed.), *Proceedings of Econometric Society Australasian Meeting*, CD Rom, Australian National University, Australia.

Billah, M.B. and M.L. King, 2000, Time series model selection and forecasting via optimal penalty estimation, *Pakistan Journal of Statistics*, 16, 126-145.

Blum, G., 1987, An invariance principle for the changepoint problem, *Sankhya, A*, 49, 248-253.

Boldin, M.V., 1982, Estimation of the distribution of noise in an autoregressive scheme, *Theory of Probability and its Applications* 27, 866-871.

Boukai, B., 1993, A nonparametric bootstrapped estimate of the changepoint, *Journal of Nonparametric Statistics* 3, 123-134.

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

Breiman, L. and D. Freedman, 1983, How many variables should be entered in a regression equation? *Journal of the American Statistical Association* 78, 131-136.

Brendt, E., B. Hall, R. Hall and J. Hausman, 1974, Estimation and inference in nonlinear structural models, *Annals of Economic and Social Measurement* 3, 653-665.

Breusch, T.S. and A.P. Pagan, 1979, A simple test for heteroscedasticity and random coefficient variation, *Econometrica* 47, 1287-1294.

Broemeling, L.D. and J.H. Chin Choy, 1980, Some Bayesian inferences for a changing linear model, *Technometrics* 2, 71-78.

Brooks, R.D. and M.L. King, 1994, Hypothesis testing of varying coefficient regression models: Procedures and applications, *Pakistan Journal of Statistics* 10, 359-374.

Brown, R.L., J. Durbin and J.M. Evans, 1975, Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society* B, 149-163.

Brownstone, D., 1990, How to "data mine" if you must, bootstrapping Stein-rule model selection procedures, *Technical Report MBS*-08, Irvine Research Unit in Mathematical Behavioural Sciences, UC Urvine.

Bryant, E.C., 1960, *Statistical Analysis,* McGraw-Hill, New York.

Bunke, H. and B. Droge, 1985, A stepwise procedure for the selection of non-linear models, *Mathematische Operationsforschung und Statistik, Series Statistics* 16, 35-45.

Carnevali, P., L. Coletti, and S. Patarnello, 1985, Image processing by simulated annealing, *IBM Journal of Research and Development* 29, 569-579.

Chow, G., 1960, Tests of equality between subsets of coefficients in two linear regressions, *Econometrica* 28, 591-605.

Chu, C.S.J., K. Hornik and C.M. Kuan, 1995, MOSUM tests for parameter constancy, *Biometrika* 82, 603-617.

Conerly, M.D. and E.R. Mansfield, 1989, An approximate test for comparing independent regression models with unequal error variances, *Journal of Econometrics* 40, 239-259.

Corana, A., M. Marchesi, C. Martini and S. Ridella, 1987, Minimizing multinomial functions of continuous variables with the simulated annealing algorithm, *ACM Transactions of Mathematical Software* 13, 262-280.

Cramer, J.S., 1986, *Econometric Applications of Maximum Likelihood Methods*, Cambridge University Press, New York.

Csorgo, M. and L. Horvath, 1987, Nonparametric tests for the changepoint problem, *Journal of Statistical Planning and Inference* 17, 1-9.

Csorgo, M. and L. Horvath, 1988, Invariance principles for changepoint problems, *Journal of Multivariate Analysis* 27, 151-168.

Darkhovsky, B.S., 1976, A non-parametric method for the posterior detection of the "disorder" time of a sequence of independent random variables, *Theory of Probability and its Applications* 21, 178-183.

Davidson, R. and J.G. MacKinnon, 1984, Model specification tests based on artificial linear regressions, *International Economic Review* 25, 485-501.

Davies, R.B., 1977, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* 64, 247-254.

Davies, R.B., 1987, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* 74, 33-43.

Derwent, D., 1988, A better way to control population, *Nature* 331, 575-578.

Deshayes, J. and D. Picard, 1982, Tests of disorder regression, asymptotic comparison, *Theory of Probability and its Applications* 27, 100-115.

Dufour, J.M., 1982, Recursive stability analysis of linear regression relationships, *Journal of Econometrics* 19, 31-76.

Durbin, J. and G.S. Watson, 1951, Testing for serial correlation in least squares regression II, *Biometrika* 38, 159-178.

Engle, R.F. and T. Bollerslev, 1986, Modelling the persistence of conditional variances, *Econometric Reviews* 5, 1-50.

Evans, M.A. and M.L. King, 1985, A point optimal test for heteroscedastic disturbances, *Journal of Econometrics* 27, 163-178.

