1986

Topics In Time Series Analysis And Forecasting

Thian San Kheoh

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

TOPICS IN TIME SERIES ANALYSIS AND FORECASTING

by

Thian San Kheoh

Department of Statistical and

Actuarial Sciences

Submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Faculty of Graduate Studies

The University of Western Ontario

London, Ontario

June, 1986

ISBN 0-315-33038-4

ABSTRACT

This thesis contains new developments in various topics in time series analysis and forecasting. These topics include: model selection, estimation, forecasting and diagnostic checking.

In the area of model selection, finite and large sample properties of the commonly used selection criteria, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), are discussed. In the finite case, the study is limited to the two sample problem. The exact probability of selection is obtained for finite samples. The risk of each criterion is evaluated in the two sample situation. Empirical evidence regarding these risks are given for autoregressive processes. The asymptotic distribution of the $\hat{h}$ is given, where $\hat{h}$ is the estimate of the number of extra parameters in the model selected by the AIC criterion. This derivation is based on large sample properties of the likelihood ratio test statistic. The asymptotic distribution of the AIC in PAR models is also discussed.

In estimation, an explicit expression for the efficiency of strongly consistent estimates for the ARMA(1,1) model is derived. Empirical efficiency and the empirical estimate are examined by simulation.

On the topic of forecasting, the asymptotic variance of the forecast error is derived for an autoregressive model of first order. In the derivation, the estimated parameter is not assumed to be independent of the data. The variance of the one-step forecast error is also derived for the fractional noise model.

In the last topic, empirical results for portmanteau test statistics are studied. It is shown that the modified Portmanteau test of Ljung and Box (1980) outperforms the modified test of Li and McLeod (1981). In testing for whiteness, the modified Portmanteau test is shown to have lower power than the cumulative periodogram test against both fractional noise and standard ARMA alternatives.

ACKNOWLEDGEMENTS

I wish to thank my parents for their patience and sacrifice throughout the course of my study.

I also wish to express my greatest appreciation to my supervisor, Dr. A. I. McLeod, for his suggestions, encouragement and patience throughout the entire period of my research. Without his guidance, this thesis would not have been possible.

Many thanks to Dr. K. W. Hipel, the Department of Statistics and the Faculty of Graduate Studies for financial support.

I would like to thank Dr. R. J. Kulperger and Professor I. B. MacNeill for reading the thesis. Also, thanks to all former and present graduate students and friends for making my entire stay at Western very enjoyable.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 5

LIST OF FIGURES

CHAPTER 2

CHAPTER 1

INTRODUCTION

Four main topics in time series analysis are discussed in this dissertation; *viz.*, model selection, estimation, forecasting and diagnostic checking.

There is no question that model selection is an important topic for any statistical modeller. Many statistical model selection criteria have been proposed and some have been widely used. In chapter 2, finite sample properties of the Akaike Information Criterion (AIC) (see Akaike, 1973, 1974) and the Bayesian Information Criterion (BIC) (see Schwarz, 1978) are examined. In particular, the investigation derives the finite sample distribution of these criteria and their risks for the two sample problem. In this case, it is shown that neither AIC nor BIC dominates. For the autoregressive case, empirical evidence on the performance of both criteria, in terms of their forecast error, is given. It is shown that the forecast error for the AIC is larger than the BIC. Next, a general result on the asymptotic distribution of the AIC selecting a model with h extra parameters is derived. This result is a generalization of the result of Shibata, (1976). As an illustration, the asymptotic distribution of the AIC selecting a model with extra parameters in the periodic autoregressive (PAR) model.

In chapter 3, an explicit expression for the asymptotic efficiency for a strongly consistent estimator (Hannan, 1975) is obtained for the

ARMA(1,1) model. For this model, this estimation technique is shown to be asymptotically inefficient relative to the maximum likelihood estimator. The empirical efficiency for this estimator relative to the maximum likelihood estimator is examined by simulation experiments.

Forecasting is the ultimate use of a chosen model. It is desired for one to obtained a forecast as close as possible to the value assumed in future by the process. In other words, it is hoped that the forecast error will be small. Many previous authors have derived the asymptotic mean square error of the forecast for the general autoregressive process. The result obtained by previous authors, such as Akaike (1970), assumes that the parameter estimates are independent of data used for forecasting. This implies that the model is calibrated on one set of data and used for forecasting on another independent set of data. However, in practice, the model is often calibrated on data $z_1, \ldots, z_n$ and then used to forecast $z_n, z_{n+1}, \ldots$. In chapter 4, an explicit expression of the variance of the ℓ-step ahead forecast error is derived for an autoregressive process of order one (AR(1)). In the derivation, the aforementioned dependency of the parameter estimate is taken into consideration. It is shown that the variance of the forecast error depends on the parameter and a smaller variance is obtained when dependency is taken into account. The variance of the one-step ahead forecast error is also derived for the fractional difference noise model, FARMA(0,d,0). In this case, the variance does not depend on the parameter d in this model. A comparison of this result with those for the AR(1) model and for the autoregressive-moving average (ARMA) model is discussed. It is shown that the result for the AR(1) model does not generalize to other models such as ARMA(0,0) and FARMA(0,d,0).

Several test statistics commonly used in diagnostic checking are empirically examined in chapter 5. The means and variances for various portmanteau statistics were compared (see for eg. Box and Pierce, 1970; Ljung and Box, 1978; Li and McLeod, 1981). In particular, exact means for the white noise process were compared. The performance of these statistics for autoregressive processes were also examined: in particular their means, type I errors and powers of detecting a misspecification were considered. It is seen that the modified portmanteau statistics of Ljung and Box, and Li and McLeod give close estimates to the mean and type I error. The Ljung and Box statistic is seen to provide higher power in general. Also, the statistic of McLeod and Li (1983) which uses squared residuals is examined for type I error and power. It is seen from this study that this statistic does not perform well in linear time series models. The modified portmanteau statistics and the cumulative periodogram (CUP) test commonly used in testing for whiteness are also examined. In the investigation, the AR model, the MA model and the FARMA(0,d,0) were used in a simulation to determine which test statistic perform best. The CUP test is seen to perform better than either of the modified portmanteau statistics in the ARMA and FARMA models.

REFERENCES

Akaike, H. (1970), "Statistical Predictor Identification". *Annals of the Institute of Statistical Mathematics* **22**, 203–217.

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle". *2nd International Symposium on Information Theory.* (eds. B. N. Petrov and F. Csaki), Akademiai Kiado, Budapest, 267–281.

Akaike, H. (1974), "A New Look at the Statistical Model Identification". *IEEE Transactions on Automatic Control* **AC-19**, 716–723.

Box, G. E. P. and Pierce, D. A. (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Models". *Journal of the American Statistical Association* **65**, 1509–1526.

Hannan, E. J. (1975), "The Estimation of ARMA Models". *The Annals of Statistics* **3**, 975–981.

Li, W. K. and McLeod, A. I. (1981), "Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models". *Journal of the Royal Statistical Society (B)* **43**, 231–239.

Ljung, G. M. and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models". *Biometrika* **65**, 297–303.

McLeod, A. I. and Li, W. K. (1983), "Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations". *Journal of Time Series Analysis* **4**, 269–273.

Schwarz, G. (1978), "Estimating the Dimension of a Model". *The Annals of Statistics* **6**, 461–464.

Shibata, R. (1976), "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion". *Biometrika* **63**, 117–126.

CHAPTER 2

MODEL SELECTION USING AIC AND BIC

2.1 INTRODUCTION

Consider a zero mean, stationary time series $\{z_t;\ t = 1, 2, \ldots\}$. An autoregressive model of order k for this series, denoted by AR(k), is defined by

$$z_t = \phi_1(k)z_{t-1} + \cdots + \phi_k(k)z_{t-k} + a_t, \qquad k \geq 1, \tag{2.1}$$

where $\phi_i(k)$, $i = 1, 2, \ldots, k$ are real parameters to be estimated. For stationarity, it is required that the roots of $\phi(B) = 0$ are less than unity, where $\phi(B) = \phi_1(k)B - \cdots - \phi_k(k)B^k$ and where B is the backward shift operator such that $B^i z_t = z_{t-i}$. The sequence $\{a_t;\ t = 1, 2, \ldots\}$ are assumed to be independent, identically and normally distributed with mean zero and variance σ_a^2.

Let $\{\hat{\phi}_i(k);\ i = 1, 2, \ldots, k\}$ denote the set of estimated parameters of an AR(k) model based on n observations $z_1, \ldots, z_n$ and assume that k is bounded above by K. The parameters in (2.1) can be estimated by solving the Yule-Walker equations:

$$c_j = \sum_{i=1}^{k} \hat{\phi}_i(k)c_{j-i}, \qquad j = 1, 2, \ldots, k,$$

where

$$c_m = \begin{cases} \frac{1}{n}\sum_{t=K+1}^{n} z_t z_{t-m} & m \geq 0 \\ c_{-m} & m < 0. \end{cases}$$

Then an estimator of the variance of a_t is given by,

$$\hat{\sigma}_a^2(k) = \frac{1}{n} \sum_{t=K+1}^{n} \{z_t - \hat{\phi}_1(k)z_{t-1} - \cdots - \hat{\phi}_k(k)z_{t-k}\}^2. \tag{2.2}$$

There exist several model selection criteria. Akaike (1969, 1970) developed the Final Prediction Error (FPE) which is defined by

$$\text{FPE}(k) = \left(\frac{n+k}{n-k}\right)\hat{\sigma}_a^2(k), \qquad k = 0, 1, \ldots, K. \tag{2.3}$$

where $\hat{\sigma}_a^2(k)$ is as defined in (2.2). In this procedure, the model is identified as an AR(k) if FPE is minimized at k, for $0 \le k \le K$. If the true model is an AR(k), then FPE(h) is an unbiased estimate of the one-step prediction error variance, $\sigma_a^2(1 + h/n)$, provided that $h \ge k$. For $h < k$, on the other hand, $\hat{\sigma}_a^2(h) >> \sigma_a^2$. Thus use of the FPE in model selection achieves a balance between model adequacy ($\hat{\sigma}_a^2(h) \approx \sigma_a^2$) and model parsimony ($h$ should be as small as possible without violating model adequacy).

Perhaps the most widely used selection criterion is the Akaike Information Criterion (AIC) (Akaike 1973, 1974) defined by

$$\begin{aligned} \text{AIC}(k) &= -2\log(\text{maximum likelihood}) + 2k \\ &= n\log\{\hat{\sigma}_a^2(k)\} + 2k. \end{aligned} \tag{2.4}$$

The model with minimum AIC(k) is selected for $0 \le k \le K$. The minimum AIC model has the maximum estimated entropy. Since $\log \text{FPE}(k) = n^{-1}\text{AIC}(k) + O(1/n^2)$, the asymptotic behaviour of these two criteria is identical. In practice, the AIC criterion has been widely used in determining the order of a model. However, this criterion has been shown by Shibata (1976) and subsequently by Woodroofe

(1982) to be inconsistent in nested modelling situations. It is inconsistent in the sense that there is a nonzero probability of selecting an order that is higher than the true order. Duong (1984) studied the AIC from the point of view of ranking and selection procedures.

Independently, Akaike (1978, 1979) and Schwarz (1978) proposed another selection criterion based on Bayesian considerations. The Bayesian Information Criterion (BIC) is given by

$$\mathrm{BIC}(k) = n \log\{\hat{\sigma}_a^2(k)\} + k \log(n), \tag{2.5}$$

and the model with minimum BIC(k), $0 \leq k \leq K$ is selected. The minimum BIC model has the greatest posterior probability.

Other criteria have also been proposed. Parzen (1974) suggested the Criterion Autoregressive Transfer Function (CAT) which is also inconsistent in the nested situation. Hannan and Quinn (1979) proposed a strongly consistent criterion. This criterion is obtained by replacing $2k$ in (2.4) by $2kc \log\log(n)$ where $c > 1$. The asymptotic relationships between these criteria can be found in Priestley (1981, pp. 372–376). Some comments on AIC and BIC were given in the paper by Stone (1979). Stone (1977) discussed the asymptotic relationship between AIC and cross-validation.

In this chapter, the performance of AIC and BIC is of particular interest. Section 2.2 considers the small sample properties of each criterion. A comparison between the criteria is discussed based on the likelihood ratio test statistic for the two sample problem. Section 2.3 discusses the risk in using AIC and BIC in the two sample problem. The empirical risk for AIC and BIC in the time series context is given in section 2.4. A simulation experiment was carried out for the autoregressive processes. In section 2.5, the asymptotic distribution of model selection using the AIC is derived. The derivation is based on the likelihood ratio test statistic. Section 2.6 discusses an application of the result of section 2.5 to the periodic autoregressive model.

2.2 FINITE SAMPLE COMPARISON OF AIC AND BIC FOR THE TWO SAMPLE PROBLEM

Let $x = \{x_1, \ldots, x_n\}$ and $y = \{y_1, \ldots, y_n\}$ denote two random samples of size n. Suppose that two models are available. In model A, the two random samples are assumed to be independent identically normally distributed with common mean, μ, and common variance, σ^2. The maximum likelihood estimators of μ and σ^2, denoted by $\hat{\mu}_A$ and $\hat{\sigma}_A^2$, are respectively,

$$\hat{\mu}_A = \frac{\bar{x} + \bar{y}}{2}$$

and

$$\hat{\sigma}_A^2 = \frac{1}{2n}\left\{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2 + \frac{n}{2}(\bar{x} - \bar{y})^2\right\} .$$

In model B, the random samples are assumed to be distributed normally with different means, μ_x and μ_y, and common variance, σ^2. Let the corresponding maximum likelihood estimators be denoted by $\hat{\mu}_x$, $\hat{\mu}_y$ and $\hat{\sigma}_B^2$. These estimators are,

$$\hat{\mu}_x = \bar{x},$$

$$\hat{\mu}_y = \bar{y},$$

and

$$\hat{\sigma}_B^2 = \frac{1}{2n}\left\{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2\right\}.$$

Suppose the following hypotheses are of interest: $H_o : \mu_x = \mu_y$ versus $H_1 : \mu_x \neq \mu_y$ and denote the likelihood ratio test statistic by λ. Then λ is defined as follows:

$$\lambda = \frac{sup L_A}{sup L_B}$$

$$= \left(\frac{\hat{\sigma}_B^2}{\hat{\sigma}_A^2} \right)^n$$

$$= \left(1 + \frac{t^2}{2(n-1)} \right)^{-n},$$

where $sup L_A$ denotes the supremum of the likelihood function under model A; $sup L_B$ is similarly defined and

$$t^2 = \frac{n}{2} \frac{(\overline{x} - \overline{y})^2}{\left(\sum (x_i - \overline{x})^2 + \sum (y_i - \overline{y})^2 \right) / 2(n-1)}.$$

It is well-known that t has the Student's t distribution with $2(n-1)$ degrees of freedom.

Both the AIC and BIC are commonly used for model discrimination in time series. The large sample properties of the AIC are known for autoregressive models (Shibata, 1976), but little is known about its properties in small samples. This is also true for the BIC. Hence, it is of interest to investigate the small sample properties of these criteria.

Case 1. If model A is the true model, then

$$Pr\{AIC_A \leq AIC_B \mid A\}, \tag{2.6}$$

is the probability that the AIC will select model A when model A is the true model. From the definition of AIC given in equation (2.4), (2.6) becomes,

$$\Pr\{2n \log \hat{\sigma}_A^2 + 4 \le 2n \log \hat{\sigma}_B^2 + 6\}$$

$$= \Pr\left\{ 2n \log\left(\frac{\hat{\sigma}_A}{\hat{\sigma}_B^2}\right) \le 2 \right\}$$

$$= \Pr\left\{ \frac{\hat{\sigma}_A}{\hat{\sigma}_B^2} \le \exp\left(\frac{1}{n}\right) \right\}$$

$$= \Pr\left\{ 1 + \frac{t^2}{2(n-1)} \le \exp\left(\frac{1}{n}\right) \right\}$$

$$= \Pr\left\{ t^2 \le 2(n-1)\left[\exp\left(\frac{1}{n}\right) - 1\right] \right\}. \qquad (2.7)$$

Note that,

$$\lim_{n\to\infty} 2(n-1)\Big(\exp(1/n) - 1\Big) = 2.$$

It follows that for large n, the probability of selecting the correct model (A in the present case) is $\Pr\{Z^2 \le 2\} \approx 0.843$, where Z is a standard normal variable. Hence the AIC is inconsistent for large n. This agrees with the result of Shibata (1976) ~~for~~ AR(p) models.

Now for the BIC criterion, an equation similar to (2.6) can be given.

$$\Pr\{\mathrm{BIC}_A \le \mathrm{BIC}_B \mid A\}$$

$$= \Pr\{2n \log \hat{\sigma}_A^2 + 2 \log n \le 2n \log \hat{\sigma}_B^2 + 3 \log n\}$$

$$= \Pr\left\{ \frac{\hat{\sigma}_A^2}{\hat{\sigma}_B^2} \le \exp\left(\frac{\log n}{2n}\right) \right\}$$

$$= \Pr\left\{ 1 + \frac{t^2}{2(n-1)} \le \exp\left(\frac{\log n}{2n}\right) \right\}$$

$$= \Pr\left\{ t^2 \le 2(n-1)\left[\exp\left(\frac{\log n}{2n}\right) - 1\right] \right\}. \qquad (2.8)$$

Equation (2.8) gives the probability of the BIC selecting model A when model A is the true model. It follows from (2.8) that this probability is one for large n since,

$$\lim_{n \to \infty} 2(n-1)\left[\exp\left(\frac{\log n}{2n}\right) - 1\right] = \infty.$$

Hence the BIC is consistent.

Case 2. If model B is the true model, then it can be easily shown as in (2.7) and (2.8) that if AIC is used for model selection, then

$$\Pr\left\{ \mathrm{AIC_A} \le \mathrm{AIC_B} \mid \mathrm{B} \right\}$$

$$= \Pr\left\{ t'^2\left(\frac{\theta_o}{\sigma}\sqrt{\frac{n}{2}}\right) \le 2(n-1)\left[\exp\left(\frac{1}{n}\right) - 1\right] \right\}, \qquad (2.9)$$

where t' is the noncentral t-distribution with $2(n-1)$ degrees of freedom and the noncentrality parameter is given as $(n\theta_o^2/2\sigma^2)^{1/2}$; $\theta_o = \mu_x - \mu_y$. Similarly, if BIC is used for model selection, then

$$\Pr\left\{ \mathrm{BIC_A} \le \mathrm{BIC_B} \mid \mathrm{B} \right\}$$

$$= \Pr\left\{ t'^2\left(\frac{\theta_o}{\sigma}\sqrt{\frac{n}{2}}\right) \le 2(n-1)\left[\exp\left(\frac{\log n}{2n}\right) - 1\right] \right\}. \qquad (2.10)$$

From equations (2.7) to (2.10), the exact probabilities for the AIC and BIC can be tabulated. Table 2.1 gives the probability of selecting model A for various combinations of n and θ_o. The IMSL subroutine named MDTN was used to calculate the probabilities of the noncentral-t distribution. This table shows that when $\theta_o = 0$ and $n > 5$, AIC has lower probability of selecting model A than the BIC; however, for $n < 5$, the probability of BIC selecting model A is lower than the AIC. As n increases, both AIC and BIC attain their asymptotic values. When $\theta_o > 0$, the BIC has higher probability of selecting model A than the AIC, indicating that the BIC has a smaller chance of selecting the true model than the AIC. For $n < 5$, the AIC provides higher probability than the BIC. As θ_o increases, it is seen that the probabilities decrease and that each criterion selects the true model virtually every time.

TABLE 2.1

Probability of selecting model A

n		θ_o 0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
5	AIC	0.780	0.759	0.700	0.610	0.502	0.387	0.280	0.190	0.119	0.070	0.038
	BIC	0.729	0.707	0.645	0.553	0.444	0.333	0.233	0.152	0.092	0.051	0.027
10	AIC	0.814	0.773	0.660	0.502	0.336	0.195	0.098	0.042	0.015	0.005	0.001
	BIC	0.844	0.806	0.698	0.543	0.374	0.226	0.118	0.053	0.020	0.006	0.002
20	AIC	0.829	0.749	0.543	0.306	0.129	0.040	0.009	0.001			
	BIC	0.906	0.844	0.666	0.426	0.210	0.077	0.020	0.004			
50	AIC	0.837	0.648	0.276	0.056	0.005	0.000					
	BIC	0.949	0.831	0.489	0.154	0.022	0.001					
100	AIC	0.840	0.495	0.078	0.002							
	BIC	0.967	0.766	0.248	0.019							
200	AIC	0.841	0.278	0.005								
	BIC	0.978	0.618	0.045								
500	AIC	0.842	0.001									
	BIC	0.991	0.033									

2.3 THE RISK OF AIC AND BIC

The exact probabilities derived in section 2 can be used to determine which criterion provides the smaller risk.

Consider again the situation where two random samples, x and y, each of size n are available. In model A, $\theta_o = 0$ was hypothesized where $\theta_o = \mu_x - \mu_y$. In model B, $\theta_o \neq 0$ was hypothesized. Let $\hat{\theta}$ be the estimator of θ_o. Then the maximum likelihood estimators are as follows:

$$\hat{\mu}_x = \bar{x} \quad \text{if model B is selected,}$$

$$\hat{\mu}_y = \bar{y} \quad \text{if model B is selected,}$$

$$\hat{\mu}_x = \hat{\mu}_y = \frac{\bar{x} + \bar{y}}{2} \quad \text{if model A is selected.}$$

Hence, $\hat{\theta} = \bar{x} - \bar{y}$ if model B is selected and $\hat{\theta} = 0$ if model A is selected.

The mean square error (MSE) of the estimator, $R(\hat{\theta})$, is defined as

$$R(\hat{\theta}) = \langle(\hat{\theta} - \theta_o)^2\rangle$$

$$= \langle(\hat{\theta} - \theta_o)^2 \mid \mathrm{A}\rangle P_{\mathrm{A}} + \langle(\hat{\theta} - \theta_o)^2 \mid \mathrm{B}\rangle P_{\mathrm{B}}, \qquad (2.11)$$

where $\langle\cdot\rangle$ denotes the mathematical expectation. The right hand side of (2.11) is a conditional statement, the expectation in the first term denotes the MSE conditional on model A being selected and P_{A} denotes the probability that model A is selected. Similarly, the second term is interpreted as the MSE conditional on model B being selected and P_{B} denotes the probability that model B is selected. Note that $R(\hat{\theta})$ is also the risk of $\hat{\theta}$ under squared error loss function, hence the two terms can be used interchangeably. From here on, it will be refered to as the risk. The expectations in (2.11) are:

$$\langle(\hat{\theta} - \theta_o)^2 \mid \mathrm{A}\rangle = \theta_o^2,$$

and

$$\langle(\hat{\theta} - \theta_o)^2 \mid \mathrm{B}\rangle = \langle(\bar{x} - \bar{y} - \theta_o)^2\rangle$$

$$= \langle(\bar{x} - \mu_x)^2\rangle + \langle(\bar{y} - \mu_y)^2\rangle$$

$$= \frac{2\sigma^2}{n}.$$

The general expression of (2.11) is then given by,

$$R(\hat{\theta}) = \theta_o^2 P_{\mathrm{A}} + \frac{2\sigma^2}{n} P_{\mathrm{B}}. \tag{2.12}$$

Thus (2.12) provides the risk of the estimator $\hat{\theta}$ depending on the selection criterion used since P_{A} and P_{B} are obtained depending on the criterion used. Note that $R(\hat{\theta})$ given in (2.12) is equivalent (see appendix A2.1) to the risk function $\bar{R}(\hat{\mu}_{\mathrm{A}}, \hat{\mu}_{\mathrm{B}})$ given by

$$\bar{R}(\hat{\mu}_x, \hat{\mu}_y) = \langle(\hat{\mu}_x - \mu_x)^2 + (\hat{\mu}_y - \mu_y)^2\rangle.$$

The point of intersection for $R(\hat{\theta})$ when using AIC and BIC can be obtained by equating $R(\hat{\theta} \mid \mathrm{A})$ and $R(\hat{\theta} \mid \mathrm{B})$ where $R(\hat{\theta} \mid \mathrm{A})$ denotes the risk of $\hat{\theta}$ when the AIC is used for model selection and $R(\hat{\theta} ; \mathrm{B})$ denotes the risk of $\hat{\theta}$ when BIC is used. Solving this equation immediately gives $\theta_o = 2\sigma^2/n$.

The values of the $R(\hat{\theta})$ were obtained and are plotted in Figures 2.1 to 2.4 with plots pertaining to different n's, $n = 5,\ 10,\ 20,\ 30$, with values of θ_o ranging from 0 to 2; σ^2 was set at 1.0. These figures clearly indicate neither criterion dominates the other. For $n > 5$ and θ_o less than $2\sigma^2/n$, the BIC is seen to give smaller risk, otherwise, BIC provides larger risk. However, for $n \leq 5$ and $\theta_o < 2\sigma^2/n$, the AIC gives smaller risk, otherwise, AIC gives larger risk. These figures also show that the risk decreases as sample size is increased and that as θ_o becomes large, both criteria have the same risk.

To examine the behaviour when $\theta_o = 0$, consider the asymptotic risk of $\hat{\theta}$ given by,

$$R_{asy}(\hat{\theta} \mid \theta_o = \theta) = \lim_{n \to \infty} \left\{ n \left[\theta^2 P_{\mathrm{A}} + \frac{2\sigma^2}{n} P_{\mathrm{B}} \right] \right\}.$$

Suppose that the AIC is used for model selection then,

$$R_{asy}(\hat{\theta} \mid \theta_o = 0) = 2\sigma^2 \lim_{n \to \infty} \Pr(\mathrm{AIC_B} \leq \mathrm{AIC_A} \mid \theta_o = 0)$$

$$= 2\sigma^2 c, \qquad (2.13)$$

where c is some constant, $c > 0$. If $\theta_o = \theta \neq 0$,

$$R_{asy}(\hat{\theta} \mid \theta_o \neq 0) = \theta^2 \lim_{n\to\infty} \left\{ n \Pr(\mathrm{AIC_A} \le \mathrm{AIC_B} \mid \theta_o = \theta) \right\}$$
$$+ 2\sigma^2 \lim_{n\to\infty} \left\{ 1 - \Pr(\mathrm{AIC_A} \le \mathrm{AIC_B} \mid \theta_o = \theta) \right\}$$

$$= 2\sigma^2. \tag{2.14}$$

In the case of BIC, if $\theta_o = 0$,

$$R_{asy}(\hat{\theta} \mid \theta_o = 0) = 2\sigma^2 \lim_{n\to\infty} \Pr(\mathrm{BIC_B} \le \mathrm{BIC_A} \mid \theta_o = 0)$$

$$= 0, \tag{2.15}$$

and if $\theta_o = \theta \neq 0$,

$$R_{asy}(\hat{\theta} \mid \theta_o \neq 0) = \theta^2 \lim_{n\to\infty} \left\{ n \Pr(\mathrm{BIC_A} \le \mathrm{BIC_B} \mid \theta_o = \theta) \right\}$$
$$+ 2\sigma^2 \lim_{n\to\infty} \left\{ 1 - \Pr(\mathrm{BIC_A} \le \mathrm{BIC_B} \mid \theta_o = \theta) \right\}$$

$$= 2\sigma^2. \tag{2.16}$$

Note that $n \Pr(\mathrm{AIC_A} \le \mathrm{AIC_B} \mid \theta_o \neq 0)$ and $n \Pr(\mathrm{BIC_A} \le \mathrm{BIC_B} \mid \theta_o \neq 0)$ are both zero in the limit as $n \to \infty$. This can be shown by centering the noncentral-t (Johnson and Kotz, 1970, p. 204) with the noncentrality parameter and by applying the Chebyshev's inequality.

FIGURE 2.1

MSE OF AIC AND BIC N=6

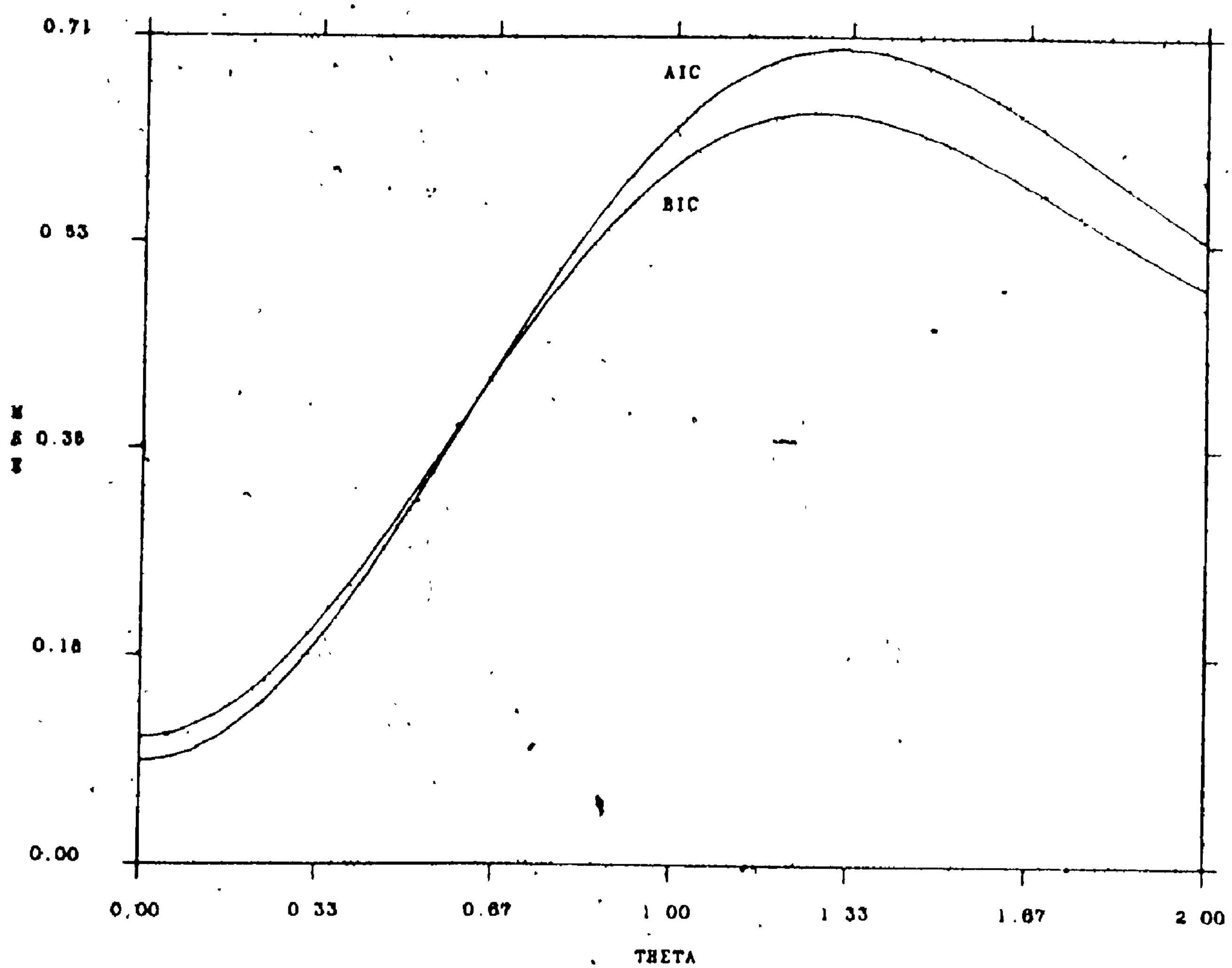


FIGURE 2.2

MSE OF AIC AND BIC: N=10

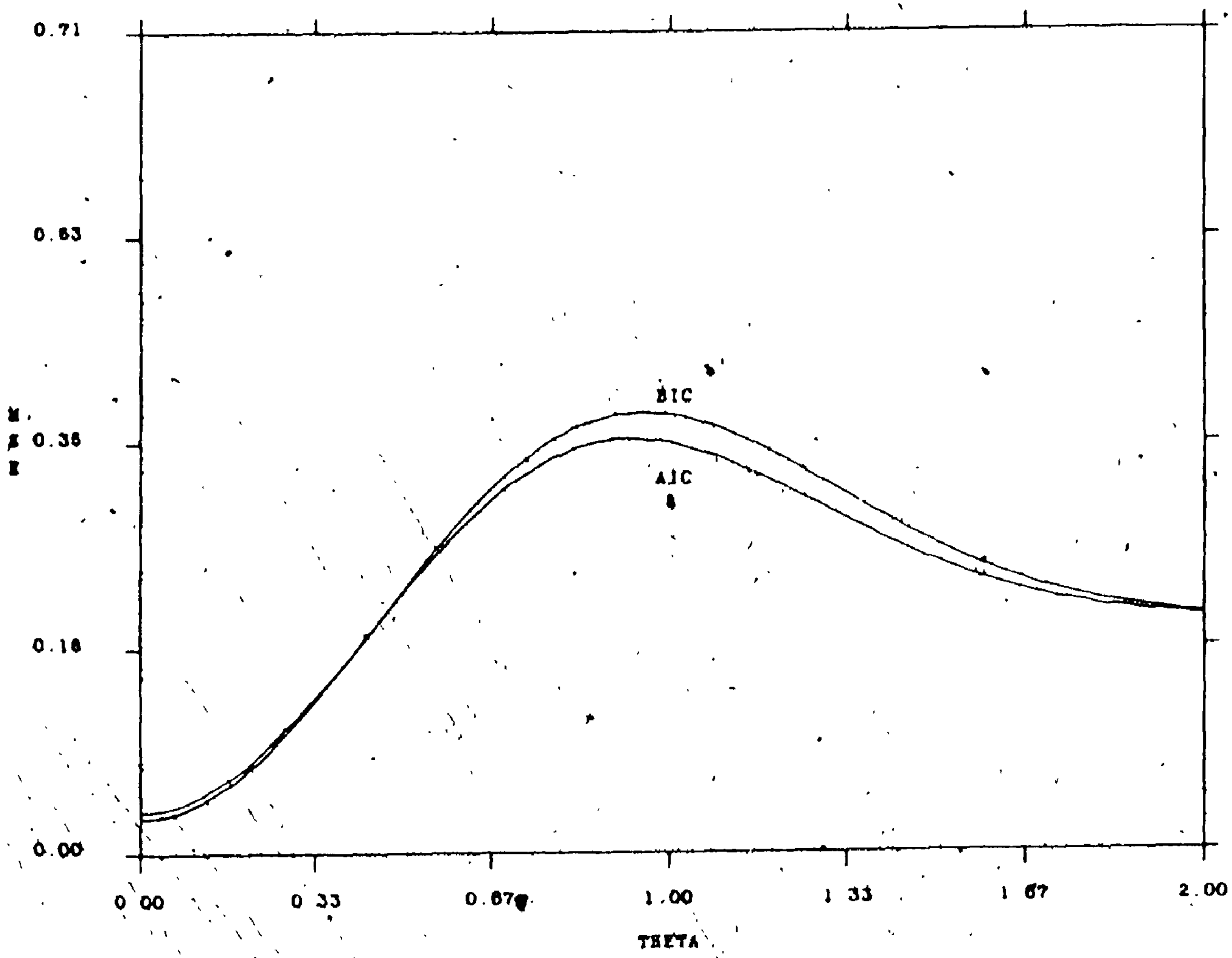


FIGURE 2.3

MSE OF AIC AND BIC: N=20.

