DTIC FILE COPY

(2)

NUSC Technical Report 8341
28 July 1988

Applications of the Generalized Likelihood to Time Series Analysis

AD-A199 101

Gerald R. Swope
Submarine Sonar Department

Naval Underwater Systems Center
Newport, Rhode Island / New London, Connecticut

88 9 2 034

PREFACE

This study was prepared as a dissertation in partial fulfillment of the requirements for the Doctor of Philosophy in Operations Research and Statistics at Rensselaer Polytechnic Institute. This research was jointly funded under NUSC Project No. A60050, "Towed Array Sonar Technology," Principal Investigator, Dr. Norman L. Owsley, Program Manager, Dr. T. Warfield (ONT-231); NUSC Project No. A75035 (Independent Research and Development); and Project No. 621Y00.

The author acknowledges the contributions of those individuals who supported the completion of this research. Specifically, he acknowledges the assistance of the following groups and individuals:

From Rensselaer Polytechnic Institute: members of the Operations Research and Statistics Department for their constructive comments throughout the course of this study; Professors J. W. Wilkinson, M. Kupferschmid, and H. Kaufman for serving on his advisory committee; Professor A. S. Paulson for the significant assistance, guidance, time, and effort devoted to this entire research project.

From the Naval Underwater Systems Center: the Long-Term Training Committee for their financial support of this research; Drs. F. J. Kingsbury and W. A. Von Winkle for providing financial support of this research; Dr. N. L. Owsley and Mr. J. F. Law for providing an environment to perform this research; and the Word Processing Staff of the Integrated Data Systems, especially Deidre Eccleston, Dorreen Henley, and Sandy Marceau, for typing this thesis.

REVIEWED AND APPROVED: 28 July 1988

F. J. Kingsbury

F. J. KINGSBURY
HEAD: SUBMARINE SONAR DEPARTMENT

W A VonWinkle

W. A. VON WINKLE
ASSOCIATE TECHNICAL DIRECTOR FOR RESEARCH & TECHNOLOGY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AD-A199 101

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NUSC TR 8341	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Underwater Systems Center 6b. OFFICE SYMBOL (If applicable) 2111	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) New London Laboratory New London, CT 06320	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Technology 8b. OFFICE SYMBOL (If applicable) 231	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) 800 North Quincy St. Arlington, VA 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. | PROJECT NO. A60050 | TASK NO. | WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

APPLICATIONS OF THE GENERALIZED LIKELIHOOD TO TIME SERIES ANALYSIS

12. PERSONAL AUTHOR(S)

Gerald R. Swope

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Ph.D. Thesis	FROM ____ TO ____	1988 July 28	290

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Autoregression; Autoregressive-Moving Average; Divergence; Factorial Designs; Generalized Likelihood; Kalman Filter

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Model-critical procedures provide a means to scrutinize an assumed parametric statistical model by varying the way the data are processed for repeated fits to the model. The criticism of the data is accomplished using the generalized likelihood function for the assumed probability density of the data. The degree of criticism is controlled by a user specified constant, c. The model-critical parameter estimates are obtained by maximization of the generalized likelihood function. When c=0, no criticism is performed and maximum likelihood estimates are obtained.

Model-critical estimation procedures are presented for univariate and multivariate autoregressive-moving average processes. The procedures use a Kalman filter in evaluating the generalized likelihood function. A model selection criterion, based on the generalized likelihood, is also presented. A statistical test of fit for multivariate Gaussianity

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Gerald R. Swope	22b. TELEPHONE (Include Area Code) (203) 440-5439 | 22c. OFFICE SYMBOL 2111

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

18. (Cont'd)

Linear Regression
Model-Critical Estimation
Model Selection
Test for Multivariate Normality

19. (Cont'd)

is presented; the test compares the model-critical and maximum likelihood estimates of the covariance matrix.

The test is shown to have excellent power against symmetric alternatives and good power against nonsymmetric alternatives. The test is shown to be insensitive to the underlying structural model and, as such, the test can be applied to the residuals from a structural model such as an autoregression. The model-critical parameter c controls the sensitivity of the test to the assumed parametric model. The test of fit is unique among tests for normality because of its insensitivity to the underlying structural model. For other tests of normality to be applied to the residuals from a structured model, percentage points must be obtained for the particular model under consideration.

CONTENTS

CONTENTS (Cont'd)

LIST OF TABLES

LIST OF TABLES (Cont'd)

LIST OF TABLES (Cont'd)

LIST OF TABLES (Cont'd)

LIST OF TABLES (Cont'd)

LIST OF TABLES (Cont'd)

LIST OF TABLES (Cont'd)

LIST OF TABLES (Cont'd)

LIST OF FIGURES

LIST OF FIGURES (Cont'd)

LIST OF FIGURES (Cont'd)

LIST OF FIGURES (Cont'd)

PART 1

INTRODUCTION

Statistical models are as the name implies only models of reality. As such, they only approximate the true process and in general, more than one model can be used to describe a given set of data. Under these conditions, the experimenter must determine models which adequately describe the observed data. Depending on how the model is to be used, one of the candidate models is chosen. In this thesis, the term model refers to the structural and distributional description of the data. The nominal distributional model is the normal distribution; the structural models are parametric models such as linear regression. This thesis examines model-critical procedures which scrutinize the data and the assumed model by varying the way the data is processed during model fitting. Some aspects of model-critical analysis have been presented by Presser (1980), Paulson, Presser and Lawrence (1981), Paulson and Delaney (1981), Paulson, Presser and Nicklin (1982), Paulson and Delehanty (1983), and Paulson and Swope (1987). All but the last reference use the term self-critical in place of model-critical. Since we are criticizing a model, the term model-critical is adopted here.

Model-critical analysis is based on the generalized likelihood (Paulson and Delehanty, 1983) for a random sample x_1, x_2, ... x_n

$$L_c(\Theta) = \frac{1}{c} \sum_{i=1}^{n} \left[\frac{f^c(x_i;\Theta)}{Q^a(\Theta)} - 1 \right] \tag{1.1}$$

where

$a = c/(1+c)$,

$f(x_i;\Theta)$ is the assumed probability density of the data evaluated at observation x_i,

$$Q(\Theta) = \int_{R_p} f^{1+c}(x_i;\Theta)\, dx \tag{1.2}$$

is the information generating function for $f(x;\Theta)$ (Golumb, 1966, and Paulson and Delehanty, 1983), and c is the model-critical parameter. The model-critical estimate for Θ is the value of $\Theta(c)$ which maximizes (1.1). The estimate $\Theta(c)$ is a robust estimate for Θ with the degree of robustness controlled by c. It will be shown in Part 2 that $L_o(\Theta)$ is the usual log likelihood for Θ and that $\Theta(o)$ is the maximum likelihood estimate of Θ. Differentiating $L_c(\Theta)$ with respect to Θ and setting the result equal to zero yields

$$\frac{\partial L_c(\Theta)}{\partial \Theta} = \sum_{i=1}^{n} \frac{f^c(x_i;\Theta)}{Q^a(\Theta)} \left[\frac{\partial \log f(x_i;\Theta)}{\partial \Theta} - \left(\frac{1}{1+c}\right) \frac{\partial \log Q(\Theta)}{\partial \Theta} \right] = 0 \tag{1.3}$$

which is a necessary condition that must be satisfied by $\Theta(c)$. For the models discussed in Part 2, equation (1.3) is used to obtain $\Theta(c)$.

From (1.3) it can be seen that each term in the sum is weighted by the c^{th} power of the assumed probability density of the data evaluated at x_i. The value of c determines the amount and type of weighting used in (1.3). Positive values of c downweight outlying observations and negative values of c downweight inlying observations. This weighting produces a criticism of the data and the assumed model $f(x;\Theta)$ that can indicate if any model assumptions have been violated.

Outliers are an example of a violation of the model assumptions on the data. It is noted that an observation is an outlier only with respect to the assumed underlying model; if a different model is used, the observation may no longer be an outlier. Unlike unstructured data, where an outlier "sticks out", the structure of a model can hide the outlier. If multiple outliers are present, they can compensate each other (Barnett and Lewis, 1978, Chapter 7). In time series where the observations are not independent, outliers need only be large with respect to the error process to seriously affect the parameter estimates (Kleiner, Martin, and Thomson, 1979). With outliers of this magnitude, they may not show up in plots of the data. Fox (1972) considers two outlier models for time series. The additive outlier is a gross error at a single observation. The innovations outlier is a large value in the error process due to a heavy tailed error distribution. It is noted that both types of outliers can occur with independent as well as dependent observations. With dependent observations, the innovative outlier will affect subsequent observations due to the correlation between observations. A host of robust or resistant procedures are available to reduce the effect of outliers on the parameter estimates (See Andrews et al., 1972, and Martin and Thomson, 1982). The previous discussion has focused on outlying contamination; however, inlying or short-tailed contamination can also be a concern (Hogg, 1974).

From the above discussion, it can be seen that robustness and goodness of fit are related. Model-critical analysis uses this relationship to examine models for a given set of data. The analysis compares maximum likelihood and robust parameter estimates. Clearly, the robust and maximum likelihood estimates must estimate the same quantity if they are to be comparable. From the derivation of $L_c(\theta)$, the estimate $\theta(c)$ is a consistent estimate of θ (Delehanty, 1983). Thus, $\theta(c)$ and $\theta(o)$ are two consistent estimates of θ. If the data and the assumed model are internally consistent, then $\theta(o)$ and $\theta(c)$ should be approximately equal over a range of c values. However, if the data and the assumed model are not consistent, then $\theta(c)$ will change considerably as c increases. Large changes in parameter estimates $\theta(c)$ indicate that the model requires closer examination. As an M-estimator, robust weights are obtained as part of the estimation process. For $c \neq 0$, these critical weights can be used to flag questionable observations. Examination of the weights aids the analyst in evaluating the model. For example, small weights indicate outlying contamination when $c > 0$. Although effective for analyzing models, the above procedure is quite subjective. In order to make critical analysis more precise, Delehanty (1983) and Hwang (1984) have presented goodness of fit statistics to test for multivariate normality that compare the maximum likelihood and model-critical estimate of the covariance matrix. These test statistics like other tests for Gaussianity were developed for data without structure other than a mean vector and covariance matrix. Since most data analysis involves models with additional structure, it would be desirable to have a test of fit which can be applied to structured as well as unstructured models. In this way, the test could be applied to the residuals of a structured model. Gentleman and Wilk (1975) have applied the Shapiro-Wilk test (Shapiro and Wilk, 1965) to two-way layout models; however, before using the test, percentage points of the statistic had to be tabulated for the particular model. This is undesirable especially if a number of different models are to be examined. Ideally one would like a test that can be developed for unstructured data and also be applicable to structured data.

Part 2 discusses model-critical estimation and presents procedures to obtain model-critical estimates for linear regression, autoregression and two way layout models; some complementary material can be found in Delehanty (1983). Also, a model-critical selection procedure is presented; it is a generalization of the selection criteria of Akaike (1974), and Hannan and Quinn (1979) which are special cases of the model-critical procedure. For data contaminated with outliers, it is shown that the criterion selects the model which fits the bulk of the data.

Parts 3 and 4 present model-critical estimation procedures for univariate and multivariate autoregressive-moving average (ARMA) processes, respectively. Harvey and Phillips (1979) and Jones (1980) have presented Kalman filter algorithms to calculate the log likelihood function for Gaussian ARMA models. This algorithm is extended to enable the calculation of the generalized likelihood function. Model-critical parameter estimates are obtained by maximization of the generalized likelihood function. Since the Kalman filter processes each observation individually, it is ideal for downweighting observations inconsistent with the assumed model. Using the Kalman filter algorithm, there are only a few differences in the calculation of $L_o(\Theta)$ and $L_c(\Theta)$; therefore, the same computer program can be used to calculate $L_o(\Theta)$ and $L_c(\Theta)$ with a switch to indicate the calculation of $L_o(\Theta)$ or $L_c(\Theta)$.

In Part 5, a test for multivariate normality is presented; the test is derived from the Kullback divergence (Kullback, 1959). The test is shown to have excellent power against symmetric alternatives and good power against nonsymmetric alternatives. It is shown that the test is insensitive to the underlying structural model; therefore, the test can be applied to the residuals from a structured model. The sensitivity to the parametric model is controlled by the model-critical parameter c. The ability to apply the test to structured as well as unstructured models makes the test unique among tests for normality which can only be applied to unstructured data. Part 6 applies the estimation, model selection and test procedures to experimental data.

The following are some notation conventions. Lower case letters denote vectors and capital letters denote matrices. Since scalars are a special type of vector or matrix, both lower and upper case letters will be used to denote scalars; in general, lower case letters will denote scalars. In general, the vector of parameters Θ will not be included in the arguments of a probability density; for example, $f(x)$ will denote $f(x;\Theta)$. The estimates of a parameter Θ, for example, will be denoted by $\hat{\Theta}$, $\Theta(o)$, or $\Theta(c)$, where $\Theta(o)$ and $\Theta(c)$ denote the maximum likelihood and model-critical estimates, respectively.

PART 2

MODEL-CRITICAL PROCEDURES

2.1 Introduction

Model-critical procedures provide the analyst with a means to analyze a proposed model for a set of observed data. Since there is usually more than one model which can be used to describe a set of data, the procedures allow for the selection of a model using a model-critical analogue of the Akaike selection criterion. If the data are contaminated with outliers for example, the model-critical selection criterion will select the model which describes the bulk of the data. Section 2.2 defines the generalized likelihood function which is used to obtain model-critical parameter estimates. Section 2.3 presents model-critical estimation procedures for a number of widely used models; some complementary material can be found in Delehanty (1983). Section 2.4 derives the model-critical selection criterion which is analyzed in Section 2.5 using simulated autoregressive processes with and without outliers.

2.2 Generalized Likelihood

Let the p x 1 vectors $x_1, x_2, \ldots, x_n$ constitute a random sample from the p-variate Gaussian distribution denoted $N_p(m,D)$ with mean vector m and covariance matrix D, and with probability density

$$f(x) = |2\pi D|^{-1/2}\exp[-(x-m)^T D^{-1}(x-m)/2] . \qquad (2.1)$$

The information generating function of $f(x)$ is defined by (Golumb, 1966; Paulson and Delehanty, 1983)

$$Q(m,D,c) = \int_{R_p} f^c(x)\, f(x)\, dx \qquad (2.2)$$

for c contained in some nondegenerate neighborhood of $c = 0$. The expression of (2.2) can be explicitly and directly evaluated as

$$Q(m,D,c) = [\,|2\pi D|^c (1+c)^p\,]^{-1/2}\ ,\ c > -1\ . \qquad (2.3)$$

All mutual self-information quantities can be obtained directly from (2.3), e.g., the entropy of $f(x)$ is given by

$$-\,Q_c(m,D,0) = (p + \log\,|2\pi D|\,)/2 \qquad (2.4)$$

where $Q_c(m,D,0)$ represents the first partial derivative of Q with respect to c and c set to 0.

The generalized likelihood for m and D given the density (2.1) and the random sample x_1, x_2, ..., x_n is (Paulson and Delehanty, 1983)

$$L_c(m,D) = (1/c) \sum_{i=1}^{n} \left[f^c(x_i)/Q^*(m,D,c) - 1\right] \qquad (2.5)$$

where $c > -1$ and $Q^*(m,D,c) = Q(m,D,c)^{c/(1+c)}$. It is easily shown by expansion or by L'Hospital's rule that

$$\lim_{c\to 0} L_c(m,D) = \sum_{i=1}^{n} \log f(x_i) = L_o(m,D)$$

is the usual log likelihood. The model-critical estimates for m and D are the values m(c) and D(c) which maximize (2.5). The estimates m(c) and D(c) are the solutions to the system of equations

$$\frac{\partial L_c}{\partial \Theta} = \sum_{i=1}^{n} \frac{f_i^c}{Q^a}\left[(1+c)\ \frac{\partial \log f_i}{\partial \Theta} - \frac{\partial \log Q}{\partial \Theta}\right] = 0 \tag{2.6}$$

for $\Theta = m$ and D, and where the arguments of $f_i = f(x_i)$ and Q have been suppressed for notational convenience. Each term in (2.6) is weighted by the c^{th} power of $f(x_i)$, and this affects the estimation of m and D by downweighting terms in (2.6) corresponding to small values of f_i^c. Using (2.6) the following set of implicit estimation equations for m and D are obtained.

$$m = \sum_{i=1}^{n} w_i x_i \tag{2.7a}$$

$$D = (1+c) \sum_{i=1}^{n} w_i (x_i - m)(x_i - m)^T \qquad (2.7b)$$

where

$$w_k = f_k^c(x_k) / \sum_{i=1}^{n} f^c(x_i) \qquad (2.7c)$$

The estimates for m and D are the values m(c) and D(c) which satisfy system (2.7). Clearly, when c = 0, m(0) and D(0) are the usual maximum likelihood estimates. The specification of the user-provided constant c is based on sample size n, dimension p, and the character of the sample. Part 5 discusses appropriate values of c; an additional discussion can be found in Paulson and Delehanty (1983). Equations (2.7) define a family of estimates for m and D indexed on c, m(c) and D(c), with m(0) and D(0) the maximum likelihood estimates for m and D.

When $c \neq 0$, the weighting w_i is determined from the assumed multivariate density and the data $x_1, x_2, \ldots, x_n$. Data that are not consistent with the multivariate Gaussian assumption will receive small weights w_i. This affects the estimates m(c) and D(c) by downweighting data which are not consistent with the model. It is noted that all the data are used in the estimation process, the influence of each observation on the estimates m(c) and D(c) being determined by its weight w_i. If the data and the multivariate Gaussian model are internally consistent, then m(c) and D(c) will be approximately equal to m(0) and C(0). For $c \neq 0$, the procedure estimates the parameters m and D for the most Gaussian-like cluster in the data. For $c > 0$, outlying observations will be downweighted and, for $c < 0$, inlying observations will be downweighted. The choice of c determines how the data is processed.

As an illustration, generalized likelihood parameter estimates are presented for a bivariate sample taken from Anderson (1984, p. 97). Table 2.1 lists the 25 observations from Anderson plus 5 outliers which have been appended. A scatter plot of the data is shown in Figure 2.1 where an x signifies one of the original 25 observations and a y signifies an outlier. The parameter estimates m(c) and D(c) for c = 0 (maximum likelihood), 0.1, 0.2, 0.3, and 0.4 are shown in Table 2.2. As c increases from 0 to 0.4, the estimated mean vector changes little, whereas, the covariance structure changes considerably. The estimated correlation coefficient increases from 0.45 to 0.88 and the estimated standard deviations $\hat{\sigma}_1 = D_{11}^{1/2}$ and $\hat{\sigma}_2 = D_{22}^{1/2}$ decrease from 12.4 and 9.3 to 9.1 and 5.5, respectively. For c = 0.4, the parameter estimates are closer to the estimates from the uncontaminated data than are the maximum likelihood estimates.

For c = 0.1, 0.2, 0.3, and 0.4, Table 2.1 shows the unnormalized weights $\tilde{w}_i = \exp(-c(x_i - m(c))^T D(c)^{-1}(x_i - m(c)))$ of each observation; all the weights are one for c = 0. For observations 25 to 30, the corresponding weights decrease rapidly as c increases and only

TABLE 2.1

Anderson Data and Weights $100xw_i$
for c = 0.1, 0.2, 0.3, and 0.4

	x_1	x_2	0.1	0.2	0.3	0.4
1	179	145	98	95	92	88
2	201	152	91	70	39	24
3	185	149	100	99	99	99
4	188	149	99	96	91	87
5	171	142	92	85	76	66
6	192	152	98	95	90	86
7	190	149	99	93	81	73
8	189	152	99	99	98	96
9	197	159	94	87	74	60
10	187	151	100	100	99	99
11	186	148	99	97	92	89
12	174	147	95	89	74	62
13	185	152	100	99	94	88
14	195	157	96	91	83	74
15	187	158	95	84	41	30
16	161	130	73	50	25	10
17	183	158	93	74	32	13
18	173	148	94	84	61	44
19	182	146	99	97	93	89
20	165	137	84	70	55	39
21	185	152	100	99	94	88
22	178	147	98	96	92	88
23	176	143	96	91	85	78
24	200	158	92	84	73	62
25	187	150	100	99	98	98
26	200	130	52	5	0	0
27	200	135	64	12	0	0
28	165	160	66	16	0	0
29	195	170	77	41	5	0
30	220	170	63	37	17	7

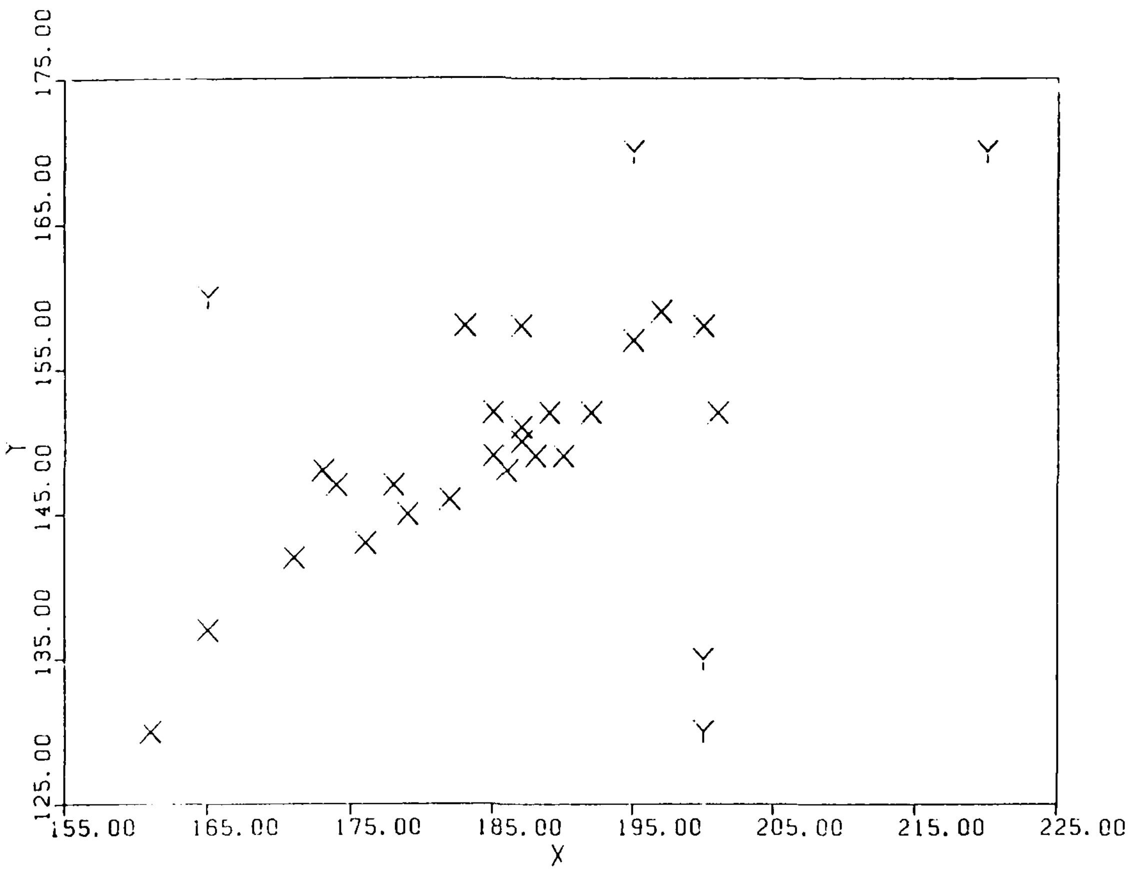

FIGURE 2.1. Scatter Plot of the Anderson Data with Five Outliers

TABLE 2.2

Maximum Likelihood (c = 0) and Model-Critical Parameter Estimates for the Anderson Data

	c				
	0	0.1	0.2	0.3	0.4
$\hat{\mu}_1$	185.9	185.5	185.1	184.9	184.9
$\hat{\mu}_2$	149.9	150.0	150.2	149.8	149.6
$\hat{\sigma}_1$	12.4	11.9	11.1	10.3	9.7
$\hat{\sigma}_2$	9.3	8.8	7.7	6.6	6.0
$\hat{\rho}_{12}$	0.45	0.52	0.75	0.86	0.87

observation 29 and 30 have weights different from zero when $c = 0.3$. The five appended outliers are perceived as not being consistent with the original 25 observations and the assumed single Gaussian population. The estimation procedure clusters the data in the sense that the original 25 observations are retained as a single population and the outliers are more or less ignored.

2.3 Models With Additional Structure

In this section, models with structure in addition to a mean vector and covariance matrix are examined. That is, the observations are of the form

$$y_i = h(x_i;\Theta) + \varepsilon_i \tag{3.1}$$

where

y_i is a p x 1 vector of observations,

x_i is a q x 1 vector of concomitant variables,

Θ is a q x 1 vector of parameters to be estimated from the data, and

ε_i is a p x 1 vector of errors.

The errors ε_i are assumed to be independent and identically distributed Gaussian random variables with zero mean and covariance matrix D. The model $h(x_i;\Theta) = m$ was examined in Section 2.2.

The probability density of the errors is

$$f(\varepsilon_i) = |2\pi D|^{-1/2}\exp(-\varepsilon_i D^{-1}\varepsilon_i/2)$$

$$= |2\pi D|^{-1/2}\exp(-(y_i - h(x_i;\Theta))^T D^{-1}(y_i - h(x_i;\Theta))/2) \quad . \quad (3.2)$$

The generalized likelihood without the constant term denoted L(c) is

$$L(c) = \frac{1}{c}\sum_{i=1}^{n} \left|\frac{1+c}{2\pi D}\right|^{a} \exp(-c(y_i-h(x_i;\Theta))^T D^{-1}(y_i-h(x_i;\Theta))/2) \qquad (3.3)$$

where

$$a = 0.5c/(1+c).$$

For the proposed model $h(x;\Theta)$ in L(c), the model-critical estimates of Θ and D are obtained by maximizing (3.3) over Θ and D. For many models, setting equal to zero the derivatives of L(c) with respect to Θ and D yields a set of implicit equations which can be solved via a fixed point algorithm. Autoregressive-moving average (ARMA) models cannot be solved via a fixed point algorithm. Since ARMA models require a different estimation procedure, they are discussed separately in Parts 3 and 4.

2.3.1 Linear Regression

For multivariate linear regression, $h(x_i;A) = Ax_i$ and (3.1) becomes

$$y_i = Ax_i + \varepsilon_i \quad . \qquad (3.4)$$

For the regression model, $L(c)$ is obtained by substituting Ax_i for $h(x_i;\theta)$ in (3.3). Setting equal to zero the derivatives of $L(c)$ with respect to A and D, the model-critical estimation equations for A and D are

$$A(c) = \left[\sum_{i=1}^{n} w_i y_i x_i^T\right]\left[\sum_{i=1}^{n} w_i x_i x_i^T\right]^{-1} \tag{3.5}$$

and

$$D(c) = (1+c)/w. \sum_{i=1}^{n} w_i (y_i - A(c)x_i)(y_i - A(c)x_i)^T \tag{3.6}$$

where

$$w_i = \exp(-c(y_i - A(c)x_i)^T D(c)^{-1}(y_i - A(c)x_i)/2) \tag{3.7}$$

and

$$w. = \sum_{i=1}^{n} w_i .$$

As an example, consider the abrasion resistance of rubber data (Suich and Derringer, 1977) in Table 2.3. The model considered is

$$y_i = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_1^2 + \alpha_3 x_2 + \alpha_4 x_2^2 + \alpha_5 x_1 x_2 + \epsilon_i . \tag{3.8}$$

TABLE 2.3

Abrasion Resistance of Rubber Data and Critical Weights for Linear Regression Example

y	x_1	x_2	c = 0.2 weight	c = 0.3 weight	c = 0.4 weight
83	1.0	-1.0	0.55	0.26	0.04
113	1.0	1.0	0.85	0.81	0.83
92	-1.0	1.0	0.99	0.98	0.97
82	-1.0	-1.0	0.94	0.89	0.79
100	0.0	0.0	0.97	0.96	0.96
96	0.0	0.0	0.97	0.92	0.77
98	0.0	0.0	1.00	1.00	0.97
95	0.0	1.5	0.95	0.97	0.98
80	0.0	-1.5	1.00	0.99	0.92
100	1.5	0.0	0.62	0.25	0.01
92	-1.5	0.0	0.94	0.95	0.98
87	1.5	-1.5	0.88	0.90	0.94
110	1.5	1.5	0.92	0.82	0.41
79	-1.5	-1.5	0.97	0.93	0.85
84	-1.5	1.5	0.98	0.96	0.88
79	-1.5	1.5	0.90	0.85	0.77
75	-1.5	-1.5	0.96	0.95	0.92
117	1.5	1.5	0.88	0.84	0.92
84	1.5	-1.5	0.99	1.00	0.92
100	0.0	0.0	0.97	0.96	0.96
102	0.0	0.0	0.87	0.82	0.75
96	0.0	0.0	0.97	0.92	0.77

TABLE 2.4

Maximum Likelihood (c = o) and Model-Critical Parameter Estimates for the Rubber Abrasion Resistance Data

c	a_0	a_1	a_2	a_3	a_4	a_5	s
-0.1	97.6	5.80	-0.17	6.09	-3.80	2.86	3.36
0.0	97.7	5.87	-0.11	6.04	-3.88	2.82	3.43
0.1	97.8	5.96	-0.01	5.97	-4.00	2.78	3.46
0.2	98.0	6.09	0.15	5.90	-4.19	2.72	3.43
0.3	98.4	6.47	0.64	5.80	-4.70	2.66	3.14
0.4	98.9	7.07	1.26	5.83	-5.26	2.72	2.55

The independent variables are x_1, silica filler level, and x_2, coupling agent level. Table 2.4 presents the maximum likelihood and model-critical estimates for c = -0.1, 0.0, 0.1, 0.2, 0.3, and 0.4. Over the range of c values, the coefficients a_1, a_2, and a_4 change considerably; a_1, a_2, and a_4 are the coefficients of x_1, x_1^2, and x_2^2, respectively. The critical weights shown in Table 2.3 indicate that the data for observations 1 and 10 should be examined further which we will do in Part 6.

2.3.2 Autoregression

For the multivariate autoregressive (AR) model of order m, (3.1) can be expressed as

$$y_i = \sum_{k=1}^{m} A_k \, y_{i-k} + \varepsilon_i \tag{3.9}$$

for i = m+1, m+2, ..., n. The model-critical estimates for D and A_k, k=1, 2, ..., m are obtained by differentiating L(c) with respect to D and A_k, k=1, 2, ..., m; setting the derivatives equal to zero; and solving for D and A_k, k=1, 2, ..., m. The model-critical estimation equations are

$$A(c) = C^{-1} b \tag{3.10}$$

and

$$D(c) = \left(\frac{1+c}{w.}\right) \sum_{i=m+1}^{n} w_i(y_i - \sum_{k=1}^{m} A_k(c)y_{i-k})(y_i - \sum_{k=1}^{m} A_k(c)y_{i-k})^T \tag{3.11}$$

where

$$A(c) = [A_1(c),\ A_2(c),\ \ldots,\ A_m(c)]^T\ . \tag{3.12}$$

The rs^{th} entry of C is

$$c_{rs} = \sum_{i=m+1}^{n} w_i y_{i-r} y_{i-s}^T \tag{3.13}$$

for r, s = 1, 2, ..., m, the r^{th} entry of b is

$$b_r = \sum_{i=m+1}^{n} w_i y_i y_{i-r}^T\ . \tag{3.14}$$

where

$$w_i = \exp(-c(y_i - \sum_{k=1}^{m} A_i(c)y_{i-k})^T D(c)^{-1}(y_i - \sum_{k=1}^{m} A_i(c)y_{i-k})/2) \tag{3.15}$$

and

$$w. = \sum_{i=m+1}^{n} w_i \ .$$

The estimates A(0), D(0) and A(c), D(c), $c \neq 0$ are conditional maximum likelihood and conditional model-critical estimates since they are conditioned on the first m observations; however, the estimates will still be referred to as maximum likelihood or model-critical estimates. A procedure to estimate the full maximum likelihood and model-critical estimates for ARMA models will be presented in Parts 3 and 4.

As an illustration, model-critical estimates are presented for a simulated Gaussian AR(4) process (ARMA(4,0)) with representation

$$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.652x_{t-4} + \epsilon_t$$

where the ϵ_t are independent identically distributed with zero mean and unit variance. Figure 2.2 is a plot of the realization used to the obtain parameter estimates. Table 2.5 contains the parameter estimates for c = 0, 0.1, 0.2, 0.3, and 0.4. It can be seen that the model-critical and maximum likelihood estimates are approximately equal. Next, additive outliers were added to four observations selected at random in the realization shown in Figure 2.2. The

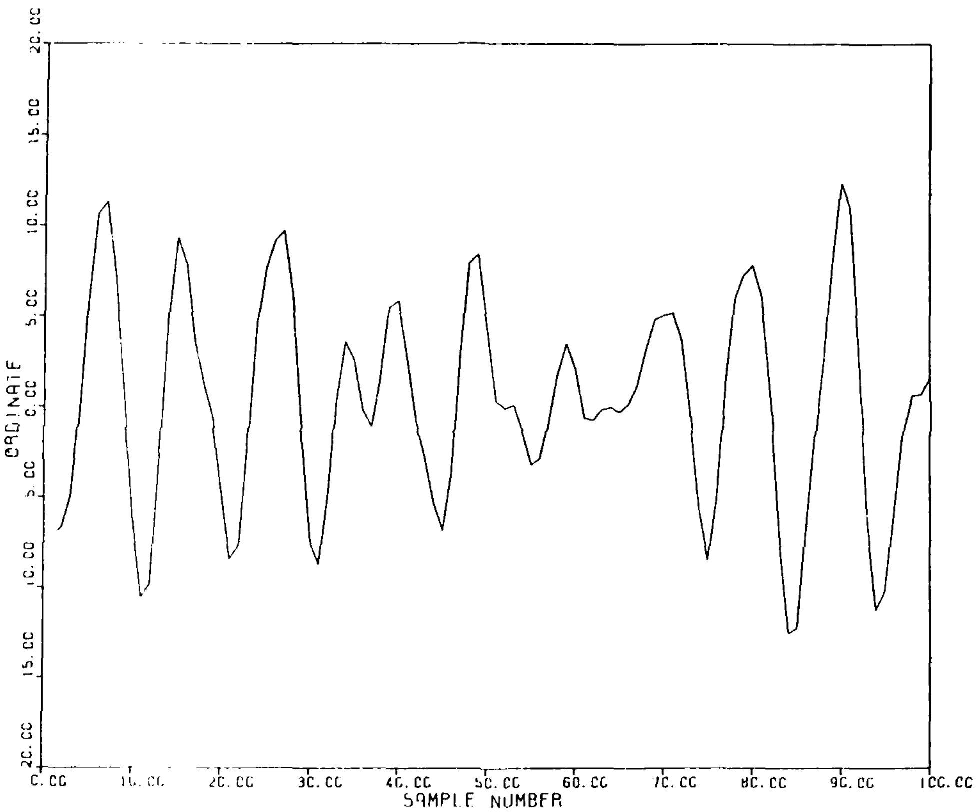

FIGURE 2.2. A Simulated AR(4) Process with Innovations Distributed Normal(0,1)

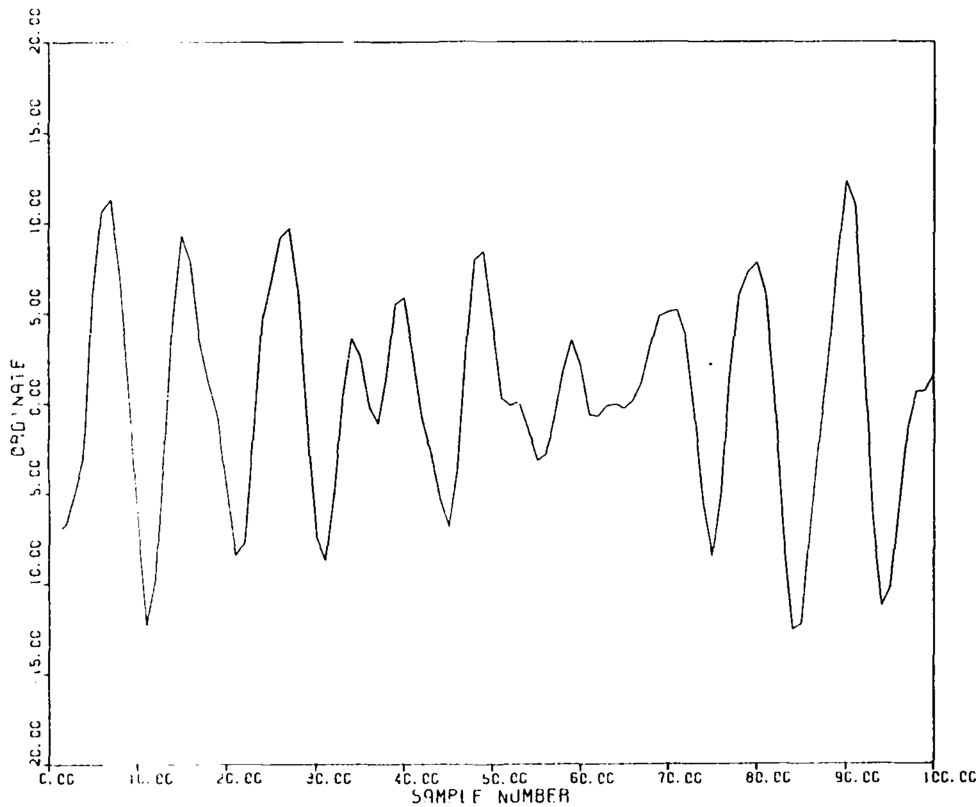

FIGURE 2.3. A Simulated AR(4) Process with Innovations Distributed Normal(0,1), and Four Outliers

outliers are independent, identically distributed Gaussian random variables with zero mean and variance 2; also, the outliers are independent of x_t. The additive outlier model (Martin, 1981) is given by $y_t = x_t + v_t$ where x_t is the AR(4) process and v_t is the outlier. For this example, $v_t \neq 0$ for only four observations. Figure 2.3 is a plot of the data in Figure 2.2 with outliers added to four observations; without a priori knowledge, one would not suspect that outliers are present. Martin (1981) notes that for time series, the outliers need only be large relative to the innovations process ϵ_t to seriously affect the parameter estimates. With outliers of this magnitude, they may not stand out as they do when the observations are independent. This is clearly seen in Table 2.6, which presents the maximum likelihood and model-critical parameter estimates for the data in Figure 2.3. As c increases, the model-critical estimates approach the true values; the improvement in the estimates follows from the downweighting of the outliers in model-critical estimation. For the data with outliers, Figure 2.4 contains plots of the model-critical and maximum likelihood spectrum. The critical spectrum is closer to the true spectrum (the solid line) than the maximum likelihood spectrum estimate. For the data with outliers, the maximum likelihood spectrum contains more high frequency components than the critical spectrum. This is the case because maximum likelihood estimation fits all the data, whereas critical estimation fits the bulk of the data without outliers.

TABLE 2.5

Maximum Likelihood (c = o) and Model-Critical (c ≠ o) Parameter Estimates for a Simulated Univaraite AR(4) Process with True Parameters α_1 = 2.0625, α_2 = -2.4325, α_3 = 1.5845, α_4 = -0.652, and σ = 1; Sample Size = 100

c	a_1	a_2	a_3	a_4	s^2
0.0	2.1346	-2.5685	1.7113	-0.7092	0.9100
0.1	2.1190	-2.5438	1.6906	-0.7094	0.9137
0.2	2.0984	-2.5040	1.6527	-0.6981	0.8915
0.3	2.0841	-2.4820	1.6353	-0.6999	0.8663
0.4	2.0631	-2.4400	1.5962	-0.6914	0.8241

TABLE 2.6

Maximum Likelihood (c = o) and Model-Critical (c ≠ o) Parameter Estimates for a Simulated Univaraite AR(4) Process with Four Additive Outliers; True Parameters α_1 = 2.0625, α_2 = -2.4325, α_3 = 1.5845, α_4 = -0.652, and σ = 1; Sample Size = 100

c	a_1	a_2	a_3	a_4	s^2
0.0	1.7661	-1.7643	0.9633	-0.4409	2.0344
0.1	1.8477	-1.9116	1.0799	-0.4651	1.7752
0.2	1.9556	-2.1520	1.2968	-0.5410	1.4180
0.3	2.0342	-2.3064	1.4354	-0.5912	1.1620
0.4	2.0391	-2.3498	1.4765	-0.6102	1.0740

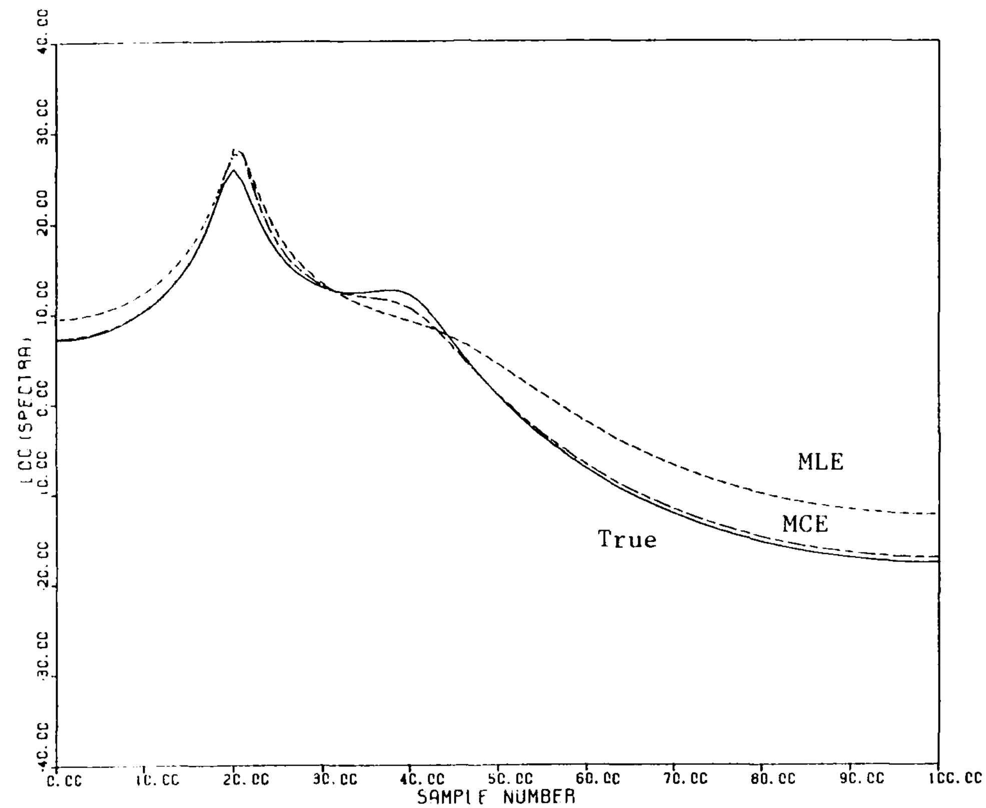

FIGURE 2.4. The Maximum Likelihood and Model-Critical(c = 0.4) Spectrum Estimates for the Simulated AR(4) Process with Four Outliers, and the True Spectrum

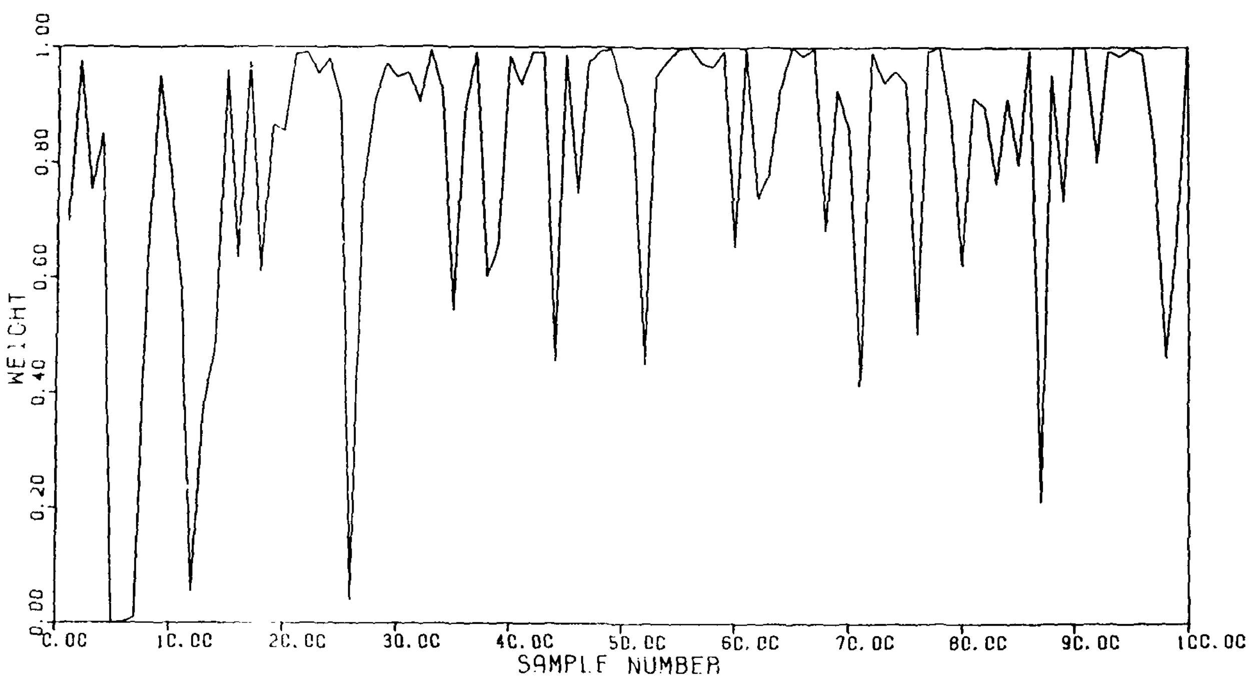

FIGURE 2.5. The Model-Critical Weights for the Simulated AR(4) Process with Innovations Distributed Normal(0,1), and Four Outliers; c = 0.4

For c = 0.4, Figure 2.5 is a plot of the model-critical weights

$$w_t = \exp\left(-c\left(x_t - \sum_{1}^{p} a_k(c)x_{t-k}\right)^2 / 2s^2(c)\right) \tag{3.16}$$

where $s^2(c)$ and $a_k(c)$ are calculated using (3.10) and (3.11). Since the critical weights are a measure of fit between the data and the model, analysis of the weights is an integral part of the modeling process. In Figure 2.5, the small weights at observations 5, 6, 11, and 25 indicate that the model which describes the bulk of the data does not give a good fit to these observations. In fact, some of the subsequent weights are small due to the dependence between observations. Small weights alert the experimenter to the fact that further analysis of the data and model may be necessary.

2.3.3 Factorial Designs

Model-critical estimates for analysis of variance models are presented here. Since the notation of the general analysis of variance model becomes tedious, the multivariate two-way layout with interaction will be used to illustrate model-critical procedures for factorial designs. The model is

$$y_{ijk} = \mu + \alpha_i + B_j + \gamma_{ij} + \varepsilon_{ijk} \tag{3.17}$$

for $i = 1, 2, \ldots, I$; $j = 1, 2, \ldots, J$; and

$k = 1, 2, \ldots, n_{ij}$.

As with the previous models, a system of equations can be obtained by substituting (3.17) into (3.3) and differentiating with respect to μ, α_i, β_j, γ_{ij} for $i = 1, 2, \ldots, I$; $j = 1, 2, \ldots, J$. The system of equations obtained by setting the above derivatives equal to zero is not of full rank. The following constraint equations will yield a full rank system of equations.

$$\sum_i \alpha_i w_{i..} = 0$$

$$\sum_j \beta_j w_{.j.} = 0 \tag{3.18}$$

$$\sum_i \gamma_{ij} w_{ij.} = 0 \quad \text{for all } j \,,$$

and

$$\sum_j \gamma_{ij} w_{ij.} = 0 \quad \text{for all } i \,,$$

where the dot indicates summation over a subscript. With the above constraints, the following recursion relations for the estimation of μ, γ_i, β_j, and γ_{ij} are

$$\mu(c) = \sum_i \sum_j \sum_k y_{ijk} w_{ijk} / w_{...}$$

$$\alpha_i(c) = \sum_j \sum_k (y_{ijk} - \mu(c) - \beta_j(c)) w_{ijk} / w_{i..} , \quad 1 \le i \le I - 1$$

$$\alpha_I(c) = - \sum_{i=1}^{I-1} \alpha_i(c) w_{i..} / w_{I..} , \tag{3.19}$$

$$\beta_j(c) = \sum_i \sum_k (y_{ijk} - \mu(c) - \alpha_i(c)) w_{ijk} / w_{.j.} , \quad 1 \le j \le J - 1$$

$$\beta_J(c) = - \sum_{j=1}^{J-1} \beta_j w_{.j.} / w_{.J.} ,$$

$$\gamma_{ij}(c) = \sum_{k=1}^{n_{ij}} (y_{ijk} - \mu(c) - \alpha_i(c) - \beta_j(c)) w_{ijk} / w_{ij.} , \quad \begin{matrix} 1 \le i \le I - 1 , \\ 1 \le j \le J - 1 , \end{matrix}$$

$$\gamma_{Ij}(c) = - \sum_{i=1}^{I-1} \gamma_{ij}(c) w_{ij.} / w_{Ij.} , \quad 1 \le j \le J - 1$$

$$\gamma_{iJ}(c) = - \sum_{j-1}^{J-1} \gamma_{ij}(c) w_{ij.} / w_{iJ.} , \quad 1 \le i \le I - 1$$

$$\gamma_{IJ}(c) = \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \gamma_{ij}(c) w_{ij.} / w_{IJ.} , \text{ and}$$

$$R(c) = (1 + c)/w_{...} \sum_i \sum_j \sum_k w_{ijk}\, e_{ijk}\, e_{ijk}^T \tag{3.20}$$

where

$$e_{ijk} = y_{ijk} - \mu(c) - \alpha_i(c) - \beta_j(c) - \gamma_{ij}(c)$$

and

$$w_{ijk} = \exp(-\, c\, e_{ijk}^T R(c)^{-1} e_{ijk}/2)\ .$$

Unless indicated specifically, all summations are over the range of the indicated subscripts. A fixed point algorithm is used to obtain the parameter estimates. Where parameter estimates appear explicitly on the right side of a relation, the current estimate is used. The weights are updated after all the effects and error covariance matrix are estimated.

As an example, parameter estimates are obtained for a univariate two-way layout with interaction. Table 2.7 presents survival time data which are taken from Box and Cox (1964). The parameter estimates are obtained using the system of equations (3.19) and (3.20). Table 2.8 presents the maximum likelihood and model-critical parameter estimates. Almost all the parameter estimates change as c increases from 0 to 0.3. For c = 0.3, the model-critical weights are presented in Table 2.9. The small weights in cells (1,2,2) (i.e., poison 1, treatment 2, and

TABLE 2.7

Survival Time Data for Three Poisons and Four Treatments

Poison	Treatment 1	2	3	4
1	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
2	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
3	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33

TABLE 2.8

Maximum Likelihood (c = 0) and Model-Critical (c ≠ 0) Parameter Estimates for the Survival Data

	c = 0	c = 0.1	c = 0.3
μ	0.479	0.461	0.412
α_1	0.138	0.148	0.168
α_2	0.650(-1)	0.535(-1)	0.300(-1)
α_3	-0.203	-0.190	-0.146
β_1	-0.165	-0.151	-0.110
β_2	0.197	0.186	0.133
β_3	-0.869(-1)	-0.744(-1)	-0.565
β_4	0.548(-1)	0.590(-1)	0.770(-1)
γ_{11}	-0.398(-1)	-0.460(-1)	-0.493(-1)
γ_{21}	-0.592(-1)	-0.443(-1)	-0.108(-1)
γ_{31}	0.989(-1)	0.892(-1)	0.539(-1)
γ_{12}	0.652(-1)	0.778(-1)	0.107
γ_{22}	0.733(-1)	0.646	-0.230(-1)
γ_{32}	-0.138	-0.122	-0.642(-1)
γ_{13}	0.369(-1)	0.288(-1)	-0.259(-1)
γ_{23}	-0.825(-1)	-0.656(-1)	-0.108(-1)
γ_{33}	0.456(-1)	0.380(-1)	0.257(-1)
γ_{14}	-0.623(-1)	-0.561(-1)	-0.147(-1)
γ_{24}	0.683(-1)	0.715(-1)	0.660(-1)
γ_{34}	-0.604(-2)	-0.541(-2)	-0.185(-1)
s	0.129	0.117	0.647(-1)

TABLE 2.9

Model-Critical Weights for the Survival Data, c = 0.3

Poison		Treatment 1	2	3	4
		0.65	1.00	0.83	0.28
		0.97	0.07	0.91	0.85
	1	0.95	0.89	0.57	0.99
		1.00	0.70	0.10	0.98
Poison					
		0.95	0.01	0.86	0.98
		0.97	0.89	0.98	0.00
	2	0.81	0.87	0.86	0.58
		0.74	0.00	0.98	0.23
		1.00	0.96	1.00	0.98
		1.00	0.96	0.99	0.96
	3	0.97	0.93	1.00	0.99
		0.99	0.93	0.99	1.00

replication 2), (1,3,4), (1,4,1), (2,2,1), (2,2,4), (2,4,2), and (2,4,4) indicate that the joint structure of the two-way layout with interaction model and independent, identically distributed Gaussian errors is not consistent. As with the other examples, this data will be examined further in Part 6.

2.4 Model-Critical Selection

Until now it has been assumed that the true model for the data is known. In practice, the model is selected from a number of competing models. A number of procedures exist for selecting a model; two widely used selection criteria are the Akaike information criterion (AIC) (Akaike, 1974) and Mallows' C_p statistic (Mallows, 1973). Although not restricted to a particular class of models, the AIC has been used primarily in time series modeling. The C_p statistic has been widely used in linear regression modeling. Other model selection procedures are discussed in Draper and Smith (1966), and Daniel and Wood (1980).

Using the generalized likelihood, we derive a model-critical selection criterion which is the model-critical analogue to the AIC. It is assumed that all the necessary regularity conditions are satisfied. The following derivation is for univariate data; however, there is nothing in the derivation which restricts the criterion to univariate data. Let $L_c(\Theta)$ denote the generalized likelihood evaluated at Θ, where Θ is a p-dimensional parameter vector. Let $\hat{\Theta}$ denote the value of Θ which maximizes $L_c(\Theta)$. Let

$$H_o = - E \left[\frac{\partial^2 L_c(\Theta)}{\partial\Theta_i \ \partial\Theta_j} \right]_{\Theta=\Theta_o} \tag{4.1}$$

and

$$V_o = E \left[\frac{\partial L_c(\Theta)}{\partial\Theta_i} \ \frac{\partial L_c(\Theta)}{\partial\Theta_j}^T \right]_{\Theta=\Theta_o} \tag{4.2}$$

where Θ_o denotes the true parameter vector.

If the correct model is used in $L_c(\Theta)$, then $\sqrt{n}(\hat{\Theta} - \Theta_o)$ is asymptotically normally distributed with zero mean and covariance matrix $H_o^{-1} \ V_o \ H_o^{-1}$ (see Delehanty, 1983).