Evans, M.A. and M.L. King, 1988, A further class of tests for heteroscedasticity, *Journal of Econometrics* 37, 265-276.

Fair, R.C. and R. Shiller, 1990, Comparing information in forecasts from econometric models, *American Economic Review* 80, 375-389.

Farley, J.U. and M.J. Hinich, 1970, A test for a shifting slope coefficient in a linear model, *Journal of American Statistical Association* 65, 1320-1329.

Farley, J.U., M. Hinch, and T. McGuire, 1975, Some comparison of tests for a shift in the slopes of a multivariate linear time series model, *Journal of Econometrics* 3, 297-318.

Feder, P.I., 1975a, The log likelihood ratio in segmented regression, *Annals of Statistics* 3, 84-97.

Feder, P.I., 1975b, On asymptotic distribution theory in segmented regression problems identified case, *Annals of Statistics* 3, 49-83.

Ferger, D., 1994a, An extension of the Csorgo-Horvath functional limit theorem and its applications to changepoint problems, *Journal of Multivariate Analysis* 51, 338-351.

Ferger, D., 1994b, On the power of nonparametric changepoint tests, *Metrika* 41, 277-292.

Ferger, D., 1994c, Asymptotic distribution theory of changepoint estimators and confidence intervals based on bootstrap approximation, *Mathematical Methods of Statistics* 3, 362-378.

Finch, S.J., N.R. Mendell and H.C. Thode, 1989, Probabilistic measures of accuracy of a numerical search for a global maximum, *Journal of the American Statistical Association* 84, 1020-1023.

Forbes, C.S., M.L. King, and A. Morgan, 1995, A small sample variable selection procedure, in C.S. Forbes, P. Kofman and T.R.L. Fry eds., *Proceedings of the 1995 Econometrics Conference at Monash*, Monash University, 343-360.

Fox, G.C., 1988a, A graphical approach to load balancing and sparse matrix vector multiplication on the hypercube, in M. Schultz, editor, *Numerical Algorithms for Modern Parallel Computer Architectures*, Springer-Verlag, New York, 37-62.

Fox, G.C., 1988b, A review of automatic load balancing and decomposition methods for the hypercube, in M. Schultz, editor, *Numerical Algorithms for Modern Parallel Computer Architectures*, Springer-Verlag, New York, 63-76.

Fox, K.J., 1995, Model selection criteria: A reference source, An unpublished manuscript, University of New South Wales.

Franses, P.H., 1989, The distance between regression models and its impact on model selection, *Applied Mathematics and Computation* 34, 1-6.

Fraser, D.A.S., 1967, Data transformations and the linear model, *Annals of Mathematical Statistics* 38, 1456-1465.

GAUSS, 1997, *GAUSS Constrained Optimization: Application Module*, Aptech Systems, Inc., Washington.

Gaver, K.M. and M.S. Geisel, 1974, Discriminating among alternative models, Bayesian and non-Bayesian methods, In: Paul Zarembka ed., *Frontiers of Econometrics*, Academic Press, New York, 49-77.

Geisser, S., 1975, The predictive sample reuse method with applications, *Journal of the American Statistical Association* 70, 320-328.

Geman, S. and D. Geman, 1984, Stochastic relaxation, Gibbs distribution and the Bayesian restoration in images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721-741.

Geman, S. and C.R. Hwang, 1986, Diffusions for global optimization, *SIAM Journal of Control and Optimization* 24, 1031-1043.

Gidas, B., 1985, Global optimization via the Langevin equations. In *Proceedings of the 24th IEEE Conference on Decision and Control,* IEEE, New York, 774-778.

Goffe, W.L., G.D. Ferrier and J. Rogers, 1994, Global optimization of statistical functions with simulated annealing, *Journal of Econometrics* 60, 65-99.

Goldberger, A.S., 1962, Best linear unbiased prediction in the generalized linear regression model, *Journal of the American Statistical Association* 57, 369-375.

Goldfeld, S.M., 1976, The case of the missing money, *Brookings Papers on Economic Activities* 3, 588-635.

Golub, G.H., M. Heath and G. Wahba, 1979, Generalized cross validation as a method for choosing a good ridge parameter, *Technometrics* 21, 215-223.

Gombay, E. and L. Horvath, 1990, Asymptotic distributions of maximum likelihood tests for change in the mean, *Biometrika* 77, 411-414.

Gombay, E. and L. Horvath, 1994, Limit theorems for change in linear regression, *Journal of Multivariate Analysis* 48, 43-69.

Gombay, E. and L. Horvath, 1995, An application of U-statistics to changepoint analysis, *Acta Scientiarum Naturaliam* 60, 345-357.