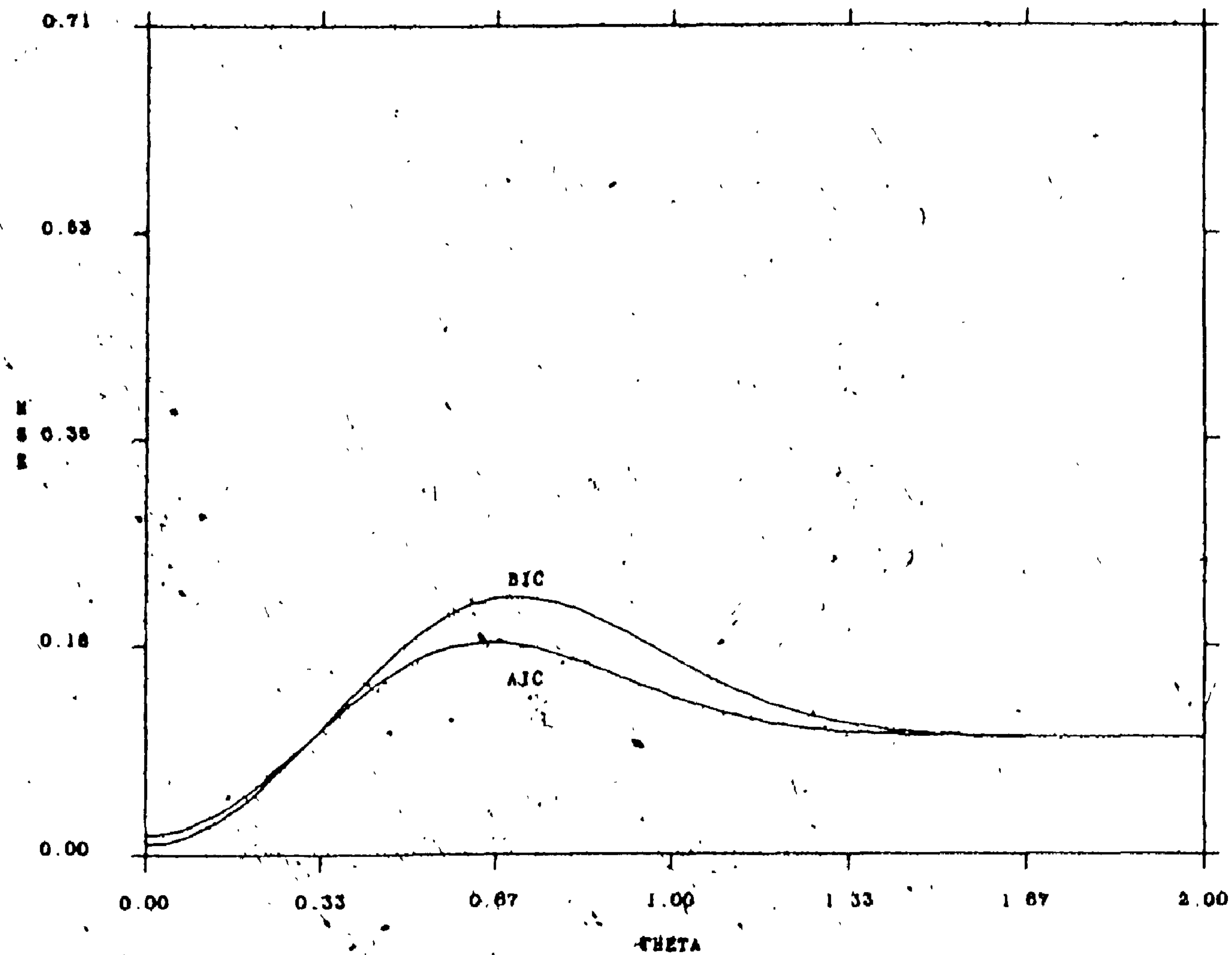
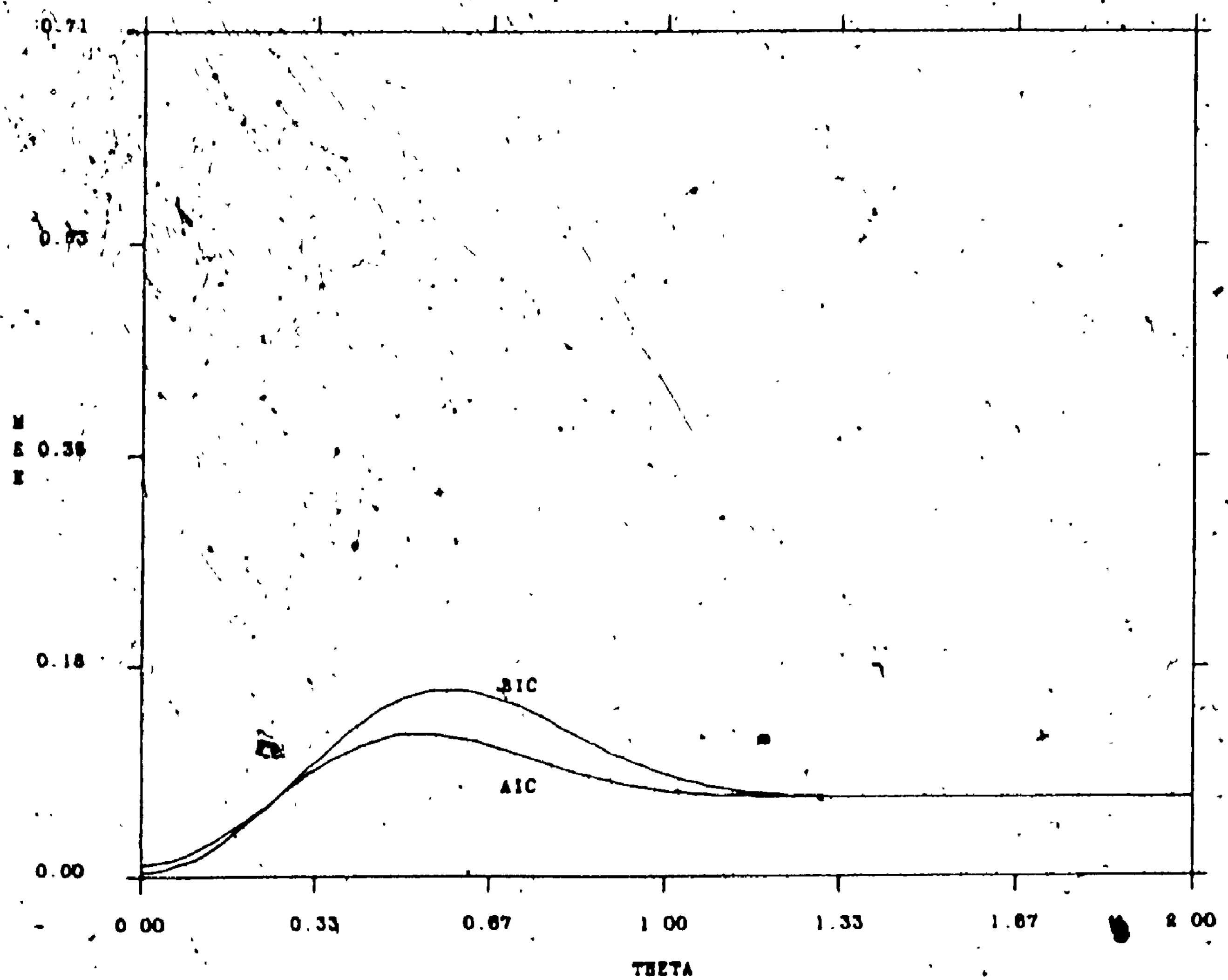


FIGURE 2.4

MSE OF AIC AND BIC. N=30



2.4 EMPIRICAL MEAN SQUARE ERROR OF AIC AND BIC FOR AUTOREGRESSIVE PROCESSES

Thus far our investigations have been restricted to the two sample problem, but at least some idea of how the two criteria behave in small samples is now available. How they behave in small samples when observations are correlated is a difficult problem. This section presents some empirical evidence regarding the risks of AIC and BIC in large samples.

An AR(1) time series with mean zero and innovation variance $\sigma_a^2 = 1.0$ was generated using the algorithm described in McLeod and Hipel (1978). The random number generator Super Duper (Marsaglia, 1976), in conjunction with the Box-Mueller method, was used to generate the $NID(0,1)$ variates. The sample size, n, was set successively at 50, 100, 200 and the autoregressive parameter, ϕ, was set successively at 0.0, ±0.1, ±0.2, ±0.3, ±0.4, ±0.5, ±0.6, ±0.7, ±0.8, ±0.9. For each of the 10,000 replications, autoregressions of various orders were selected according to the two criteria, AIC and BIC. The order of the fitted autoregression has an upper bound K, where K was set successively at 5, 10, 15. The mean square error of the one-step ahead forecast (MSE) in each replication is calculated. The MSE is given by,

$$\mathrm{MSE} = \sum_{i=0}^{p}\sum_{j=0}^{p} \hat{\phi}_i \hat{\phi}_j \gamma_{|i-j|}, \tag{2.17}$$

where p is the order of the fitted autoregression; $\gamma_k = \phi^k$ and $\hat{\phi}_i$ is the estimate of ϕ_i of an AR(p) process and $\hat{\phi}_0 = -1$. For the derivation of MSE, refer to appendix A2.2. It may also be shown that the average MSE is equivalent to the risk function,

$$R(\tilde{\beta}) = \langle(\hat{\beta} - \beta)' V_{\hat{\beta}}^{-1} (\hat{\beta} - \beta)\rangle,$$

where $V_{\hat{\beta}}$ is the covariance matrix of $\hat{\beta}$;

$$\hat{\beta} = \begin{cases} (\hat{\mu}, \hat{\phi}_1, \ldots, \hat{\phi}_K) & \text{if the mean is estimated,} \\ (\hat{\phi}_1, \ldots, \hat{\phi}_K) & \text{if the mean is assumed known.} \end{cases}$$

If an AR(p) with $p < K$ is selected, then $\hat{\phi}_{p+1} = \cdots = \hat{\phi}_K = 0$.

Figures 2.5 to 2.13 contain different graphs of the average MSE. In each figure, slightly different behaviour of the MSE is shown for various n or K. These figures show that as the sample size increases, the average MSE decreases and the difference in average MSE between AIC and BIC becomes small. However for all n, the BIC appears to dominate the AIC unlike the situation in the two sample problem where for some parameter value, θ_o, the AIC was better. These figures also show that if K is set too large, then AIC yields a larger MSE. However this is not the case for the BIC, which remains approximately the same level for all K. This is perhaps due to the inconsistency property of the AIC. The heading IFMEAN=1 indicates that the sample mean of the time series is subtracted while calculating the sample autocorrelation functions. For $n = 50$, these figures show some abnormal behaviour at $\phi = 0.9$; that is, it shows higher MSE than do the rest of the parameters.

Figures 2.14 to 2.22 are basically the same as figures 2.5 to 2.13 with the exception that the sample mean of the time series is not subtracted (IFMEAN=0) while calculating the sample autocorrelation functions; this is equivalent to assuming either the mean is zero or the true mean is known and has been subtracted from the series. The latter figures are more symmetric and easier to interpret. The interpretation of these plots are similar to the above but they also illustrate the importance of the sample mean.

FIGURE 2.5

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=50: K=5: IFMEAN=1: NUMBER OF REPLICATIONS=10000

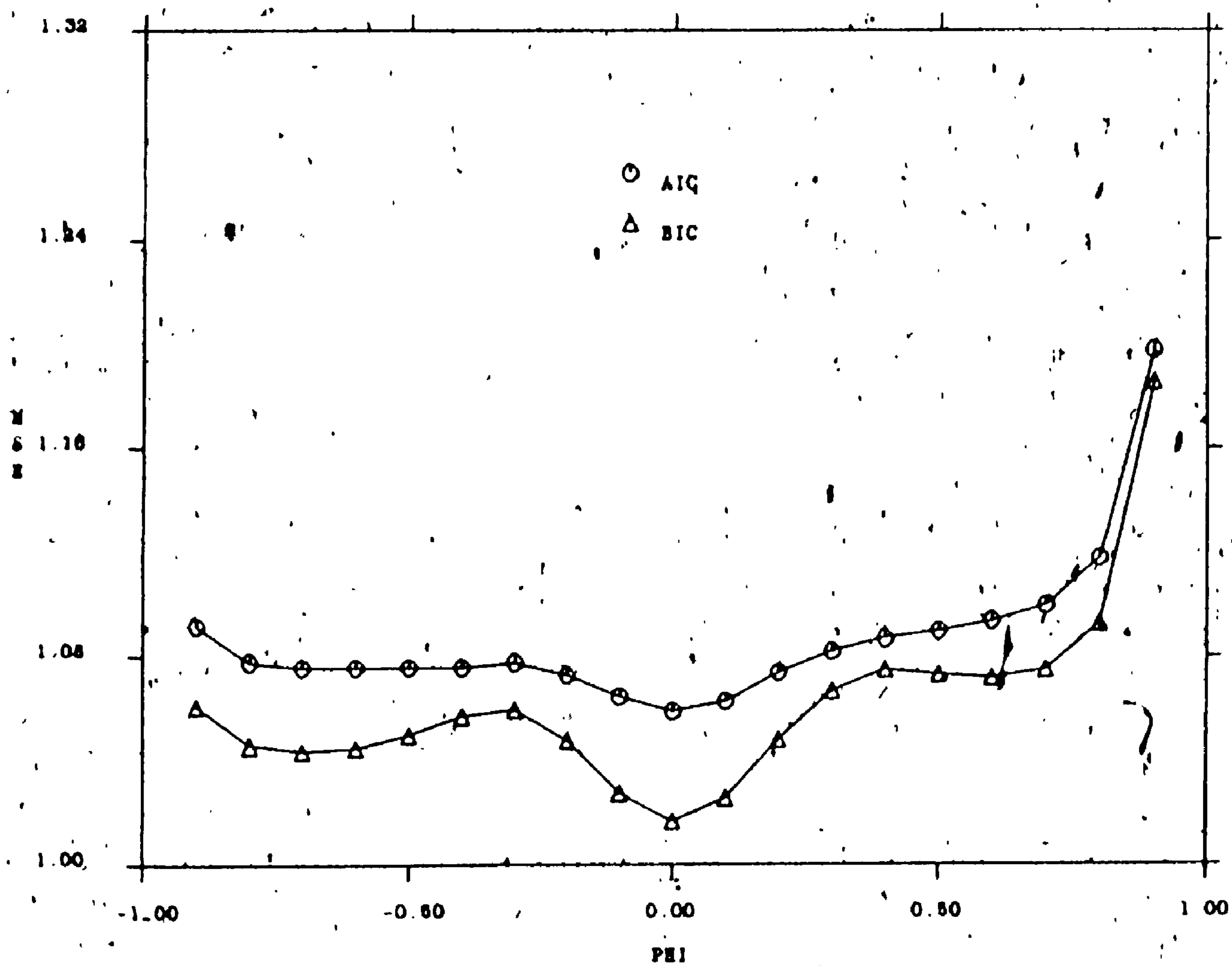


FIGURE 2.6

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=50; K=10; IFMEAN=1. NUMBER OF REPLICATIONS=10000.

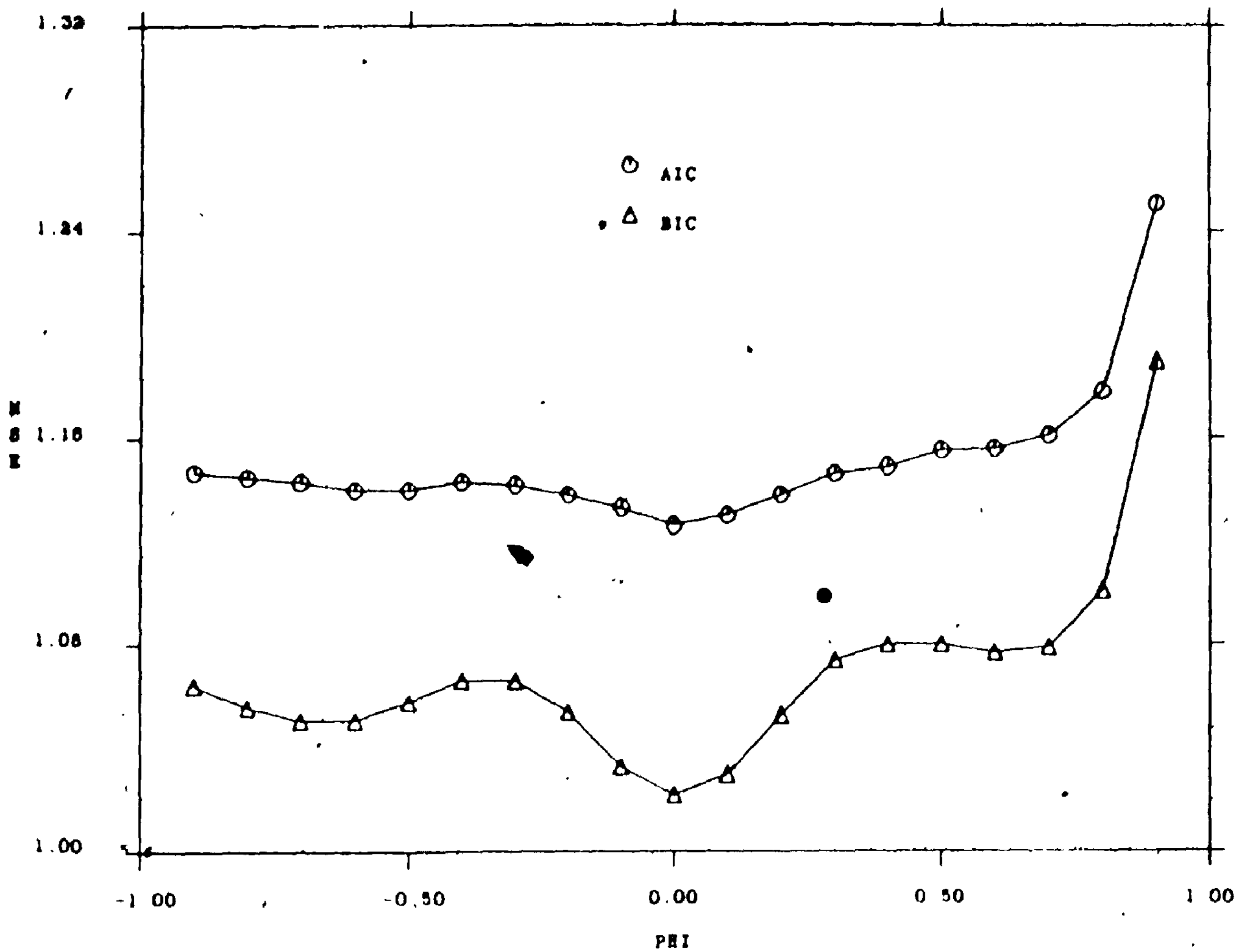


FIGURE 2.7

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=50; K=16; IFMEAN=1. NUMBER OF REPLICATIONS=10000

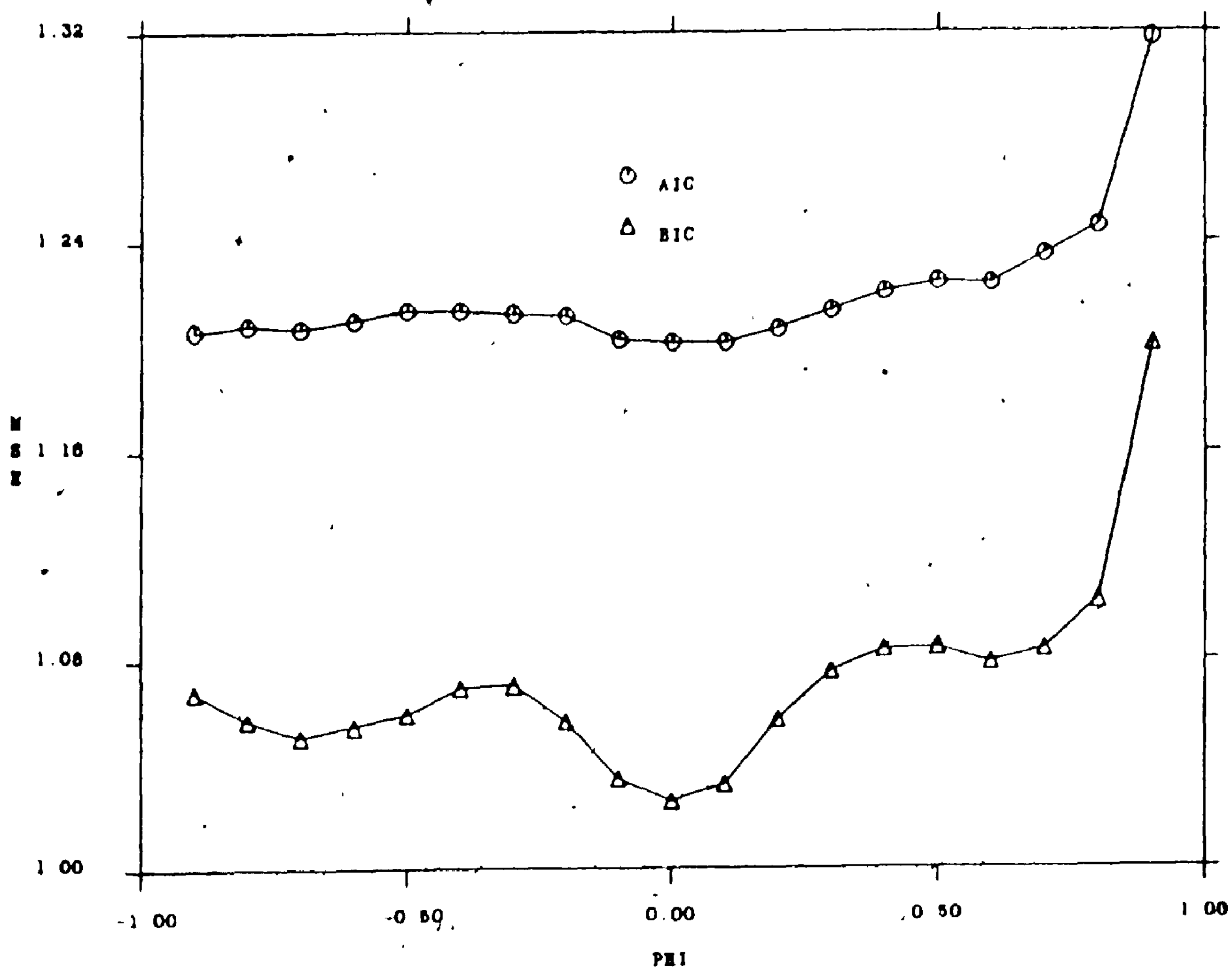


FIGURE 2.8

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=100; K=5; IFMEAN=1: NUMBER OF REPLICATIONS=10000.

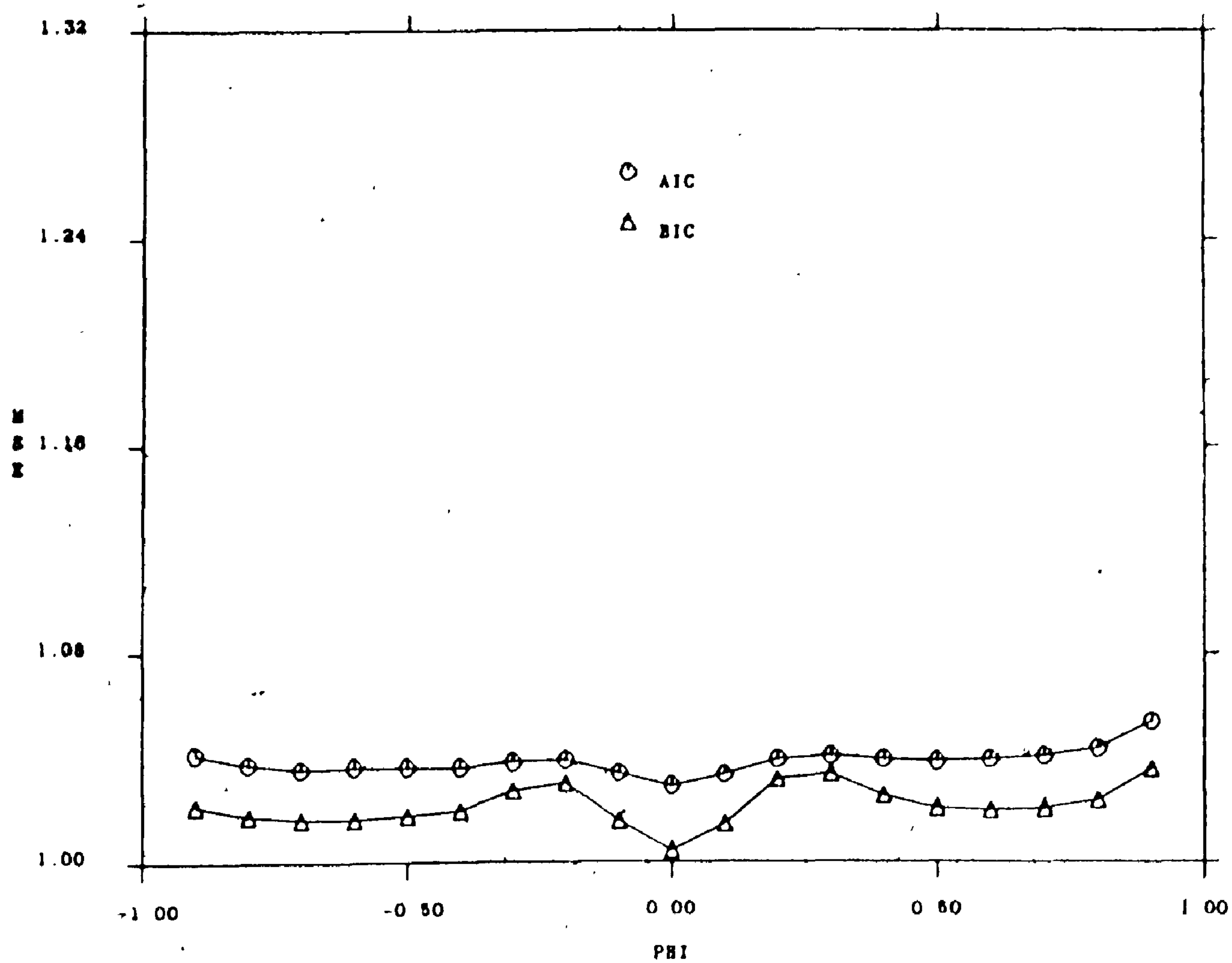


FIGURE 2.9

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=100: K=10: IFMEAN=1: NUMBER OF REPLICATIONS=10000

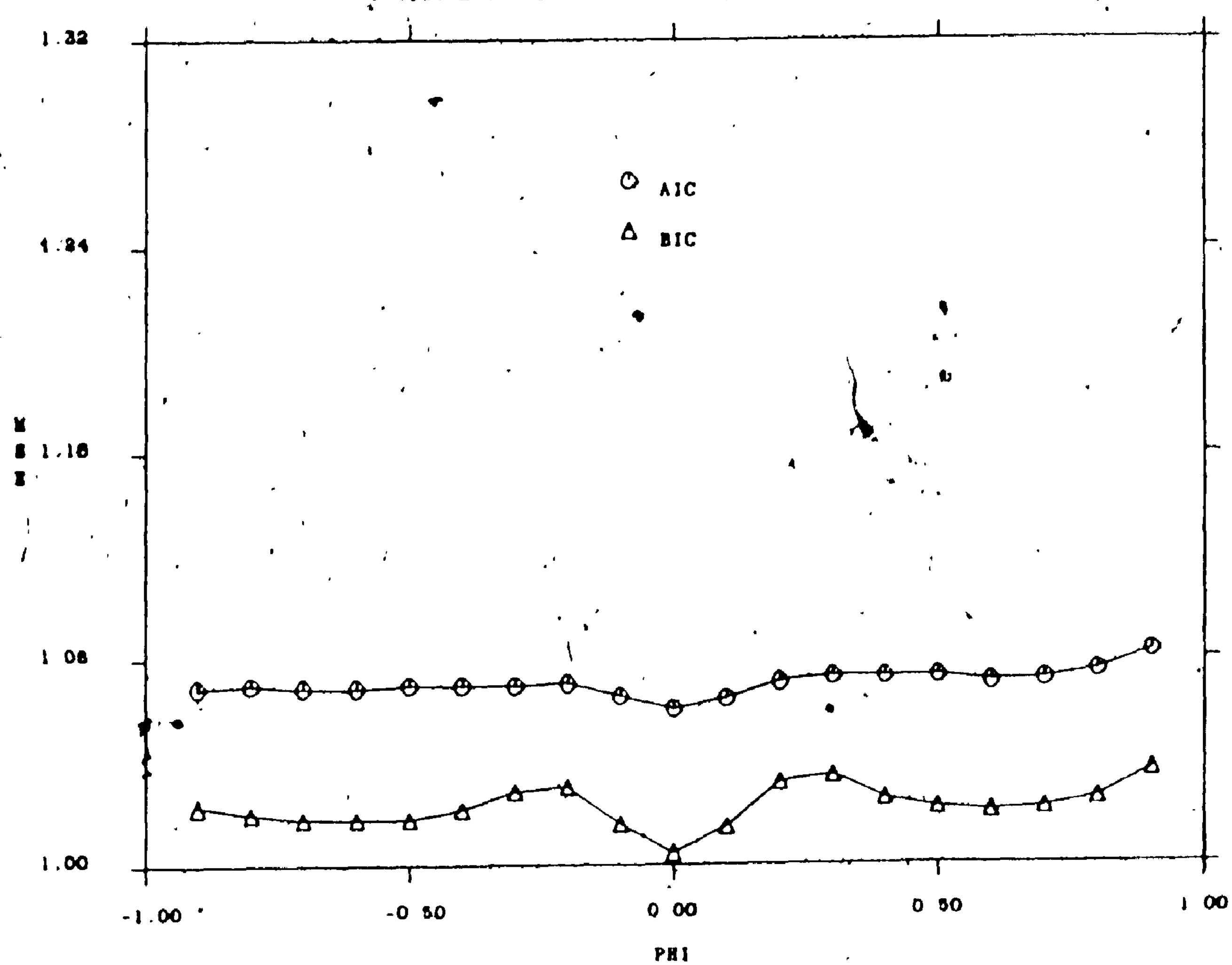


FIGURE 2.10

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=100: K=15: IFMEAN=1: NUMBER OF REPLICATIONS=10000

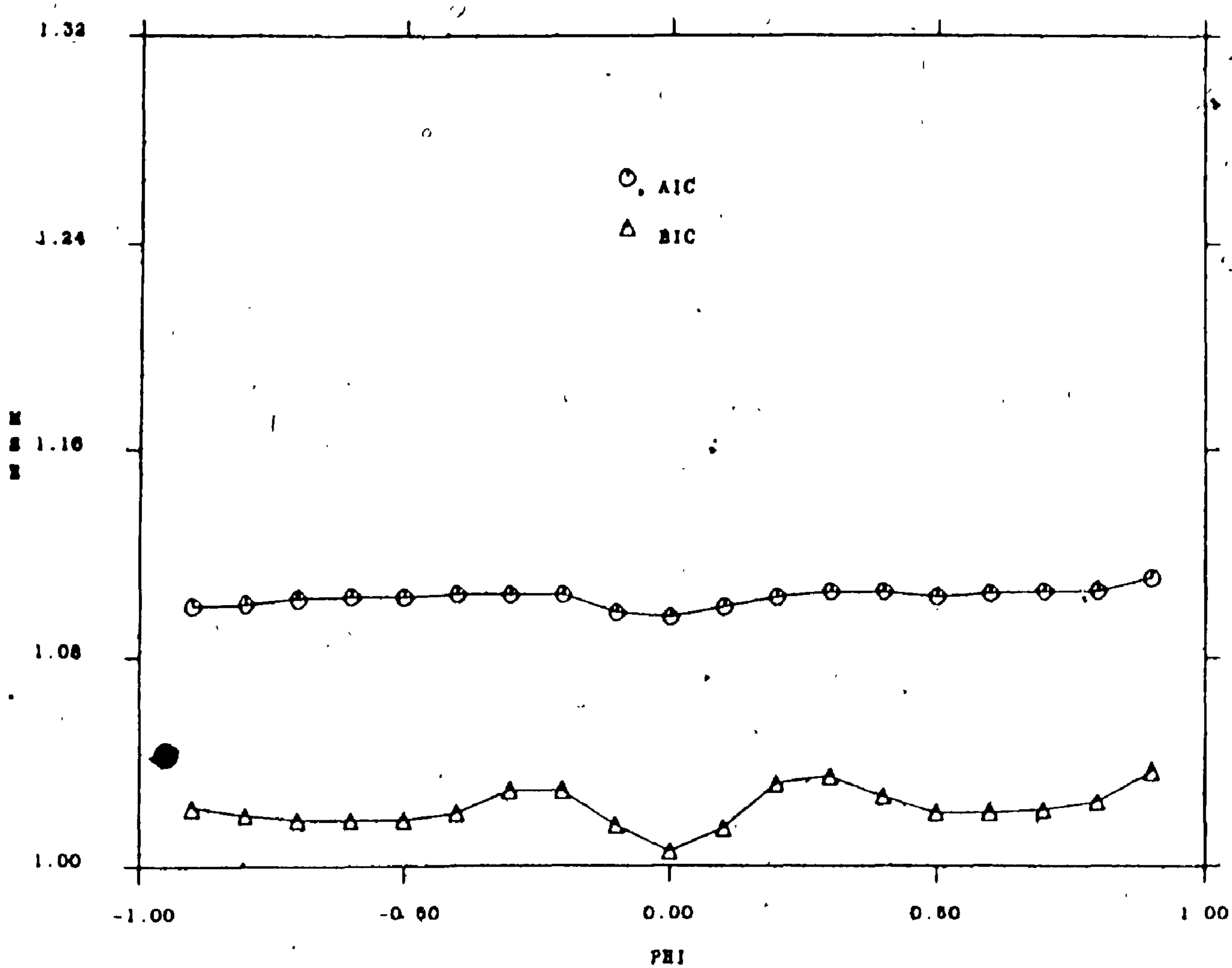


FIGURE 2.11

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=200: K=6: IFMEAN=1: NUMBER OF REPLICATIONS=10000