Approximating $L_c(\Theta_o)$ as a truncated Taylor series about $\hat{\Theta}$ yields

$$L_c(\Theta_o) = L_c(\hat{\Theta}) + \frac{1}{2} (\hat{\Theta} - \Theta_o)^T \left[\frac{\partial^2 L_c(\hat{\Theta})}{\partial\Theta_i \ \partial\Theta_j} \right] (\hat{\Theta} - \Theta_o) \tag{4.3}$$

since $\partial L_c(\hat{\Theta})/\partial\Theta = 0$. Equation (4.3) can be rearranged to obtain

$$2(L_c(\hat{\Theta}) - L_c(\Theta_o)) = (\hat{\Theta} - \Theta_o)^T \ \hat{H} \ (\hat{\Theta} - \Theta_o) \tag{4.4}$$

where $\hat{H}$ is (4.1) evaluated at $\hat{\Theta}$. Delehanty (1983) has shown that $2(L_c(\hat{\Theta}) - L_c(\Theta_c))$ has the same asymptotic distribution as $z^T V_o^{1/2} H_o^{-1} V_0^{1/2} z$, where z is a multivariate normal vector with zero mean and covariance matrix equal to the identity matrix. For Gaussian data,

$$H_o = A(1 + c)^{-3/2} C^{-1} \tag{4.5a}$$

and

$$V_o = A^2 (1 + 2c)^{-3/2} C^{-1} \tag{4.5b}$$

where C depends on the underlying model and

$$A = \left[\frac{(1 + c)}{2\pi\sigma_o^2}\right]^{c/2(1 + c)} \tag{4.6}$$

From the above discussion, it follows that

$$2(L_c(\hat{\Theta}) - L_c(\Theta_o))/(A((1 + c)/(1 + 2c))^{3/2}) \tag{4.7a}$$

or

$$(\hat{\Theta}-\Theta_o)^T \hat{H} (\hat{\Theta}-\Theta_o)/(A((1 + c)/(1 + 2c))^{3/2}) \tag{4.7b}$$

has an asymptotic chi-square distribution with $p = p_0$ degrees of freedom, where p_0 is the dimension of Θ_0. In fact, (4.7b) is asymptotically the sum of p_0 independent chi-square random variables, each with one degree of freedom. Since $(1/n)\hat{H}$ converges to H_0 in probability, $\upsilon = \hat{H}^{1/2}(\hat{\Theta} - \Theta_0)$ is asymptotically distributed the same as $A^{1/2}\ ((1 + c)/(1 + 2c))^{3/4}\ z$ where the entries of z are independent normal random variables with zero mean and unit variance. That is, the entries of υ are asymptotically standard normal random variables. Since $\hat{\Theta}$ converges to Θ_0 in probability at the true model, we submit without proof that the entries υ_i of $\hat{H}^{1/2}(\hat{\Theta} - \Theta_0)/\sqrt{n}$ obey the law of the iterated logarithm (Heyde, 1973). That is, for large n,

$$\upsilon_i = A^{1/2}\left(\frac{1+c}{1+2c}\right)^{3/4} \delta_i(n) \left(\frac{2\log\log n}{n}\right)^{1/2} \qquad (4.8)$$

where $-1 < \delta_i(n) < 1$. Squaring and summing the entries υ_i yields

$$\upsilon^T\upsilon = \sum_{i=1}^{p} \upsilon_i^2 = A\left(\frac{1+c}{1+2c}\right)^{3/2} \delta(n)\left(\frac{2p\log\log n}{n}\right) \qquad (4.9)$$

where

$$\delta(n) = \sum_{i=1}^{p} \delta_i(n)^2/p \ .$$

Using (4.4), we obtain

$$\frac{2}{n}\left(L_c(\hat{\Theta}) - L_c(\Theta_o)\right) = A\left(\frac{1+c}{1+2c}\right)^{3/2} \delta(n) \left(\frac{2p\log\log n}{n}\right). \tag{4.10}$$

where $o < \delta(n) < 1$ and $p = p_o$. Let S_p and S_{p+q} be parameter spaces with dimensions p and p+q, respectively, and let S_p be a subset of S_{p+q}. If $\hat{\Theta}_p$ and $\hat{\Theta}_{p+q}$ are the values of Θ which maximize $L_c(\hat{\Theta})$ over S_p and S_{p+q}, respectively, then $L_c(\hat{\Theta}_p) \leq L_c(\hat{\Theta}_{p+q})$ and

$$-2(L_c(\hat{\Theta}_p) - L_c(\Theta_o)) \geq -2(L_c(\hat{\Theta}_{p+q}) - L_c(\Theta_o)) \; . \tag{4.11}$$

From (4.11), we have that $-2(L_c(\hat{\Theta}) - L_c(\Theta_o))$ is non-increasing as the parameter space increases. It is noted that (4.11) is true when Θ_o is in both S_p and S_{p+q}; the reduction in $-2L_c(\hat{\Theta})$ being due to the effect of estimating additional parameters.

Eliminating $\delta(n)$ from (4.10) and rearranging yields

$$-\frac{2}{n}(L_c(\hat{\Theta}) - L_c(\Theta_o)) + A\left(\frac{1+c}{1+2c}\right)^{3/2} \left(\frac{2p\log\log n}{n}\right) > 0 \tag{4.12}$$

for $p > p_o$. From the above discussion, a model selection criterion is to select the model which minimizes the left side of (4.12). The term $L_c(\Theta_o)$ can be ignored since it contains only the true parameters; however, the term A involves the true, but unknown, variance σ_o^2. Using

$$L_c(\hat{\Theta}) = \frac{1}{c}\sum_{t=1}^{n}\left\{\left(\frac{1+c}{2\pi\hat{\sigma}^2}\right)^{c/2(1+c)} \exp\left[-c\,(x_t - h_t(\hat{\Theta}))^2/2\hat{\sigma}^2\right]-1\right\} , \tag{4.13}$$

$$-\frac{2(L_c(\hat{\Theta}) - L_c(\Theta_o))}{A} =$$

$$-\frac{2}{c}\sum_{t=1}^{n}\left[\hat{\sigma}^2/\sigma_o^2\right.{}^{-c/2(1+c)} \exp(-c(x_t - h_t(\hat{\Theta}))^2/2\hat{\sigma}^2)$$

$$\left. - \exp\left(-c(x_t - h_t(\Theta_o))^2/2\sigma_o^2\right)\right] . \tag{4.14}$$

Approximating

$$\left(\hat{\sigma}^2/\sigma_o^2\right)^{-c/2(1+c)} \approx \left[1 - \left(\frac{c}{2(1+c)}\right)\log(\hat{\sigma}^2/\sigma_o^2) + \frac{1}{2}\left(\left(\frac{c}{2(1+c)}\right)\log(\hat{\sigma}^2/\sigma_o^2)\right)^2\right]$$

for small c and substituting into (4.14) yields

$$-\frac{2(L_c(\hat{\Theta}) - L_c(\Theta_o))}{A} \approx$$

$$\sum_{t=1}^{n}\left[\frac{2}{c}(w_t - \hat{w}_t) + \frac{w_t}{(1+c)}\left(\log \hat{\sigma}^2/\sigma_o^2\right) - c\hat{w}_t\left(\frac{\log(\hat{\sigma}^2/\sigma_o^2)}{2(1+c)}\right)^2\right] . \tag{4.15}$$

where

$$w_t = \exp(-\ c(x_t - h_t(\Theta_o))^2/2\sigma_o^2)$$

and

$$\hat{w}_t = \exp(-\ c(x_t\ -\ h_t(\hat{\Theta}))_2/2\hat{\sigma}^2).$$

For small c and large n, the first and third terms of (4.15) will be small; thus, $-2(L_c(\hat{\Theta})\ -\ L_c(\Theta_o))/A$ can be approximated by

$$(1/(1\ +\ c))\ \log(\hat{\sigma}^2/\sigma_o^2)\sum_{t=1}^{n} w_t\ .$$

At the true model $\hat{w}_t$ converges to $(1+c)^{-1/2}$ for Gaussian data. From the above discussion, it follows that

$$-\ 2(L_c(\hat{\Theta})\ -\ L_c(\Theta_o))/A \approx \frac{n\log(\hat{\sigma}^2/\sigma_o^2)}{(1\ +\ c)^{3/2}}\ +\ t(c,p) \qquad (4.16)$$

where t(c,p) contains terms involving powers of c and the number of parameters p. Using (4.12) and (4.16), a model-critical selection criterion is to select the model which minimizes

$$PSIC(p,c) = \log \sigma^2(c) + 2(s(c)p/n)\log\log n \qquad (4.17)$$

where $s(c) > [(1 + c)^2/(1 + 2c)]^{3/2}$. The restriction on s(c) is required to offset the effect of t(c,p) in (4.16). The form of (4.17) is a generalization of the Hannan-Quinn (1979) procedure and reduces to their procedure for c = 0.

2.5 Monte Carlo Analysis of PSIC

In the above discussion, it was assumed that the data were Gaussian with no contamination such as outliers and that the data were processed using a small value of c which reflects the amount of criticism. A Monte Carlo study was performed to evaluate the PSIC criterion and to examine the sensitivity to the assumptions. The analysis examined autoregressive models of order p = 1, 2, 4, and 8 with innovations from the Gaussian, Gamma, uniform and t distributions; for p = 8 only Gaussian innovations were considered. For the critical parameter c, the values used in the estimation and selection procedure were c = 0.1, 0.2, 0.3, and 0.4. The model selection procedure was also applied to a Gaussian AR(p) process with four additive outliers for p = 1, 2, and 4. A Gaussian AR(8) process with eight additive outliers was also examined. The locations of the outliers in the time series were randomly selected using the uniform distribution. The outliers had a Gaussian distribution with zero mean and variance σ_v^2. The model order was selected for 1000 realizations of an autoregressive process with sample sizes of 50, 100 or 200. For the Hannan-Quinn procedure, denoted H-Q, a value of $s(0) = 1$ was used and for the model-critical procedure PSIC(p,c), $c > 0$, the values of $s(c) = 1$ and $s(c) = (1 + c)^2$ were used. For an AR(4) process, order selection results are presented in Tables 2.10 to 2.20. The experimental results indicate that the selection procedures are not sensitive to the innovations distribution. The H-Q and PSIC selection criteria performed better than the AIC when no outliers are present. However, when additive outliers are present, the PSIC selection procedure improves as c increases as a result of the outliers being downweighted. Since maximum likelihood estimation weights all data equally, a higher order model is required to fit the AR(p) process with outliers. The downweighting of outliers in model-critical estimation results in obtaining the best AR(p) model which fits the bulk of the data. The results for AR(1) and AR(2) processes are similar to the AR(4) results shown in the tables. The effect of s(c) is seen in the results of PSIC 1 and PSIC 2. Using $s(c) = 1$, PSIC 1 is not as effective as H-Q; for $s(c) = (1 + c)^2$, PSIC 2 outperforms H-Q. When the data contain outliers, both PSIC 1 and PSIC 2 become more effective as c increases. Our experiments have shown that $s(c) = (1 + c)^2$ produces the largest s(c) which results in few selections of underfit models. A more conservative s(c) is $(1 + c)$ which will yield results between that of PSIC 1 and PSIC 2.

TABLE 2.10

Frequencies of the Order Selected for the Model
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.652x_{t-4} + e_t$
with e_t Distributed Normal(0,1) and Sample Size 50

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	0	2	467	117	89	75	78	60	112	
H-Q	2	2	637	113	72	56	41	27	50	
PSIC 1	2	2	594	126	78	59	50	32	57	C = 0.1
PSIC 2	2	3	691	119	60	39	35	24	27	
PSIC 1	1	3	523	122	79	67	58	52	95	C = 0.2
PSIC 2	3	6	715	119	48	26	32	19	32	
PSIC 1	0	6	430	109	84	69	65	72	165	C = 0.3
PSIC 2	5	8	707	105	46	29	33	27	40	
PSIC 1	0	5	316	96	80	72	78	108	245	C = 0.4
PSIC 2	7	15	647	77	49	31	39	35	100	

TABLE 2.11

Frequencies of the Order Selected for the Model
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.652x_{t-4} + e_t$
with e_t Distributed Normal(0,1) and Sample Size 100

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	0	0	544	139	77	63	57	47	73	
H-Q	0	0	766	109	57	28	12	14	14	
PSIC 1	0	0	728	124	57	36	19	17	19	C = 0.1
PSIC 2	0	0	803	107	50	21	6	9	4	
PSIC 1	0	0	677	130	71	43	27	23	29	C = 0.2
PSIC 2	0	0	842	92	37	17	5	4	3	
PSIC 1	0	0	592	149	88	47	40	33	51	C = 0.3
PSIC 2	0	0	879	72	26	12	5	6	0	
PSIC 1	0	0	511	153	95	63	58	49	71	C = 0.4
PSIC 2	0	0	883	63	24	13	4	9	4	

TABLE 2.12

Frequencies of the Order Selected for the Model
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.652x_{t-4} + e_t$
with e_t Distributed Normal(0,1) and Sample Size 200

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	0	0	613	130	75	67	32	39	44	
H-Q	0	0	828	91	39	20	9	6	7	
PSIC 1	0	0	810	102	43	26	8	5	6	C = 0.1
PSIC 2	0	0	860	90	32	8	4	2	4	
PSIC 1	0	0	759	124	55	35	11	7	9	C = 0.2
PSIC 2	0	0	890	76	23	6	2	1	2	
PSIC 1	0	0	698	133	71	50	18	16	14	C = 0.3
PSIC 2	0	0	912	61	18	4	2	2	1	
PSIC 1	0	0	646	135	79	58	31	29	22	C = 0.4
PSIC 2	0	0	912	62	17	4	2	2	1	

TABLE 2.13

Frequencies of the Order Selected for the Model
$y_t = x_t + v_t$ Where
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.625x_{t-4} + e_t$
with e_t Distributed Normal(0,1), v_t Distributed Normal(0,4),
4 Values of $v_t \neq 0$, and Sample Size 50

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	39	25	179	285	123	81	83	50	135	
H-Q	70	35	255	315	111	60	61	26	67	
PSIC 1	52	36	261	297	108	66	65	30	85	C = 0.1
PSIC 2	81	43	301	318	91	47	47	17	55	
PSIC 1	42	31	248	252	102	66	74	55	130	C = 0.2
PSIC 2	85	44	339	278	92	40	37	27	58	
PSIC 1	34	21	220	198	97	69	81	94	186	C = 0.3
PSIC 2	87	44	372	219	77	39	41	41	80	
PSIC 1	26	18	172	134	87	60	81	121	301	C = 0.4
PSIC 2	93	50	349	167	71	27	47	55	141	

TABLE 2.14

Frequencies of the Order Selected for the Model
$y_t = x_t + v_t$ Where
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.625x_{t-4} + e_t$
with e_t Distributed Normal(0,1), v_t Distributed Normal(0,4),
4 Values of $v_t \neq 0$, and Sample Size 100

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	16	4	56	493	161	69	63	57	81	
H-Q	54	6	97	572	122	48	37	31	38	
PSIC 1	24	3	137	563	98	50	32	38	55	C = 0.1
PSIC 2	37	8	175	570	79	42	21	29	39	
PSIC 1	13	3	258	404	78	68	49	50	68	C = 0.2
PSIC 2	23	9	334	437	72	44	22	23	36	
PSIC 1	7	2	350	297	86	68	58	58	74	C = 0.3
PSIC 2	16	6	504	330	57	27	23	17	20	
PSIC 1	2	3	367	217	89	64	73	80	105	C = 0.4
PSIC 2	10	7	594	276	48	16	21	12	16	

TABLE 2.15

Frequencies of the Order Selected for the Model
$y_t = x_t + v_t$ Where
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.625x_{t-4} + e_t$
with e_t Distributed Normal(0,1), v_t Distributed Normal(0,4),
4 Values of $v_t \neq 0$, and Sample Size 200

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	0	0	51	559	148	83	52	45	62	
H-Q	1	0	88	723	93	42	23	11	19	
PSIC 1	0	0	313	464	108	46	26	16	27	C = 0.1
PSIC 2	0	0	363	472	84	28	21	10	22	
PSIC 1	0	0	566	229	82	44	23	30	26	C = 0.2
PSIC 2	0	0	681	226	49	14	14	11	5	
PSIC 1	0	0	629	161	83	41	30	30	26	C = 0.3
PSIC 2	0	0	806	143	29	8	7	5	2	
PSIC 1	0	0	600	148	88	52	33	45	34	C = 0.4
PSIC 2	0	0	849	116	23	6	5	1	0	

TABLE 2.16

Frequencies of the Order Selected for the Model
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.625x_{t-4} + e_t$
with e_t Distributed Uniform [-5/2, 5/2], and Sample Size 100

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	0	0	561	151	81	49	58	40	60	
H-Q	0	0	804	104	45	20	12	7	8	
PSIC 1	0	0	758	114	52	30	22	12	12	C = 0.1
PSIC 2	0	0	837	87	43	15	6	7	5	
PSIC 1	0	0	701	126	61	34	29	20	29	C = 0.2
PSIC 2	0	0	860	75	41	13	5	3	3	
PSIC 1	0	0	622	137	72	42	46	32	49	C = 0.3
PSIC 2	0	1	878	64	32	14	5	2	4	
PSIC 1	0	1	532	150	83	58	59	44	73	C = 0.4
PSIC 2	0	2	883	60	29	14	5	2	5	

TABLE 2.17

Frequencies of the Order Selected for the Model
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.625x_{t-4} + e_t$
with e_t Distributed Gamma (3,1), and Sample Size 100

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	0	0	578	125	77	72	54	41	53	
H-Q	0	0	765	96	51	37	24	12	15	
PSIC 1	0	0	742	104	59	45	20	16	14	C = 0.1
PSIC 2	0	0	818	86	50	26	9	9	2	
PSIC 1	0	0	679	127	71	49	30	21	23	C = 0.2
PSIC 2	0	0	842	90	36	19	5	8	0	
PSIC 1	0	0	623	126	73	61	33	41	43	C = 0.3
PSIC 2	0	0	869	73	31	13	6	4	4	
PSIC 1	0	0	527	130	79	72	48	61	83	C = 0.4
PSIC 2	0	1	877	66	24	13	6	5	8	

TABLE 2.18

Frequencies of the Order Selected for the Model
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.625x_{t-4} + e_t$
with e_t Distributed Gamma (1.25,1), and Sample Size 100

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	0	0	583	114	83	66	56	42	56	
H-Q	0	0	750	102	54	38	31	7	18	
PSIC 1	0	0	705	129	67	43	30	12	14	C = 0.1
PSIC 2	0	0	790	107	47	28	14	6	8	
PSIC 1	0	0	644	144	82	43	32	27	28	C = 0.2
PSIC 2	0	0	802	113	45	17	15	2	6	
PSIC 1	0	0	586	144	90	52	41	41	46	C = 0.3
PSIC 2	0	0	840	95	35	9	11	6	4	
PSIC 1	0	0	507	151	88	66	54	56	78	C = 0.4
PSIC 2	0	0	864	76	31	8	10	5	6	

TABLE 2.19

Frequencies of the Order Selected for the Model
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.625x_{t-4} + e_t$
with e_t Distributed t(10), and Sample Size 100

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	0	0	543	122	93	64	54	57	67	
H-Q	0	0	752	107	56	36	19	21	9	
PSIC 1	0	0	722	116	66	36	22	16	22	C = 0.1
PSIC 2	0	0	799	102	55	23	12	6	3	
PSIC 1	0	0	652	135	75	48	33	19	38	C = 0.2
PSIC 2	0	0	821	109	45	12	9	1	3	
PSIC 1	0	0	571	144	85	62	43	34	61	C = 0.3
PSIC 2	0	0	839	95	41	8	10	4	3	
PSIC 1	0	0	474	150	87	67	55	58	109	C = 0.4
PSIC 2	0	1	846	79	40	10	11	7	6	

TABLE 2.20

Frequencies of the Order Selected for the Model
$x_t = 2.0625x_{t-1} - 2.4325x_{t-2} + 1.5845x_{t-3} - 0.625x_{t-4} + e_t$
with e_t Distributed t(3), and Sample Size 100

	≤ 2	3	4	5	6	7	8	9	≥ 10	
AIC	0	0	592	121	74	56	47	47	63	
H-Q	0	0	753	100	47	36	23	20	21	
PSIC 1	0	0	658	131	64	60	33	21	33	C = 0.1
PSIC 2	0	0	748	113	46	42	19	14	18	
PSIC 1	0	0	609	136	83	58	46	31	37	C = 0.2
PSIC 2	0	0	757	125	42	32	19	9	16	
PSIC 1	0	0	517	149	90	78	63	38	65	C = 0.3
PSIC 2	0	0	772	122	47	20	17	6	16	
PSIC 1	0	0	432	148	90	85	78	69	98	C = 0.4
PSIC 2	0	0	787	107	42	22	18	9	15	

The PSIC and H-Q model selection procedures are affected by the location of the roots of the characteristic equation:

$$1 - \sum_{h=1}^{p} \alpha_h \, z^{-h} = 0 \; . \tag{5.1}$$

For small to moderate sample sizes, these procedures tend to underestimate the model order if the roots of (5.1) are "well" within the unit circle; as the roots of (5.1) get closer to the unit circle, the procedures tend to estimate the true order. The effect of the location of the zeroes of (5.1) on the model selection process was examined by selecting the zeroes to be uniformly distributed over some annulus or group of annuli inside the unit circle. The strategy of drawing models in this way allows us to obtain results for a multitude of models within a certain class as opposed to repeatedly drawing realizations from the same model. For example, one could determine N distinct pth-order models by drawing N p-triplets from inside the unit circle $|z| = 1$ in order to compare one procedure such as H-Q against PSIC. We report in detail on one annulus examined closely, namely $0.875 \leq |z| < 0.99$, where z is a root of (5.1). Simulations were performed for AR(8) processes. The selection results for 1000 realizations of Gaussian AR(8) processes are presented in Table 2.21 and 2.22. For N = 200, the AR(8) process was examined both with and without outliers. These results are similar to the case of a fixed model.

TABLE 2.21

Frequencies of the Order Selected For Gaussian AR(8) Models Where the Roots, z, of the Characteristic Equation Satisfy $0.875 < |z| < 0.99$, Sample Size 200, and $c = 0.3$

	≤ 5	6	7	8	9	10	11	12	≥ 13
AIC	0	0	0	444	132	68	73	97	186
H-Q	0	0	0	621	94	57	54	64	110
PSIC 1	0	0	1	524	136	61	69	80	129
PSIC 2	1	1	0	715	74	31	44	68	66

TABLE 2.22

Frequencies of the Order Selected For Gaussian AR(8) Models with 8 Additive Outliers Where the Roots, z, of the Characteristic Equation Satisfy $0.875 < |z| < 0.99$, Sample Size 200, and $c = 0.3$

	≤ 5	6	7	8	9	10	11	12	≥ 13
AIC	4	4	4	75	131	210	191	185	196
H-Q	23	12	22	119	204	246	164	124	86
PSIC 1	13	16	21	153	220	181	135	115	145
PSIC 2	36	33	40	260	269	173	81	59	49

2.6 Discussion

The derivation of (4.17) assumed univariate data. The PSIC criterion can be extended to r-variate models by

$$\mathrm{PSIC}(p,c) = \log|D(c)| + (2ps(c)/n)\log\log n \tag{6.1}$$

where D(c) is the model-critical estimate of the error variance-covariance, p is the number of parameters in the model, and s(c) $> [(1 + c)^2/(1 + 2c)]^{((r/2)+1)}$. Since a process, especially multivariate, can have more than one representation, constraints on the structure of the model are, in general, required in order to effectively use a selection criteria such as (6.1). Model constraints are a major issue in multivariate model identification and are not discussed here (see Denham 1974; Dunsmuir and Hannan 1976; and Hannan 1969, 1981, and 1982).

As an asymptotic result, the law of the iterated logarithm requires a large sample size. For small sample sizes, the term loglogn in (4.17) and (6.1) could be eliminated; the result is an Akaike-like selection procedure which is less sensitive to underestimating the model order. If the model order is underestimated, the parameter estimates for the resulting model will be biased (see Draper and Smith, 1966). This may have serious consequences if the model is to be used for prediction.

If the model order is overestimated too many parameters will be estimated and some efficiency will be lost. For prediction, overestimating the model order is not as serious as underestimating the model order. Also, goodness of fit tests will be affected more seriously by underestimating the model order than by overestimating the model order.

PART 3

MODEL-CRITICAL ESTIMATION FOR UNIVARIATE ARMA(p,q) MODELS

3.1 Introduction

In Part 2, procedures to obtain model-critical parameter estimates for autoregressive (AR) processes were presented. Model-critical estimation procedures for autogregressive-moving average (ARMA) models are presented in this part. As in autoregressive models, outliers or model anomalies can be masked by the structure of the ARMA model. The AR(4) example in Part 2 illustrates that plots of the data may not illuminate model difficulties. The autocorrelations which are used to obtain the parameter estimates tend to accommodate outliers or model anomalies (Barnett and Lewis, 1978). Unlike the independent observation case where an outlier usually does not affect other observations, an outlier in an ARMA process may or may not affect other (usually subsequent) observations depending on the mechanism which produced the outlier.

Harvey and Phillips (1979) and Jones (1980) have presented Kalman filter algorithms to calculate the exact likelihood of a Gaussian ARMA process. The recursive procedure of the Kalman filter makes it ideal for model-critical analysis, since data inconsistent with the model can be downweighted during the estimation of the model parameters. Since the usual log likelihood function is a special case of the generalized likelihood function, virtually the same computer program can be used to calculate the log likelihood and generalized likelihood functions.

Nonlinear optimization programs can be used to maximize the likelihood functions and thereby obtain the maximum likelihood and model-critical parameter estimates.

3.2 Generalized Likelihood for ARMA Models

Let $x_1, x_2, \ldots, x_n$ be a realization of a zero mean ARMA(p,q) process with representation

$$x_t = \sum_{k=1}^{p} \alpha_k x_{t-k} + \varepsilon_t + \sum_{k=1}^{q} \beta_k \varepsilon_{t-k} , \tag{2.1}$$

where ε_t is distributed $N(0, \sigma^2)$ and $E[\varepsilon_t \varepsilon_s] = 0$ for $t \neq s$. For stationarity and invertability, the roots of the characteristic equations

$$1 - \sum_{k=1}^{p} \alpha_k z^k = 0 \tag{2.2a}$$

and

$$1 + \sum_{k=1}^{p} \beta_k z^k = 0 \tag{2.2b}$$

are assumed to lie outside the unit circle, $|z| = 1$. Using the multiplication rule, the log likelihood $L_0(\alpha, \beta, \sigma^2)$ can be expressed as

$$L_0(\alpha, \beta, \sigma^2) = \sum_{n=1}^{n} \log f_t(x_t | \tilde{x}_{t-1}, \alpha, \beta, \sigma^2) \tag{2.3}$$

where

$$\alpha^T = (\alpha_1, \alpha_2, \ldots, \alpha_p) ,$$

$$\beta^T = (\beta_1, \beta_2, \ldots, \beta_q) ,$$

$$\tilde{x}_{t-1} = (x_1, x_2, \ldots, x_{t-1}) ,$$

$$f_1(x_1 | \tilde{x}_0, \alpha, \beta, \sigma^2) = f_1(x_1 | \alpha, \beta, \sigma^2) ,$$

$$f_t(x_t | \tilde{x}_{t-1}, \alpha, \beta, \sigma^2) = (2\pi \sigma_t^2)^{-1/2} \exp\left[-(x_t - x(t|t-1))^2 / 2\sigma_t^2\right], \tag{2.4}$$

$$x(t|t-1) = E[x_t | \tilde{x}_{t-1}, \alpha, \beta, \sigma^2] , \tag{2.5}$$

and

$$\sigma_t^2 = \mathrm{Var}[x_t | \tilde{x}_{t-1}, \alpha, \beta, \sigma^2] . \tag{2.6}$$

Using equations (2.4) to (2.6), the log likelihood of equation (2.3) can be written as

$$L_0(\alpha, \beta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{n} \log \sigma_t^2 - \frac{1}{2} \sum_{t=1}^{n} (x_t - x(t|t-1))^2 / \sigma_t^2 . \tag{2.7}$$

The model critical parameter estimates are a(c), b(c) which are vectors of the autoregressive and moving average parameters, respectively, and $s^2(c)$ which maximize the generalized likelihood function

$$L_c(\alpha, \beta, \sigma^2) = \frac{1}{c} \sum_{t=1}^{n} \left\{ \frac{f_t^c(x_t\ x_{t-1}, \alpha, \beta, \sigma^2)}{Q_t^a(\alpha, \beta, \sigma^2, c)} - 1 \right\} \qquad (2.8)$$

where

$f_t(x_t|\tilde{x}_{t-1}, \alpha, \beta, \sigma^2)$ is defined by (2.4) ,

$$Q_t(\alpha, \beta, \sigma^2, c) = [(1 + c)(2\pi\sigma_t^2)^c]^{-1/2} , \qquad (2.9)$$

$a = c/(1 + c)$, and c is the model-critical parameter. The function $Q_t(\alpha, \beta, \sigma^2, c)$ is the information generating function of the density (2.4) and defined as in (2.2.2). In the next section, it will be shown that, given α and β, expressions for $x(t|t-1)$ and σ_t^2 can be obtained; with $x(t|t-1)$ and σ_t^2 available, $L_0(\alpha, \beta, \alpha^2)$ and $L_c(\alpha, \beta, \sigma^2)$ can be evaluated. Unlike the case of independent observations, $f_t(\tilde{x}_t|x_{t-1}, \alpha, \beta, \sigma^2)$ and $Q_t(\alpha, \beta, \sigma^2, c)$ depend on the observation t.

For the remainder of this part, f_t and Q_t will denote $f_t(x_t|\tilde{x}_{t-1}, \alpha, \beta, \sigma^2)$ and $Q_t(\alpha, \beta, \sigma^2, c)$, respectively. The generalized likelihood $L_c(\alpha, \beta, \sigma^2)$ without the constant term $(-n/c)$ will be denoted by L(c). The model-critical estimate for $\theta^T = (\alpha^T, \beta^T, \sigma^2)$ is the solution $\theta(c)$ to

$$\frac{\partial L(c)}{\partial \Theta} = 0 = \sum_{t=1}^{n} \frac{f_t^c}{Q_t^a} \left[\frac{\partial \log f_t}{\partial \Theta} - \frac{1}{(1 + c)} \frac{\partial \log Q_t}{\partial \Theta} \right]. \qquad (2.10)$$

The following discussion applies to the selected ARMA(p,q) model. As in the independent sample case. each term in (2.10) is weighted by f_t^c/Q_t^a. For f_t and Q_t defined above,

$$f_t^c/Q_t^a = ((1+c)/2\pi\sigma_t^2)^{a/2} \exp\left[- c(x_t - x(t|t-1))^2/2\sigma_t^2 \right]. \qquad (2.11)$$

Data corresponding to large $(x_t - x(t|t-1))^2/\sigma_t^2$ will be downweighted in an adaptive manner as the estimation process iterates to find a solution; the degree of downweighting being determined by the value of c. Model-critical estimation filters out non-Gaussian influences and finds the "best" ARMA(p,q) model consistent with Gaussianity. Data inconsistent with the Gaussian ARMA(p,q) model will receive small weights since $(x_t - x(t|t-1))^2$ will be large. As the value of $c > 0$ is increased, the procedure is more critical of the joint character of the data and the assumed model. If the data and the Gaussian ARMA(p,q) model are internally consistent, then the model-critical estimates a(c), b(c), and $s^2(c)$ and the maximum likelihood estimates a(o), b(o), and $s^2(o)$ will be approximately equal. Our experiments have shown that, if the model is ARMA(p,q) and the innovations are symmetric heavy tailed non-Gaussian, a(c) and b(c) will not differ much from a(o) and b(o) as c increases. However, $s^2(c)$ will differ considerably from $s^2(o)$ as c increases. These facts provide a measure as to the consistency between the data and the model. If p and q are known, the estimates $s^2(c)$ and $s^2(o)$ can be used to examine the innovations structure; this will be discussed further in Part 5.

3.3 <u>Evaluation of the Generalized Likelihood</u>

In the previous section, expressions were obtained for the log likelihood (2.7) and the generalized likelihood (2.8). For ARMA models, it is not practical to obtain the parameter estimates by solving the system of equations defined by (2.10) since the equations are nonlinear and complicated, especially the equations for model-critical estimates. It will be shown that given a, b and s^2 the values of $x(t|t-1)$ and σ_t^2 can be calculated for t = 1,2, ..., n; hence, (2.7) and (2.8) can be evaluated. Being able to evaluate (2.7) and (2.8), they can be maximized by searching over a, b, and s^2 via nonlinear optimization methods. A Kalman filter will be used to obtain $x(t|t-1)$ and σ_t^2 recursively. Since the Kalman filter processes each observation individually, it is ideal for model-critical estimation because it allows for data inconsistent with the model to be downweighted during the estimation process. The maximum likelihood estimates a(o), b(o), and $s^2(o)$ and the model-critical estimates a(c), b(c), and $s^2(c)$ can be examined to provide insight into the adequacy of the model. The following is a brief discussion of the

Kalman filter algorithm which can be found in a number of references (see Kalman, 1960 and Gelb, 1974).

Let the observation y_t be given by the <u>measurement equation</u>

$$y_t = z^T w_t + u_t \tag{3.1}$$

where y_t is the observed value, z is a k x 1 vector of fixed known values, w_t is the k x 1 state vector of the system, and u_t is the measurement error. The measurement error, u_t, is usually assumed to be zero mean Gaussian with variance R. The <u>state equation</u> is given by

$$w_t = Aw_{t-1} + Be_t \tag{3.2}$$

where A is a k x k <u>transition</u> matrix of known values, B is a k x m matrix of known values, and e_t is a m x 1 vector of normally distributed random variables with $E[e_t] = 0$, $E[e_t e_t^T] = \sigma^2 Q$ and $E[e_t e_s^T] = 0$ for all $t \neq s$. The known matrix Q is assumed to be positive definite. Further, $E[u_t] = 0$, $E[u_t^2] = R$, $E[u_t u_s] = 0$ for $t \neq s$, and $E[u_t e_s] = 0$ for all t and s. Given the measurements $y_1, y_2, \ldots, y_{t-1}$, let $w(t-1|t-1)$ denote the minimum mean square estimate of w_{t-1}. Let $\sigma^2 P(t-1|t-1)$ denote the estimation error covariance matrix, where $P(t-1|t-1)$ is known. That is,

$$E[(w_{t-1} - w(t-1|t-1))\,(w_{t-1} - w(t-1|t-1))^T] = \sigma^2 P(t-1|t-1).$$

Let $w(t|t-1)$ and $P(t|t-1)$ denote the predicted values of w_t and P_t given $y_1, y_2, \ldots, y_{t-1}$. These quantities are given by the prediction equations

$$w(t|t-1) = Aw(t-1|t-1) \tag{3.3}$$

and

$$P(t|t-1) = AP(t-1|t-1)A^T + BQB^T \tag{3.4}$$

Using the t^{th} observation, y_t, $w(t|t)$ and $P(t|t)$ are obtained by the update equations

$$w(t|t) = w(t|t-1) + P(t|t-1)zb_t^{-1}(y_{t-1} - w(t|t-1)) \tag{3.5}$$

and

$$P(t|t) = P(t|t-1) - P(t|t-1)zb_t^{-1}z^TP(t|t-1) \tag{3.6}$$

where

$$b_t = z^TP(t|t-1)z + R. \tag{3.7}$$

The ARMA(p,q) model in equation (2.1) can be written as

$$x_t = \sum_{k=1}^{r} \alpha_k x_{t-k} + \epsilon_t + \sum_{k=1}^{r} \beta_k \epsilon_{t-k} \tag{3.8}$$

where $r = \max(p, q+1)$.

Alternatively, (3.8) can be expressed in Markovian form by

$$w_t = \begin{bmatrix} \alpha_1 & 1 & 0 & \dots & 0 \\ \alpha_2 & 0 & 1 & \dots & 0 \\ . & . & 0 & \dots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ \alpha_{r-1} & 0 & . & \dots & 1 \\ \alpha_r & 0 & 0 & \dots & 0 \end{bmatrix} w_{t-1} + \begin{bmatrix} 1 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_{r-1} \end{bmatrix} \epsilon_t \, , \tag{3.9}$$

where the first element of w_t is x_t. Equation (3.9) is the state equation (3.2) in the state space formulation of the ARMA(p,q) model. The measurement equation corresponding to (3.1) is

$$x_t = z^T w_t \tag{3.10}$$

where

$$z^T = (1, 0, 0, \dots, 0_r).$$

For ARMA(p, q) models without measurement error, $R = 0$, $Q = 1$,

$$A = \begin{bmatrix} \alpha_1 & 1 & 0 & 0 & \dots & 0 \\ \alpha_2 & 0 & 1 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \alpha_{r-1} & 0 & 0 & 0 & \dots & 1 \\ \alpha_r & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \tag{3.11}$$

and

$$B = \begin{bmatrix} 1 \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{r-1} \end{bmatrix}. \tag{3.12}$$

To start the recursions in equations (3.3) to (3.7), initial values are needed for $w(0|0)$ and $P(0|0)$. When no observations are available, the minimum mean square estimate of w_0 is zero; therefore, $w(0|0)$ is set equal to zero. Since $P(0|0) = E[w_0 w_0^T]$, the covariance matrix of the state vector, the value of $P(0|0)$ is the solution P of the equation

$$P = APA^T + BQB^T \ . \tag{3.13}$$

Equation (3.13) follows using equation (3.2) and noting the stationarity assumptions on the process (see Kalman, 1960). Using the fact that $APA^T = (A \otimes A)\mathrm{vec}(P)$ (Neudecker, 1969), (3.13) can be rewritten as

$$[I - A \otimes A]\mathrm{vec}(P) = \mathrm{vec}(BB^T) \tag{3.14}$$

where $\otimes$ indicates the Kronecker product. $[I - A \otimes A]$, $\mathrm{vec}(P)$ and $\mathrm{vec}(BB^T)$ are defined by

$$[I - A \times A] = \begin{bmatrix} (I-\alpha_1 A) & -A & 0 & \dots & 0 & 0 \\ -\alpha_2 A & I & -A & \dots & 0 & 0 \\ . & . & . & & . & . \\ . & . & . & & . & . \\ . & . & . & & . & . \\ -\alpha_{r-1} A & 0 & 0 & \dots & I & -A \\ -\alpha_r A & 0 & 0 & \dots & 0 & I \end{bmatrix}, \tag{3.15}$$

$$\mathrm{vec}(BB^T) = (B^T, \ \beta_1 B^T, \ \dots, \ \beta_r B^T)^T \ , \tag{3.16}$$

and

$$\mathrm{vec}(P) = (p_1^T, \ p_2^T, \ \dots, \ p_r^T)^T \tag{3.17}$$

where p_k is the k^{th} column of P. The $r^2 \times r^2$ matrix $[I - A \otimes A]$ does not have to be inverted. The system of equations (3.14) can be transformed so that the $r^2 \times r^2$ coefficient matrix is block lower triangular. In this form, only an r x r matrix inversion is required to obtain p_1; the remaining columns of P are obtained recursively. With $w(0|0)$ and $P(0|0)$ defined, equations (3.3) to (3.7) can be computed for each observation $y_t = x_t$ for t = 1,2, ...,n. Thus, given α, β and σ^2, $L_0(\alpha, \beta, \sigma^2)$ and $L_c(\alpha, \beta, \sigma^2)$ can be computed by setting $\sigma_t^2 = \sigma^2 b_t$ and $x(t|t-1) = z^T w(t|t-1)$ in equations (2.7) and (2.8). If $\sigma_t^2 = b_t$, $L_0(\alpha, \beta, \sigma^2) = L_0(\alpha, \beta)$ can be maximized by searching over α and β, and σ^2 is estimated by

$$s^2 = \frac{1}{n} \sum_{t=1}^{n} (x_t - x(t|t-1))^2 / b_t. \tag{3.18}$$

However, to maximize $L_c(\alpha, \beta, \sigma^2)$ requires searching over α, β, and σ^2; this follows by noting that an explicit expression for s^2 cannot be obtained from equation (2.10). In fact, the estimates a(c), b(c), and $s^2(c)$ are interrelated. The Kalman filter algorithm as presented includes the possibility of observation error. Thus, for ARMA models with observation error $u_t \neq o$, the additional parameter R, the variance of u_t, must be estimated. That is, the log likelihood and generalized likelihood must be maximized with respect to α, β, σ^2, and R. The observations are $y_t = x_t + u_t$, where x_t is the ARMA process of (2.1) or (3.8). Then, y_t replaces x_t in (2.7) and

(2.8); $\tilde{y}_{t-1}$ replaces $\tilde{x}_{t-1}$ in (2.8). The inclusion of observation error in the model can result in a more parsimonious model representation (see Box and Jenkins, 1970).

Since L(c) is nonlinear for ARMA(p,q) processes, the Ellipsoid algorithm of Kupferschmid and Ecker (1984) for nonlinear optimization is used to calculate the maximum of L(c). Without a priori information, the initial $a_i(c)$'s and $b_i(c)$'s are set to zero and, for $c \neq 0$, $s^2(c)$ is set equal to one half the sample variance of the data. An ellipsoid about these initial values must be given such that the ellipsoid contains the true parameters. In general, the algorithm will terminate at the maximum of L(c) inside the ellipsoid. If it is suspected that L(c) evaluated at a(c), b(c), and $s^2(c)$ is not the global maximum, the algorithm can be applied again using the current values of a(c), b(c), and $s^2(c)$ as the center of a new ellipsoid. In general, the size of the new ellipsoid will be smaller than the previous ellipsoid. In our experiments, we maximized L(0) by searching over α and β; σ^2 was estimated using equation (3.18). These estimates (denoted a(0), b(0) and $s^2(0)$) were used as initial values for the maximization of L(c). The difference between calculating L(0) and L(c) can be examined using equations (2.7), (2.8) and (2.11). L(0) requires calculating $\log \sigma_t^2$ and $(x_t - x(t|t-1))^2/2\sigma_t^2$, whereas L(c) requires calculating $(\sigma_t^2)^{-a/2}$ and $\exp[-c(x_t - x(t|t-1))^2/2\sigma_t^2]$.

Since the above are the only variations between calculating L(0) and L(c), the same computer program can be used for both calculations with a switch to indicate the evaluation of L(0) or L(c).

To ensure that the parameter estimates satisfy the stationarity and invertibility criteria in (2.2), we reparameterize in terms of the partial autoregressive coefficients $\tilde{a}_k$ and the partial moving average coefficients, $\tilde{b}_k$ which have values in the open interval (-1,1). The autoregressive and moving average coefficients are calculated by the Levinson-Durbin (1960) recursion (as cited in Jones, 1980). For $i = 1,2, \ldots, p$; $j = 1,2,\ldots, q$

$$a_i^{(i)} = \tilde{a}_i \tag{3.19a}$$

$$b_j^{(j)} = \tilde{b}_j \tag{3.19b}$$

and for $i > 1$, $j > 1$

$$a_k^{(i)} = a_k^{(i-1)} - \tilde{a}_k\, a_{i-k}^{(i-1)}, \text{ for } k = 1,2,\ldots, i-1, \tag{3.20a}$$

$$b_k^{(j)} = b_k^{(j-1)} + \tilde{b}_k\, b_{j-k}^{(j-1)}, \text{ for } k = 1,2,\ldots, j-1\ . \tag{3.20b}$$

The autoregressive coefficients are $a_k = a_k^{(p)}$ for $k = 1,2, \ldots, p$ and the moving average coefficients are $b_k = b_k^{(q)}$ for $k = 1,2,\ldots, q$.

The $\tilde{a}_k$ and $\tilde{b}_k$ can be constrained to the interval (-1,1) by adding the constraints

$$\begin{aligned} -1 + \tilde{a}_k &< 0 \\ -1 - \tilde{a}_k &< 0 \\ -1 + \tilde{b}_k &< 0 \\ -1 - \tilde{b}_k &< 0 \end{aligned} \tag{3.21}$$

in the Ellipsoid algorithm.

3.4 An ARMA(2,1) Example

As an illustration, maximum likelihood and model-critical parameter estimates are obtained for a simulated ARMA(2,1) process with representation

$$x_t = 1.52x_{t-1} - 0.9x_{t-2} + e_t + 0.4e_{t-1} \tag{4.1}$$

where e_t has zero mean and variance σ^2. Also, $E[e_t e_s] = 0$ for $t \neq s$. Figure 3.1 is a plot of a realization of x_t. Table 3.1 shows the parameter estimates for c = 0, 0.1, 0.2, 0.3, and 0.4 when e_t has a normal distribution with $\sigma^2 = 1$. It can be seen that the model-critical and maximum likelihood estimates are approximately the same. Next, parameter estimates were obtained for x_t with e_t having a t-distribution with five degrees of freedom (Figure 3.2).

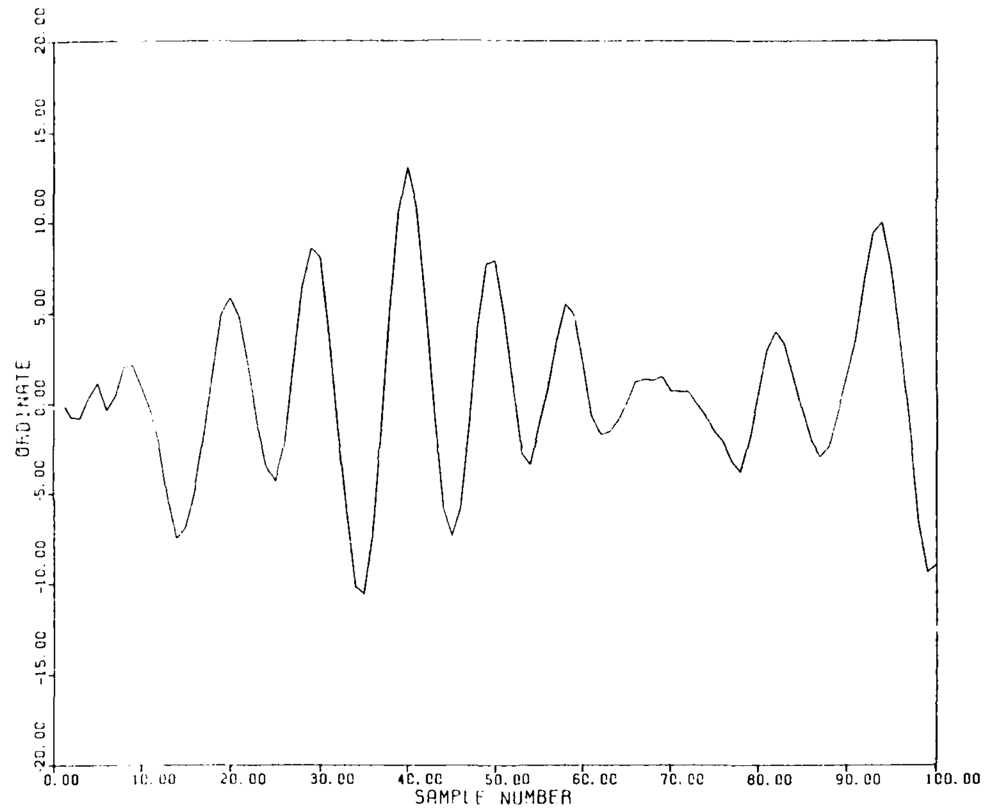

FIGURE 3.1 A Simulated ARMA(2,1) Process with Innovations Distributed Normal(0,1)

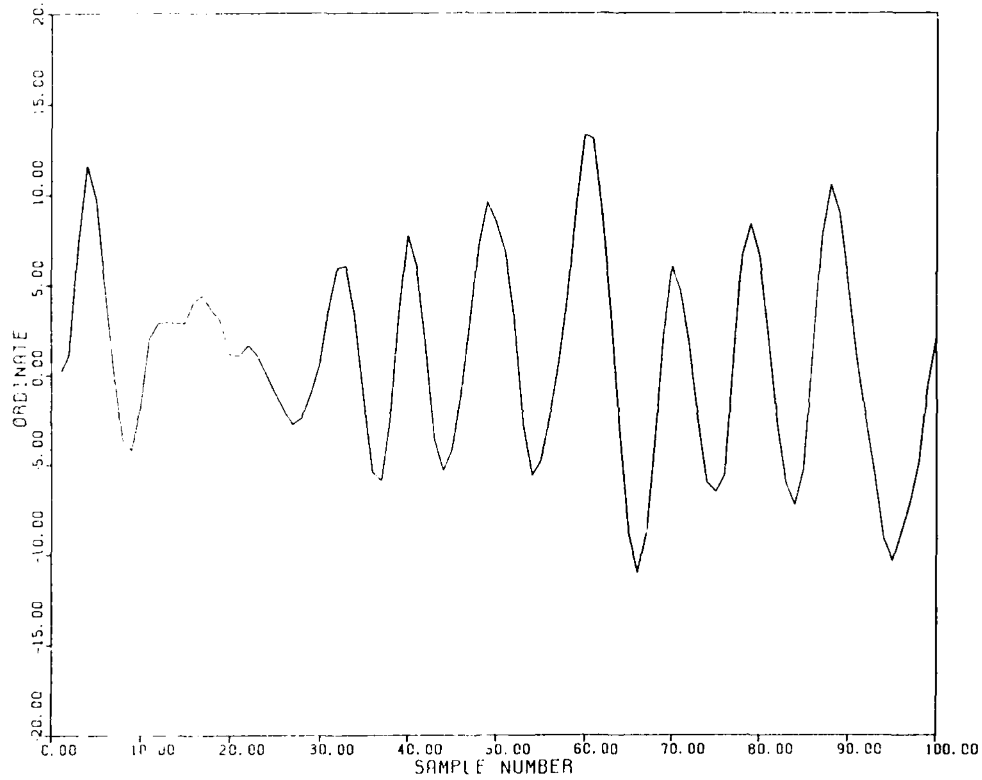

FIGURE 3.2 A Simulated ARMA(2,1) Process with Innovations Distributed t(5)

TABLE 3.1

Maximum Likelihood ($c = 0$) and Model-Critical ($c \neq 0$) Parameter Estimates for a Simulated ARMA(2,1) Process with Innovations Distributed N(0,1), and No Outliers

c	a_1	a_2	b_1	s^2
0.0	1.5912	-0.9255	0.3253	0.6260
0.1	1.5924	-0.9272	0.3366	0.5983
0.2	1.5929	-0.9285	0.3355	0.5591
0.3	1.5950	-0.9307	0.3296	0.5209
0.4	1.5948	-0.9303	0.3205	0.4945

TABLE 3.2

Maximum Likelihood ($c = 0$) and Model-Critical ($c \neq 0$) Parameter Estimates for a Simulated ARMA(2,1) Process with Innovations Distributed t(5), and No Outliers

c	a_1	a_2	b_1	s^2
0.0	1.4599	-0.8179	0.5314	1.5432
0.1	1.4676	-0.8214	0.4963	1.4018
0.2	1.4756	-0.8235	0.4747	1.2777
0.3	1.4807	-0.8274	0.4550	1.1761
0.4	1.4836	-0.8287	0.4467	1.0934

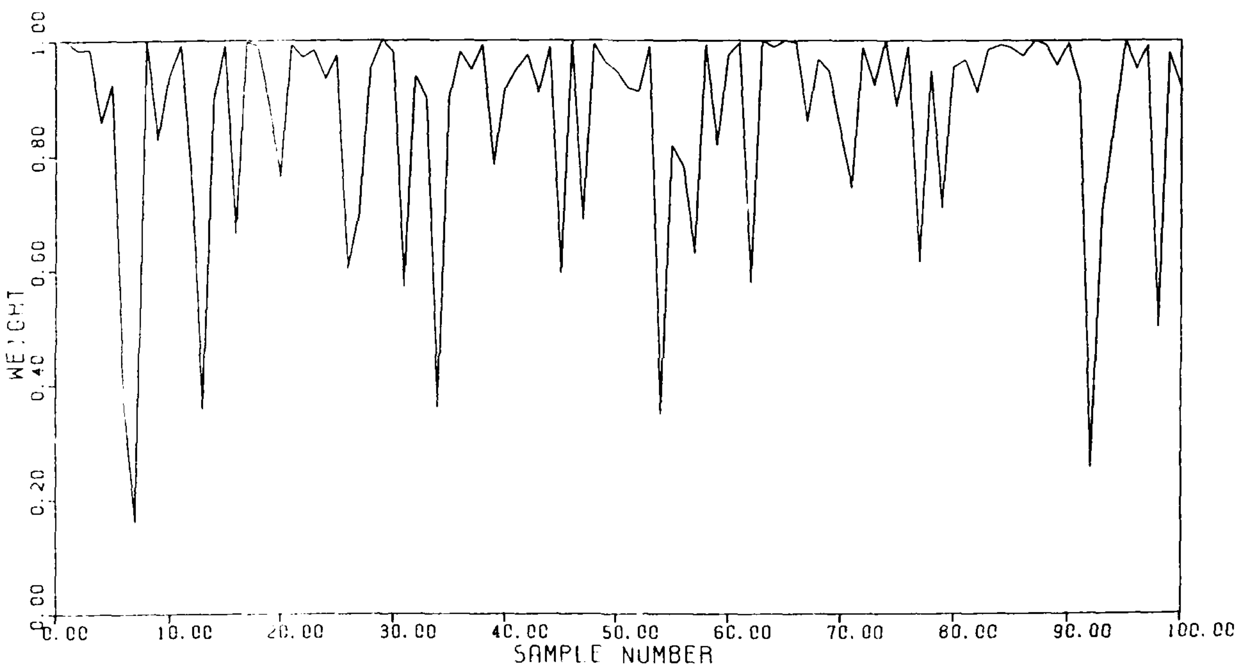

FIGURE 3.3 Model-Critical Weights for the Simulated ARMA(2,1) Process with Innovations Distributed Normal(0,1); c = 0.3

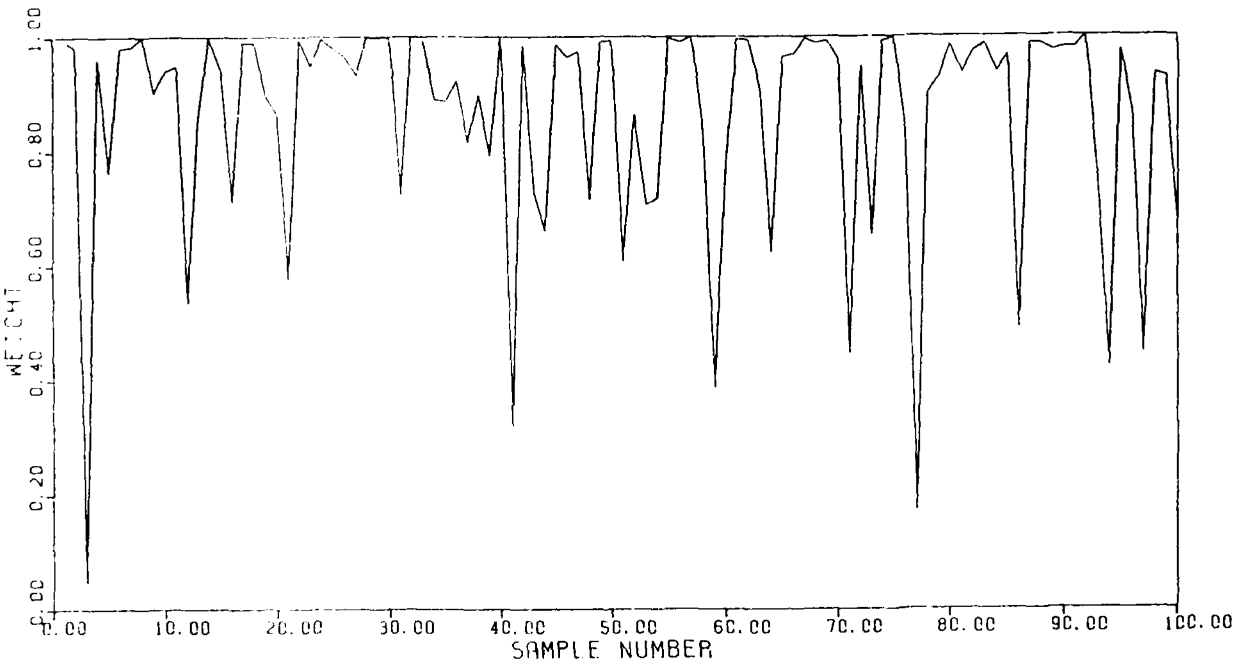

FIGURE 3.4 Model-Critical Weights for the Simulated ARMA(2,1) Process with Innovations Distributed $t(5)$; $c = 0.3$

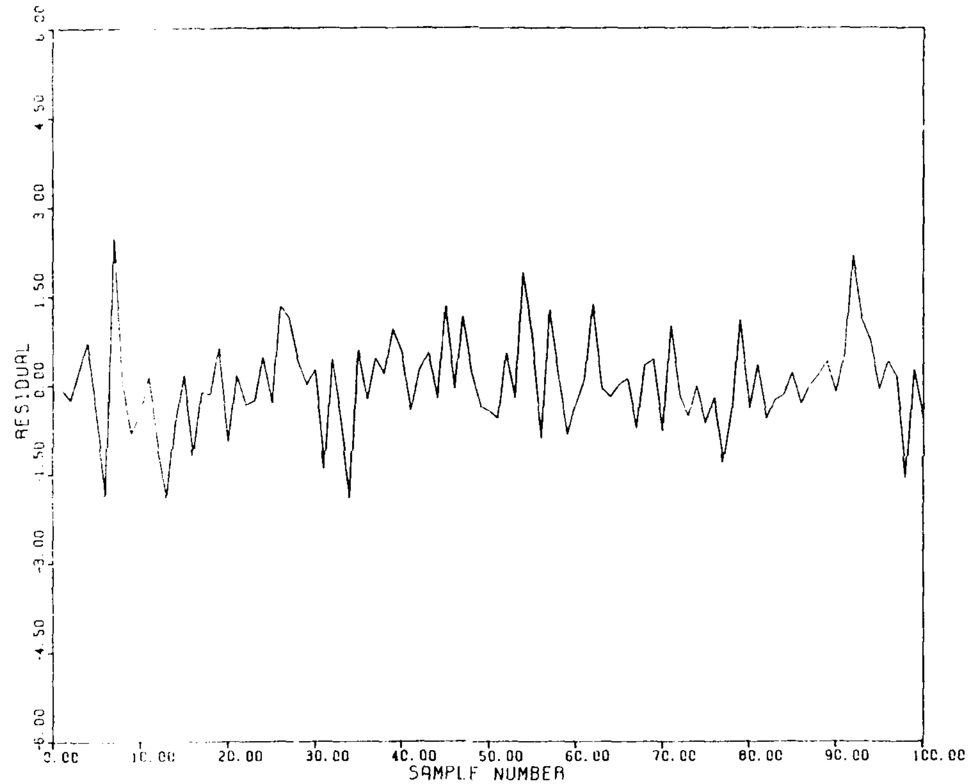

FIGURE 3.5 Maximum Likelihood Residuals for the Simulated ARMA(2,1) Process with Innovations Distributed Normal(0,1)

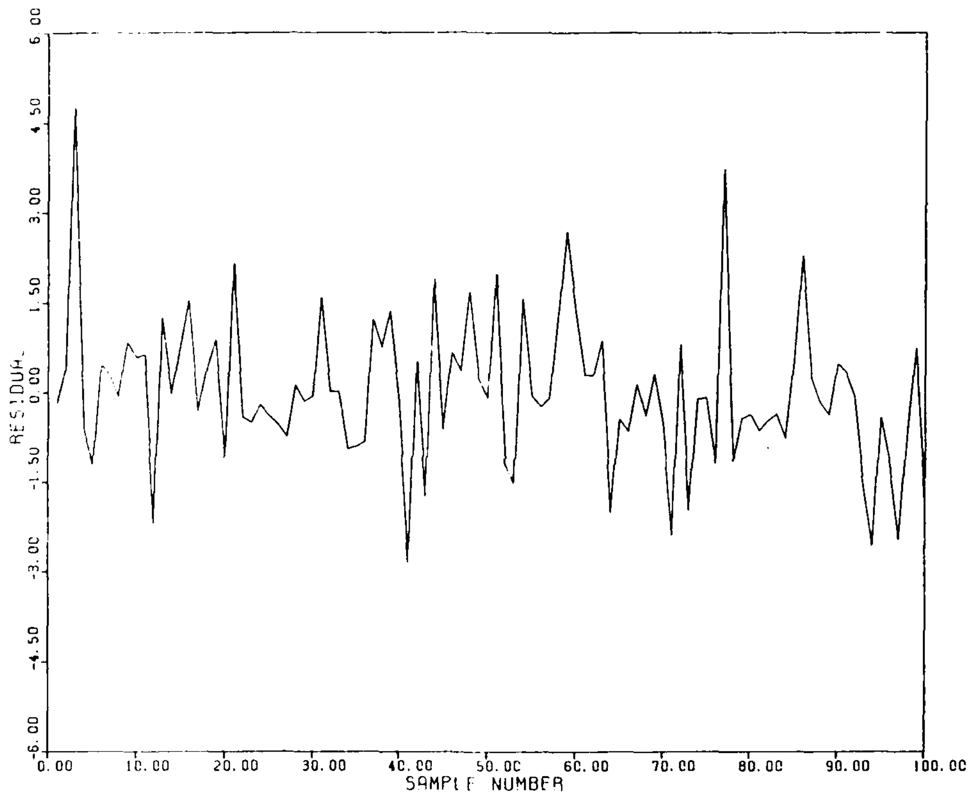

FIGURE 3.6 Maximum Likelihood Residuals for the Simulated ARMA(2,1) Process with Innovations Distributed t(5)

This corresponds to the case of innovative outliers of Fox (1972), and Martin and Thomson (1982). Table 3.2 presents the parameter estimates for c = 0, 0.1, 0.2, 0.3, and 0.4. As in the case when e_t is normal, the ARMA parameter estimates (a_1, a_2, b_1) do not change much as c increases. However, when e_t has a t-distribution, the estimate of the innovations variance σ^2 decreases considerably as c increases. The decrease in the value of $s^2(c)$ as c increases results from the downweighting of e_t values which are in the tails of the t-distribution. For c = 0.3, Figures 3.3 and 3.4 are plots of the weights $w_t = \exp(- c(x_t - x(t|x-t))^2/\sigma^2 b_t)$ for the ARMA(2,1) example with normal and t-distributed errors, respectively. Also, the maximum likelihood residuals for these two examples are shown in Figures 3.5 and 3.6. In both examples, the weights and residuals do not indicate any obvious problems with the data.