Gombay, E., L. Horvath and M. Huskova, 1996, Estimates and tests for change in the variance, *Statistics and Decisions* 14, 145-151.

Gorman, J.W. and R.J. Toman, 1966, Selection of variables for fitting equations to data, *Technometrics* 8, 27-51.

Granger, C.J.W., M.L. King and H. White, 1995, Comments on testing economic theories and the use of model selection criteria, *Journal of Econometrics* 67, 173-187.

Grassa, A.A., 1989, *Econometric Model Selection: A New Approach*, Kluwer Academic Press, Boston.

Grose, S.D. and M.L. King, 1994, The use of information criteria for model selection between models with equal numbers of parameters, Paper presented at the *Australasian Meeting of the Econometric Society*, Armidale, Australia.

Haccou, P., E. Meelis and S.A. van de Geer, 1988, The likelihood ratio test for the changepoint problem for exponentially distributed random variable, *Stochastic Processes and Their Applications* 27, 121-139.

Hackl, P., 1980, *Testing the Constancy of Regression Models over Time*, Vandenhoeck and Ruprecht, Gottingen.

Hajek, B., 1988, Cooling schedules for optimal annealing, *Mathematics of Operational Research* 13, 311-329.

Hannan, E.J. and B.G. Quinn, 1979, The determination of the order of an autoregression, *Journal of the Royal Statistical Society* B, 190-195.

Hannan, E.J., 1980, The estimation of the order of an ARMA process, *Annals of Statistics* 8, 1071-1081.

Hansen, B. E., 1991, A comparison of tests for parameter instability, An examination of asymptotic local power, unpublished manuscript, Department of Economics, University of Rochester.

Hao, K., 1994, Testing for structural change in linear regression models, Unpublished Ph.D. Thesis, Department of Econometrics and Business Statistics, Monash University, Australia.

Hawkins, D. M., 1977, Testing a sequence of observations for a shift in location, *Journal of the American Statistical Association* 72, 180-186.

Henderson, R., 1990, A problem with the likelihood ratio test for a change-point hazard rate model, *Biometrika* 77, 835-843.

Hendry, D.F., 1989, *PC-GIVE, An Interactive Econometric Modelling System*, University of Oxford.

Hildreth, C. and J.P. Houck, 1968, Some estimators for a linear model with random coefficients, *Journal of the American Statistical Association* 68, 584-595.

Hinkley, D.V., 1969, Inference about the intersection in two-phase regression, *Biometrika* 56, 495-504.

Hinkley, D.V. and E.A. Hinkley, 1970, Inference about the changepoint in a sequence of binomial random variables, *Biometrika* 57, 477-488.

Hocking, R.R., 1976, The analysis and selection of variables in linear regression, *Biometrics* 32, 1-49.

Holley, R.A., S. Kusuoka and D.W. Stroock, 1989, Asymptotics of the spectral gap with applications to the theory of simulated annealing, *Journal of Functional Analysis* 83, 333-347.

Horvath, L., M, Huskova, and M. Serbinowska, 1997, Estimators for the time of change in linear models, *Statistics* 29, 109-130.

Horvath, L. and M. Serbinowska, 1995, Testing for changes in multinomial observations: the Lindisfarne Scribes problem, *Scandinavian Journal of Statistics* 22, 371-384.

Horvath, L. and Q.M. Shao, 1996, Limit theorems for maximum of standardized U-statistics with an application, *Annals of Statistics* 24, 2266-2279.

Hossain, Z.M. and M.L. King, 1998, Model selection when a key parameter is constrained to be in an interval, Working Paper 15/98, Department of Econometrics and Business Statistics, Monash University, Australia.

Hsu, D.A., 1977, Tests for variance shift at an unknown time point, *Applied Statistics* 26, 279-284.

Hu, I. and A.L. Rukhin, 1995, A lower bound for error probability in changepoint estimation, *Statistica Sinica* 5, 319-331.

Hughes, A.W., 1997, Improved model selection based on AIC-type criteria, Unpublished Ph.D. Thesis, Department of Econometrics and Business Statistics, Monash University, Australia.

Hughes, A.W., and M.L. King, 1994, One-sided model selection procedures, Paper presented to the Australasian Meetings of the Econometric Society, Armidale, Australia.

Hurvich, C.M. and C.L. Tsai, 1989, Regression and time series model selection in small samples, *Biometrika* 76, 297-307.

Hurvich, C.M. and C.L. Tsai, 1990, The impact of model selection on inference in linear regression, *The American Statistician* 44, 214-217.