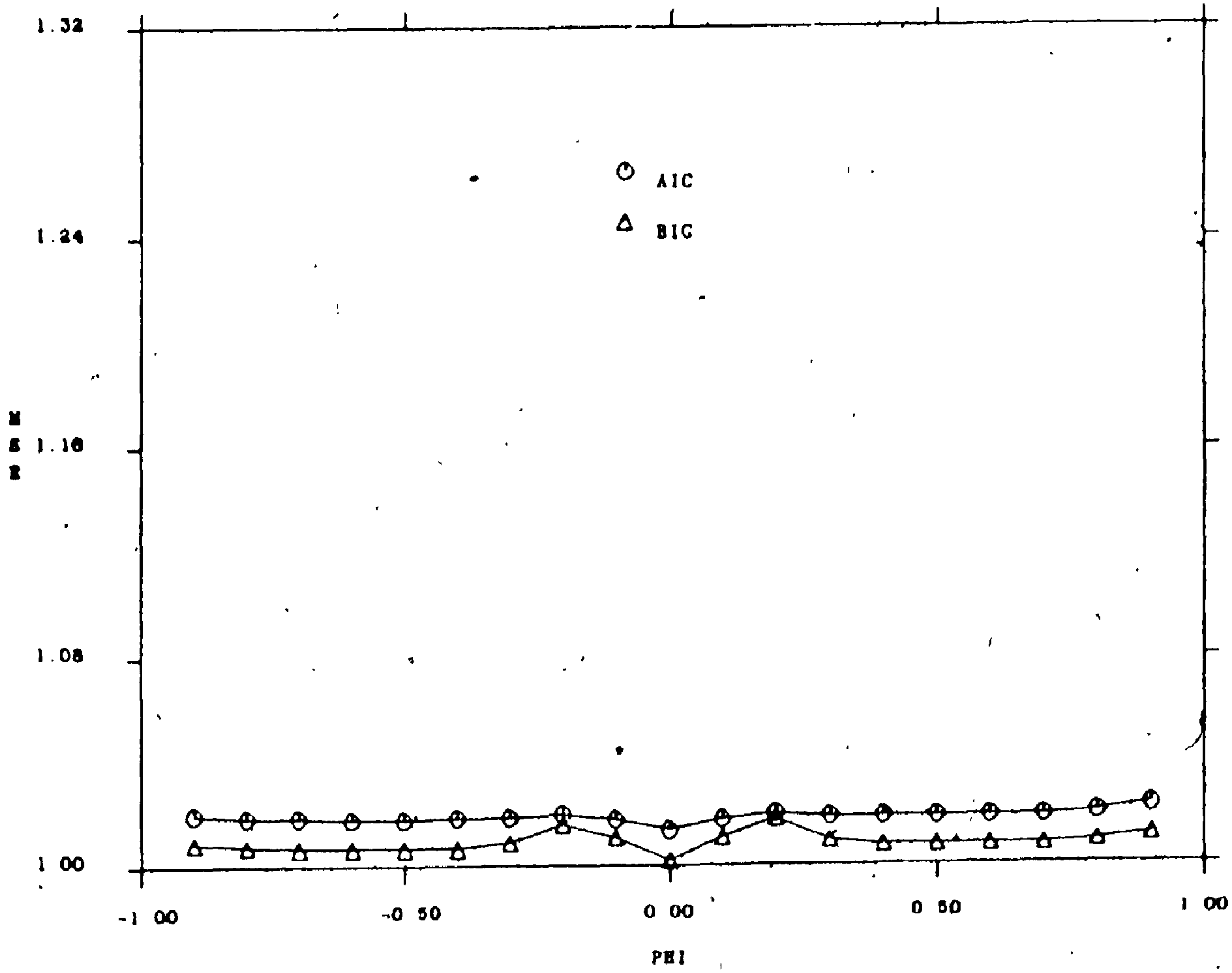


FIGURE 2.12

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=200: K=10: IFMEAN=1: NUMBER OF REPLICATIONS=10000

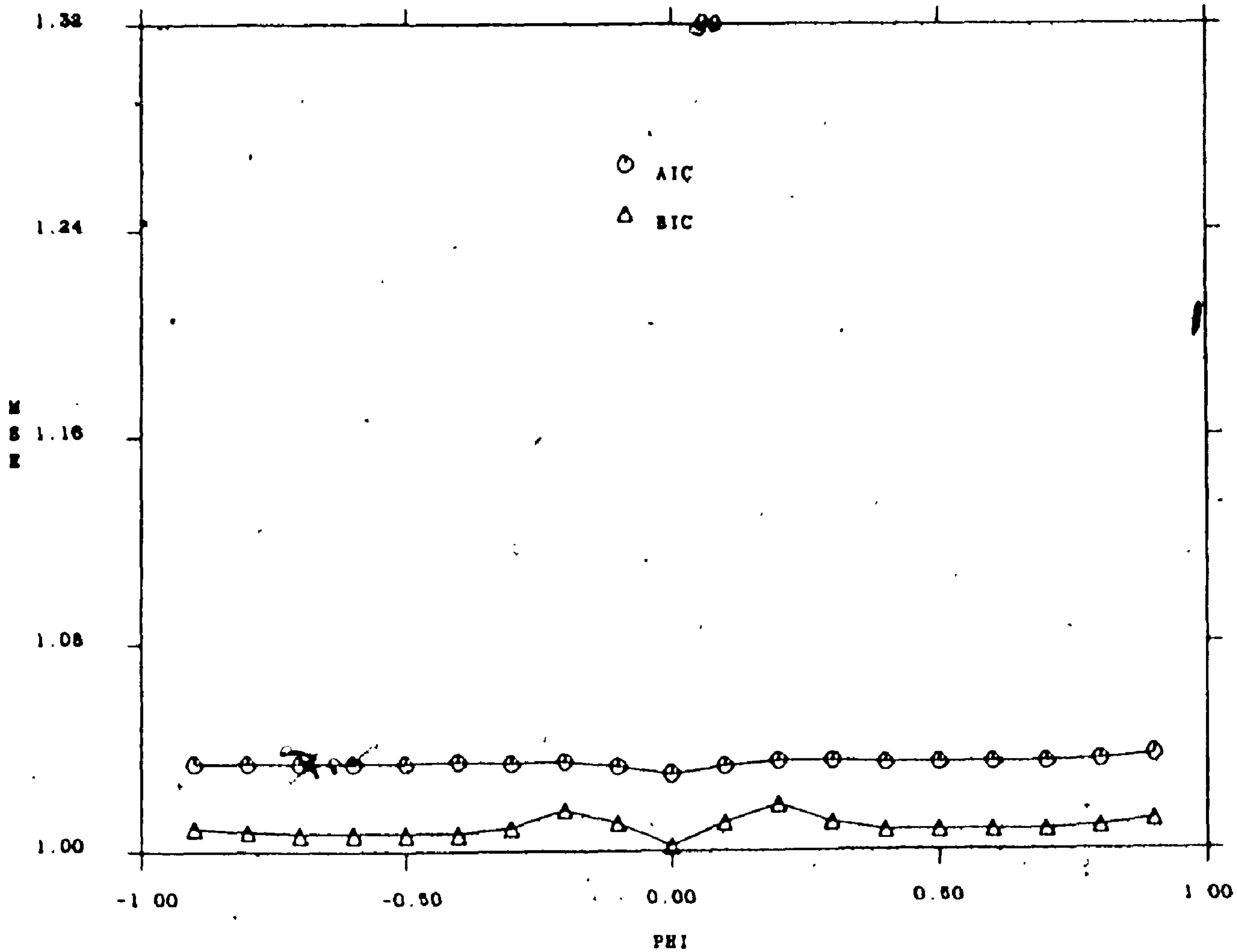


FIGURE 2.13

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=200: K=15: IFMEAN=1: NUMBER OF REPLICATIONS=10000

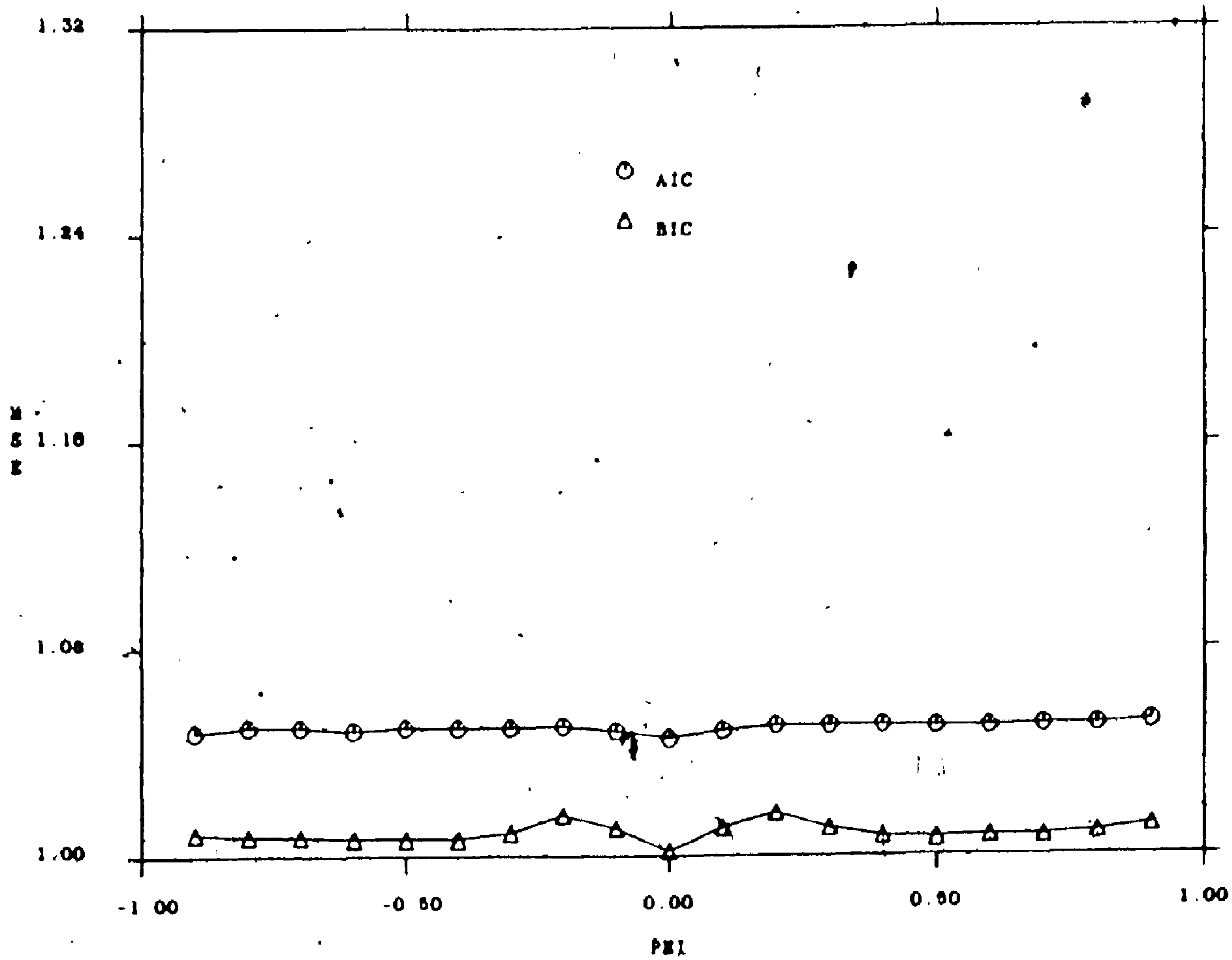


FIGURE 2.14

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N-50: K-5: IFMEAN-0: NUMBER OF REPLICATIONS-10000

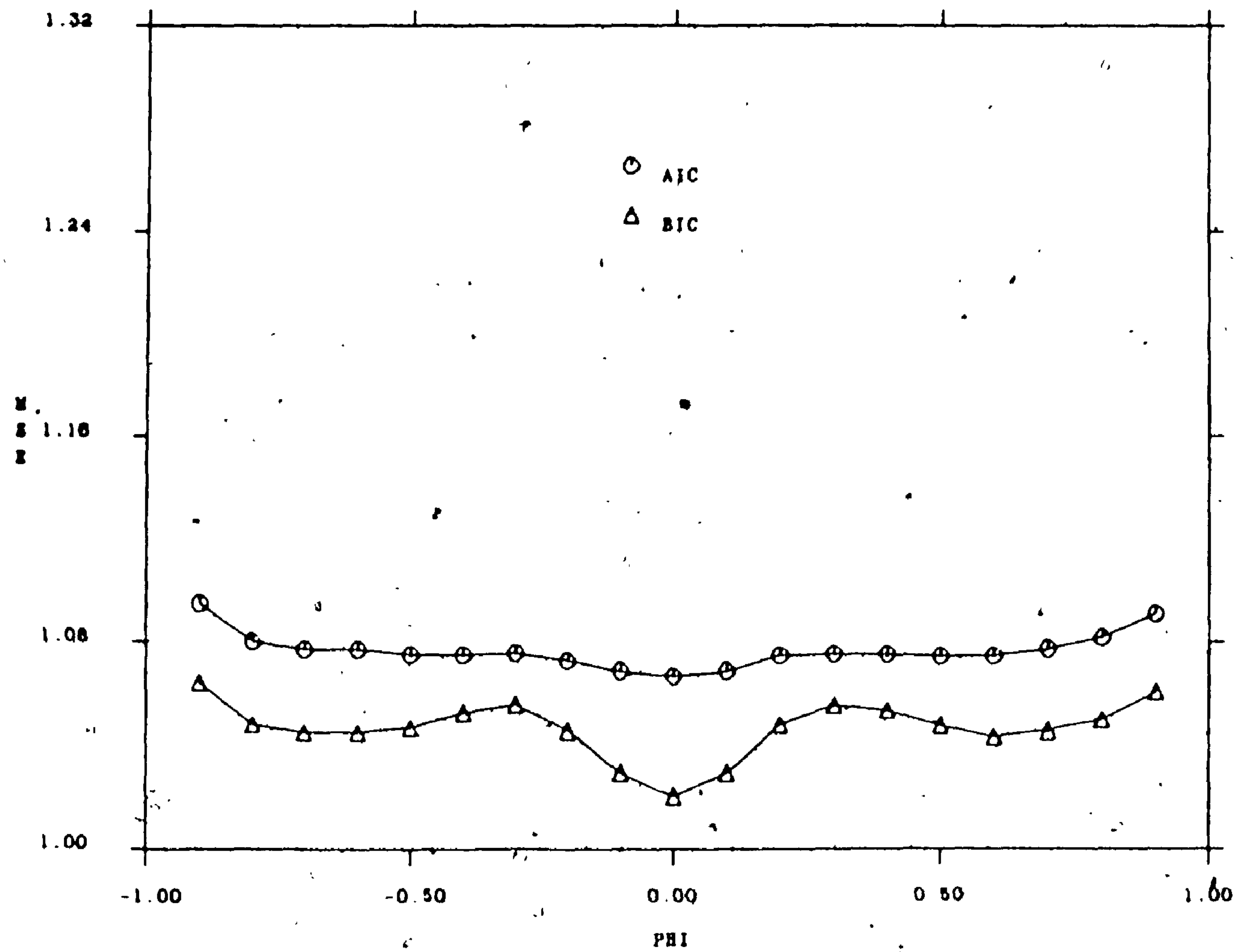


FIGURE 2.15

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=50. K=10: IFMEAN=0: NUMBER OF REPLICATIONS=10000

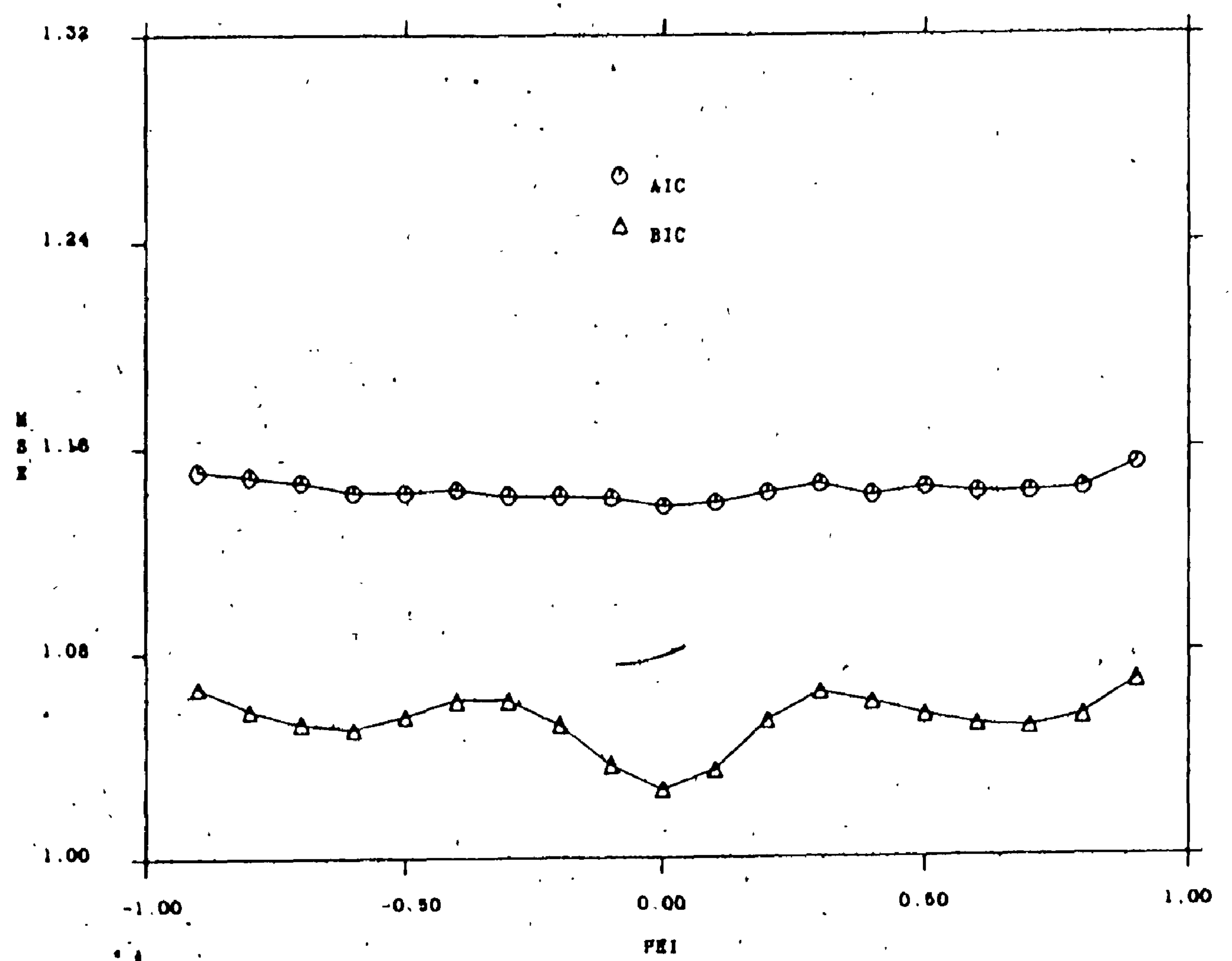


FIGURE 2.16

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=50; K=16; IFMEAN=0; NUMBER OF REPLICATIONS=10000

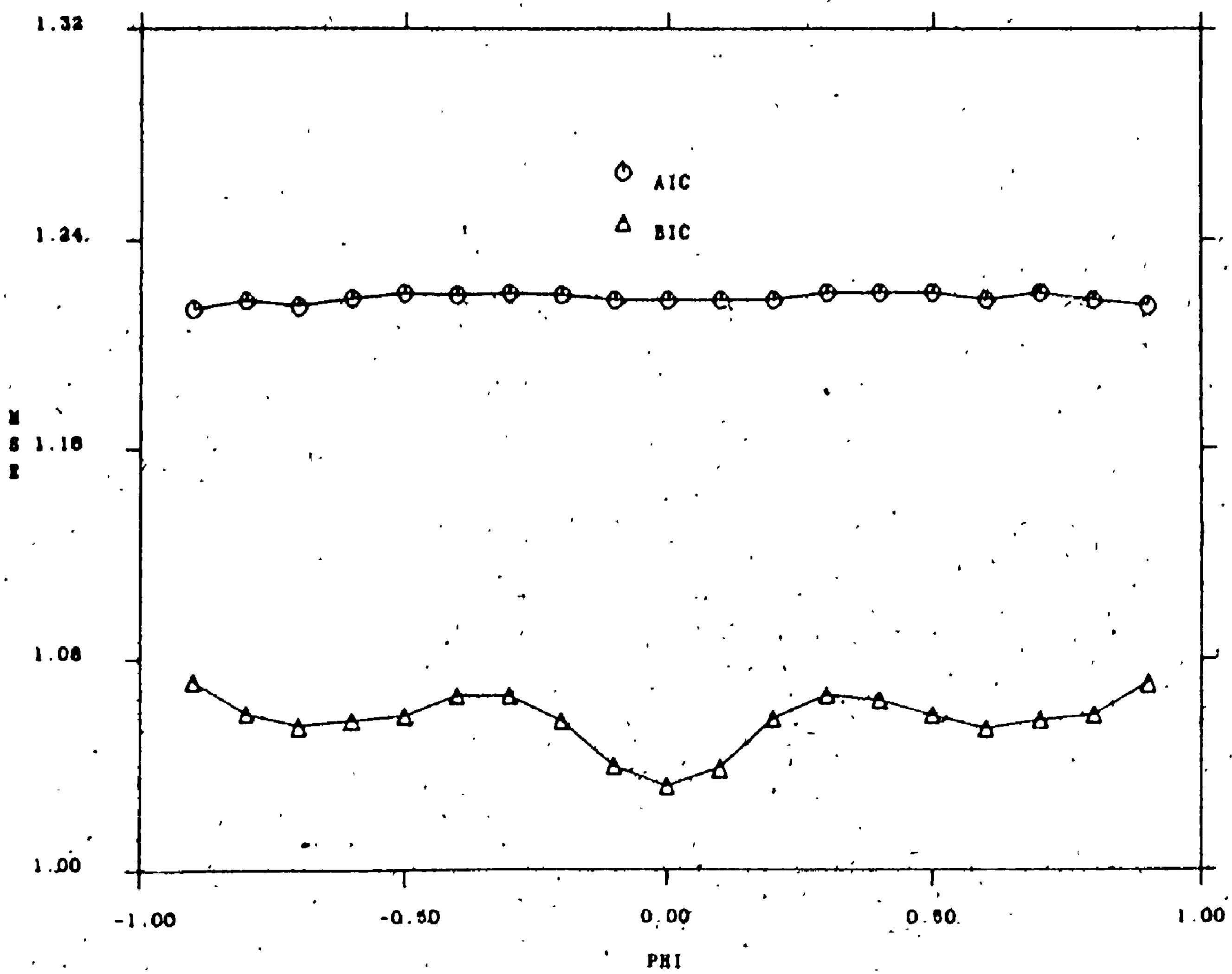


FIGURE 2.17

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=100: K=6: IFMEAN=0: NUMBER OF REPLICATIONS=10000

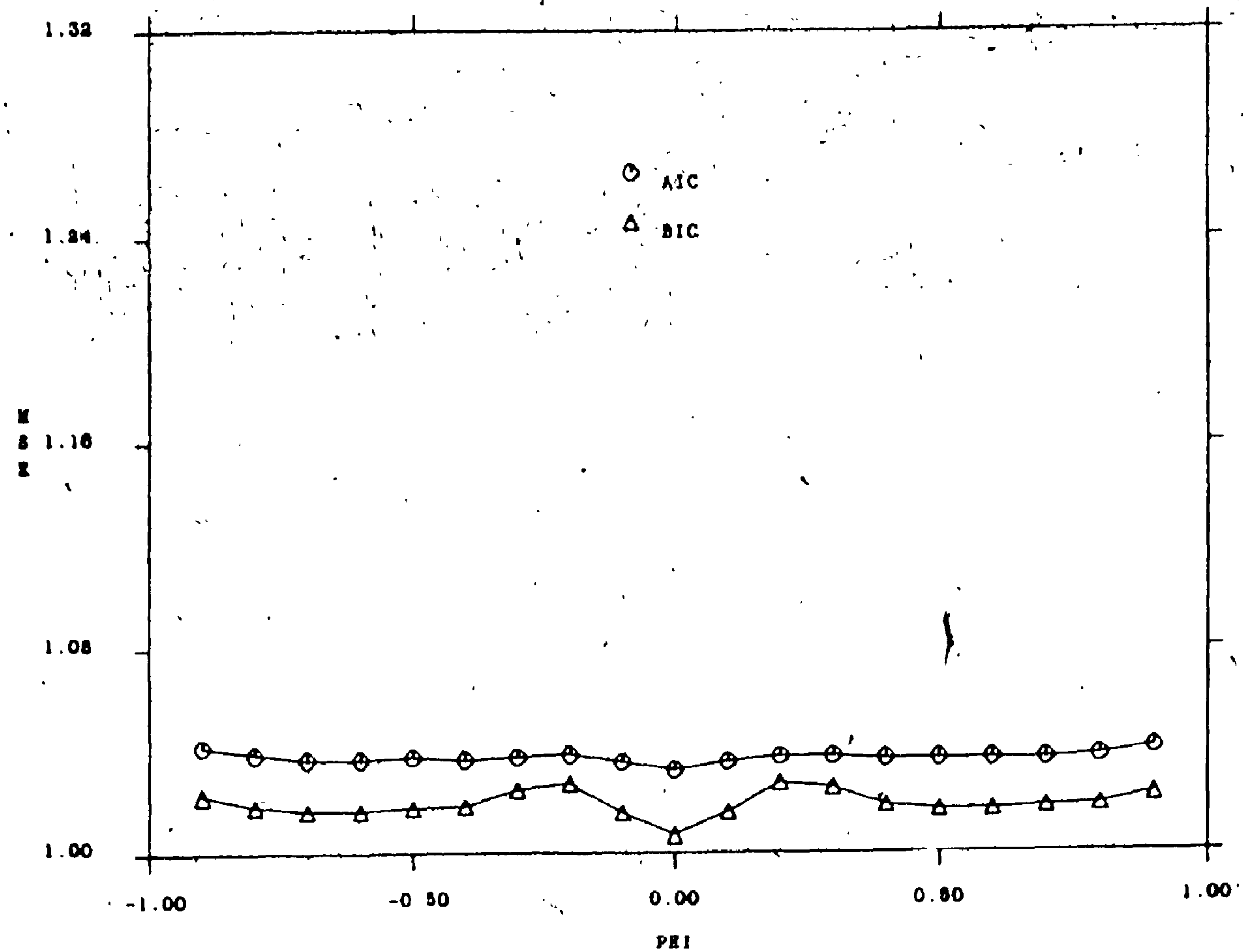


FIGURE 2.18

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=100; K=10; IFMEAN=0; NUMBER OF REPLICATIONS=10000

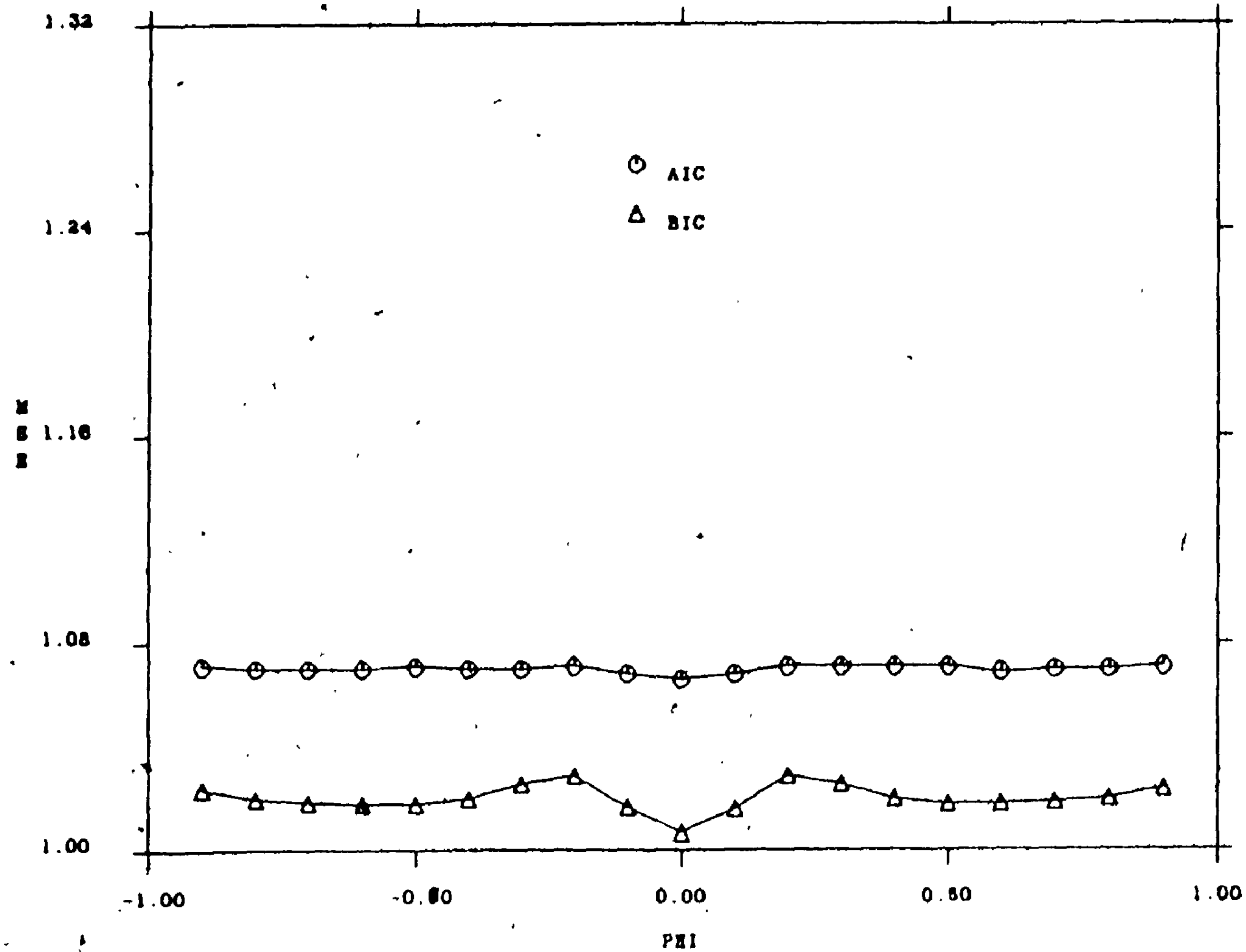


FIGURE 2.19

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=100; K=15; IFMEAN=0; NUMBER OF REPLICATIONS=10000