Next, four additive outliers (Fox, 1972) were added to the two realizations described above; plots of the realizations are shown in Figures 3.7 and 3.8. As in the plot of the AR(4) process with outliers, the outliers are not obvious. For c = 0, 0.1, 0.2, 0.3, and 0.4, Tables 3.3 and 3.4 present the parameter estimates. In both cases, $b_1(c)$ and $s^2(c)$ change considerably as c varies between 0 and 0.4. The changes in the moving average term $b_1(c)$ result from the additive outliers being downweighted. In both cases, the parameter estimate $b_1(c)$ moves in the direction of the true value. The presence of additive outliers in the data breaks up the structure of the moving

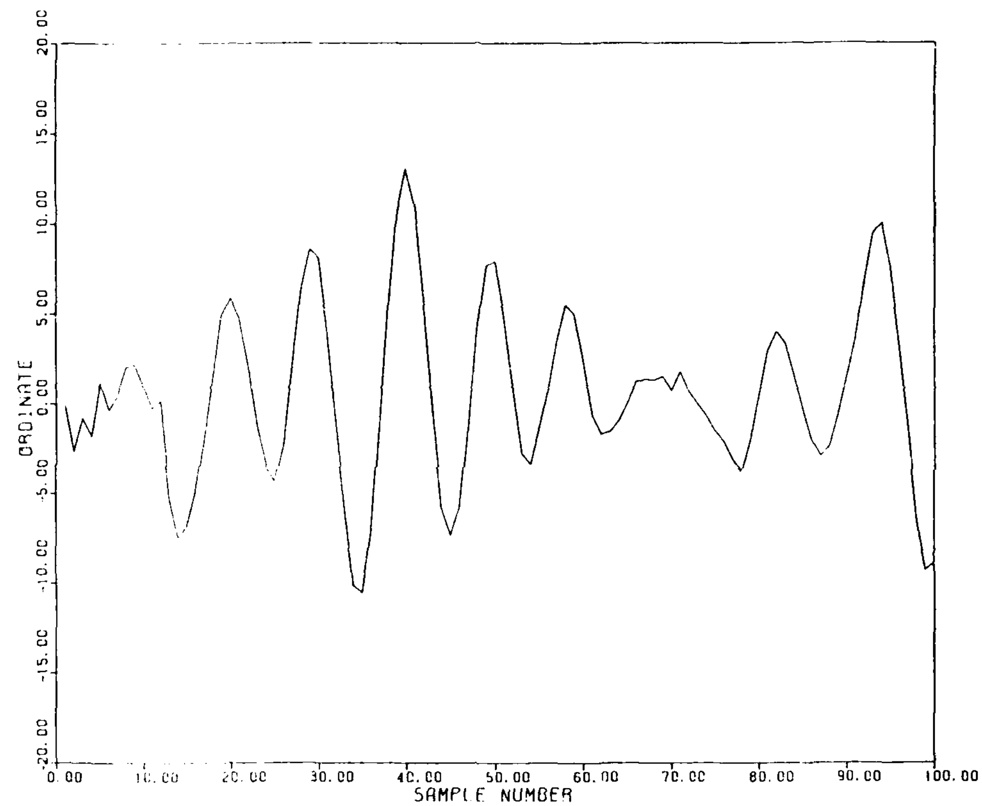

FIGURE 3.7 A Simulated ARMA(2,1) Process with Innovations Distributed Normal(0,1), and Four Additive Outliers

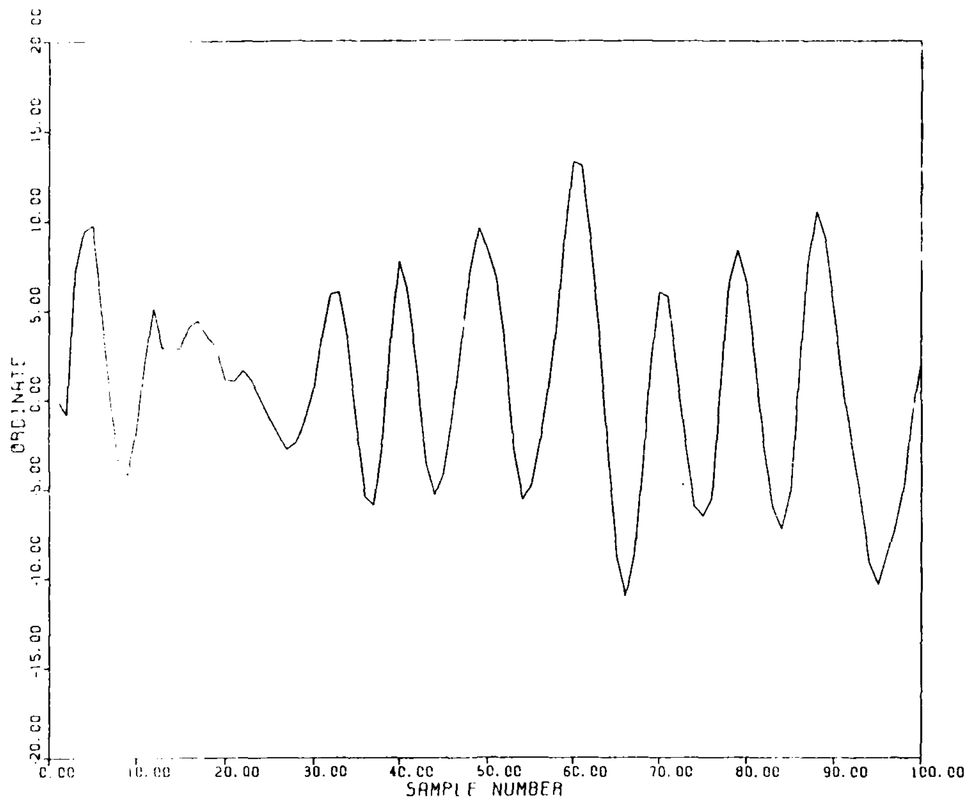

FIGURE 3.8 A Simulated ARMA(2,1) Process with Innovations Distributed t(5), and Four Additive Outliers

TABLE 3.3

Maximum Likelihood (c = 0) and Model-Critical (c ≠ 0) Parameter Estimates for a Simulated ARMA(2,1) Process with Innovations Distributed N(0,1), and 4 Additive Outliers

c	a_1	a_2	b_1	s^2
0.0	1.6063	-0.9339	-0.2227	1.3875
0.1	1.6040	-0.9370	-0.1556	1.0855
0.2	1.6066	-0.9433	-0.0925	0.8626
0.3	1.6045	-0.9426	-0.0381	0.7237
0.4	1.5996	-0.9403	0.0226	0.6309

TABLE 3.4

Maximum Likelihood (c = 0) and Model-Critical (c ≠ 0) Parameter Estimates for a Simulated ARMA(2,1) Process with Innovations Distributed t(5), and 4 Additive Outliers

c	a_1	a_2	b_1	s^2
0.0	1.4903	-0.8437	0.0943	2.4281
0.1	1.5064	-0.8553	0.0992	1.8670
0.2	1.5146	-0.8625	0.1263	1.6718
0.3	1.5222	-0.8679	0.1623	1.5568
0.4	1.5251	-0.8690	0.2088	1.4104

average part of the process as seen by $b_1(c)$; as a result, the unexplained moving average variation is summarized in $s^2(c)$. The model-critical weights for these two examples are shown in Figures 3.9 and 3.10. The plots indicate the presence of the outliers around observations 4 and 13.

The maximum likelihood and model-critical residuals are shown in Figures 3.11 and 3.12 for the example of Gaussian innovations with additive outliers. The residuals also indicate potential problems about samples 4 and 12.

3.5 Discussion

A Kalman filter algorithm has been used to obtain model-critical parameter estimates for univariate ARMA models. The examples show that model-critical procedures provide insight into the joint character of the data and the assumed ARMA model. The selection criterion of (2.4.17) can be used to select the ARMA model where p is replaced by (p + q). Part 4 extends the model-critical procedure discussed here to multivariate ARMA models. Part 5 presents a goodness of fit test which makes the comparison between maximum likelihood and model-critical parameter estimates more precise.

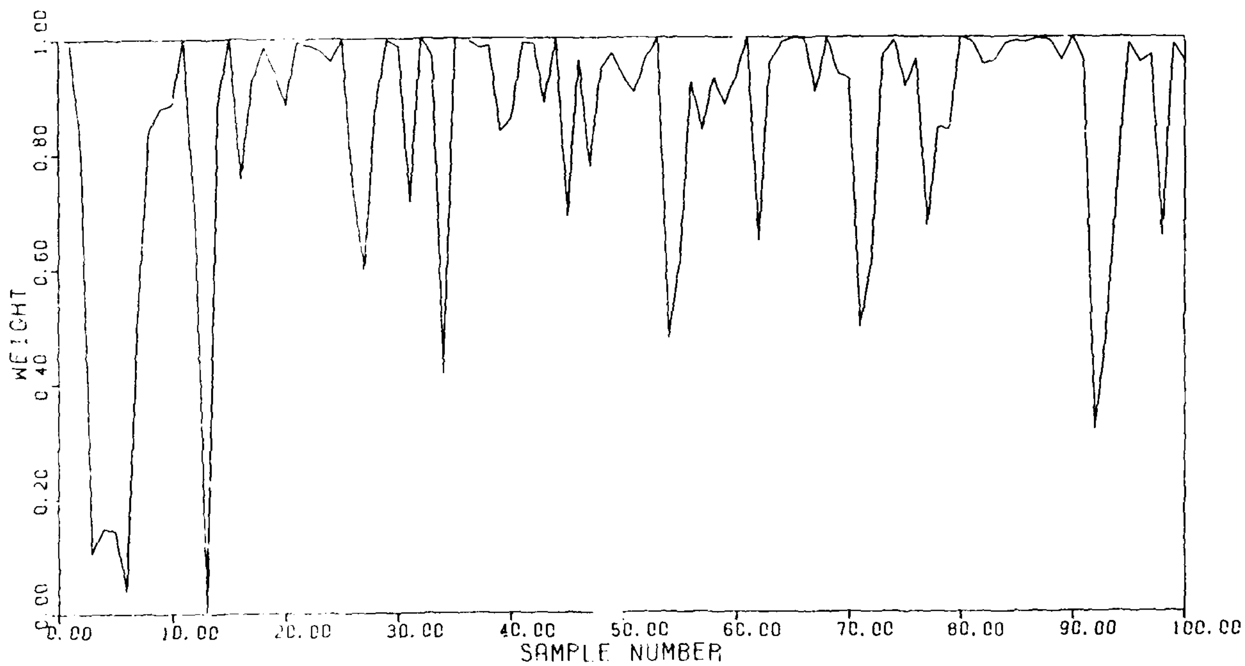

FIGURE 3.9 Model-Critical Weights for the Simulated ARMA(2,1) Process with Innovations Distributed Normal(0,1), and Four Additive Outliers; $c = 0.3$

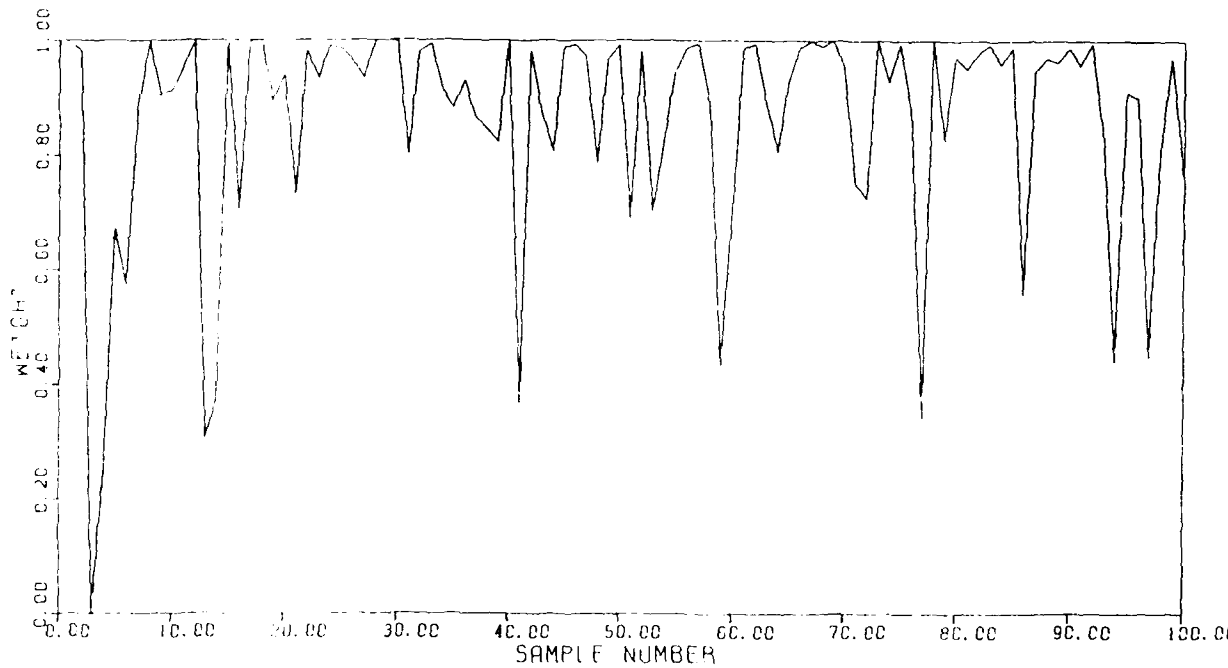

FIGURE 3.10 Model-Critical Weights for the Simulated ARMA(2,1) Process with Innovations Distributed t(5) and Four Additive Outliers; c = 0.3

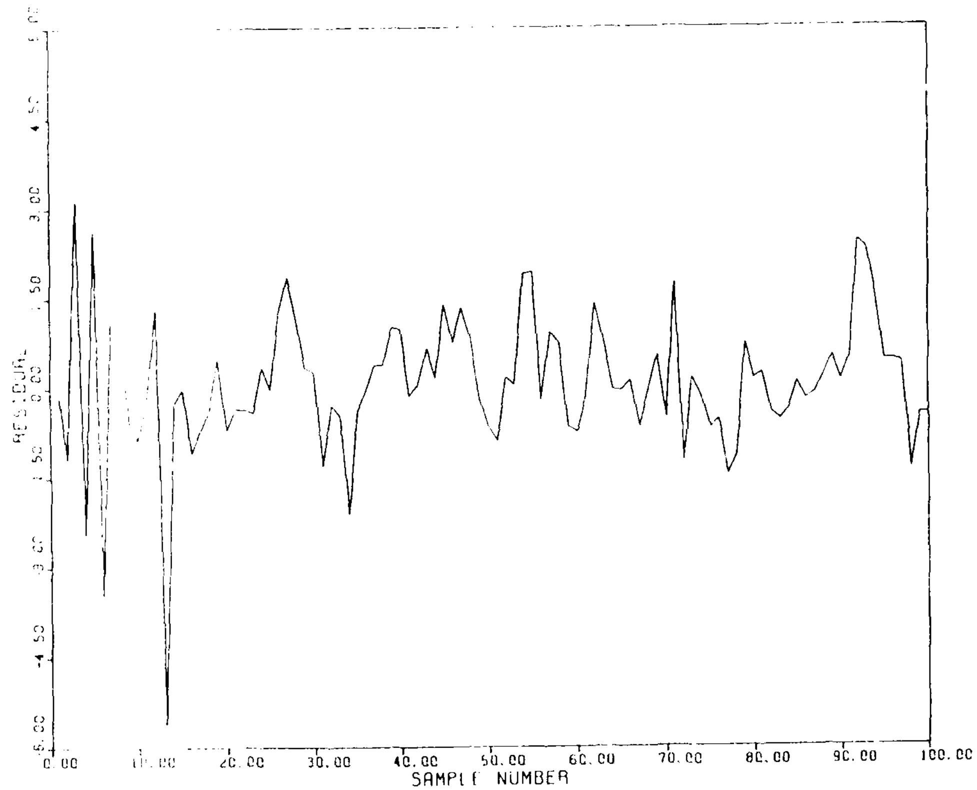

FIGURE 3.11 Maximum Likelihood Residuals for the Simulated ARMA(2,1) Process with Innovations Distributed Normal(0,1), and Four Additive Outliers

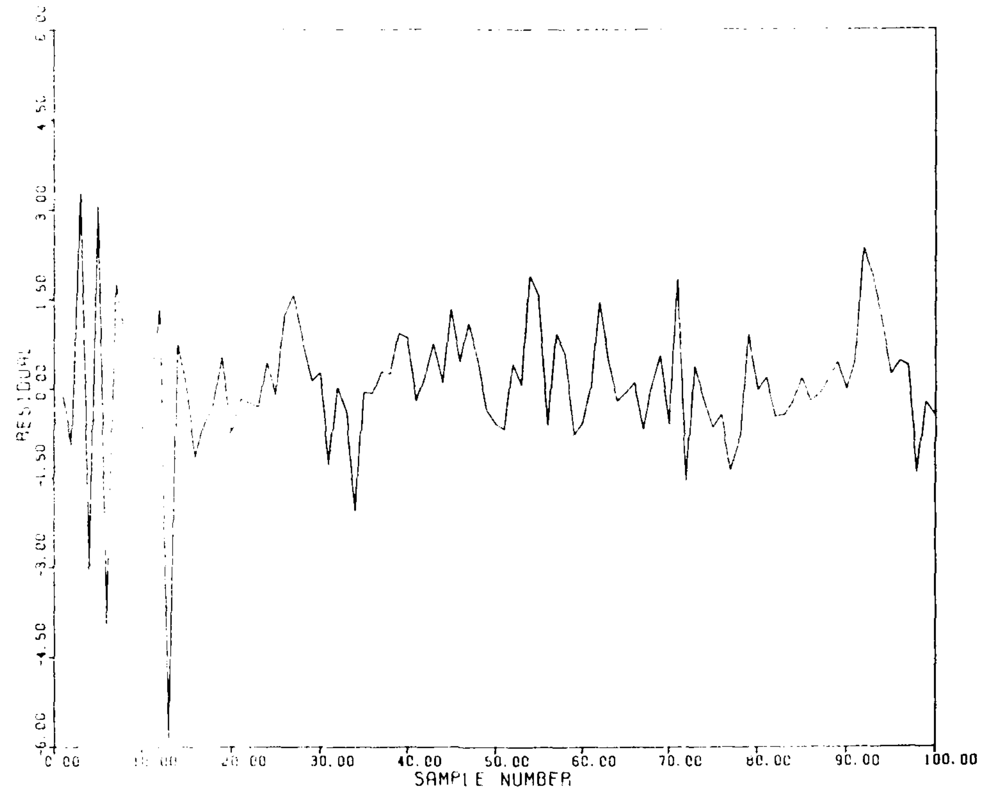

FIGURE 3.12 Model-Critical Residuals for the Simulated ARMA(2,1) Process with Innovations Distributed Normal(0,1), and Four Additive Outliers; c = 0.3

PART 4

MODEL-CRITICAL ESTIMATION FOR MULTIVARIATE ARMA MODELS

4.1 Introduction

In this part, the results of Part 3 are extended to multivariate ARMA models. Unlike univariate models, multivariate ARMA models have not received as much attention, primarily due to the complexity of multivariate processes. Estimation procedures for multivariate ARMA models have been presented by Wilson (1973), Osborn (1977) and Nicholls (1976); some theoretical background has been presented by Dunsmuir and Hannan (1976). Goodness of fit analysis for multivariate models has received even less attention (see Hosking, 1980, and Poskitt and Tremayne, 1982). From the examples in Parts 2 and 3, it can be seen that parametric and distributional deficiencies in the data can be masked by the structure of the process. This issue is complicated further by the interaction of the entries in a vector process. As in Parts 2 and 3, model-critical procedures will provide a means to assess the adequacy of a multivariate ARMA model by subjecting the data and the model to varying amounts of criticism.

For multivariate observations, it is straightforward to extend the Kalman filter algorithm described in Section 3.3 for the evaluation of the log likelihood and generalized likelihood functions. The nonlinear optimization algorithm of Kupferschmid and Ecker is used to maximize the likelihood functions. As in the univariate case, the same basic computer program can be used to calculate the log likelihood and generalized likelihood functions.

4.2 Generalized Likelihood for Multivariate ARMA Models

Let the m-vectors $x_1, x_2, \ldots, x_n$ be a realization of the autoregressive-moving average, ARMA(p,q), process with representation

$$x_t = \sum_{k=1}^{p} A_k x_{t-k} + \varepsilon_t + \sum_{k=1}^{q} B_k \varepsilon_{t-k} \tag{2.1}$$

where ε_t is distributed multivariate Gaussian with zero mean vector and covariance matrix D, and $E[\varepsilon_t \varepsilon_s^T] = 0$ for $t \neq s$. For stationarity and invertability, the roots of the determinants

$$\left| I - \sum_{k=1}^{p} A_k z^k \right| = 0$$

and (2.2)

$$\left| I + \sum_{k=1}^{q} B_k z^k \right| = 0$$

are assumed to lie outside the unit circle, $|z| = 1$. Using the multiplication rule, the log likelihood $L_0(x_1, x_2, \ldots, x_n | A, B, D)$ can be written as

$$L_o(x_1,x_2,\ldots x_n \mid A,B,D) = \sum_{t=1}^{n} \log f_t(x_t \mid \tilde{x}_{t-1},A,B,D) \tag{2.3}$$

where

$$\tilde{x}_{t-1} = (x_1, x_2, \ldots, x_{t-1}),$$

$$A = (A_1,A_2,\ldots,A_p),$$

$$B = (B_1,B_2,\ldots,B_q),$$

$$f_1(x_1 \mid \tilde{x}_o,A,B,D) = f_1(x_1 \mid A,B,D),$$

$$f_t(x_t \mid \tilde{x}_{t-1},A,B,D) = |2\pi F_t|^{-1/2}\exp[-\tfrac{1}{2}(x_t-x(t|t-1))^T F_t^{-1}(x_t-x(t|t-1))], \tag{2.4}$$

$$x(t|t-1) = E[x_t|\tilde{x}_{t-1},A,B,D], \tag{2.5}$$

and

$$F_t = \mathrm{Var}[x_t|\tilde{x}_{t-1},A,B,D] . \tag{2.6}$$

By using equations (2.4) through (2.6), the log likelihood equation (2.3) can be written as

$$L_o(x_1,x_2,\ldots,x_n|A,B,D) = -\frac{nm}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{n} \log |F_t| - \frac{1}{2}\sum_{t=1}^{n} (x_t - x(t|t-1))^T F_t^{-1}(x_t-x(t|t-1)) . \tag{2.7}$$

The model-critical parameter estimates are the A(c), B(c), and D(c) which maximize the generalized likelihood function

$$L_c(x_1,x_2,\ldots x_n|A,B,D) = \tag{2.8}$$

$$\frac{1}{c}\sum_{t=1}^{n}\left\{\frac{f_t^{\,c}(x_t|\,\tilde{x}_{t-1},A,B,D)}{Q_t^a(A,B,D,c)} - 1\right\}$$

where

$f_t(x_t\,|\,\tilde{x}_{t-1},A,B,D)$ is defined by (2.4),

$$Q_t(A,B,D,c) = [(1+c)^m\,|2\pi F_t|^{\,c}]^{-1/2}\ , \tag{2.9}$$

$a = c/(1+c)$ and $Q_t(A,B,D,c)$ is the information generating function of the density (2.4) as defined by (2.2.2).

For notational convenience, f_t and Q_t will denote $f_t(x_t\,|\,\tilde{x}_{t-1},A,B,D)$ and $Q_t(A,B,D,c)$, respectively, throughout the remainder of this part. Letting L(c) denote $L_c(x_1,\ x_2,\ \ldots,\ x_n\ A,B,D)$ without the constant term, the estimate for $\Theta = (A,B,D)$ is the solution $\Theta(c)$ to

$$\frac{\partial L(c)}{\partial\Theta} = 0 = \sum_{t=1}^{n}\frac{f_t^{\,c}}{Q_t^{\,a}}\left[\frac{\partial\log f_t}{\partial\Theta} - \frac{1}{(1+c)}\ \frac{\partial\log Q_t}{\partial\Theta}\right]. \tag{2.10}$$

It can be seen that each term in equation (2.10) is weighted by f_t^c/Q_t^a; for f_t and Q_t as defined above,

$$f_t^{\,c}/Q_t^{\,a} = \left(\frac{1+c}{|2\pi F_t|}\right)^{a/2} \exp\left[-\frac{c}{2}\,(x_t - x(t|t-1))^T\, F_t^{-1}(x_t - x(t|t-1))\right] \quad (2.11)$$

Thus, data corresponding to large values of $(x_t - x(t|t-1))^T F_t^{-1}(x_t - x(t|t-1))$ will be downweighted in the estimation process, the degree of downweighting being determined by the value of c. Model-critical estimation is a robust procedure that filters out non-Gaussian influences and finds the "best" multivariate ARMA(p,q) process consistent with Gaussianity. Unlike other robust procedures which estimate location (A,B) and scale (D) parameters separately, model-critical procedures estimate all parameters jointly using the assumed parametric and distributional form of the model. Using the distributional as well as the parametric model yields a natural framework for analyzing goodness of fit. The model-critical parameter c is similar to the robustness constant used by other robust procedures. For $c > 0$, data inconsistent with the Gaussian ARMA(p,q) model will receive small weights f_t^c/Q_t^a since the quadratic form in the exponential will be large. Values of $c > 0$ lead to criticism of outliers or heavy tailed distributions, whereas values of $c < 0$ lead to criticism of inliers or short tailed distributions. As the value of $|c|$ is increased, the procedure is more critical of the data and the model; hence, the use of the term model-critical. If the data and the Gaussian ARMA(p,q) model are consistent, then the model-critical estimates $\Theta(c)$ and the maximum likelihood estimates $\Theta(0)$ will be approximately equal; this provides a subjective measure of fit between the model and the data. Our experiments have shown that, for an ARMA(p,q) process with heavy tailed innovations, the estimates A(c) and B(c) will be close to the estimates A(0) and B(0) as c increases. However, D(c) and D(0) will differ considerably as c increases. When additive outliers (Fox, 1972) are present, A(c), B(c), and D(c) will differ from A(0), B(0), and D(0), respectively.

The additive outlier model (Martin and Thompson, 1982) is

$$y_t = x_t + v_t \tag{2.12}$$

where x_t is the ARMA process and v_t is the outlier. There are two common classes of additive outliers. The first is where $v_t \neq 0$ for few observations and v_t is large relative to e_t. The second class is where $v_t \neq 0$ for every observation and the process v_t has zero mean and covariance matrix R. In the latter situation, the covariance matrix of v_t, R, must also be estimated.

4.3 The Kalman Filter

The engineering literature abounds with applications of the Kalman filter such as missile tracking and parameter estimation. However, the statistical literature contains fewer applications. Harvey and

Phillips (1979) and Jones (1980) have presented Kalman filter algorithms to calculate the Gaussian maximum likelihood estimates for univariate ARMA models. Harvey and Phillips (1976) briefly discuss a Kalman filter algorithm to obtain maximum likelihood estimates for multivariate ARMA models. In general, the use of equation (2.10) to estimate Θ is not practical since expressions for $\partial \log f_t / \partial \Theta$ and $\partial \log Q_t / \partial \Theta$ are difficult to obtain explicitly for ARMA models. Instead, the vector Θ is estimated using equation (2.7) or (2.8) via a Kalman filter algorithm. The Kalman filter is used to calculate the value of L(c) and L(0) given Θ; the estimate of Θ is obtained by maximizing L(c) or L(0) using nonlinear optimization methods.

The following is a brief discussion of the Kalman filter algorithm which can be found in a number of references (see Kalman, 1960, or Gelb, 1974). Let the observation x_t be defined by the linear system

$$x_t = Z^T s_t + v_t \qquad \text{(measurement equation)} \qquad (3.1)$$

$$s_t = A s_{t-1} + B \varepsilon_t \qquad \text{(state equation)} \qquad (3.2)$$

where

x_t is a m x 1 vector of observations,

Z^T is a m x k matrix of known values,

s_t is a k x 1 vector of unknown state parameters, A is a k x k (transition) matrix,

B is a k x g matrix ($g \leq m$),

v_t is a m x 1 vector of Gaussian random variables with

$E[v_t] = 0$,

$E[v_t\ v_t^T] = R$ and $E[v_i\ v_j^T] = 0$ for all $i \neq j$,

ϵ_t is a g x 1 vector of Gaussian random variables with

$E[\epsilon_t] = 0$,

$E[\epsilon_t\ \epsilon_t^T] = C$ and $E[\epsilon_i\ \epsilon_j^T] = 0$ for $i \neq j$, and

$E[\epsilon_i\ v_j^T] = 0$ for all i and j.

Given the measurements $x_1, x_2, \ldots, x_{t-1}$, let $s(t-1|t-1)$ denote the minimum mean square estimate of s_{t-1}. Let $P(t-1|t-1)$ denote the estimation error covariance matrix where $P(t-1|t-1)$ is known. That is,

$$E[(s_{t-1} - s(t-1|t-1))(s_{t-1} - s(t-1|t-1))^T] = P(t-1|t-1) \quad .$$

Let $s(t|t-1)$ and $P(t|t-1)$ denote the predicted value of s_t and its corresponding error covariance matrix given $x_1, x_2, \ldots, x_{t-1}$; these quantities are given by the <u>prediction equations</u>

$$s(t|t-1) = As(t-1|t-1) \tag{3.3}$$

$$P(t|t-1) = AP(t-1|t-1)A^T + BCB^T \quad . \tag{3.4}$$

Using the information in the t^{th} observation x_t, $s(t|t)$ and $P(t|t)$ are obtained by the <u>update equations</u>

$$s(t|t) = s(t|t-1) + P(t|t-1)Z\, F_t^{-1}(x_t - Z^T s(t|t-1)) \tag{3.5}$$

$$P(t|t) = P(t|t-1) - P(t|t-1)Z\, F_t^{-1} Z^T P(t|t-1) \tag{3.6}$$

where

$$F_t = Z^T P(t|t-1)Z + R \tag{3.7}$$

The ARMA (p,q) model in equation (2.1) can be rewritten as

$$x_t = \sum_{k=1}^{r} A_k x_{t-k} + \varepsilon_t + \sum_{k=1}^{r} B_k \varepsilon_{t-k} \tag{3.8}$$

where $r = \max(p,q+1)$. Alternately, equation (3.8) can be expressed as the first order system

$$s_t = \begin{bmatrix} A_1 & I_m & 0 & \ldots & 0 & 0 \\ A_2 & 0 & I_m & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ A_{r-1} & 0 & 0 & \ldots & 0 & I_m \\ A_r & 0 & 0 & \ldots & 0 & 0 \end{bmatrix} s_{t-1} + \begin{bmatrix} I \\ B_1 \\ B_2 \\ \cdot \\ \cdot \\ \cdot \\ B_{r-1} \end{bmatrix} \varepsilon_t \qquad (3.9)$$

where the first m entries of s_t are the entries of x_t, I_m is the m x m identify matrix, and 0 is the m x m matrix of zero entries. Equation (3.9) is the state equation (3.2) in the state space formulation of the ARMA(p,q) model. The measurement equation corresponding to (3.1) is

$$x_t = Z^T s_t + v_t \qquad (3.10)$$

where

$$Z^T = (I_m, 0, \ldots, 0) .$$

There are r (m x m) matrix entries in Z and the product $Z^T s_t$ is equal to the first m entries of s_t. In the above formulation, if A_1, A_2, ..., A_r, B_1, B_2, ..., B_{r-1}, C, R, s(0|0) and P(0|0) are available, equations (3.3) to (3.6) can be used to compute s(t|t-1), P(t|t-1), s(t|t) and P(t|t) for t = 1,2,...,n. The quantities F_t and $(x_t - Z^T s(t|t-1))$ are used in equations (2.7) and (2.8) to evaluate L(0) and L(c). For the model of equation (2.1), R = 0 and C = D where all the matrices are m x m.

In order to start the recursions in equation (3.3) to (3.6), initial values are needed for s(0|0) and P(0|0). The value of s(0|0) is set equal to zero since this is the minimum mean square error estimate of s_o. From the stationarity of the first order autoregressive process s_t, the value of P(0|0) is the solution P_o of

$$P_o = TP_oT^T + UCU^T \tag{3.11a}$$

where

$$T = \begin{bmatrix} A_1 & I_m & 0 & \cdots & 0 \\ A_2 & 0 & I_m & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ A_{r-1} & 0 & 0 & \cdots & I_m \\ A_r & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{3.11b}$$

$$U = \begin{bmatrix} I_m \\ B_1 \\ B_2 \\ \vdots \\ B_{r-1} \end{bmatrix} \tag{3.11c}$$

Harvey and Phillips (1979) point out that $P(0|0)$ can be found by solving the linear system

$$[I - T \otimes T]\ \mathrm{vec}(P_o) = \mathrm{vec}(UCU^T) \tag{3.12}$$

for P_o (see Neudecker, 1969). The symbol $\otimes$ indicates the Kronecker product and the symbol vec(•) indicates forming a vector from the columns of the matrix. For example,

$$\mathrm{vec}(C) = \begin{bmatrix} C_1 \\ C_2 \\ \cdot \\ \cdot \\ \cdot \\ C_m \end{bmatrix}$$

where C_i is the i^{th} column of C. Using the definition of the Kronecker product and noting the form of T in (3.11b), $[I - T \otimes T]$ can be written as

$$\begin{bmatrix} (I - A_1 \otimes T) & -I_m \otimes T & 0 & \dots & 0 & 0 \\ -A_2 \otimes T & I & -I_m \otimes T & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ -A_{r-1} \otimes T & 0 & 0 & \dots & I & -I_m \otimes T \\ -A_r \otimes T & 0 & 0 & \dots & 0 & I \end{bmatrix} . \tag{3.13}$$

where I is the rm^2 identity matrix. It is easy to see that (3.13) can be transformed into the lower block diagonal form of (3.14).

$$\begin{bmatrix} (I - \sum_{i=1}^{r} A_i \otimes T^i) & 0 & 0 & \cdots & 0 & 0 \\ - \sum_{i=2}^{r} A_i \otimes T^i & I & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ -(A_{r-1} \otimes T + A_r \otimes T^2) & 0 & 0 & \cdots & I & 0 \\ -A_r \otimes T & 0 & 0 & \cdots & 0 & I \end{bmatrix} \tag{3.14}$$

$Vec(UCU^T)$ can be written as

$$Vec(UCU^T) = \begin{bmatrix} Vec(UC) \\ Vec(UCB_1^T) \\ \vdots \\ Vec(UCB_{r-1}^T) \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{bmatrix} . \tag{3.15}$$

Performing the same operations to (3.15) that were used to transform (3.13) to (3.14) yields

$$\begin{bmatrix} \sum_{i=1}^{r} (I \otimes T^{i-1})\, b_i \\ \sum_{i=2}^{r} (I \otimes T^{i-2})\, b_i \\ \vdots \\ b_{r-1} + (I \otimes T)\, b_r \\ b_r \end{bmatrix} . \tag{3.16}$$

Now P_o is an rm x rm matrix which we write as

$$P_o = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1r} \\ P_{21} & P_{22} & \cdots & P_{2r} \\ \vdots & \vdots & & \vdots \\ P_{r1} & P_{r2} & \cdots & P_{rr} \end{bmatrix} \tag{3.17}$$

where P_{ij} is an m x m matrix and $P_{ij} = P_{ji}^T$.

Let $P_i = \mathrm{vec} \begin{bmatrix} P_{1i} \\ P_{2i} \\ \vdots \\ P_{ri} \end{bmatrix}$ then using (3.14) and (3.17) we can recursively solve for the vector P_i, $i = 1, 2, \ldots, r$. That is

$$
\begin{aligned}
P_1 &= [I - \sum_{i=1}^{r} (A_i \otimes I^i)]^{-1} [\sum_{i=1}^{r} (I \otimes T^{i-1})\, b_i] \\
P_2 &= \sum_{i=2}^{r} [(A_i \otimes T^{i-1})\, P_1 + (I \otimes T^{i-2})\, b_i] \\
&\vdots \\
P_k &= \sum_{i=k}^{r} [(A_i \otimes T^{i-k+1})\, P_1 + (I \otimes T^{i-k})\, b_i] \\
&\vdots \\
P_r &= (A_r \otimes T)\, P_1 + b_r .
\end{aligned}
\tag{3.18}
$$

From the above recursion, it can be seen that only a $rm^2 \times rm^2$ matrix inversion is required rather than a $(rm)^2 \times (rm)^2$ matrix inversion.

With A, B, D, $s(0|0)$ and $P(0|0)$ available, $L(0)$, (2.6), and $L(c)$, (2.7), can be evaluated using equations (3.3) to (3.7) and can be maximized by searching over A, B, and D. Since $L(c)$, $c \geq 0$, is nonlinear for ARMA(p,q) models, $L(c)$ is maximized using the Ellipsoid algorithm of Ecker and Kupferschmid (1984) for nonlinear optimization. Without a priori information, the initial values for $A(c)$ and $B(c)$ are set equal to zero and $D(c)$ is set equal to one-half the sample covariance of the data. An ellipsoid about these initial values must be given such that the ellipsoid contains the optimal parameters. In general, the algorithm will terminate at the maximum of L(c) inside the ellipsoid. If it is suspected that L(c) evaluated at A(c), B(c) and D(c) is not the global maximum, the algorithm can be applied again using the current values of A(c), B(c) and D(c) as the center of a new ellipsoid; the size of the ellipsoid depends on how close the experimenter believes the current parameter values are to the optimal parameter values. In our experience, the maximum likelihood estimates A(0), B(0) and D(0) obtained from L(0) are used as the initial values of A(c), B(c) and D(c) for the maximization of L(c). This usually results in fewer iterations being required to obtain the final estimates of A(c), B(c) and D(c).

The innovations covariance matrix D is not estimated directly; since D is positive definite, D can be expressed as $D = LL^T$ where L is lower triangular. The matrix L is estimated by the algorithm and $= LL^T$. This procedure eliminates the need to add a constraint to ensure that D be positive definite. However, constraints may be needed to ensure that $|F_t| > 0$ for $t = 1,2,\ldots, t_{pq}$ for t_{pq} depending on p and q. Since the Ellipsoid algorithm is a constrained optimization procedure, the above constraints are easily incorporated by defining the constraints $g_t(A,B,D) = -|F_t| < 0$ for $t = 1, 2, \ldots, t_{pq}$. These constraints will aid in obtaining parameter estimates A(c) and B(c) corresponding to a stationary process.

For a multivariate ARMA process, model identification is a problem since the representation (2.1) is not necessarily unique. A discussion of identification and canonical forms for multivariate processes can be found in Hannan (1969), Mayne (1972), Denham (1974), and Dunsmuir and Hannan (1976). Identifiability constraints can be added to the Ellipsoid algorithm to assure a unique representation (2.1).

Examination of (2.7), (2.8) and (2.11) reveals that evaluations of (2.11) and log $|F_t|$ are the only differences in calculating L(c) and L(0); hence, the same basic computer program can be used to calculate L(0) or L(c).

The Kalman filter formulation allows for observation errors. That is, x_t is equal to the true ARMA process plus the observation errors v_t. In this situation, the parameter R, the observation error covariance matrix, must also be estimated. Since R is positive definite, $R = \tilde{R}\,\tilde{R}^T$ where $\tilde{R}$ is lower triangular. As in the estimation of D, $\tilde{R}$ is estimated by the algorithm and $\hat{\tilde{R}} = \hat{\tilde{R}}\,\hat{\tilde{R}}^T$. As in the univariate case, the use of the observation noise in the model can result in a more parsimonious model.

The Kalman filter provides a natural way to accommodate missing data. If an observation y_t is missing, the state vector $s(t|t)$ is set to $s(t|t-1)$, the predicted state vector for time t. Similarly, the error covariance matrix $P(t|t)$ is set to the predicted value

$P(t|t-1)$. Since y_t is missing, the functions $L(0)$ and $L(c)$ are not updated at time t. Clearly, the procedure can be applied to any number of missing observations; however, if a large number of observations are missing, the algorithm will essentially reinitialize at the next available observation.

The estimation of A and B requires specifying p and q. In general, p and q are not known; they must be obtained via a selection procedure such as the PSIC selection criterion (2.4.17). For multivariate processes, the PSIC criteria becomes

$$PSIC(p,q) = \log|D(c)| + \frac{2(p+q)m^2}{n} s(c) \log\log n \qquad (3.19)$$

where $s(c) > [(1 + c)^2/(1 + 2c)]^{(m/2 + 1)}$. For $c = 0$, the PSIC criteria reduces to the criterion of Hannan (1981). With additive outliers in the data, the PSIC selection criterion will select the model which describes the bulk of the data since data inconsistent with the model will be downweighted and thereby will have a reduced contribution to the model selected.

The computational burden of the multivariate Kalman filter is considerably greater than for the univariate Kalman filter. We have taken advantage of the structure of (3.9) to eliminate matrix multiplies involving zero matrices. The $m \times m$ matrix F_t must be inverted for each t. If the process is invertible, the system will tend towards steady state as t increases and F_t will tend toward the matrix $\hat{D}$. Consequently, for sufficiently large $t = \tilde{t}$, we can replace F_t^{-1} by $\hat{D}^{-1}$ and $|F_t|$ by $|\hat{D}|$ for $t > \tilde{t}$.

For the ARMA process without observation noise ($R = 0$), the matrix $P(t|t)$ has the form

$$P(t|t) = \begin{bmatrix} 0 & 0^T_{m(r-1)} \\ & \\ 0_{m(r-1)} & \tilde{P}(t|t) \end{bmatrix}$$

for $t = 1, 2, \ldots, n$, where $\tilde{P}(t|t)$ is a $m(r-1) \times m(r-1)$ matrix. Taking advantage of the special form of matrix T,

$$TP(t|t)T^T = \begin{bmatrix} \tilde{P}(t|t) & 0_{m(r-1)} \\ & \\ 0^T_{m(r-1)} & 0 \end{bmatrix}$$

which eliminates the need to perform any matrix multiplications in obtaining $P(t|t-1)$ from $P(t-1|t-1)$ via (3.4).

4.4 An Illustrative Example

We present an example to illustrate critical estimation for multivariate ARMA models. Consider the two-dimensional ARMA(1,1) process with representation

$$x_t = Ax_{t-1} + e_t + Be_{t-1} \tag{4.1}$$

where

$$A = \begin{bmatrix} 0.8 & 0.6 \\ -0.5 & 0.75 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.65 & -0.4 \\ 0.4 & 0.75 \end{bmatrix},$$

and e_t has a multivariate Gaussian distribution with zero mean vector and covariance matrix equal to the identity matrix. Further, $E[e_t\, e_s^T] = 0$ for $t \neq s$. A realization of the above process containing 120 samples was simulated and is shown in Figures 4.1a and 4.1b. The maximum likelihood and model-critical parameter estimates are presented in Table 4.1 for c = 0.025, 0.1, 0.2, and 0.3. Since the data and the model are internally consistent, the parameter estimates change little as c increases. Figure 4.2 is a plot of the critical weights

$$w_t = \exp[-c\ (x_t - x(t|t-1))^T\ F_t^{-1}\ (x_t - x(t|t-1)/2]\ . \tag{4.2}$$

Since by definition the model-critical weights are a measure of fit between the data and model, examination of the weights is an integral part of the modeling process. The weights shown in Figure 4.2 do not indicate any deficiencies in the model.

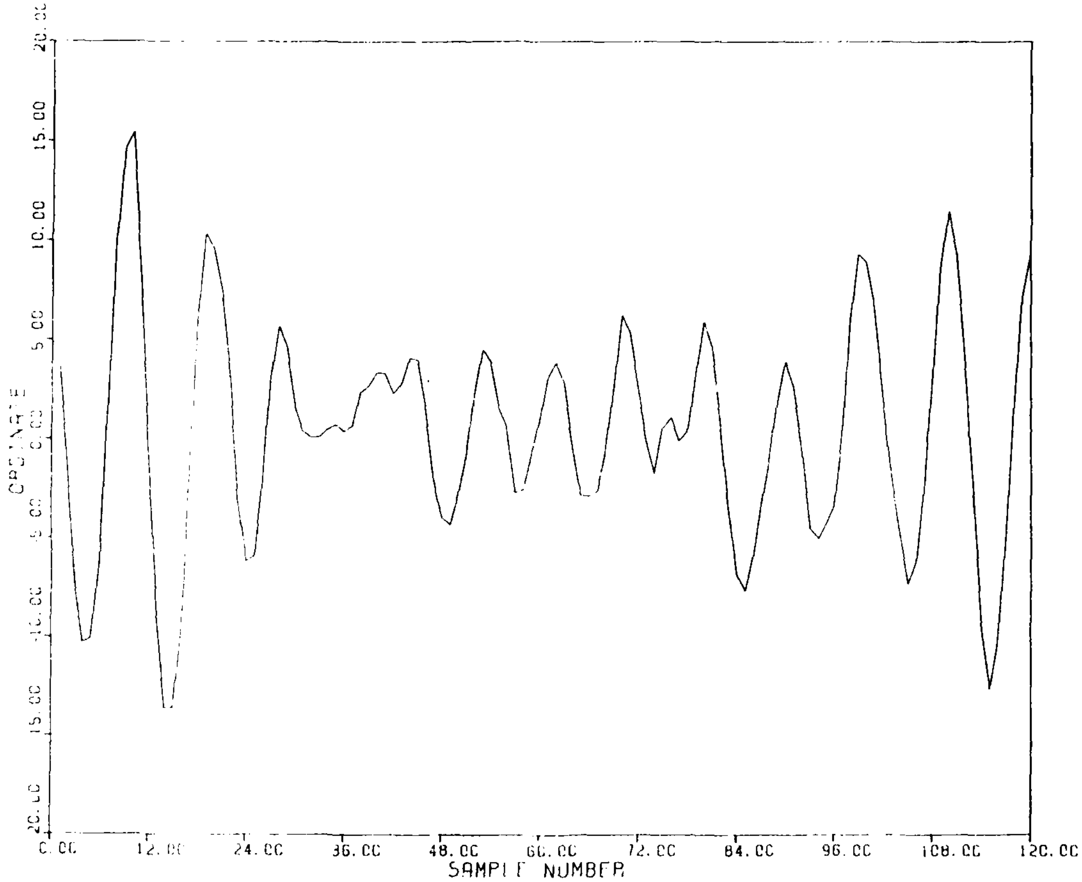

FIGURE 4.1a A Simulated Bivariate ARMA(1,1) Process with Innovations Distributed Normal $N_2(0,I)$; Series 1

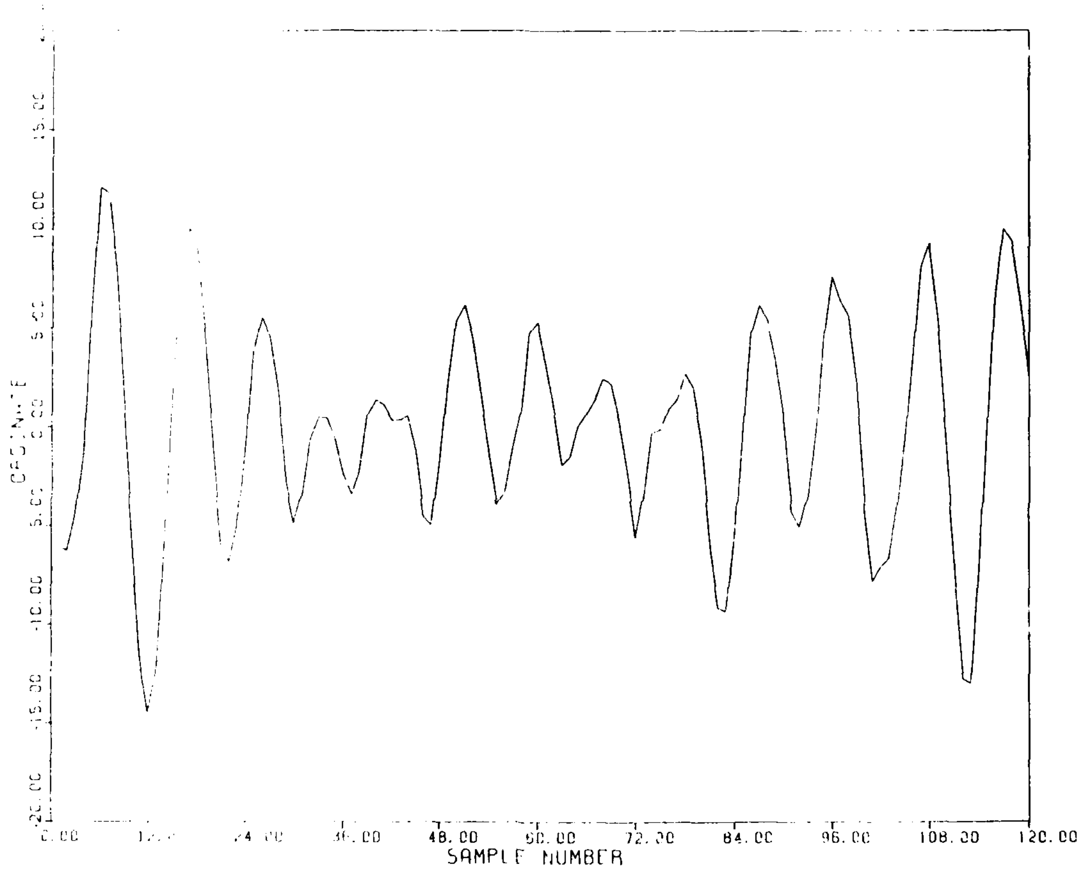

FIGURE 4.1b A Simulated Bivariate ARMA(1,1) Process with Innovations Distributed Normal $N_2(0,I)$; Series 2

TABLE 4.1

Maximum Likelihood ($c = 0$) and Model-Critical ($c \neq 0$) Parameter Estimates for a Simulated Bivariate ARMA(1,1) Process with Innovations Distributed Normal $N_2(0,I)$

c	0	0.025	0.1	0.2	0.3
a_{11}	0.828	0.825	0.819	0.816	0.823
a_{21}	-0.552	-0.538	-0.547	-0.550	-0.543
a_{12}	0.583	0.583	0.580	0.580	0.587
a_{22}	0.740	0.744	0.754	0.747	0.742
b_{11}	0.642	0.639	0.637	0.636	0.634
b_{21}	0.533	0.533	0.526	0.519	0.515
b_{12}	-0.447	-0.447	-0.451	-0.453	-0.453
b_{22}	0.727	0.729	0.732	0.736	0.741
d_{11}	0.836	0.836	0.832	0.825	0.812
d_{21}	0.034	0.037	0.038	0.041	0.043
d_{22}	0.810	0.814	0.826	0.839	0.852

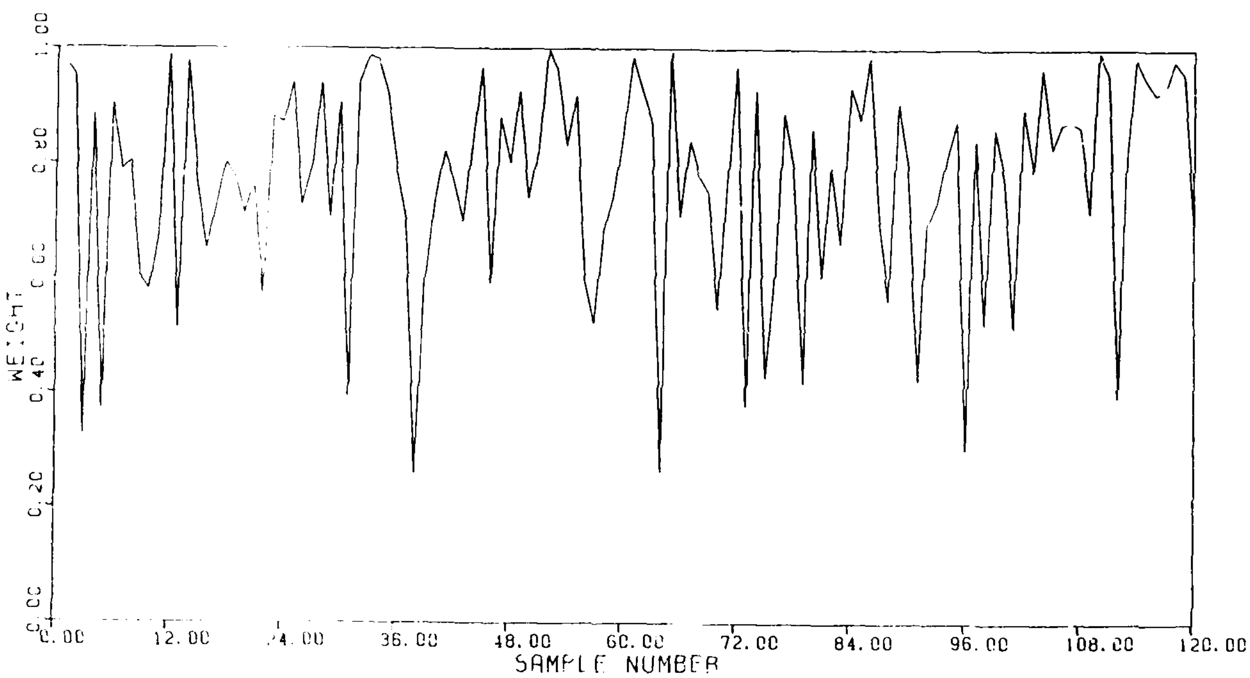

FIGURE 4.2 Model-Critical Weights for the Bivariate ARMA(1,1) Process with Innovations Distributed Normal $N_2(0,1)$; c = 0.3

To examine critical estimation of an ARMA process with outliers, four additive outliers were added at random to the realization discussed above; plots of the two series are shown in Figures 4.3a and 4.3b. The outliers were distributed multivariate Gaussian with zero mean vector and covariance matrix 2I where I is the 2 x 2 identity matrix. Each outlier is independent of x_t and the other outliers. Table 4.2 presents the maximum likelihood (c = 0) and model-critical parameter estimates for the ARMA(1,1) process with additive outliers. As in the univariate example, the outliers are not obvious from the plots, as can be seen by comparing Figures 4.3a and 4.3b with Figures 4.1a and 4.1b, respectively.