Hurvich, C.M. and C.L. Tsai, 1993, A corrected Akaike information criterion for vector autoregressive model selection, *Journal of Time Series Analysis* 14, 271-279.

Huskova, M., 1988a, Recursive M-tests for detection of change, *Sequential Analysis* 7, 75-90.

Huskova, M., 1988b, Adaptive procedures for detection of change, *Statistics and Decisions* 6, 137-148.

Huskova, M., 1990a, Some asymptotic results foil robust procedures for testing the constancy of regression models over time, *Kybernetika*, 26, 392-403.

Huskova, M., 1990b, Asymptotics for robust MOSUM, *Comment, Mathematic, University Carolinae*, 31, 345-356.

Huskova, M., 1991, Recursive M-tests for the change point problem, In: *Economic, Structural Change: Analysis and forecasting*, Springer-Verlag, New York, 13-34.

Huskova, M., 1994, Nonparametric procedures for detecting a change in simple linear regression models, *Journal of Applied Statistical* Science 4, 337-358.

Huskova, M., 1996, Estimation of a change in linear models, *Statistics and Probability Letters* 26, 13-24.

Ilmakunnas, P. and H. Tsurumi, 1985, Testing the Lucas hypothesis on output-inflation trade offs, *Journal of Business and Economic Statistics* 3, 43-53.

Ingber, L., 1989, Very fast simulated re-annealing, *Mathematical and Computer Modelling* 12, 967-973.

James, B.J., K.L. James and D. Siegmund, 1987, Tests for a changepoint, *Biometrika* 74, 71-84.

Jeffreys, H., 1967, *Theory of Probability*, 4th ed., Oxford University Press, Oxford.

Jeng, F.C. and J.W. Woods, 1990, Simulated annealing in compound Gaussian random fields, *IEEE Transaction on Information Theory* 36, 94-107.

Johnson, D., C. Aragon, L. Mcgedch and C. Schevon, 1990, Optimization by simulated annealing: An experimental evaluation, Part I, Graph partitioning, *Operation Research* 37, 865-892.

Johnson, D., C. Aragon, L. Mcgedch, and C. Schevon, 1991, Optimization by simulated annealing, An experimental evaluation, Part II, Graph coloring and number partitioning, *Operation Research* 39, 378-406.

Johnson, D., C. Aragon, L. Mcgedch, and C. Schevon, 1992, Optimization by simulated annealing, An experimental evaluation, Part III, The travelling salesman problem, Unpublished manuscript.

Johnson, R.A. and M. Bagshaw, 1974, The effect of serial correlation on the performance of CUSUM tests, *Technometrics* 16, 103-112.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lutkepohl and T.C. Lee, 1985, *The Theory and Practice of Econometrics*, 2nd ed., John Wiley, New York.

Kabaila, P., 1995, The effect of model selection on confidence regions and prediction regions, *Econometric Theory* 11, 537-549.

Kalbfleish, J. and D. Sprott, 1970, Application of likelihood methods to models involving large numbers of parameters (with discussion), *Journal of the Royal Statistical Society* Series B, 175-208.

Kim, H.J and D. Siegmund, 1989, The likelihood ratio test for a change-point in simple linear regression, *Biometrika* 76, 409-423.

Kim, H.J. and L. Cai, 1993, Robustness of the likelihood ratio test for a change in simple linear regression, *Journal of the American Statistical Association* 88, 864-871.

King, M.L. and G. Bose, 2000, Finding optimal penalties for model selection in the linear regression model, mimeo, Department of Econometrics and Business Statistics, Monash University, Australia.

King, M.L., C.S. Forbes and A. Morgan, 1995, Improved small sample model selection procedures, *paper presented at the World Congress of the Econometric Society*, Tokyo, Japan.

King, M.L. and T.S. Shively, 1993, Locally optimal testing when a nuisance parameter is present only under the alternative, *The Review of Economics and Statistics* 75, 1-7.

Kirkpatrick, S.C.D., M.P. Gelatt and M.P. Vecchi, 1982, Optimization by simulated annealing, *IBM Research Report*, RC 9355.

Kirkpatrick, S.C.D., M.P. Gelatt and M.P. Vecchi, 1983, Optimization by simulated annealing, *Science* 220, 671-680.

Kohn, R., 1983, Consistent estimation of minimal subset dimension, *Econometrica* 51, 367-376.

Koul, H.L. and S. Levental, 1989, Weak convergence of the residual empirical process in explosive autoregression, *Annals of Statistics* 17, 1784-1794.

Kozin, F., and F. Nakajima, 1980, The order determination problem for linear time varying AR models, *IEEE Transactions on Automatic Control* 25, 250-257.