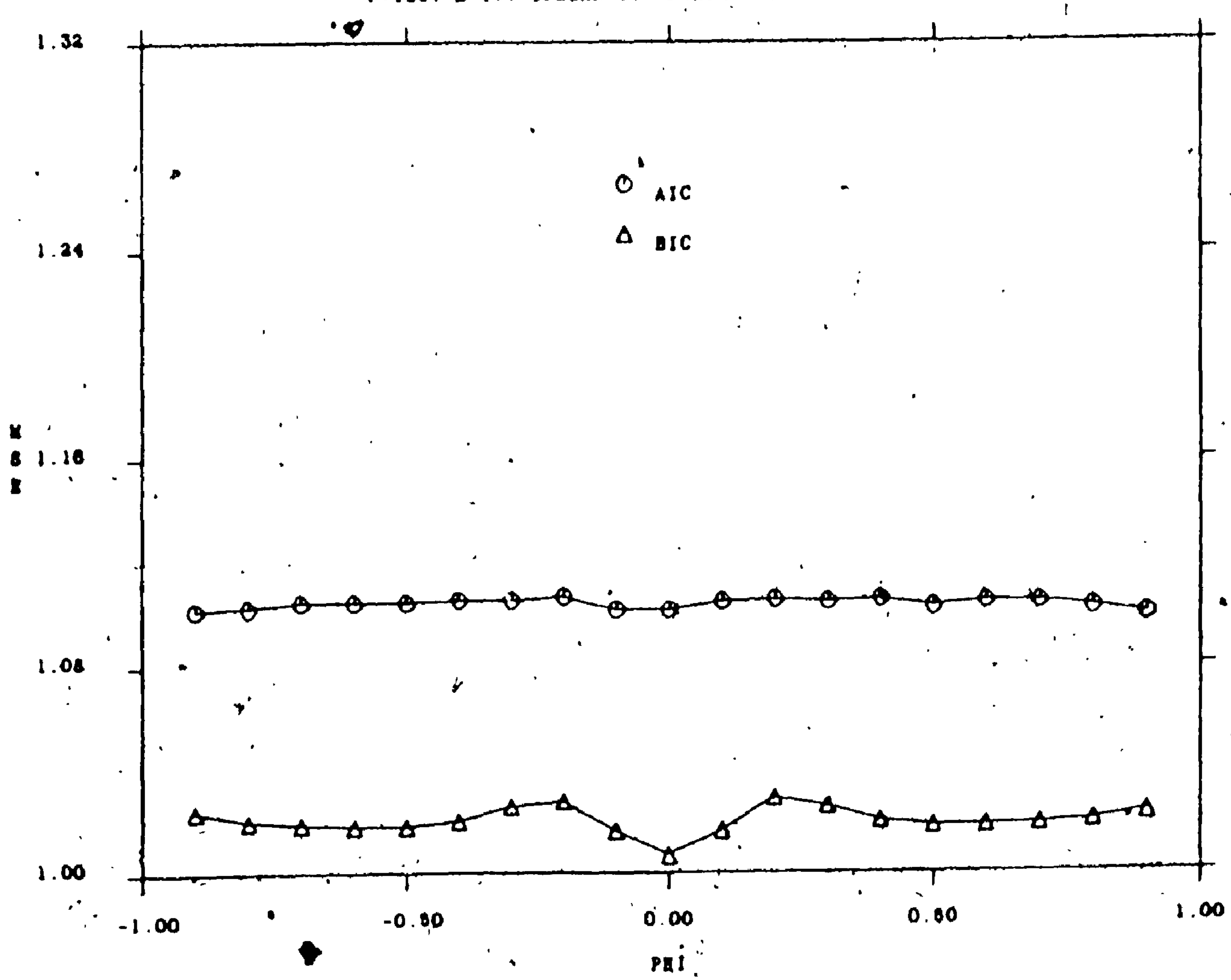


FIGURE 2.20

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=200. K=5: IFMEAN=0: NUMBER OF REPLICATIONS=10000

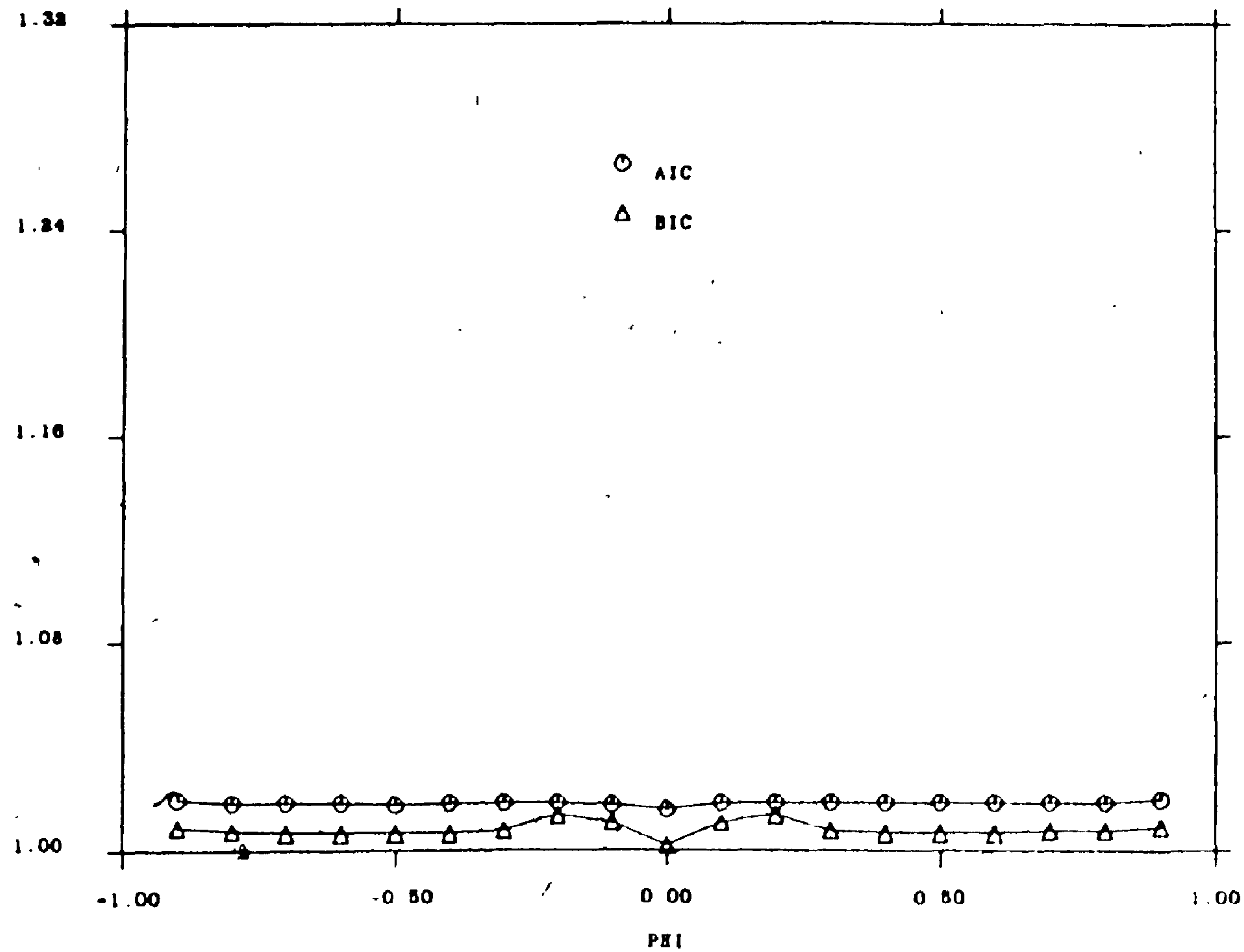


FIGURE 2.21

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=200 K=10 ITMEAN=0 NUMBER OF REPLICATIONS=10000

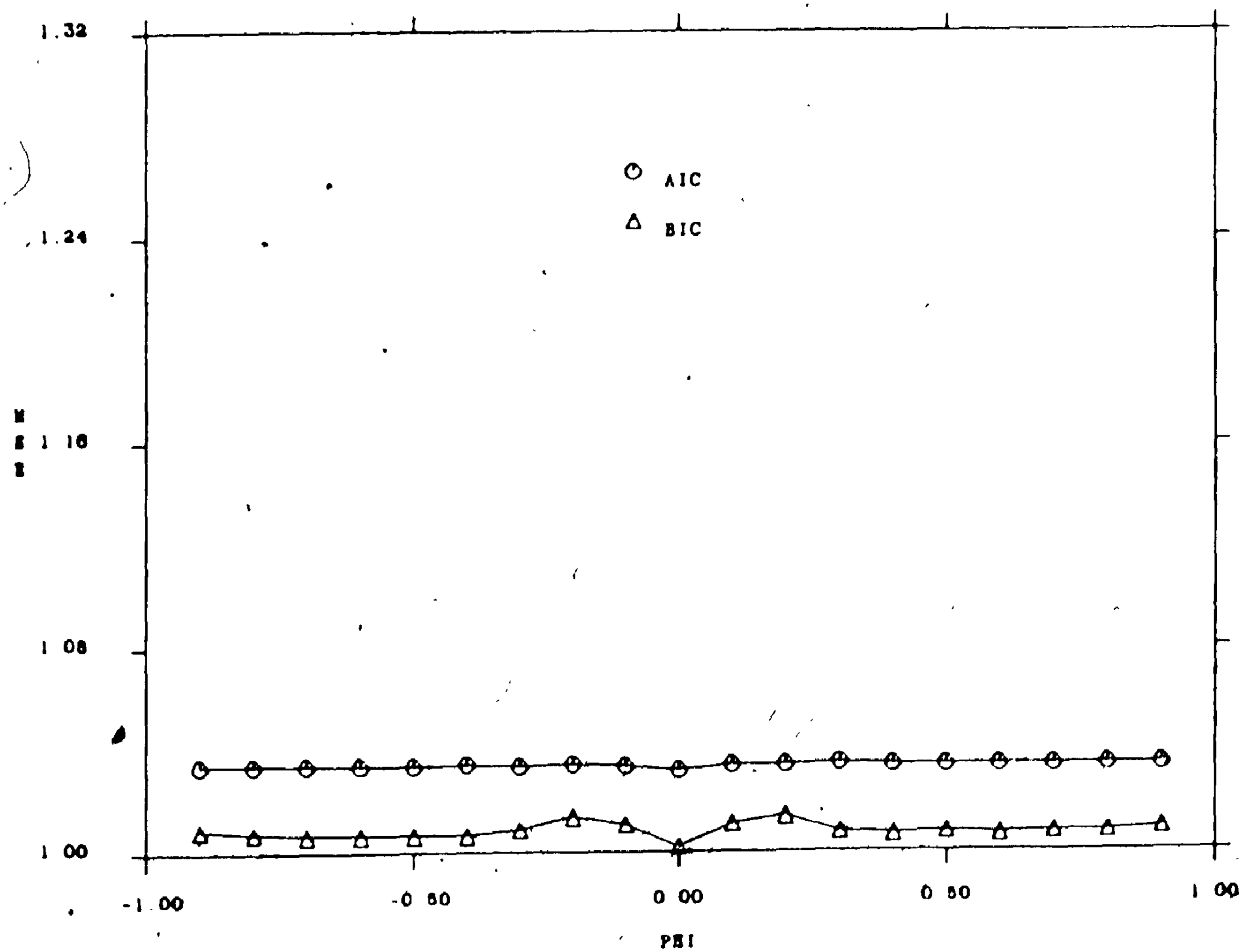
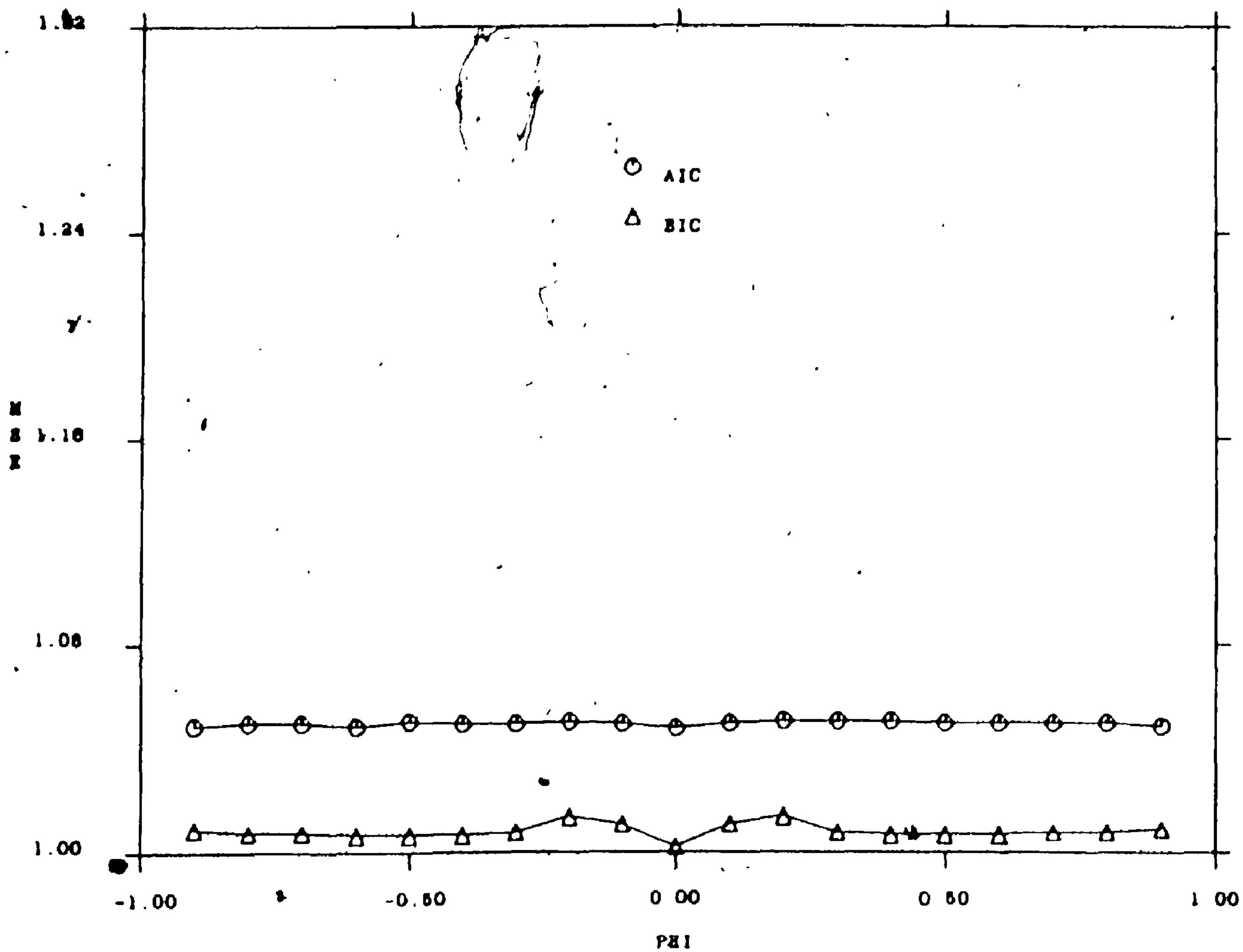


FIGURE 2.22

EMPIRICAL AVERAGE MSE FOR AIC AND BIC IN AR(1) MODELS

N=200: K=16: IFMEAN=0: NUMBER OF REPLICATIONS=10000



2.5 ASYMPTOTIC DISTRIBUTION OF MODEL SELECTION WITH THE AIC

Consider a realization $\{z_1, z_2, \ldots, z_n\}$ having a distribution $f(\cdot;\theta)$, where the parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_{k+k_o})'$ is unknown. Let Ω represent the $k+k_o$ dimensional space of θ, $\Omega \subset \Re^{k+k_o}$, and let $\Omega_j \subset \Omega$, where $\Omega_j = \{\theta \in \Omega : \theta_i = 0,\ j+1 \leq i \leq k+k_o\}$. To test the null hypothesis $H_j : \theta \in \Omega_j$ versus the alternative hypothesis $H_\Omega : \theta \in \Omega$, the likelihood ratio test, λ, defined earlier is,

$$\lambda = \frac{L_{\Omega_j}}{L_\Omega},$$

where L_{Ω_j} is the maximized likelihood function under the null hypothesis and L_Ω is the maximized likelihood function under the alternative. A well-known result of Wilks (1963) is that, under a number of regularity conditions, for a random sample of size n, $-2\log\lambda$ is asymptotically distributed as χ^2 with $(k+k_o-j)$ degrees of freedom if H_j is true. If H_j is false, then $\lambda \rightarrow 0$ as $n \rightarrow \infty$ and so the test is consistent. In the context of time series models, Whittle (1961) has shown that, under fairly general regularity conditions which include, for example, the ARMA(p,q) case, the likelihood ratio test is valid for nested models.

Assume that the regularity conditions of Whittle (1961) are satisfied. Let the true model consists of k_o parameters, $\theta \in \Omega_{k_o}$ and let the alternative model be the super-model consisting of $(h+k_o)$ parameters. Then the AIC for the model with h extra parameters may be written,

$$\text{AIC}_h = -2\log(\text{maximum likelihood}) + 2h + const., \qquad h = 0, \ldots, k. \quad (2.18)$$

For some suitable choice of constant and for large n,

$$\begin{aligned}\text{AIC}_h &= 2[\log L_{\Omega_{k_o}} - \log L_{\Omega_{k_o+h}}] + 2h \\ &= 2h - \chi_h^2, \end{aligned} \quad (2.19)$$

where L is the maximized likelihood function with respect to the appropriate parameter space. Consider fitting nested models with a parameter of dimension $k_o + h$, $h = 0, 1, \ldots, k$. Let $\hat{h}$ be a random variable which takes on the value h, the number of extra parameters of the selected model using the AIC criterion. Then the event $\hat{h} = h$ is equivalent to

$$\text{AIC}_h = \min(0, \text{AIC}_1, \ldots, \text{AIC}_k). \quad (2.20)$$

Let

$$S_j = \sum_{i=1}^{j} \left(2 - X_i\right),$$

where X_i are distributed independently as a χ_1^2 random variable. Hence,

$$\begin{aligned}\Pr(\hat{h} = h) &= \Pr\left(\min(S_0, S_1, \ldots, S_k) = S_h\right) \\ &= \Pr\left(S_0 - S_h > 0, S_1 - S_h > 0, \ldots, S_{h-1} - S_h > 0,\right. \\ &\qquad \left. S_{h+1} - S_h > 0, \ldots, S_k - S_h > 0\right), \end{aligned} \quad (2.21)$$

where $S_0 = 0$. For $j < h$,

$$S_j - S_h = \left(2j - \sum_{i=1}^{j} X_i\right) - \left(2h - \sum_{i=1}^{h} X_i\right)$$

$$= 2(j - h) + \sum_{i=j+1}^{h} X_i,$$

and if $j > h$,

$$S_j - S_h = \left(2j - \sum_{i=1}^{j} X_i\right) - \left(2h - \sum_{i=1}^{h} X_i\right)$$

$$= 2(j - h) - \sum_{i=h+1}^{j} X_i.$$

Hence, from the independence of X_i, (2.21) becomes,

$$\Pr(\hat{h} = h) = \Pr\left(S_0 - S_h > 0, \ldots, S_{h-1} - S_h > 0\right)$$

$$\times \Pr\left(S_{h+1} - S_h > 0, \ldots, S_k - S_h > 0\right). \qquad (2.22)$$

Let

$$A_1 = S_{h-1} - S_h$$

$$\vdots$$

$$A_h = S_0 - S_h$$

$$A_h - A_{h+1} = S_{h+1} - S_h$$

$$\vdots$$

$$A_h - A_k = S_k - S_h.$$

Then (2.22) becomes,

$$\Pr(\hat{h} = h) = \Pr(A_m > 0,\ 1 \le m \le h)$$

$$\times \Pr(A_k - A_m > 0,\ h + 1 \le m \le k)$$

$$= \Pr(A_m > 0,\ 1 \le m \le h)$$

$$\times \Pr(A_m - A_h \le 0,\ h + 1 \le m \le k). \qquad (2.23)$$

Before stating the main theorem, the following lemma is useful in the proof of the theorem. The proof of the lemma is given by Spitzer (1956).

Lemma. Let $x_1, \ldots, x_n$ be identically, independently distributed random variables and $s_k = x_1 + \cdots + x_k$, $1 \le k \le n$. Then the probabilities

$$p_n = \Pr(s_1 > 0, \ldots, s_n > 0) \text{ and } q_n = \Pr(s_1 \le 0, \ldots, s_n \le 0)$$

can be represented as

$$p_n = \sum\nolimits_n^{*} \left\{ \prod_{i=1}^{n} \frac{1}{r_i!} \left(\frac{\alpha_i}{i} \right)^{r_i} \right\},$$

and

$$q_n = \sum\nolimits_n^{*} \left\{ \prod_{i=1}^{n} \frac{1}{r_i!} \left(\frac{1-\alpha_i}{i} \right)^{r_i} \right\},$$

where $\alpha_i = \Pr(s_i > 0)$, $i = 1, \ldots, n$ and where the summation $\sum_n^{*}$ extends over all n-tuples $(r_1, \ldots, r_n)$ of nonnegative integers with the property $r_1 + 2r_2 + \cdots + nr_n = n$.

Now the main theorem can be stated.

Theorem 1. The distribution of $\hat{h}$ as $n \to \infty$ is,

$$\Pr(\hat{h} = h) = p_h q_{k-h} \qquad 0 \le h \le k, \tag{2.24}$$

where p_i, q_i are as defined in the lemma.

Proof. Suppose that Ω^{k+k_o} is the $k + k_o$ parameter space. The AIC for the model with h extra parameters is given in (2.18). If the true model consists of k_o parameters, then using the AIC to select a model with h extra parameters and for some appropriate choice of constant,

$$\mathrm{AIC}_h = 2[\log L_{\Omega_{k_o}} - \log L_{\Omega_{k_o+h}}] + 2h. \tag{2.25}$$

By the large sample result of the likelihood ratio test, equation (2.25) can be written as,

$$\mathrm{AIC}_h = 2h - \chi^2_h.$$

Let $\hat{h}$ be defined as before and let $S_j = \sum_1^j (2 - X_i)$, where X_i are chi-square random variables each with one degree of freedom. The probability that AIC selects a model with h extra parameters is given by,

$$\Pr(\hat{h} = h) = \Pr(\mathrm{AIC}_h = \min(0, \mathrm{AIC}_1, \ldots, \mathrm{AIC}_k)).$$

The result now follows from equations (2.21) to (2.23) and the lemma.

Remark 1. This theorem is seen to be a generalization of the results of Shibata (1976).

Theorem 1 shows that the probability of selecting h extra parameters is asymptotically the same for all models under very general conditions.

Remark 2. Letting $k = 1$, equation (2.23) with $h = 0$ is equivalent to equation (2.6) as $n \to \infty$. That is, $\Pr\{\hat{h} = 0\} \approx 0.843$ in this situation.

As an illustration of Theorem 1, consider the fractional ARMA models of Li and McLeod (1986). If the choice is between an

ARMA(p,q) model and an ARMA(p,q) with fractional differencing and if the true model is just the ARMA(p,q) model, the AIC will select the correct model about 84% of the time when n is large.

Remark 3. In the nested modelling situation, $\hat{h}$, does not converge to zero in probability as the sample size, n, becomes large and hence the AIC may be considered inconsistent. However, while the AIC may select a non-parsimonious model it will never select an inadequate model when n is large. This property follows from the consistency of the likelihood ratio test. The consistency property was illustrated in section 2.2 where it was found that $P_A \rightarrow 0$ as $n \rightarrow \infty$ when $\theta_o \neq 0$.

2.6 PERIODIC AUTOREGRESSIVE MODEL SELECTION

In this section, the application of Theorem 1 to periodic autoregressive models is described.

Time series data containing seasonal periods often arise in practice. Several techniques for choosing an appropriate stochastic model have been proposed. In particular, the Periodic Autoregressive (PAR) model is of interest here. Noakes *et al.* (1985) applied this model to riverflow time series.

Consider $\{z(t),\ t = 1, 2, \ldots\}$, to be a realization of a seasonal time series with period s. The index t can be viewed as a function of the year and the season. Hence t can be expressed as $t = (T-1)s + m$ where T denotes the Tth year, $T = 1, 2, \ldots, N$ and m denotes the mth season, $m = 1, 2, \ldots, s$. Then the process $z(\cdot)$ is said to be a periodic autoregression of period s and order $(p_1, \ldots, p_s)$ if for all integers t,

$$z(t) + \sum_{j=1}^{p_t} \alpha_t(j) z(t-j) = a(t) \quad ,$$

where $\{a(t)\}$ are uncorrelated with $\langle a(t) \rangle = 0$ and $\langle a^2(t) \rangle = \sigma_t^2$; and $p_t = p_{t+s}$, $\sigma_t^2 = \sigma_{t+s}^2$ and $\alpha_t(j) = \alpha_{t+s}(j)$, $j = 1, 2, \ldots, p_t$. The subscript t obeys modulo s arithmetic. It is also assumed that $z(\cdot)$ is a zero mean stationary time series. The covariance kernel is then defined as

$$R(k, t) = \langle z(k) z(t) \rangle = R(k+s, t+s).$$

Let $x(t)$ be a s-dimensional vector of an AR(p) process, then $x(t)$

is represented by

$$x(t) + \sum_{j=1}^{p} A(j)x(t-j) = u(t),$$

where $\{u(t)\}$ are uncorrelated with mean zero and $\text{cov}(u(t)) = \sum$. Define the jth coordinate of $x(t)$ as follows:

$$x_j(t) = z((t-1)s + j).$$

Thus the PAR model is a special case of the multivariate AR process. Pagano (1978) demonstrated the relationship between $x(\cdot)$ and $z(\cdot)$.

One way of estimating the parameters of the periodic autoregression model, $\alpha_m(j)$ and σ_m^2, $j = 1, 2, \ldots, p_m$, $m = 1, 2, \ldots, s$, is by solving the modified Yule-Walker equations:

$$R(m, m-v) + \sum_{j=1}^{p_m} \alpha_m(j) R(m-j, m-v) = \delta_{v0}\sigma_m^2, \qquad v \geq 0, \;\; m = 1, 2, \ldots, s, \tag{2.26}$$

where δ_{v0} is the kronecker delta and $R(\cdot)$ in (2.26) is the covariance kernel defined above. In practice, one replaces $R(\cdot)$ in (2.26) by

$$R_n(m, v) = \frac{1}{n} \sum_{j=1}^{h} z(m + js) z(v + js)$$

where $h = [n - \max(m, v)/s]$. Pagano (1978) has also shown that if $z(\cdot)$ is Gaussian, the estimates, $\hat{\alpha}_m = (\hat{\alpha}_m(1), \ldots, \hat{\alpha}_m(p_m))'$, $m = 1, 2, \ldots, s$ for $\alpha_m = (\alpha_m(1), \ldots, \alpha_m(p_m))'$, obtained by this method are consistent and the parameters can be estimated independently of the model of any other season. Furthermore, the information matrix is block diagonal; hence the estimated parameters are statistically independent of any other season and one can analyze each season separately.

The idea of this periodic autoregressive model can be extended to the periodic autoregressive-moving average (PARMA) model (see Tiao and Grupe, 1980). However, estimating the parameters can be tedious due to nonlinearity in the parameters (Vecchia, 1983).

In practice, the order of autoregression, p_m, may vary from season to season. The order of the AR model of each season, p_m, $m = 1, 2, \ldots, s$, can be determined using the AIC (see Sakai, 1982, for the definition of AIC). The AIC for season m is given by

$$\mathrm{AIC}(p_m) = n \log \hat{\sigma}_a^2(p_m) + 2p_m, \qquad m = 1, 2, \ldots, s$$

where $\hat{\sigma}_a^2(p_m)$ is the estimator of $\sigma_a^2(p_m)$, the residual variance of season m of order p_m. Once the order of the AR model is determined for each season, the final AIC value for the PAR model becomes:

$$\mathrm{AIC}(p_1, \ldots, p_s) = \sum_{i=1}^{s} \mathrm{AIC}(p_i) \;.$$

Let $\hat{h}_m$ be a random variable that the AIC selects a model with h_m extra parameters for season m and let the parameter space for season m be $\Omega^{k_{o,m} + k_m}$. Then we have the following.

Corollary. For season m of a periodic autoregression, the distribution of $\hat{h}_m$ as $N \to \infty$ is given by,

$$\Pr(\hat{h}_m = h_m) = p_{h_m} q_{k_m - h_m} \qquad 0 \le h_m \le k_m,$$

where $p_0 = q_0 = 1$, p_i and q_i are defined in the lemma in section 2.5 with $\alpha_i = \Pr(\chi_i^2 > 2i)$.

Proof. The proof follows from Theorem 1 by letting N tend to infinity instead of n.

2.7 CONCLUSION

It is shown that for the two sample problem, the BIC is not uniformly better than the AIC for a quadratic loss function although the BIC, in general, selects the true model more frequently than the AIC. This investigation also suggest that BIC tends to select models with fewer parameters (Table 2.1). Although this investigation is only based on the two sample situation, it has provided useful information on the behaviour of these criteria in the small sample situation. For the AR(1) model, empirical results showed BIC dominates the AIC. This simulation study has also indicated that one should not set an upper bound, K, too large. The larger K is, the larger the expected mean square error of forecast.

The asymptotic distribution for the model selected by the AIC was also derived. This derivation uses the large sample properties of the likelihood ratio test. It is seen that this derivation is simpler and more general than that of Shibata (1976).

APPENDIX A2.1

Derivation of $\tilde{R}(\hat{\mu}_x, \hat{\mu}_y)$ in section 2.3

Let $\theta_o = \mu_x - \mu_y$ where μ_x and μ_y denote the true mean values of x and y respectively and denote their corresponding maximum likelihood estimators by $\hat{\mu}_x$ and $\hat{\mu}_y$. Depending on the model selected, the estimators are given follows:

$$\hat{\mu}_x = \overline{x} \quad \text{if model B is selected,} \tag{A2.1}$$

$$\hat{\mu}_y = \overline{y} \quad \text{if model B is selected,} \tag{A2.2}$$

$$\hat{\mu}_x = \hat{\mu}_y = \frac{\overline{x} + \overline{y}}{2} \quad \text{if model A is selected.} \tag{A2.3}$$

Now, the risk function of $\hat{\mu}_x$ and $\hat{\mu}_y$ is given by,

$$\begin{aligned}\tilde{R}(\hat{\mu}_x, \hat{\mu}_y) &= \langle(\hat{\mu}_x - \mu_x)^2 + (\hat{\mu}_y - \mu_y)^2\rangle \\ &= \langle(\hat{\mu}_x - \mu_x)^2 + (\hat{\mu}_y - \mu_y)^2 \mid \text{A}\rangle P_{\text{A}} \\ &\quad + \langle(\hat{\mu}_x - \mu_x)^2 + (\hat{\mu}_y - \mu_y)^2 \mid \text{B}\rangle P_{\text{B}}.\end{aligned} \tag{A2.4}$$

Equation (A2.4) has an interpretation similar to equation (2.11) in section 2.3. If model A is selected, then using (A2.3),

$$\begin{aligned}\langle(\hat{\mu}_x - \mu_x)^2 + (\hat{\mu}_y - \mu_y)^2 \mid \text{A}\rangle &= \left\langle\left(\frac{\overline{x} + \overline{y}}{2} - \mu_x\right)^2 + \left(\frac{\overline{x} + \overline{y}}{2} - \mu_y\right)^2\right\rangle \\ &= 2\left(\frac{\theta_o^2}{4} + \frac{\sigma^2}{2n}\right) \\ &= \frac{\theta_o^2}{2} + \frac{\sigma^2}{n}.\end{aligned} \tag{A2.5}$$

This last equation is obtained since

$$V\left(\frac{\overline{x} + \overline{y}}{2}\right) = \frac{\sigma^2}{2n}$$

and

$$Bias = \frac{\mu_x + \mu_y}{2} - \frac{2\mu_x}{2} = \frac{\mu_y - \mu_x}{2}.$$

If model B is selected, then using (A2.1) and (A2.2),

$$\langle(\hat{\mu}_x - \mu_x)^2 + (\hat{\mu}_y - \mu_y)^2 ; B\rangle = \frac{2\sigma^2}{n}. \qquad (A2.6)$$

Hence, from (A2.5) and (A2.6), the general expression for the risk function is,

$$\begin{aligned} \hat{R}(\hat{\mu}_x, \hat{\mu}_y) &= \left(\frac{\theta_o^2}{2} + \frac{\sigma^2}{n}\right) P_A + \frac{2\sigma^2}{n} P_B \\ &= \frac{\theta_o^2}{2} P_A + \frac{\sigma^2}{n} P_B + \frac{\sigma^2}{n} \\ &= 2R(\hat{\theta}) + constant, \end{aligned}$$

where $R(\hat{\theta})$ is as defined in section 2.3.

APPENDIX A2.2

Derivation of the MSE in equation (2.17)

The MSE denotes the expected mean square of the one-step prediction error when an AR(p) model is used to forecast.

Consider the time series, $z_1, \ldots, z_n$, which is assumed to be generated by the AR(1) model

$$z_t = \phi z_{t-1} + a_t,$$

where $a_t \sim NID(0, \sigma_a^2)$. Then if an AR($p$) with fitted parameters $\hat{\phi}_1, \ldots, \hat{\phi}_p$ is used to make the one-step ahead forecasts, the prediction error for the forecast of z_t is

$$\hat{e}_{t,p} = z_t - \hat{\phi}_1 z_{t-1} - \cdots - \hat{\phi}_p z_{t-p}.$$

Then,

$$\begin{aligned}
\text{MSE} &= \lim_{m \to \infty} \frac{1}{m} \sum_{t=n+1}^{m+n} \langle \hat{e}_{t,p}^2 \mid z_1, \ldots, z_n \rangle \\
&= \lim_{m \to \infty} \frac{1}{m} \sum_{t=n+1}^{m+n} \langle (z_t - \hat{\phi}_1 z_{t-1} - \cdots - \hat{\phi}_p z_{t-p})^2 \mid z_1, \ldots, z_n \rangle \\
&= \sum_{i=0}^{p} \sum_{j=0}^{p} \hat{\phi}_i \hat{\phi}_j \gamma_{|i-j|} \\
&= \sum_{i=0}^{p} \sum_{j=0}^{p} \hat{\phi}_i \hat{\phi}_j \phi^{|i-j|} \gamma_0,
\end{aligned}$$

where $\gamma_0 = \sigma_a^2 / (1 - \phi^2)$.

REFERENCES

Akaike, H. (1969), "Fitting Autoregressive Models for Prediction". *Annals of the Institute of Statistical Mathematics* **21**, 243–247.

Akaike, H. (1970), "Statistical Predictor Identification". *Annals of the Institute of Statistical Mathematics* **22**, 203–217.

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle". *2nd International Symposium on Information Theory*, (eds. B. N. Petrov and F. Csaki), Akademiai Kiado, Budapest, 267–281.

Akaike, H. (1974), "A New Look at the Statistical Model Identification". *IEEE Transactions on Automatic Control* **AC-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. (1979), "A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting". *Biometrika* **66**, 237–242.

Duong, Q. P. (1984), "On the Choice of the Order of Autoregressive Models: A Ranking and Selection Approach". *Journal of Time Series Analysis* **5**, 145–157.

Hannan, E. J. and Quinn, B. G. (1979), "The Determination of the Order of an Autoregression". *Journal of the Royal Statistical Society (B)* **41**, 190–195.

Johnson, N. L. and Kotz, S. (1970), *Continuous Univariate Distributions-2*. Boston: Houghton Mifflin Company.

Li, W. K. and McLeod, A. I. (1986), "Fractional Time Series Modelling". *Biometrika* **73**, 217–221.

Marsaglia, G. (1976), "Random Number Generation". In Encyclopedia of Computer Science, (ed. A. Ralston). New York: Petrocelli and Charter, 1192–1197.