For this example, the moving average and covariance matrix parameters are the only parameters which change considerably as c increases. For c = 0.3, it can be seen that the model-critical parameter estimates are approximately the same as those obtained from uncontaminated realization. The change in the moving average and covariance matrix estimates as c increases from 0 to 0.4 results from the downweighting of the outliers. For c = 0, all the data are weighted equally; this results in the variation caused by the outliers being summarized in the covariance matrix. The outliers break up the moving average part of the process.

From another perspective, as c increases the estimation procedure becomes increasingly more critical of the data and model. This

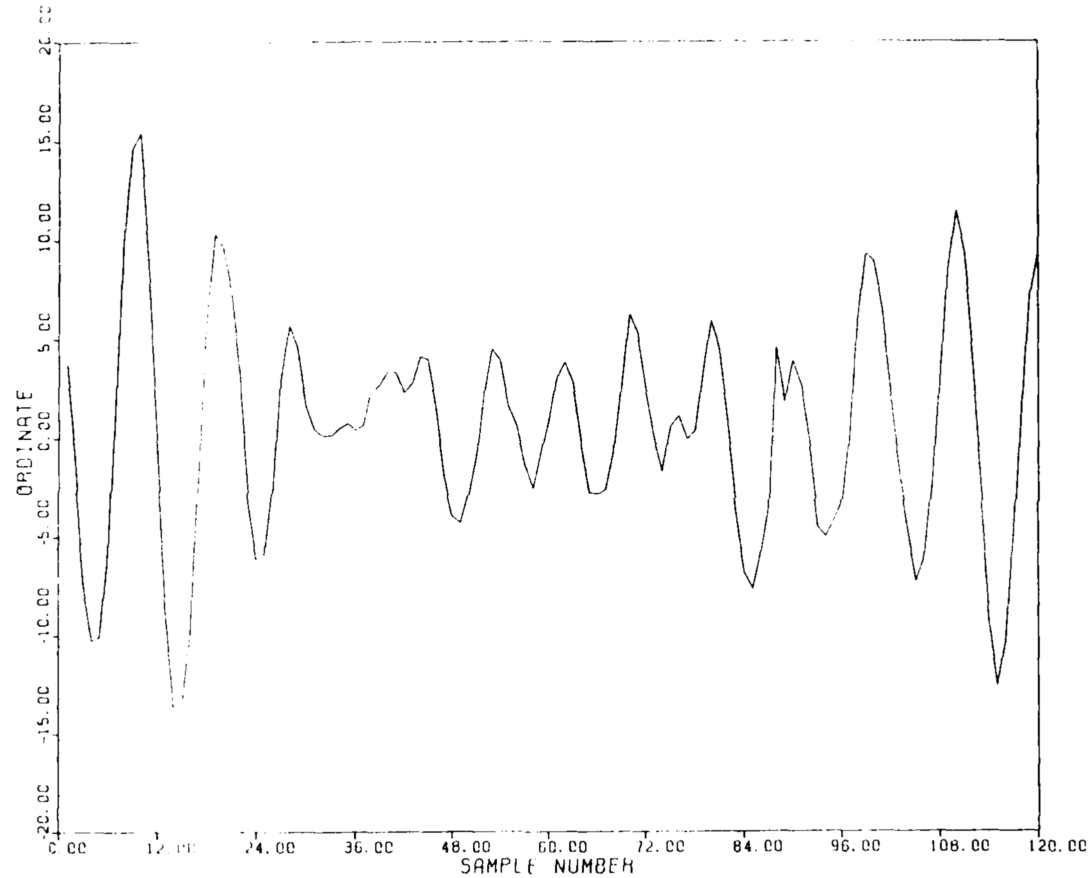

FIGURE 4.3a A Simulated Bivariate ARMA(1,1) Process with Innovations Distributed Normal $N_2(0,I)$, and Four Additive Outliers; Series 1

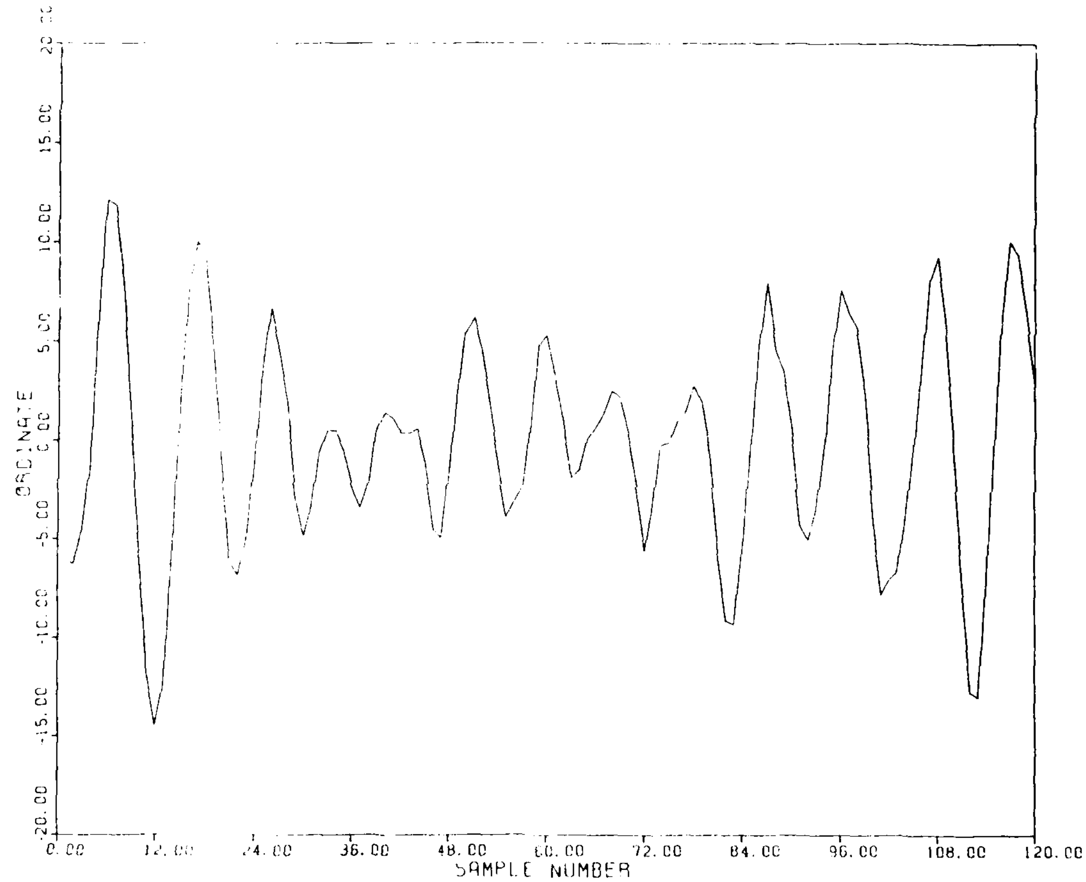

FIGURE 4.3b A Simulated Bivariate ARMA(1,1) Process with Innovations Distributed Normal $N_2(0,1)$, and Four Additive Outliers; Series 2

TABLE 4.2

Maximum Likelihood (c = 0) and Model-Critical (c ≠ 0) Parameter Estimates for a Simulated Bivariate ARMA(1,1) Process with Innovations Distributed Normal $N_2(0,I)$, and 4 Additive Outliers

c	0	0.025	0.1	0.2	0.3
a_{11}	0.826	0.824	0.832	0.828	0.820
a_{21}	-0.550	-0.546	-0.546	-0.539	-0.547
a_{12}	0.583	0.585	0.585	0.588	0.580
a_{22}	0.751	0.750	0.750	0.763	0.761
b_{11}	0.156	0.180	0.267	0.464	0.540
b_{21}	0.412	0.399	0.360	0.369	0.415
b_{12}	-0.096	-0.123	-0.210	-0.332	-0.351
b_{22}	0.485	0.499	0.532	0.516	0.522
d_{11}	1.498	1.425	1.192	0.938	0.855
d_{21}	-0.087	-0.065	0	0.018	0.008
d_{22}	1.219	1.198	1.127	1.071	1.037

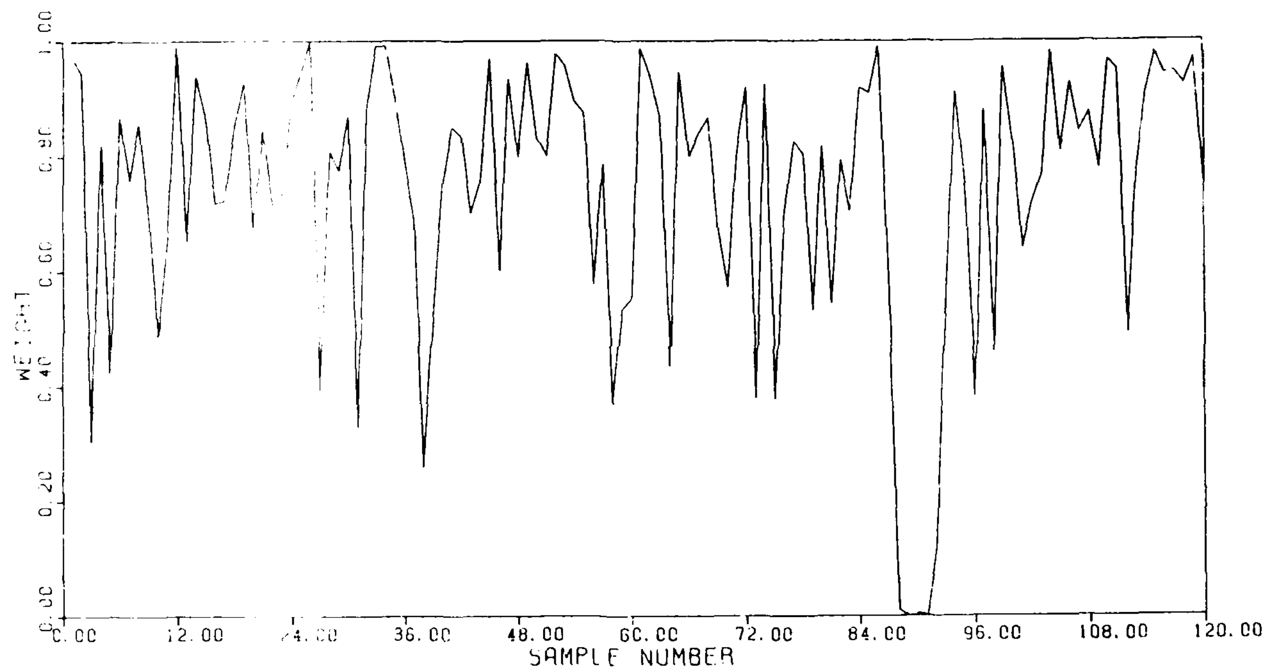

FIGURE 4.4 Model-Critical Weights for the Bivariate ARMA(1,1) Process with Innovations Distributed Normal $N_2(0,I)$, and Four Additive Outliers; $c = 0.3$

criticism can be seen by examining the critical weights which are shown in Figure 4.4. The small weights about observation 88 indicate that the model which describes the bulk of the data does not give a good fit to these observations. Since the observations in an ARMA process are not independent, a single additive outlier can result in the downweighting of neighboring observations. This is seen in the small weights about observation 88; observations 87 and 88 are both contaminated by additive outliers. The outliers at these observations result in the downweighting of samples 88, 89, 90, 91 and 92. This downweighting is necessary to reduce the effect of the outlier on $x(t|t-1)$.

To examine the effects of innovative outliers, a realization of (4.1) was examined where the entries of e_t were independent and identically t-distributed random variables with 5 degrees of freedom. Plots of the realization are shown in Figures 4.5a and 4.5b. Table 4.3 presents the maximum likelihood and model-critical estimates for $c = 0.025$, 0.1, 0.2 and 0.3. The location parameters $A(c)$ and $B(c)$ are approximately the same for $c = 0$, 0.025, 0.1, 0.2 and 0.3. However, the covariance matrix $D(c)$ changes considerably as c increases.

These examples provide insight into the analysis of ARMA models. If the location and scale parameters are approximately constant over a range of c, then the parametric and distributional model are

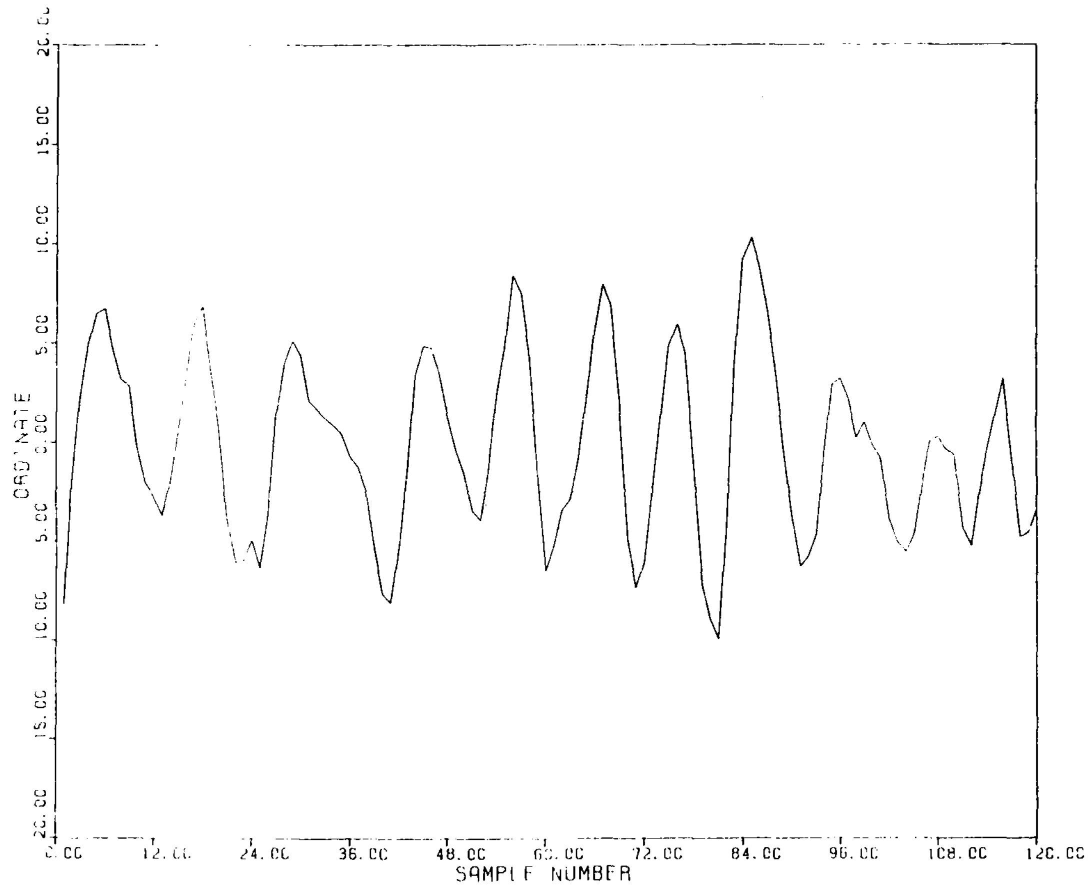

FIGURE 4.5a A Simulated Bivariate ARMA(1,1) Process with Innovations Distributed $t(5)$; Series 1

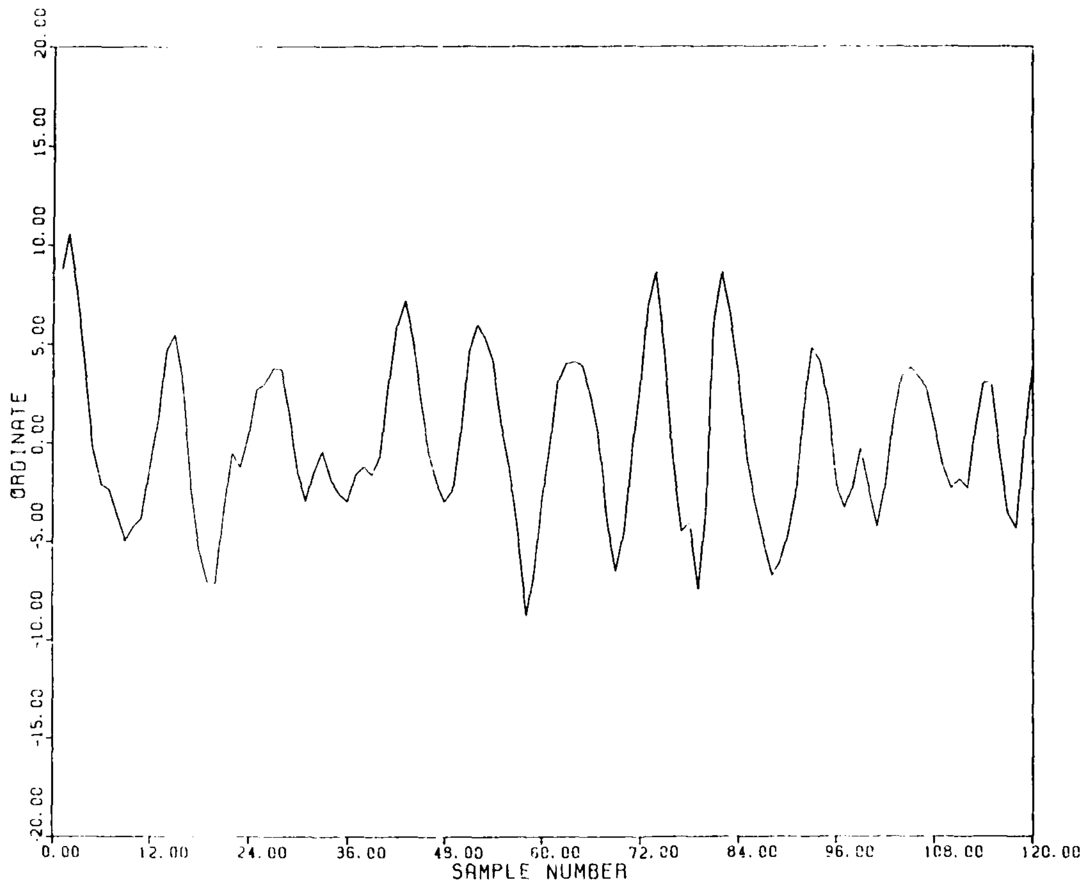

FIGURE 4.5b A Simulated Bivariate ARMA(1,1) Process with Innovations Distributed t(5); Series 2

TABLE 4.3

Maximum Likelihood ($c = 0$) and Model-Critical ($c \neq 0$) Parameter Estimates for a Simulated Bivariate ARMA(1,1) Process with Innovations Distributed t(5)

c	0	0.025	0.1	0.2	0.3
a_{11}	0.847	0.849	0.853	0.855	0.856
a_{21}	-0.485	-0.492	-0.506	-0.512	-0.513
a_{12}	0.577	0.586	0.571	0.567	0.556
a_{22}	0.614	0.607	0.606	0.606	0.606
b_{11}	0.532	0.537	0.545	0.554	0.561
b_{21}	0.376	0.374	0.369	0.375	0.385
b_{12}	-0.409	-0.409	-0.420	-0.422	-0.425
b_{22}	0.700	0.710	0.730	0.741	0.739
d_{11}	1.376	1.356	1.294	1.216	1.151
d_{21}	0.104	0.093	0.064	0.042	0.033
d_{22}	1.353	1.296	1.144	1.011	0.928

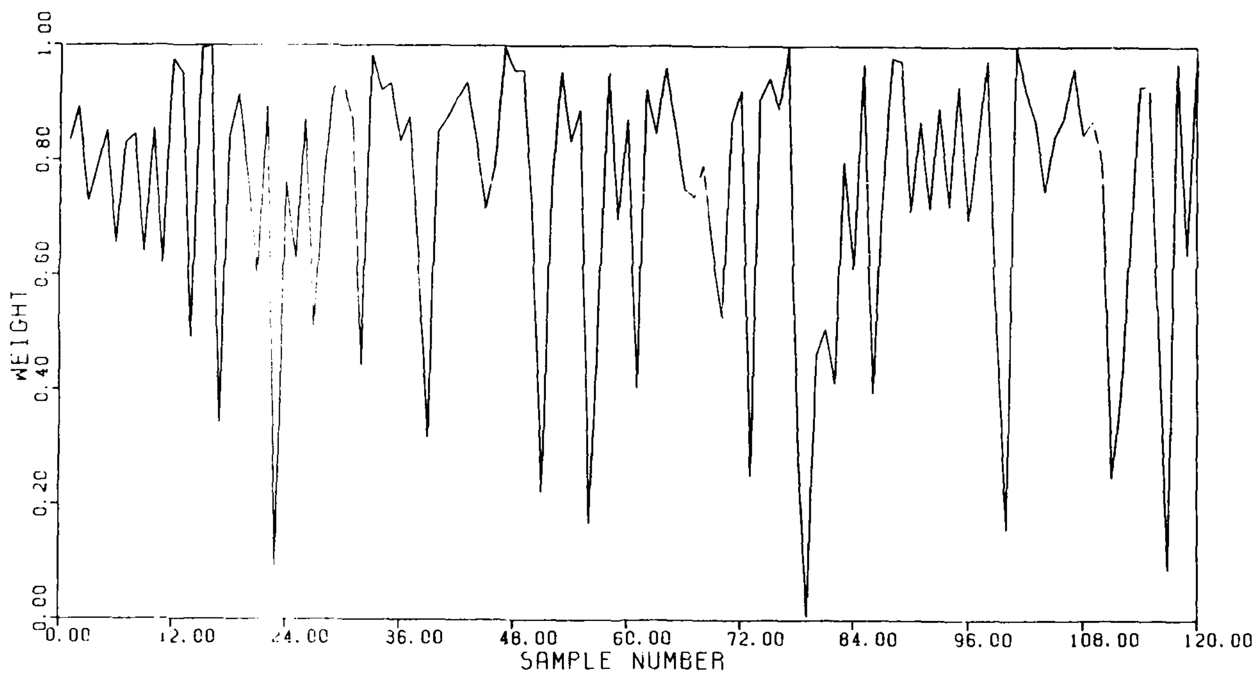

FIGURE 4.6 Model-Critical Weights for the Bivariate ARMA(1,1) Process with Innovations Distributed t(5); c = 0.3

reasonable. If the location parameters remain approximately constant over a range of c values but the scale parameters do not, then the Gaussian error model is suspect. Heavy tailed distributions yield variances that decrease as c increases, whereas the opposite occurs for short tailed distributions. Finally, if both scale and location parameters change considerably over a range of c values, then the presence of additive outliers is suspected.

4.5 Summary

Model-critical procedures have been presented for the analysis of multivariate ARMA models. These procedures provide a means to assess whether the observed data and the assumed model are internally consistent. A Kalman filter algorithm is used to obtain the model-critical parameter estimates. Since the samples are processed individually, the algorithm allows for data inconsistent with the model to be downweighted during the estimation process; the inconsistent data are identified by the critical weights which can aid the modeling process. The PSIC selection criterion can be used to select an ARMA model from a set of candidate models. In Part 5, a test for multivariate normality which compares $D(0)$ and $D(c)$ will be presented. It will be shown that the test can be applied to residuals after fitting with a linear model.

PART 5

A TEST FOR MULTIVARIATE NORMALITY BASED ON THE GENERALIZED LIKELIHOOD AND DIVERGENCE

5.1 Introduction

The Shapiro-Wilk test (Shapiro and Wilk, 1965) for univariate normality W is based on a comparison of two different estimates of the variance; if the random sample $x_1, x_2, \ldots, x_n$ is Gaussian, the two estimators should be close and their ratio should be unity, apart from sampling error. The distribution of

$$W = \left(\sum a_i x_i\right)^2 / \sum\left(x_i - \bar{x}\right)^2 \qquad (1.1)$$

is analytically intractable and must be developed by simulation. Here a_j are tabulated constants (Shapiro, 1980) and $x = (1/n)\sum x_i$. This idea may be extended to the multivariate setting by analogy, i.e., by determining two estimators of the covariance matrix and by determining a sensible way of comparing these estimators. We shall also require that the measure of closeness which institutes the test statistic be such that it is applicable to the case of a single sample without structure as well as the structured case. For example, given the concomitant variables $z_{i1}, z_{i2}, \ldots, z_{iq}$, the x_i have a combined systematic and error structure given by

$$x_i = h(z_{i1}, z_{i2}, \ldots, z_{iq}; \Theta_1, \Theta_2, \ldots, \Theta_q) + \varepsilon_i \qquad (1.2)$$

where h is (tentatively) functionally specified, apart from the parameters $\Theta_1, \Theta_2, \ldots, \Theta_q$ which are to be estimated from the data, and the ε_i are independent, identically distributed p-variate Gaussian variates with mean vector 0 and positive definite covariance matrix D. For the case of a single sample without structure, h of (1.2) reduces to h = m, for example, so that the x_i are p-variate Gaussian with mean m and covariance matrix D, denoted $N_p(m,D)$, with density

$$f(x) = \left|2\pi D\right|^{-1/2}\exp(-(x-m)^T D^{-1}(x-m)/2) \ . \tag{1.3}$$

The model representation allows for regression models, experimental design models, general linear models, and nonlinear models. Our test for p-variate normality will be constructed from consideration of the divergence (Kullback, 1959, Chapters 1, 2, and 8) which is discussed in the next section.

5.2 Information Statistics

From an information theory perspective, m(o), D(o) and m(c), D(c) of (2.2.7) can be thought of as the mean and covariance matrix for two normal distributions with probability densities f_o and f_c, respectively. The divergence J(o,c) per observation between the two densities f_o and f_c is defined by

$$J(o,c) = \int_{R_p} (f_o(x) - f_c(x))\log(f_o(x)/f_c(x))dx \tag{2.1}$$

where R_p denotes the p-dimensional Euclidean space (Kullback, 1959, Chapter 1). The quantity $J(o,c) \geq 0$ and is equal to zero if and only if $f_o = f_c$ (Kullback, 1959, Chapter 2). For n independent observations, the divergence is

$$J(o,c:n) = \sum_{i=1}^{n} J_i(o,c) \tag{2.2}$$

where $J_i(o,c)$ is the divergence in the i^{th} observation. If the observations are identically distributed, then $J(o,c:n) = nJ(o,c)$ (Kullback, 1959, Chapter 2).

By definition, $J(o,c)$ is a measure of the difference between two distributions with densities f_o and f_c. For the Gaussian probability densities f_o and f_c, it is straight forward to evaluate (2.1), which yields

$$\begin{aligned} J(o,c) = {} & \frac{1}{2}\, tr[D(o)D(c)^{-1} + D(c)D(o)^{-1}] - p \\ & + \frac{1}{2}\,(m(o) - m(c))^T [D(o)^{-1} + D(c)^{-1}]\,(m(o) - m(c)) \end{aligned} \tag{2.3}$$

where "tr" denotes the trace of a matrix. Let

$$J_1 = \frac{1}{2}\, tr[D(o)D(c)^{-1} + D(c)D(o)^{-1}] - p \tag{2.4}$$

and

$$J_2 = \frac{1}{2}(m(o) - m(c))^T[D(o)^{-1} + D(c)^{-1}](m(o) - m(c)) \qquad (2.5)$$

then $J(o,c)$ can be written as $J(o,c) = J_1 + J_2$. The term J_1 is a measure of the differences between the covariance matrices $D(o)$ and $D(c)$, whereas the term J_2 provides a measure of the differences between the means $m(o)$ and $m(c)$ with respect to $D(o)$ and $D(c)$. That is, $J(o,c)$ measures the difference between two normal distributions by comparing their means and covariances as seen in (2.3) through (2.5). If the data $x_1, x_2, \ldots, x_n$ are multivariate normal, then $m(c) \approx m(o)$ and $D(c) \approx D(o)$; hence, $J(o,c)$ will be small. If the data are not Gaussian, $D(c)$ will differ considerably from $D(o)$; however, $m(c)$ and $m(o)$ may or may not differ depending on the nature of the non-normality as seen by the example in Section 2.2. The expression for $J(o,c)$ in (2.2) is for unstructured Gaussian data. For data with additional structure as in (1.2), $J(o,c)$ can still be written as $J_1 + J_2$, where J_1 is defined by (2.4) and J_2 is defined by (2.5) with $m(o)$ and $m(c)$ replaced by $h_i(\Theta(o))$ and $h_i(\Theta(c))$, respectively. The quantity $h_i(\hat{\Theta}) = h(z_{i1}, z_{i2}, \ldots, z_{iq}; \hat{\Theta})$, where $\hat{\Theta} = (\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_q)^T$. Thus, the term J_1 is the same for data with or without additional structure, but J_2 is not.

From the above discussion, a family of tests for multivariate normality indexed by c are

$$T_1(c) = nJ_1 = n/2\left\{tr[D(o)D(c)^{-1} + D(c)D(o)^{-1}] - 2p\right\} . \qquad (2.6)$$

For each value of c, $T_1(c)$ produces a test statistic similar to the Shapiro-Wilk test statistic. Since D(c) and D(o) are estimates of the covariance matrix D for structured and unstructured models, the definition of $T_1(c)$ indicates that the effects of estimating structural parameters on D(c) and D(o) should approximately cancel. The same does not apply to J_2 since it depends on the concomitant variables $z_{i1}, z_{i2}, \ldots, z_{iq}$ and the model parameters $\theta_1, \theta_2, \ldots, \theta_q$. For example, for the univariate linear regression model $h_i(\theta) = \theta_o + \theta_1 z_i = \theta^T v_i$, where $\theta = (\theta_0, \theta_1)^T$ and $v_i = (1, z_i)^T$,

$$J_{2i} = v_i^T(\theta(o) - \theta(c))\ (s^{-2}(o) + s^{-2}(c))\ (\theta(o) - \theta(c))^T v_i,$$

for a single observation. For the entire sample,

$$J_2 = \sum_{i=1}^{n} v_i^T(\theta(o) - \theta(c))\ (s^{-2}(o) + s^{-2}(c))\ (\theta(o) - (\theta(c))^T v_i$$

which shows that J_2 depends on the predictor variables v_i. In fact, for structured data, J_2 is a measure of the differences between the predicted values $h_i(\theta(o))$ and $h_i(\theta(c))$. In Section 4, it will be shown that $T_1(c)$ is insensitive to the underlying model; thus, $T_1(c)$ provides a test of multivariate normality for the residuals from an assumed model such as a linear model.

Like the Shapiro-Wilk test statistic, the distribution of $T_1(c)$ must be obtained via Monte Carlo simulations because it is otherwise intractable. Since large values of $T_1(c) \geq 0$ indicate non-normality, only the upper percentage points of the test statistic are required. Tables 5.1 to 5.8 (a-e) contain percentage points of $T_1(c)$ for $p = 1(1)6$, 8, and 10; sample sizes n = 10, 20, 24, 30, 40, 60, and 120;and c values dependent on the dimension p. For p = 5, 6, 8, and 10, the smallest sample size used was n = 24. For each n and p, the percentage points of $T_1(c)$ via Monte Carlo simulation were based on 10,000 samples from $N_p(m,D)$. Since the procedures are affine invariant (Delaney, 1979), m = 0 and D = I were used in the analysis.

TABLE 5.1a. Upper Tail Percentage Points for the Statistic $T_1(c) \times 100$, $p = 1$

c	n	PERCENT 75	80	85	90	95	97.5	99
0.05	10	0.396	0.448	0.505	0.589	0.717	0.876	1.69
	20	0.505	0.588	0.693	0.838	1.19	2.17	4.43
	24	0.542	0.636	0.756	0.930	1.34	2.34	4.97
	30	0.565	0.674	0.807	1.01	1.47	2.55	5.23
	40	0.625	0.750	0.899	1.16	1.75	3.04	5.69
	60	0.684	0.822	1.00	1.28	1.91	3.20	6.38
	120	0.780	0.947	1.170	1.52	2.21	3.33	5.63
0.04	10	0.254	0.288	0.325	0.377	0.457	0.54[illegible]	0.99
	20	0.324	0.377	0.444	0.538	0.751	1.34	2.72
	24	0.349	0.409	0.485	0.596	0.844	1.44	3.05
	30	0.362	0.433	0.518	0.647	0.926	1.59	3.25
	40	0.402	0.482	0.578	0.739	1.11	1.90	3.62
	60	0.440	0.529	0.643	0.821	1.22	2.02	4.07
	120	0.505	0.613	0.755	0.979	1.42	2.16	3.61
0.03	10	0.143	0.163	0.184	0.213	0.256	0.302	0.517
	20	0.183	0.213	0.250	0.303	0.418	0.723	1.45
	24	0.197	0.231	0.273	0.335	0.471	0.780	1.64
	30	0.206	0.244	0.293	0.365	0.516	0.874	1.78
	40	0.228	0.272	0.326	0.416	0.615	1.06	1.96
	60	0.250	0.299	0.363	0.463	0.685	1.14	2.26
	120	0.287	0.347	0.429	0.555	0.805	1.23	2.05
0.025	10	0.0998	0.113	0.128	0.148	0.177	0.210	0.347
	20	0.127	0.148	0.174	0.211	0.289	0.493	0.983
	24	0.137	0.161	0.189	0.233	0.325	0.531	1.11
	30	0.143	0.170	0.203	0.253	0.360	0.596	1.22
	40	0.159	0.189	0.228	0.289	0.429	0.729	1.36
	60	0.174	0.209	0.253	0.323	0.476	0.788	1.56
	120	0.200	0.242	0.299	0.387	0.561	0.849	1.42

TABLE 5.1b. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 1$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.02	10	0.640	0.724	0.821	0.946	1.13	1.34	2.14
	20	0.815	0.948	1.11	1.35	1.85	3.10	6.16
	24	0.881	1.03	1.21	1.49	2.06	3.35	6.91
	30	0.921	1.09	1.30	1.62	2.29	3.75	7.69
	40	1.02	1.21	1.46	1.84	2.74	4.60	8.63
	60	1.12	1.34	1.62	2.07	3.05	5.01	10.0
	120	1.28	1.56	1.92	2.48	3.58	5.45	9.13
0.015	10	0.361	0.408	0.463	0.533	0.638	0.753	1.16
	20	0.460	0.535	0.626	0.761	1.04	1.71	3.39
	24	0.498	0.581	0.685	0.837	1.16	1.84	3.82
	30	0.519	0.616	0.734	0.910	1.29	2.08	4.24
	40	0.573	0.683	0.824	1.04	1.53	2.57	4.79
	60	0.633	0.754	0.916	1.16	1.71	2.81	5.63
	120	0.726	0.879	1.09	1.40	2.01	3.07	5.18
0.01	10	0.161	0.182	0.206	0.236	0.284	0.333	0.497
	20	0.204	0.238	0.278	0.338	0.460	0.746	1.48
	24	0.222	0.259	0.306	0.372	0.511	0.799	1.67
	30	0.231	0.275	0.327	0.404	0.571	0.912	1.86
	40	0.256	0.305	0.367	0.462	0.679	1.14	2.09
	60	0.282	0.337	0.406	0.519	0.760	1.24	2.48
	120	0.324	0.393	0.486	0.623	0.895	1.36	2.31
0.0075	10	0.0906	0.102	0.116	0.133	0.159	0.187	0.274
	20	0.115	0.134	0.157	0.190	0.257	0.416	0.820
	24	0.125	0.146	0.172	0.209	0.286	0.446	0.932
	30	0.130	0.155	0.184	0.228	0.321	0.509	1.04
	40	0.144	0.172	0.207	0.260	0.382	0.634	1.17
	60	0.159	0.190	0.229	0.293	0.428	0.697	1.39
	120	0.182	0.222	0.274	0.351	0.503	0.767	1.30

TABLE 5.1c. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 1$

c	n	PERCENT 75	80	85	90	95	97.5	99
-0.0075	10	0.0916	0.103	0.117	0.134	0.160	0.185	0.246
	20	0.116	0.135	0.158	0.190	0.253	0.389	0.760
	24	0.125	0.146	0.173	0.210	0.284	0.423	0.877
	30	0.131	0.156	0.185	0.229	0.316	0.490	0.998
	40	0.146	0.173	0.209	0.261	0.379	0.617	1.14
	60	0.161	0.191	0.232	0.294	0.428	0.686	1.37
	120	0.184	0.224	0.278	0.355	0.504	0.776	1.33
-0.01	10	0.163	0.184	0.209	0.238	0.284	0.329	0.430
	20	0.207	0.239	0.282	0.338	0.449	0.685	1.34
	24	0.223	0.260	0.307	0.374	0.504	0.746	1.54
	30	0.234	0.278	0.329	0.407	0.562	0.865	1.76
	40	0.260	0.308	0.371	0.463	0.673	1.09	2.01
	60	0.286	0.340	0.412	0.523	0.759	1.22	2.44
	120	0.327	0.400	0.496	0.632	0.899	1.39	2.36
-0.015	10	0.368	0.415	0.470	0.536	0.637	0.739	0.934
	20	0.467	0.538	0.633	0.763	1.01	1.51	2.94
	24	0.504	0.586	0.692	0.842	1.13	1.65	3.41
	30	0.528	0.627	0.742	0.916	1.26	1.92	3.89
	40	0.585	0.694	0.838	1.04	1.52	2.43	4.45
	60	0.646	0.766	0.928	1.18	1.71	2.72	5.46
	120	0.741	0.902	1.12	1.43	2.03	3.14	5.30
-0.02	10	0.656	0.739	0.836	0.955	1.13	1.31	1.62
	20	0.831	0.959	1.13	1.36	1.79	2.64	5.12
	24	0.898	1.04	1.23	1.50	2.00	2.89	5.95
	30	0.940	1.12	1.32	1.64	2.24	3.38	6.79
	40	1.04	1.24	1.49	1.85	2.70	4.27	7.84
	60	1.15	1.36	1.65	2.09	3.04	4.83	9.59
	120	1.33	16.1	2.00	2.55	3.63	5.56	9.43

TABLE 5.1d. Upper Tail Percentage Points for the Statistic $T_1(c) \times 100$, $p = 1$

c	n	75	80	85	90	95	97.5	99
		PERCENT						
-0.025	10	0.103	0.116	0.131	0.150	0.177	0.205	0.248
	20	0.130	0.150	0.177	0.213	0.279	0.407	0.782
	24	0.141	0.164	0.193	0.234	0.312	0.445	0.912
	30	0.148	0.175	0.207	0.255	0.348	0.522	1.04
	40	0.163	0.194	0.233	0.289	0.419	0.659	1.22
	60	0.180	0.214	0.260	0.328	0.472	0.754	1.49
	120	0.208	0.252	0.314	0.398	0.571	0.872	1.47
-0.03	10	0.148	0.167	0.189	0.216	0.256	0.296	0.351
	20	0.188	0.217	0.255	0.307	0.402	0.576	1.10
	24	0.204	0.237	0.278	0.338	0.450	0.632	1.29
	30	0.213	0.253	0.299	0.367	0.502	0.738	1.48
	40	0.236	0.280	0.337	0.416	0.601	0.939	1.74
	60	0.260	0.308	0.375	0.475	0.679	1.08	2.13
	120	0.302	0.364	0.454	0.575	0.824	1.26	2.12
-0.04	10	0.266	0.298	0.338	0.385	0.458	0.522	0.608
	20	0.336	0.388	0.455	0.548	0.711	0.987	1.86
	24	0.362	0.424	0.497	0.604	0.793	1.09	2.21
	30	0.381	0.451	0.534	0.656	0.890	1.27	2.53
	40	0.421	0.502	0.599	0.739	1.06	1.63	3.04
	60	0.466	0.552	0.670	0.843	1.21	1.89	3.75
	120	0.542	0.656	0.815	1.03	1.48	2.25	3.77
-0.05	10	0.418	0.470	0.531	0.604	0.720	0.820	0.924
	20	0.527	0.608	0.715	0.859	1.10	1.50	2.79
	24	0.570	0.666	0.782	0.943	1.24	1.65	3.32
	30	0.598	0.709	0.837	1.03	1.38	1.95	3.83
	40	0.664	0.790	0.941	1.16	1.65	2.49	4.65
	60	0.733	0.868	1.05	1.32	1.89	2.92	5.73
	120	0.854	1.04	1.28	1.62	2.31	3.55	5.87

TABLE 5.1e. Upper Tail Percentage Points for the Statistic $T_1(c)$, $p = 1$

c	n	PERCENT 75	80	85	90	95	97.5	99
0.3	10	0.133	0.152	0.181	0.233	0.745	2.11	8.16
	20	0.173	0.205	0.248	0.338	0.770	1.46	3.19
	24	0.181	0.215	0.267	0.361	0.767	1.42	2.90
	30	0.190	0.231	0.284	0.379	0.767	1.41	2.46
	40	0.200	0.243	0.304	0.402	0.746	1.32	2.20
	60	0.213	0.259	0.320	0.413	0.647	1.11	1.83
	120	0.220	0.275	0.343	0.449	0.669	0.961	1.50
0.2	10	0.0592	0.0677	0.0784	0.0954	0.177	0.476	1.33
	20	0.0782	0.0919	0.110	0.140	0.269	0.544	1.29
	24	0.0821	0.0975	0.119	0.153	0.286	0.559	1.24
	30	0.0872	0.104	0.127	0.164	0.311	0.577	1.13
	40	0.0924	0.112	0.138	0.182	0.317	0.583	1.03
	60	0.0999	0.121	0.149	0.192	0.295	0.523	0.928
	120	0.0106	0.131	0.164	0.212	0.322	0.471	0.748
0.1	10	0.0153	0.0174	0.0198	0.0234	0.0307	0.0522	0.116
	20	0.0199	0.0234	0.0279	0.0342	0.0510	0.0951	0.232
	24	0.0210	0.0247	0.0297	0.0374	0.0556	0.113	0.253
	30	0.0227	0.0269	0.0321	0.0405	0.0660	0.126	0.263
	40	0.0242	0.0292	0.0352	0.0448	0.0723	0.134	0.248
	60	0.0265	0.0317	0.0390	0.0499	0.0741	0.132	0.236
	120	0.0285	0.0355	0.0455	0.0569	0.0848	0.127	0.241
-0.1	10	0.0171	0.0193	0.0220	0.0249	0.0296	0.0336	0.0382
	20	0.0221	0.0254	0.0296	0.0351	0.0439	0.0553	0.0930
	24	0.0231	0.0269	0.0315	0.0387	0.0497	0.0635	0.112
	30	0.0253	0.0296	0.0348	0.0418	0.0542	0.0713	0.135
	40	0.0270	0.0319	0.0380	0.0469	0.0629	0.0908	0.168
	60	0.0303	0.0361	0.0434	0.0546	0.0747	0.110	0.199
	120	0.0341	0.0415	0.0516	0.0650	0.0946	0.136	0.244

TABLE 5.2a. Upper Tail Percentage Points for the Statistic
$T_1(c) \times 100$, $p = 2$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.025	10	0.283	0.305	0.332	0.368	0.440	0.553	0.760
	20	0.397	0.440	0.496	0.601	0.952	1.46	2.39
	24	0.420	0.464	0.527	0.641	0.965	1.54	2.46
	30	0.459	0.512	0.593	0.737	1.16	1.79	2.90
	40	0.495	0.558	0.655	0.800	1.25	1.89	3.01
	60	0.571	0.640	0.746	0.925	1.44	2.19	3.40
	120	0.667	0.760	0.869	1.05	1.54	2.26	3.46
0.02	10	0.181	0.196	0.212	0.235	0.278	0.344	0.470
	20	0.254	0.282	0.317	0.381	0.597	0.913	1.48
	24	0.268	0.296	0.336	0.408	0.609	0.966	1.54
	30	0.294	0.327	0.378	0.468	0.731	1.13	1.75
	40	0.317	0.357	0.419	0.510	0.790	1.20	1.90
	60	0.366	0.411	0.478	0.591	0.913	1.39	2.17
	120	0.429	0.488	0.558	0.674	0.984	1.45	2.24
0.015	10	0.102	0.110	0.120	0.132	0.155	0.188	0.255
	20	0.143	0.158	0.178	0.213	0.328	0.500	0.806
	24	0.151	0.166	0.189	0.228	0.337	0.531	0.843
	30	0.165	0.184	0.213	0.262	0.404	0.625	0.955
	40	0.178	0.201	0.235	0.286	0.440	0.672	1.06
	60	0.206	0.231	0.268	0.332	0.510	0.779	1.22
	120	0.242	0.275	0.314	0.380	0.555	0.816	1.26
0.01	10	0.0453	0.0489	0.0529	0.0585	0.0683	0.0814	0.110
	20	0.0634	0.0701	0.0790	0.0940	0.143	0.217	0.346
	24	0.0670	0.0739	0.0840	0.100	0.148	0.232	0.366
	30	0.0733	0.0817	0.0943	0.116	0.178	0.272	0.421
	40	0.0794	0.0894	0.104	0.127	0.193	0.296	0.463
	60	0.0918	0.103	0.119	0.148	0.226	0.342	0.538
	120	0.108	0.123	0.140	0.169	0.247	0.364	0.565

TABLE 5 2b. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 2$

c	n	PERCENT 75	80	85	90	95	97.5	99
0.008	10	0.290	0.314	0.339	0.374	0.436	0.518	0.696
	20	0.406	0.449	0.504	0.600	0.909	1.37	2.19
	24	0.429	0.474	0.538	0.642	0.935	1.47	2.32
	30	0.469	0.523	0.604	0.739	1.13	1.73	2.92
	40	0.509	0.572	0.665	0.808	1.23	1.89	2.94
	60	0.588	0.659	0.762	0.943	1.44	2.19	3.43
	120	0.622	0.787	0.896	1.08	1.58	2.33	3.61
0.006	10	0.163	0.176	0.191	0.211	0.245	0.289	0.387
	20	0.228	0.252	0.283	0.336	0.506	0.765	1.21
	24	0.241	0.266	0.302	0.360	0.521	0.820	1.29
	30	0.264	0.294	0.340	0.414	0.632	0.964	1.62
	40	0.287	0.322	0.374	0.453	0.689	1.06	1.65
	60	0.331	0.371	0.429	0.529	0.807	1.23	1.92
	120	0.390	0.443	0.504	0.611	0.891	1.31	2.03
0.004	10	0.0726	0.0782	0.0849	0.0936	0.109	0.127	0.170
	20	0.101	0.112	0.126	0.149	0.223	0.337	0.533
	24	0.107	0.118	0.134	0.160	0.230	0.361	0.568
	30	0.118	0.131	0.151	0.184	0.279	0.425	0.714
	40	0.127	0.143	0.166	0.201	0.305	0.468	0.727
	60	0.147	0.165	0.191	0.234	0.358	0.542	0.851
	120	0.173	0.197	0.225	0.272	0.396	0.582	0.900
0.002	10	0.0181	0.0195	0.0212	0.0234	0.0272	0.0315	0.0418
	20	0.0253	0.0280	0.0313	0.0371	0.0552	0.0834	0.132
	24	0.0268	0.0296	0.0335	0.0399	0.0572	0.0895	0.141
	30	0.0294	0.0327	0.0376	0.0458	0.0692	0.106	0.177
	40	0.0318	0.0358	0.0416	0.0501	0.0758	0.116	0.181
	60	0.0368	0.0412	0.0477	0.0585	0.0894	0.135	0.212
	120	0.0434	0.0493	0.0561	0.0682	0.0991	0.146	0.225

TABLE 5.2c. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 2$

c	n	PERCENT 75	80	85	90	95	97.5	99
-0.002	10	0.0182	0.0196	0.0212	0.0233	0.0271	0.0310	0.0406
	20	0.0253	0.0279	0.0313	0.0369	0.0543	0.0817	0.128
	24	0.0268	0.0296	0.0335	0.0398	0.0562	0.0882	0.138
	30	0.0294	0.0327	0.0376	0.0456	0.0685	0.104	0.173
	40	0.0319	0.0359	0.0416	0.0501	0.0754	0.115	0.179
	60	0.0369	0.0411	0.0476	0.0585	0.0892	0.135	0.210
	120	0.0435	0.0495	0.0562	0.0685	0.0992	0.146	0.227
-0.004	10	0.0726	0.0784	0.0849	0.0932	0.108	0.124	0.160
	20	0.101	0.112	0.125	0.147	0.216	0.324	0.508
	24	0.107	0.118	0.134	0.159	0.224	0.349	0.549
	30	0.118	0.131	0.150	0.182	0.273	0.416	0.686
	40	0.128	0.143	0.166	0.200	0.300	0.459	0.711
	60	0.148	0.165	0.190	0.234	0.356	0.536	0.839
	120	0.174	0.198	0.225	0.274	0.397	0.586	0.911
-0.006	10	0.164	0.176	0.191	0.210	0.243	0.277	0.355
	20	0.228	0.251	0.281	0.331	0.483	0.720	1.13
	24	0.241	0.266	0.301	0.356	0.501	0.779	1.23
	30	0.265	0.294	0.338	0.408	0.610	0.930	1.53
	40	0.288	0.322	0.374	0.450	0.674	1.02	1.59
	60	0.332	0.371	0.428	0.526	0.801	1.20	1.88
	120	0.393	0.446	0.507	0.617	0.895	1.32	2.06
-0.008	10	0.291	0.314	0.340	0.373	0.432	0.492	0.621
	20	0.405	0.446	0.500	0.588	0.853	1.27	1.98
	24	0.428	0.472	0.535	0.632	0.888	1.37	2.16
	30	0.471	0.522	0.600	0.724	1.08	1.65	2.69
	40	0.512	0.574	0.664	0.799	1.19	1.82	2.80
	60	0.591	0.660	0.760	0.934	1.42	2.13	3.34
	120	0.699	0.794	0.901	1.10	1.60	2.34	3.67

TABLE 5.2d. Upper Tail Percentage Points for the Statistic $T_1(c) \times 100$, $p = 2$

c	n	75	80	85	90	95	97.5	99
		PERCENT						
-0.01	10	0.0455	0.0491	0.0531	0.0583	0.0674	0.0766	0.0959
	20	0.0633	0.0698	0.0782	0.0916	0.132	0.196	0.307
	24	0.0669	0.0738	0.0834	0.0988	0.138	0.212	0.336
	30	0.0737	0.0816	0.0938	0.113	0.167	0.256	0.415
	40	0.0800	0.0897	0.104	0.125	0.186	0.283	0.434
	60	0.0924	0.103	0.119	0.146	0.222	0.331	0.521
	120	0.109	0.124	0.141	0.172	0.250	0.366	0.575
-0.015	10	0.102	0.110	0.120	0.131	0.151	0.170	0.208
	20	0.142	0.157	0.175	0.205	0.291	0.430	0.672
	24	0.151	0.166	0.188	0.221	0.306	0.467	0.732
	30	0.166	0.184	0.211	0.254	0.371	0.568	0.911
	40	0.180	0.202	0.233	0.280	0.414	0.633	0.961
	60	0.208	0.233	0.267	0.328	0.498	0.741	1.17
	120	0.247	0.280	0.318	0.387	0.561	0.822	1.30
-0.02	10	0.183	0.196	0.213	0.233	0.267	0.301	0.359
	20	0.253	0.278	0.310	0.364	0.508	0.745	1.17
	24	0.268	0.296	0.334	0.390	0.535	0.809	1.27
	30	0.295	0.327	0.375	0.450	0.649	0.991	1.58
	40	0.321	0.360	0.412	0.496	0.728	1.11	1.68
	60	0.370	0.414	0.476	0.583	0.879	1.31	2.06
	120	0.440	0.499	0.568	0.690	1.00	1.46	2.30
-0.025	10	0.286	0.308	0.334	0.364	0.415	0.466	0.551
	20	0.397	0.435	0.485	0.567	0.779	1.14	1.77
	24	0.418	0.463	0.521	0.605	0.825	1.24	1.95
	30	0.461	0.510	0.585	0.701	1.00	1.52	2.42
	40	0.502	0.561	0.642	0.772	1.13	1.70	2.58
	60	0.578	0.647	0.744	0.911	1.37	2.02	3.19
	120	0.691	0.784	0.890	1.08	1.56	2.28	3.59

TABLE 5.2e. Upper Tail Percentage Points for the Statistic $T_1(c)$, p = 2

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.3	10	0.546	0.875	1.72	5.65	14.4	31.5	68.0
	20	0.664	0.845	1.20	1.84	3.51	5.82	10.7
	24	0.687	0.854	1.15	1.68	2.91	4.75	8.97
	30	0.716	0.857	1.14	1.64	2.75	4.16	6.91
	40	0.740	0.870	1.10	1.56	2.40	3.45	5.22
	60	0.776	0.895	1.10	1.43	2.10	2.95	4.14
	120	0.792	0.901	1.05	1.31	1.82	2.45	3.19
0.2	10	0.194	0.223	0.291	0.505	1.53	1.14	2.26
	20	0.267	0.316	0.409	0.619	1.24	2.06	4.30
	24	0.287	0.334	0.422	0.628	1.14	1.92	3.42
	30	0.305	0.356	0.446	0.650	1.13	1.85	3.08
	40	0.328	0.385	0.469	0.642	1.07	1.57	2.54
	60	0.352	0.404	0.484	0.643	0.970	1.38	1.98
	120	0.374	0.427	0.498	0.610	0.869	1.19	1.69
0.1	10	0.0457	0.0499	0.0555	0.0650	0.106	0.170	0.273
	20	0.0628	0.0705	0.0827	0.110	0.203	0.331	0.655
	24	0.0686	0.0775	0.0901	0.121	0.207	0.345	0.603
	30	0.0737	0.0848	0.101	0.136	0.230	0.362	0.635
	40	0.0820	0.0932	0.111	0.145	0.238	0.355	0.607
	60	0.0908	0.103	0.122	0.157	0.239	0.350	0.512
	120	0.0998	0.113	0.132	0.163	0.229	0.321	0.500
-0.1	10	0.0480	0.0516	0.0554	0.0604	0.0676	0.0738	0.0808
	20	0.0643	0.0700	0.0770	0.0865	0.106	0.134	0.192
	24	0.0689	0.0755	0.0834	0.0953	0.119	0.159	0.226
	30	0.0757	0.0837	0.0932	0.108	0.137	0.185	0.286
	40	0.0844	0.0929	0.105	0.124	0.162	0.224	0.339
	60	0.0959	0.108	0.122	0.146	0.197	0.278	0.414
	120	0.114	0.128	0.147	0.177	0.243	0.350	0.559

TABLE 5.3a. Upper Tail Percentage Points the Statistic $T_1(c) \times 1000$, p = 3

c	n	75	80	85	90	95	97.5	99
		PERCENT						
0.025	10	0.560	0.591	0.628	0.681	0.785	0.899	1.06
	20	0.831	0.908	1.04	1.26	1.87	2.54	3.75
	24	0.910	1.00	1.15	1.42	2.13	2.97	4.36
	30	1.01	1.12	1.29	1.62	2.40	3.40	4.96
	40	1.11	1.24	1.45	1.80	2.67	3.67	5.58
	60	1.26	1.41	1.65	2.03	2.89	4.09	6.29
	120	1.44	1.61	1.85	2.22	2.96	3.95	5.78
0.02	10	0.358	0.377	0.401	0.433	0.496	0.560	0.657
	20	0.531	0.579	0.656	0.799	1.16	1.57	2.31
	24	0.579	0.640	0.725	0.900	1.33	1.85	2.70
	30	0.642	0.716	0.819	1.02	1.50	2.13	3.08
	40	0.710	0.795	0.916	1.14	1.68	2.32	3.51
	60	0.802	0.900	1.05	1.29	1.84	2.61	3.99
	120	0.925	1.03	1.18	1.42	1.90	2.54	3.73
0.015	10	0.201	0.212	0.225	0.242	0.276	0.308	0.360
	20	0.298	0.324	0.365	0.442	0.640	0.858	1.25
	24	0.325	0.359	0.405	0.499	0.730	1.01	1.47
	30	0.361	0.400	0.458	0.569	0.830	1.17	1.69
	40	0.399	0.446	0.511	0.635	0.932	1.28	1.92
	60	0.451	0.506	0.590	0.723	1.03	1.46	2.23
	120	0.522	0.584	0.667	0.800	1.07	1.44	2.10
0.01	10	0.0892	0.0940	0.0997	0.107	0.121	0.134	0.155
	20	0.132	0.143	0.161	0.194	0.277	0.370	0.533
	24	0.144	0.158	0.178	0.218	0.318	0.435	0.634
	30	0.160	0.177	0.202	0.250	0.362	0.508	0.726
	40	0.177	0.197	0.226	0.279	0.411	0.564	0.839
	60	0.232	0.260	0.296	0.356	0.454	0.643	0.984
	120	0.232	0.260	0.296	0.356	0.476	0.639	0.939