Kramer, W., W. Ploberger, and A. Raimund, 1988, Testing for structural change in dynamic models, *Econometrica* 56, 1355-1369.

Kushner, H. J., 1985, Asymptotic global behavior for stochastic approximations and diffusions with slowly decreasing noise effects, Global minimization via Monte Carlo, Technical Report LCDS 85-7, Department of Applied Mathematics, Brown University.

Kwek, K.T., 1999, *Model Selection for a Class of Conditional Heteroscedastic Processes*, Unpublished Ph.D. Thesis, Department of Econometrics and Business Statistics, Monash University, Australia.

Kwek, K.T. and M.L. King, 1997a, Model selection and optimal penalty estimation for conditional heteroscedastic processes, in P. Bardsley and V.L. Martin eds., *Proceedings of the Econometric Society Australasian Meeting*, University of Melbourne, Australia.

Kwek, K.T. and M.L. King, 1997b, Model Selection of ARCH Processes using iterated optimal penalties, *Proceedings of the Third Annual Doctoral Research Conference*, Faculty of Business and Economics, Monash University, Australia.

Kwek, K. T. and M.L. King, 1998, Information criteria in conditional heteroscedastic models, a Bayesian prior approach to penalty function building, in C.L. Skeels ed., *Proceedings of the Econometric Society Australasian Meeting*, CD Rom, Australian National University, Canberra, Australia.

Laud, P., L.M. Berliner and P.K. Goel, 1989, A stochastic probing algorithm for global optimization In *Computer Science and Statistics, Proceedings of the 21st Symposium on the Interface* (K. Berk and L. Malone, eds.), American Statistical Association Alexandria, VA, 220-229.

Leamer, E.E., 1978, *Specification Searches*, Wiley, New York.

Leamer, E.E., 1979, Information criteria for choice of regression models, a comment, *Econometrica* 47, 507-510.

Linhart, H. and W. Zucchini, 1986, *Model Selection*, John Wiley and Sons, New York.

Lo, A.W., and W.K. Newey, 1985, A large-sample Chow test for the linear simultaneous equation, *Economic Letters* 18, 351-353.

Loader, C.R., 1992, A log-linear model for a Poisson process changepoint, *Annals of Statistics*, 20, 1391-1411.

Lombard, F., 1983, Asymptotic distributions of rank statistics in the changepoint problem, *South African Statistics Journal* 17, 83-105.

Lombard, F., 1988, Detecting change points by Fourier analysis, *Technometrics* 30, 305-310.

Lombard, F. and J.D. Hart, 1994, The analysis of changepoint data with dependent errors, In: *Changepoint Problems*, IMS Lecture Notes-Monograph Series, 23, 194-209.

Lutkepohl, H., 1988, Prediction tests for structural stability, *Journal of Econometrics* 39, 267-296.

Lutkepohl, H., 1989, Prediction tests for structural stability of multiple time series, *Journal of Business and Economic Statistics* 7, 129-135.

MacKinnon, J.G., 1983, Model specification tests against non-nested alternatives, with discussion and reply, *Econometric Reviews* 2, 85-157.

Maddala, G.S., 1992, *Introduction to Econometrics*, Macmillan Publishing Company, New York.

Makridakis, S., 1986, The art and science of forecasting, An assessment and future direction, *International Journal of Forecasting* 2, 15-39.

Mallows, C.L., 1964, Choosing variables in a linear regression, a graphical aid. Paper presented at the Central Regional Meeting of the Institute of Mathematical Statistics Manhattan, Kansas.

Mallows, C.L., 1973, Some comments on C_p, *Technometrics* 15, 661-676.

McAleer, M., 1987, Specification tests for separate models: a survey, In M.L. King and D.E.A. Giles eds., *Specification Analysis in the Linear Model, Essays in Honour of Donald Cochrane*, Routledge and Kegan Paul, New York, 146-196.

McCabe, B.P.M., 1988, A multiple decision theory analysis of structural stability in regression, *Econometric Theory* 4, 499-508.

McCabe, B.P.M. and M.J. Harrison, 1980, Testing the constancy of regression relationships over time using least squares residuals, *Applied Statistics* 29, 142-148.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, 1953, Equation of state calculations by fast computing machines, *Journal of Chemical Physics* 21, 1087-1092.

Meyer, R.K. and C.J. Nachtsheim, 1988, Constructing exact D-optimal experimental designs by simulated annealing, *American Journal of Mathematical and Management Science* 8, 329-359.

Miao, B.Q., 1988, Inference in a model with at most one slope-changepoint, *Journal of Multivariate Analysis* 27, 375-391.