McLeod, A. I. and Hipel, K. W. (1978), "Simulation Procedures for Box-Jenkins Models". *Water Resources Research* **15**, 969–975.

Noakes, D. J., McLeod, A. I. and Hipel, K. W. (1985), "Forecasting Monthly Riverflow Time Series". *International Journal of Forecasting* **1**, 179–190.

Pagano, M. (1978), "On Periodic and Multiple Autoregressions". *The Annals of Statistics* **6**, 1310–1317.

Parzen, E. (1974), "Some Recent Advances in Time Series Modelling". *IEEE Transactions on Automatic Control* **AC-19**, 723–730.

Priestley, M. B. (1981), *Spectral Analysis and Time Series, Vol. 1: Univariate Series*. New York: Academic Press.

Sakai, H. (1982), "Circular Lattice Filtering Using Pagano's Method". *IEEE Transactions on Acoustics, Speech, and Signal Processing* **ASSP-30**, 279–287.

Schwarz, G. (1978), "Estimating the Dimension of a Model". *The Annals of Statistics* **6**, 461–464.

Shibata, R. (1976), "Selection of the Order of an Autoregressive Model by Akaike's Information Criterion". *Biometrika* **63**, 117–126.

Spitzer, F. (1956), "A Combinatorial Lemma and its Application to Probability Theory". *Transactions of the American Mathematical Society* **82**, 323–339.

Stone, M. (1977), "An Asymptotic Equivalence of Choice of Model by Cross-validation and Akaike's Criterion". *Journal of the Royal Statistical Society (B)* **39**, 44–47.

Stone, M. (1979), "Comments on Model Selection Criteria of Akaike and Schwarz". *Journal of the Royal Statistical Society (B)* **41**, 276–278.

Tiao, G. C. and Grupe, M. R. (1980), "Hidden Periodic Autoregressive-Moving Average Models in Time Series Data". *Biometrika* **67**, 365–373.

Vecchia, A. V., Jr. (1983), Aggregation and Estimation for Periodic Autoregressive-Moving Average Models. Unpublished Ph.D. dissertation, Colorado State University, U.S.A.

Whittle, P. (1961), "Gaussian Estimation in Stationary Time Series". *Bulletin of the International Statistical Institute* **34**, 105–128.

Wilks, S. S. (1963), *Mathematical Statistics*. New York: Wiley.

Woodroofe, M. (1982), "On Model Selection and the Arc Sine Laws". *The Annals of Statistics* **10**, 1182–1194.

CHAPTER 3

ON THE EFFICIENCY OF A STRONGLY CONSISTENT ESTIMATOR IN ARMA MODELS

3.1 INTRODUCTION

Given the time series $\{z_t; t \geq 1\}$, the mixed autoregressive-moving average model of order p and q respectively, ARMA(p,q), is defined to be

$$\phi(B)z_t = \theta(B)a_t, \tag{3.1}$$

where $\phi(B)$ is as defined in chapter 2; $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$; $\{a_t\}$ are assumed to be independent, identically and normally distributed with mean zero and variance σ_a^2. The characteristic roots of $\phi(B) = 0$ and $\theta(B) = 0$ are assumed to lie outside the unit circle and it is further assumed that there are no common roots. It is further assumed that the time series has zero mean, i.e., $\langle z_t \rangle = 0$, where $\langle \cdot \rangle$ denotes mathematical expectation.

Consider a realization of n observations. Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$. For a pure autoregression of order p, denoted by AR(p), the vector of parameters $\boldsymbol{\phi}$ can be determined by solving the Yule-Walker equation

$$P_p \boldsymbol{\phi} = \boldsymbol{\rho}_p, \tag{3.2}$$

where

$$P_p = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \\ \rho_{p-1} & \rho_{p-2} & & 1 \end{bmatrix},$$

$$\boldsymbol{\rho}_p = (\rho_1, \cdots, \rho_p)'$$

and

$$\rho_k = \frac{\langle z_t z_{t-k} \rangle}{\langle z_t^2 \rangle}, \qquad k = 1, 2, \ldots, p.$$

Then $\bar{\boldsymbol{\phi}}$, the Yule-Walker estimate, can be obtained from (3.2) by simply replacing ρ_k by its estimate r_k, the sample autocorrelation function defined by

$$r_k = \frac{c_k}{c_o},$$

where $c_k = n^{-1} \sum_{t=1}^{n-k} z_t z_{t+k}$.

Hannan (1975) showed that in the ARMA(p, q) model the autoregressive parameters, $\phi_1, \ldots, \phi_p$, estimated by solving

$$\sum_{i=0}^{p} \bar{\phi}_i c_{i-j} = 0, \qquad j = q+1, \ldots, q+p, \tag{3.3}$$

where $\bar{\phi}_0 = -1$, are strongly consistent. As pointed out by Tsay and Tiao (1984), the estimates $\bar{\phi}_1, \ldots, \bar{\phi}_p$ are asymptotically equivalent to the iterative ordinary least squares estimates. In the next section, the asymptotic efficiency of $\bar{\phi}_1$ in the ARMA(1,1) model is derived.

3.2 EFFICIENCY OF $\tilde{\phi}_1$

The ARMA(p,q) model in (3.1) can be written as

$$z_t - \phi_1 z_{t-1} - \cdots - \phi_p z_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}.$$

If one multiplies by z_{t-k} and takes expected values, one obtains the following:

$$\gamma_k - \phi_1 \gamma_{k-1} - \cdots - \phi_p \gamma_{k-p} = \gamma_{za}(k) - \theta_1 \gamma_{za}(k-1) - \cdots - \theta_q \gamma_{za}(k-q), \quad (3.4)$$

where γ_k is the covariance function between the series z_t and z_{t-k} defined by $\gamma_k = \langle z_t z_{t-k} \rangle$, and $\gamma_{za}(k)$ is the cross covariance function between z_t and a_t, and is defined by $\gamma_{za}(k) = \langle z_{t-k} a_t \rangle$. Upon dividing by γ_0, equation (3.4) can be written,

$$\rho_k - \phi_1 \rho_{k-1} - \cdots - \phi_p \rho_{k-p} = 0, \qquad k \geq q+1.$$

The estimate $\tilde{\phi}$ is obtained by solving this equation and by replacing ρ_k by its sample estimate r_k.

For the ARMA(1,1) model,

$$\tilde{\phi}_1 = \frac{r_2}{r_1} = \frac{c_2}{c_1}.$$

Expanding in Taylor series up to first order terms yields:

$$\tilde{\phi}_1 = \phi_1 + (c_2 - \gamma_2) \cdot \frac{1}{\gamma_1} - (c_1 - \gamma_1) \frac{\gamma_2}{\gamma_1^2}.$$

It is well known (Mann and Wald, 1943) that this expansion ignores terms of order $O(1/n)$. Then

$$V(\tilde{\phi}_1) = \frac{1}{\gamma_1^2} V(c_2) + \frac{\phi_1^2}{\gamma_1^2} V(c_1) - \frac{2\phi_1}{\gamma_1^2} \operatorname{cov}(c_1, \ c_2). \quad (3.5)$$

It is easily shown that (see appendix A3.1 and A3.2), apart from terms $O(1/n^2)$,

$$V(c_1) = \frac{1}{n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-1} \left\{ \gamma_{t-s}^2 + \gamma_{t-s-1} \gamma_{t+1-s} \right\}$$

$$= \frac{1}{n} \left\{ \gamma_0^2 + 2\phi_1 \gamma_0 \gamma_1 + \gamma_1^2 + \frac{2\gamma_1^2}{1-\phi_1^2}(1+\phi_1^2) \right\}, \qquad (3.6)$$

$$V(c_2) = \frac{1}{n^2} \sum_{t=1}^{n-2} \sum_{s=1}^{n-2} \left\{ \gamma_{t-s}^2 + \gamma_{t-s-2} \gamma_{t+1-s} \right\}$$

$$= \frac{1}{n} \left\{ \gamma_0^2 + 3\phi_1^2 \gamma_1^2 + 2\phi_1^3 \gamma_0 \gamma_1 + \frac{2\gamma_1^2}{1-\phi_1^2}(1+\phi_1^4) \right\}, \qquad (3.7)$$

and

$$\mathrm{cov}(c_1, c_2) = \frac{1}{n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-2} \left\{ \gamma_{t-s} \gamma_{t-s-1} + \gamma_{t-s-2} \gamma_{t+1-s} \right\}$$

$$= \frac{2}{n} \left\{ \phi_1 \gamma_1^2 + \gamma_1 (1+\phi_1^2) \left(\gamma_0 + \frac{\phi_1 \gamma_1}{1-\phi_1^2} \right) \right\}. \qquad (3.8)$$

If one substitutes equations (3.6), (3.7) and (3.8) into (3.5) one may obtain the following:

$$V(\tilde{\phi}_1) = \frac{1}{n} \left\{ \frac{1+\phi_1^2}{\rho_1^2} - \frac{4\phi_1}{\rho_1} + 2 \right\},$$

where

$$\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}.$$

Let $\hat{\phi}_1$ denote the maximum likelihood estimate of ϕ_1. Box and Jenkins (1976, p. 242) give the asymptotic variance of $\hat{\phi}_1$ as follows:

$$V(\hat{\phi}_1) = \frac{(1-\phi_1\theta_1)^2(1-\phi_1^2)}{n(\phi_1-\theta_1)^2}.$$

Hence the asymptotic efficiency of $\tilde{\phi}_1$ relative to $\hat{\phi}_1$ is

$$\text{Eff} = \frac{V(\hat{\phi}_1)}{V(\tilde{\phi}_1)}$$

$$= \frac{\rho_1^2(1-\phi_1\theta_1)^2(1-\phi_1^2)}{(\phi_1-\theta_1)^2(2\rho_1^2-4\rho_1\phi_1+\phi_1^2+1)}.$$

It follows that,

$$\lim_{\theta_1\to 0} \text{Eff} = 1.$$

Hence $\tilde{\phi}_1$ is as efficient as $\hat{\phi}_1$ when θ_1 is close to zero. The asymptotic efficiency for various models is presented in Table 3.1. This table shows that $\tilde{\phi}_1$ is indeed as efficient as the maximum likelihood estimator when θ_1 is near zero, but that the efficiency is very poor when both ϕ_1 and θ_1 are close to negative one or when both ϕ_1 and θ_1 are close to positive one.

TABLE 3.1

Asymptotic efficiency of $\tilde{\phi}_1$ relative to $\hat{\phi}_1$

	θ_1						
ϕ_1	–0.90	–0.60	–0.30	0.00	0.30	0.60	0.90
–0.95	0.003	0.288	0.908	1.000	0.983	0.961	0.944
–0.75	0.012	0.174	0.739	1.000	0.919	0.815	0.736
–0.50	0.045	0.215	0.688	1.000	0.846	0.652	0.516
–0.25	0.107	0.290	0.696	1.000	0.783	0.510	0.340
0.00	0.204	0.389	0.731	—	0.731	0.389	0.204
0.25	0.340	0.510	0.783	1.000	0.696	0.290	0.107
0.50	0.516	0.652	0.846	1.000	0.688	0.215	0.045
0.75	0.736	0.815	0.919	1.000	0.739	0.174	0.012
0.95	0.944	0.961	0.983	1.000	0.908	0.288	0.003

3.3 FINITE SAMPLE SIMULATION OF AR(1) PROCESS

From Table 3.1, it is seen that when the process is a pure autoregression, $\tilde{\phi}_1$ is fully efficient asymptotically. To examine the efficiency in finite samples, AR(1) processes were simulated with parameter ϕ_1 successively equal to ± 0.3, ± 0.6, ± 0.9 and with sample size n successively equal to 50, 100, 200, 400, 800. The random number generator Super Duper (Marsaglia, 1976), in conjunction with the Box-Mueller method was used to generate the $N(0,1)$ variates. For each combination of ϕ_1 and n, 10000 realizations were simulated to obtain $\hat{\phi}_{1,i}$ and $\tilde{\phi}_{1,i}$, $i = 1, 2, \ldots, 10000$. The estimated parameters were calculated as $\hat{\phi}_1 = r_1$ and $\tilde{\phi}_1 = r_2/r_1$ and the corresponding sample variances of $\hat{\phi}_1$ and $\tilde{\phi}_1$ were calculated as,

$$S_{11} = \frac{1}{10000} \sum_i \{\hat{\phi}_{1,i} - avg(\hat{\phi}_1)\}^2$$

and

$$S_{22} = \frac{1}{10000} \sum_i \{\tilde{\phi}_{1,i} - avg(\tilde{\phi}_1)\}^2,$$

where $avg(\hat{\phi}_1) = \sum_i \hat{\phi}_{1,i}/10000$ and $avg(\tilde{\phi}_1) = \sum_i \tilde{\phi}_{1,i}/10000$. The efficiency is then obtained by taking the ratio of the sample variances,

$$\text{Eff} = \frac{S_{11}}{S_{22}}.$$

The variance of the efficiency can be obtained by expanding Eff in Taylor series which yields,

$$\hat{V}(\text{Eff}) = \frac{S^2}{S_{22}^2\, 10000},$$

where $S^2 = S_{11} + S_{22}\text{Eff}^2 - 2S_{12}\text{Eff}$ and S_{12} is the sample covariance between $\left(\hat{\phi}_{1,i} - avg(\hat{\phi}_1)\right)$ and $\left(\bar{\phi}_{1,i} - avg(\bar{\phi}_1)\right)$.

Table 3.2 shows the simulation results with the standard errors given in parentheses. The table shows that $\bar{\phi}$ is quite inefficient when ϕ_1 is near zero and has higher efficiency near ± 1.

Table 3.3 presents the empirical estimates of ϕ_1 using estimators $\hat{\phi}_1$ and $\bar{\phi}_1$. The results in this table are obtained from the same simulation experiment as in Table 3.2. From this table, it is clear that $\bar{\phi}_1$ underestimates ϕ_1 and $\hat{\phi}_1$ is seen to provide a better estimate. Both estimators provide close estimates for ϕ_1 when n is large.

TABLE 3.2

Empirical efficiency of an AR(1) process

Number of replications: 10000

	Φ_1					
n	−0.9	−0.6	−0.3	0.3	0.6	0.9
50	0.615 (0.041)	0.159 (0.012)	0.001 (0.000)	0.000 (0.000)	0.121 (0.009)	0.547 (0.034)
100	0.723 (0.070)	0.297 (0.030)	0.016 (0.002)	0.002 (0.000)	0.276 (0.027)	0.689 (0.060)
200	0.765 (0.103)	0.331 (0.047)	0.070 (0.010)	0.062 (0.009)	0.318 (0.044)	0.756 (0.097)
400	0.792 (0.154)	0.339 (0.067)	0.083 (0.017)	0.079 (0.016)	0.343 (0.069)	0.781 (0.146)
800	0.794 (0.217)	0.356 (0.100)	0.084 (0.024)	0.086 (0.024)	0.344 (0.097)	0.801 (0.209)

TABLE 3.3

Empirical estimates of ϕ_1 of an AR(1) process

Number of replications: 10000

		ϕ_1					
n		-0.9	-0.6	-0.3	0.3	0.6	0.9
50	$\hat{\phi}_1$	−0.851 (0.001)	−0.575 (0.001)	−0.298 (0.001)	0.257 (0.001)	0.531 (0.001)	0.793 (0.001)
	$\tilde{\phi}_1$	−0.838 (0.001)	−0.520 (0.003)	−0.140 (0.057)	0.131 (0.071)	0.473 (0.003)	0.771 (0.001)
100	$\hat{\phi}_1$	−0.874 (0.001)	−0.585 (0.001)	−0.298 (0.001)	0.278 (0.001)	0.566 (0.001)	0.849 (0.001)
	$\tilde{\phi}_1$	−0.869 (0.001)	−0.563 (0.001)	−0.226 (0.007)	0.221 (0.020)	0.541 (0.002)	0.842 (0.001)
200	$\hat{\phi}_1$	−0.887 (0.000)	−0.594 (0.001)	−0.299 (0.001)	0.289 (0.001)	0.583 (0.001)	0.876 (0.000)
	$\tilde{\phi}_1$	−0.885 (0.000)	−0.583 (0.001)	−0.269 (0.003)	0.255 (0.003)	0.572 (0.001)	0.873 (0.000)
400	$\hat{\phi}_1$	−0.894 (0.000)	−0.597 (0.000)	−0.299 (0.000)	0.294 (0.000)	0.591 (0.000)	0.888 (0.000)
	$\tilde{\phi}_1$	−0.893 (0.000)	−0.592 (0.001)	−0.285 (0.002)	0.278 (0.002)	0.586 (0.001)	0.887 (0.000)
800	$\hat{\phi}_1$	−0.897 (0.000)	−0.598 (0.000)	−0.299 (0.000)	0.297 (0.000)	0.595 (0.000)	0.894 (0.000)
	$\tilde{\phi}_1$	−0.897 (0.000)	−0.596 (0.000)	−0.292 (0.001)	0.290 (0.001)	0.593 (0.000)	0.894 ([illegible])

3.4 CONCLUSION

An explicit expression for the asymptotic efficiency of the strongly consistent estimator $\bar{\phi}_1$ of ϕ_1 of an ARMA(1,1) model is obtained. Although the asymptotic efficiency is one when the moving average parameter is zero, the finite sample simulation results show that the estimator $\bar{\phi}_1$ is not as efficient as the maximum likelihood estimator, $\hat{\phi}_1$, even when the sample size is as large as 800. In particular, the efficiency is very poor when ϕ_1 is near the origin.

The result of section 2 also applies to the estimator of θ_1 in an ARMA(1,1) model given by,

$$\bar{\theta}_1 = \frac{ri(2)}{ri(1)},$$

where $ri(\cdot)$ denotes the inverse autocorrelation function.

APPENDIX A3.1

Derivation of $V(c_1)$ in (3.6)

The variance of c_2 is not given since the exact method can be used to evaluate $V(c_2)$. Denote the cumulant operator by "*cum*". The properties of cumulants given in Brillinger (1975) are used below.

$$V(c_1) = cum(c_1,\ c_1)$$

$$= \frac{1}{n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-1} cum(z_t z_{t+1},\ z_s z_{s+1})$$

$$= \frac{1}{n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-1} \Big\{ cum(z_t,\ z_{t+1},\ z_s,\ z_{s+1}) + cum(z_t,\ z_s) cum(z_{t+1},\ z_{s+1}) + cum(z_t,\ z_{s+1}) cum(z_{t+1},\ z_s) \Big\} \quad \text{(A3.1)}$$

Equation (A3.1) follows from the fact that $\langle z_t \rangle = 0$ for all $t = 1, 2, \ldots, n$. Note that $cum(z_t,\ z_s) = \gamma_{t-s}$ where γ_k is the theoretical autocovariance function at lag k. Also, since normality is assumed, all cumulants of order higher than two are zero. Hence, (A3.1) is reduced to

$$V(c_1) = \frac{1}{n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-1} \left\{ \gamma_{t-s}^2 + \gamma_{t-s-1} \gamma_{t+1-s} \right\}. \quad \text{(A3.2)}$$

Now, the first term on the right side of (A3.2) is

$$\frac{1}{n^2}\sum_{t=1}^{n-1}\sum_{s=1}^{n-1}\gamma_{t-s}^2 = \frac{1}{n^2}\left\{\sum_{t=1}^{n-1}\gamma_{t-1}^2 + \sum_{t=1}^{n-1}\gamma_{t-2}^2 + \cdots + \sum_{t=1}^{n-1}\gamma_{t-n+1}^2\right\}$$

$$= \frac{1}{n^2}\left\{(n-1)\gamma_0^2 + 2(n-2)\gamma_1^2 + \cdots + 2\gamma_{n-2}^2\right\}$$

$$= \frac{n-1}{n^2}\gamma_0^2 + \frac{2\gamma_1^2}{n^2}\left\{(n-2) + (n-3)\phi_1^2 + \cdots + \phi_1^{2(n-3)}\right\} \quad \text{(A3.3)}$$

$$= \frac{n-1}{n^2}\gamma_0^2 + \frac{2\gamma_1^2}{n^2}\left\{\frac{(n-2)(1-\phi_1^{2(n-2)})}{1-\phi_1^2} - \frac{\phi_1^2\{1-(n-2)\phi_1^{2(n-3)} + (n-3)\phi_1^{2(n-2)}\}}{(1-\phi_1^2)^2}\right\} \quad \text{(A3.4)}$$

Equation (A3.3) is obtained from the fact that $\gamma_k = \phi_1^{k-1}\gamma_1$, $k \geq 2$. Ignoring terms with ϕ_1^n and $1/n^2$ in (A3.4), then

$$\frac{1}{n^2}\sum_{t=1}^{n-1}\sum_{s=1}^{n-1}\gamma_{t-s}^2 \doteq \frac{\gamma_0^2}{n} + \frac{2\gamma_1^2}{n(1-\phi_1^2)} \quad \text{(A3.5)}$$

The second term in equation (A3.2) is

$$\frac{1}{n^2}\sum_{t=1}^{n-1}\sum_{s=1}^{n-1}\gamma_{t-s-1}\gamma_{t+1-s}$$

$$= \frac{1}{n^2}\left\{\sum_{t=1}^{n-1}\gamma_{t-2}\gamma_t + \sum_{t=1}^{n-1}\gamma_{t-3}\gamma_{t-1} + \cdots + \sum_{t=1}^{n-1}\gamma_{t-n}\gamma_{t-n+2}\right\}$$

$$= \frac{1}{n^2}\left\{(n-1)\gamma_1^2 + 2(n-2)\gamma_0\gamma_2 + 2(n-3)\gamma_1\gamma_3 + \cdots + 2\gamma_{n-3}\gamma_{n-1}\right\}$$

$$= \frac{n-1}{n^2}\gamma_1^2 + \frac{2(n-2)\phi_1\gamma_0\gamma_1}{n^2}$$

$$+ \frac{2\phi_1^2\gamma_1^2}{n^2}\left\{ (n-3) + (n-4)\phi_1^2 + \cdots + \phi_1^{2(n-4)} \right\},$$

$$= \frac{n-1}{n^2}\phi_1^2 + \frac{2(n-2)\phi_1\gamma_0\gamma_1}{n^2} + \frac{2\phi_1^2\gamma_1^2}{n^2}\left\{ \frac{(n-3)(1-\phi_1^{2(n-3)})}{1-\phi_1^2} \right.$$

$$\left. - \frac{\phi_1^2\{1-(n-3)\phi_1^{2(n-4)} + (n-4)\phi_1^{2(n-3)}\}}{(1-\phi_1^2)^2} \right\} \qquad \text{(A3.6)}$$

Now, if terms with ϕ_1^n and $1/n^2$ in (A3.6) are ignored, then

$$\frac{1}{n^2}\sum_{t=1}^{n-1}\sum_{s=1}^{n-1}\gamma_{t-s-1}\gamma_{t+1-s} = \frac{\gamma_1^2}{n} + \frac{2\phi_1\gamma_0\gamma_1}{n} + \frac{2\phi_1^2\gamma_1^2}{n(1-\phi_1^2)}. \qquad \text{(A3.7)}$$

Combining (A3.5) and (A3.7) gives

$$V(c_1) = \frac{1}{n}\left\{ \gamma_0^2 + \frac{2\gamma_1^2}{1-\phi_1^2} + \gamma_1^2 + 2\phi_1\gamma_0\gamma_1 + \frac{2\phi_1^2\gamma_1^2}{1-\phi_1^2} \right\}$$

$$= \frac{1}{n}\left\{ \gamma_0^2 + 2\phi_1\gamma_0\gamma_1 + \frac{2\gamma_1^2(1+\phi_1^2)}{1-\phi_1^2} \right\}$$

which is equation (3.6). Equation (3.7) can be obtained similarly.

APPENDIX A3.2

Derivation of $cov(c_1, c_2)$ in (3.8)

A technique similar to that used in appendix A3.1 is used in the following derivation.

$$cov(c_1, c_2) = cum(c_1, c_2)$$

$$= \frac{1}{n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-2} cum(z_t z_{t+1}, z_s z_{s+2})$$

$$= \frac{1}{n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-2} \Big\{ cum(z_t, z_{t+1}, z_s, z_{s+2}) + cum(z_t, z_s)cum(z_{t+1}, z_{s+2}) + cum(z_t, z_{s+2})cum(z_s, z_{t+1}) \Big\}$$

$$= \frac{1}{n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-2} \Big\{ \gamma_{t-s}\gamma_{t-s-1} + \gamma_{t-s-2}\gamma_{t+1-s} \Big\}. \qquad \text{(A3.8)}$$

(A3.8) is obtained because $\langle z_t \rangle = 0$ and because of the Gaussian assumption. Now the first term in (A3.8) is

$$\frac{1}{n^2} \sum_{t=1}^{n-1} \sum_{s=1}^{n-2} \gamma_{t-s}\gamma_{t-s-1}$$

$$= \frac{1}{n^2} \Big\{ \sum_{s=1}^{n-2} \gamma_{1-s}\gamma_{-s} + \sum_{s=1}^{n-2} \gamma_{2-s}\gamma_{1-s} + \cdots + \sum_{s=1}^{n-2} \gamma_{n-1-s}\gamma_{n-2-s} \Big\}$$

$$= \frac{1}{n^2} \Big\{ 2(n-2)\gamma_0\gamma_1 + 2(n-3)\gamma_2\gamma_1 + \cdots + 2\gamma_{n-2}\gamma_{n-3} \Big\}$$

$$= \frac{2(n-2)\gamma_0\gamma_1}{n^2} + \frac{2\phi_1\gamma_1^2}{n^2}\left\{(n-3) + (n-4)\phi_1^2 + \cdots + \phi_1^{2(n-4)}\right\}$$

$$= \frac{2(n-2)\gamma_0\gamma_1}{n^2} + \frac{2\phi_1\gamma_1^2}{n^2}\left\{\frac{(n-3)(1-\phi_1^{2(n-3)})}{1-\phi_1^2} - \frac{\phi_1^2\{1-(n-3)\phi_1^{2(n-4)} + (n-4)\phi_1^{2(n-3)}\}}{(1-\phi_1^2)^2}\right\}. \quad \text{(A3.9)}$$

If one ignores terms with $1/n^2$ and ϕ_1^n in (A3.9), then

$$\frac{1}{n^2}\sum_{t=1}^{n-1}\sum_{s=1}^{n-2}\gamma_{t-s}\gamma_{t-s-1} = \frac{2\gamma_0\gamma_1}{n} + \frac{2\phi_1\gamma_1^2}{n(1-\phi_1^2)}. \quad \text{(A3.10)}$$

The second term of (A3.8) is

$$\frac{1}{n^2}\sum_{t=1}^{n-1}\sum_{s=1}^{n-2}\gamma_{t-s-2}\gamma_{t+1-s}$$

$$= \frac{1}{n^2}\left\{\sum_{s=1}^{n-2}\gamma_{-s-1}\gamma_{2-s} + \sum_{s=1}^{n-2}\gamma_{-s}\gamma_{3-s} + \cdots + \sum_{s=1}^{n-2}\gamma_{n-3-s}\gamma_{n-s}\right\}$$

$$= \frac{1}{n^2}\left\{(n-2)\gamma_1\gamma_2 + (n-3)\gamma_0\gamma_3 + (n-4)\gamma_1\gamma_4 + \cdots + \gamma_{n-4}\gamma_{n-1}\right\}.$$

$$= \frac{2\phi_1(n-2)\gamma_1^2}{n^2} + \frac{2\phi_1^2(n-3)\gamma_0\gamma_1}{n^2} + \frac{2\phi_1^3\gamma_1^2}{n^2}\left\{(n-4) + (n-6)\phi_1^2 + \cdots + \phi_1^{2(n-5)}\right\}$$

$$= \frac{2\phi_1(n-2)\gamma_1^2}{n^2} + \frac{2\phi_1^2(n-3)\gamma_0\gamma_1}{n^2} + \frac{2\phi_1^2\gamma_1^2}{n^2}\left\{\frac{(n-4)(1-\phi_1^{2(n-4)})}{1-\phi_1^2} - \frac{\phi_1^2\{1-(n-4)\phi_1^{2(n-5)} + (n-5)\phi_1^{2(n-4)}\}}{(1-\phi_1^2)^2}\right\}. \quad \text{(A3.11)}$$

If one ignores terms with ϕ_1^n and $1/n^2$, one obtains

$$\frac{1}{n^2}\sum_{t=1}^{n-1}\sum_{s=1}^{n-2}\gamma_{t-s-2}\gamma_{t+1-s} = \frac{1}{n}\left\{2\phi_1\gamma_1^2 + 2\phi_1^2\gamma_0\gamma_1 + \frac{2\phi_1^3\gamma_1^2}{1-\phi_1^2}\right\}. \qquad \text{(A3.12)}$$

If one puts (A3.10) and (A3.12) together, then (3.9) is obtained as follows:

$$cov(c_1, c_2) = \frac{2}{n}\left\{\gamma_0\gamma_1 + \frac{\phi_1\gamma_1^2}{1-\phi_1^2} + \phi_1\gamma_1^2 + \phi_1^2\gamma_0\gamma_1 + \frac{\phi_1^3\gamma_1^2}{1-\phi_1^2}\right\}$$

$$= \frac{2}{n}\left\{\phi_1\gamma_1^2 + \gamma_1^2(1+\phi_1^2)\left\{\gamma_0 + \frac{\phi_1\gamma_1}{1-\phi_1^2}\right\}\right\}.$$

REFERENCES

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control.* San Francisco: Holden-Day.

Brillinger, D. R. (1975), *Time Series, Data Analysis and Theory.* New York: Holt, Rinehart and Winston.

Hannan, E. J. (1975), "The Estimation of ARMA Models". *Annals of Statistics* **3**, 975–981.

Mann, H. B. and Wald, A. (1943), "On Stochastic Limit and Order relationships". *Annals of Mathematical Statistics* **14**, 217–226.

Marsaglia, G. (1976), "Random Number Generation". In Encyclopedia of Computer Science (ed. A. Ralston), New York: Petrocelli and Charter, 1192–1197.

Tsay, R. S. and Tiao, G. C. (1984), "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models". *Journal of the American Statistical Association* **79**, 84–96.

CHAPTER 4

ON THE FORECAST ERROR VARIANCE OF FITTED TIME SERIES MODELS

4.1 INTRODUCTION

For a given stochastic process, the structure is generally not known. One way of estimating this structure is by the use of finite parameters models. Based on a set of observations, the number of parameters can be determined and estimated. Estimation techniques such as maximum likelihood and least squares methods are generally used to estimate the parameters. Both methods are known to be asymptotically equivalent if the process is Gaussian. The ultimate use of the chosen model is for forecasting.

The linear least-squares theory of prediction assumes that observations from the infinite past are available. In practice, this is not a realistic assumption. However, as pointed out by Whittle (1983), the linear least-squares predictor remains valid for the autoregressive process in the finite sample situation, provided the order of the autoregression is smaller than the number of observations.

Several researchers have derived the asymptotic mean square error (a.m.s.e.) of the forecast for estimated finite parameter models. Akaike (1970) showed that for autoregressive models of order p (AR(p)), the variance of the one-step forecast is $\sigma_a^2(1 + p/n) + O_p(1/n^2)$, where n is the number of observations and σ_a^2 is the variance of the innovations. Bloomfield (1972) discussed aspects of nonparametric and parametric procedures for forecasts of time series and derived their a.m.s.e. of forecasts for a linear time series model. Using a Taylor series expansion, Bhansali (1974) derived the a.m.s.e. of forecasts of more than one-step ahead for a general autoregressive model. For the same model, Yamamoto (1976) provides an expression for the a.m.s.e. of the forecast. In the multivariate case, Baillie (1979) derived the a.m.s.e. of the forecasts of more than one-step ahead for the general vector autoregressive model. The a.m.s.e. of a multistep forecast for econometric models was given by Schmidt (1974); this was followed up by Schmidt (1977) in a finite sample study. The effect of using non-parsimonious time series models for forecasting was studied by Ledolter and Abraham (1981).

All the above studies of the a.m.s.e. of forecast are based on the assumption that the parameters of the model are estimated from an independent set of observations which have a similar covariance structure to the realization in which prediction is to be made. Such asymptotic results are important but it is also important to investigate the effect on the forecast error when this independence assumption is not made. Hence, when deriving the a.m.s.e., one should incorporate the dependence on the estimated parameter in the forecast. Phillips (1979) has examined the effect of this dependency on the sampling distribution of the forecast errors when the innovation sequence of an

AR(1) process is independent, identically distributed with mean zero and variance σ_a^2. Phillips showed that the conditional distribution of the forecast errors given the final period observation is skewed. Fuller and Hasza (1980, 1981), and Yamamoto (1981) have also examined this dependency.

Let $z_1, \ldots, z_n$ be a realization of n observations of a time series and consider the forecast error for the autoregressive-moving average model with both p and q being zero (ARMA(0,0)). This model is given by

$$z_t = \mu + a_t \ .$$

Then the one step ahead forecast is

$$\hat{z}_n(1) = \hat{\mu}$$

where $\hat{\mu} = n^{-1} \sum z_t$. The one step ahead forecast error is

$$\hat{e}_n(1) = z_{n+1} - \hat{z}_n(1)$$

$$= \mu + a_{n+1} - \hat{\mu}$$

$$= (\mu - \hat{\mu}) + a_{n+1} \ .$$

Then, the variance is

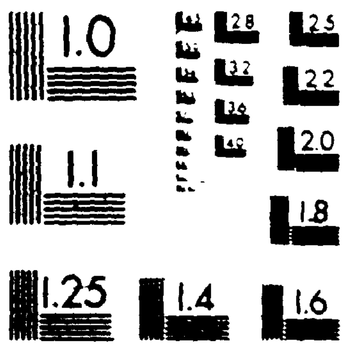
$$V\left(\hat{e}_t(1)\right) = V\left(\mu - \hat{\mu}\right) + \sigma_a^2$$

$$= \frac{\sigma_a^2}{n} + \sigma_a^2$$

$$= \sigma_a^2\left(1 + \frac{1}{n}\right)$$

It is clear from this last equation that there is no difference at all between the two situations. However, it is shown in the next section that this is not the case for the AR(1) process and an explicit expression for the variance of the ℓ-step ($\ell \geq 1$) ahead forecast error is derived. In section 4.3, the one-step ahead forecast error for the fractional noise model is also examined.