TABLE 5.3b. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 3$

c	n	75	80	85	90	95	97.5	99
		PERCENT						
0.008	10	0.571	0.602	0.637	0.684	0.771	0.851	0.982
	20	0.843	0.915	1.02	1.23	1.76	2.33	3.36
	24	0.922	1.01	1.14	1.39	2.02	2.75	4.01
	30	1.02	1.13	1.29	1.60	2.30	3.22	4.60
	40	1.13	1.26	1.44	1.78	2.60	3.59	5.33
	60	1.29	1.43	1.67	2.04	2.89	4.10	6.28
	120	1.49	1.67	1.90	2.29	3.06	4.10	6.02
0.006	10	0.321	0.339	0.358	0.385	0.432	0.475	0.548
	20	0.474	0.514	0.575	0.689	0.979	1.30	1.86
	24	0.517	0.568	0.637	0.778	1.12	1.53	2.22
	30	0.574	0.634	0.721	0.893	1.29	1.80	2.55
	40	0.637	0.708	0.810	0.994	1.46	2.01	2.98
	60	0.723	0.805	0.936	1.14	1.62	2.30	3.52
	120	0.837	0.938	1.07	1.28	1.72	2.31	3.39
0.004	10	0.143	0.151	0.159	0.171	0.191	0.210	0.241
	20	0.211	0.228	0.255	0.304	0.431	0.571	0.814
	24	0.229	0.252	0.282	0.343	0.496	0.672	0.977
	30	0.255	0.281	0.319	0.395	0.568	0.792	1.12
	40	0.282	0.315	0.359	0.440	0.644	0.890	1.31
	60	0.321	0.357	0.415	0.506	0.719	1.02	1.56
	120	0.372	0.417	0.475	0.571	0.764	1.03	1.51
0.002	10	0.0357	0.0376	0.0398	0.0427	0.0478	0.0521	0.0598
	20	0.0526	0.0568	0.0635	0.0756	0.107	0.141	0.201
	24	0.0572	0.0629	0.0703	0.0853	0.123	0.166	0.241
	30	0.0636	0.0702	0.0797	0.0983	0.141	0.196	0.276
	40	0.0706	0.0795	0.0894	0.110	0.160	0.221	0.325
	60	0.0802	0.0893	0.104	0.126	0.179	0.254	0.389
	120	0.093	0.104	0.119	0.143	0.191	0.257	0.378

TABLE 5.3c. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 3$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
-0.002	10	0.0358	0.0376	0.0398	0.0426	0.0476	0.0514	0.0585
	20	0.0525	0.0567	0.0632	0.0748	0.105	0.138	0.196
	24	0.0571	0.0626	0.0699	0.0846	0.121	0.163	0.235
	30	0.0633	0.0699	0.0793	0.0972	0.139	0.193	0.271
	40	0.0704	0.0782	0.0890	0.109	0.158	0.217	0.320
	60	0.0802	0.0892	0.103	0.126	0.178	0.253	0.387
	120	0.0934	0.105	0.119	0.143	0.191	0.258	0.380
-0.004	10	0.143	0.150	0.159	0.170	0.190	0.204	0.231
	20	0.210	0.227	0.252	0.298	0.416	0.546	0.771
	24	0.228	0.250	0.279	0.337	0.478	0.646	0.931
	30	0.253	0.279	0.316	0.387	0.535	0.764	1.08
	40	0.281	0.312	0.355	0.433	0.629	0.861	1.27
	60	0.320	0.357	0.413	0.501	0.710	1.01	1.54
	120	0.374	0.419	0.476	0.572	0.766	1.03	1.52
-0.006	10	0.322	0.338	0.359	0.384	0.426	0.460	0.516
	20	0.472	0.510	0.566	0.664	0.930	1.21	1.71
	24	0.513	0.561	0.625	0.754	1.07	1.44	2.07
	30	0.569	0.627	0.709	0.868	1.24	1.70	2.40
	40	0.633	0.701	0.798	0.971	1.41	1.92	2.83
	60	0.721	0.801	0.927	1.12	1.59	2.25	3.44
	120	0.843	0.943	1.07	1.29	1.72	2.33	3.41
-0.008	10	0.572	0.601	0.636	0.681	0.755	0.814	0.914
	20	0.838	0.905	1.00	1.18	1.64	2.13	3.00
	24	0.912	0.995	1.11	1.33	1.89	2.53	3.64
	30	1.01	1.11	1.26	1.54	2.18	3.01	4.23
	40	1.12	1.24	1.41	1.72	2.49	3.39	4.98
	60	1.28	1.42	1.65	1.99	2.82	3.99	6.08
	120	1.50	1.68	1.90	2.29	3.06	4.14	6.06

TABLE 5.3d. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 3$

c	n	75	80	85	90	95	97.5	99
		PERCENT						
-0.01	10	0.0894	0.0939	0.0994	0.106	0.118	0.127	0.142
	20	0.130	0.141	0.156	0.182	0.254	0.329	0.462
	24	0.142	0.155	0.173	0.206	0.293	0.392	0.562
	30	0.158	0.174	0.196	0.239	0.338	0.465	0.655
	40	0.176	0.194	0.220	0.267	0.387	0.526	0.772
	60	0.200	0.222	0.257	0.311	0.439	0.622	0.947
	120	0.235	0.262	0.297	0.357	0.478	0.648	0.948
-0.015	10	0.201	0.211	0.224	0.239	0.264	0.283	0.315
	20	0.294	0.316	0.349	0.405	0.559	0.718	1.00
	24	0.320	0.348	0.385	0.459	0.647	0.861	1.22
	30	0.354	0.389	0.438	0.531	0.745	1.03	1.43
	40	0.394	0.436	0.494	0.596	0.859	1.16	1.70
	60	0.451	0.501	0.576	0.699	0.980	1.39	2.11
	120	0.530	0.591	0.670	0.806	1.07	1.46	2.14
-0.02	10	0.358	0.375	0.399	0.425	0.467	0.501	0.552
	20	0.523	0.561	0.617	0.717	0.971	1.25	1.72
	24	0.569	0.618	0.680	0.809	1.13	1.50	2.12
	30	0.628	0.689	0.774	0.931	1.30	1.78	2.48
	40	0.701	0.772	0.873	1.05	1.51	2.04	2.94
	60	0.802	0.890	1.02	1.24	1.73	2.44	3.69
	120	0.944	1.05	1.19	1.43	1.91	2.60	3.80
-0.025	10	0.561	0.588	0.623	0.663	0.729	0.777	0.851
	20	0.815	0.874	0.960	1.11	1.48	1.89	2.61
	24	0.886	0.964	1.06	1.25	1.73	2.28	3.21
	30	0.978	1.07	1.20	1.44	2.00	2.73	3.79
	40	1.09	1.21	1.36	1.63	2.32	3.13	4.51
	60	1.25	1.39	1.59	1.92	2.67	3.77	5.68
	120	1.48	1.65	1.87	2.24	3.00	4.07	5.94

TABLE 5.3e. Upper Tail Percentage Points for the Statistic
$T_1(c) \times 1000$, $p = 3$

c	n	PERCENT 75	80	85	90	95	97.5	99
0.3	10	9.53	14.2	21.7	42.1	107.0	267.0	793.0
	20	2.46	3.29	4.51	6.73	12.3	19.9	35.4
	24	2.24	2.87	3.76	5.35	8.57	13.8	22.2
	30	2.09	2.60	3.29	4.42	6.99	10.5	16.0
	40	1.93	2.29	2.79	3.32	5.13	7.19	10.8
	60	1.83	2.12	2.50	3.14	4.32	5.75	7.80
	120	1.72	1.92	2.19	2.60	3.49	4.43	5.70
0.2	10	0.600	0.814	1.35	6.95	23.1	38.0	82.0
	20	0.763	0.972	1.27	1.88	3.50	6.10	11.0
	24	0.787	0.973	1.25	1.80	3.03	4.97	8.62
	30	0.783	0.935	1.20	1.65	2.66	4.04	6.36
	40	0.792	0.934	1.14	1.46	2.15	3.10	4.52
	60	0.803	0.933	1.11	1.38	1.95	2.63	3.64
	120	0.811	0.905	1.04	1.24	1.66	2.07	2.75
0.1	10	0.0959	0.104	0.117	0.144	0.207	0.280	0.407
	20	0.145	0.168	0.206	0.277	0.454	0.748	1.22
	24	0.158	0.181	0.222	0.300	0.490	0.785	1.38
	30	0.169	0.194	0.235	0.314	0.486	0.727	1.25
	40	0.182	0.208	0.250	0.314	0.472	0.682	1.03
	60	0.197	0.225	0.268	0.336	0.485	0.665	0.914
	120	0.216	0.241	0.274	0.329	0.438	0.560	0.767
-0.1	10	0.0940	0.0991	0.104	0.111	0.120	0.129	0.138
	20	0.130	0.138	0.149	0.164	0.195	0.229	0.287
	24	0.141	0.151	0.164	0.183	0.221	0.280	0.370
	30	0.157	0.169	0.185	0.208	0.258	0.320	0.447
	40	0.173	0.189	0.209	0.240	0.303	0.390	0.510
	60	0.200	0.221	0.248	0.293	0.388	0.493	0.644
	120	0.243	0.268	0.302	0.354	0.477	0.633	0.839

TABLE 5.4a. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, p = 4

c	n	75	80	85	90	95	97.5	99
		PERCENT						
0.01	10	0.148	0.154	0.160	0.169	0.181	0.195	0.213
	20	0.237	0.257	0.287	0.345	0.461	0.603	0.817
	24	0.260	0.285	0.324	0.397	0.541	0.707	0.977
	30	0.291	0.322	0.367	0.443	0.599	0.784	1.10
	40	0.325	0.361	0.414	0.503	0.704	0.929	1.29
	60	0.372	0.412	0.469	0.555	0.780	1.03	1.42
	120	0.425	0.467	0.531	0.630	0.815	1.05	1.45
0.008	10	0.0945	0.0982	0.103	0.108	0.116	0.124	0.135
	20	0.151	0.164	0.182	0.219	0.291	0.380	0.514
	24	0.166	0.182	0.206	0.252	0.342	0.447	0.614
	30	0.186	0.206	0.234	0.282	0.380	0.496	0.697
	40	0.208	0.230	0.264	0.320	0.448	0.589	0.818
	60	0.238	0.263	0.300	0.354	0.498	0.659	0.908
	120	0.272	0.299	0.340	0.404	0.521	0.675	0.930
0.006	10	0.0531	0.0552	0.0576	0.0607	0.0649	0.0695	0.0757
	20	0.0848	0.0916	0.102	0.122	0.162	0.211	0.284
	24	0.0932	0.102	0.115	0.140	0.190	0.248	0.340
	30	0.104	0.115	0.131	0.157	0.212	0.276	0.387
	40	0.117	0.129	0.148	0.179	0.250	0.329	0.456
	60	0.134	0.148	0.168	0.199	0.279	0.369	0.509
	120	0.153	0.168	0.191	0.227	0.294	0.380	0.524
0.005	10	0.0369	0.0383	0.0400	0.0421	0.0450	0.0481	0.0523
	20	0.0588	0.0635	0.0706	0.0847	0.122	0.146	0.195
	24	0.0646	0.0705	0.0797	0.0969	0.131	0.172	0.234
	30	0.0724	0.0798	0.0906	0.109	0.147	0.191	0.267
	40	0.0810	0.0896	0.102	0.124	0.173	0.228	0.315
	60	0.929	0.102	0.117	0.138	0.194	0.255	0.353
	120	0.106	0.117	0.132	0.158	0.204	0.264	0.364

TABLE 5.4b. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 4$

c	n	PERCENT 75	80	85	90	95	97.5	99
0.004	10	0.236	0.245	0.256	0.270	0.288	0.307	0.334
	20	0.376	0.406	0.450	0.539	0.712	0.927	1.24
	24	0.413	0.450	0.509	0.618	0.836	1.09	1.49
	30	0.463	0.510	0.579	0.696	0.935	1.21	1.70
	40	0.518	0.573	0.655	0.794	1.10	1.45	2.01
	60	0.594	0.655	0.745	0.881	1.24	1.63	2.25
	120	0.681	0.747	0.848	1.01	1.31	1.70	2.33
0.003	10	0.133	0.138	0.144	0.152	0.162	0.172	0.188
	20	0.211	0.228	0.253	0.302	0.399	0.518	0.690
	24	0.232	0.253	0.285	0.346	0.468	0.609	0.830
	30	0.260	0.286	0.325	0.390	0.524	0.679	0.950
	40	0.291	0.322	0.367	0.446	0.618	0.814	1.12
	60	0.334	0.368	0.419	0.494	0.695	0.914	1.26
	120	0.383	0.420	0.477	0.569	0.733	0.954	1.31
0.002	10	0.0589	0.0613	0.0639	0.0673	0.0719	0.0765	0.0830
	20	0.0937	0.101	0.112	0.134	0.176	0.229	0.304
	24	0.103	0.112	0.126	0.153	0.207	0.269	0.366
	30	0.115	0.127	0.144	0.173	0.232	0.300	0.420
	40	0.129	0.143	0.163	0.198	0.274	0.360	0.497
	60	0.148	0.163	0.186	0.219	0.309	0.405	0.561
	120	0.170	0.187	0.212	0.252	0.326	0.424	0.581
0.001	10	0.0147	0.0153	0.0160	0.0168	0.0179	0.0191	0.0207
	20	0.0234	0.0253	0.0279	0.0333	0.0438	0.0568	0.0752
	24	0.0257	0.0280	0.0315	0.0382	0.0514	0.0669	0.0908
	30	0.0288	0.0317	0.0359	0.0430	0.0577	0.0747	0.104
	40	0.0322	0.0356	0.0407	0.0493	0.0682	0.0898	0.124
	60	0.0371	0.0408	0.0464	0.0548	0.0771	0.101	0.140
	120	0.0425	0.0467	0.0530	0.0632	0.0816	0.106	0.145

TABLE 5.4c. Upper Tail Percentage Points for the Statistic
$T_1(c) \times 1000$, $p = 4$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
-0.001	10	0.0147	0.0153	0.0160	0.0168	0.0179	0.0190	0.0205
	20	0.0233	0.0252	0.0280	0.0330	0.0432	0.0560	0.0739
	24	0.0256	0.0279	0.0314	0.0379	0.0508	0.0660	0.0894
	30	0.0288	0.0316	0.0357	0.0428	0.0572	0.0740	0.103
	40	0.0322	0.0356	0.0406	0.0490	0.0677	0.0891	0.122
	60	0.0370	0.0407	0.0462	0.0545	0.0768	0.100	0.139
	120	0.0426	0.0467	0.0530	0.0631	0.0817	0.106	0.145
-0.002	10	0.0589	0.0613	0.0638	0.0672	0.0716	0.0760	0.0818
	20	0.0933	0.101	0.111	0.131	0.172	0.222	0.294
	24	0.102	0.111	0.125	0.151	0.202	0.262	0.355
	30	0.115	0.126	0.142	0.171	0.228	0.294	0.409
	40	0.128	0.142	0.162	0.196	0.270	0.355	0.488
	60	0.148	0.163	0.185	0.218	0.306	0.401	0.556
	120	0.170	0.187	0.212	0.252	0.327	0.425	0.582
-0.003	10	0.133	0.138	0.144	0.151	0.161	0.171	0.184
	20	0.210	0.226	0.249	0.294	0.385	0.497	0.656
	24	0.230	0.250	0.280	0.339	0.453	0.587	0.792
	30	0.258	0.283	0.319	0.382	0.511	0.659	0.916
	40	0.289	0.319	0.364	0.438	0.606	0.795	1.09
	60	0.333	0.366	0.415	0.490	0.689	0.899	1.25
	120	0.383	0.421	0.477	0.568	0.737	0.956	1.31
-0.004	10	0.236	0.245	0.255	0.269	0.286	0.303	0.326
	20	0.372	0.401	0.442	0.521	0.682	0.879	1.16
	24	0.408	0.444	0.497	0.600	0.801	1.04	1.40
	30	0.458	0.502	0.567	0.678	0.905	1.16	1.62
	40	0.513	0.566	0.650	0.777	1.07	1.41	1.93
	60	0.592	0.650	0.737	0.869	1.22	1.59	2.22
	120	0.682	0.748	0.848	1.01	1.31	1.70	2.33

TABLE 5.4d. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 4$

c	n	PERCENT 75	80	85	90	95	97.5	99
-0.005	10	0.0369	0.0383	0.0399	0.0419	0.0447	0.0473	0.0509
	20	0.0580	0.0625	0.0689	0.0812	0.106	0.137	0.180
	24	0.0638	0.0692	0.0775	0.0935	0.125	0.161	0.217
	30	0.0715	0.0783	0.0833	0.106	0.141	0.181	0.252
	40	0.0801	0.0883	0.101	0.121	0.167	0.219	0.300
	60	0.0924	0.101	0.115	0.135	0.191	0.248	0.345
	120	0.107	0.117	0.132	0.158	0.204	0.265	0.363
-0.006	10	0.0531	0.0552	0.0574	0.0604	0.0643	0.0680	0.0731
	20	0.0835	0.0899	0.0989	0.116	0.152	0.196	0.257
	24	0.0917	0.0996	0.111	0.134	0.178	0.230	0.310
	30	0.103	0.113	0.127	0.152	0.202	0.259	0.360
	40	0.115	0.127	0.145	0.174	0.239	0.314	0.430
	60	0.133	0.146	0.166	0.195	0.274	0.357	0.496
	120	0.153	0.168	0.191	0.227	0.294	0.382	0.522
-0.008	10	0.0944	0.0980	0.102	0.107	0.114	0.121	0.129
	20	0.148	0.160	0.175	0.206	0.267	0.343	0.450
	24	0.163	0.176	0.197	0.237	0.315	0.404	0.542
	30	0.182	0.199	0.224	0.268	0.356	0.456	0.633
	40	0.204	0.225	0.256	0.308	0.423	0.553	0.758
	60	0.236	0.259	0.294	0.345	0.486	0.633	0.878
	120	0.237	0.299	0.340	0.403	0.524	0.680	0.928
-0.01	10	0.148	0.153	0.160	0.168	0.178	0.188	0.201
	20	0.230	0.248	0.272	0.319	0.414	0.530	0.694
	24	0.254	0.275	0.307	0.368	0.487	0.623	0.834
	30	0.284	0.311	0.349	0.416	0.551	0.706	0.978
	40	0.318	0.351	0.399	0.478	0.656	0.857	1.17
	60	0.368	0.404	0.458	0.539	0.757	0.984	1.37
	120	0.427	0.468	0.530	0.629	0.818	1.06	1.45

TABLE 5.4e. Upper Tail Percentage Points for the Statistic $T_1(c)$, $p = 4$

PERCENT

c	n	75	80	85	90	95	97.5	99
0.3	24	6.43	8.05	10.5	14.8	25.0	40.1	71.4
	30	5.27	6.36	7.95	10.4	15.7	23.1	36.5
	40	4.54	5.30	6.32	7.92	11.0	14.5	21.1
	60	3.86	4.38	5.09	6.17	8.21	10.5	13.9
	120	3.53	3.90	4.35	5.07	6.37	7.64	9.65
0.2	24	1.91	2.37	2.97	4.10	6.88	10.7	17.8
	30	1.84	2.19	2.72	3.67	5.50	7.92	12.0
	40	1.79	2.09	2.48	3.16	4.51	6.01	8.38
	60	1.67	1.89	2.19	2.68	3.66	4.81	6.44
	120	1.61	1.79	2.00	2.36	3.00	3.64	4.58
0.1	24	0.325	0.379	0.454	0.583	0.915	1.37	2.29
	30	0.349	0.404	0.485	0.639	0.954	1.44	2.19
	40	0.378	0.431	0.511	0.644	0.951	1.36	2.04
	60	0.396	0.441	0.512	0.626	0.870	1.19	1.72
	120	0.419	0.463	0.524	0.613	0.796	1.02	1.32
-0.1	24	0.241	0.255	0.273	0.300	0.344	0.409	0.506
	30	0.268	0.285	0.307	0.343	0.416	0.506	0.633
	40	0.307	0.330	0.358	0.404	0.505	0.636	0.824
	60	0.356	0.384	0.424	0.478	0.603	0.763	1.04
	120	0.441	0.478	0.526	0.613	0.795	1.02	1.37

TABLE 5.5a. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 5$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.004	24	0.664	0.723	0.806	0.937	1.22	1.50	1.88
	30	0.750	0.820	0.920	1.08	1.39	1.73	2.27
	40	0.851	0.937	1.05	1.23	1.64	2.07	2.67
	60	0.970	1.06	1.19	1.41	1.83	2.33	3.10
	120	1.14	1.24	1.38	1.61	2.06	2.54	3.23
0.0035	24	0.508	0.553	0.616	0.715	0.931	1.14	1.44
	30	0.573	0.627	0.703	0.829	1.06	1.32	1.73
	40	0.651	0.717	0.806	0.941	1.25	1.58	2.04
	60	0.742	0.810	0.912	1.08	1.40	1.78	2.37
	120	0.870	0.947	1.06	1.23	1.57	1.95	2.47
0.003	24	0.373	0.406	0.452	0.524	0.682	0.838	1.05
	30	0.421	0.460	0.516	0.608	0.778	0.966	1.27
	40	0.478	0.526	0.592	0.691	0.917	1.16	1.49
	60	0.545	0.595	0.669	0.722	1.02	1.30	1.74
	120	0.639	0.696	0.778	0.905	1.16	1.43	1.81
0.0025	24	0.259	0.281	0.313	0.363	0.472	0.580	0.726
	30	0.292	0.319	0.358	0.421	0.539	0.669	0.876
	40	0.331	0.365	0.410	0.479	0.635	0.884	1.03
	60	0.378	0.413	0.464	0.550	0.710	0.905	1.20
	120	0.444	0.483	0.541	0.628	0.802	0.993	1.26

TABLE 5.5b. Upper Tail Percentage Points for the Statistic $T_1(c) \times 10{,}000$, p = 5

c	n	75	80	85	90	95	97.5	99
		PERCENT						
0.002	24	1.66	1.80	2.00	2.32	3.01	3.70	4.63
	30	1.86	2.04	2.29	2.69	3.44	4.27	5.59
	40	2.12	2.33	2.62	3.06	4.06	5.14	6.59
	60	2.42	2.64	2.97	3.51	4.54	5.78	7.68
	120	2.84	3.09	3.46	4.02	5.13	6.36	8.05
0.0015	24	0.930	1.01	1.12	1.30	1.69	2.07	2.59
	30	1.05	1.14	1.28	1.51	1.93	2.39	3.14
	40	1.19	1.31	1.47	1.72	2.28	2.89	3.70
	60	1.36	1.48	1.67	1.97	2.55	3.25	4.31
	120	1.60	1.74	1.95	2.26	2.89	3.58	4.53
0.001	24	0.413	0.448	0.498	0.577	0.748	0.917	1.15
	30	0.465	0.508	0.570	0.670	0.856	1.06	1.39
	40	0.529	0.582	0.653	0.763	1.01	1.28	1.64
	60	0.605	0.659	0.741	0.876	1.13	1.44	1.91
	120	0.709	0.774	0.864	1.01	1.28	1.59	2.01
0.0005	24	0.103	0.112	0.124	0.144	0.187	0.228	0.286
	30	0.116	0.127	0.142	0.167	0.213	0.269	0.346
	40	0.132	0.145	0.163	0.190	0.252	0.319	0.409
	60	0.151	0.165	0.185	0.219	0.282	0.360	0.478
	120	0.177	0.194	0.216	0.251	0.321	0.397	0.503

TABLE 5.5c. Upper Tail Percentage Points for the Statistic $T_1(c) \times 10{,}000$, p = 5

c	n	75	80	85	90	95	97.5	99
		PERCENT						
-0.0005	24	0.103	0.112	0.124	0.143	0.186	0.227	0.283
	30	0.116	0.127	0.142	0.166	0.212	0.263	0.344
	40	0.132	0.145	0.163	0.190	0.251	0.317	0.406
	60	0.151	0.164	0.185	0.218	0.282	0.359	0.476
	120	0.177	0.194	0.216	0.251	0.320	0.396	0.502
-0.001	24	0.411	0.446	0.496	0.572	0.740	0.903	1.13
	30	0.464	0.506	0.567	0.664	0.846	1.05	1.37
	40	0.527	0.580	0.650	0.759	1.00	1.27	1.62
	60	0.603	0.657	0.738	0.872	1.13	1.43	1.90
	120	0.709	0.774	0.863	1.01	1.28	1.59	2.01
-0.0015	24	0.924	1.00	1.11	1.28	1.66	2.03	2.52
	30	1.04	1.14	1.27	1.49	1.90	2.35	3.07
	40	1.18	1.30	1.46	1.70	2.25	2.84	3.64
	60	1.36	1.48	1.66	1.96	2.53	3.22	4.27
	120	1.60	1.74	1.94	2.26	2.88	3.57	4.53
-0.002	24	1.64	1.78	1.98	2.28	2.94	3.59	4.47
	30	1.85	2.02	2.26	2.65	3.37	4.17	5.43
	40	2.10	2.31	2.59	3.03	3.98	5.04	6.44
	60	2.41	2.62	2.95	3.48	4.49	5.71	7.57
	120	2.84	3.10	3.45	4.01	5.11	6.35	8.06

TABLE 5.5d. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 5$

c	n	75	80	85	90	95	97.5	99
		PERCENT						
-0.0025	24	0.256	0.277	0.308	0.356	0.458	0.559	0.695
	30	0.289	0.315	0.353	0.412	0.526	0.650	0.846
	40	0.328	0.361	0.404	0.472	0.621	0.786	1.00
	60	0.376	0.410	0.460	0.544	0.701	0.891	1.18
	120	0.443	0.484	0.539	0.627	0.799	0.992	1.26
-0.003	24	0.368	0.399	0.443	0.511	0.658	0.802	0.998
	30	0.416	0.452	0.507	0.593	0.755	0.932	1.21
	40	0.472	0.519	0.582	0.679	0.892	1.13	1.44
	60	0.541	0.590	0.662	0.782	1.01	1.28	1.70
	120	0.638	0.697	0.776	0.903	1.15	1.43	1.81
-0.0035	24	0.500	0.542	0.602	0.694	0.893	1.09	1.35
	30	0.566	0.615	0.689	0.805	1.03	1.27	1.65
	40	0.643	0.705	0.790	0.923	1.21	1.53	1.96
	60	0.736	0.802	0.901	1.06	1.37	1.74	2.31
	120	0.869	0.948	1.06	1.23	1.57	1.94	2.46
-0.004	24	0.653	0.707	0.785	0.905	1.16	1.42	1.76
	30	0.738	0.802	0.899	1.05	1.34	1.65	2.14
	40	0.838	0.921	1.03	1.20	1.58	2.00	2.55
	60	0.961	1.05	1.17	1.39	1.79	2.27	3.01
	120	1.13	1.24	1.38	1.60	2.05	2.54	3.21

TABLE 5.5e. Upper Tail Percentage Points for the Statistic $T_1(c)$, $p = 5$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.3	24	18.3	23.2	31.6	47.7	89.8	167.0	317.0
	30	12.2	14.8	18.3	24.3	40.3	64.1	120.0
	40	9.03	10.5	12.4	15.4	21.7	28.7	40.7
	60	7.24	8.16	9.40	11.2	14.7	18.4	24.0
	120	6.03	6.59	7.36	8.51	10.4	12.5	15.2
0.2	24	4.18	5.19	6.70	9.02	14.9	22.3	35.0
	30	3.68	4.39	5.40	7.04	10.5	14.3	22.3
	40	3.32	3.80	4.50	5.63	7.71	10.3	14.2
	60	3.03	3.38	3.90	4.68	6.32	7.92	10.2
	120	2.74	2.99	3.33	3.84	4.78	5.73	6.94
0.1	24	0.585	0.676	0.814	1.07	1.60	2.43	4.18
	30	0.629	0.728	0.869	1.12	1.58	2.28	3.47
	40	0.660	0.750	0.878	1.10	1.56	2.12	3.25
	60	0.678	0.765	0.876	1.07	1.46	1.88	2.57
	120	0.696	0.762	0.855	0.995	1.26	1.53	1.96
-0.1	24	0.371	0.389	0.412	0.443	0.506	0.572	0.665
	30	0.418	0.441	0.469	0.511	0.594	0.686	0.823
	40	0.475	0.502	0.541	0.606	0.726	0.873	1.08
	60	0.563	0.602	0.656	0.736	0.896	1.11	1.38
	120	0.702	0.757	0.829	0.941	1.18	1.45	1.88

TABLE 5.6a. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 6$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.004	24	0.980	1.06	1.17	1.34	1.64	1.99	2.46
	30	1.12	1.22	1.35	1.57	1.94	2.39	3.12
	40	1.29	1.40	1.58	1.83	2.33	2.82	3.64
	60	1.48	1.63	1.81	2.09	2.67	3.30	4.25
	120	1.72	1.86	2.05	2.35	2.93	3.55	4.61
0.0035	24	0.749	0.810	0.894	1.02	1.25	1.51	1.88
	30	0.855	0.931	1.04	1.20	1.48	1.83	2.38
	40	0.985	1.07	1.20	1.40	1.78	2.15	2.78
	60	1.13	1.24	1.38	1.60	2.04	2.52	3.25
	120	1.32	1.42	1.57	1.80	2.24	2.71	3.53
0.003	24	0.550	0.594	0.655	0.748	0.915	1.11	1.37
	30	0.627	0.683	0.760	0.877	1.09	1.34	1.74
	40	0.723	0.786	0.884	1.03	1.31	1.58	2.03
	60	0.832	0.914	1.02	1.17	1.50	1.85	2.38
	120	0.968	1.04	1.15	1.32	1.65	1.99	2.59
0.0025	24	0.381	0.412	0.454	0.518	0.633	0.765	0.947
	30	0.435	0.473	0.526	0.607	0.752	0.926	1.20
	40	0.502	0.545	0.613	0.711	0.905	1.09	1.41
	60	0.577	0.634	0.705	0.815	1.04	1.28	1.65
	120	0.672	0.725	0.799	0.918	1.14	1.38	1.80

TABLE 5.6b. Upper Tail Percentage Points for the Statistic $T_1(c) \times 10{,}000$, $p = 6$

c	n	75	80	85	90	95	97.5	99
		PERCENT						
0.002	24	2.43	2.63	2.90	3.31	4.04	4.87	6.03
	30	2.78	3.02	3.36	3.87	4.80	5.91	7.67
	40	3.21	3.49	3.91	4.54	5.78	6.98	8.97
	60	3.69	4.05	4.51	5.21	6.65	8.19	10.5
	120	4.30	4.64	5.11	5.88	7.32	8.86	11.5
0.0015	24	1.37	1.48	1.63	1.86	2.27	2.72	3.37
	30	1.56	1.70	1.89	2.18	2.69	3.31	4.30
	40	1.80	1.96	2.20	2.55	3.24	3.91	5.03
	60	2.08	2.28	2.53	2.92	3.73	4.60	5.92
	120	2.42	2.61	2.87	3.31	4.12	4.98	6.47
0.001	24	0.607	0.655	0.723	0.823	1.00	1.21	1.49
	30	0.693	0.753	0.838	0.964	1.19	1.47	1.90
	40	0.799	0.869	0.974	1.13	1.44	1.74	2.23
	60	0.922	1.01	1.12	1.30	1.66	2.04	2.63
	120	1.07	1.16	1.28	1.47	1.83	2.21	2.88
0.0005	24	0.152	0.163	0.180	0.205	0.250	0.301	0.371
	30	0.173	0.188	0.209	0.240	0.297	0.365	0.474
	40	0.200	0.217	0.243	0.282	0.358	0.433	0.555
	60	0.230	0.253	0.281	0.324	0.413	0.510	0.656
	120	0.269	0.290	0.319	0.367	0.457	0.552	0.720

TABLE 5.6c. Upper Tail Percentage Points for the Statistic $T_1(c) \times 10{,}000$, $p = 6$

c	n	75	80	85	90	95	97.5	99
		PERCENT						
-0.0005	24	0.151	0.163	0.180	0.204	0.248	0.299	0.368
	30	0.172	0.187	0.208	0.239	0.296	0.362	0.470
	40	0.199	0.217	0.243	0.281	0.356	0.430	0.551
	60	0.230	0.252	0.280	0.323	0.412	0.508	0.653
	120	0.269	0.290	0.319	0.367	0.458	0.551	0.720
-0.001	24	0.604	0.650	0.716	0.815	0.991	1.19	1.46
	30	0.689	0.749	0.831	0.955	1.18	1.44	1.87
	40	0.796	0.866	0.969	1.12	1.42	1.71	2.20
	60	0.918	1.01	1.12	1.29	1.64	2.03	2.61
	120	1.07	1.16	1.27	1.47	1.83	2.20	2.88
-0.0015	24	1.36	1.46	1.61	1.83	2.22	2.67	3.28
	30	1.55	1.68	1.86	2.14	2.64	3.24	4.20
	40	1.79	1.95	2.18	2.52	3.19	3.85	4.93
	60	2.06	2.26	2.52	2.91	3.69	4.56	5.86
	120	2.42	2.61	2.87	3.30	4.11	4.96	6.47
-0.002	24	2.41	2.59	2.85	3.24	3.94	4.73	5.80
	30	2.75	2.98	3.31	3.80	4.69	5.74	7.43
	40	3.18	3.46	3.86	4.47	5.66	6.82	8.74
	60	3.66	4.02	4.47	5.16	6.56	8.09	10.4
	120	4.30	4.63	5.10	5.86	7.31	8.82	11.5

TABLE 5.6d. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 6$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
-0.0025	24	0.376	0.404	0.445	0.505	0.614	0.736	0.903
	30	0.429	0.465	0.516	0.593	0.731	0.894	1.16
	40	0.496	0.539	0.603	0.697	0.883	1.03	1.36
	60	0.572	0.627	0.698	0.805	1.02	1.26	1.62
	120	0.671	0.720	0.797	0.916	1.14	1.38	1.80
-0.003	24	0.540	0.581	0.640	0.726	0.882	1.06	1.26
	30	0.617	0.670	0.742	0.852	1.05	1.28	1.66
	40	0.713	0.776	0.866	1.00	1.27	1.52	1.95
	60	0.824	0.903	1.00	1.16	1.47	1.81	2.33
	120	0.966	1.04	1.15	1.32	1.64	1.98	2.59
-0.0035	24	0.734	0.789	0.869	0.985	1.20	1.43	1.75
	30	0.838	0.910	1.01	1.16	1.43	1.73	2.25
	40	0.970	1.05	1.18	1.36	1.72	2.07	2.65
	60	1.12	1.23	1.37	1.57	2.00	2.46	3.16
	120	1.31	1.42	1.56	1.79	2.23	2.69	3.52
-0.004	24	0.958	1.03	1.13	1.28	1.56	1.86	2.28
	30	1.09	1.19	1.31	1.51	1.86	2.27	2.93
	40	1.27	1.38	1.53	1.77	2.24	2.70	3.45
	60	1.46	1.60	1.78	2.05	2.61	3.21	4.12
	120	1.72	1.85	2.03	2.34	2.92	3.52	4.60

TABLE 5.6e. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, p = 6

c	n	75	80	85	90	95	97.5	99
		PERCENT						
0.2	24	8.67	11.1	14.3	20.3	34.0	52.6	91.0
	30	6.97	8.11	9.87	13.1	18.7	26.1	39.1
	40	5.85	6.69	7.79	9.44	12.9	16.7	23.1
	60	5.00	5.55	6.31	7.39	9.43	11.4	14.4
	120	4.39	4.76	5.21	5.91	7.06	8.11	9.55
0.1	24	1.02	1.17	1.41	1.85	2.91	4.31	7.17
	30	1.07	1.21	1.44	1.79	2.64	3.76	5.62
	40	1.09	1.23	1.42	1.76	2.40	3.18	4.68
	60	1.08	1.21	1.39	1.65	2.13	2.65	3.45
	120	1.10	1.19	1.32	1.52	1.83	2.16	2.59
0.05	24	0.187	0.208	0.238	0.286	0.387	0.508	0.691
	30	0.209	0.232	0.264	0.315	0.429	0.564	0.766
	40	0.230	0.255	0.289	0.346	0.457	0.584	0.812
	60	0.247	0.275	0.311	0.371	0.469	0.584	0.766
	120	0.273	0.298	0.329	0.376	0.458	0.547	0.696
-0.1	24	0.534	0.558	0.585	0.626	0.697	0.783	0.912
	30	0.606	0.635	0.669	0.719	0.826	0.951	1.12
	40	0.689	0.730	0.783	0.855	1.01	1.18	1.43
	60	0.820	0.873	0.942	1.07	1.27	1.48	1.82
	120	1.05	1.12	1.21	1.38	1.65	1.99	2.45

TABLE 5.7a. Upper Tail Percentage Points for the Statistic $T_1(c) \times 100$, $p = 8$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.004	24	0.183	0.195	0.210	0.231	0.269	0.304	0.358
	30	0.216	0.232	0.253	0.283	0.338	0.397	0.474
	40	0.252	0.271	0.295	0.334	0.402	0.482	0.584
	60	0.293	0.318	0.346	0.391	0.471	0.561	0.693
	120	0.339	0.362	0.394	0.436	0.524	0.618	0.757
0.0035	24	0.140	0.149	0.161	0.177	0.205	0.232	0.273
	30	0.165	0.177	0.193	0.216	0.258	0.303	0.361
	40	0.192	0.207	0.225	0.255	0.307	0.368	0.445
	60	0.224	0.243	0.265	0.299	0.360	0.429	0.529
	120	0.260	0.277	0.301	0.334	0.401	0.473	0.580
0.003	24	0.103	0.109	0.118	0.129	0.150	0.170	0.200
	30	0.121	0.130	0.141	0.158	0.189	0.222	0.264
	40	0.141	0.152	0.165	0.187	0.225	0.269	0.326
	60	0.164	0.178	0.194	0.219	0.264	0.314	0.387
	120	0.191	0.204	0.221	0.245	0.294	0.347	0.426
0.0025	24	0.0710	0.0755	0.0815	0.0896	0.104	0.117	0.138
	30	0.0838	0.0899	0.0978	0.110	0.131	0.153	0.182
	40	0.0979	0.105	0.114	0.129	0.156	0.186	0.225
	60	0.114	0.123	0.135	0.152	0.183	0.218	0.268
	120	0.132	0.141	0.154	0.170	0.204	0.241	0.295

TABLE 5.7b. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 8$

c	n	PERCENT 75	80	85	90	95	97.5	99
0.002	24	0.454	0.482	0.520	0.572	0.663	0.748	0.878
	30	0.535	0.574	0.625	0.700	0.832	0.977	1.16
	40	0.625	0.671	0.730	0.826	0.994	1.19	1.44
	60	0.728	0.789	0.861	0.972	1.17	1.39	1.71
	120	0.846	0.904	0.982	1.09	1.31	1.54	1.89
0.0015	24	0.255	0.271	0.292	0.321	0.372	0.419	0.492
	30	0.300	0.322	0.350	0.392	0.467	0.547	0.650
	40	0.351	0.376	0.410	0.463	0.557	0.665	0.804
	60	0.409	0.443	0.484	0.546	0.655	0.780	0.959
	120	0.476	0.508	0.552	0.611	0.735	0.868	1.06
0.001	24	0.113	0.120	0.129	0.142	0.165	0.186	0.218
	30	0.133	0.143	0.155	0.174	0.207	0.242	0.287
	40	0.156	0.167	0.182	0.205	0.247	0.295	0.356
	60	0.182	0.197	0.215	0.242	0.290	0.346	0.425
	120	0.211	0.226	0.245	0.272	0.326	0.385	0.471
0.0005	24	0.0282	0.0299	0.0322	0.0354	0.0410	0.0462	0.0541
	30	0.0332	0.0356	0.0388	0.4342	0.0515	0.0603	0.0716
	40	0.0389	0.0417	0.0453	0.0512	0.0615	0.0734	0.0885
	60	0.0453	0.0491	0.0535	0.0604	0.0725	0.0863	0.106
	120	0.0528	0.0565	0.0613	0.0678	0.0815	0.0963	0.118

TABLE 5.7c. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, p = 8

c	n	75	80	85	90	95	97.5	99
		PERCENT						
-0.0005	24	0.0281	0.0298	0.0321	0.0352	0.0408	0.0458	0.0537
	30	0.0331	0.0355	0.0386	0.0431	0.0511	0.0598	0.0709
	40	0.0387	0.0415	0.0451	0.0509	0.0611	0.0728	0.0879
	60	0.0452	0.0489	0.0534	0.0601	0.0722	0.0859	0.105
	120	0.0528	0.0564	0.0612	0.0677	0.0814	0.0961	0.177
-0.001	24	0.112	0.119	0.128	0.140	0.163	0.182	0.214
	30	0.132	0.142	0.154	0.172	0.204	0.238	0.282
	40	0.155	0.166	0.180	0.203	0.244	0.290	0.350
	60	0.181	0.195	0.213	0.240	0.288	0.343	0.420
	120	0.211	0.226	0.245	0.271	0.326	0.384	0.468
-0.0015	24	0.252	0.267	0.287	0.314	0.364	0.409	0.478
	30	0.297	0.318	0.346	0.386	0.456	0.534	0.632
	40	0.347	0.372	0.404	0.456	0.546	0.651	0.783
	60	0.406	0.439	0.479	0.539	0.647	0.769	0.941
	120	0.475	0.507	0.551	0.608	0.732	0.864	1.05
-0.002	24	0.446	0.473	0.509	0.557	0.645	0.724	0.846
	30	0.527	0.564	0.613	0.684	0.808	0.946	1.12
	40	0.616	0.661	0.717	0.810	0.969	1.16	1.39
	60	0.721	0.779	0.850	0.957	1.15	1.36	1.67
	120	0.843	0.901	0.978	1.08	1.30	1.54	1.87

TABLE 5.7d. Upper Tail Percentage Points for the Statistic $T_1(c) \times 100$, $p = 8$

c	n	75	80	85	90	95	97.5	99
		PERCENT						
-0.0025	24	0.0696	0.0738	0.0793	0.0868	0.101	0.113	0.132
	30	0.0822	0.0879	0.0956	0.106	0.126	0.147	0.174
	40	0.0961	0.103	0.112	0.126	0.151	0.180	0.216
	60	0.112	0.121	0.133	0.149	0.179	0.213	0.260
	120	0.132	0.141	0.153	0.169	0.203	0.240	0.293
-0.003	24	0.100	0.106	0.114	0.125	0.144	0.162	0.189
	30	0.118	0.126	0.137	0.153	0.181	0.211	0.249
	40	0.138	0.148	0.161	0.181	0.217	0.258	0.309
	60	0.162	0.175	0.191	0.215	0.257	0.306	0.373
	120	0.190	0.203	0.220	0.243	0.293	0.345	0.421
-0.0035	24	0.136	0.144	0.155	0.169	0.196	0.219	0.255
	30	0.160	0.171	0.186	0.207	0.245	0.287	0.338
	40	0.188	0.201	0.218	0.246	0.294	0.351	0.420
	60	0.220	0.237	0.259	0.292	0.349	0.415	0.506
	120	0.258	0.276	0.299	0.330	0.398	0.469	0.574
-0.004	24	0.177	0.188	0.202	0.220	0.255	0.285	0.332
	30	0.209	0.224	0.243	0.270	0.319	0.373	0.440
	40	0.245	0.262	0.284	0.321	0.382	0.456	0.545
	60	0.287	0.310	0.338	0.380	0.455	0.541	0.659
	120	0.337	0.360	0.391	0.432	0.519	0.613	0.749

TABLE 5.7e. Upper Tail Percentage Points for the Statistic $T_1(c)$, $p = 8$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.1	24	2.62	3.09	3.73	4.93	8.01	12.4	23.2
	30	2.60	2.98	3.53	4.47	6.64	9.85	15.6
	40	2.58	2088	3.34	4.02	5.47	7.29	10.5
	60	2.44	2.69	2.99	3.48	4.36	5.37	6.75
	120	2.29	2.45	2.66	2.96	3.50	4.01	4.90
0.075	24	1.07	1.21	1.39	1.71	2.42	3.29	5.20
	30	1.16	1.30	1.51	1.83	2.56	3.57	5.37
	40	1.25	1.38	1.58	1.88	2.50	3.28	4.74
	60	1.27	1.39	1.55	1.81	2.25	2.79	3.57
	120	1.27	1.36	1.48	1.64	1.97	2.24	2.82
-0.05	24	0.243	0.252	0.263	0.281	0.311	0.338	0.373
	30	0.282	0.296	0.313	0.339	0.383	0.427	0.484
	40	0.338	0.357	0.380	0.415	0.475	0.542	0.637
	60	0.405	0.433	0.464	0.512	0.596	0.685	0.812
	120	0.511	0.542	0.585	0.648	0.755	0.870	1.06
-0.1	24	0.901	0.923	0.955	0.993	1.07	1.12	1.18
	30	1.04	1.07	1.11	1.18	1.27	1.35	1.47
	40	1.23	1.28	1.34	1.43	1.57	1.71	1.88
	60	1.50	1.57	1.67	1.80	2.01	2.23	3.14
	120	1.97	2.07	2.22	2.42	2.76	3.15	3.65

TABLE 5.8a. Upper Tail Percentage Points for the Statistic $T_1(c) \times 100$, p = 10

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.004	24	0.288	0.303	0.323	0.346	0.386	0.420	0.472
	30	0.347	0.368	0.395	0.438	0.506	0.566	0.668
	40	0.421	0.447	0.483	0.535	0.623	0.717	0.862
	60	0.504	0.533	0.578	0.639	0.756	0.882	1.04
	120	0.597	0.635	0.683	0.747	0.861	0.987	1.16
0.0035	24	0.220	0.232	0.246	0.264	0.294	0.320	0.359
	30	0.265	0.281	0.302	0.334	0.386	0.431	0.509
	40	0.322	0.342	0.369	0.408	0.475	0.546	0.657
	60	0.386	0.408	0.442	0.489	0.577	0.673	0.795
	120	0.457	0.486	0.522	0.572	0.658	0.755	0.890
0.003	24	0.161	0.170	0.180	0.193	0.215	0.234	0.262
	30	0.194	0.206	0.221	0.245	0.282	0.315	0.371
	40	0.236	0.250	0.270	0.299	0.348	0.400	0.480
	60	0.283	0.299	0.324	0.358	0.423	0.493	0.582
	120	0.335	0.357	0.384	0.420	0.483	0.554	0.653
0.0025	24	0.111	0.117	0.125	0.134	0.149	0.162	0.181
	30	0.135	0.142	0.153	0.169	0.195	0.218	0.256
	40	0.163	0.173	0.187	0.207	0.241	0.276	0.331
	60	0.196	0.207	0.225	0.248	0.293	0.341	0.403
	120	0.233	0.248	0.266	0.291	0.335	0.384	0.452

TABLE 5.8b. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, p = 10

		PERCENT						
c	n	75	80	85	90	95	97.5	99
0.002	24	0.711	0.749	0.796	0.853	0.950	1.03	1.16
	30	0.859	0.909	0.976	1.08	1.24	1.39	1.63
	40	1.04	1.11	1.19	1.32	1.54	1.76	2.11
	60	1.25	1.32	1.43	1.59	1.87	2.18	2.57
	120	1.49	1.58	1.70	1.86	2.14	2.46	2.89
0.0015	24	0.399	0.420	0.446	0.478	0.532	0.578	0.647
	30	0.482	0.510	0.547	0.605	0.696	0.777	0.910
	40	0.585	0.621	0.670	0.741	0.861	0.985	1.18
	60	0.703	0.743	0.805	0.890	1.05	1.22	1.44
	120	0.837	0.890	0.956	1.05	1.20	1.38	1.62
0.001	24	0.177	0.186	0.198	0.212	0.235	0.256	0.286
	30	0.214	0.226	0.243	0.268	0.308	0.344	0.402
	40	0.260	0.275	0.297	0.328	0.381	0.436	0.522
	60	0.312	0.329	0.357	0.395	0.464	0.540	0.637
	120	0.372	0.395	0.425	0.464	0.534	0.613	0.721
0.0005	24	0.0441	0.0464	0.0493	0.0528	0.0586	0.0637	0.0713
	30	0.0533	0.0564	0.0605	0.0667	0.0766	0.0856	0.0999
	40	0.0647	0.0686	0.0740	0.0819	0.0949	0.109	1.130
	60	0.0778	0.0822	0.0891	0.0984	0.116	0.135	0.159
	120	0.0928	0.0988	0.106	0.116	0.133	0.153	0.180

TABLE 5.8c. Upper Tail Percentage Points for the Statistic
$T_1(c) \times 1000$, $p = 10$

		PERCENT						
c	n	75	80	85	90	95	97.5	99
-0.0005	24	0.0439	0.0461	0.0490	0.0524	0.0582	0.0633	0.0706
	30	0.0530	0.0560	0.0601	0.0663	0.0760	0.0848	0.0987
	40	0.0646	0.0682	0.0735	0.0814	0.0943	0.108	0.129
	60	0.0775	0.0819	0.0888	0.0980	0.115	0.134	0.157
	120	0.0927	0.0986	0.106	0.116	0.133	0.153	0.180
-0.001	24	0.175	0.184	0.195	0.209	0.232	0.252	0.281
	30	0.212	0.224	0.240	0.264	0.303	0.337	0.393
	40	0.257	0.272	0.293	0.324	0.376	0.428	0.513
	60	0.310	0.327	0.354	0.391	0.459	0.534	0.627
	120	0.371	0.394	0.423	0.463	0.532	0.611	0.718
-0.0015	24	0.393	0.413	0.438	0.468	0.519	0.564	0.629
	30	0.475	0.502	0.538	0.593	0.678	0.756	0.878
	40	0.577	0.611	0.658	0.728	0.842	0.959	1.15
	60	0.696	0.734	0.796	0.878	1.03	1.20	1.41
	120	0.833	0.886	0.951	1.04	1.20	1.37	1.61
-0.002	24	0.697	0.731	0.776	0.829	0.919	0.999	1.11
	30	0.841	0.890	0.953	1.05	1.20	1.34	1.55
	40	1.02	1.08	1.17	1.29	1.49	1.70	2.03
	60	1.24	1.30	1.41	1.56	1.83	2.12	2.49
	120	1.48	1.58	1.69	1.85	2.12	2.44	2.86

TABLE 5.8d. Upper Tail Percentage Points for the Statistic $T_1(c) \times 1000$, $p = 10$

c	n	75	80	85	90	95	97.5	99
		PERCENT						
-0.0025	24	0.109	0.114	0.121	0.129	0.143	0.156	0.173
	30	0.131	0.139	0.149	0.164	0.187	0.208	0.241
	40	0.160	0.169	0.182	0.201	0.232	0.264	0.316
	60	0.193	0.203	0.220	0.243	0.284	0.330	0.388
	120	0.231	0.246	0.264	0.288	0.331	0.380	0.447
-0.003	24	0.156	0.164	0.174	0.185	0.205	0.223	0.248
	30	0.189	0.199	0.213	0.235	0.268	0.298	0.345
	40	0.229	0.243	0.261	0.288	0.333	0.379	0.453
	60	0.277	0.292	0.316	0.348	0.409	0.474	0.557
	120	0.333	0.354	0.380	0.415	0.477	C.547	0.643
-0.0035	24	0.212	0.222	0.236	0.251	0.278	0.303	0.336
	30	0.256	0.270	0.289	0.318	0.363	0.404	0.468
	40	0.312	0.330	0.354	0.391	0.452	0.514	0.614
	60	0.376	0.397	0.430	0.473	0.556	0.644	0.755
	120	0.453	0.481	0.516	0.564	0.648	0.744	0.874
-0.004	24	0.276	0.290	0.307	0.327	0.362	0.394	0.437
	30	0.334	0.352	0.377	0.414	0.472	0.525	0.607
	40	0.406	0.429	0.462	0.509	0.588	0.669	0.798
	60	0.491	0.517	0.560	0.616	0.724	0.838	0.983
	120	0.591	0.628	0.674	0.736	0.846	0.970	1.14

TABLE 5.8e. Upper Tail Percentage Points for the Statistic $T_1(c)$, p = 10

c	n	75	80	85	90	95	97.5	99
		PERCENT						
0.1	24	6.48	7.90	10.1	14.7	26.1	42.9	66.1
	30	6.07	7.07	8.53	11.3	17.2	24.5	36.2
	40	5.35	6.01	6.93	8.46	11.6	15.2	21.2
	60	4.72	5.14	5.72	6.55	8.01	9.93	12.5
	120	4.20	4.45	4.79	5.24	6.08	6.85	8.04
0.075	24	2.10	2.38	2.80	3.51	5.08	7.76	13.7
	30	2.34	2.61	3.03	3.78	5.31	7.51	12.1
	40	2.40	2.65	3.02	3.62	4.80	6.35	8.94
	60	2.38	2.58	2.86	3.26	3.97	4.98	6.20
	120	2.30	2.45	2.62	2.88	3.34	3.80	4.52
-0.05	24	0.367	0.377	0.391	0.413	0.445	0.476	0.509
	30	0.438	0.456	0.477	0.507	0.559	0.605	0.662
	40	0.534	0.559	0.589	0.635	0.707	0.788	0.883
	60	0.668	0.700	0.741	0.789	0.905	1.03	1.20
	120	0.861	0.906	0.959	1.06	1.20	1.33	1.59
-0.1	24	1.33	1.35	1.38	1.42	1.49	1.55	1.61
	30	1.56	1.59	1.64	1.70	1.80	1.89	1.98
	40	1.89	1.95	2.01	2.10	2.25	2.38	2.54
	60	2.39	2.48	2.58	2.73	2.95	3.17	3.43
	120	3.23	3.38	3.55	3.79	4.21	4.61	5.18

5.3 A Power Study

A power study was performed to determine how well $T_1(c)$ could detect non-normality. The following definitions of non-Gaussian multivariate distributions were used in the power study. Suppose $(x_1, x_2, \ldots, x_p)^T \sim N(0,R)$ where R is a positive-definite correlation matrix.

(1) If $y_i = \sum_{j=1}^{r} x_{ij}^2$, $i = 1, 2, \ldots, p$. Then $(y_1, y_2, \ldots, y_p)^T$ follows a p-variate chi-square distribution with r degrees of freedom.

(2) If w has a chi-square distribution with r degrees of freedom, w is independent of the x_i's, and $t_i = x_i/\sqrt{w/r}$, then $(t_1, t_2, \ldots, t_p)^T$ follows a p-variate, t-distribution on r degrees of freedom.

(3) If $y_i = \exp[x_i]$ $i = 1, 2, \ldots, p$ then $(y_1, y_2, \ldots, y_p)^T$ follows a p-variate lognormal distribution.

(4) If, in a sample of n p-variate vectors, $\lambda = n_1/n$ is the fraction from $N_p(m_1, D_1)$ and $(1 - \lambda)$ is the fraction from $N_p(m_2, D_2)$, then $(x_1, x_2, \ldots, x_p)$ is from the normal mixture $\lambda N_p(m_1, D_1) + (1 - \lambda)N_p(m_2, D_2)$.

For the analysis, the matrix R was the p-variate identity for p = 1, 2, or 5. Chi-square distributions with 2, 4, 6, 10, and 14 degrees of freedom were examined; t-distributions with 3, 5, 7, and 9 degrees of freedom were examined. Four Gaussian mixture distributions were analyzed. For $p = 1$, the parameters defining the mixture distributions were (1) $m_1 = 0$, $m_2 = 2$, $D_1 = D_2 = 1$, and $\lambda = 0.75$; (2) $m_1 = 0$, $m_2 = 2$, $D_1 = D_2 = 1$, and $\lambda = 0.5$; (3) $m_1 = m_2 = 0$, $D_1 = 1$, $D_2 = 4$, and $\lambda = 0.5$; and (4) $m_1 = 0$, $m_2 = 1$, $D_1 = D_2 = 1$, and $\lambda = 0.75$. For $p = 2$, the parameters defining the mixture distributions were

(1) $m_1 = [2,2]^T$, $m_2 = [0,0]^T$,

$$D_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \text{ and } \lambda = 0.25 ;$$

(2) $m_1 = [2,2]^T$, $m_2 = [0,0]^T$,

$$D_1 = D_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \lambda = 0.5 ;$$

(3) $m_1 = [1,1]^T$, $m_2 = [0,0]^T$,

$$D_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 1 & -0.75 \\ -0.75 & 1 \end{bmatrix}$$

and $\lambda = 0.25$;

(4) $m_1 = [1,1]^T$, $m_2 = [0,0]^T$,

$$D_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \text{ and } \lambda = 0.25 .$$

For $p = 5$, the Gaussian mixture parameters were

(1) $m_1 = [4,4,4,4,4]^T$, $m_2 = [0,0,0,0,0]^T$,

$$D_1 = I_5 , \quad D_2 = \begin{bmatrix} 1.0 & & & & \\ 0.5 & 1.0 & & & \\ 0.25 & 0.5 & 1.0 & & \\ 0.125 & 0.25 & 0.5 & 1.0 & \\ 0.0625 & 0.125 & 0.25 & 0.5 & 1.0 \end{bmatrix} ,$$

and $\lambda = 0.25$;

(2) $m_1 = [3,3,3,3,3]^T$, $m_2 = [0,0,0,0,0]^T$,

$D_1 = D_2 = I_5$, and $\lambda = 0.5$;

(3) $m_1 = m_2 = [0,0,0,0,0]^T$, $D_1 = I$, $D_2 = 4I$

and $\lambda = 0.5$;

(4) $m_1 = [2,2,2,2,2]^T$, $m_2 = [0,0,0,0,0]^T$, $D_1 = I_5$,

$$D_2 = \begin{bmatrix} 1.0 & & & & \\ 0.5 & 1.0 & & & \\ 0.25 & 0.5 & 1.0 & & \\ 0.125 & 0.25 & 0.5 & 1.0 & \\ 0.0625 & 0.625 & 0.25 & 0.5 & 1.0 \end{bmatrix} , \text{ and } \lambda = 0.25.$$

These mixture distributions are denoted by MX1, MX2, MX3, and MX4. Based on 2000 trials, the above distributions were examined for sample sizes 20, 50, and 100. Tables 5.9 to 5.16 (a-d) present the power results for $\alpha = 0.1$, 0.05, 0.025, and 0.01.