Miller, R. and D. Siegmund, 1982, Maximally selected chi square statistics, *Biometrics* 38, 1011-1016.

Mills, J.A. and K. Prasad, 1992, A comparison of model selection criteria, *Econometric Reviews* 11, 201-233.

Muller, H.G., 1992, Change points in nonparametric regression analysis, *Annals of Statistics* 20, 737-761.

Nishii, R., 1988, Maximum likelihood principle and model selection when the true model is unspecified, *Journal of Multivariate Analysis* 27, 392-403.

Nyblom, J., 1989, Testing for the constancy of parameters over time, *Journal of the American Statistical Association* 84, 223-230.

Ohtani, K. and T. Toyoda, 1985, Small sample properties of tests of equality between sets of coefficients in two linear regressions under heteroscedasticity, *International Economic Review* 26, 37-44.

Otten, R.H.J.M., and L.P.P.P. van Ginneken, 1989, *The Annealing Algorithm*, Kluwer Academic, Boston, MA.

Ozaki, T., 1977, On the order determination of ARIMA models, *Applied Statistics* 26, 290-301.

Ozaki, T. and H. Tong, 1975, On the fitting of non-stationary autoregressive models in time series analysis, *proceeding of the 8-th Hawaii International Conference on System Science*, Western Periodical Company.

Pagan, A.R. and G.W. Schwert, 1990, Alternative models for conditional stock volatility, *Journal of Econometrics* 45, 267-290.

Page, E.S., 1954, Continuous inspection schemes, *Biometrika* 11, 1-15.

Page, E.S., 1955, A test for a change in a parameter occurring at an unknown point, *Biometrika* 42, 523-527.

Pettitt, A.N., 1979, A non-parametric approach to the changepoint problem, *Applied Statistics* 28, 126-135.

Pettitt, A.N., 1980, A simple cumulative sum type statistic for the changepoint problem with zero-one observations, *Biometrika* 67, 79-84.

Pincus, M.A., 1970, Monte Carlo method for the approximate solution of certain types of constrained optimization problems, *Operation Research* 18, 1225-1228.

Ploberger, W. and W. Kramer, 1989, The local power of the CUSUM and CUSUM of squares tests, *Econometric Theory* 6, 335-347.

Ploberger, W. and W. Kramer, 1992, The CUSUM test with OLS residuals, *Econometrica* 60, 271-285.

Ploberger, W., W. Kramer, and R. Alt, 1989, A modification of the CUSUM test in the linear regression model with lagged dependent variables, *Empirical Economics* 14, 1-11.

Ploberger, W., W. Kramer and K. Kontrous, 1989, A new test for structural stability in the linear regression model, *Journal of Econometrics* 40, 307-318.

Potscher, B.M., 1991, Effects of model selection on inference, *Econometric Theory* 7, 163-185.

Potscher, B.M. and A.J. Novak, 1996, The distribution of estimators after model selection, large and small sample results, *Mimeo, Institute fur Statistik, Operation Research and Computerverfahren*, Universitat Wein.

Praagman, J., 1988, Bahadur efficiency of rank tests for the change-point problem, *Annals of Statistics* 16, 198-217.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1986, *Numerical Recipes, the Art of Scientific Computing*, Cambridge University, Press, Cambridge, England.

Pronzato, L., E. Walter, A. Venot, and J.F. Lebruchec, 1984, A general-purpose optimizer, implementation and applications, *Mathematics and Computers in Simulation*, 24, 412-422.

Quandt, R., 1958, The estimation of the parameters of a linear regression system obeying two separate regimes, *Journal of the American Statistical Association* 53, 873-880.

Quandt, R., 1960, Tests of the hypothesis that a linear regression system obeys two separate regimes, *Journal of the American Statistical Association* 55, 324-330.

Quinn, B.G., 1988, A note on AIC order determination for multivariate autoregressions, *Journal of Time Series Analysis* 9, 241-245.

Rahman, M.S., G.K. Bose and M.L. King, 1998, Improved penalty functions for information criteria based model selection, in C.L. Skeels (ed.), *Proceedings of the Econometric Society Australasian Meeting*, CD Rom, Australian National University, Australia.

Rahman, M.S. and M.L. King, 1997, Marginal likelihood based score tests of regression disturbances in the presence of nuisance parameters, *Journal of Econometrics* 82, 63-80.

Rissanen, J., 1986, Stochastic complexity and modelling, *Annals of Statistics* 14, 1080-1100.

Rissanen, J., 1987, Stochastic complexity and modelling, *Journal of the Royal Statistical Society* B, 223-239 and 252-265 (with discussions).