2 OF DE 2



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

4.2 VARIANCE OF ℓ-STEP AHEAD FORECAST ERROR OF AN AR(1) PROCESS

In this section, an explicit expression for the variance of the ℓ-step ahead forecast error of an AR(1) process is obtained. The effect on the variance of the forecast error when the parameter is estimated from the same data set upon which the forecast is based is taken into consideration.

Consider a realization of n observations $\{z_1, \ldots, z_n\}$. An autoregressive process of order one (AR(1)) is defined by the equation

$$z_t - \phi z_{t-1} = a_t \quad ,$$

where the elements of the sequence $\{a_t\}$ are assumed to be independent, identically and normally distributed with mean zero and variance σ_a^2. It is assumed here that the process is stationary, i.e., $|\phi| < 1$.

The forecast error can be written as follows:

$$e_t(\ell) = a_{t+\ell} + \phi a_{t+\ell-1} + \phi^2 a_{t+\ell-2} + \cdots + \phi^{\ell-1} a_{t+1}$$

$$= \sum_{j=0}^{\ell-1} \phi^j a_{t+\ell-j} \quad , \qquad (4.1)$$

where $e_t(\ell)$ denotes the ℓ-step ahead forecast error at origin time t, $t \geq n$. Let $\dot{\phi}$ be some arbitrary value of ϕ and let $\hat{\phi}$ be an asymptotically efficient estimate of ϕ based on n observations. Let $\dot{e}_t(\ell)$ be the value in (4.1) when $\phi = \dot{\phi}$ and $\hat{e}_t^{(n)}(\ell)$ be the estimate of $e_t(\ell)$, obtained by setting $\phi = \hat{\phi}$. Then the Taylor expansion up to first order about $\dot{\phi} = \phi$ and evaluated at $\hat{\phi}$, is given by

$$\hat{e}_t^{(n)}(\ell) = e_t(\ell) + (\hat{\phi} - \phi)\left(\frac{\partial \dot{e}_t(\ell)}{\partial \dot{\phi}}\right) , \tag{4.2}$$

where $(\partial/\partial\dot{\phi})(\dot{e}_t(\ell))$ is the derivative of $\dot{e}_t(\ell)$ with respect to $\dot{\phi}$ in (4.1).

Define the auxiliary time series $\{v_t\}$ by the equation

$$v_t - \phi v_{t-1} = -a_t \ .$$

Then $(\partial/\partial\phi)(a_t) = v_{t-1}$. Hence, the derivative of $e_t(\ell)$ with respect to ϕ is

$$\frac{\partial e_t^{(n)}(\ell)}{\partial\phi} = \frac{\partial}{\partial\phi}(a_{t+\ell}) + \frac{\partial}{\partial\phi}(\phi a_{t+\ell-1}) + \cdots + \frac{\partial}{\partial\phi}(\phi^{\ell-1}a_{t+1})$$

$$= v_{t+\ell-1} + a_{t+\ell-1} + \phi v_{t+\ell-2} + 2\phi a_{t+\ell-2} + \cdots$$
$$+ (\ell-1)\phi^{\ell-2}a_{t+1} + \phi^{\ell-1}v_t$$

$$= (v_{t+\ell-1} + \phi v_{t+\ell-2} + \cdots + \phi^{\ell-1}v_t)$$
$$+ (a_{t+\ell-1} + 2\phi a_{t+\ell-2} + \cdots + (\ell-1)\phi^{\ell-2}a_{t+1})$$

$$= A_t + R_t \tag{4.3}$$

where $A_t = \sum_{i=0}^{\ell-1}\phi^i v_{t+\ell-1-i}$ and $R_t = \sum_{k=0}^{\ell-2}(k+1)\phi^k a_{t+\ell-1-k}$. It can be shown that both A_t and R_t are $O_p(1)$ (appendix A4.1). From a lemma given by McLeod (1978),

$$\hat{\phi} - \phi = I^{-1} s_c + O_p(\frac{1}{n}) \quad , \tag{4.4}$$

where I is the large sample information matrix per observation and s_c is

$$s_c = -\frac{1}{n} \sum_{t=1}^{n} a_t v_{t-1} \quad .$$

Now from (4.3) and (4.4), equation (4.2) can be rewritten as follows:

$$\hat{e}_t^{(n)}(\ell) = e_t(\ell) + A_t I^{-1} s_c + R_t I^{-1} s_c + O_p(\frac{1}{n}) \quad . \tag{4.5}$$

Note that for an AR(1) process, $I^{-1} = (1-\phi^2)/\sigma_a^2$. Hence, apart from terms of order $O_p(1/n^2)$, the variance of (4.5) is

$$V\left(\hat{e}_t^{(n)}(\ell)\right) = V\left(e_t(\ell) + \frac{A_t(1-\phi^2)s_c}{\sigma_a^2} + \frac{R_t(1-\phi^2)s_c}{\sigma_a^2}\right)$$

$$\begin{aligned} &= V(e_t(\ell)) + \frac{(1-\phi^2)^2}{\sigma_a^4} V(A_t s_c) + \frac{(1-\phi^2)^2}{\sigma_a^4} V(R_t s_c) \\ &+ \frac{2(1-\phi^2)^2}{\sigma_a^4} cov(A_t s_c, \ R_t s_c) + \frac{2(1-\phi^2)}{\sigma_a^2} cov(e_t(\ell), \ R_t s_c) \\ &+ \frac{2(1-\phi^2)}{\sigma_a^2} cov(e_t(\ell), \ A_t s_c) \quad . \end{aligned} \tag{4.6}$$

It is easily shown that the last two terms on the right side of (4.6) are zero since

$$cov(e_t(\ell), \ R_t s_c) = cov\Big(\sum_{j=0}^{\ell-1} \phi^j a_{t+\ell-j}, \\ -\frac{1}{n}\sum_{k=0}^{\ell-2}\sum_{d=1}^{n}(k+1)\phi^k a_{t+\ell-1-k} a_d v_{d-1}\Big)$$

$$= -\frac{1}{n}\sum_{j=0}^{\ell-1}\sum_{k=0}^{\ell-2}\sum_{d=1}^{n}(k+1)\phi^k \langle a_{t+\ell-j}\, a_{t+\ell-1-k}\, a_d v_{d-1}\rangle$$

$$= 0.$$

The above equation is obtained by the fourth moment result and the fact that $\langle a_d v_{d-1}\rangle = 0$ for all d. Similarly, it can be shown that the last term in (4.6) is zero. Hence, (4.6) becomes

$$V(\hat{e}_t^{(n)}(\ell)) = V(e_t(\ell)) + \frac{(1-\phi^2)^2}{\sigma_a^4}V(A_t s_c) \\ + \frac{(1-\phi^2)^2}{\sigma_a^4}V(R_t s_c) + \frac{2(1-\phi^2)^2}{\sigma_a^4}cov(A_t s_c, \ R_t s_c) \quad . \quad (4.7)$$

Each term on the right side of (4.7) can now be evaluated individually. The first term in (4.7) can be easily shown to be

$$V(e_t(\ell)) = V\Big(\sum_{j=0}^{\ell-1}\phi^j a_{t+\ell-j}\Big)$$

$$= \sum_{j=0}^{\ell-1}\phi^{2j}\sigma_a^2$$

$$= \sigma_a^2\Big(\frac{1-\phi^{2\ell}}{1-\phi^2}\Big) \quad . \quad (4.8)$$

The second, third and fourth terms on the right side of (4.7) are evaluated in appendices A4.2, A4.3 and A4.4 respectively. From the results in the appendices, the second term may be shown to be

$$\begin{aligned}\frac{(1-\phi^2)^2}{\sigma_a^4}V(A_t s_c) = {} & \frac{\sigma_a^2}{n}\left(\frac{1-\phi^{2\ell}}{1-\phi^2}\right) + \frac{2\sigma_a^2}{n}\sum_{i=1}^{\ell-1} i\phi^{2i} \\ & + \frac{2\ell^2\sigma_a^2\phi^{2t-2n+2\ell-2}(1-\phi^{2n})}{n^2} \\ & + \frac{8\ell^2\sigma_a^2\phi^{2t-2n+2\ell}(1-\phi^{2n-2})}{n^2(1-\phi^2)} \\ & - \frac{8\ell^2(n-1)\sigma_a^2\phi^{2t+2\ell-2}}{n^2} \\ & + \frac{2\ell^2\sigma_a^2\phi^{2t-2n+2\ell}(1-\phi^{2n})}{n^2(1-\phi^2)}. \end{aligned} \tag{4.9}$$

Similarly the third term of (4.7) may be shown to be

$$\frac{(1-\phi^2)^2}{\sigma_a^4}V(R_t s_c) = \frac{\sigma_a^2(1-\phi^2)}{n}\left(\sum_{k=0}^{\ell-2} k^2\phi^{2k} + 2\sum_{k=0}^{\ell-2} k\phi^{2k} + \frac{1-\phi^{2(\ell-1)}}{1-\phi^2}\right). \tag{4.10}$$

Finally the last term of (4.7) may be shown to be

$$\frac{2(1-\phi^2)^2}{\sigma_a^4}cov(A_t s_c,\ R_t s_c) = -\frac{2\ell(1-\phi^2)\sigma_a^2}{n}\left(\sum_{k=1}^{\ell-2} k\phi^{2k} + \frac{1-\phi^{2(\ell-1)}}{1-\phi^2}\right). \tag{4.11}$$

Expressions (4.8), (4.9), (4.10) and (4.11), may be used to evaluate equation (4.7) as follows:

$$\begin{aligned} V(\hat{e}_t^{(n)}(\ell)) = {} & \sigma_a^2\left(\frac{1-\phi^{2\ell}}{1-\phi^2}\right) + \frac{\sigma_a^2}{n}\left(\frac{1-\phi^{2\ell}}{1-\phi^2}\right) + \frac{2\sigma_a^2}{n}\sum_{i=1}^{\ell-1} i\phi^{2i} \\ & + \frac{2\ell\sigma_a^2\phi^{2t-2n+2\ell-2}(1-\phi^{2n})}{n^2} + \frac{8\ell^2\sigma_a^2\phi^{2t-2n+2\ell}(1-\phi^{2n-2})}{n^2(1-\phi^2)} \\ & - \frac{8\ell^2\sigma_a^2\phi^{2t+2\ell-2}}{n^2} + \frac{2\ell^2\sigma_a^2\phi^{2t-2n+2\ell}(1-\phi^{2n})}{n^2(1-\phi^2)} \end{aligned}$$

$$+\frac{\sigma_a^2(1-\phi^2)}{n}\Big(\sum_{k=0}^{\ell-2}k^2\phi^{2k}+2\sum_{k=0}^{\ell-2}k\phi^{2k}+\frac{1-\phi^{2(\ell-1)}}{1-\phi^2}\Big)$$

$$-\frac{2\ell(1-\phi^2)\sigma_a^2}{n}\Big(\sum_{k=0}^{\ell-2}k\phi^{2k}+\frac{1-\phi^{2(\ell-1)}}{1-\phi^2}\Big) \quad (4.12)$$

If one ignores terms which are $O(1/n^2)$, then

$$V(\hat{e}_t^{(n)}(\ell)) = \sigma_a^2\Big(\frac{1-\phi^{2\ell}}{1-\phi^2}\Big)+\frac{\sigma_a^2}{n}\Big\{\Big(\frac{1-\phi^{2\ell}}{1-\phi^2}\Big)-8\ell^2\phi^{2t+2\ell-2}$$

$$+2\sum_{i=1}^{\ell-1}i\phi^{2i}+(1-\phi^2)\Big(\sum_{k=0}^{\ell-2}k^2\phi^{2k}+2\sum_{k=0}^{\ell-2}k\phi^{2k}+\frac{1-\phi^{2(\ell-1)}}{1-\phi^2}\Big)$$

$$-2\ell(1-\phi^2)\Big(\sum_{k=0}^{\ell-2}k\phi^{2k}+\frac{1-\phi^{2(\ell-1)}}{1-\phi^2}\Big)\Big\} \quad (4.13)$$

It can be seen from (4.13) that if n increases to infinity, (4.13) yields the same expression as given by Box and Jenkins (1976, p. 151),

$$V(\hat{e}_t^{(n)}(\ell)) = \sigma_a^2\Big(\frac{1-\phi^{2\ell}}{1-\phi^2}\Big) \qquad \ell \geq 1 , \quad (4.14)$$

and as ℓ increases, (4.14) achieves a constant variance of $\sigma_a^2/(1-\phi^2)$. Letting $\ell = 1$ in (4.13), gives

$$V(\hat{e}_t^{(n)}(1)) = \sigma_a^2+\frac{\sigma_a^2}{n}(1-8\phi^{2t}) \quad (4.15)$$

Equation (4.15) is obtained due to the fact that $\sum_{k=0}^{\ell-1}k\phi^{2k}$, $\sum_{k=0}^{\ell-2}k\phi^{2k}$ and $\sum_{k=0}^{\ell-2}k^2\phi^{2k}$ are zero when $\ell = 1$. Clearly as t tends to infinity, (4.15) becomes

$$V(\hat{e}_t^{(n)}(1)) = \sigma_a^2(1+\frac{1}{n}) ,$$

which agrees with the well known result of Akaike (1970). However for small t the variance as given in (4.15) is slightly smaller.

4.3 VARIANCE OF THE ONE-STEP FORECAST ERROR IN THE FRACTIONAL NOISE MODEL

Fractional ARMA models have been studied by authors such as Hosking (1981) and Li and McLeod (1986) and are useful in modelling riverflow time series. This model is similar to the ARIMA(p,d,q) model except that d is allowed to take on nonintegral values. When the parameter d lies in $0 < d < 1/2$, the process is capable of modelling long-term persistence. Hence, d is assumed to lie in this range. Consider the Fractional Noise model, FARMA$(0,d,0)$, given by

$$\nabla^d z_t = a_t,$$

or equivalently by,

$$z_t = \nabla^{-d} a_t,$$

where $\nabla^d = (1 - B)^d$ and B is as defined in chapter 2. Expansion into moving-average form yields the following:

$$z_t = \sum_{j=0}^{\infty} (-1)^j \binom{-d}{j} a_{t-j}.$$

Jimenez and McLeod (1986) showed that for some finite m,

$$z_t = \sum_{j=0}^{m} (-1)^j \binom{-d}{j} a_{t-j} + O_p\left(\frac{1}{m^{1/2}}\right).$$

If only n observations $z_1, \ldots, z_n$ are available, then, the one-step ahead forecast is

$$z_t(1) = \sum_{j=1}^{n} (-1)^j \binom{-d}{j} a_{t+1-j} + O_p\left(\frac{1}{n^{1/2}}\right), \qquad t \geq n,$$

and the corresponding forecast error is

$$e_t(1) = a_{t+1} + O_p\left(\frac{1}{n^{1/2}}\right).$$

Let $\hat{e}_t^{(n)}(1)$ denote the estimator of $e_t(1)$ given observations $z_1, \ldots, z_n$. If one expands in Taylor series up to first order terms and neglects terms which are smaller than $O_p(1/n^{1/2})$, one obtains

$$\begin{aligned}\hat{e}_t^{(n)}(1) &= e_t(1) + (\hat{d} - d)\left(\frac{\partial a_{t+1}}{\partial d}\right) \\ &= a_{t+1} + I^{-1} s_c\left(\frac{\partial a_{t+1}}{\partial d}\right). \qquad (4.16)\end{aligned}$$

It can be shown that $(\partial^2/\partial d^2)(a_t) = O_p(1)$ (see appendix A4.5). From Li and McLeod. (1986), $\delta_t = (\partial/\partial d)(a_t) = (\log \nabla) a_{t+1}$; $\log \nabla = -B - \frac{1}{2}B^2 - \frac{1}{3}B^3 - \cdots$ and $V(\delta_t) = \frac{1}{6}\pi^2\sigma_a^2$. Hence, $I^{-1} = 6/(\pi^2\sigma_a^2)$. Li (1981) shows that $s_c = -n^{-1}\sum a_b \delta_{b-1}$. Hence (4.16) can be rewritten as follows:

$$V\left(\hat{e}_t^{(n)}(1)\right) = V\left(a_{t+1}\right) + \left(\frac{6}{n\pi^2\sigma_a^2}\right)^2 V\left(\sum_{b=1}^{n} a_b \delta_{b-1} \delta_t\right). \qquad (4.17)$$

Now $V(a_{t+1}) = \sigma_a^2$ and we may write

$$V\left(\sum_b a_b \delta_{b-1} \delta_t\right) = \sum_{b=1}^{n}\sum_{c=1}^{n} cum\left(a_b \delta_{b-1} \delta_t,\ \ a_c \delta_{c-1} \delta_t\right). \qquad (4.18)$$

The following results are useful in the derivation of the variance.

(i)

$$\begin{aligned}\langle a_b \delta_t \rangle &= -\sum_{k=1}^{\infty} \frac{1}{k} \langle a_b a_{t+1-k} \rangle \\ &= -\frac{1}{t+1-b}\sigma_a^2.\end{aligned}$$

(ii)

$$\langle \delta_{b-1}\delta_t \rangle = \langle \sum_{k=1}^{\infty} \frac{1}{k} a_{b-k} \quad \sum_{j=1}^{\infty} \frac{1}{j} a_{t+1-j} \rangle$$

$$= \sum_{k=1}^{\infty} \frac{1}{k(t+1-b+k)} \sigma_a^2$$

$$= -\sum_{k=1}^{\infty} \frac{1}{t+1-b} \left(\frac{1}{t+1-b+k} - \frac{1}{k} \right) \sigma_a^2$$

$$= \frac{1}{t+1-b} \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-b} \right) \sigma_a^2.$$

(iii)

$$\langle \delta_t^2 \rangle = \frac{\pi^2}{6} \sigma_a^2.$$

(iv)

$$\langle \delta_{c-1}\delta_t \rangle = \langle \sum_{k=1}^{\infty} \frac{1}{k} a_{c-k} \quad \sum_{j=1}^{\infty} \frac{1}{j} a_{t+1-j} \rangle$$

$$= \sum_{k} \frac{1}{k(t+1-c+k)} \sigma_a^2$$

$$= \frac{1}{t+1-c} \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-c} \right) \sigma_a^2.$$

(v)

$$\langle a_c \delta_{b-1} \rangle = -\sum_{k=1}^{\infty} \frac{1}{k} \langle a_c a_{b-k} \rangle \qquad b > c$$

$$= -\sum_{k=1}^{\infty} \frac{1}{k} \langle a_c a_{b-k} \rangle$$

$$= -\frac{1}{b-c} \sigma_a^2.$$

(vi)

$$\sum_{b,c} \langle \delta_{b-1}\delta_{c-1} \rangle = \sum_{b} \langle \delta_b^2 \rangle + 2 \sum_{b=2}^{n} \sum_{c=1}^{b-1} \langle \delta_b \delta_c \rangle$$

$$= \sum_{b} \langle \delta_b^2 \rangle + 2 \sum_{b=2}^{n} \sum_{c=1}^{b-1} \frac{1}{b-c} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{b-c} \right) \sigma_a^2.$$

If one applies both the above results and the results from appendix A4.6 and neglect terms which are smaller order than $O(1/n)$, one obtains the following:

$$V\left(\hat{e}_t^{(n)}(1)\right) = \sigma_a^2\left(1 + \frac{1}{n}\right). \tag{4.19}$$

Remark: Equation (4.15) shows that $V\left(\hat{e}_t^{(n)}(1)\right)$ depends on t in the AR(1) model but equation (4.19) shows that $V\left(\hat{e}_t^{(n)}(1)\right)$ does not depend on t in the FARMA(0,d,0) model (neglecting terms which are smaller order than $1/n$).

4.4 CONCLUSION

Section 4.2 demonstrates that in the first order AR model, the variance of the one-step ahead forecast error is shown to be affected by the estimated parameter when the parameter is estimated from the same data set upon which the forecasting is based. The effect of estimating the parameter is a reduction in the variance of the forecast error. It is shown that when the forecast origin is increased ($t \to \infty$), the usual asymptotic result remains valid. Essentially, the case of $t \to \infty$ is similar to that where an independent data set having similar characteristics is available for estimating the parameters. The normality assumption was assumed; if this assumption had not been made, then one would be required to consider all product moments of order larger than two; the derivation then becomes intractable. An extension to the case of AR(p) or ARMA(p,q) models using this method is also seen to be intractable. It is seen that in the fractional noise model, the variance of the forecast error does not depend on the parameter d.

APPENDIX A4.1

Derivation of the result $A_t = O_p(1)$

Since $\langle A_t^2 \rangle$ involves only squared and cross product terms, hence $\langle A_t^2 \rangle < \infty$. It follows by Chebyshev's inequality that

$$\Pr\{| A_t |> C\} \leq \frac{\langle A_t^2 \rangle}{C^2}.$$

Let $\epsilon = \langle A_t^2 \rangle / C^2$ and $g_n = 1$. Then

$$\Pr\{| A_t |> M_\epsilon g_n\} \leq \epsilon,$$

where $M_\epsilon = \{\langle A_t^2 \rangle / \epsilon\}^{1/2}$. Therefore A_t is $O_p(1)$.

APPENDIX A4.2

Derivation of the second term in (4.7)

To evaluate (4.7), one requires the properties of cumulants (Brillinger, 1975). Denote the cumulant operator by "cum", and note that $cum(x,y) = cov(x,y)$. Then

$$\frac{(1-\phi^2)^2}{\sigma_a^4} V(A_t s_c)$$
$$= \frac{(1-\phi^2)^2}{n^2\sigma_a^4} \sum_{i,j=0}^{\ell-1} \sum_{b,d=1}^{n} \phi^{i+j}\, cum(v_{t+\ell-1-i}a_d v_{d-1}, v_{t+\ell-1-j}a_b v_{b-1}) \quad \text{(A4.1)}$$

Using Theorem 2.3.2 of Brillinger and note that only those indecomposable partitions with elements of size two need to be considered, since the process is Gaussian and $\langle v_t\rangle = \langle a_t\rangle = 0$. Therefore,

$$\begin{aligned} &cum(v_{t+\ell-1-i}a_d v_{d-1},\ v_{t+\ell-1-j}a_b v_{b-1}) \\ &\quad = cum(v_{t+\ell-1-i},\ a_d)\Big\{cum(v_{d-1},\ a_b)cum(v_{t+\ell-1-j},\ v_{b-1}) \\ &\qquad\qquad + cum(v_{d-1},\ v_{b-1})cum(v_{t+\ell-1-j},\ a_b)\Big\} \\ &\quad + cum(v_{t+\ell-1-i},\ v_{d-1})\Big\{cum(a_d,\ a_b)cum(v_{t+\ell-1-j},\ v_{b-1}) \\ &\qquad\qquad + cum(a_d,\ v_{b-1})cum(v_{t+\ell-1-j},\ a_b)\Big\} \\ &\quad + cum(v_{t+\ell-1-i},\ v_{t+\ell-1-j})cum(a_d,\ a_b)cum(v_{d-1},\ v_{b-1}) \\ &\quad + cum(v_{t+\ell-1-i},\ a_b)\Big\{cum(a_d,\ v_{t+\ell-1-j})cum(v_{d-1},\ v_{b-1}) \\ &\qquad\qquad + cum(a_d,\ v_{b-1})cum(v_{d-1},\ v_{t+\ell-1-j})\Big\} \\ &\quad + cum(v_{t+\ell-1-i},\ v_{b-1})\Big\{cum(a_d,\ v_{t+\ell-1-j})cum(v_{d-1},\ a_b) \\ &\qquad\qquad + cum(a_d,\ a_b)cum(v_{d-1},\ v_{t+\ell-1-j})\Big\}. \end{aligned} \quad \text{(A4.2)}$$

After gathering terms, (A4.1) becomes

$$\frac{(1-\phi^2)^2}{\sigma_a^4} V(A_t s_c)$$

$$= \frac{(1-\phi^2)^2}{n^2\sigma_a^4}\Bigg\{ \sum_{i,j=0}^{\ell-1} \sum_{d=1}^{n} \phi^{i+j} cov(v_{t+\ell-1-i},\ v_{t+\ell-1-j}) cov(a_d,\ a_d) \times cov(v_{d-1},\ v_{d-1})$$

$$+ 2 \sum_{i,j=0}^{\ell-1} \sum_{b,d=1}^{n} \phi^{i+j} cov(v_{t+\ell-1-i},\ a_b) cov(a_d,\ v_{t+\ell-1-j}) \times cov(v_{d-1},\ v_{b-1})$$

$$+ 4 \sum_{i,j=0}^{\ell-1} \sum_{b=2}^{n} \sum_{d=1}^{b-1} \phi^{i+j} cov(v_{t+\ell-1-i},\ a_b) cov(a_d,\ v_{b-1}) \times cov(v_{d-1},\ v_{t+\ell-1-j})$$

$$+ 2 \sum_{i,j=0}^{\ell-1} \sum_{d=1}^{n} \phi^{i+j} cov(v_{t+\ell-1-i},\ v_{d-1}) cov(a_d,\ a_d) \times cov(v_{d-1},\ v_{t+\ell-1-j}) \Bigg\}. \qquad \text{(A4.3)}$$

The first term on the right side of (A4.3) is

$$\sum_{i,j=0}^{\ell-1} \sum_{d=1}^{n} \phi^{i+j} cov(v_{t+\ell-1-i},\ v_{t+\ell-1-j}) cov(a_d,\ a_d) cov(v_{d-1},\ v_{d-1})$$

$$= \sum_{i=0}^{\ell-1} \sum_{d=1}^{n} \phi^{2i} cov(v_{t+\ell-1-i},\ v_{t+\ell-1-i}) cov(a_d,\ a_d) cov(v_{d-1},\ v_{d-1})$$

$$+ 2 \sum_{i=1}^{\ell-1} \sum_{j=0}^{i-1} \sum_{d=1}^{n} \phi^{i+j} cov(v_{t+\ell-1-i},\ v_{t+\ell-1-j}) cov(a_d,\ a_d) \times cov(v_{d-1},\ v_{d-1})$$

$$= \sum_{i=0}^{\ell-1} \sum_{d=1}^{n} \phi^{2i} \left(\frac{\sigma_a^2}{1-\phi^2}\right)^2 \sigma_a^2 + 2 \sum_{i=1}^{\ell-1} \sum_{j=0}^{i-1} \sum_{d=1}^{n} \phi^{i+j} \sigma_a^2 \left(\frac{\phi^{j-i}\sigma_a^2}{1-\phi^2}\right)\left(\frac{\sigma_a^2}{1-\phi^2}\right)$$

$$= \frac{n\sigma_a^6}{1-\phi^2}\sum_{i=0}^{\ell-1}\phi^{2i} + \frac{2\sigma_a^6}{(1-\phi^2)^2}\sum_{i=1}^{\ell-1} i\phi^{2i}$$

$$= \frac{n\sigma_a^6(1-\phi^{2\ell})}{(1-\phi^2)^3} + \frac{2n\sigma_a^6}{(1-\phi^2)^2}\sum_{i=1}^{\ell-1} i\phi^{2i} \qquad \text{(A4.4)}$$

The second term of (A4.3) is

$$2\sum_{i,j=0}^{\ell-1}\sum_{b,d=1}^{n}\phi^{i+j}\,cov(v_{t+\ell-1-i},\ a_b)cov(a_d,\ v_{t+\ell-1-j})cov(v_{d-1},\ v_{b-1})$$

$$= 2\sum_{i,j=0}^{\ell-1}\sum_{d=1}^{n}\phi^{i+j}\,cov(v_{t+\ell-1-i},\ a_d)cov(a_d,\ v_{t+\ell-1-j})cov(v_{d-1},\ v_{d-1})$$

$$+ 4\sum_{i,j=0}^{\ell-1}\sum_{d=2}^{n}\sum_{b=1}^{d-1}\phi^{i+j}\,cov(v_{t+\ell-1-i},\ a_b)cov(a_d,\ v_{t+\ell-1-j}) \times cov(v_{d-1},\ v_{b-1})$$

$$= 2\sum_{i,j=0}^{\ell-1}\sum_{d=1}^{n}\phi^{i+j}\left(-\phi^{t+\ell-i-d-1}\sigma_a^2\right)\left(-\phi^{t+\ell-j-d-1}\sigma_a^2\right)\left(\frac{\sigma_a^2}{1-\phi^2}\right)$$

$$+ 4\sum_{i,j}^{\ell-1}\sum_{d=2}^{n}\sum_{b=1}^{d-1}\phi^{i+j}\left(-\phi^{t+\ell-i-b-1}\sigma_a^2\right)\left(-\phi^{t+\ell-j-d-1}\sigma_a^2\right)\left(\frac{\phi^{d-b}\sigma_a^2}{1-\phi^2}\right)$$

$$= \frac{2\ell^2\sigma_a^6\phi^{2t+2\ell-2}}{1-\phi^2}\sum_{d=1}^{n}\left(\frac{1}{\phi^2}\right)^d + \frac{4\ell^2\sigma_a^6\phi^{2t+2\ell-2}}{1-\phi^2}\sum_{d=2}^{n}\sum_{b=1}^{d-1}\left(\frac{1}{\phi^2}\right)^b$$

$$= \frac{2\ell^2\sigma_a^6\phi^{2t-2n+2\ell-2}(1-\phi^{2n})}{(1-\phi^2)^2} + \frac{4\ell^2\sigma_a^6\phi^{2t-2n+2\ell}(1-\phi^{2n-2})}{(1-\phi^2)^3}$$

$$- \frac{4\ell^2(n-1)\sigma_a^6\phi^{2t+2\ell-2}}{(1-\phi^2)^2} \qquad \text{(A4.5)}$$

The third term of (A4.3) is

$$4\sum_{i,j=0}^{\ell-1}\sum_{b=2}^{n}\sum_{d=1}^{b-1}\phi^{i+j}cov(v_{t+\ell-1-i},\ a_b)cov(a_d,\ v_{b-1})cov(v_{d-1},\ v_{t+\ell-1-j})$$

$$= 4\sum_{i,j=0}^{\ell-1}\sum_{b=2}^{n}\sum_{d=1}^{b-1}\phi^{i+j}\left(-\phi^{t+\ell-1-i-b}\sigma_a^2\right)\left(-\phi^{b-1-d}\sigma_a^2\right) \times\left(\frac{\phi^{t+\ell-1-j-d+1}\sigma_a^2}{1-\phi^2}\right)$$

$$= \frac{4\ell^2\sigma_a^6\phi^{2t+2\ell-2}}{1-\phi^2}\sum_{b=2}^{n}\sum_{d=1}^{b-1}\left(\frac{1}{\phi^2}\right)^d$$

$$= \frac{4\ell^2\sigma_a^6\phi^{2t-2n+2\ell}(1-\phi^{2n-2})}{(1-\phi^2)^3} - \frac{4\ell^2\sigma_a^6(n-1)\phi^{2t+2\ell-2}}{(1-\phi^2)^2} . \qquad (A4.6)$$

The last term of (A4.3) is

$$2\sum_{i,j=0}^{\ell-1}\sum_{d=1}^{n}\phi^{i+j}cov(v_{t+\ell-1-i},\ v_{d-1})cov(a_d,\ a_d)cov(v_{d-1},\ v_{t+\ell-1-j})$$

$$= 2\sum_{i,j=0}^{\ell-1}\sum_{d=1}^{n}\phi^{i+j}\left(\frac{\phi^{t+\ell-1-i-d+1}\sigma_a^2}{1-\phi^2}\right)\sigma_a^2\left(\frac{\phi^{t+\ell-1-j-d+1}\sigma_a^2}{1-\phi^2}\right)$$

$$= \frac{2\ell^2\sigma_a^6\phi^{2t+2\ell}}{(1-\phi^2)^2}\sum_{d=1}^{n}\left(\frac{1}{\phi^2}\right)^d$$

$$= \frac{2\ell^2\sigma_a^6\phi^{2t-2n+2\ell}(1-\phi^{2n})}{(1-\phi^2)^3} . \qquad (A4.7)$$

If one substitutes (A4.4), (A4.5), (A4.6) and (A4.7) into (A4.3) one obtain equation (4.9).

APPENDIX A4.3

Derivation of the third term of (4.7)

The method of evaluation of third term in (4.7) is similar to the method used in appendix A4.1.

$$\frac{(1-\phi^2)^2}{\sigma_a^4} V(R_t s_c)$$

$$= \frac{(1-\phi^2)^2}{n^2\sigma_a^4} \sum_{i,k=0}^{\ell-2} \sum_{b,d=1}^{n} (k+1)(i+1)\phi^{k+i}$$
$$\times cov(a_{t+\ell-1-k}\, a_d v_{d-1},\ a_{t+\ell-1-i}\, a_b v_{b-1})$$

$$= \frac{(1-\phi^2)^2}{n^2\sigma_a^4} \sum_{i,k=0}^{\ell-2} \sum_{b,d=1}^{n} (k+1)(i+1)\phi^{i+k} cov(a_{t+\ell-1-k},\ a_{t+\ell-1-i}) cov(a_d,\ a_b)$$
$$\times cov(v_{b-1},\ v_{d-1})$$

$$= \frac{(1-\phi^2)^2}{n^2\sigma_a^4} \sum_{k=0}^{\ell-2} \sum_{d=1}^{n} (k+1)^2\phi^{2k} cov(a_{t+\ell-1-k},\ a_{t+\ell-1-k}) cov(a_d,\ a_d)$$
$$\times cov(v_{d-1},\ v_{d-1})$$

$$= \frac{(1-\phi^2)^2}{n^2\sigma_a^4} \sum_{k=0}^{\ell-2} \sum_{d=1}^{n} (k+1)^2\phi^{2k}\sigma_a^4\left(\frac{\sigma_a^2}{1-\phi^2}\right)$$

$$= \frac{\sigma_a^2(1-\phi^2)}{n}\left\{\sum_{k=0}^{\ell-2} k^2\phi^{2k} + 2\sum_{k=0}^{\ell-2} k\phi^{2k} + \frac{1-\phi^{2(\ell-1)}}{1-\phi^2}\right\}. \qquad \text{(A4.8)}$$

Hence, equation (4.10) is obtained.