Comparing Tables 5.9 to 5.16 (a and b) with the results of a power study carried out by Paulson, Roohan, Hwang and Fuller (1987), it can be seen that the proposed test has excellent power against heavy-tailed symmetric alternatives and good power against nonsymmetric alternatives. In general, the power of $T_1(c)$ increases as c increases since the procedure becomes more critical of observations in the tails of the distribution. The exception is the second mixture distribution, MX2, where the power of $T_1(c)$ increases as c decreases. For MX2, an equal number of observations are drawn from Gaussian distributions with equal covariance matricies but with unequal means. The observations from MX2 that are near the mean $0.5(m_1 + m_2)$ can be considered as inliers of the distribution. Since the model-critical procedure becomes increasingly critical of inliers as $c < 0$ decreases, the power of $T_1(c)$ increases as c decreases. Finally, it will be noted that for $|c| > 0.1$ the power does not increase dramatically in general. The implication is that, for hypothesis testing, small values of $|c|$ are adequate. In the next section, it will be shown that $T_1(c)$ is a good test of normality for the residuals from a linear model when $|c| \leq 0.1$. In this way, $T_1(c)$ provides a joint test of fit for the joint character of the errors and the assumed parametric model.

TABLE 5.9a

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.1$, $n = 20$ and $p = 1$

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	0.80	0.77	0.72	0.69	0.67	0.65	0.63	0.59
T(9)	0.21	0.20	0.19	0.18	0.17	0.17	0.16	0.15
T(7)	0.25	0.24	0.22	0.21	0.20	0.19	0.18	0.17
T(5)	0.34	0.32	0.29	0.28	0.27	0.26	0.25	0.23
T(3)	0.51	0.48	0.46	0.44	0.42	0.42	0.40	0.37
$X^2(14)$	0.20	0.19	0.18	0.18	0.18	0.18	0.17	0.16
$X^2(10)$	0.25	0.24	0.23	0.22	0.21	0.21	0.20	0.19
$X^2(6)$	0.32	0.31	0.28	0.27	0.26	0.25	0.24	0.22
$X^2(4)$	0.40	0.39	0.36	0.34	0.33	0.31	0.30	0.27
$X^2(2)$	0.58	0.55	0.51	0.49	0.47	0.46	0.42	0.39
MX1	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
MX2	0.05	0.08	0.12	0.14	0.15	0.16	0.17	0.18
MX3	0.30	0.27	0.23	0.21	0.19	0.18	0.17	0.15
MX4	0.08	0.09	0.10	0.10	0.10	0.10	0.10	0.10

TABLE 5.9b

Power of the Test for Normality $T_1(c)$ at
Significance Level $\alpha = 0.05$, n = 20 and p = 1

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	0.75	0.73	0.68	0.66	0.64	0.63	0.60	0.56
T(9)	0.14	0.14	0.14	0.13	0.13	0.13	0.12	0.11
T(7)	0.18	0.18	0.17	0.16	0.16	0.15	0.14	0.13
T(5)	0.25	0.26	0.25	0.24	0.23	0.23	0.21	0.19
T(3)	0.43	0.43	0.42	0.40	0.39	0.38	0.36	0.34
$X^2(14)$	0.13	0.13	0.13	0.12	0.12	0.12	0.11	0.11
$X^2(10)$	0.18	0.18	0.17	0.16	0.15	0.15	0.13	0.13
$X^2(6)$	0.23	0.23	0.21	0.20	0.20	0.19	0.18	0.17
$X^2(4)$	0.32	0.31	0.28	0.27	0.26	0.25	0.24	0.22
$X^2(2)$	0.50	0.49	0.46	0.44	0.41	0.40	0.38	0.35
MX1	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
MX2	0.01	0.01	0.01	0.03	0.04	0.05	0.07	0.09
MX3	0.19	0.18	0.18	0.16	0.15	0.14	0.13	0.11
MX4	0.05	0.04	0.04	0.05	0.04	0.05	0.05	0.06

TABLE 5.9c

Power of the Test for Normality $T_1(c)$ at
Significance Level α = 0.025, n = 20 and p = 1

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	0.69	0.66	0.63	0.60	0.58	0.58	0.56	0.53
T(9)	0.10	0.10	0.10	0.09	0.09	0.09	0.09	0.08
T(7)	0.13	0.13	0.13	0.12	0.12	0.12	0.12	0.11
T(5)	0.19	0.19	0.19	0.18	0.18	0.18	0.18	0.17
T(3)	0.37	0.36	0.35	0.34	0.33	0.33	0.32	0.30
$X^2(14)$	0.09	0.09	0.09	0.08	0.08	0.08	0.08	0.07
$X^2(10)$	0.12	0.12	0.11	0.10	0.10	0.10	0.09	0.09
$X^2(6)$	0.17	0.17	0.17	0.15	0.15	0.14	0.13	0.16
$X^2(4)$	0.26	0.24	0.23	0.21	0.21	0.20	0.19	0.18
$X^2(2)$	0.44	0.42	0.38	0.36	0.34	0.34	0.33	0.31
MX1	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
MX2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
MX3	0.12	0.12	0.12	0.10	0.10	0.10	0.10	0.09
MX4	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.02

TABLE 5.9d

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.01$, n = 20 and p = 1

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	0.62	0.57	0.52	0.50	0.49	0.48	0.47	0.45
T(9)	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05
T(7)	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
T(5)	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
T(3)	0.27	0.26	0.25	0.25	0.25	0.25	0.25	0.24
$X^2(14)$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04
$X^2(10)$	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05
$X^2(6)$	0.12	0.11	0.10	0.10	0.10	0.09	0.09	0.09
$X^2(4)$	0.18	0.18	0.16	0.14	0.13	0.13	0.13	0.12
$X^2(2)$	0.35	0.32	0.28	0.27	0.26	0.26	0.26	0.24
MX1	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MX2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MX3	0.07	0.06	0.05	0.05	0.05	0.05	0.05	0.05
MX4	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.01

TABLE 5.10a

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.1$, $n = 50$ and $p = 1$

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	0.99	0.98	0.97	0.96	0.95	0.95	0.93	0.92
T(9)	0.30	0.30	0.29	0.28	0.27	0.26	0.25	0.24
T(7)	0.30	0.30	0.29	0.36	0.35	0.35	0.34	0.31
T(5)	0.55	0.54	0.52	0.51	0.50	0.50	0.49	0.47
T(3)	0.79	0.78	0.76	0.75	0.74	0.74	0.73	0.71
$X^2(14)$	0.31	0.30	0.29	0.28	0.27	0.27	0.26	0.25
$X^2(10)$	0.37	0.36	0.35	0.34	0.32	0.32	0.31	0.29
$X^2(6)$	0.50	0.48	0.45	0.44	0.42	0.41	0.40	0.38
$X^2(4)$	0.64	0.61	0.57	0.55	0.53	0.53	0.50	0.47
$X^2(2)$	0.85	0.83	0.79	0.76	0.74	0.73	0.71	0.68
MX1	0.07	0.07	0.08	0.07	0.07	0.08	0.08	0.08
MX2	0.14	0.15	0.18	0.18	0.20	0.21	0.22	0.24
MX3	0.44	0.41	0.38	0.36	0.34	0.34	0.32	0.28
MX4	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.11

TABLE 5.10b

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.05$, $n = 50$ and $p = 1$

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	0.98	0.97	0.95	0.95	0.94	0.93	0.92	0.91
T(9)	0.23	0.23	0.23	0.22	0.22	0.21	0.21	0.20
T(7)	0.32	0.33	0.31	0.30	0.30	0.29	0.28	0.27
T(5)	0.48	0.48	0.47	0.46	0.45	0.45	0.43	0.42
T(3)	0.74	0.73	0.72	0.72	0.71	0.70	0.69	0.67
$X^2(14)$	0.22	0.22	0.21	0.21	0.20	0.20	0.20	0.19
$X^2(10)$	0.29	0.30	0.28	0.27	0.26	0.25	0.25	0.24
$X^2(6)$	0.43	0.42	0.39	0.38	0.36	0.36	0.34	0.32
$X^2(4)$	0.56	0.54	0.52	0.49	0.47	0.47	0.45	0.43
$X^2(2)$	0.82	0.78	0.74	0.72	0.70	0.69	0.67	0.64
MX1	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
MX2	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.11
MX3	0.35	0.33	0.31	0.29	0.28	0.27	0.25	0.24
MX4	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.05

TABLE 5.10c

Power of the Test for Normality $T_1(c)$ at Significance Level α = 0.025, n = 50 and p = 1

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	0.96	0.95	0.93	0.92	0.91	0.90	0.90	0.88
T(9)	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
T(7)	0.24	0.24	0.24	0.24	0.23	0.23	0.23	0.23
T(5)	0.39	0.39	0.39	0.38	0.38	0.38	0.38	0.37
T(3)	0.68	0.67	0.65	0.64	0.63	0.63	0.63	0.62
$X^2(14)$	0.16	0.16	0.15	0.15	0.14	0.14	0.14	0.14
$X^2(10)$	0.23	0.22	0.21	0.21	0.20	0.20	0.19	0.19
$X^2(6)$	0.33	0.33	0.31	0.30	0.29	0.29	0.28	0.27
$X^2(4)$	0.48	0.46	0.43	0.41	0.40	0.40	0.39	0.37
$X^2(2)$	0.76	0.72	0.68	0.66	0.64	0.63	0.61	0.59
MX1	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01
MX2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
MX3	0.25	0.23	0.21	0.20	0.20	0.19	0.19	0.18
MX4	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

TABLE 5.10d

Power of the Test for Normality $T_1(c)$ at Significance Level α = 0.01, n = 50 and p = 1

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	0.95	0.93	0.90	0.87	0.86	0.86	0.84	0.83
T(9)	0.10	0.11	0.11	0.10	0.10	0.10	0.10	0.10
T(7)	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
T(5)	0.32	0.31	0.30	0.30	0.29	0.29	0.29	0.29
T(3)	0.60	0.59	0.56	0.55	0.54	0.54	0.53	0.52
$X^2(14)$	0.10	0.10	0.10	0.09	0.09	0.09	0.09	0.09
$X^2(10)$	0.15	0.16	0.15	0.14	0.13	0.13	0.13	0.13
$X^2(6)$	0.25	0.25	0.23	0.22	0.21	0.21	0.20	0.19
$X^2(4)$	0.40	0.38	0.34	0.32	0.31	0.30	0.29	0.28
$X^2(2)$	0.69	0.65	0.60	0.56	0.54	0.53	0.52	0.50
MX1	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00
MX2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MX3	0.16	0.14	0.13	0.12	0.11	0.11	0.11	0.11
MX4	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00

TABLE 5.11a

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.1$, $n = 100$ and $p = 1$

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99
T(9)	0.42	0.43	0.42	0.41	0.40	0.40	0.39	0.38
T(7)	0.57	0.57	0.55	0.53	0.52	0.52	0.52	0.51
T(5)	0.77	0.76	0.74	0.73	0.73	0.73	0.72	0.70
T(3)	0.95	0.95	0.94	0.94	0.93	0.93	0.93	0.92
$X^2(14)$	0.39	0.40	0.38	0.37	0.36	0.36	0.35	0.34
$X^2(10)$	0.50	0.50	0.49	0.48	0.47	0.46	0.45	0.43
$X^2(6)$	0.68	0.65	0.62	0.60	0.59	0.59	0.57	0.55
$X^2(4)$	0.83	0.82	0.79	0.77	0.76	0.75	0.73	0.71
$X^2(2)$	0.98	0.97	0.95	0.94	0.94	0.93	0.92	0.92
MX1	0.08	0.08	0.08	0.07	0.07	0.08	0.08	0.08
MX2	0.29	0.30	0.31	0.31	0.32	0.32	0.33	0.36
MX3	0.63	0.60	0.57	0.54	0.53	0.52	0.50	0.48
MX4	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.11

TABLE 5.11b

Power of the Test for Normality $T_1(c)$ at
Significance Level $\alpha = 0.05$, $n = 100$ and $p = 1$

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99
T(9)	0.35	0.35	0.36	0.35	0.35	0.34	0.33	0.33
T(7)	0.49	0.49	0.48	0.48	0.47	0.47	0.45	0.44
T(5)	0.71	0.70	0.70	0.69	0.68	0.68	0.67	0.66
T(3)	0.94	0.93	0.93	0.93	0.92	0.92	0.91	0.90
$X^2(14)$	0.31	0.32	0.32	0.31	0.30	0.30	0.29	0.27
$X^2(10)$	0.41	0.43	0.43	0.42	0.41	0.40	0.39	0.37
$X^2(6)$	0.60	0.59	0.57	0.56	0.55	0.54	0.52	0.50
$X^2(4)$	0.79	0.78	0.75	0.73	0.71	0.71	0.69	0.66
$X^2(2)$	0.97	0.96	0.94	0.93	0.92	0.91	0.90	0.88
MX1	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.03
MX2	0.10	0.10	0.12	0.12	0.13	0.14	0.14	0.17
MX3	0.55	0.53	0.49	0.47	0.46	0.45	0.42	0.40
MX4	0.03	0.03	0.03	0.05	0.04	0.04	0.04	0.04

TABLE 5.11c

Power of the Test for Normality $T_1(c)$ at
Significance Level $\alpha = 0.025$, $n = 100$ and $p = 1$

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.28	0.28	0.29	0.28	0.28	0.27	0.27	0.27
T(7)	0.41	0.42	0.41	0.40	0.40	0.39	0.39	0.38
T(5)	0.64	0.65	0.64	0.63	0.62	0.62	0.61	0.61
T(3)	0.92	0.91	0.91	0.90	0.89	0.89	0.88	0.88
$X^2(14)$	0.25	0.25	0.25	0.24	0.23	0.23	0.23	0.22
$X^2(10)$	0.35	0.36	0.35	0.35	0.33	0.33	0.32	0.31
$X^2(6)$	0.54	0.53	0.52	0.50	0.48	0.47	0.46	0.44
$X^2(4)$	0.75	0.72	0.69	0.67	0.65	0.64	0.62	0.60
$X^2(2)$	0.96	0.94	0.92	0.91	0.89	0.89	0.87	0.86
MX1	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MX2	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.04
MX3	0.46	0.42	0.39	0.37	0.36	0.35	0.34	0.33
MX4	0.01	0.01	0.01	0.03	0.03	0.03	0.03	0.01

TABLE 5.11d

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.01$, $n = 100$ and $p = 1$

	c							
Alternative	0.3	0.2	0.1	0.05	0.01	-0.01	-0.05	-0.1
Lognormal	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.98
T(9)	0.21	0.22	0.21	0.20	0.20	0.20	0.19	0.19
T(7)	0.31	0.32	0.32	0.30	0.30	0.30	0.30	0.28
T(5)	0.57	0.56	0.54	0.53	0.52	0.51	0.51	0.50
T(3)	0.88	0.87	0.86	0.84	0.84	0.83	0.83	0.81
$X^2(14)$	0.17	0.18	0.18	0.16	0.16	0.16	0.15	0.15
$X^2(10)$	0.28	0.28	0.27	0.25	0.25	0.24	0.24	0.23
$X^2(6)$	0.46	0.45	0.43	0.40	0.39	0.38	0.37	0.35
$X^2(4)$	0.67	0.64	0.61	0.58	0.56	0.54	0.53	0.50
$X^2(2)$	0.94	0.92	0.89	0.86	0.84	0.83	0.81	0.79
MX1	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MX2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MX3	0.33	0.30	0.28	0.25	0.23	0.23	0.22	0.21
MX4	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00

TABLE 5.12a

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.1$, $n = 20$ and $p = 2$

	c							
Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	0.91	0.89	0.86	0.83	0.82	0.81	0.80	0.73
T(9)	0.27	0.27	0.26	0.23	0.23	0.23	0.22	0.20
T(7)	0.36	0.36	0.35	0.32	0.31	0.31	0.29	0.25
T(5)	0.45	0.45	0.43	0.40	0.39	0.39	0.38	0.32
T(3)	0.66	0.65	0.64	0.60	0.59	0.58	0.57	0.52
$X^2(14)$	0.24	0.24	0.24	0.21	0.21	0.21	0.21	0.19
$X^2(10)$	0.27	0.26	0.26	0.23	0.23	0.23	0.23	0.21
$X^2(6)$	0.38	0.37	0.36	0.33	0.32	0.32	0.31	0.27
$X^2(4)$	0.48	0.46	0.45	0.41	0.40	0.39	0.38	0.35
$X^2(2)$	0.71	0.68	0.64	0.59	0.59	0.58	0.56	0.48
MX1	0.14	0.14	0.13	0.12	0.12	0.11	0.11	0.11
MX2	0.05	0.06	0.07	0.11	0.12	0.13	0.14	0.21
MX3	0.33	0.29	0.27	0.23	0.22	0.22	0.20	0.17
MX4	0.13	0.13	0.13	0.12	0.12	0.12	0.11	0.11

TABLE 5.12b

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.05$, $n = 20$ and $p = 2$

	c							
Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	0.87	0.84	0.80	0.77	0.75	0.75	0.74	0.70
T(9)	0.18	0.18	0.17	0.17	0.17	0.17	0.17	0.15
T(7)	0.25	0.25	0.24	0.22	0.22	0.22	0.22	0.19
T(5)	0.35	0.35	0.34	0.32	0.32	0.31	0.31	0.28
T(3)	0.57	0.56	0.54	0.52	0.52	0.52	0.52	0.48
$X^2(14)$	0.16	0.17	0.16	0.16	0.16	0.16	0.16	0.14
$X^2(10)$	0.18	0.17	0.17	0.16	0.16	0.16	0.16	0.15
$X^2(6)$	0.28	0.28	0.26	0.23	0.23	0.23	0.23	0.21
$X^2(4)$	0.36	0.36	0.34	0.32	0.31	0.31	0.31	0.28
$X^2(2)$	0.62	0.58	0.54	0.49	0.49	0.48	0.47	0.42
MX1	0.08	0.07	0.07	0.07	0.07	0.07	0.07	0.06
MX2	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.08
MX3	0.23	0.19	0.17	0.15	0.15	0.14	0.14	0.12
MX4	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.07

TABLE 5.12c

Power of the Test for Normality $T_1(c)$ at
Significance Level $\alpha = 0.025$, $n = 20$ and $p = 2$

c

Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	0.84	0.80	0.74	0.70	0.70	0.69	0.68	0.66
T(9)	0.12	0.13	0.12	0.11	0.11	0.11	0.11	0.11
T(7)	0.17	0.18	0.17	0.16	0.16	0.16	0.16	0.15
T(5)	0.28	0.28	0.26	0.25	0.25	0.25	0.25	0.24
T(3)	0.50	0.48	0.47	0.45	0.44	0.44	0.44	0.43
$X^2(14)$	0.10	0.11	0.10	0.09	0.09	0.09	0.09	0.10
$X^2(10)$	0.12	0.13	0.12	0.11	0.11	0.11	0.11	0.10
$X^2(6)$	0.20	0.20	0.19	0.17	0.17	0.16	0.16	0.16
$X^2(4)$	0.29	0.29	0.27	0.24	0.24	0.23	0.23	0.22
$X^2(2)$	0.55	0.50	0.44	0.40	0.39	0.39	0.38	0.36
MX1	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
MX2	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.02
MX3	0.16	0.13	0.11	0.09	0.09	0.09	0.09	0.08
MX4	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.04

TABLE 5.12d

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.01$, n = 20 and p = 2

	c							
Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	0.79	0.71	0.65	0.61	0.61	0.61	0.61	0.58
T(9)	0.07	0.08	0.07	0.07	0.07	0.07	0.07	0.07
T(7)	0.11	0.11	0.10	0.10	0.10	0.10	0.10	0.10
T(5)	0.18	0.19	0.18	0.18	0.18	0.18	0.18	0.17
T(3)	0.40	0.39	0.36	0.35	0.35	0.35	0.35	0.35
$X^2(14)$	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.05
$X^2(10)$	0.07	0.08	0.07	0.07	0.07	0.07	0.07	0.06
$X^2(6)$	0.13	0.12	0.11	0.11	0.11	0.11	0.11	0.10
$X^2(4)$	0.20	0.18	0.16	0.16	0.16	0.16	0.15	0.15
$X^2(2)$	0.45	0.38	0.32	0.30	0.30	0.30	0.30	0.29
MX1	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
MX2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MX3	0.10	0.07	0.05	0.05	0.05	0.05	0.05	0.05
MX4	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

TABLE 5.13a

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.1$, $n = 50$ and $p = 2$

	c							
Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.98
T(9)	0.44	0.44	0.44	0.43	0.43	0.42	0.42	0.38
T(7)	0.58	0.58	0.57	0.56	0.55	0.55	0.54	0.49
T(5)	0.75	0.75	0.74	0.73	0.72	0.72	0.72	0.68
T(3)	0.94	0.94	0.93	0.92	0.92	0.92	0.92	0.90
$X^2(14)$	0.35	0.36	0.36	0.36	0.35	0.35	0.35	0.32
$X^2(10)$	0.45	0.46	0.44	0.43	0.43	0.42	0.41	0.38
$X^2(6)$	0.62	0.62	0.59	0.57	0.56	0.56	0.55	0.50
$X^2(4)$	0.75	0.74	0.71	0.69	0.68	0.68	0.67	0.62
$X^2(2)$	0.95	0.93	0.91	0.90	0.89	0.89	0.88	0.84
MX1	0.16	0.16	0.15	0.15	0.15	0.15	0.14	0.13
MX2	0.06	0.06	0.09	0.17	0.19	0.20	0.22	0.26
MX3	0.58	0.52	0.47	0.43	0.41	0.40	0.38	0.31
MX4	0.12	0.12	0.12	0.13	0.13	0.12	0.12	0.11

TABLE 5.13b

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.05$, $n = 50$ and $p = 2$

	c							
Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	1.00	0.99	0.99	0.98	0.98	0.98	0.98	0.97
T(9)	0.34	0.34	0.34	0.34	0.34	0.34	0.33	0.31
T(7)	0.47	0.48	0.48	0.46	0.46	0.46	0.45	0.42
T(5)	0.67	0.66	0.66	0.65	0.65	0.65	0.64	0.62
T(3)	0.92	0.91	0.90	0.90	0.90	0.89	0.89	0.88
$X^2(14)$	0.25	0.26	0.26	0.26	0.26	0.26	0.26	0.24
$X^2(10)$	0.35	0.35	0.34	0.33	0.33	0.32	0.32	0.30
$X^2(6)$	0.52	0.51	0.49	0.48	0.47	0.47	0.46	0.43
$X^2(4)$	0.68	0.66	0.64	0.62	0.61	0.60	0.60	0.56
$X^2(2)$	0.92	0.90	0.87	0.84	0.84	0.83	0.83	0.79
MX1	0.10	0.08	0.08	0.09	0.09	0.09	0.09	0.08
MX2	0.02	0.03	0.03	0.03	0.03	0.04	0.04	0.07
MX3	0.47	0.41	0.33	0.30	0.29	0.29	0.28	0.24
MX4	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07

TABLE 5.13c

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.025$, $n = 50$ and $p = 2$

	c							
Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	0.99	0.99	0.98	0.97	0.97	0.97	0.97	0.96
T(9)	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.25
T(7)	0.39	0.40	0.40	0.38	0.38	0.38	0.38	0.37
T(5)	0.58	0.59	0.58	0.56	0.56	0.56	0.55	0.54
T(3)	0.88	0.87	0.86	0.85	0.85	0.85	0.85	0.83
$X^2(14)$	0.18	0.19	0.20	0.19	0.18	0.18	0.18	0.18
$X^2(10)$	0.26	0.27	0.26	0.25	0.25	0.25	0.24	0.23
$X^2(6)$	0.43	0.43	0.41	0.39	0.38	0.38	0.38	0.36
$X^2(4)$	0.60	0.59	0.56	0.53	0.53	0.53	0.52	0.49
$X^2(2)$	0.90	0.86	0.82	0.79	0.78	0.78	0.77	0.73
MX1	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05
MX2	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
MX3	0.37	0.30	0.24	0.21	0.20	0.19	0.19	0.16
MX4	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

TABLE 5.13d

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.01$, $n = 50$ and $p = 2$

	c							
Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	0.99	0.98	0.97	0.96	0.95	0.95	0.95	0.93
T(9)	0.17	0.19	0.19	0.19	0.19	0.19	0.18	0.18
T(7)	0.30	0.30	0.30	0.29	0.29	0.29	0.29	0.28
T(5)	0.48	0.48	0.47	0.46	0.46	0.45	0.45	0.44
T(3)	0.83	0.82	0.80	0.79	0.78	0.78	0.78	0.77
$X^2(14)$	0.12	0.13	0.13	0.12	0.12	0.12	0.12	0.12
$X^2(10)$	0.18	0.19	0.18	0.18	0.18	0.18	0.18	0.17
$X^2(6)$	0.33	0.33	0.32	0.31	0.30	0.30	0.30	0.28
$X^2(4)$	0.50	0.49	0.47	0.44	0.43	0.43	0.43	0.40
$X^2(2)$	0.84	0.80	0.75	0.71	0.70	0.69	0.68	0.64
MX1	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
MX2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
MX3	0.27	0.20	0.15	0.12	0.12	0.11	0.11	0.09
MX4	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

TABLE 5.14a

Power of the Test for Normality $T_1(c)$ at
Significance Level $\alpha = 0.1$, $n = 100$ and $p = 2$

	c							
Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.68	0.67	0.66	0.65	0.64	0.64	0.64	0.60
T(7)	0.81	0.79	0.78	0.78	0.77	0.77	0.77	0.74
T(5)	0.94	0.94	0.93	0.92	0.92	0.92	0.91	0.90
T(3)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$X^2(14)$	0.48	0.48	0.48	0.47	0.47	0.46	0.45	0.42
$X^2(10)$	0.61	0.60	0.59	0.58	0.57	0.57	0.57	0.53
$X^2(6)$	0.83	0.82	0.80	0.78	0.77	0.77	0.76	0.72
$X^2(4)$	0.93	0.93	0.91	0.89	0.88	0.88	0.87	0.85
$X^2(2)$	1.00	1.00	0.99	0.99	0.99	0.99	0.98	0.98
MX1	0.23	0.22	0.22	0.22	0.22	0.22	0.22	0.20
MX2	0.30	0.30	0.33	0.40	0.41	0.42	0.43	0.48
MX3	0.83	0.77	0.68	0.61	0.59	0.57	0.55	0.47
MX4	0.18	0.17	0.17	0.18	0.18	0.18	0.18	0.18

TABLE 5.14b

Power of the Test for Normality $T_1(c)$ at
Significance Level $\alpha = 0.05$, $n = 100$ and $p = 2$

c

Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.58	0.58	0.59	0.55	0.55	0.55	0.54	0.59
T(7)	0.74	0.73	0.73	0.70	0.69	0.69	0.69	0.73
T(5)	0.90	0.90	0.90	0.88	0.88	0.88	0.88	0.90
T(3)	1.00	1.00	1.00	0.99	0.99	0.99	0.99	1.00
$X^2(14)$	0.38	0.39	0.40	0.37	0.36	0.36	0.36	0.40
$X^2(10)$	0.51	0.51	0.52	0.49	0.48	0.47	0.47	0.52
$X^2(6)$	0.75	0.75	0.75	0.71	0.70	0.70	0.69	0.75
$X^2(4)$	0.90	0.89	0.88	0.85	0.84	0.84	0.83	0.88
$X^2(2)$	1.00	0.99	0.99	0.98	0.98	0.98	0.98	0.99
MX1	0.14	0.15	0.16	0.14	0.14	0.14	0.14	0.18
MX2	0.06	0.06	0.10	0.09	0.09	0.10	0.10	0.15
MX3	0.76	0.67	0.59	0.48	0.45	0.45	0.42	0.34
MX4	0.10	0.11	0.13	0.12	0.12	0.12	0.12	0.11

TABLE 5.14c

Power of the Test for Normality $T_1(c)$ at
Significance Level $\alpha = 0.025$, $n = 100$ and $p = 2$

c

Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.47	0.49	0.47	0.45	0.44	0.44	0.44	0.42
T(7)	0.65	0.66	0.65	0.62	0.61	0.60	0.60	0.58
T(5)	0.87	0.87	0.85	0.83	0.82	0.82	0.82	0.80
T(3)	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.98
$X^2(14)$	0.29	0.31	0.31	0.29	0.29	0.28	0.28	0.27
$X^2(10)$	0.42	0.44	0.43	0.40	0.40	0.39	0.39	0.36
$X^2(6)$	0.68	0.69	0.66	0.63	0.62	0.61	0.61	0.56
$X^2(4)$	0.86	0.86	0.83	0.79	0.79	0.78	0.77	0.73
$X^2(2)$	0.99	0.99	0.98	0.97	0.97	0.96	0.96	0.94
MX1	0.09	0.10	0.10	0.09	0.09	0.09	0.09	0.08
MX2	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.02
MX3	0.65	0.58	0.45	0.34	0.32	0.30	0.30	0.24
MX4	0.06	0.07	0.08	0.07	0.07	0.07	0.07	0.07

TABLE 5.14d

Power of the Test for Normality $T_1(c)$ at
Significance Level α = 0.01, n = 100 and p = 2

	c							
Alternative	0.3	0.2	0.1	0.025	0.006	-0.006	-0.025	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.38	0.37	0.36	0.33	0.33	0.33	0.32	0.32
T(7)	0.58	0.57	0.55	0.52	0.51	0.51	0.51	0.50
T(5)	0.82	0.80	0.79	0.75	0.75	0.75	0.74	0.72
T(3)	0.99	0.99	0.98	0.98	0.98	0.97	0.97	0.97
$X^2(14)$	0.22	0.22	0.22	0.20	0.19	0.19	0.19	0.18
$X^2(10)$	0.35	0.35	0.34	0.31	0.30	0.30	0.29	0.28
$X^2(6)$	0.61	0.59	0.56	0.51	0.50	0.50	0.48	0.46
$X^2(4)$	0.82	0.79	0.76	0.70	0.69	0.68	0.68	0.65
$X^2(2)$	0.99	0.99	0.98	0.94	0.94	0.93	0.93	0.94
MX1	0.06	0.06	0.05	0.05	0.04	0.04	0.04	0.04
MX2	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.01
MX3	0.57	0.44	0.31	0.20	0.20	0.18	0.17	0.14
MX4	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.04

TABLE 5.15a

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.1$, $n = 50$ and $p = 5$

	c							
Alternative	0.3	0.2	0.1	0.004	0.002	-0.002	-0.004	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.74	0.75	0.74	0.71	0.71	0.71	0.71	0.68
T(7)	0.85	0.86	0.85	0.83	0.83	0.83	0.83	0.79
T(5)	0.96	0.96	0.95	0.94	0.94	0.94	0.94	0.92
T(3)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
$X^2(14)$	0.35	0.40	0.39	0.37	0.37	0.37	0.37	0.36
$X^2(10)$	0.47	0.53	0.52	0.49	0.49	0.49	0.49	0.47
$X^2(6)$	0.70	0.74	0.70	0.65	0.65	0.65	0.65	0.61
$X^2(4)$	0.87	0.88	0.84	0.80	0.80	0.80	0.80	0.75
$X^2(2)$	0.99	0.99	0.98	0.97	0.97	0.97	0.97	0.94
MX1	0.64	0.48	0.41	0.34	0.34	0.34	0.34	0.30
MX2	0.06	0.07	0.08	0.10	0.10	0.11	0.11	0.24
MX3	0.96	0.94	0.91	0.86	0.86	0.86	0.86	0.75
MX4	0.36	0.36	0.34	0.31	0.31	0.31	0.31	0.29

TABLE 5.15b

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.05$, $n = 50$ and $p = 5$

	c							
Alternative	0.3	0.2	0.1	0.004	0.002	-0.002	-0.004	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.61	0.67	0.62	0.61	0.61	0.61	0.61	0.58
T(7)	0.77	0.81	0.78	0.75	0.75	0.75	0.75	0.71
T(5)	0.92	0.94	0.92	0.91	0.91	0.91	0.91	0.89
T(3)	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.98
$X^2(14)$	0.22	0.30	0.28	0.27	0.27	0.27	0.27	0.26
$X^2(10)$	0.32	0.41	0.37	0.36	0.36	0.36	0.36	0.34
$X^2(6)$	0.57	0.63	0.58	0.54	0.54	0.54	0.54	0.50
$X^2(4)$	0.79	0.81	0.75	0.71	0.71	0.71	0.71	0.66
$X^2(2)$	0.99	0.98	0.96	0.93	0.93	0.93	0.93	0.90
MX1	0.55	0.38	0.27	0.23	0.23	0.23	0.23	0.21
MX2	0.03	0.04	0.04	0.04	0.05	0.05	0.05	0.08
MX3	0.92	0.89	0.82	0.75	0.75	0.74	0.74	0.63
MX4	0.25	0.25	0.22	0.21	0.21	0.20	0.20	0.20

TABLE 5.15c

Power of the Test for Normality $T_1(c)$ at
Significance Level $\alpha = 0.025$, n = 50 and p = 5

c

Alternative	0.3	0.2	0.1	0.004	0.002	-0.002	-0.004	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.49	0.57	0.54	0.52	0.52	0.52	0.52	0.47
T(7)	0.66	0.73	0.69	0.66	0.66	0.66	0.66	0.62
T(5)	0.88	0.90	0.89	0.87	0.87	0.87	0.87	0.83
T(3)	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98
$X^2(14)$	0.14	0.20	0.20	0.19	0.19	0.19	0.19	0.18
$X^2(10)$	0.22	0.29	0.28	0.27	0.27	0.27	0.27	0.25
$X^2(6)$	0.44	0.52	0.47	0.44	0.44	0.44	0.44	0.39
$X^2(4)$	0.70	0.72	0.67	0.62	0.61	0.61	0.61	0.55
$X^2(2)$	0.97	0.97	0.93	0.89	0.89	0.89	0.89	0.85
MX1	0.48	0.27	0.19	0.16	0.16	0.16	0.16	0.13
MX2	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.03
MX3	0.86	0.82	0.71	0.61	0.61	0.61	0.61	0.36
MX4	0.17	0.18	0.15	0.14	0.14	0.14	0.14	0.12

TABLE 5.15d

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.01$, $n = 50$ and $p = 5$

	c							
Alternative	0.3	0.2	0.1	0.004	0.002	-0.002	-0.004	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
T(9)	0.36	0.43	0.40	0.41	0.41	0.40	0.40	0.37
T(7)	0.53	0.60	0.57	0.56	0.56	0.56	0.56	0.52
T(5)	0.80	0.85	0.82	0.80	0.80	0.79	0.79	0.76
T(3)	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.96
$X^2(14)$	0.07	0.12	0.12	0.13	0.13	0.13	0.13	0.11
$X^2(10)$	0.13	0.19	0.19	0.19	0.19	0.19	0.19	0.16
$X^2(6)$	0.31	0.38	0.34	0.33	0.33	0.33	0.33	0.29
$X^2(4)$	0.58	0.61	0.53	0.51	0.51	0.50	0.50	0.45
$X^2(2)$	0.95	0.93	0.88	0.84	0.84	0.84	0.84	0.78
MX1	0.43	0.17	0.11	0.10	0.10	0.10	0.10	0.08
MX2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MX3	0.77	0.72	0.53	0.45	0.45	0.45	0.45	0.36
MX4	0.10	0.10	0.08	0.08	0.08	0.08	0.08	0.07

TABLE 5.16a

Power of the Test for Normality $T_1(c)$ at
Significance Level α = 0.1, n = 100 and p = 5

c

Alternative	0.3	0.2	0.1	0.004	0.002	-0.002	-0.004	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.94	0.94	0.93	0.92	0.92	0.92	0.92	0.89
T(7)	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.97
T(5)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(3)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$X^2(14)$	0.55	0.59	0.58	0.55	0.55	0.55	0.55	0.52
$X^2(10)$	0.72	0.75	0.73	0.69	0.69	0.69	0.68	0.64
$X^2(6)$	0.92	0.92	0.90	0.87	0.87	0.87	0.87	0.83
$X^2(4)$	0.98	0.98	0.98	0.96	0.96	0.96	0.96	0.93
$X^2(2)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MX1	0.81	0.69	0.58	0.50	0.49	0.49	0.49	0.43
MX2	0.13	0.14	0.15	0.19	0.18	0.19	0.19	0.35
MX3	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.97
MX4	0.54	0.52	0.48	0.46	0.45	0.45	0.45	0.41

TABLE 5.16b

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.05$, $n = 100$ and $p = 5$

	c							
Alternative	0.3	0.2	0.1	0.004	0.002	-0.002	-0.004	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.90	0.91	0.90	0.88	0.88	0.88	0.87	0.85
T(7)	0.97	0.98	0.97	0.96	0.96	0.96	0.96	0.94
T(5)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
T(3)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$X^2(14)$	0.43	0.47	0.47	0.44	0.44	0.44	0.44	0.41
$X^2(10)$	0.60	0.63	0.63	0.59	0.59	0.59	0.58	0.54
$X^2(6)$	0.87	0.87	0.85	0.81	0.81	0.80	0.80	0.75
$X^2(4)$	0.97	0.97	0.96	0.93	0.93	0.93	0.93	0.90
$X^2(2)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
MX1	0.75	0.58	0.46	0.36	0.36	0.36	0.36	0.31
MX2	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.11
MX3	1.00	1.00	0.99	0.98	0.98	0.98	0.97	0.94
MX4	0.41	0.40	0.38	0.32	0.32	0.32	0.32	0.28

TABLE 5.16c

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.025$, $n = 100$ and $p = 5$

c

Alternative	0.3	0.2	0.1	0.004	0.002	-0.002	-0.004	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.85	0.86	0.86	0.81	0.81	0.81	0.81	0.78
T(7)	0.95	0.96	0.95	0.93	0.93	0.93	0.93	0.91
T(5)	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.98
T(3)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$X^2(14)$	0.32	0.36	0.37	0.34	0.34	0.34	0.34	0.31
$X^2(10)$	0.49	0.55	0.54	0.49	0.49	0.49	0.49	0.44
$X^2(6)$	0.80	0.82	0.80	0.72	0.72	0.72	0.72	0.66
$X^2(4)$	0.95	0.95	0.93	0.90	0.90	0.90	0.90	0.84
$X^2(2)$	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99
MX1	0.67	0.48	0.37	0.27	0.27	0.27	0.26	0.21
MX2	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.04
MX3	1.00	1.00	0.99	0.95	0.95	0.95	0.95	0.88
MX4	0.32	0.31	0.28	0.22	0.22	0.22	0.22	0.20

TABLE 5.16d

Power of the Test for Normality $T_1(c)$ at Significance Level $\alpha = 0.01$, n = 100 and p = 5

	c							
Alternative	0.3	0.2	0.1	0.004	0.002	-0.002	-0.004	-0.1
Lognormal	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
T(9)	0.77	0.80	0.77	0.74	0.74	0.73	0.73	0.69
T(7)	0.91	0.93	0.92	0.88	0.88	0.88	0.88	0.84
T(5)	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.97
T(3)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$X^2(14)$	0.22	0.27	0.27	0.25	0.25	0.25	0.25	0.21
$X^2(10)$	0.38	0.43	0.41	0.37	0.37	0.37	0.37	0.32
$X^2(6)$	0.71	0.75	0.70	0.63	0.62	0.62	0.62	0.56
$X^2(4)$	0.92	0.93	0.89	0.82	0.82	0.82	0.82	0.76
$X^2(2)$	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.97
MX1	0.58	0.36	0.24	0.18	0.18	0.18	0.18	0.13
MX2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02
MX3	0.99	0.99	0.96	0.89	0.89	0.88	0.88	0.76
MX4	0.22	0.22	0.18	0.15	0.15	0.15	0.15	0.11

5.4 Analysis of Linear Models

In this section, it will be shown that the multivariate normality test statistic $T_1(c)$ can be applied to models such as linear regression that have additional structure. This is an extraordinary new result since most other tests for normality cannot be used when the data have structure other than a mean and covariance. The Shapiro-Wilk test has been applied to two-way layouts (see Gentleman and Wilk, 1975); however, for the particular two-way layout, percentage points were tabulated for the test statistic W. It is not practical to tabulate percentage points for each model considered. By adjusting c, it will be seen that the test statistic $T_1(c)$ can be made insensitive to the underlying model. In this way, $T_1(c)$ will provide an approximate test of normality for the residuals from a linear model. This result is especially useful for multivariate models where probability plots are not as effective.

In Section 2.3, a number of models for the errors were presented with the errors having the form

$$\epsilon_i = y_i - h(x_i;\Theta) \tag{4.1}$$

where

- ϵ_i is a p x 1 vector of errors,
- y_i is a p x 1 vector of observations,
- x_i is a q x 1 vector of concomitant variables, and
- Θ is a q x 1 vector of parameters to be estimated from the data.

The errors ϵ_i are assumed to be independent and identically distributed Gaussian random variables with zero mean and covariance matrix R. It is this assumption that the statistic $T_1(c)$ will examine. If $h(x_i,\Theta) = m$, then (4.1) is the model used in section 5.2.

For the following, the density of the errors is

$$f(\epsilon_i) = |2\pi R|^{-1/2}\exp(-\epsilon_i^T R^{-1}\epsilon_i/2) \tag{4.2}$$

$$= |2\pi R|^{-1/2}\exp\left[-(y_i - h(x_i;\Theta))^T R^{-1}(y_i - h(x_i;\Theta))/2\right].$$

The generalized likelihood $L_c(\Theta)$ without the constant term, denoted L(c) as in Section 2.3, is

$$L(c) = \left(\frac{1}{c}\right)\sum_{i=1}^{n}\left|\frac{1+c}{2\pi R}\right|^{a}\exp(-c(y_i-h(x_i;\Theta))^T R^{-1}(y_i-h(x_i;\Theta))/2) \tag{4.3}$$

$$= \left(\frac{1}{c}\right)\sum_{i=1}^{n}\left|\frac{1+c}{2\pi R}\right|^{a}\exp(-c\epsilon_i^T R^{-1}\epsilon_i/2)$$

where $a = 0.5c/(1+c)$.

For a proposed model $h(x;\Theta)$ in L(c), model-critical estimates of Θ and R are obtained by maximizing (4.3) over Θ and R. For many models, setting equal to zero the derivatives of L(c) with respect to Θ and R yields a set of implicit equations which can be solved via a fixed point algorithm (see Sections 2.2 and 2.3). As with the unstructured case, estimates R(o) and R(c) of R can be used to obtain the test statistic

$$\tilde{T}_1(c) = n/2 \left\{ \mathrm{tr}[R(c)R(o)^{-1} + R(o)R(c)^{-1}] - 2p \right\} . \tag{4.4}$$

The tilde is used to indicate the test statistic obtained from a structured model. Since $\tilde{T}_1(c)$ is defined similarly to $T_1(c)$, the statistic $\tilde{T}_1(c)$ or a function of $\tilde{T}_1(c)$ may have approximately the same distribution as $T_1(c)$. The definition of $\tilde{T}_1(c)$ indicates that the effects on R(o) and R(c) due to estimating additional parameters should approximately cancel. If $T_1(c)$ and $\tilde{T}_1(c)$ have the same distribution, $\tilde{T}_1(c)$ could be used to test the normality of the residuals in a linear model using the percentage points of $T_1(c)$. This would eliminate the need to tabulate percentage points of $\tilde{T}_1(c)$ for each model.

5.5 Monte Carlo Analysis of $\tilde{T}_1(c)$

Monte Carlo simulation was used to analyze $\tilde{T}_1(c)$ as a test of normality of the residuals from a linear model, the emphasis being on univariate models. For linear regressions, autoregressions, and two-way layouts, percentage points based on 10,000 trials were obtained as a function of c and sample size. For linear regressions, one to seven parameter models were examined for sample sizes of 32, 64, and 128. The autoregressive models had one to eight parameters and sample sizes of 40, 60, and 120. For both types of models, the c-values used were -0.05, -0.025, -0.01, 0.01, 0.025 and 0.05. For c = -0.05, -0.025, 0.025, and 0.05, Figures 5.1 to 5.12 are plots of the 90 and 95 percentage points of $\tilde{T}_1(c)$ as a function of model order for p-parameter linear regressions. The solid line in each plot is the tabulated percentage point of $T_1(c)$. The plots show that the percentage points increase (decrease) as the number of parameters increases for $c > o$ $(c < o)$. Also, the rate of increase $(c > o)$ or decrease $(c < o)$ increases with the magnitude of c. Thus, the critical parameter c controls the behavior of the percentage points of $\tilde{T}_1(c)$. That is, for a given model, the percentage points of $\tilde{T}_1(c)$ can be "tuned" to those of $T_1(c)$ by an appropriate choice of c. However, the change in percentage points does not dramatically alter the size of the test, since the percentage of observations of $\tilde{T}_1(c)$ exceeding the α^{th} percentage point of $T_1(c)$ is approximately $1 - \alpha$. Tables 5.17 to 5.22 present the size of the test results for p-parameter linear regressions and autoregressions. Tables 5.23 and

5.24 show the size of the test results for two-way layouts as a function of c. From the tables and the above discussion, it can be seen that the test $\tilde{T}_1(c)$ tends to be liberal for $c > o$ and conservative for $c < o$. In fact, as c decreases, there is a smooth transition from a liberal to a conservative test as seen in Table 5.24. As a general guideline, the value of $|c|$ should be decreased as the number of parameters increases, the dimension of the data increases, or the sample size decreases. The analysis shows that $T_1(c)$ and $\tilde{T}_1(c)$ have approximately the same distribution for small values of c; hereafter, $T_1(c)$ will denote the test statistic of (4.4) regardless of the underlying parametric model.

The power study for $T_1(c)$ in section 5.3 showed that the power of $T_1(c)$ decreases slightly as $|c|$ decreases. For univariate two-way layouts and autoregressions, the power of $T_1(c)$ for t and chi-square distributed errors was examined. Tables 5.25 to 5.27 present the power results for sample size 40, 60, and 120; as a reference, the power of $T_1(c)$ for unstructured data is also presented. The results indicate that the power of $T_1(c)$ for structured and unstructured models is approximately constant. The greatest loss of power is for two-way layout models with interaction; this is not entirely unexpected due to the correlation between residuals. The slight loss in power is acceptable in order to be able to test the normality of the residuals in a model. The real benefit of the test statistic $T_1(c)$ is in analyzing the residuals from multivariate models.

The choice of appropriate c value for testing is not clear. Even with the guidelines given previously, the choice of critical parameter is still more subjective than objective. Choosing an effective c value depends on the user's experience with model-critical methods. Table 5.28 lists c values which we feel are conservative for testing residuals. Although the choice of c can be troublesome, it is this flexibility of c which can make the test statistic $T_1(c)$ insensitive to the structural model; this makes $T_1(c)$ unique among tests for Gaussianity since no other test can be "tuned" to the model.

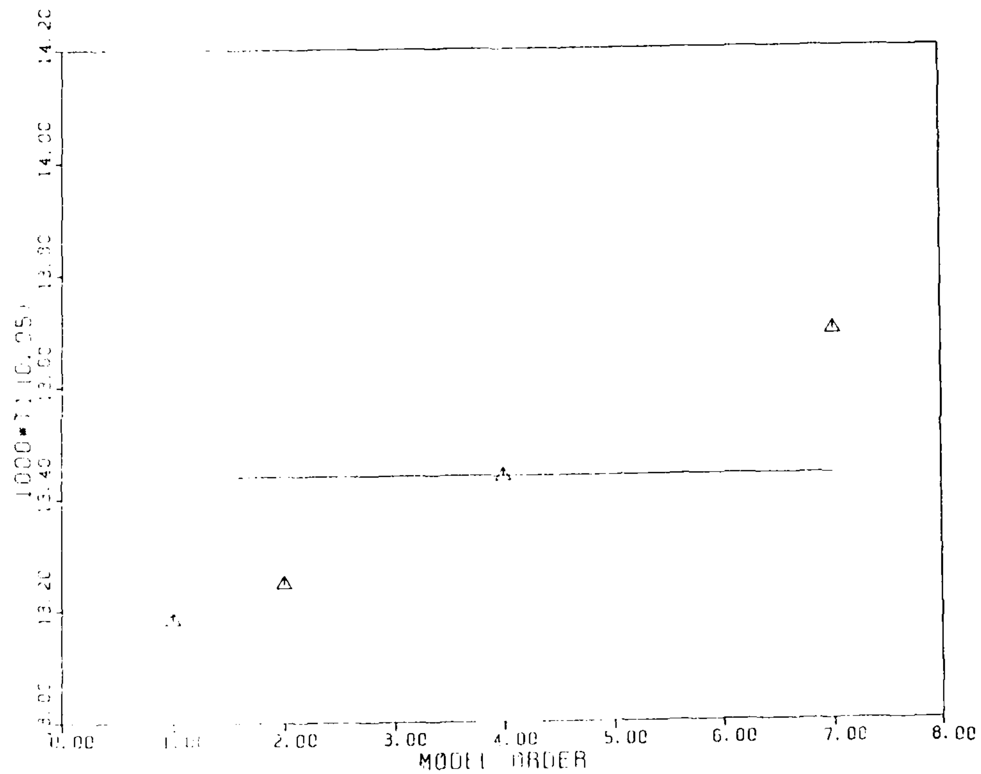

FIGURE 5.1. The 90 Percent Point of $T_1(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = 0.05

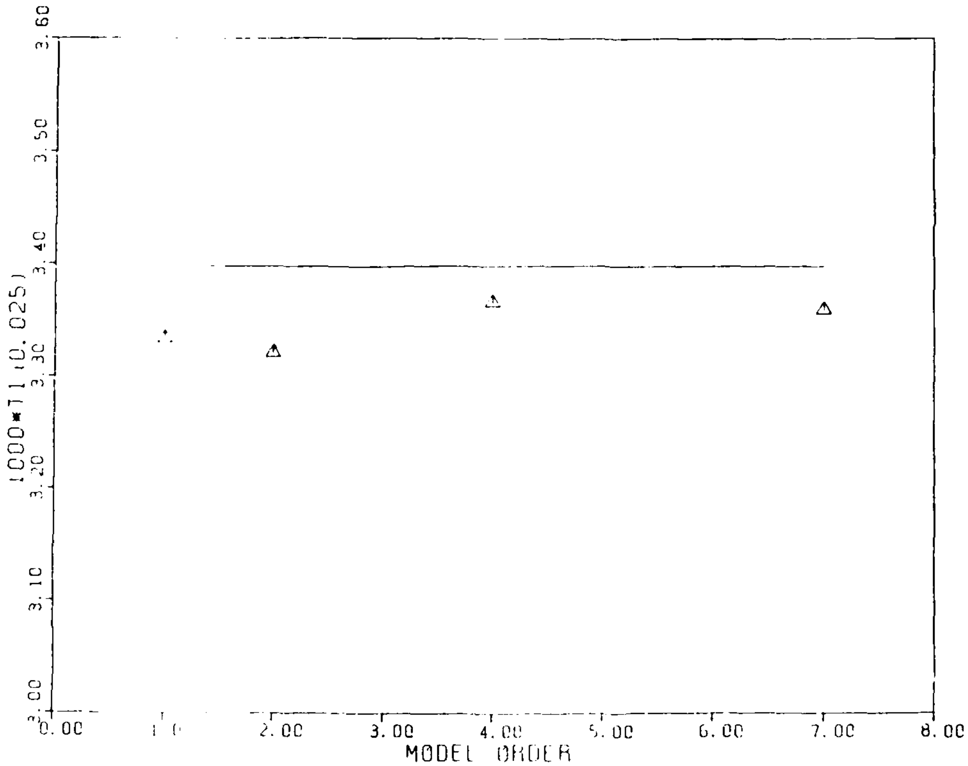

FIGURE 5.2. The 90 Percent Point of $T_1(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = 0.025

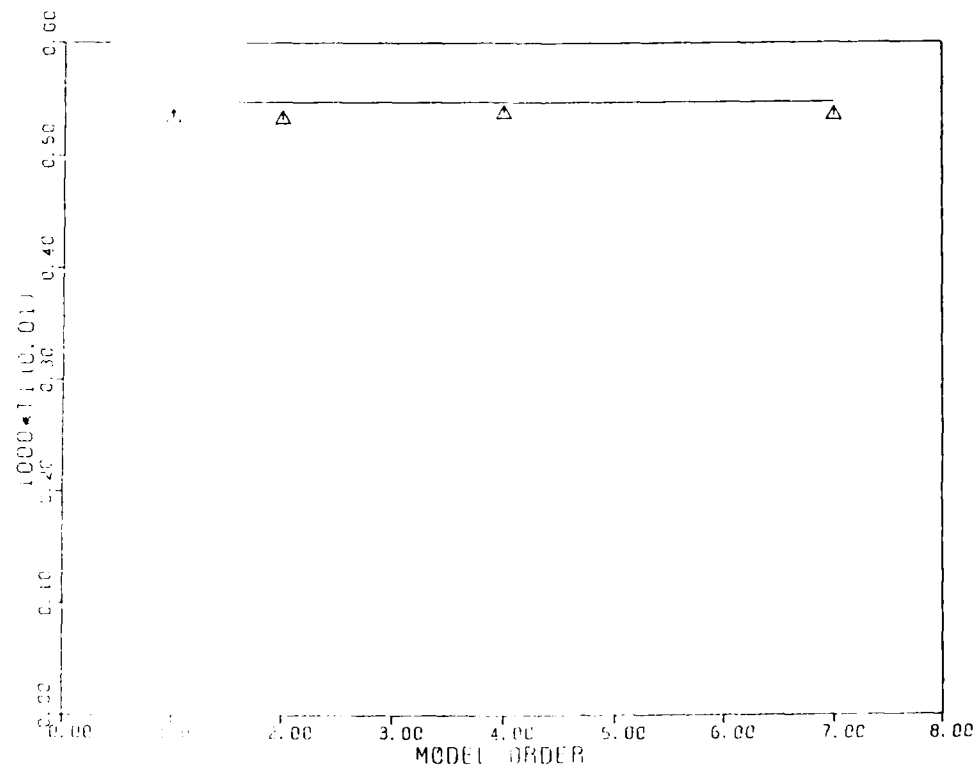

FIGURE 5.3. The 90 Percent Point of $T_I(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = 0.01

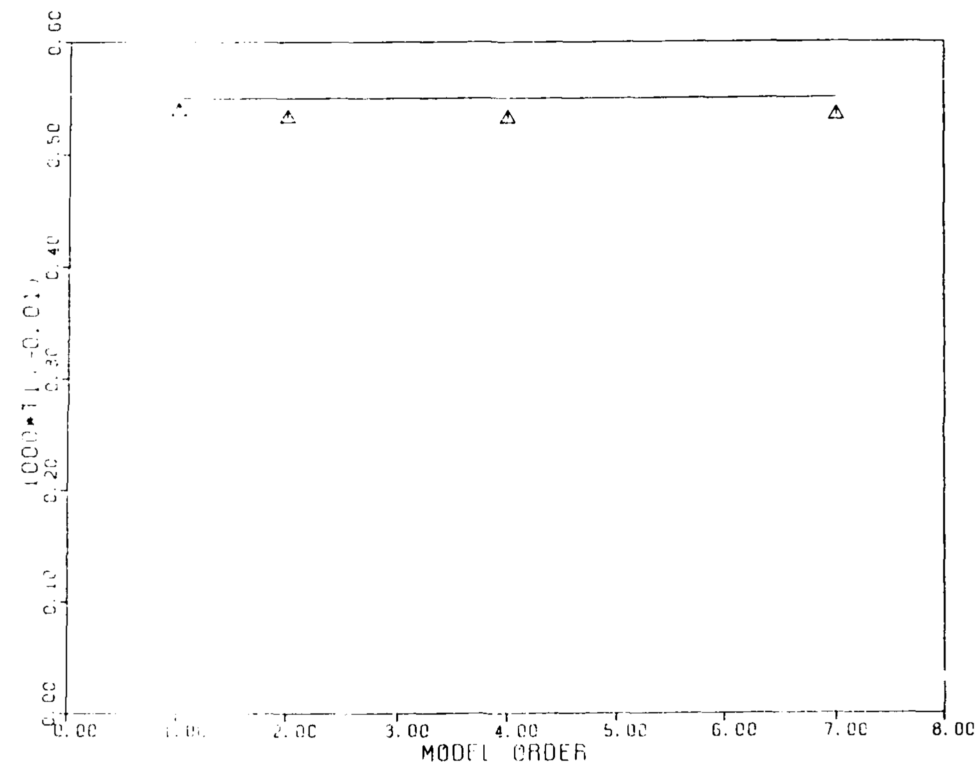

FIGURE 5.4. The 90 Percent Point of $T_1(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = -0.01