Rissanen, J., 1988, Stochastic complexity and the MDL principle, *Econometric Reviews* 7, 85-102.

Ronchetti, E., 1985, Robust model selection in regression, *Statistics and Probability Letters* 3, 21-23.

Rosenberg, B., 1973, The analysis of a cross-section of time series by stochastically convergent parameter regression, *Annals of Economic and Social Measurement* 2, 399-428.

Rothman, D., 1968, Letter to the editor, *Technometrics* 10, 432.

Roy, S.N., 1953, On a heuristic method of test construction and its use in multivariate analysis, *Annals of Mathematical Statistics* 24, 220-238.

Roy, S.N., R. Gnanadesikan and J.N. Srivastava, 1971, *Analysis and Design of Certain Quantitative Multiresponse Experiments*, Oxford, Pergamon Press.

Rukhin, A.L., 1994, Asymptotic minimaxity in the changepoint problem, In: *Changepoint Problems*, IMS Lecture Notes, Monograph Series 23, 284-291.

Sakai, H., 1981, Asymptotic distribution of the order selected by AIC in multivariate autoregressive model fitting, *International Journal of Control* 33, 175-180.

Schmidt, P., 1971, Methods of choosing among alternative models, *Econometrics Workshop Paper* 7004, Michigan State University.

Schmidt, P., 1973, Calculating the power of the minimum standard error choice criterion, *International Economic Review* 14, 253-255.

Schmidt, P., 1974, Choosing among alternative linear regression models, *Atlantic Economic Journal* 2, 7-13.

Schmidt, P., 1975, Choosing among alternative linear regression models, a correction and some further results, *Atlantic Economic Journal* 3, 61-63.

Schmidt, P. and R. Sickles, 1977, Some further evidence on the use of Chow test under heteroscedasticity, *Econometrica* 45, 1293-1298.

Schwarz, G., 1978, Estimating the dimension of a model, *Annals of Statistics* 6, 461-464.

Seber, G.A.F. and C.J. Wild, 1989, *Nonlinear Regression*, Chapter 9, Wiley Series in Probability and Mathematical Statistics.

Sen, P.K., 1980, Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebeite* 52, 203-218.

Sen, P.K., 1985, *Theory and Applications of Sequential Nonparametrics*, SIAM, Philadelphia.

Sen, A. and M.S. Srivastava, 1975, On tests for detecting change in mean, *Annals of Statistics* 3, 98-108.

Shibata, R., 1976, Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika* 63, 117-126.

Shibata, R., 1980, Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, *Annals of Statistics* 8, 147-164.

Shibata, R., 1981, An optimal selection of regression variables, *Biometrika* 68, 45-54.

Shibata, R., 1986, Consistency of model selection and parameter estimation, In J. Gani and M.B. Priestley eds., *Essays in Time Series and Allied Processes*, *Journal of Applied Probability* 23A, 127-141.

Sin, C.Y. and H. White, 1992, Information criteria for selecting parameter models, a misspecification analysis *Working Paper*, Department of Economics, University of California, San Diego.

Sin, C.Y. and H. White, 1996, Information criteria for selecting possibly misspecified parametric models, *Journal of Econometrics* 71, 207-225.

Smith, A.F.M., 1975, A Bayesian approach to inference about a change-point in a sequence of random variables, *Biometrika* 62, 407-416.

Soderstrom, T., 1977, On model structure testing in system identification, *International Journal of Control* 26, 1-18.

Srivastava, M.S. and K.J. Worsley, 1986, Likelihood ratio tests for a change in the multivariate normal mean, *Journal of the American Statistical Association* 81, 199-204.

Stone, M., 1974, Cross-validatory choice and assessment of statistical predictors. *Journal of the Royal Statistical Society Series B*, 111-147.

Stone, M., 1977, Asymptotics for and against cross-validation, *Biometrika* 64, 29-35.

Stone, M., 1979, Comments on model selection criteria of Akaike and Schwarz, *Journal of the Royal Statistical Society* B, 276-278.

Stute, W., 1996, Changepoint problems under random censorship, *Statistics* 27, 255-265.

Szu, H. and R. Hartley, 1987, Fast simulated annealing, *Physics Letters* A, 157-162.

Tan, S.L. and M.L. King, 1994, Testing structural change when the changepoint is unknown, Mimeo, Department of Econometrics and Business Statistics, Monash University, Australia.

Tang, S.M. and I.B. MacNeill, 1993, The effect of serial correlation on tests for parameter change at unknown time, *Annals of Statistics* 21, 552-575.