APPENDIX A4.4

Derivation of the fourth term of (4.7)

A similar technique to that used in the previous appendices is used in the following derivation.

$$\frac{2(1-\phi^2)^2}{\sigma_a^4} cov(A_t s_c,\ R_t s_c)$$

$$= \frac{2(1-\phi^2)^2}{n^2\sigma_a^4} \sum_{i=0}^{\ell-1}\sum_{k=0}^{\ell-2}\sum_{b,d=1}^{n} (k+1)\phi^{i+k} cov(v_{t+\ell-1-i}a_d v_{d-1},\ v_{t+\ell-1-k}a_b v_{b-1})$$

$$= \frac{2(1-\phi^2)^2}{n^2\sigma_a^4} \sum_{i=0}^{\ell-1}\sum_{k=0}^{\ell-2}\sum_{d=1}^{n} (k+1)\phi^{i+k} cov(v_{t+\ell-1-i},\ a_{t+\ell-1-k}) cov(a_d,\ a_d) \times cov(v_{d-1},\ v_{d-1})$$

$$= \frac{2(1-\phi^2)^2}{n\sigma_a^4} \sum_{i=0}^{\ell-1}\sum_{k=0}^{\ell-2}\sum_{d=1}^{n} (k+1)\phi^{i+k} \left(-\phi^{k-i}\sigma_a^2\right)\sigma_a^2\left(\frac{\sigma_a^2}{1-\phi^2}\right)$$

$$= -\frac{2\ell(1-\phi^2)\sigma_a^2}{n}\sum_{k=0}^{\ell-2}(k+1)\phi^{2k}$$

$$= -\frac{2\ell(1-\phi^2)\sigma_a^2}{n}\left\{\sum_{k=0}^{\ell-2} k\phi^{2k} + \frac{1-\phi^{2(\ell-1)}}{1-\phi^2}\right\}, \qquad (A4.9)$$

which is equation (4.11).

APPENDIX A4.5

Derivation of the result $(\partial^2/\partial d^2)(a_t) = O_p(1)$

To show that the second derivative of a_t with respect to d is a bounded quantity, note that

$$\begin{aligned}\frac{\partial^2 a_t}{\partial d^2} &= (\log \nabla)^2 a_t \\ &= (\log \nabla)\Big(-\sum_{k=1}^{\infty} \frac{1}{k} a_{t-k}\Big) \\ &= \sum_{k=2}^{\infty} c_k a_{t-k},\end{aligned}$$

where

$$c_k = \sum_{j=1}^{k-1} \frac{1}{j(k-j)}.$$

Now, c_k can be rewritten as,

$$\begin{aligned}c_k &= \sum_{j=1}^{k-1} \Big(\frac{1}{j} + \frac{1}{k-j}\Big)\frac{1}{k} \\ &= \frac{2}{k}\sum_{j=1}^{k-1} \frac{1}{j} \\ &\approx \frac{2}{k} \log k.\end{aligned}$$

Since $\langle a_{t-j} a_{t-k}\rangle = 0$ for $k \neq j$ and $\langle a_t^2\rangle = \sigma_a^2$, then (ignoring the constant quantity),

$$\Big\langle\Big(\sum_{k=2}^{\infty} \frac{\log k}{k} a_{t-k}\Big)^2\Big\rangle = \sum_{k=2}^{\infty} \Big(\frac{\log k}{k}\Big)^2 \sigma_a^2,$$

and $\sum_2^{\infty} (\log k/k)^2$ is a convergent series. The result follows from Chebyshev's inequality.

APPENDIX A4.6

Derivation of the variance in (4.18)

In the derivation, results (i) to (vi) in section 4.2 and Theorem 2.3.2 of Brillinger (1975) will be used. First, note that all those nonzero cumulants of order two are given and that $cum(x, y) = \langle xy \rangle$. Hence,

$$
\begin{aligned}
& cum(a_b \delta_{b-1} \delta_t,\ a_c \delta_{c-1} \delta_t) \\
&= \langle a_b \delta_t \rangle \Big[\langle \delta_{b-1} a_c \rangle \langle \delta_{c-1} \delta_t \rangle + \langle \delta_{b-1} \delta_{c-1} \rangle \langle a_c \delta_t \rangle \Big] \\
&\quad + \langle a_b a_c \rangle \Big[\langle \delta_{b-1} \delta_t \rangle \langle \delta_{c-1} \delta_t \rangle + \langle \delta_{b-1} \delta_{c-1} \rangle \langle \delta_t \delta_t \rangle \\
&\qquad + \langle \delta_{b-1} \delta_t \rangle \langle \delta_t \delta_{c-1} \rangle \Big] \\
&\quad + \langle a_b \delta_{c-1} \rangle \Big[\langle \delta_{b-1} \delta_t \rangle \langle a_c \delta_t \rangle + \langle \delta_{b-1} \delta_t \rangle \langle \delta_t a_c \rangle \Big] \\
&\quad + \langle a_b \delta_t \rangle \Big[\langle \delta_{b-1} \delta_{c-1} \rangle \langle \delta_t a_c \rangle + \langle \delta_{b-1} a_c \rangle \langle \delta_t \delta_{c-1} \rangle \Big].
\end{aligned}
$$

If one incorporates the sums and evaluates each of the terms individually using results (i) to (vi) in section 2, the following results are obtained.

1.

$$
\begin{aligned}
& \sum_{b,c} \langle a_b \delta_t \rangle \langle \delta_{b-1} a_c \rangle \langle \delta_{c-1} \delta_t \rangle \\
&= \sum_{b=2}^{n} \sum_{c=1}^{b-1} \left(\frac{1}{t+1-b} \right) \left(\frac{1}{b-c} \right) \left(\frac{1}{t+1-c} \right) \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-c} \right) \sigma_a^6 .
\end{aligned}
$$

2.

$$\sum_{b,c} \langle a_b \delta_t \rangle \langle \delta_{b-1} \delta_{c-1} \rangle \langle a_c \delta_t \rangle$$

$$= \frac{\sigma_a^6 \pi^2}{6} \sum_b \left(\frac{1}{t+1-b} \right)^2$$

$$+ 2\sigma_a^6 \sum_{b=2}^{n} \sum_{c=1}^{b-1} \left(\frac{1}{t+1-b} \right) \left(\frac{1}{t+1-c} \right) \left(\frac{1}{b-c} \right) \left(1 + \frac{1}{2} + \cdots + \frac{1}{b-c} \right).$$

3.

$$\sum_{b,c} \langle a_b a_c \rangle \langle \delta_{b-1} \delta_t \rangle \langle \delta_{c-1} \delta_t \rangle$$

$$= \sum_b \left\{ \frac{1}{t+1-b} \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-b} \right) \right\}^2 \sigma_a^6.$$

4.

$$\sum_{b,c} \langle a_b a_c \rangle \langle \delta_{b-1} \delta_{c-1} \rangle \langle \delta_t \delta_t \rangle = n\sigma_a^6 \left(\frac{\pi^2}{6} \right)^2.$$

5.

$$\sum_{b,c} \langle a_b a_c \rangle \langle \delta_{b-1} \delta_t \rangle \langle \delta_t \delta_{c-1} \rangle$$

$$= \sum_b \left\{ \frac{1}{t+1-b} \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-b} \right) \right\}^2 \sigma_a^6.$$

6.

$$\sum_{b,c} \langle a_b \delta_{c-1} \rangle \langle \delta_{b-1} \delta_t \rangle \langle a_c \delta_t \rangle$$

$$= \sum_{c=2}^{n} \sum_{b=1}^{c-1} \left(\frac{1}{c-b} \right) \left(\frac{1}{t+1-c} \right) \left(\frac{1}{t+1-b} \right) \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-b} \right) \sigma_a^6.$$

7.

$$\sum_{b,c} \langle a_b \delta_{c-1} \rangle \langle \delta_{b-1} \delta_t \rangle \langle \delta_t a_c \rangle$$

$$= \sum_{c=2}^{n} \sum_{b=1}^{c-1} \left(\frac{1}{c-b} \right) \left(\frac{1}{t+1-c} \right) \left(\frac{1}{t+1-b} \right) \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-b} \right) \sigma_a^6.$$

8.

$$\sum_{b,c} \langle a_b \delta_t \rangle \langle \delta_{b-1} \delta_{c-1} \rangle \langle \delta_t a_c \rangle$$

$$= \frac{\sigma_a^6 \pi^2}{6} \sum_b \left(\frac{1}{t+1-b} \right)^2$$

$$+ 2\sigma_a^6 \sum_{b=2}^{n} \sum_{c=1}^{b-1} \left(\frac{1}{t+1-b} \right) \left(\frac{1}{t+1-c} \right) \left(\frac{1}{b-c} \right) \left(1 + \frac{1}{2} + \cdots + \frac{1}{b-c} \right).$$

9.

$$\sum_{b,c} \langle a_b \delta_t \rangle \langle \delta_{b-1} a_c \rangle \langle \delta_t \delta_{c-1} \rangle$$

$$= \sum_{b=2}^{n} \sum_{c=1}^{b-1} \left(\frac{1}{t+1-b} \right) \left(\frac{1}{b-c} \right) \left(\frac{1}{t+1-c} \right) \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-c} \right) \sigma_a^6.$$

Hence combining 1 to 9 and collecting terms, one obtains the following result:

$$\sum_{b,c} cum(a_b \delta_{b-1} \delta_t,\ \ a_c \delta_{c-1} \delta_t)$$

$$= n\sigma_a^6 \left(\frac{\pi^2}{6} \right)^2 + 2\sigma_a^6 \left[\frac{\pi^2}{6} \sum_b \left(\frac{1}{t+1-b} \right)^2 \right.$$

$$+ \sum_b \left\{ \frac{1}{t+1-b} \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-b} \right) \right\}^2$$

$$+ 2 \sum_{b=2}^{n} \sum_{c=1}^{b-1} \left(\frac{1}{b-c} \right) \left(\frac{1}{t+1-b} \right) \left(\frac{1}{t+1-c} \right) \left(1 + \frac{1}{2} + \cdots + \frac{1}{t+1-c} \right)$$

$$\left. + 2 \sum_{b=2}^{n} \sum_{c=1}^{b-1} \left(\frac{1}{t+1-c} \right) \left(\frac{1}{t+1-b} \right) \left(\frac{1}{b-c} \right) \left(1 + \frac{1}{2} + \cdots + \frac{1}{b-c} \right) \right].$$

REFERENCES

Akaike, H. (1970), "Statistical Predictor Identification". *Annals of the Institute of Statistical Mathematics*, **22**, 203–217.

Baillie, R. T. (1979), "Asymptotic Prediction Mean Squared Error for Vector Autoregressive Models". *Biometrika* **66**, 675–678.

Bhansali, R. J. (1974), "Asymptotic Mean-Square Error of Predicting More Than One-Step Ahead Using the Regression Method". *Applied Statistics*, **23**, 35–42.

Bloomfield, P. (1972), "On the Error of Prediction of a Time Series". *Biometrika* **59**, 501–507.

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Brillinger, D. R. (1975), *Time Series, Data Analysis and Theory*. New York: Holt, Rinehart and Winston.

Fuller, W. A. and Hasza, D. P. (1980), "Predictors for the First-Order Autoregressive Process". *Journal of Econometrics* **13**, 139–157.

Fuller, W. A. and Hasza, D. P. (1981), "Properties of Predictors for Autoregressive Time Series". *Journal of the American Statistical Association* **76**, 155–161.

Hosking, J. R. M. (1981), "Fractional Differencing". *Biometrika* **68**, 165–176.

Jimenez, C. and McLeod, A. I. (1986), Working Paper.

Ledolter, J. and Abraham, B. (1981), "Parsimony and its Importance in Time Series Forecasting". *Technometrics* **23**, 411–414.

Li, W. K. (1981), Topics in Time Series Modelling. Unpublished Ph.D. dissertation, The University of Western Ontario, London, Canada.

Li, W. K. and McLeod, A. I. (1986), "Fractional Time Series Modelling". *Biometrika* **73**, 217–221.

McLeod, A. I. (1978), "On the Distribution of Residual Autocorrelations in Box-Jenkins Models". *Journal of the Royal Statistical Society (B)* **40**, 296–302.

Phillips, P. C. B. (1979), "The Sampling Distribution of Forecasts from a First-Order Autoregression". *Journal of Econometrics* **9**, 241–261.

Schmidt, P. (1974), "The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model". *Econometrica* **42**, 303–309.

Schmidt, P. (1977), "Some Small Sample Evidence on the Distribution of Dynamic Simulation Forecasts". *Econometrica* **45**, 997–1005.

Whittle, P. (1983), *Prediction and Regulation by Linear Least-Square Method.* Minneapolis: University of Minnesota Press.

Yamamoto, T. (1976), "Asymptotic Mean Square Prediction Error for an Autorregressive Model with Estimated Coefficients". *Applied Statistics* **25**, 123–127.

Yamamoto, T. (1981), "Predictions of Multivariate Autoregressive-Moving Average Models". *Biometrika* **68**, 485–492.

CHAPTER 5

EMPIRICAL POWER COMPARISON OF PORTMANTEAU GOODNESS - OF - FIT TESTS

5.1 INTRODUCTION

Consider n observations, $z_1, z_2, \ldots, z_n$, of a stationary autoregressive-moving average model

$$\phi(B)\{z_t - \mu\} = \theta(B)a_t, \tag{5.1}$$

where $\phi(B)$ and $\theta(B)$ are as defined in chapter 2; $\mu = \langle z_t \rangle$ and $\{a_t\}$ is a sequence of independent and identically $N(0, \sigma_a^2)$ random variables.

The modelling strategy of Box and Jenkins (1976) is comprised of three stages: identification, estimation and diagnostic checking. Once a model has been selected and calibrated, the estimated residuals, $\{\hat{a}_t\}$, obtained should have the property of being white noise provided that the model is adequate. The sequence $\{a_t\}$ is unknown in general and hence one is required to examine the estimated residuals. An overall test of model adequacy was proposed by Box and Pierce (1970). These authors obtained the distribution of the residual autocorrelations

$$\hat{r}_k = \frac{\sum_{t=k+1}^{n} \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^{n} \hat{a}_t^2} \qquad k = 1, 2, \ldots, M ,$$

where M is some upper bound, $M < n$, and suggested the well known portmanteau test statistic defined by

$$Q = n \sum_{k=1}^{M} \hat{r}_k^2 . \tag{5.2}$$

It was shown by Box and Pierce (1970) that the portmanteau statistic Q, under the null hypothesis that the selected model is adequate, is asymptotically distributed as $\chi^2(M-p-q)$ provided that M and n are sufficiently large (see also McLeod, 1978). Davies *et al.* (1977) reported evidence that the test statistic Q does not provide satisfactory solution to model discrimination. In particular, Davies *et al.* showed by simulation that the test statistic Q gives lower significance levels than its preset level. A modified version of the Q statistic was then proposed by Davies *et al.*, and it is given by

$$Q^* = n(n+2)\sum_{k=1}^{M} \hat{r}_k^2/(n-k) \quad . \tag{5.3}$$

The asymptotic distribution of Q^* is $\chi^2(M-p-q)$. The statistic Q^* was also independently studied by Ljung and Box (1978) and it was shown by simulation that Q^* does provide a much better test of model adequacy. Although the mean of Q^* is much closer to its asymptotic mean, the variance of Q^* is shown by Ljung and Box to be much larger than its asymptotic variance. A study of the power of Q^* was reported in a paper by Davies and Newbold (1979). In their paper, simulation studies showed that the power of Q^* increases as the variance of forecast error increases. Li and McLeod (1981) proposed another form of the modified portmanteau statistic, namely,

$$Q^{**} = n\sum_{k=1}^{M} \hat{r}_k^2 + \frac{M(M+1)}{2n} \quad . \tag{5.4}$$

It was shown that Q^{**} is also asymptotically distributed as $\chi^2(M-p-q)$. The significance level of this modified portmanteau statistic for the multivariate model was shown in their paper to be better than that of Q. This modified statistic is known to have mean close to the true value and a variance that is smaller than Q^*. Hence it is of interest to compare the performance of Q^{**} and Q^* in detecting model misspecification.

Other tests of model adequacy have been proposed. McLeod (1977, 1978) and Ansley and Newbold (1979) suggested examination of the first few residual autocorrelations by computing the statistic

$$t_k = \hat{r}_k^* / st.\ err.(\hat{r}_k^*) \qquad k = 1, 2, \ldots$$

where $\hat{r}_k^* = \{(n+2)/(n-k)\}^{1/2}\hat{r}_k$ and *st. err.* denotes the estimated standard error. McLeod and Li (1983) discussed another test statistic based on the squared-residuals; this statistic is:

$$Q_{aa} = n(n+2)\sum_{k=1}^{M} \hat{r}_{aa}^2(k)/(n-k) \ ,$$

where $\hat{r}_{aa}(k) = \sum_{t=k+1}^{n}(\hat{a}_t^2 - \hat{\sigma}^2)(\hat{a}_{t-k}^2 - \hat{\sigma}^2)/\sum_{t=1}^{n}(\hat{a}_t^2 - \hat{\sigma}^2)^2$ and $\hat{\sigma}^2 = \sum_{t=1}^{n}\hat{a}_t^2/n$. It was shown by McLeod and Li (1983) that Q_{aa} is asymptotically $\chi^2(M)$. It was indicated that this test statistic can be used in nonlinear time series modelling (see references therein); a test statistic based on the Lagrange multiplier procedure was proposed by Godfrey (1979) but this test procedure was shown by Newbold (1980) to be asymptotically equivalent to examining the first few residuals. Other studies of tests of adequacy are given by Clarke and Godolphin (1982).

In the next section, comparisons of the means of the three statistics Q, Q^* and Q^{**} will be discussed. The exact means are tabulated by assuming that the time series are Gaussian white noise. Comparison is also carried out by simulation for autoregressive models. Section 5.3 presents the empirical significance level of the three test statistics. The simulated time series are of autoregressive type. Section 5.4 examines the empirical power of these three statistics. In section 5.5, an examination of Q_{aa} is given; both its significance levels and power are examined. Section 5.6 presents some empirical results on the test statistics Q^*, Q^{**} and the cumulative periodogram when they are used in testing for whiteness.

5.2 COMPARISONS OF MEANS OF Q, Q^* AND Q^{**}

It is of interest to compare the means of the test statistics Q, Q^* and Q^{**}. Dufour and Roy (1985) obtained exact results for the sample autocorrelations for both normal and nonnormal time series. In particular, equation (2.3) in Dufour and Roy (1985) can be used to compute the exact means of the portmanteau statistics proposed. This equation is

$$\langle r_k^2 \rangle = \frac{1}{n(n-1)(n-2)(n-3)} \Big[\{-n^3 + (k+3)n^2 - k(n+6k)\} \langle s_4 / s_2^2 \rangle \\ + \{n^2(n-k-4) + 3(n-k) + 3k(n+k)\} \Big], \tag{5.5}$$

where $\langle \cdot \rangle$ denotes the mathematical expectation. Assuming that the elements of the time series $z_1, z_2, \ldots, z_n$ are normally and independently distributed, then

$$\langle \frac{s_4}{s_2^2} \rangle = \frac{3(n-1)}{n(n+1)}, \tag{5.6}$$

where $s_k = \sum_{i=1}^{n} z_i^k$, $k \geq 1$. Equation (5.6) is also given by Moran (1948). If one substitutes (5.6) into (5.5), an expression for $\langle r_k^2 \rangle$ can be obtained. Table 5.1 shows the exact means of Q, Q^* and Q^{**} for different values of n and M when the time series is white noise. Under this assumption, the asymptotic means for the statistics Q, Q^* and Q^{**} are $\langle \chi_M^2 \rangle = M$. The exact means of Q and Q^* were also given in Dufour and Roy (1986). It is clear from the table that the finite sample mean of Q is much lower than M. The statistic Q^* is clearly seen to be consistently greater than M while the statistic Q^{**}

is seen to be consistently lower than M. However, these differences are small for both Q^* and Q^{**}.

Table 5.1 provides only a picture of the situation when the time series is white noise. This model is generally not realistic in practical situations. It is more realistic to examine models such as the autoregressive models. Table 5.2 presents the empirical means and variances of the three portmanteau statistics Q, Q^* and Q^{**}, for the autoregressive model of first order. The random number generator Super Duper (Marsaglia, 1976) in conjunction with the Box-Mueller method was used to generate the $NID(0,1)$ variates. The algorithm of McLeod and Hipel (1978) was then used to generate the AR(1) process. For each of the 1000 replications, different combinations of n, ϕ and M were used. For various simulation experiments; the sample size n ranged over the set 50, 100, 200 and 500; the parameter ϕ assumed values 0.1, 0.3, 0.5, 0.7 and 0.9 and M was 10, 20 or 30. The parameter ϕ was estimated by solving the Yule-Walker equations. In this case, the asymptotic mean and variance of the statistics are $M-1$ and $2(M-1)$ respectively. Examining the column of empirical means in this table reveals that the statistic Q has finite sample means generally much lower than its asymptotic means. From the same table, Q^* is seen to overestimate the means more often than Q^{**}. In many instances, Q^{**} is seen to provide estimates that are close to the asymptotic means of Q^{**}. But as n increases, both Q^* and Q^{**} estimate the means fairly accurately. The variances of Q

and Q^{**} are exactly the same, since Q and Q^{**} differ only by a constant. Clearly, the variances of Q^{*} are larger than Q^{**} in all cases. The variance of Q^{*}, is seen from the table to be nearly twice as large as Q^{**} when n is small. However, they improve with increasing sample size. It can be noted that Q^{**} underestimates the asymptotic variance and Q^{*} overestimates it.

TABLE 5.1

Exact means of Q, Q^* and Q^{**} for white noise

n	M	Q	Q^*	Q^{**}
10	1	0.81	1.08	0.91
	3	2.16	3.25	2.76
	5	3.20	5.50	4.70
20	1	0.90	1.05	0.95
	3	2.56	3.13	2.86
	5	4.03	5.22	4.78
	10	6.91	10.50	9.66
30	1	0.93	1.03	0.97
	3	2.70	3.09	2.90
	5	4.34	5.15	4.84
	10	7.88	10.30	9.72
	15	10.65	15.51	14.65
50	1	0.96	1.02	0.98
	3	2.82	3.06	2.94
	5	4.60	5.09	4.90
	10	8.71	10.18	9.81
	15	12.33	15.27	14.73
	25	18.14	25.51	24.64
100	1	0.98	1.01	0.99
	3	2.91	3.03	2.97
	5	4.80	5.05	4.95
	10	9.35	10.09	9.90
	15	13.66	15.14	14.86
	25	21.52	25.23	24.77
	50	36.88	50.51	49.63
200	1	0.99	1.00	1.00
	3	2.96	3.01	2.99
	5	4.90	5.02	4.98
	10	9.68	10.05	9.95
	15	14.33	15.07	14.93
	25	23.25	25.12	24.88
	50	43.39	50.23	49.77

TABLE 5.2

Empirical means and variances of Q, Q^* and Q^{**}

for AR(1) process with $\sigma_a^2 = 1.0$

Number of replications: 1000

			MEAN			VARIANCE		
n	ϕ	M	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
50	0.1	10	7.551	8.928	8.651	13.318	18.543	13.318
		20	14.027	18.709	18.227	27.661	48.491	27.661
		30	18.534	28.288	27.834	35.604	75.690	35.604
	0.3	10	7.511	8.871	8.611	13.335	18.538	13.335
		20	14.042	18.707	18.242	28.924	50.909	28.924
		30	18.626	28.414	27.926	37.271	78.788	37.271
	0.5	10	7.541	8.891	8.641	15.448	21.319	15.448
		20	13.904	18.481	18.104	31.935	54.948	31.935
		30	18.475	28.168	27.775	39.971	81.743	39.971
	0.7	10	7.804	9.204	8.904	16.268	22.751	16.268
		20	14.424	19.167	18.624	33.588	58.333	33.588
		30	19.028	28.920	28.328	42.419	87.636	42.419
	0.9	10	8.021	9.402	9.121	18.009	24.834	18.009
		20	14.543	19.227	18.743	35.594	61.154	35.594
		30	18.994	28.671	28.294	43.241	87.004	43.241

TABLE 5.2 (continued)

			MEAN			VARIANCE		
n	ϕ	M	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
100	0.1	10	8.299	9.007	8.849	16.560	19.518	16.560
		20	16.494	18.899	18.594	36.583	47.692	36.583
		30	23.800	28.904	28.450	55.585	80.750	55.585
	0.3	10	8.247	8.950	8.797	14.858	17.556	14.858
		20	16.585	19.013	18.685	32.653	42.972	32.653
		30	23.684	28.728	28.334	53.726	79.396	53.726
	0.5	10	8.248	8.940	8.798	16.669	19.571	16.669
		20	16.191	18.528	18.291	35.883	47.048	35.883
		30	23.330	28.305	27.980	58.067	85.763	58.067
	0.7	10	8.584	9.301	9.134	20.092	23.573	20.092
		20	16.802	19.225	18.902	41.521	53.827	41.521
		30	24.133	29.255	28.783	63.468	91.706	63.468
	0.9	10	8.714	9.425	9.264	18.689	21.947	18.938
		20	16.978	19.398	19.074	39.249	51.280	39.249
		30	24.352	29.501	29.002	65.250	96.072	65.250

TABLE 5.2 (continued)

n	ϕ	M	MEAN Q	MEAN Q^*	MEAN Q^{**}	VARIANCE Q	VARIANCE Q^*	VARIANCE Q^{**}
200	0.1	10	8.645	9.002	8.920	16.964	18.402	16.964
		20	17.563	18.768	18.613	35.620	40.724	35.620
		30	26.318	28.903	28.643	58.233	70.440	58.233
	0.3	10	8.734	9.090	9.009	18.250	19.755	18.250
		20	17.792	19.011	18.842	36.171	41.100	36.171
		30	26.437	29.017	28.762	61.586	74.282	61.586
	0.5	10	8.677	9.031	8.952	18.408	19.919	18.408
		20	17.861	19.087	18.911	40.644	46.328	40.644
		30	26.279	28.830	28.604	56.554	67.284	56.554
	0.7	10	8.474	8.816	8.749	16.852	18.224	16.852
		20	17.639	18.849	18.689	39.577	45.189	39.577
		30	26.352	28.928	28.677	63.627	76.410	63.627
	0.9	10	8.878	9.225	9.153	18.262	19.714	18.262
		20	17.892	19.095	18.942	39.314	44.606	39.314
		30	26.462	29.017	28.787	63.781	76.518	63.781

TABLE 5.2 (continued)

			MEAN			VARIANCE		
n	ϕ	M	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
500	0.1	10	9.086	9.234	9.196	18.149	18.750	18.149
		20	18.558	19.049	18.978	34.213	36.052	34.213
		30	28.313	29.369	29.243	55.745	60.133	55.745
	0.3	10	8.911	9.053	9.053	17.662	18.221	17.662
		20	18.439	18.981	18.913	36.552	38.476	36.552
		30	28.074	29.115	29.004	54.484	58.509	54.484
	0.5	10	8.947	9.089	9.057	17.262	17.807	17.262
		20	18.685	19.179	19.105	36.556	38.507	36.556
		30	28.256	29.306	29.186	55.673	59.855	55.673
	0.7	10	8.898	9.041	9.008	18.687	19.304	18.687
		20	18.281	18.762	18.701	40.338	42.466	40.338
		30	27.564	28.583	28.494	62.746	67.382	62.746
	0.9	10	8.986	9.125	9.096	16.987	17.527	16.987
		20	18.589	19.075	19.009	38.418	40.508	38.418
		30	28.140	29.179	29.070	56.549	60.797	56.549

5.3 SIGNIFICANCE LEVEL OF Q, Q^* and Q^{**}

The empirical significance levels are examined in this section. The results are given in Table 5.3. A similar simulation technique as in section 2 was used to simulate an AR(1) process with $\sigma_a^2 = 1.0$. The sample size n was set at 50, 100, 200 and 500, the parameter ϕ was set at 0.1, 0.3, 0.5, 0.7 and 0.9, and M was set at 10, 20 and 30. One thousand replications were performed. Both 5% and 10% levels of significance are provided in the table together with their standard errors enclosed in parentheses. From Table 5.3, as is expected, Q provides a rather low level of significance especially when n is small. It is clear from this table that at the preset levels of significance, Q^{**} does not estimate the preset levels accurately when n is small but it improves for larger sample size. For small n, Q^* overestimates the preset levels at few instances and in several cases the estimates are exceedingly large. Again, the estimates improve as sample size increases.

TABLE 5.3

Empirical significance level of Q, Q^* and Q^{**}

for AR(1) process with $\sigma_a^2 = 1.0$

Number of replications: 1000

			5%			10%		
n	ϕ	M	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
50	0.1	10	0.018	0.050	0.028	0.047	0.097	0.058
			(0.004)	(0.007)	(0.005)	(0.007)	(0.009)	(0.007)
		20	0.013	0.066	0.031	0.021	0.106	0.057
			(0.004)	(0.008)	(0.005)	(0.005)	(0.010)	(0.007)
		30	0.004	0.065	0.028	0.007	0.104	0.050
			(0.002)	(0.008)	(0.005)	(0.003)	(0.010)	(0.007)
	0.3	10	0.020	0.050	0.027	0.072	0.117	0.095
			(0.004)	(0.007)	(0.005)	(0.008)	(0.010)	(0.009)
		20	0.019	0.057	0.037	0.044	0.103	0.082
			(0.004)	(0.007)	(0.006)	(0.006)	(0.010)	(0.009)
		30	0.015	0.080	0.049	0.040	0.129	0.086
			(0.004)	(0.009)	(0.007)	(0.006)	(0.011)	(0.009)
	0.5	10	0.027	0.063	0.040	0.060	0.108	0.075
			(0.005)	(0.008)	(0.006)	(0.008)	(0.010)	(0.008)
		20	0.014	0.074	0.033	0.023	0.117	0.068
			(0.004)	(0.008)	(0.006)	(0.005)	(0.010)	(0.008)
		30	0.003	0.066	0.022	0.007	0.120	0.052
			(0.002)	(0.008)	(0.005)	(0.003)	(0.010)	(0.007)
	0.7	10	0.035	0.072	0.048	0.065	0.123	0.082
			(0.006)	(0.008)	(0.007)	(0.008)	(0.010)	(0.009)
		20	0.017	0.089	0.051	0.040	0.136	0.085
			(0.004)	(0.009)	(0.007)	(0.006)	(0.011)	(0.009)
		30	0.004	0.089	0.037	0.012	0.128	0.072
			(0.002)	(0.009)	(0.006)	(0.003)	(0.011)	(0.008)

TABLE 5.3 (continued)

			5%			10%		
n	ϕ	M	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
50	0.9	10	0.037	0.066	0.047	0.062	0.106	0.077
			(0.006)	(0.008)	(0.007)	(0.008)	(0.010)	(0.008)
		20	0.018	0.074	0.042	0.032	0.125	0.071
			(0.004)	(0.008)	(0.006)	(0.006)	(0.010)	(0.008)
		30	0.006	0.069	0.032	0.012	0.110	0.060
			(0.002)	(0.008)	(0.006)	(0.003)	(0.010)	(0.008)
100	0.1	10	0.034	0.057	0.038	0.070	0.107	0.087
			(0.006)	(0.007)	(0.006)	(0.008)	(0.010)	(0.009)
		20	0.032	0.061	0.047	0.057	0.107	0.074
			(0.006)	(0.008)	(0.007)	(0.007)	(0.010)	(0.008)
		30	0.025	0.069	0.048	0.039	0.118	0.084
			(0.005)	(0.008)	(0.007)	(0.006)	(0.010)	(0.009)
	0.3	10	0.020	0.050	0.027	0.072	0.117	0.095
			(0.004)	(0.007)	(0.005)	(0.008)	(0.010)	(0.009)
		20	0.019	0.057	0.037	0.044	0.103	0.082
			(0.004)	(0.007)	(0.006)	(0.006)	(0.010)	(0.009)
		30	0.015	0.080	0.049	0.040	0.129	0.086
			(0.004)	(0.009)	(0.007)	(0.006)	(0.011)	(0.009)
	0.5	10	0.040	0.051	0.043	0.067	0.098	0.082
			(0.006)	(0.007)	(0.006)	(0.008)	(0.009)	(0.009)
		20	0.028	0.064	0.041	0.054	0.092	0.076
			(0.005)	(0.008)	(0.006)	(0.007)	(0.009)	(0.008)
		30	0.020	0.065	0.044	0.038	0.121	0.075
			(0.004)	(0.008)	(0.006)	(0.006)	(0.010)	(0.008)
	0.7	10	0.052	0.077	0.063	0.097	0.133	0.113
			(0.007)	(0.008)	(0.008)	(0.009)	(0.011)	(0.010)
		20	0.040	0.085	0.065	0.075	0.127	0.099
			(0.006)	(0.009)	(0.008)	(0.008)	(0.011)	(0.009)
		30	0.030	0.086	0.056	0.048	0.140	0.097
			(0.005)	(0.009)	(0.007)	(0.007)	(0.011)	(0.009)

TABLE 5.3 (continued)

			5%			10%		
n	ϕ	M	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
100	0.9	10	0.055	0.080	0.065	0.096	0.128	0.114
			(0.007)	(0.009)	(0.008)	(0.009)	(0.011)	(0.010)
		20	0.036	0.084	0.061	0.072	0.144	0.111
			(0.006)	(0.009)	(0.008)	(0.008)	(0.011)	(0.010)
		30	0.031	0.100	0.064	0.055	0.152	0.112
			(0.005)	(0.009)	(0.008)	(0.007)	(0.011)	(0.010)
200	0.1	10	0.038	0.047	0.045	0.074	0.093	0.084
			(0.006)	(0.007)	(0.007)	(0.008)	(0.009)	(0.009)
		20	0.028	0.052	0.042	0.065	0.098	0.091
			(0.005)	(0.007)	(0.006)	(0.008)	(0.009)	(0.009)
		30	0.026	0.061	0.050	0.059	0.118	0.094
			(0.005)	(0.008)	(0.007)	(0.007)	(0.010)	(0.009)
	0.3	10	0.044	0.057	0.051	0.089	0.106	0.097
			(0.006)	(0.007)	(0.007)	(0.009)	(0.010)	(0.009)
		20	0.040	0.057	0.047	0.066	0.100	0.083
			(0.006)	(0.007)	(0.007)	(0.008)	(0.009)	(0.009)
		30	0.034	0.062	0.048	0.059	0.113	0.093
			(0.006)	(0.008)	(0.007)	(0.007)	(0.010)	(0.009)
	0.5	10	0.040	0.057	0.047	0.093	0.106	0.098
			(0.006)	(0.007)	(0.007)	(0.009)	(0.010)	(0.009)
		20	0.041	0.068	0.050	0.077	0.114	0.096
			(0.006)	(0.008)	(0.007)	(0.008)	(0.010)	(0.009)
		30	0.032	0.060	0.045	0.056	0.094	0.078
			(0.006)	(0.008)	(0.007)	(0.007)	(0.009)	(0.008)
	0.7	10	0.044	0.052	0.047	0.080	0.095	0.085
			(0.006)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
		20	0.045	0.061	0.051	0.081	0.115	0.096
			(0.007)	(0.008)	(0.007)	(0.009)	(0.010)	(0.009)
		30	0.040	0.075	0.064	0.074	0.133	0.111
			(0.006)	(0.008)	(0.008)	(0.008)	(0.011)	(0.010)

TABLE 5.3 (continued)

n	ϕ	M	5% Q	5% Q^*	5% Q^{**}	10% Q	10% Q^*	10% Q^{**}
200	0.9	10	0.047	0.059	0.055	0.104	0.125	0.115
			(0.007)	(0.007)	(0.007)	(0.010)	(0.010)	(0.010)
		20	0.050	0.069	0.064	0.080	0.116	0.102
			(0.007)	(0.008)	(0.008)	(0.009)	(0.010)	(0.010)
		30	0.037	0.073	0.063	0.071	0.127	0.111
			(0.006)	(0.008)	(0.008)	(0.008)	(0.011)	(0.010)
500	0.1	10	0.046	0.050	0.048	0.110	0.116	0.112
			(0.007)	(0.007)	(0.007)	(0.010)	(0.010)	(0.010)
		20	0.032	0.041	0.039	0.081	0.096	0.090
			(0.006)	(0.006)	(0.006)	(0.009)	(0.009)	(0.009)
		30	0.041	0.057	0.054	0.079	0.115	0.096
			(0.006)	(0.007)	(0.007)	(0.009)	(0.010)	(0.009)
	0.3	10	0.050	0.051	0.050	0.094	0.105	0.100
			(0.007)	(0.007)	(0.007)	(0.009)	(0.010)	(0.009)
		20	0.039	0.048	0.043	0.085	0.101	0.094
			(0.006)	(0.007)	(0.006)	(0.009)	(0.010)	(0.009)
		30	0.043	0.063	0.057	0.085	0.107	0.097
			(0.006)	(0.008)	(0.007)	(0.009)	(0.010)	(0.009)
	0.5	10	0.051	0.058	0.054	0.096	0.099	0.097
			(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
		20	0.052	0.059	0.055	0.093	0.104	0.100
			(0.007)	(0.007)	(0.007)	(0.009)	(0.010)	(0.009)
		30	0.047	0.060	0.057	0.096	0.117	0.110
			(0.007)	(0.008)	(0.007)	(0.009)	(0.010)	(0.010)
	0.7	10	0.051	0.054	0.052	0.095	0.101	0.096
			(0.007)	(0.007)	(0.007)	(0.009)	(0.010)	(0.009)
		20	0.043	0.052	0.049	0.084	0.100	0.094
			(0.006)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
		30	0.053	0.061	0.059	0.083	0.100	0.095
			(0.007)	(0.008)	(0.007)	(0.009)	(0.009)	(0.009)

TABLE 5.3 (continued)

			5%			10%		
n	ϕ	M	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
500	0.9	10	0.043	0.049	0.045	0.089	0.097	0.092
			(0.006)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)
		20	0.043	0.053	0.047	0.089	0.103	0.095
			(0.006)	(0.007)	(0.007)	(0.009)	(0.010)	(0.009)
		30	0.053	0.069	0.062	0.091	0.115	0.103
			(0.007)	(0.008)	(0.008)	(0.009)	(0.010)	(0.010)

5.4 EMPIRICAL POWER STUDY

An important question is, how do the statistics perform under the alternative hypothesis? This section provides some empirical evidence regarding the power of the three statistics.