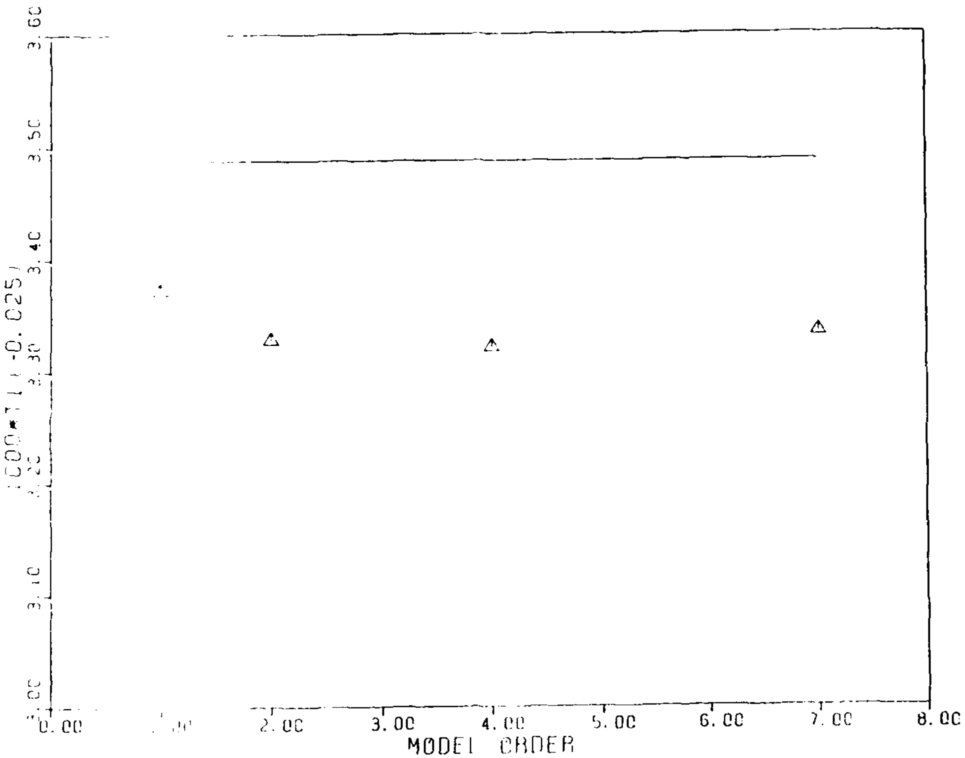

FIGURE 5.5. The 90 Percent Point of $T_1(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = -0.025

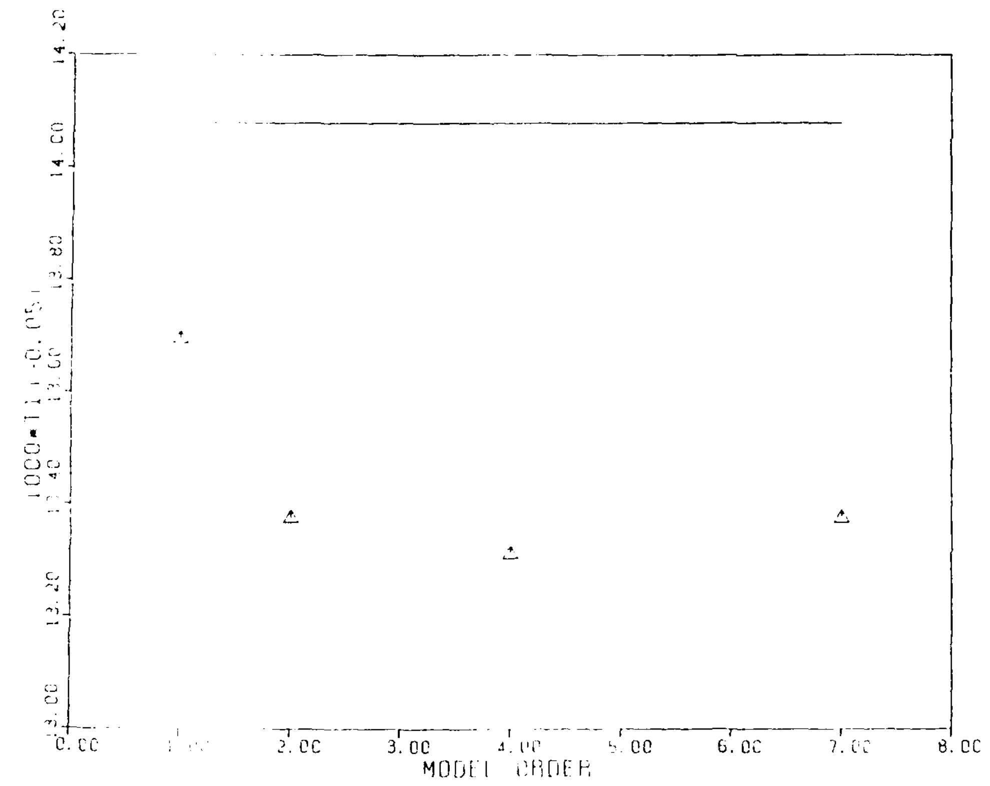

FIGURE 5.6. The 90 Percent Point of $T_1(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = -0.05

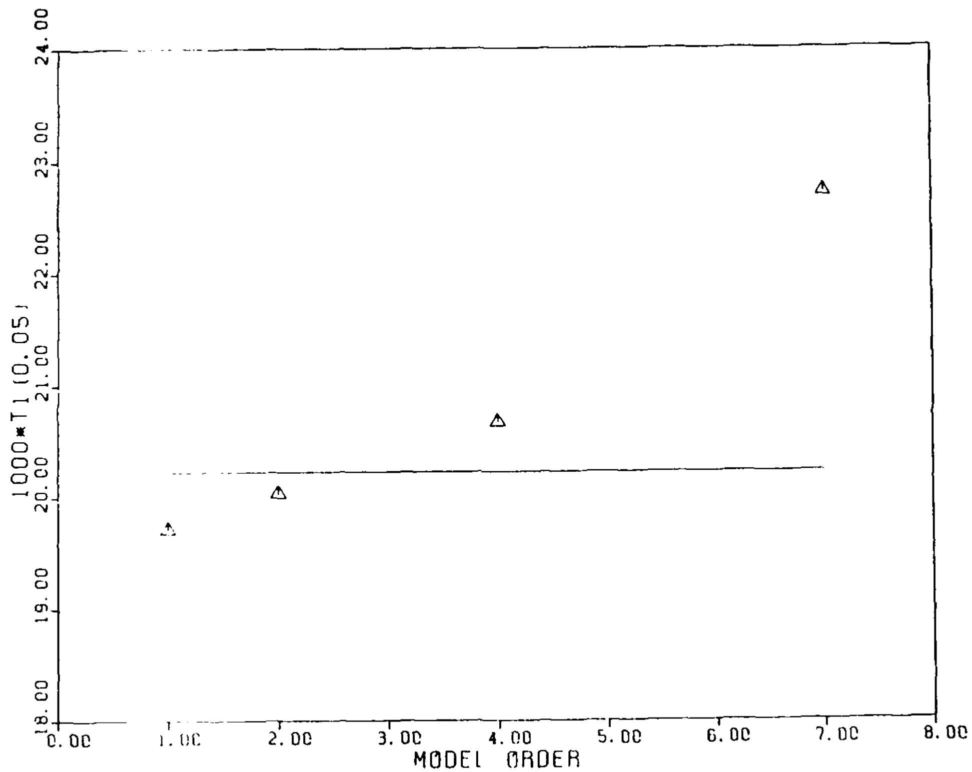

FIGURE 5.7. The 95 Percent Point of $T_1(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = 0.05

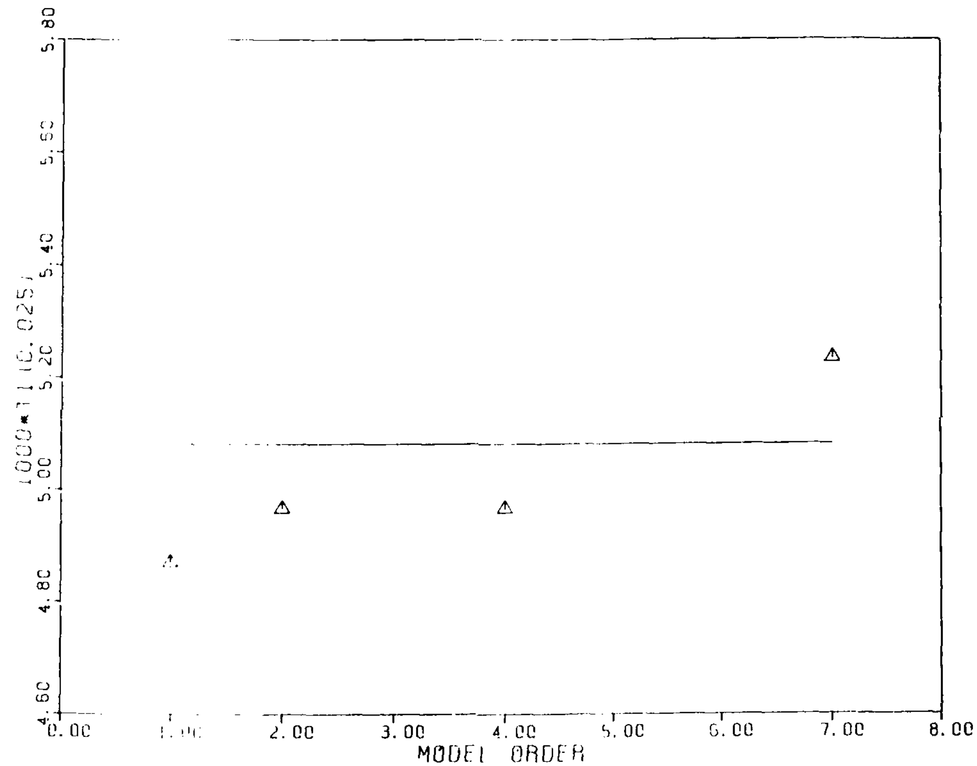

FIGURE 5.8. The 95 Percent Point of $T_1(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = 0.025

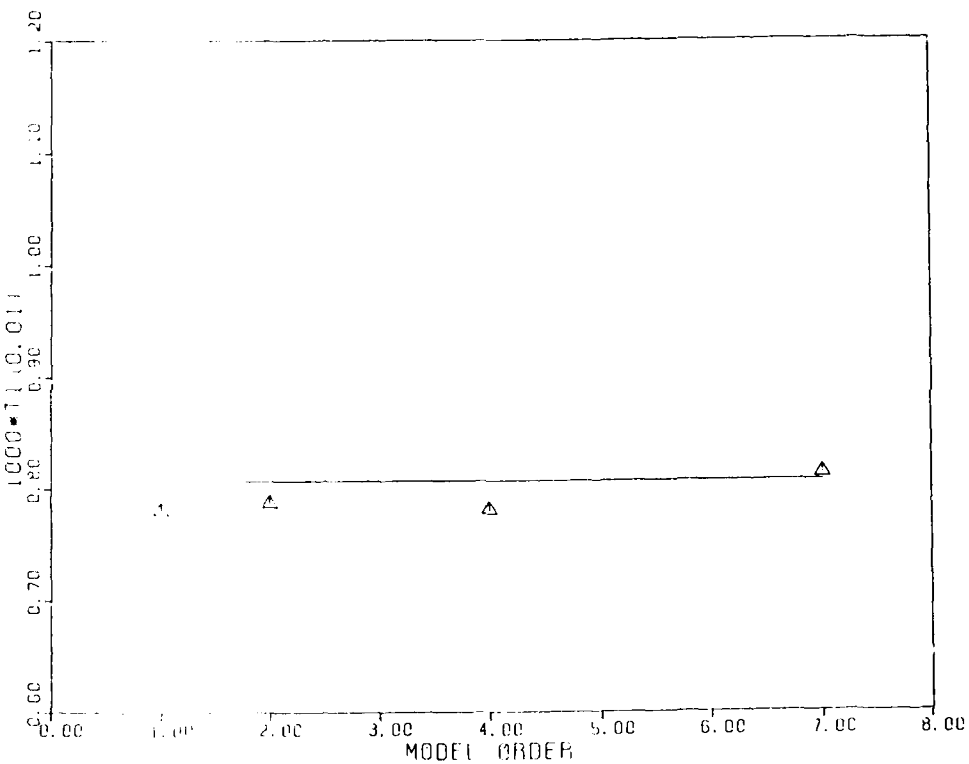

FIGURE 5.9. The 95 Percent Point of $T_I(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = 0.01

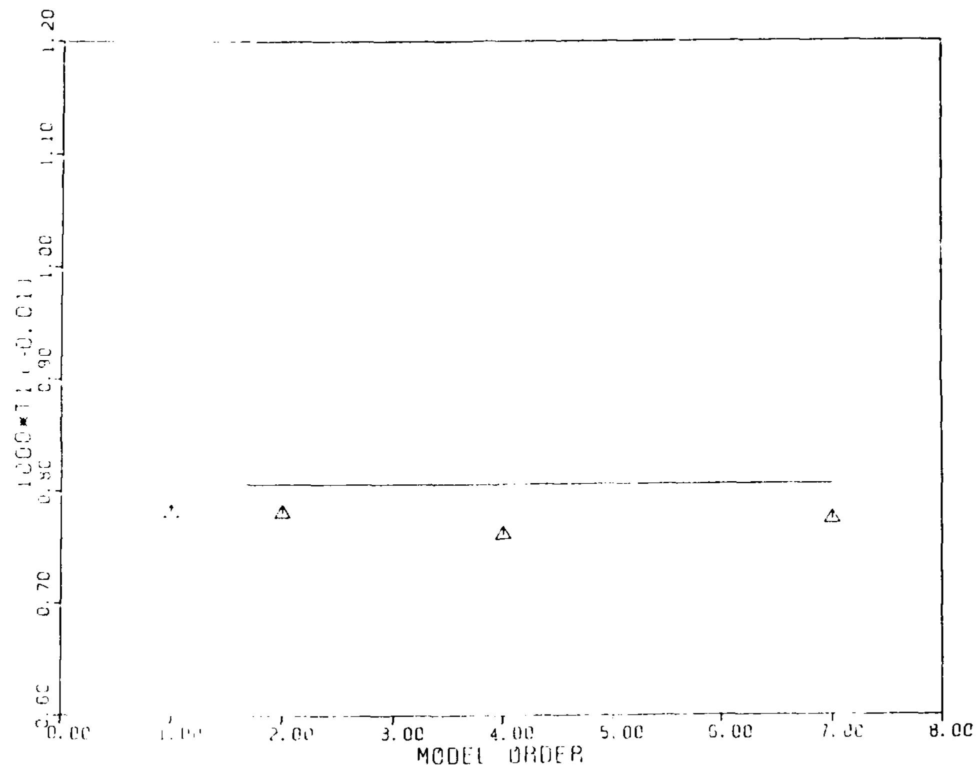

FIGURE 5.10. The 95 Percent Point of $T_1(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = -0.01

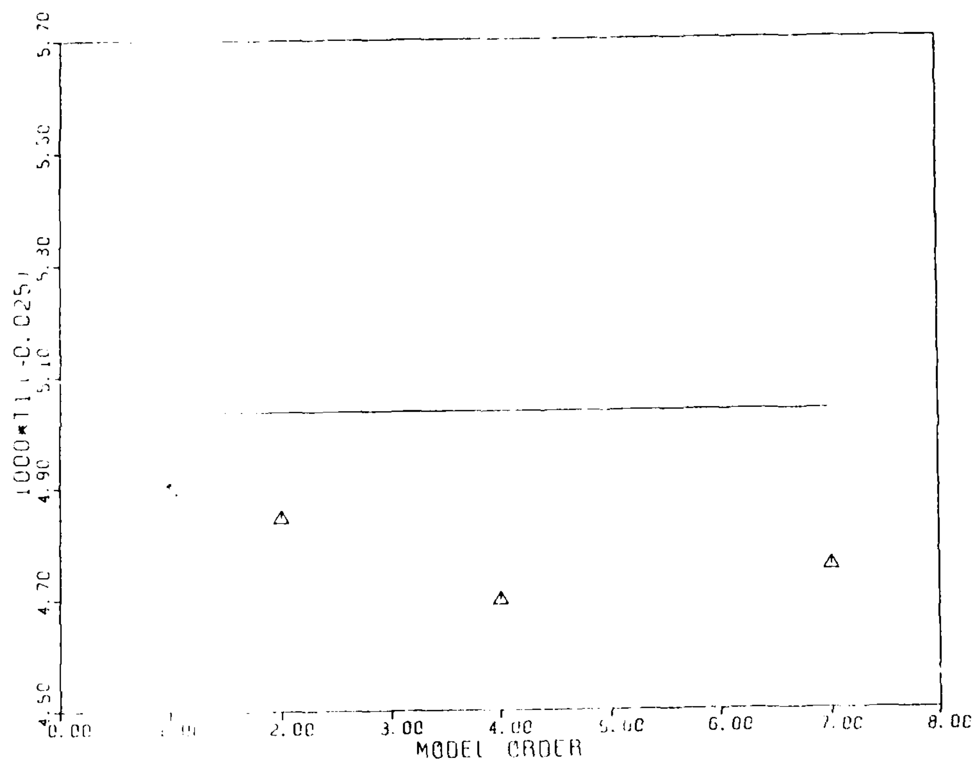

FIGURE 5.11. The 95 Percent Point of $T_1(c)$ Versus Model Order for p-Para ter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = -0.025

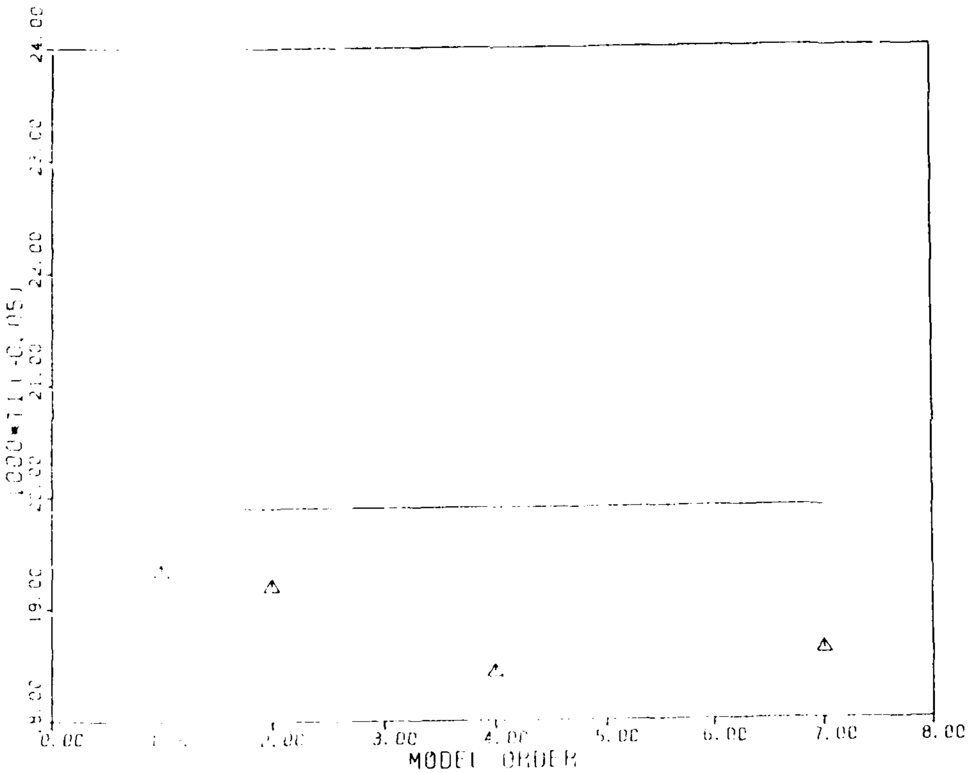

FIGURE 5.12. The 95 Percent Point of $T_1(c)$ Versus Model Order for p-Parameter Linear Regressions with Gaussian Errors, Sample Size = 64, and c = -0.05

TABLE 5.17a

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 7 Parameters, Sample Size = 128, and c = 0.025

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.099	0.051	0.025	0.009
2	0.099	0.051	0.026	0.010
4	0.100	0.051	0.025	0.010
7	0.104	0.052	0.027	0.011

TABLE 5.17b

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 7 Parameters, Sample Size = 128, and c = 0.01

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.100	0.051	0.026	0.009
2	0.098	0.050	0.026	0.010
4	0.098	0.049	0.025	0.010
7	0.104	0.050	0.026	0.010

TABLE 5.17c

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 7 Parameters, Sample Size = 128, and c = -0.01

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.098	0.051	0.025	0.009
2	0.097	0.050	0.025	0.009
4	0.097	0.050	0.025	0.009
7	0.101	0.049	0.025	0.009

TABLE 5.17d.

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 7 Parameters, Sample Size = 128, and c = -0.025

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.099	0.051	0.025	0.009
2	0.099	0.050	0.025	0.009
4	0.098	0.048	0.025	0.009
7	0.101	0.047	0.023	0.008

TABLE 5.18a.

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 7 Parameters, Sample Size = 64, and c = 0.025

Model Order	Size of the Test			
	0.1	0.05	0.025	0.01
1	0.097	0.047	0.023	0.010
2	0.096	0.048	0.025	0.010
4	0.098	0.048	0.026	0.011
7	0.097	0.052	0.029	0.011

TABLE 5.18b.

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 7 Parameters, Sample Size = 64, and c = 0.01

Model Order	Size of the Test			
	0.1	0.05	0.025	0.01
1	0.096	0.047	0.024	0.010
2	0.095	0.048	0.025	0.010
4	0.097	0.047	0.025	0.011
7	0.095	0.050	0.027	0.010

TABLE 5.18c

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 7 Parameters, Sample Size = 64, and c = -0.01

Model Order	Size of the Test 0.1	0.05	0.025	0.01
1	0.097	0.047	0.023	0.010
2	0.095	0.047	0.024	0.010
4	0.094	0.045	0.024	0.010
7	0.095	0.047	0.024	0.008

TABLE 5.18d

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 7 Parameters, Sample Size = 64, and c = -0.025

Model Order	Size of the Test 0.1	0.05	0.025	0.01
1	0.095	0.048	0.023	0.010
2	0.093	0.047	0.024	0.010
4	0.091	0.044	0.023	0.009
7	0.091	0.045	0.022	0.007

TABLE 5.19a

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 5 Parameters, Sample Size = 32, and c = 0.025

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.098	0.048	0.023	0.09
2	0.100	0.048	0.023	0.09
4	0.100	0.048	0.023	0.09
5	0.101	0.049	0.023	0.09

TABLE 5.19b

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 5 Parameters, Sample Size = 32, and c = 0.01

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.097	0.049	0.023	0.009
2	0.100	0.047	0.022	0.009
4	0.096	0.046	0.021	0.009
5	0.096	0.046	0.020	0.008

TABLE 5.19c

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 5 Parameters, Sample Size = 32, and c = -0.01

Model Order	Size of the Test 0.1	0.05	0.025	0.01
1	0.098	0.049	0.024	0.009
2	0.100	0.046	0.022	0.009
4	0.095	0.044	0.020	0.007
5	0.095	0.043	0.018	0.006

TABLE 5.19d.

Size of the Test $T_1(c)$ Versus Model Order for Linear Regressions with 1, 2, 4, and 5 Parameters, Sample Size = 32, and c = -0.025

Model Order	Size of the Test 0.1	0.05	0.025	0.01
1	0.101	0.048	0.024	0.009
2	0.103	0.046	0.022	0.009
4	0.097	0.042	0.018	0.007
5	0.098	0.040	0.016	0.005

TABLE 5.20a

Size of the Test $T_1(c)$ Versus Model Order for Autoregressions with 1, 2, 4, and 8 Parameters, Sample Size = 120, and c = 0.025

Model Order	Size of the Test 0.1	0.05	0.025	0.01
1	0.104	0.048	0.023	0.009
2	0.096	0.049	0.024	0.010
4	0.099	0.053	0.027	0.010
8	0.104	0.056	0.029	0.012

TABLE 5.20b

Size of the Test $T_1(c)$ Versus Model Order for Autoregressions with 1, 2, 4, and 8 Parameters, Sample Size = 120, and c = -0.025

Model Order	Size of the Test 0.1	0.05	0.025	0.01
1	0.105	0.049	0.023	0.009
2	0.096	0.048	0.023	0.009
4	0.099	0.052	0.025	0.009
8	0.102	0.050	0.024	0.009

TABLE 5.21a

Size of the Test $T_1(c)$ Versus Model Order for Autoregressions with 1, 2, 4, and 8 Parameters, Sample Size = 60, and c = 0.025

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.100	0.050	0.025	0.010
2	0.104	0.054	0.025	0.008
4	0.103	0.053	0.027	0.009
8	0.107	0.061	0.032	0.011

TABLE 5.21b

Size of the Test $T_1(c)$ Versus Model Order for Autoregressions with 1, 2, 4, and 8 Parameters, Sample Size = 60, and c = -0.025

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.098	0.049	0.025	0.009
2	0.103	0.051	0.024	0.008
4	0.099	0.048	0.024	0.008
8	0.102	0.052	0.052	0.008

TABLE 5.22a

Size of the Test $T_1(c)$ Versus Model Order for Autoregressions with 1, 2, 4, and 8 Parameters, Sample Size = 40, and c = 0.025

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.100	0.048	0.025	0.011
2	0.099	0.051	0.027	0.010
4	0.102	0.053	0.030	0.013
8	0.104	0.058	0.032	0.013

TABLE 5.22b

Size of the Test $T_1(c)$ Versus Model Order for Autoregressions with 1, 2, 4, and 8 Parameters, Sample Size = 40, and c = 0.01

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.100	0.047	0.024	0.011
2	0.100	0.051	0.027	0.010
4	0.101	0.052	0.029	0.013
8	0.101	0.054	0.029	0.012

TABLE 5.22c

Size of the Test $T_1(c)$ Versus Model Order for Autoregressions with 1, 2, 4, and 8 Parameters, Sample Size = 40, and c = -0.01

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.100	0.047	0.025	0.011
2	0.100	0.051	0.027	0.010
4	0.098	0.050	0.027	0.013
8	0.098	0.049	0.025	0.010

TABLE 5.22d

Size of the Test $T_1(c)$ Versus Model Order for Autoregressions with 1, 2, 4, and 8 Parameters, Sample Size = 40, and c = -0.025

	Size of the Test			
Model Order	0.1	0.05	0.025	0.01
1	0.102	0.047	0.026	0.010
2	0.100	0.049	0.027	0.010
4	0.099	0.048	0.026	0.012
8	0.096	0.046	0.023	0.010

TABLE 5.23

Size of the Test $T_1(c)$ Versus Critical Parameter for a 3 x 4 Two-Way Layout Without Interaction, Sample Size = 24

	Size of the Test			
c	0.1	0.05	0.025	0.01
0.05	0.104	0.054	0.027	0.011
0.025	0.097	0.046	0.023	0.008
0.01	0.095	0.043	0.020	0.007
-0.01	0.090	0.039	0.017	0.006
-0.025	0.091	0.037	0.015	0.005
-0.05	0.090	0.037	0.014	0.003

TABLE 5.24a

Size of the Test $T_1(c)$ Versus Critical Parameter for a 2-Dimensional, 3 x 4 Two-Way Layout with Interaction, Sample Size = 120

	Size of the Test			
c	0.1	0.05	0.025	0.01
0.025	0.107	0.052	0.025	0.011
0.008	0.102	0.046	0.021	0.008
-0.008	0.093	0.041	0.019	0.007
-0.025	0.088	0.036	0.017	0.006

TABLE 5.24b

Size of the Test $T_1(c)$ Versus Critical Parameter for a 2-Dimensional 3 x 4 Two-Way Layout with Interaction, Sample Size = 60

	Size of the Test			
c	0.1	0.05	0.025	0.01
0.025	0.110	0.055	0.026	0.010
0.008	0.097	0.045	0.021	0.007
-0.008	0.087	0.037	0.016	0.006
-0.025	0.077	0.031	0.013	0.004

TABLE 5.25

Power Results of the Statistic $T_1(c)$ for Unstructured Data (m); Autoregressive (AR) Models with 1, 2, and 4 Parameters; a Two-Way (TW) Layout Model with and Without Interaction; Sample Size is 120; and α = 0.05

c	m	AR(1)	AR(2)	AR(4)	TW(8)	TW(20)	
0.05	0.40	0.37	0.38	0.38	0.37	0.35	
0.025	0.40	0.37	0.38	0.37	0.36	0.34	
-0.025	0.39	0.37	0.37	0.36	0.34	0.30	t(9)
-0.05	0.39	0.36	0.37	0.35	0.32	0.28	
0.05	0.76	0.75	0.75	0.75	0.74	0.70	
0.025	0.76	0.75	0.74	0.74	0.73	0.69	t(5)
-0.025	0.75	0.74	0.73	0.72	0.70	0.65	
-0.05	0.74	0.73	0.72	0.71	0.69	0.64	
0.05	0.43	0.41	0.41	0.40	0.43	0.42	
0.025	0.43	0.41	0.41	0.40	0.41	0.40	$X^2(10)$
-0.025	0.42	0.42	0.41	0.40	0.39	0.35	
-0.05	0.41	0.42	0.41	0.39	0.37	0.33	
0.05	0.78	0.75	0.75	0.74	0.75	0.72	
0.025	0.77	0.75	0.75	0.73	0.73	0.70	$X^2(4)$
-0.025	0.76	0.75	0.75	0.73	0.70	0.66	
-0.05	0.75	0.75	0.75	0.72	0.67	0.63	

TABLE 5.26

Power Results of the Statistic $T_1(c)$ for Unstructured Data (m); Autoregressive (AR) Models with 1, 2, and 4 Parameters; a Two-Way (TW) Layout Model with and Without Interaction; Sample Size is 60; and $\alpha = 0.05$

c	m	AR(1)	AR(2)	AR(4)	TW(8)	TW(20)	
0.025	0.25	0.25	0.24	0.23	0.22	0.18	
0.01	0.25	0.25	0.24	0.22	0.21	0.17	
-0.01	0.24	0.25	0.23	0.22	0.19	0.15	t(9)
-0.25	0.24	0.24	0.23	0.21	0.19	0.15	
0.025	0.50	0.50	0.51	0.47	0.46	0.41	
0.01	0.50	0.50	0.50	0.47	0.45	0.39	
-0.01	0.49	0.49	0.49	0.46	0.44	0.36	t(5)
-0.025	0.48	0.49	0.49	0.45	0.42	0.35	
0.025	0.29	0.27	0.27	0.27	0.26	0.23	
0.01	0.28	0.27	0.27	0.27	0.25	0.21	
-0.01	0.28	0.27	0.27	0.27	0.24	0.18	$X^2(10)$
-0.025	0.27	0.27	0.27	0.26	0.24	0.18	
0.025	0.56	0.53	0.52	0.51	0.48	0.42	
0.01	0.55	0.53	0.52	0.51	0.46	0.40	
-0.01	0.54	0.53	0.51	0.50	0.45	0.37	$X^2(4)$
-0.025	0.52	0.53	0.51	0.50	0.44	0.35	

TABLE 5.27

Power Results of the Statistic $T_1(c)$ for Unstructured Data (m); Autoregressive (AR) Models with 1, 2, and 4 Parameters; Sample Size is 40; and $\alpha = 0.05$

c	m	AR(1)	AR(2)	AR(4)	
0.025	0.19	0.18	0.18	0.17	
0.01	0.19	0.18	0.18	0.17	t(9)
-0.01	0.18	0.18	0.17	0.16	
-0.025	0.18	0.18	0.17	0.16	
0.025	0.38	0.38	0.37	0.34	
0.01	0.38	0.37	0.36	0.33	t(5)
-0.01	0.37	0.37	0.36	0.33	
-0.025	0.37	0.37	0.35	0.32	
0.025	0.22	0.21	0.22	0.20	
0.01	0.22	0.21	0.22	0.20	
-0.01	0.21	0.21	0.22	0.19	$X^2(10)$
-0.025	0.21	0.21	0.21	0.19	
0.025	0.43	0.39	0.40	0.36	
0.01	0.42	0.39	0.40	0.36	
-0.01	0.41	0.39	0.40	0.35	$X^2(4)$
-0.025	0.40	0.39	0.39	0.35	

TABLE 5.28

Suggested Values for the Critical Parameter c for Models with 1, 2, 4, 8 and 16 Parameters (p), and Sample Sizes 30, 40, 60 and 120; one and two Dimensional (d) Data

	Sample Size				
p	30	40	60	120	
1	0.3	0.3	0.3	0.3	
2	0.025	0.025	0.025	0.1	
4	0.02	0.02	0.025	0.05	d = 1
8	0.01	0.02	0.02	0.04	
16	0.0075	0.01	0.015	0.025	
1	0.3	0.3	0.3	0.3	
2	0.02	0.02	0.025	0.05	
4	0.01	0.015	0.02	0.05	d = 2
8	0.0075	0.01	0.015	0.025	
16	0.006	0.008	0.01	0.025	

5.6 The Effect of the Model Selected

The analysis thus far has assumed that the true model was used to fit the data. In practice, a selection criteria, such as the PSIC criterion of (2.4.17), is used to select a model from a set of candidate models. The analysis and tests of fit are applied to the selected model. To examine the effect of the fitted model on the statistic $T_1(c)$, four parameter linear regression and autoregression models were fit with models of 2, 3, 4, 5 and 6 parameters. For $c = -0.025$ and 0.025, Figures 5.13 to 5.16 are plots of the .95 percentage point of $T_1(c)$ as a function of the fitted model order.

If the data are fitted with the true model or a larger model containing the true model, the percentage points are approximately the same, as seen in the plots. For autoregressions, the plots indicate that the percentage points increase when the data are overfitted regardless of the sign of c. The plots also indicate that the test is liberal when the data are overfitted; a smaller value of c will reduce the effect on $T_1(c)$ of overfitting the data. For the four parameter linear regression data, the percentage points indicate that $T_1(c)$ is a conservative test when the data are fitted at the true or larger model. For both linear regressions and autoregressions, the figures show that the test is very conservative, in general, when the data are underfit. This is not always true, as seen in Figures 5.13 and 5.14 for the three-parameter linear regression model fit to data produced by a four-parameter linear regression model. From the plots and the above discussion, we see that overfitting the data is preferred since the test statistic becomes stable when the fitted model contains the true model.

The effect of the selected model on $T_1(c)$ was analyzed via Monte Carlo simulation by examining autoregressive models with 1, 2 and 4 parameters. For each model, $T_1(c)$ was tabulated for the selected model. The model was selected using the PSIC criterion of (2.4.17) with $s(c) = (1 + c)$. For a sample size of 40, Table 5.29 presents the size of the test results for $c = -0.025$ and 0.025; these results are in good agreement with the size of $T_1(c)$ for unstructured data. If larger sample sizes are used, the agreement between the size of the test $T_1(c)$ for structured and unstructured data improves; also, larger c values can be used.

For the above models, a power study was performed on $T_1(c)$ obtained from the selected model. The alternative distributions were the chi-square distribution with 4 and 10 degrees of freedom, and the t-distribution with 5 and 9 degrees of freedom. The power results are shown in Table 5.30. It can be seen that these results compare favorably with the results in Tables 5.25 to 5.27 where the model is known.

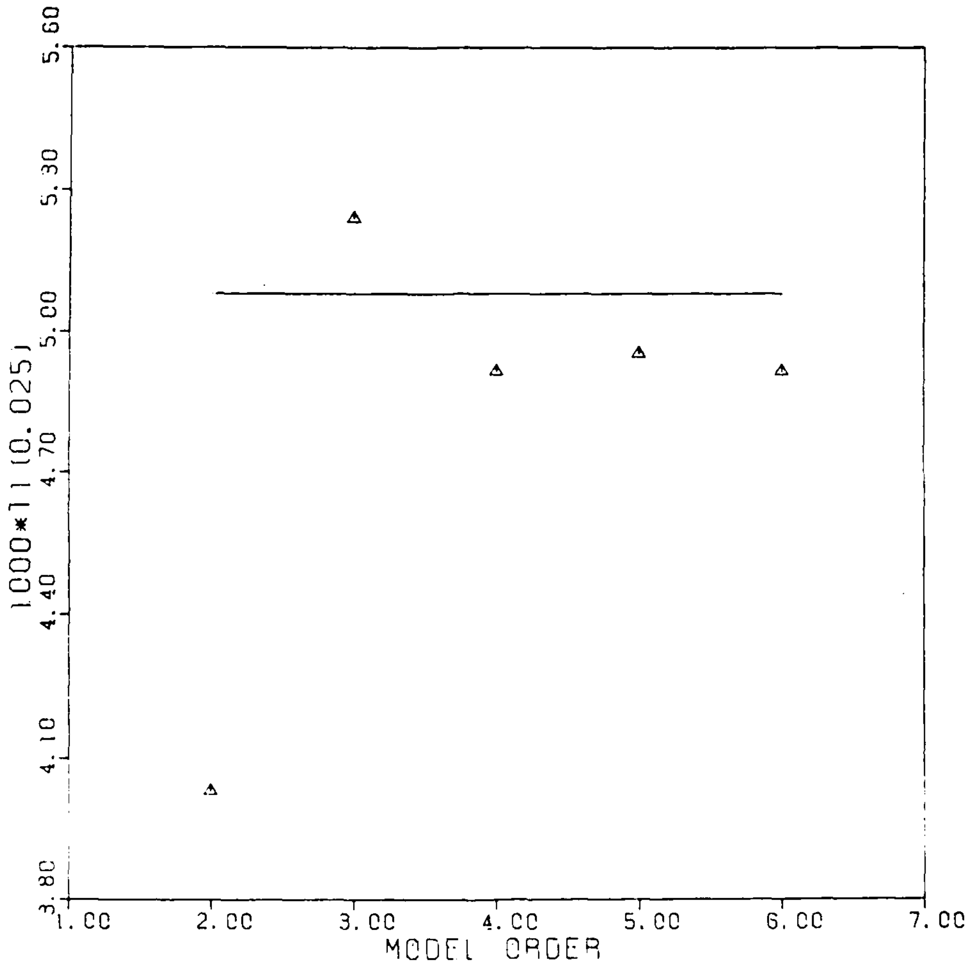

FIGURE 5.13. The 95 Percent Point of $T_1(c)$ Versus Fitted Model Order for a Four Parameter Linear Regression with Gaussian Errors, Sample Size = 64, and c = 0.025

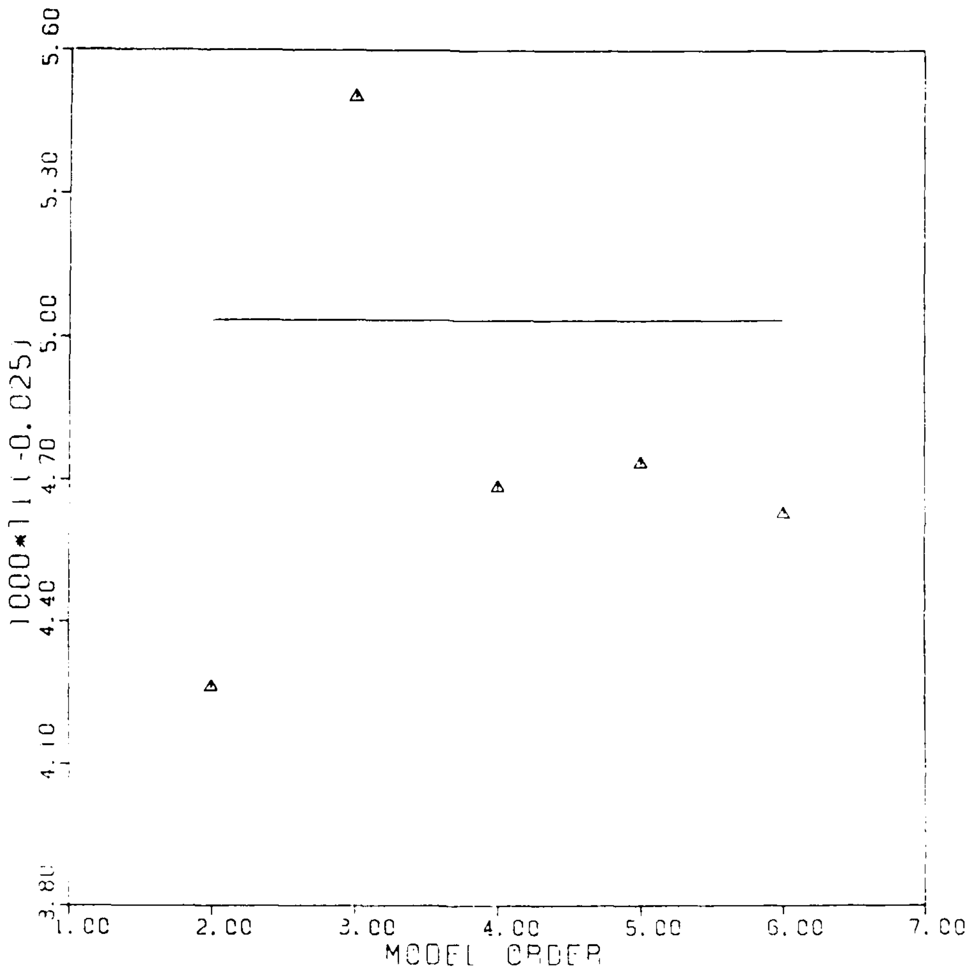

FIGURE 5.14. The 95 Percent Point of $T_1(c)$ Versus Fitted Model Order for a Four Parameter Linear Regression with Gaussian Errors, Sample Size = n4, and c = -0.025

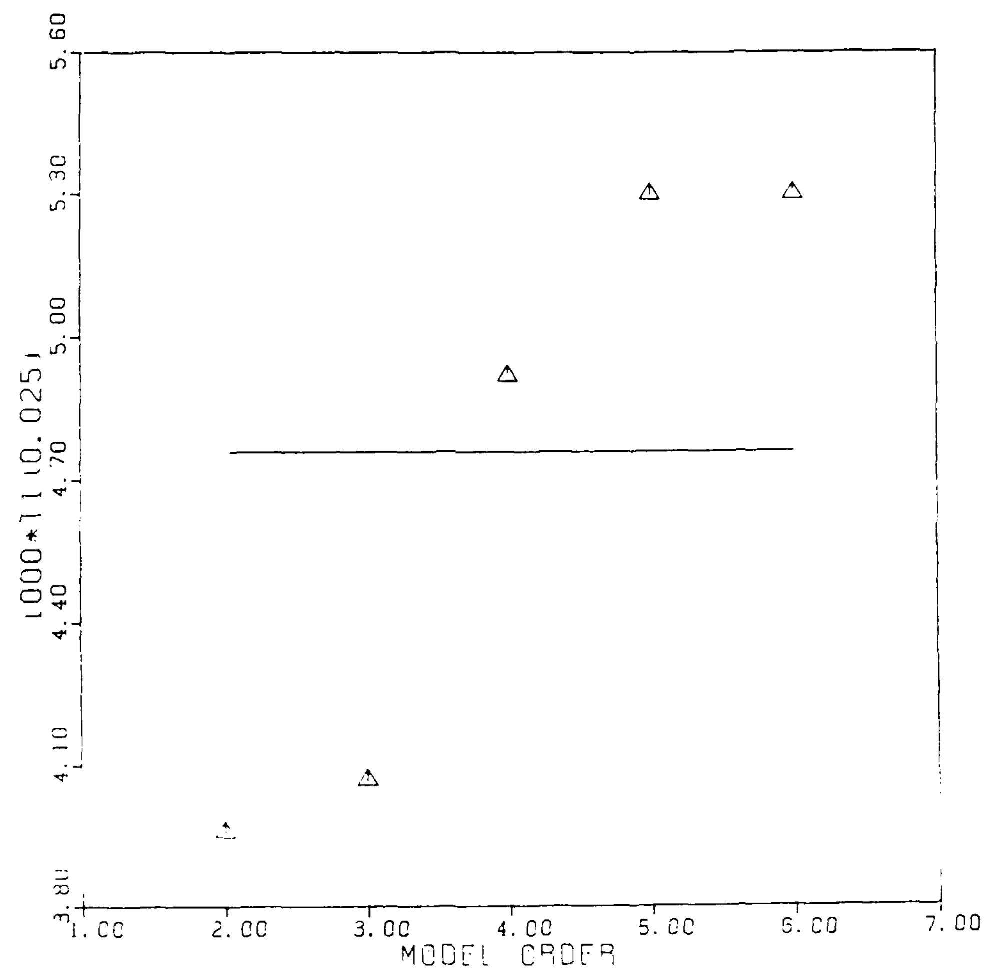

FIGURE 5.15. The 95 Percent Point of $T_1(c)$ Versus Fitted Model Order for a Four Parameter Autoregression with Gaussian Errors, Sample Size = 64, and c = 0.025

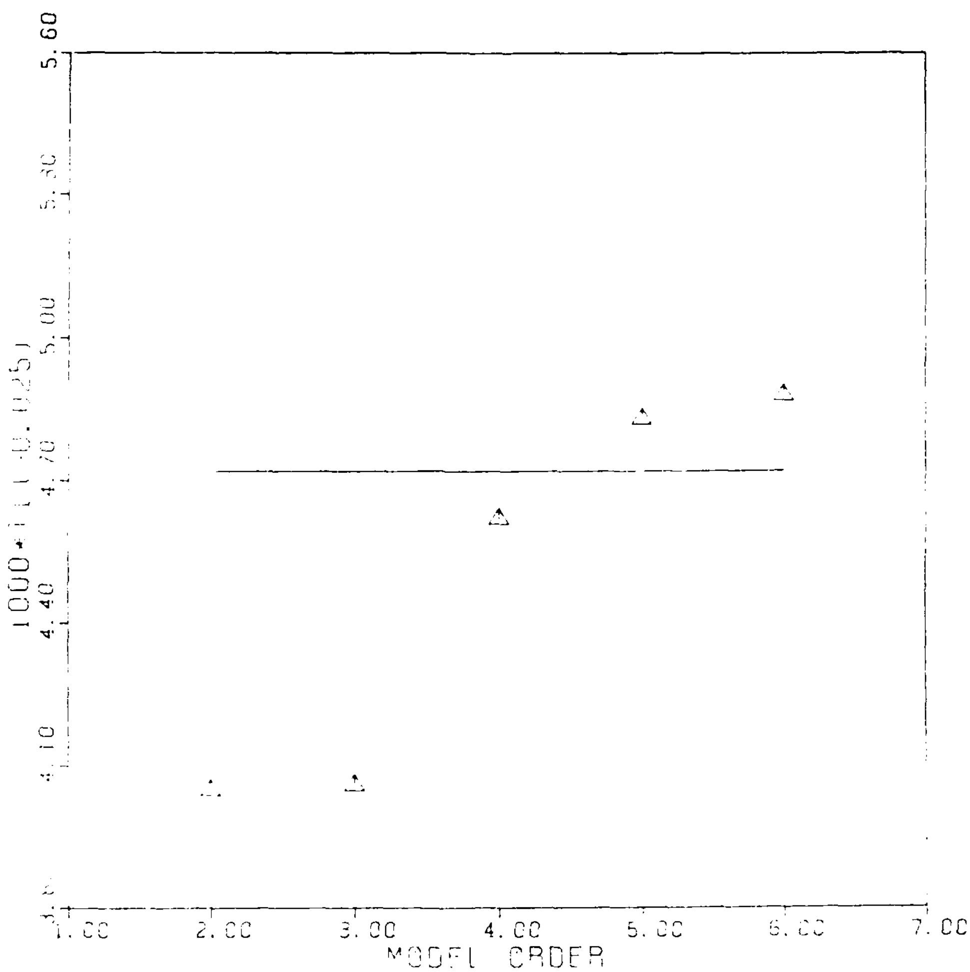

FIGURE 5.16. The 95 Percent Point of $T_1(c)$ Versus Fitted Model Order for a Four Parameter Autoregression with Gaussian Errors, Sample Size = 64, and c = -0.025

TABLE 5.29a

Size of the Test $T_1(c)$ Using the Selected Model for Autoregressive Processes with 1, 2, and 4 Parameters; Sample Size is 40; and c = 0.025

	Size of the Test			
p	0.1	0.05	0.025	0.01
1	0.104	0.052	0.027	0.012
2	0.106	0.052	0.028	0.013
4	0.109	0.058	0.032	0.014

TABLE 5.29b

Size of the Test $T_1(c)$ Using the Selected Model for Autoregressive Processes with 1, 2, and 4 Parameters; Sample Size is 40; and c = 0.01

	Size of the Test			
p	0.1	0.05	0.025	0.01
1	0.103	0.052	0.027	0.012
2	0.107	0.051	0.027	0.013
4	0.107	0.055	0.030	0.014

TABLE 5.29c

Size of the Test $T_1(c)$ Using the Selected Model for Autoregressive Processes with 1, 2, and 4 Parameters; Sample Size is 40; and c = -0.01

	Size of the Test			
p	0.1	0.05	0.025	0.01
1	0.102	0.051	0.026	0.012
2	0.105	0.050	0.026	0.012
4	0.102	0.051	0.027	0.013

TABLE 5.29d

Size of the Test $T_1(c)$ Using the Selected Model for Autoregressive Processes with 1, 2, and 4 Parameters; Sample Size is 40; and c = -0.025

	Size of the Test			
p	0.1	0.05	0.025	0.01
1	0.102	0.050	0.026	0.012
2	0.104	0.048	0.026	0.012
4	0.102	0.048	0.025	0.012

TABLE 5.30

Power Results of the Statistic $T_1(c)$ for Unstructured Data (m); the Selected Model for Autoregressive (AR) Processes with 1, 2, and 4 Parameters; Sample Size is 40; and α = 0.05

c	m	AR(1)	AR(2)	AR(4)	
0.025	0.19	0.19	0.18	0.17	
0.01	0.19	0.18	0.18	0.17	t(9)
-0.01	0.18	0.18	0.17	0.16	
-0.025	0.18	0.17	0.17	0.15	
0.025	0.38	0.36	0.36	0.33	
0.01	0.38	0.35	0.35	0.32	t(5)
-0.01	0.37	0.35	0.34	0.31	
-0.025	0.37	0.34	0.33	0.30	
0.025	0.22	0.20	0.21	0.20	
0.01	0.22	0.20	0.21	0.19	$X^2(10)$
-0.01	0.21	0.20	0.20	0.18	
-0.025	0.21	0.20	0.20	0.18	
0.025	0.43	0.37	0.38	0.35	
0.01	0.42	0.36	0.37	0.33	$X^2(4)$
-0.01	0.41	0.36	0.37	0.33	
0.025	0.40	0.35	0.36	0.32	

5.7 Discussion

The procedures presented here and in Part 2 provide a means to simultaneously analyze the goodness of fit of both the parametric and distributional form of a model. The model-critical selection criterion, PSIC, and the goodness of fit statistic, $T_1(c)$, are complementary procedures. The PSIC procedure selects the best parametric model consistent with Gaussianity. After the model is selected, the statistic $T_1(c)$ is used to test the normality of the residuals. Since both the parametric and distributional form of the model are used to obtain the test statistic $T_1(c)$, the test jointly examines both parts of the model. For testing the residuals with $T_1(c)$, small values of $|c|$ should be used. If the test rejects the normality of the residuals, larger values of $|c|$ should be used to obtain model-critical parameter estimates and model-critical weights. These estimates and weights can be used aid in determining why the test rejected normality. In Part 6, the selection and testing procedures will be applied to experimental data.

PART 6

APPLICATIONS OF MODEL-CRITICAL SELECTION AND TEST OF FIT

6.1 Introduction

In this part, our selection criterion and test of fit will be applied to some experimental data. The model-critical parameter estimates and weights will also be used in analyzing the adequacy of the assumed model.

The analysis of a set of data involves fitting the data with a set of candidate models for several values of c. The model is selected using the PSIC criterion. If the model selected depends on c, the model with the larger number of parameters is selected since underfitting the data is more serious than overfitting the data. For the selected model, the test statistic $T_1(c)$ is calculated and compared to the appropriate percentage point. If the normality of the residuals is rejected, the analysis continues by examining the model-critical parameter estimates and weights, probability plots of the residuals, and plots of the residuals versus fitted value, for a range of c values. Of these diagnostics, the critical weights are valuable in identifying inconsistencies between the data and the assumed model. In fact, regardless of the test of fit result, the model-critical parameter estimates and weights should be examined since they may expose a problem with the data and assumed model not detected by the statistic $T_1(c)$.

6.2 An ARMA(p,q) Example

The time series of 197 chemical process readings from Series A in Box and Jenkins (1970) is examined in this section. The observations were sampled every two hours and are shown in Figure 6.1. Examination of the autocorrelations and partial correlations suggests an ARMA(1,1) model (See Box and Jenkins, 1970 for details). To verify this, the data were fit with ARMA(p,q) models for $p \leq 2$ and $q \leq 2$. For $c = 0.1$, Table 6.1 is a list of the models considered, the model-critical estimate of the error variance, and the value of the model-critical selection criterion with $s(c) = (1 + c)$.

TABLE 6.1

The Models Considered, Innovations Variance Estimate, and PSIC Criterion with $s(c) = 1+c$ for the Chemical Process Data; $c = 0.1$

MODEL	$s^2(0.1)$	PSIC
ARMA(1,0)	0.102	-2.26
ARMA(0,1)	0.124	-2.07
ARMA(1,1)	0.094	-2.32
ARMA(2,0)	0.096	-2.30
ARMA(0,2)	0.108	-2.18
ARMA(2,1)	0.094	-2.31
ARMA(1,2)	0.094	-2.31
ARMA(2,2)	0.093	-2.30

From Table 6.1, it can be seen that the ARMA(1,1) model is selected by the PSIC criterion. For this model, the value of the test statistic $T_1(c)$ is 0.102 which exceeds the .95 percentage point of 0.089 (See

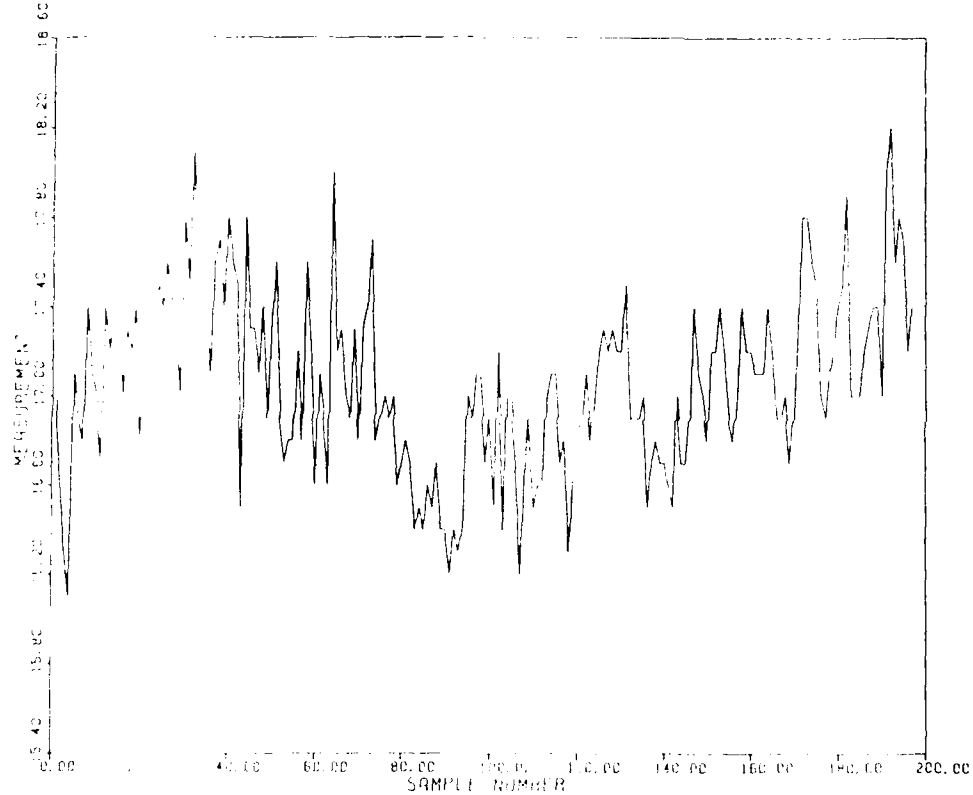

FIGURE 6.1. Chemical Process Data Sampled Every Two Hours; 197 Observations

Lawrence, Paulson, and Swope, 1986 for percentage points of $T_1(c)$ with sample sizes larger than 120); therefore, we would reject the normality of the residuals. It is noted that $T_1(0.05)$ is 0.768×10^{-2} which is less than the 0.90 percentage point of 1.60×10^{-2}. For a fixed value of c, the analysis of Part 5 shows that the percentage points of $T_1(c)$ for structured data tend toward the percentage points of $T_1(c)$ for unstructured data as the sample size increases. Also, for a fixed sample size, the goodness of fit test becomes increasingly liberal as c increases. Since the test rejected normality for c = 0.1 and not for c = 0.05, it is possible that the rejection was due to the value of c being too large. However, considering the large number of observations and the small number of parameters, we feel that the test result for c = 0.1 is valid. Thus, we must trade off some power of the test for a conservative test. Figures 6.2a to c are plots of the maximum likelihood and model-critical residuals for the ARMA(1,1) model; they indicate possible outliers at observations 43 and 64.

Our analysis continues by examining the model-critical and maximum likelihood parameter estimates for the ARMA(1,1) model; Table 6.2 presents the model-critical parameter estimates for c = 0, 0.1, 0.2, 0.3 and 0.4.