Telly, H., T.M. Liebling and A. Mocellin, 1987, Reconstruction of polycrystalline structures, a new application of combinatorial optimization, *Computing* 38, 1-11.

Theil, H., 1961, *Economic Forecasts and Policy*, Second edition, North-Holland, Amsterdam.

Theil, H., 1971, *Principles of Econometrics*, John Wiley and Sons, London.

Thompson, M.L., 1978, Selection of variables in multiple regression, *International Statistical Review* 46, 1-19 and 129-146.

Tong, H., 1975, Determination of the order of a Markov chain by Akaike's information criterion, *IEEE Transactions on Information Theory* 21, 476-480.

Tong, H., 1978, The asymptotic joint distribution of the estimated autoregressive coefficients, *International Journal of Control* 27, 801-807.

Toyoda, T., 1974, Use of the Chow tests under heteroscedasticity, *Econometrica* 42, 601-608.

Toyooka, Y., 1982, Prediction error in a linear model with estimated parameters, *Biometrika* 69, 453-459.

Tunnicliffe Wilson, G., 1989, On the use of marginal likelihood in time series model estimation, *Journal of the Royal Statistical Society* Series B, 15-27.

Van Laarhoven, P. and E. Aarts, 1987, *Simulated Annealing: Theory and Applications*, Reidel, Dordrecht.

Van Laarhoven, P., C. Boender, E. Aarts, and K.A. Rinnddy, 1989, A Bayesian approach to simulated annealing, *Probability in the Engineering and Informational Sciences* 3, 453-475.

Vetterling, W.T., W.H. Press, S.A. Teukolsky and B.P. Flannery, 1994, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Cambridge University, New York.

Vostrikova, L., 1981, Detecting "disorder" in multidimensional random processes, *Soviet Mathematics Doklady* 24, 55-59.

Vostrikova, L., 1983, Functional limit theorems for the disorder problem, *Stochastic* 9, 103-124.

Wald, A., 1943, Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Transactions of the American Mathematical Society* 54, 426-482.

Wallis, K.F., 1989, Macroeconomic forecasting: A survey, *Economic Journal* 99, 28-61.

Wasserman, P.D. and T. Schwartz, 1988, Neural networks, part 2, what are they and why is everybody so interested in them now? *IEEE Expert*, Spring, 10-15.

Watson, G.S., 1995, Detecting a change in the intercept in multiple regression, *Statistics and Probability Letters* 23, 69-72.

Watson, M. W. and R.F. Engle, 1985, Testing for regression coefficient stability with a stationary AR1 alternative, *Review of Economics and Statistics* 67, 341-346

Westlund, A., 1985, On the power of some tests for examining the stability of regression coefficients, *Computation and Simulation* 21, 275-289.

White, H., 1982a, Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1-25.

White, H., 1982b, Model specification, *Journal of Econometrics* 20, 1-157.

White, H., 1983, Nonnested models, *Journal of Econometrics* 21, 1-160.

White, H., 1990, A consistent model selection procedure based on M-testing, In: C.W.J. Granger ed., *Modelling Economic Series, Readings in Econometric Methodology*, Oxford University Press, Oxford.

Wong, D.F., H.W. Leong and C.L. Liu, 1988, *Simulated Annealing for VLSI Design*, Kluwer Academic Publishers, Boston, MA.

Woodroofe, M., 1982, On model selection and the arc sine laws, *Annals of Statistics* 10, 1182-1194.

Worsley, K.J., 1979, On the likelihood ratio test for a shift in location of normal populations, *Journal of the American Statistical Association* 74, 365-367.

Worsley, K.J., 1983, The power of likelihood ratio and cumulative sum tests for a change in a binomial probability, *Biometrika* 70, 455-464.

Yao, Q., 1993a, Tests for changepoints with epidemic alternatives, *Biometrika* 80, 179-191.

Yao, Q., 1993b, Boundary crossing probabilities of some random fields related to likelihood ratio tests for epidemic alternatives, *Journal Applied Probability* 30, 52-65.

Yao, Y.C., 1990, On the asymptotic behavior of a class of nonparametric tests for a changepoint problem, *Statistics and Probability Letters* 9, 173-177.

Young, A.S., 1982, The Bivar criterion for selecting regressors, *Technometrics* 24, 181-189.

Zarembka, P., 1974, *Frontiers in Econometrics*, Academic Press, New York.

Zellner, A., 1979, Statistical analysis of econometric models, *Journal of the American Statistical Association* 74, 628-642.

Zellner, A., 1988, Bayesian analysis in econometrics, *Journal of Econometrics* 37, 27-50.