Time series data were simulated from several autoregressive-moving average models, and autoregressive models of either order one or four were fitted. The autoregressive parameters were estimated by solving the Yule-Walker equations. The number of replications is 1000. For each of the 1000 replications, the test statistics Q, Q^* and Q^{**} were calculated. The proportion of times that the three statistics correctly rejected the null hypothesis that the true order of the model is one or four was recorded at the 5% and 10% levels of significance. The value of M was set at 20 and the sample size was set at 50, 100 and 200.

Table 5.4 provides the power of the three statistics. The table is arranged in increasing order according to the magnitude of the mean square error of one step ahead prediction (MSE) by using the incorrect model. If the fitted model is of order p, then the MSE is given as

$$\mathrm{MSE} \doteq \sum_{i=0}^{p} \sum_{j=0}^{p} \hat{\phi}_i \hat{\phi}_j \gamma_{|i-j|},$$

where γ_k is the lag k theoretical autocovariance function. The derivation of the MSE is similar to that given in appendix A2.2 at the end of chapter 2.

From Table 5.4, it can be seen that the powers of Q are lower than those of Q^* and Q^{**} in most models except when n and MSE are large. Table 5.4(a) shows that the power of Q^{**} is considerably lower than that of Q^{*} in most models but that the differences become small as MSE increases. Table 5.4(b) shows that the powers for both Q^{*} and Q^{**} are much closer. In Table 5.4(c), it is shown that as n and the MSE increase, both Q^{*} and Q^{**} correctly detect inadequacy of the fitted model virtually the same number of times. Hence, both Q^{*} and Q^{**} performed equally well when n and the MSE are large.

TABLE 5.4

Empirical power of Q, Q^* and Q^{**}

Number of replications: 1000

(a) $n = 50$, $M = 20$

Model						5%			10%		
ϕ_1	ϕ_2	θ_1	θ_2	AR	MSE	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
0.00	0.00	-0.60	-0.40	4	1.094	0.033	0.103	0.059	0.052	0.151	0.104
						(0.006)	(0.010)	(0.007)	(0.007)	(0.011)	(0.010)
0.80	0.00	0.20	0.40	4	1.095	0.047	0.103	0.074	0.062	0.158	0.113
						(0.007)	(0.010)	(0.008)	(0.008)	(0.012)	(0.010)
0.30	0.00	0.75	0.00	1	1.118	0.068	0.201	0.143	0.119	0.274	0.202
						(0.008)	(0.013)	(0.011)	(0.010)	(0.014)	(0.013)
0.40	0.00	-0.20	-0.40	1	1.127	0.088	0.235	0.172	0.148	0.312	0.248
						(0.009)	(0.013)	(0.012)	(0.011)	(0.015)	(0.014)
0.90	0.00	-0.25	0.00	1	1.158	0.164	0.308	0.242	0.212	0.392	0.318
						(0.012)	(0.015)	(0.014)	(0.013)	(0.015)	(0.015)
0.00	0.00	-0 20	0.40	1	1.201	0.157	0.338	0.263	0.224	0.438	0.362
						(0.012)	(0.015)	(0.014)	(0.013)	(0.016)	(0.015)
0.00	0.00	-0.90	-0.80	1	1.412	0.267	0.502	0.413	0.374	0.603	0.540
						(0.014)	(0.016)	(0.16)	(0.015)	(0.015)	(0.016)
1 60	-0.90	-0 80	0.00	4	2.290	0.651	0.792	0.782	0.738	0.864	0.849
						(0.015)	(0.013)	(0.013)	(0 014)	(0.011)	(0.011)
0.80	-0.40	-0.80	0.00	1	2.340	0.810	0.930	0.918	0.852	0.966	0.960
						(0.012)	(0.008)	(0.009)	(0.010)	(0.006)	(0.006)

TABLE 5.4 (continued)

(b) $n = 100,\ M = 20$

Model						5%			10%		
ϕ_1	ϕ_2	θ_1	θ_2	AR	MSE	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
0.00	0.00	-0.60	-0.40	4	1.044	0.027	0.073	0.053	0.059	0.128	0.103
						(0.005)	(0.008)	(0.007)	(0.007)	(0.011)	(0.010)
0.90	0.00	-0.25	0.00	4	1.067	0.113	0.175	0.145	0.161	0.246	0.212
						(0.010)	(0.012)	(0.011)	(0.012)	(0.014)	(0.013)
0.80	0.00	0.20	0.40	1	1.069	0.145	0.240	0.206	0.238	0.346	0.311
						(0.011)	(0.014)	(0.013)	(0.013)	(0.015)	(0.015)
0.40	0.00	-0.20	-0.40	1	1.109	0.294	0.390	0.355	0.388	0.507	0.468
						(0.014)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
0.30	0.00	0.75	0.00	1	1.111	0.188	0.297	0.247	0.285	0.409	0.373
						(0.012)	(0.014)	(0.014)	(0.014)	(0.016)	(0.015)
0.00	0.00	-0.20	0.40	1	1.195	0.494	0.607	0.584	0.610	0.721	0.700
						(0.016)	(0.015)	(0.016)	(0.015)	(0.014)	(0.014)
0.00	0.00	-0.90	-0.80	1	1.390	0.695	0.806	0.770	0.806	0.888	0.878
						(0.015)	(0.013)	(0.13)	(0.013)	(0.010)	(0.010)
1.60	-0.90	-0.80	0.00	4	1.718	0.755	0.808	0.804	0.817	0.873	0.873
						(0.014)	(0.012)	(0.013)	(0.012)	(0.011)	(0.011)
0.80	-0.40	-0.80	0.00	1	2.326	1.000	1.000	1.000	1.009	1.000	1.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

TABLE 5.4 (continued)

(c) $n = 200,\ M = 20$

Model						5%			10%		
ϕ_1	ϕ_2	θ_1	θ_2	AR	MSE	Q	Q^*	Q^{**}	Q	Q^*	Q^{**}
0.00	0.00	-0.60	-0.40	4	1.024	0.053	0.078	0.067	0.102	0.148	0.137
						(0.007)	(0.008)	(0.008)	(0.010)	(0.011)	(0.011)
0.90	0.00	-0.25	0.00	1	1.063	0.374	0.427	0.416	0.495	0.550	0.537
						(0.015)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
0.80	0.00	0.20	0.40	1	1.064	0.352	0.420	0.399	0.479	0.549	0.534
						(0.015)	(0.016)	(0.015)	(0.016)	(0.016)	(0.016)
0.40	0.00	-0.20	-0.40	1	1.101	0.631	0.690	0.678	0.735	0.776	0.770
						(0.015)	(0.015)	(0.015)	(0.014)	(0.013)	(0.013)
0.30	0.00	0.75	0.00	1	1.106	0.530	0.590	0.576	0.671	0.737	0.725
						(0.016)	(0.016)	(0.016)	(0.015)	(0.014)	(0.014)
0.00	0.00	-0.20	0.40	1	1.191	0.907	0.926	0.924	0.958	0.975	0.972
						(0.009)	(0.008)	(0.008)	(0.006)	(0.005)	(0.005)
0.00	0.00	-0.90	-0.80	1	1.384	0.995	0.996	0.996	0.997	0.999	0.999
						(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
1.60	-0.90	-0.80	0.00	4	1.449	0.909	0.982	0.927	0.945	0.956	0.957
						(0.009)	(0.008)	(0.008)	(0.007)	(0.006)	(0.006)
0.80	-0.40	-0.80	0.00	1	2.320	1.000	1.000	1.000	1.000	1.000	1.000
						(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

5.5 EMPIRICAL SIGNIFICANCE LEVEL AND POWER OF Q_{aa}

An empirical investigation of the statistic Q_{aa} is presented in this section. As mentioned earlier, this statistic was recommended for diagnostic checking in nonlinear time series models. It is empirically shown below that this statistic does not perform well in detecting misspecified linear time series models.

A method of simulation similar to that in section 5.4 was used. One thousand replications of AR(1) process were computed. For various simulation experiments; the parameter of the autoregressive process, ϕ, were: 0.1, 0.3, 0.5, 0.7, 0.9; sample sizes were: 50, 100, 200, 500; sizes of M were: 10, 20, 30. The empirical significance levels at each combination of n, M and ϕ are given in Table 5.5. This table clearly indicates that the preset levels are not closely estimated when $n = 50$ but it improves for n larger than 50.

Table 5.6 summarizes the empirical power of this statistic. Similar models to those in Table 5.4 were used to generate the time series. It is clear that this statistic performs very poorly in all the models selected. Despite the improvements for large sample size, the power is low even for $n = 200$. As mentioned above, this statistic was introduced for nonlinear rather than linear models. Thus, it is not surprising that it did not perform as well as the other portmanteau statistics discussed in the previous sections.

TABLE 5.5

Empirical significance level of Q_{aa}

Number of replications: 1000

n	ϕ	M	5%	10%
50	0.1	10	0.040	0.086
			(0.006)	(0.009)
		20	0.044	0.074
			(0.006)	(0.008)
		30	0.037	0.077
			(0.006)	(0.008)
	0.3	10	0.035	0.075
			(0.006)	(0.008)
		20	0.050	0.088
			(0.007)	(0.009)
		30	0.041	0.080
			(0.006)	(0.009)
	0.5	10	0.041	0.067
			(0.006)	(0.008)
		20	0.052	0.088
			(0.007)	(0.009)
		30	0.046	0.076
			(0.007)	(0.008)
	0.7	10	0.036	0.069
			(0.006)	(0.008)
		20	0.046	0.078
			(0.007)	(0.008)
		30	0.048	0.073
			(0.007)	(0.008)

TABLE 5.5 (continued)

n	ϕ	M	5%	10%
50	0.9	10	0.036	0.072
			(0.006)	(0.008)
		20	0.048	0.076
			(0.007)	(0.008)
		30	0.043	0.072
			(0.006)	(0.008)
100	0.1	10	0.044	0.088
			(0.006)	(0.009)
		20	0.069	0.108
			(0.008)	(0.010)
		30	0.073	0.110
			(0.008)	(0.010)
	0.3	10	0.046	0.076
			(0.007)	(0.008)
		20	0.039	0.071
			(0.006)	(0.008)
		30	0.044	0.078
			(0.006)	(0.008)
	0.5	10	0.057	0.093
			(0.007)	(0.009)
		20	0.049	0.090
			(0.007)	(0.009)
		30	0.054	0.091
			(0.007)	(0.009)
	0.7	10	0.048	0.091
			(0.007)	(0.009)
		20	0.050	0.082
			(0.007)	(0.009)
		30	0.049	0.085
			(0.007)	(0.009)

TABLE 5.5 (continued)

n	ϕ	M	5%	10%
100	0.9	10	0.033	0.073
			(0.006)	(0.008)
		20	0.052	0.096
			(0.007)	(0.009)
		30	0.053	0.085
			(0.007)	(0.009)
200	0.1	10	0.039	0.075
			(0.006)	(0.008)
		20	0.042	0.053
			(0.006)	(0.009)
		30	0.053	0.078
			(0.007)	(0.008)
	0.3	10	0.046	0.088
			(0.007)	(0.009)
		20	0.055	0.096
			(0.007)	(0.009)
		30	0.048	0.082
			(0.007)	(0.009)
	0.5	10	0.047	0.095
			(0.007)	(0.009)
		20	0.048	0.099
			(0.007)	(0.009)
		30	0.065	0.097
			(0.008)	(0.009)
	0.7	10	0.047	0.099
			(0.007)	(0.009)
		20	0.044	0.082
			(0.006)	(0.009)
		30	0.043	0.086
			(0.006)	(0.009)

TABLE 5.5 (continued)

n	ϕ	M	5%	10%
200	0.9	10	0.055 (0.007)	0.088 (0.009)
		20	0.051 (0.007)	0.083 (0.009)
		30	0.057 (0.007)	0.089 (0.009)
500	0.1	10	0.047 (0.007)	0.094 (0.009)
		20	0.055 (0.007)	0.099 (0.009)
		30	0.065 (0.008)	0.108 (0.010)
	0.3	10	0.057 (0.007)	0.102 (0.010)
		20	0.061 (0.008)	0.117 (0.010)
		30	0.062 (0.008)	0.106 (0.010)
	0.5	10	0.043 (0.006)	0.084 (0.009)
		20	0.048 (0.007)	0.091 (0.009)
		30	0.050 (0.007)	0.088 (0.009)
	0.7	10	0.049 (0.007)	0.091 (0.009)
		20	0.060 (0.008)	0.109 (0.010)
		30	0.054 (0.007)	0.097 (0.009)

TABLE 5.5 (continued)

n	ϕ	M	5%	10%
500	0.9	10	0.051 (0.007)	0.092 (0.009)
		20	0.059 (0.007)	0.089 (0.009)
		30	0.059 (0.007)	0.095 (0.009)

TABLE 5.6

Empirical power of Q_{aa}

Number of replications: 1000

(a) $n = 50$, $M = 20$

ϕ_1	ϕ_2	θ_1	θ_2	AR	MSE	5%	10%
0.00	0.00	−0.60	−0.40	4	1.094	0.037	0.063
						(0.006)	(0.008)
0.80	0.00	0.20	0.40	4	1.095	0.040	0.078
						(0.006)	(0.008)
0.30	0.00	0.75	0.00	1	1.118	0.060	0.092
						(0.008)	(0.009)
0.40	0.00	−0.20	−0.40	1	1.127	0.044	0.075
						(0.006)	(0.008)
0.90	0.00	−0.25	0.00	1	1.158	0.053	0.083
						(0.007)	(0.009)
0.00	0.00	−0.20	0.40	1	1.201	0.062	0.090
						(0.008)	(0.009)
0.00	0.00	−0.90	−0.80	1	1.412	0.065	0.110
						(0.008)	(0.010)
1.60	−0.90	−0.80	0.00	4	2.290	0.200	0.277
						(0.013)	(0.014)
0.80	−0.40	−0.80	0.00	1	2.340	0.089	0.138
						(0.009)	(0.011)

TABLE 5.6 (continued)

(b) $n = 100,\ M = 20$

ϕ_1	ϕ_2	θ_1	θ_2	AR	MSE	5%	10%
0.00	0.00	−0.60	−0.40	4	1.044	0.064	0.100
						(0.008)	(0.009)
0.90	0.00	−0.25	0.00	4	1.067	0.043	0.088
						(0.006)	(0.009)
0.80	0.00	0.20	0.40	1	1.069	0.055	0.091
						(0.007)	(0.009)
0.40	0.00	−0.20	−0.40	1	1.109	0.059	0.095
						(0.007)	(0.009)
0.30	0.00	0.75	0.00	1	1.111	0.069	0.117
						(0.008)	(0.010)
0.00	0.00	−0.20	0.40	1	1.195	0.095	0.138
						(0.009)	(0.011)
0.00	0.00	−0.90	−0.80	1	1.390	0.080	0.123
						(0.009)	(0.010)
1.60	−0.90	−0.80	0.00	4	1.718	0.217	0.296
						(0.013)	(0.014)
0.80	−0.40	−0.80	0.00	1	2.326	0.199	0.278
						(0.013)	(0.014)

TABLE 5.6 (continued)

(c) $n = 200$, $M = 20$

ϕ_1	ϕ_2	θ_1	θ_2	AR	MSE	5%	10%
0.00	0.00	−0.60	−0.40	4	1.024	0.045	0.082
						(0.007)	(0.009)
0.90	0.00	−0.25	0.00	1	1.063	0.060	0.104
						(0.008)	(0.010)
0.80	0.00	0.20	0.40	1	1.064	0.065	0.104
						(0.008)	(0.010)
0.40	0.00	−0.20	−0.40	1	1.101	0.074	0.116
						(0.008)	(0.010)
0.30	0.00	0.75	0.00	1	1.106	0.070	0.113
						(0.008)	(0.010)
0.00	0.00	−0.20	0.40	1	1.191	0.159	0.221
						(0.012)	(0.013)
0.00	0.00	−0.90	−0.80	1	1.384	0.172	0.261
						(0.012)	(0.014)
1.60	−0.90	−0.80	0.00	4	1.449	0.257	0.335
						(0.014)	(0.015)
0.80	−0.40	−0.80	0.00	1	2.320	0.472	0.558
						(0.016)	(0.016)

TABLE 5.8

Empirical power for Q^*, Q^{**} and CUP in testing

for whiteness: AR(1) process, $\sigma_a^2 = 1.0$

Number of replications: 1000

		5%			10%		
n	ϕ	Q^*	Q^{**}	CUP	Q^*	Q^{**}	CUP
50	−0.3	0.286	0.220	0.323	0.358	0.302	0.442
		(0.014)	(0.013)	(0.015)	(0.015)	(0.015)	(0.016)
	−0.1	0.114	0.071	0.043	0.165	0.113	0.079
		(0.010)	(0.006)	(0.014)	(0.012)	(0.010)	(0.009)
	0.1	0.088	0.054	0.058	0.136	0.090	0.120
		(0.009)	(0.007)	(0.007)	(0.011)	(0.009)	(0.010)
	0.3	0.275	0.219	0.385	0.362	0.296	0.510
		(0.014)	(0.013)	(0.015)	(0.015)	(0.014)	(0.016)
100	−0.3	0.440	0.412	0.695	0.555	0.524	0.805
		(0.016)	(0.016)	(0.015)	(0.016)	(0.016)	(0.013)
	−0.1	0.100	0.080	0.081	0.162	0.128	0.143
		(0.009)	(0.009)	(0.009)	(0.012)	(0.011)	(0.011)
	0.1	0.100	0.081	0.119	0.160	0.129	0.191
		(0.009)	(0.009)	(0.010)	(0.012)	(0.011)	(0.012)
	0.3	0.456	0.428	0.719	0.577	0.545	0.839
		(0.016)	(0.016)	(0.014)	(0.016)	(0.016)	(0.012)

of n and θ. In general, the CUP test is seen to have higher power. However, for $n = 50$ and 100, CUP shows lower power than Q^* at both 5% and 10% levels of significance when θ is near ± 0.1.

Table 5.8 provides the power of the three statistics when the generated process is an AR(1) with parameter $\phi = \pm 0.1, \pm 0.3$ and $n = 50, 100, 200$. A similar picture to that for the MA(1) process can be seen from this table; Q^* has higher power than Q^{**} for all the combinations of n and ϕ and the CUP test, in general, has the highest power.

The underlying process in Table 5.9 is the fractional noise model, FARMA(0,d,0). This model is defined in chapter 4 (see also Li and McLeod, 1986).

To simulate the fractional noise model, the algorithm given in Jimenez *et al.* (1986) is used. Briefly, it involves the following steps:

1. Calculate the partial linear regression coefficients, $\phi_{k,t}$, where

$$\phi_{t,t} = d/(t-d)$$

and

$$\phi_{j,t} = \frac{\phi_{j+1,t}(j+1)(t-j-d)}{(j-1-d)(t-j)}, \qquad j = t-1, \ldots, 1.$$

2. Generate $a_t \sim N(0, \prod_1^t (1 - \phi_{j,j}^2))$.

3. Compute the time series recursively using the relation

$$z_t = a_t + \phi_{1,t} z_{t-1} + \cdots + \phi_{t,t} z_0.$$

In Table 5.9(a), d was set equal to 0.2 and in Table 5.9(b), d was set equal to 0.4. Table 5.9(a) indicates that both portmanteau statistics do not perform as well as the CUP test. Again, the CUP test gives higher power than Q^* and Q^{**}. This is also true for the results in Table 5.9(b). Both Q^* and Q^{**} are comparable for sufficiently large sample size.

TABLE 5.7

Empirical power for Q^*, Q^{**} and CUP in testing

for whiteness: MA(1) process, $\sigma_a^2 = 1.0$

Number of replications: 1000

n	θ	5%			10%		
		Q^*	Q^{**}	CUP	Q^*	Q^{**}	CUP
50	−0.3	0.238	0.187	0.323	0.304	0.249	0.455
		(0.013)	(0.012)	(0.015)	(0.015)	(0.014)	(0.016)
	−0.1	0.081	0.044	0.040	0.128	0.081	0.088
		(0.009)	(0.006)	(0.006)	(0.011)	(0.009)	(0.009)
	0.1	0.107	0.058	0.035	0.146	0.102	0.062
		(0.010)	(0.007)	(0.006)	(0.011)	(0.010)	(0.006)
	0.3	0.238	0.178	0.243	0.319	0.246	0.366
		(0.013)	(0.012)	(0.014)	(0.015)	(0.014)	(0.015)
100	−0.3	0.368	0.331	0.665	0.467	0.436	0.801
		(0.015)	(0.015)	(0.015)	(0.016)	(0.016)	(0.013)
	−0.1	0.111	0.087	0.095	0.166	0.135	0.165
		(0.010)	(0.009)	(0.006)	(0.012)	(0.011)	(0.012)
	0.1	0.107	0.072	0.065	0.156	0.131	0.129
		(0.010)	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)
	0.3	0.360	0.306	0.597	0.453	0.425	0.731
		(0.015)	(0.015)	(0.016)	(0.016)	(0.016)	(0.014)

TABLE 5.7 (continued)

		5%			10%		
n	θ	Q^*	Q^{**}	CUP	Q^*	Q^{**}	CUP
200	−0.3	0.657	0.643	0.960	0.752	0.747	0.981
		(0.015)	(0.015)	(0.006)	(0.014)	(0.014)	(0.004)
	−0.1	0.115	0.099	0.201	0.196	0.179	0.319
		(0.010)	(0.009)	(0.013)	(0.013)	(0.012)	(0.015)
	0.1	0.124	0.109	0.170	0.190	0.176	0.262
		(0.010)	(0.010)	(0.012)	(0.012)	(0.012)	(0.014)
	0.3	0.642	0.627	0.949	0.753	0.743	0.975
		(0.015)	(0.015)	(0.007)	(0.014)	(0.014)	(0.005)

TABLE 5.8

Empirical power for Q^*, Q^{**} and CUP in testing

for whiteness: AR(1) process, $\sigma_a^2 = 1.0$

Number of replications: 1000

		5%			10%		
n	ϕ	Q^*	Q^{**}	CUP	Q^*	Q^{**}	CUP
50	−0.3	0.286	0.220	0.323	0.358	0.302	0.442
		(0.014)	(0.013)	(0.015)	(0.015)	(0.015)	(0.016)
	−0.1	0.114	0.071	0.043	0.165	0.113	0.079
		(0.010)	(0.006)	(0.014)	(0.012)	(0.010)	(0.009)
	0.1	0.088	0.054	0.058	0.136	0.090	0.120
		(0.009)	(0.007)	(0.007)	(0.011)	(0.009)	(0.010)
	0.3	0.275	0.219	0.385	0.362	0.296	0.510
		(0.014)	(0.013)	(0.015)	(0.015)	(0.014)	(0.016)
100	−0.3	0.440	0.412	0.695	0.555	0.524	0.805
		(0.016)	(0.016)	(0.015)	(0.016)	(0.016)	(0.013)
	−0.1	0.100	0.080	0.081	0.162	0.128	0.143
		(0.009)	(0.009)	(0.009)	(0.012)	(0.011)	(0.011)
	0.1	0.100	0.081	0.119	0.160	0.129	0.191
		(0.009)	(0.009)	(0.010)	(0.012)	(0.011)	(0.012)
	0.3	0.456	0.428	0.719	0.577	0.545	0.839
		(0.016)	(0.016)	(0.014)	(0.016)	(0.016)	(0.012)

TABLE 5.8 (continued)

n	ϕ	5% Q^*	5% Q^{**}	5% CUP	10% Q^*	10% Q^{**}	10% CUP
200	−0.3	0.750	0.743	0.966	0.829	0.827	0.985
		(0.014)	(0.014)	(0.006)	(0.012)	(0.012)	(0.004)
	−0.1	0.128	0.108	0.191	0.198	0.183	0.275
		(0.011)	(0.010)	(0.012)	(0.013)	(0.012)	(0.014)
	0.1	0.104	0.093	0.213	0.177	0.159	0.325
		(0.010)	(0.009)	(0.013)	(0.012)	(0.012)	(0.015)
	0.3	0.762	0.755	0.965	0.846	0.843	0.987
		(0.013)	(0.014)	(0.006)	(0.011)	(0.012)	(0.004)

TABLE 5.9

Empirical power for Q^*, Q^{**} and CUP in testing for whiteness: FARMA(0,d,0) process, $\sigma_a^2 = 1.0$

Number of replications: 1000

(a) $d = 0.2$

	5%			10%		
n	Q^*	Q^{**}	CUP	Q^*	Q^{**}	CUP
50	0.179	0.121	0.202	0.249	0.188	0.288
	(0.012)	(0.010)	(0.013)	(0.014)	(0.012)	(0.014)
100	0.336	0.307	0.463	0.412	0.382	0.579
	(0.015)	(0.015)	(0.016)	(0.016)	(0.015)	(0.016)
200	0.635	0.620	0.809	0.707	0.700	0.871
	(0.015)	(0.015)	(0.012)	(0.014)	(0.014)	(0.011)

(b) $d = 0.4$

50	0.587	0.532	0.720	0.642	0.607	0.793
	(0.016)	(0.016)	(0.014)	(0.015)	(0.015)	(0.013)
100	0.897	0.889	0.968	0.926	0.922	0.981
	(0.010)	(0.010)	(0.006)	(0.008)	(0.008)	(0.004)
200	0.996	0.996	1.000	0.997	0.997	1.000
	(0.002)	(0.002)	(0.000)	(0.002)	(0.002)	(0.000)

5.7 CONCLUSION

Several properties of portmanteau test statistics were discussed in this chapter. It was shown that the mean of Q is generally underestimated, the mean of Q^* is overestimated and the mean of Q^{**} is slightly below the asymptotic mean. This result was established for white noise and AR(1) processes, but it likely holds more generally.

The type I error and the power of the three statistics were also examined. It was seen that Q underestimates the preset level of significance but that Q^* and Q^{**} provide a fairly precise estimate of the type I error. In general, the power of Q^* for detecting misspecification is the highest. The power of Q was observed to be very low. The statistic Q^{**} was seen to have low power for small MSE, but it improves as the MSE increases. This was disappointing since it was hoped that Q^{**} might perform better than Q^* due to its smaller variance. However, the power of Q^* and Q^{**} are comparable for large sample size and large MSE.

The type I error and the power of Q_{aa} were also examined. As was expected, this statistic, which was designed mainly for nonlinear time series, did not detect misspecification very effectively.

The statistics Q^*, Q^{**} and the cumulative periodogram test were investigated for their performance in testing for whiteness. In both the ARMA and fractional models, CUP was seen to perform better than both Q^* and Q^{**}. Again, Q^* provided higher power than Q^{**}.

REFERENCES

Ansley, C. F. and Newbold, P. (1979), "On the Finite Sample Distribution of Residual Autocorrelations in Autoregressive-Moving Average Models". *Biometrika* **66**, 547–553.

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control.* San Francisco: Holden-Day.

Box, G. E. P. and Pierce, D. A. (1970), "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Models". *Journal of the American Statistical Association* **65**, 1509–1526.

Clarke, B. R. and Godolphin, E. J. (1982), "Comparative Power Studies for Goodness-of-Fit Tests of Time Series Models". *Journal of Time Series Analysis* **3**, 141–151.

Davies, N. and Newbold, P. (1979), "Some Power Studies of a Portmanteau Test of Time Series Model Specification". *Biometrika* **66**, 153–155.

Davies, N., Triggs, C. M. and Newbold, P. (1977), "Significance Levels of the Box-Pierce Portmanteau Statistic in Finite Samples". *Biometrika* **64**, 517–522.

Dufour, J. M. and Roy, R. (1985), "Some Robust Exact Results on Sample Autocorrelations and Tests of Randomness". *Journal of Econometrics* **29**, 257–273.

Dufour, J. M. and Roy, R. (1986), "Generalized Portmanteau Statistics and Tests of Randomness". To appear in Communications in Statistics-Theory and Methods A15, No. 10.

Godfrey, L. G. (1979), "Testing the Adequacy of a Time Series Model". *Biometrika* **66**, 67-72.

Jimenez, C., Hipel, K. W. and McLeod, A. I. (1986), "Developments in Modelling Long Term Persistence". Presented at 2nd Scientific Assembly of the International Association of Hydrological Sciences, Budapest, Hungary.

Li, W. K. and McLeod, A. I. (1981), "Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models". *Journal of the Royal Statistical Society (B)* **43**, 231-239.

Li, W. K. and McLeod, A. I. (1986), "Fractional Time Series Modelling". *Biometrika* **73**, 217-221.

Ljung, G. M. and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models". *Biometrika* **65**, 297-303.

Marsaglia, G. (1976), "Random Number Generation". In Encyclopedia of Computer Science. (ed. A. Ralston), New York: Petrocelli and Charter, 1192-1197.

McLeod, A. I. (1977), Topics in Time Series and Econometrics. Unpublished Ph.D. thesis, University of Waterloo, Canada.

McLeod, A. I. (1978), "On the Distribution of Residual Autocorrelations in Box-Jenkins Models". *Journal of the Royal Statistical Society (B)* **40**, 296–302.

McLeod, A. I., and Hipel, W. K. (1978), "Simulation Procedures for Box-Jenkins Models". *Water Resources Research* **14**, 969–975.

McLeod, A. I., and Li, W. K. (1983), "Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations". *Journal of Time Series Analysis* **4**, 269–273.

Moran, P. A. P. (1948), "Some Theorems on Time Series II. The Significance of the Serial Correlation Coefficient". *Biometrika* **35**, 255–260.

Newbold, P. (1980), "The Equivalence of Two Tests of Time Series Model Adequacy". *Biometrika* **67**, 463–465.