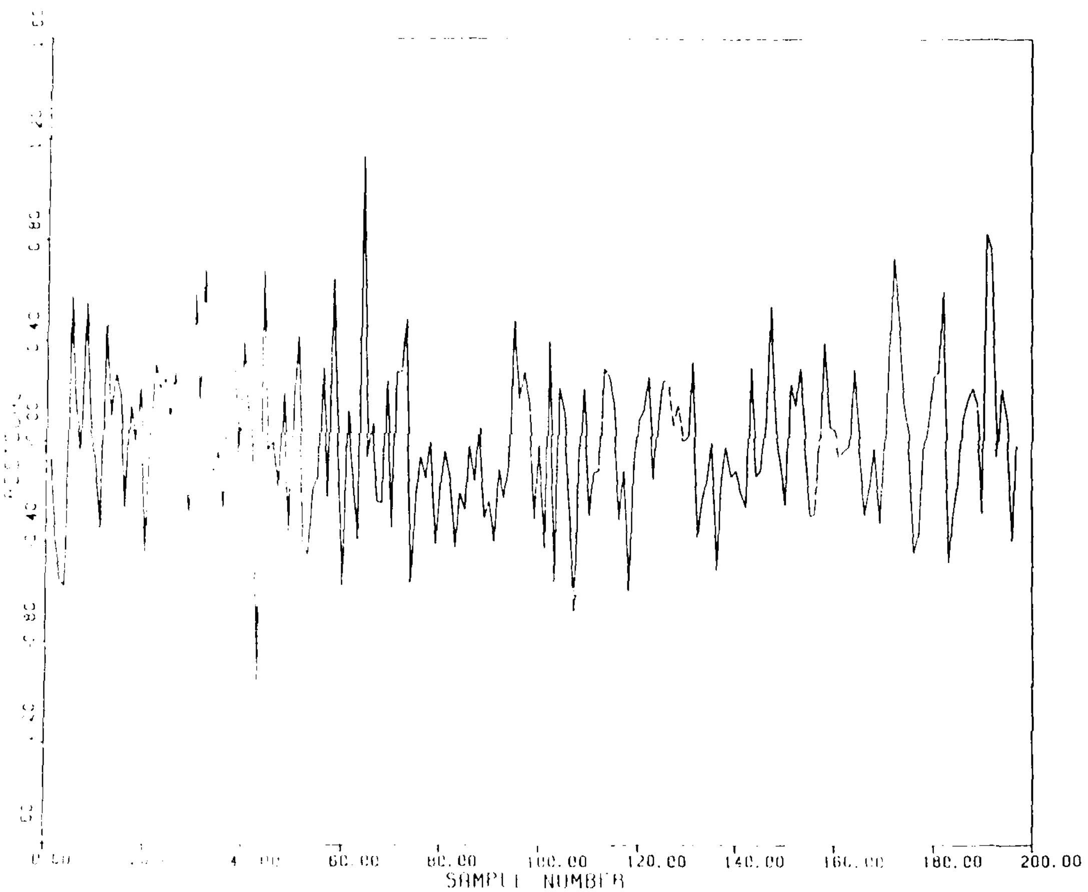

FIGURE 6.2a. Model-Critical Residuals from the ARMA(1,1) Model Fit to the Chemical Process Data; c = 0

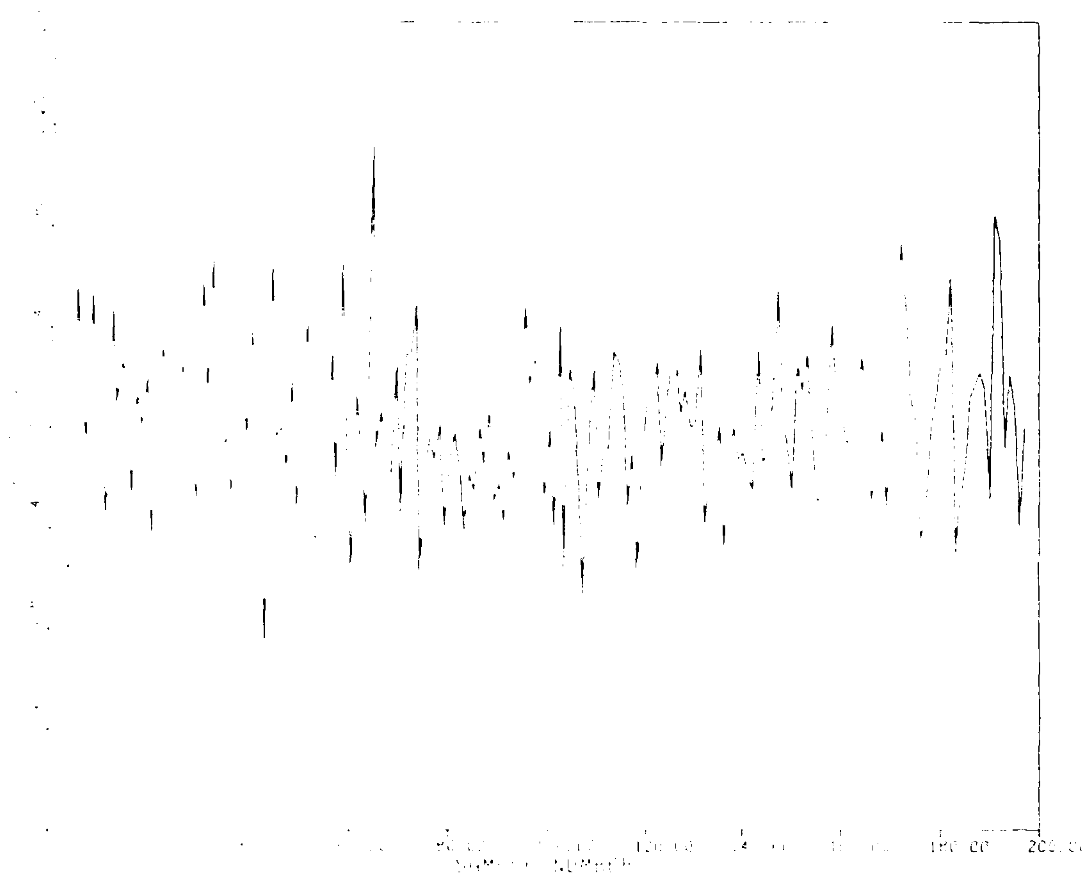

FIGURE 6.2b. Model-Critical Residuals from the ARMA(1,1) Model Fit to the Chemical Process Data; $c = 0.1$

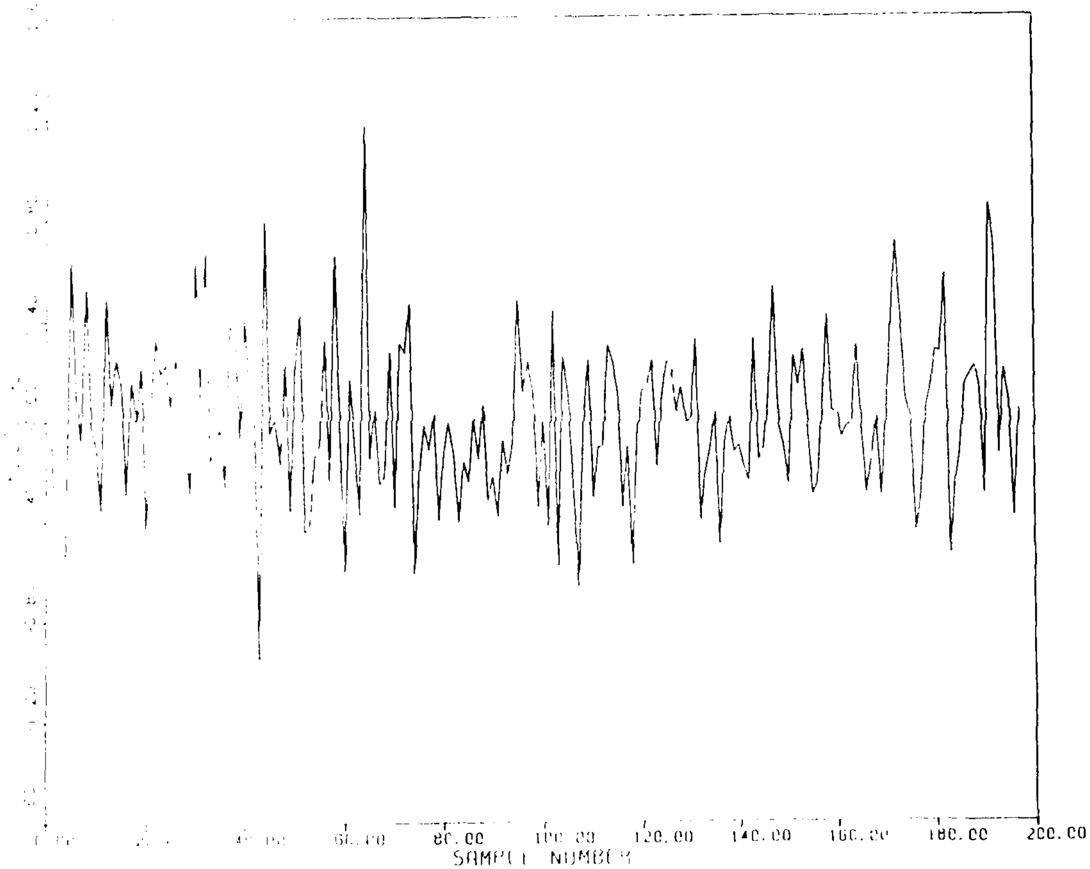

FIGURE 6.2c Model-Critical Residuals from the ARMA(1,1) Model Fit to the Chemical Process Data; c = 0.4

TABLE 6.2

Maximum Likelihood ($c = 0$) and Model-Critical ($c \neq 0$)
Estimates of the ARMA(1,1) Model used to fit the
Chemical Process Data

c	$a_1(c)$	$b_1(c)$	$s^2(c)$
0	0.908	-0.576	0.0977
0.1	0.905	-0.547	0.0946
0.2	0.892	-0.508	0.0916
0.3	0.887	-0.486	0.0886
0.4	0.887	-0.460	0.0870

Noting the analysis in Section 3.4 of the simulated ARMA(2,1) process, with t-distributed errors, the changes in $b_1(c)$ and $s^2(c)$ suggest that the residuals have a heavy-tailed distribution. Figures 6.3a and 6.3b present Gaussian probability plots of the maximum likelihood and model-critical residuals; the critical residuals were obtained using $c = 0.4$. From the plots, the residual distribution appears to have a heavy right tail and a short left tail. For $c = 0.4$, Figure 6.4 is a plot of the model-critical weights. Figures 6.3 and 6.4 indicate that observations 43 and 64 may be outliers. Inspecting data around observation 43 indicates that this observation may have been recorded incorrectly. Observation 43 has a value of 16.5, but the observations on either side have values about 17.5.

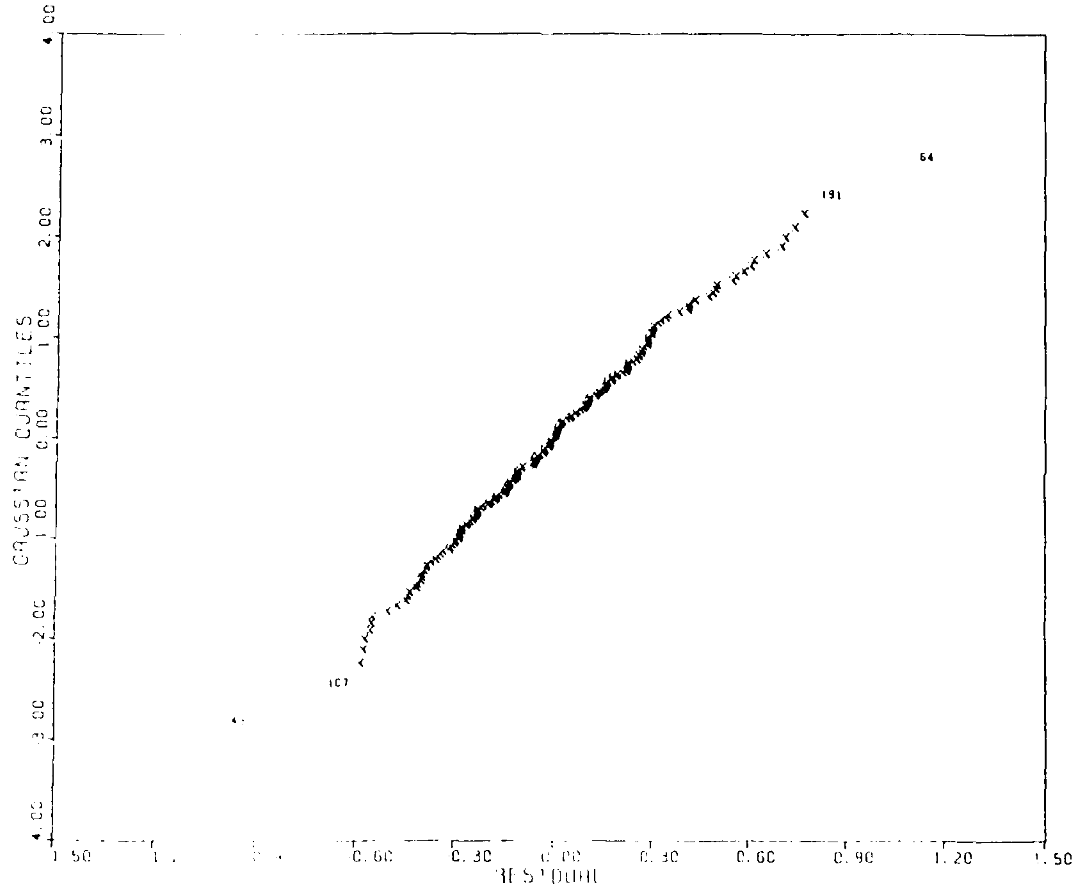

FIGURE 6.5a. Probability Plot of the Residuals from the ARMA(1,1) Model Fit to the Chemical Process Data; c = 0

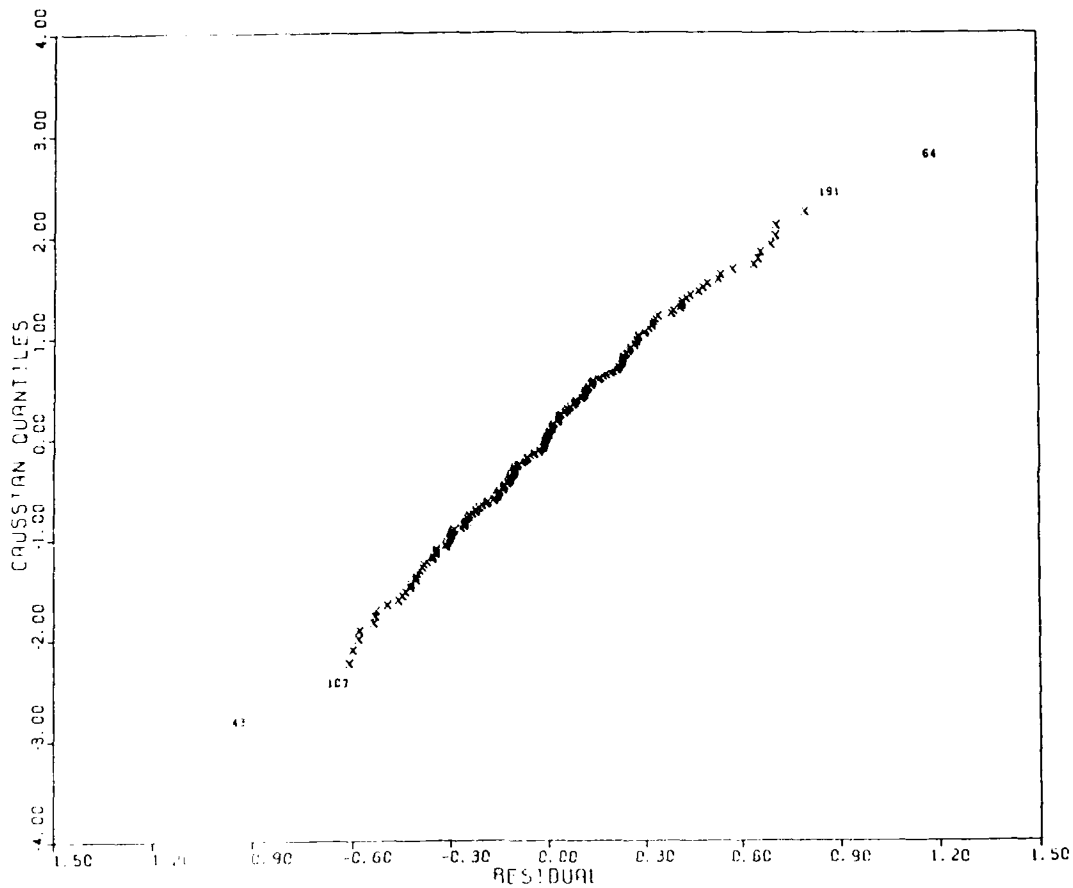

FIGURE 6.3b. Probability Plot of the Residuals from the ARMA(1,1) Model Fit to the Chemical Process Data; c = 0.4

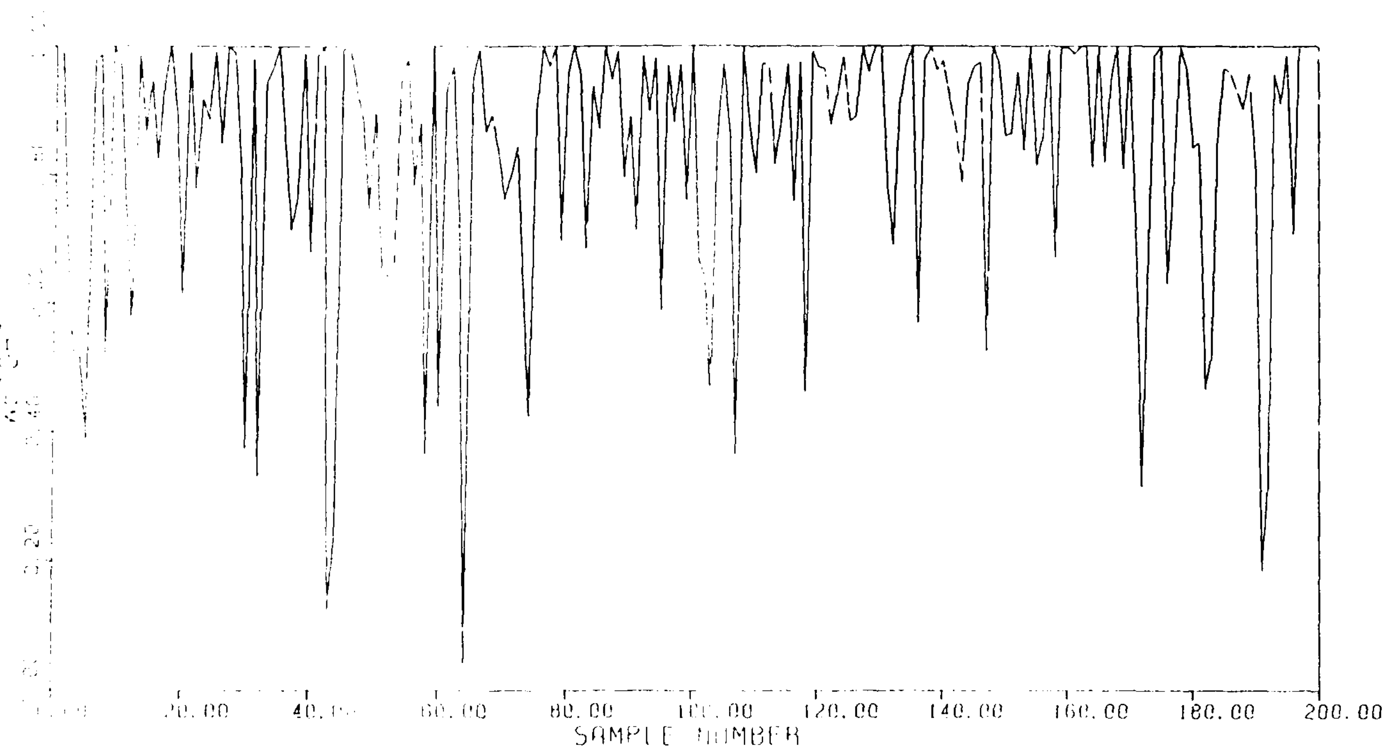

FIGURE 6.4. Model-Critical Weights from the ARMA(1,1) Model Fit to the Chemical Process Data; c = 0.4

6.3 A Linear Regression Example

In this section, we use the PSIC model selection criterion and the goodness of fit test statistic $T_1(c)$ to analyze the abrasion resistance of rubber data in Table 2.3. From (2.3.8), the full quadratic model is

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_1^2 + \alpha_3 x_2 + \alpha_4 x_2^2 + \alpha_5 x_1 x_2 + \varepsilon. \qquad (6.1)$$

For regression models with K possible independent variables, there are 2^K possible models (Daniel and Wood, 1980, Chapter 6). Using the model of (6.1), there are 32 possible regression models. For the model in (6.1), all possible regression models were examined by the PSIC selection criterion for c = -0.1, 0.1, 0.2, and 0.3. Table 6.3 presents the two best models selected by the PSIC criterion with $s(c) = (1 + c)$. From the table, it can be seen that the full model is selected only for c = 0.3.

A plot of abrasion resistance versus silica level, shown in Figure 6.5, indicates primarily a linear relationship. When c = 0.3, the downweighting of observation 1 results in a more quadratic character to the data in Figure 6.5; the Y at silica level equal to one on the plot signifies observation 1. This effect can also be seen in Table 2.4 where the coefficient of x_1^2, a_2, increases with c.

For the full quadratic model, the test statistic $T_1(c)$ was calculated at c = -0.01 and c = 0.01, and yielded 0.963×10^{-4} and

TABLE 6.3

The Variables in the Two Best Regression Models and Corresponding PSIC Values for c = -0.1, 0.1, 0.2, and 0.3; Rubber Abrasion Resistance Data

Variables in the Model	c	PSIC
$x_1, x_2, x_2^2, x_1 x_2$	-0.1	2.900
$x_1, x_1^2, x_2, x_2^2, x_1 x_2$	-0.1	2.991
$x_1, x_2, x_2^2, x_1 x_2$	0.1	3.053
$x_1, x_1^2, x_2, x_2^2, x_1 x_2$	0.1	3.167
$x_1, x_2, x_2^2, x_1 x_2$	0.2	3.120
$x_1, x_1^2, x_2, x_2^2, x_1 x_2$	0.2	3.239
$x_1, x_2, x_2^2, x_1 x_2$	0.3	3.166
$x_1, x_1^2, x_2, x_2^2, x_1 x_2$	0.3	3.161

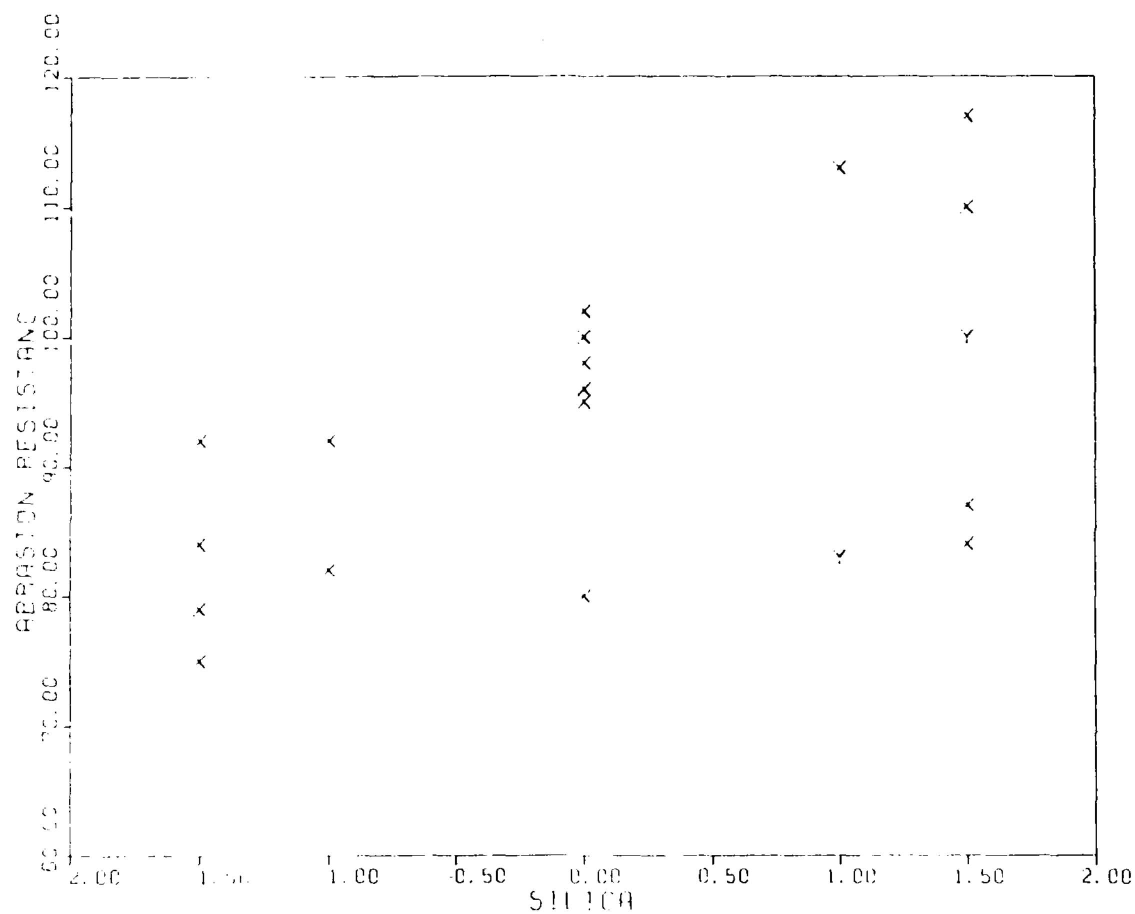

FIGURE 6.5. Abrasion Resistance of Rubber Versus Silica Level

0.859×10^{-4}, respectively. Using Table 5.1 and interpolating between $n = 20$ and 24 to obtain percentage points for $n = 22$, we see that both values of $T_1(c)$ fall well below the 0.75 percentage points of the test for $c = -0.01$ and $c = 0.01$. However, the small weights at observations 1 and 13 for $c = 0.3$, and the change in parameter estimates with c indicate that the model is still suspect. In fact, Table 2.4 shows that the estimate of the error variance increases as c increases from -0.1 to 0.1. From our experience, this indicates a short-tailed or broad-shouldered distribution. When fitting the data with a structured model, this suggests that additional variables may need to be considered in the model. Next, we examined Gaussian probability plots of the residuals for $c = 0, 0.1, 0.3$; the plots are shown in Figures 6.6a to 6.6c. The plots indicated that the residual distribution has a short right tail. For $c = 0$, observations 1 and 10 stick out on the left end of the plot in Figure 6.6a; however, the residual distribution does not appear to be distinctly non-Gaussian. The plots for $c = 0.1$ and 0.3 further amplify the outliers at observations 1 and 10.

The analysis points out that caution should be exercised when using the model. If the model is to be used for prediction, additional observations at silica and coupling agent levels equal to one should be obtained. Also, the possibility of additional variables should be explored. Suich and Derringer (1977) used this data to obtain a response surface for y as a function of x_1 and x_2. Delehanty

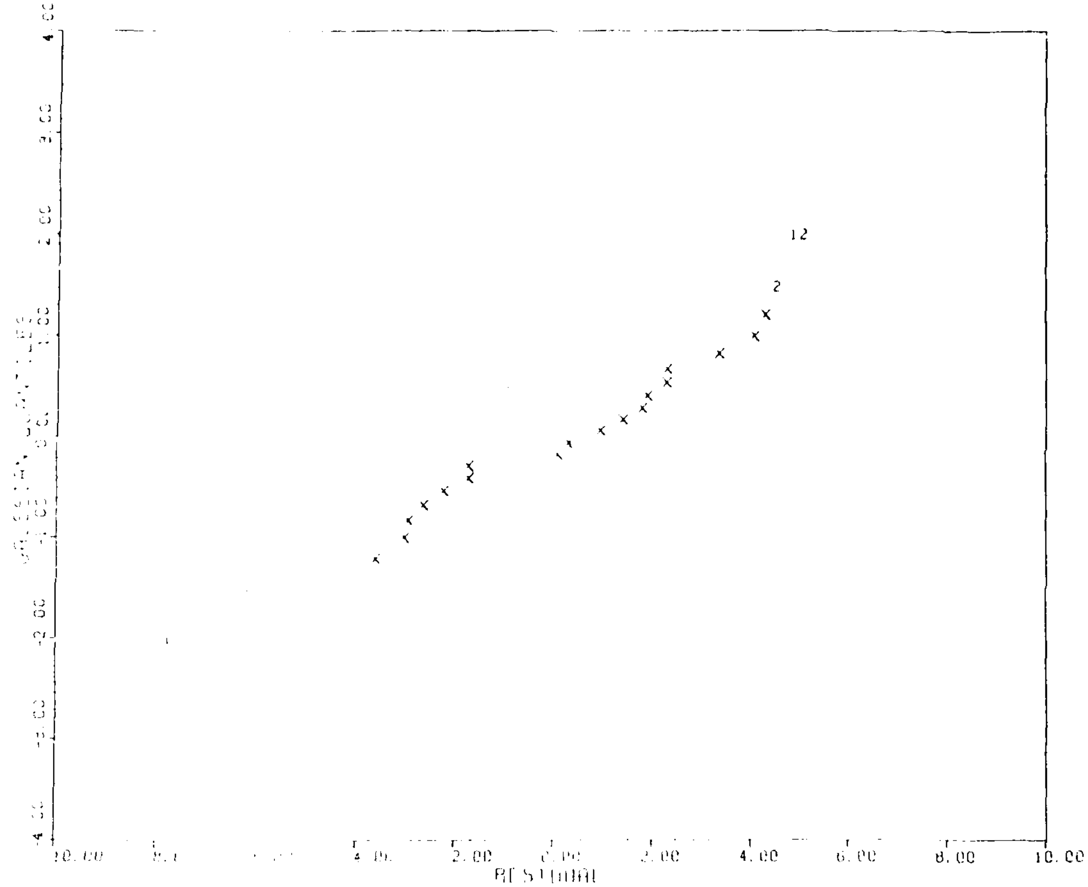

FIGURE 6.6a. Probability Plot of the Residuals from the Full Quadratic Model Fit to the Abrasion Resistance Data; c = 0

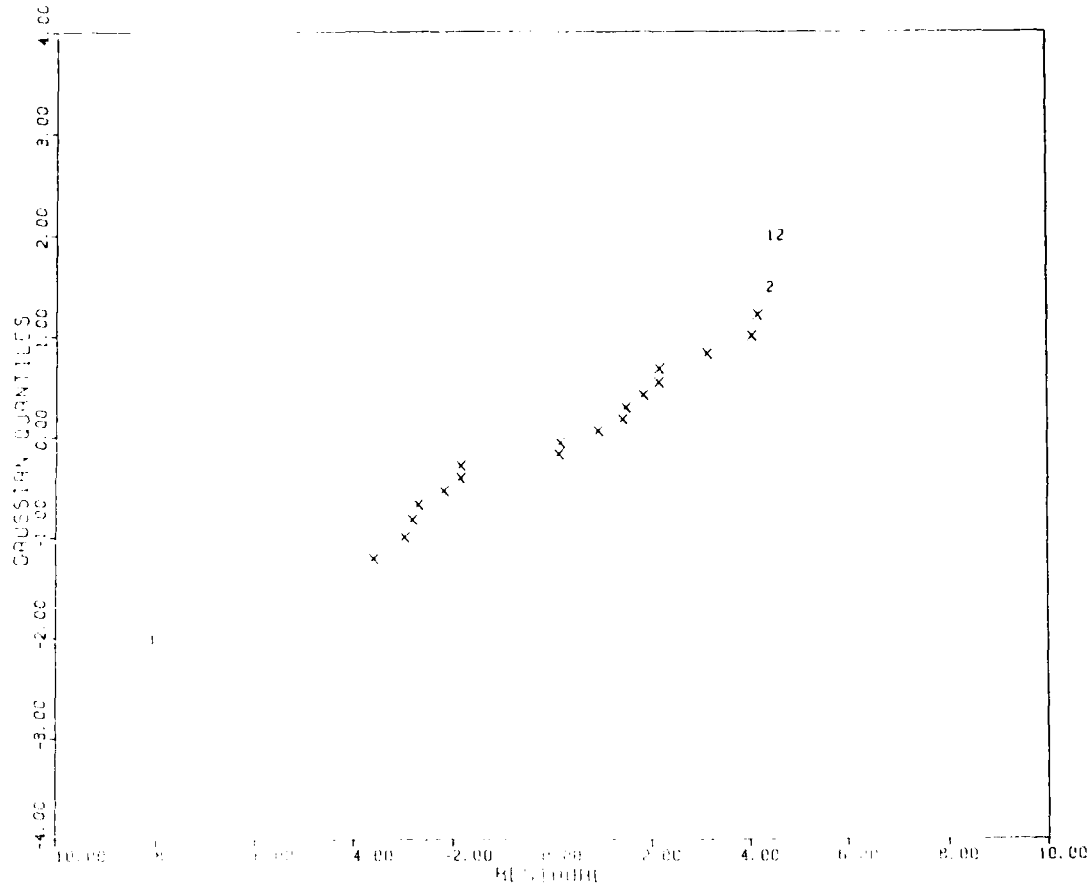

FIGURE 6.6b. Probability Plot of the Residuals from the Full Quadratic Model Fit to the Abrasion Resistance Data; $c = 0.1$

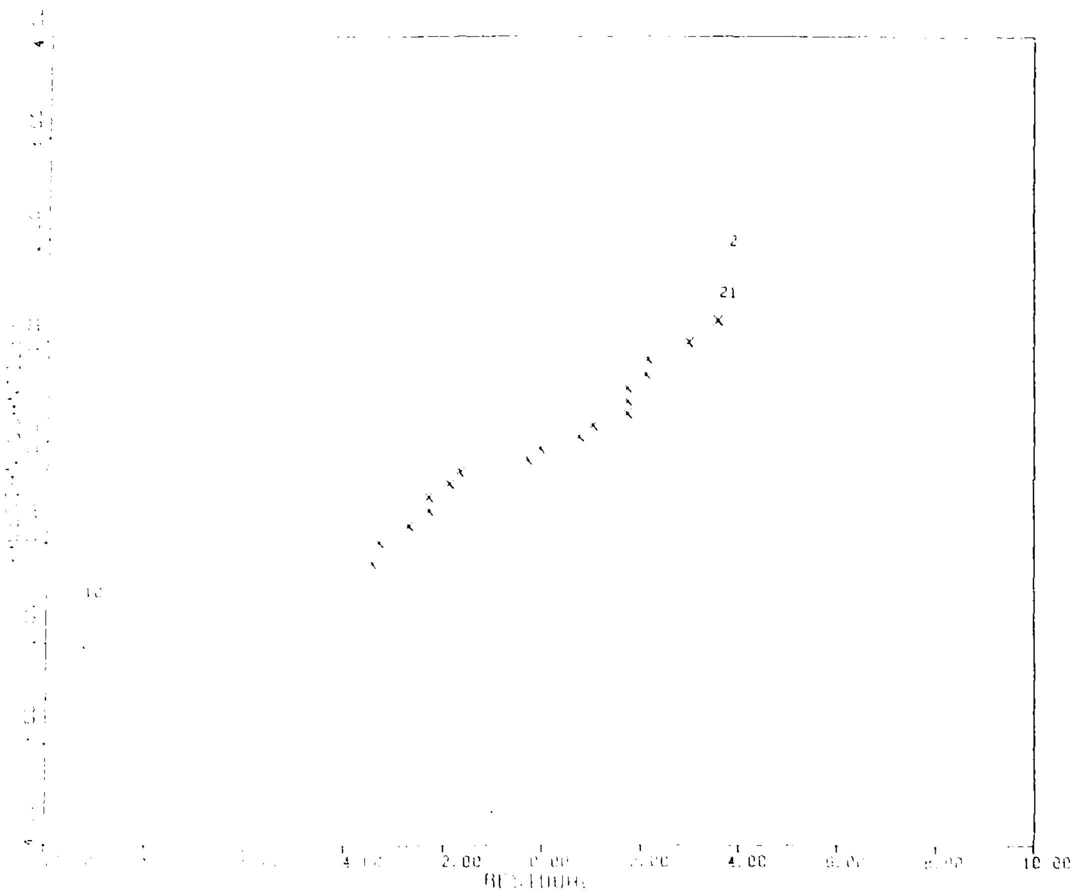

FIGURE 6.6c. Probability Plot of the Residuals from the Full Quadratic Model Fit to the Abrasion Resistance Data; c = 0.3

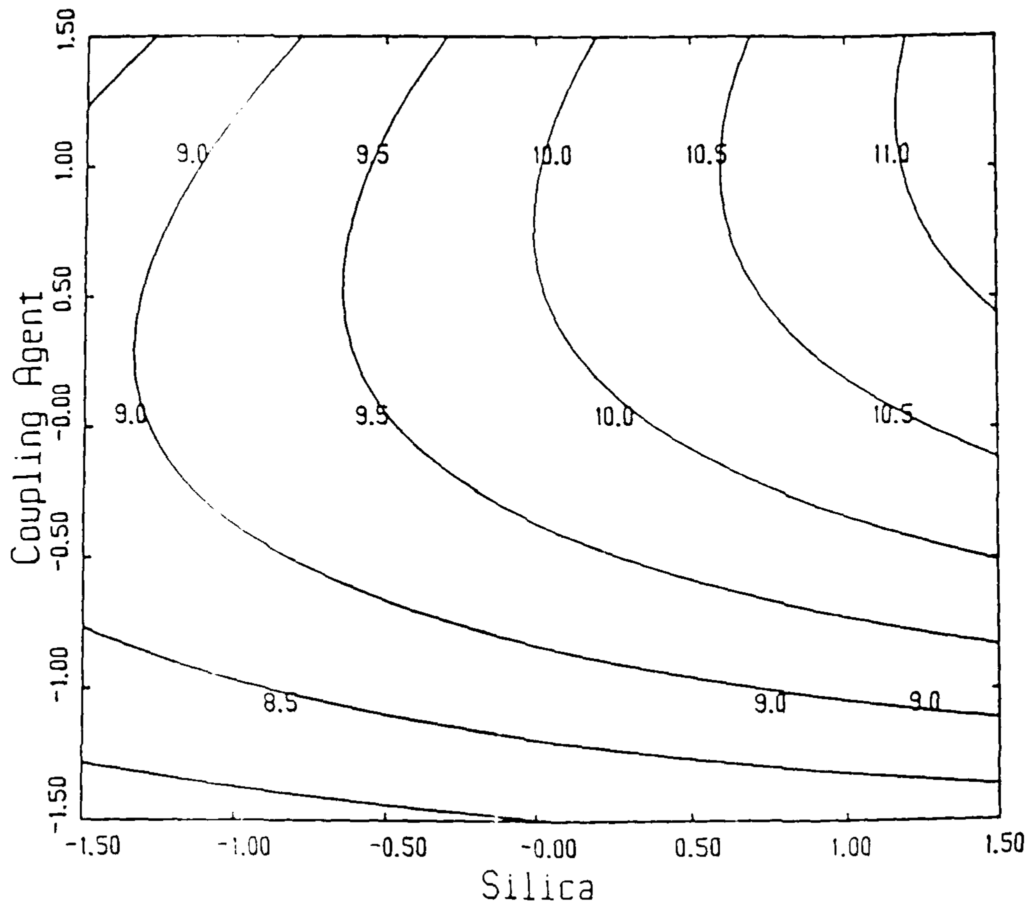

FIGURE 6.7a. Contour Plots of the Response Surface for Abrasion Resistance of Rubber Versus Silica and Coupling Agent Levels; c = 0

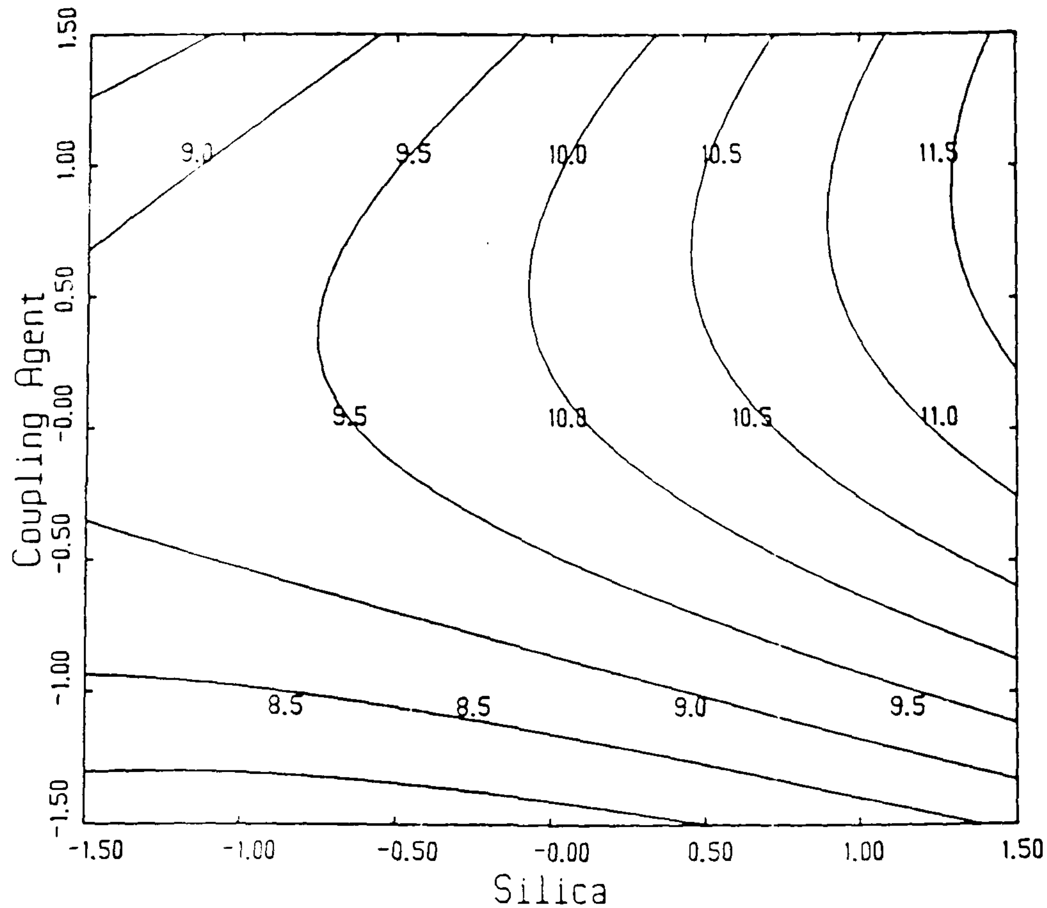

FIGURE 6.7b. Contour Plots of the Response Surface for Abrasion Resistance of Rubber Versus Silica and Coupling Agent Levels; c = 0.5

6.4 Analysis of a Two-way Layout Example

In this section, we analyze the replicated survival time data for three poisons and four treatments shown in Table 2.7. The PSIC criterion is used to select the full model with Poison, Treatment, and Poison×Treatment effects, or the additive model with only Poison and Treatment effects. For $c = 0.1$ and $s(c) = (1 + c)$, the PSIC values are -3.03 and -3.50 for the full and additive models, respectively. The additive model is selected since the reduction in the error variance does not indicate a need for the additional parameters. An interaction plot of fitted cell means for $c = 0$, 0.1, 0.2, and 0.3 is shown in Figure 6.8. The plots agree with the PSIC criterion that the additional parameters of the interaction model are not warranted.

For $c = 0.01$ and $c = 0.02$, the values of the test of fit statistic $T_1(c)$ are 0.0127 and 0.0289, respectively. Using the percentage points in Table 5.1 and interpolating between 40 and 60, the 0.99 percentage points at $n = 48$ are 0.0023 and 0.0093 for $c = 0.01$ and 0.02, respectively. Both values of the test statistic exceed the corresponding 0.99 percentage point of $T_1(c)$ and we reject the normality of the residuals. Noting the difference between the calculated test statistic and the corresponding .99 percentage point, we can see the effect of the choice of c on the value of $T_1(c)$. Gaussian probability plots of the model-critical residuals for $c = 0$ and 0.4 are shown in Figures 6.9a and b. The plots confirm the non-Gaussian character of the data. Figure 6.10 is a plot of the

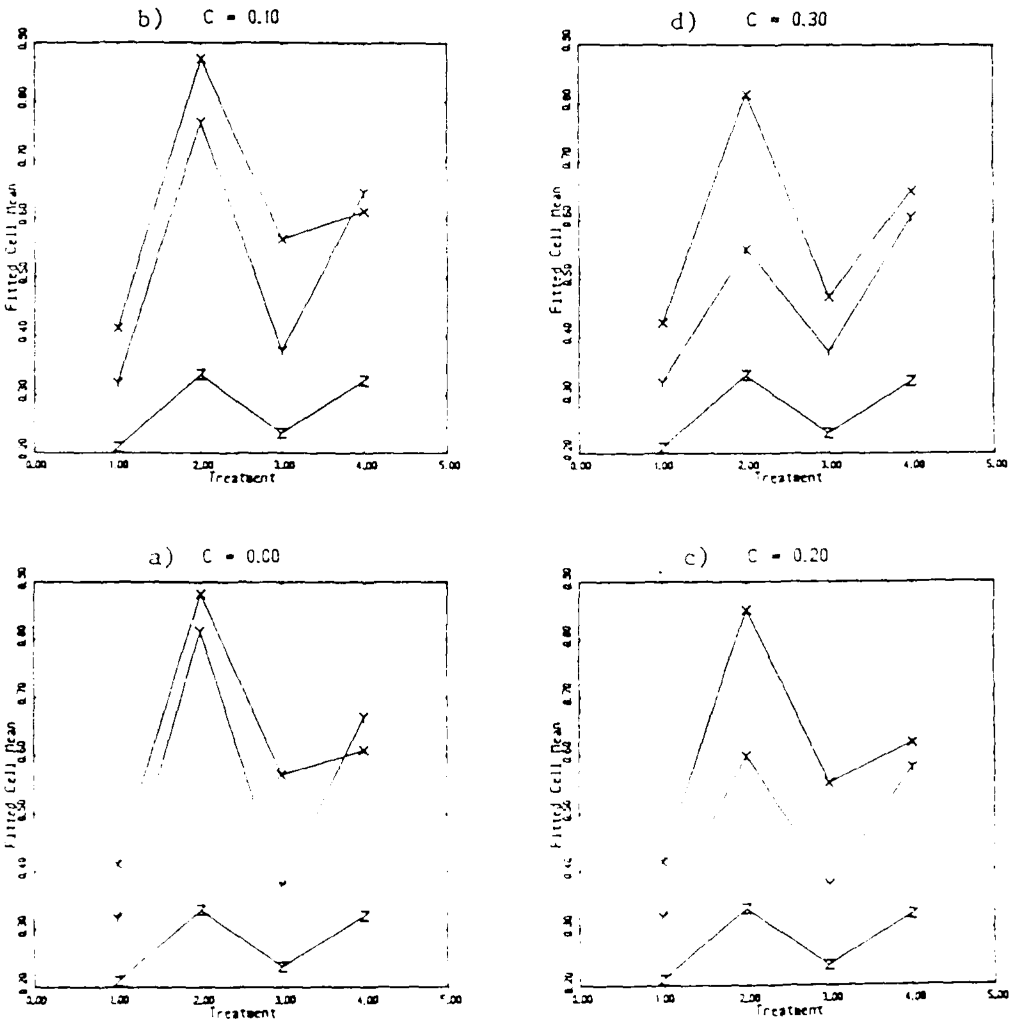

FIGURE 6.8. Interaction Plots of the Survival Time Data for c = 0, 0.1, 0.2, and 0.3

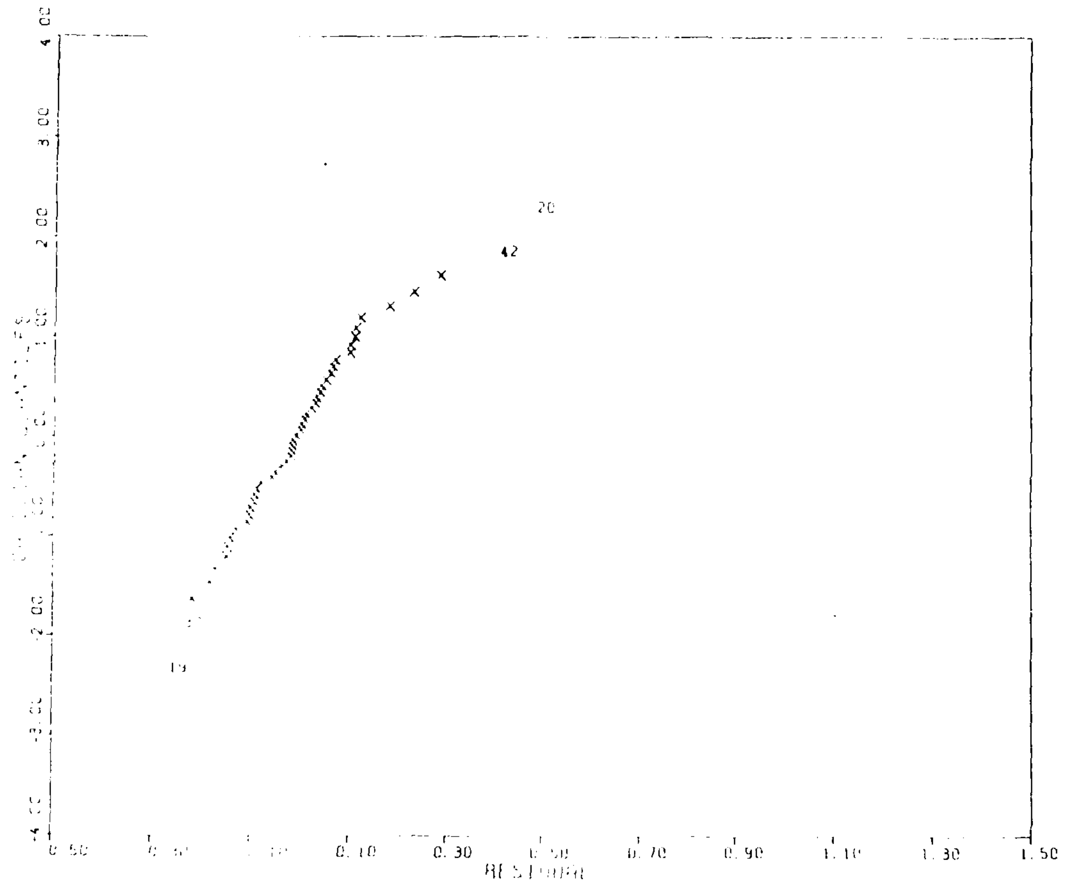

FIGURE 6.9a. Probability Plot of the Residuals from the Additive Model Fit to the Survival Time Data; c = 0

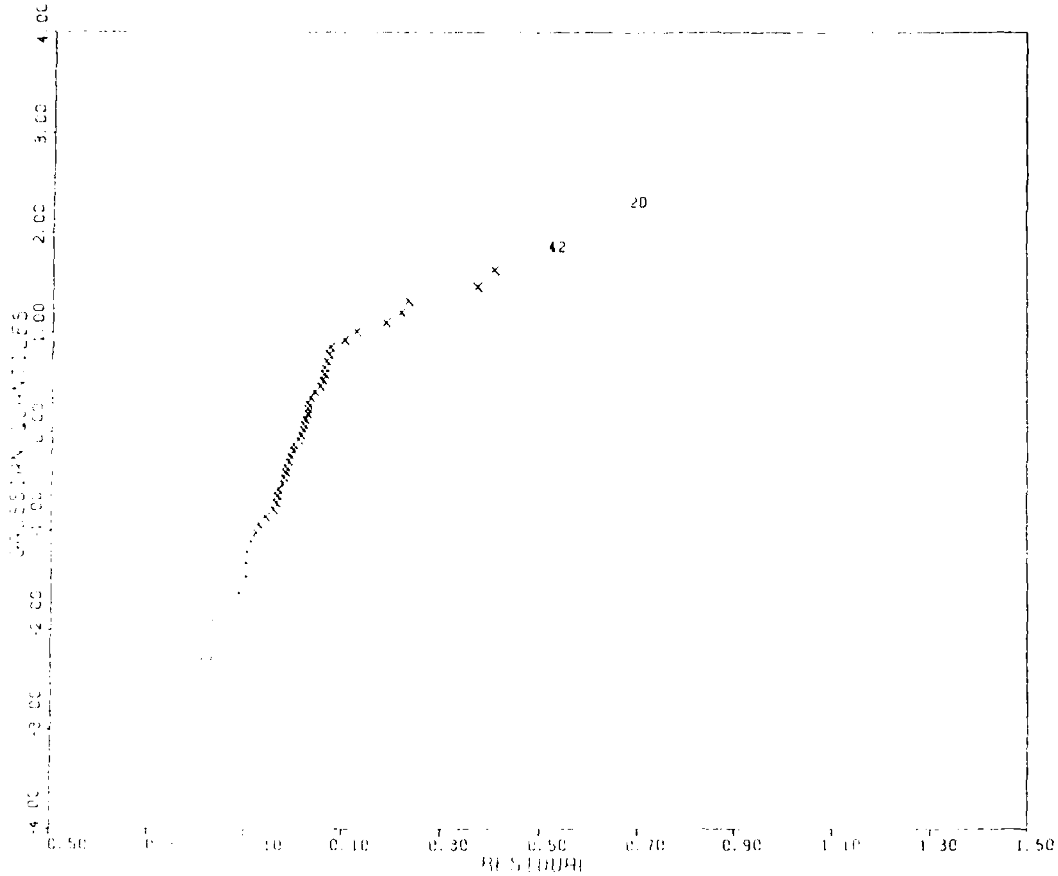

FIGURE 6.9b. Probability Plot of the Residuals from the Additive Model Fit to the Survival Time Data; c = 0.4

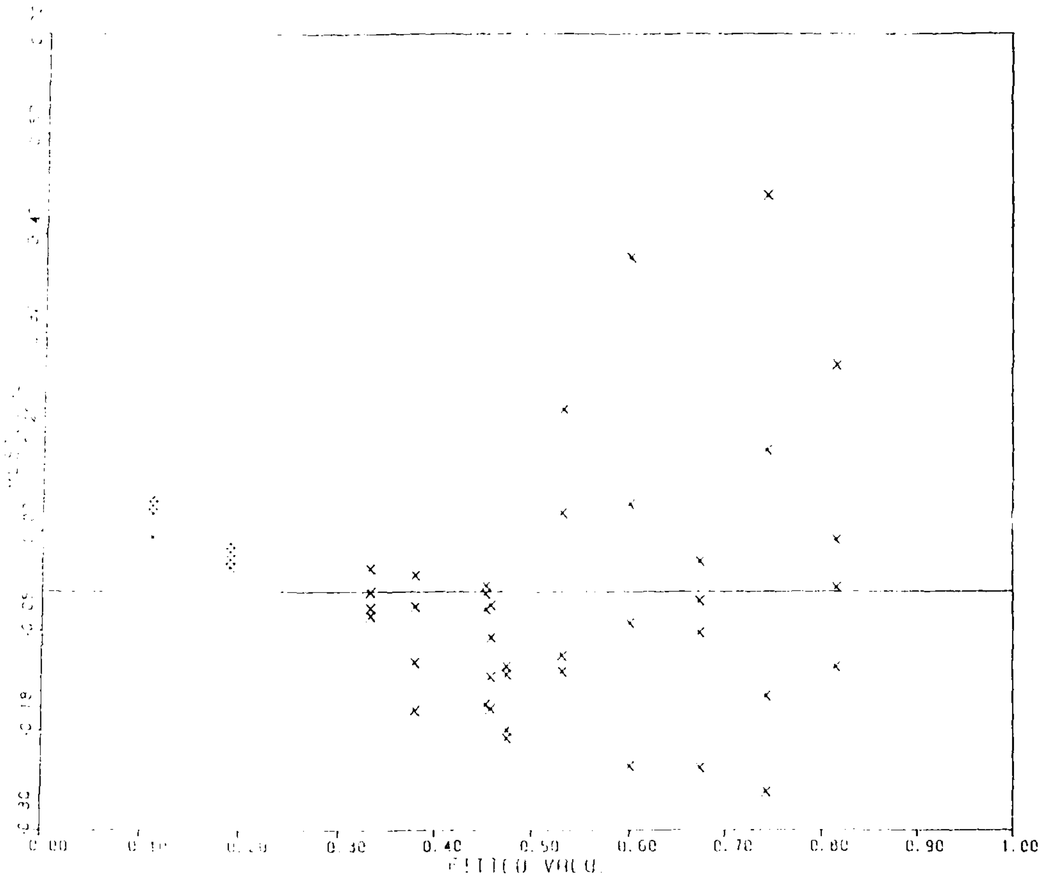

FIGURE 6.10. Residuals Versus Fitted Value for Survival Times with Additive Model; c = 0

maximum likelihood ($c = 0$) residuals versus fitted values. Figures 6.9 and 6.10 indicate the need for a transformation.

The family of Box-Cox (1964) transformations,

$$y(\lambda) = (y^{\lambda}-1) \,/\, \lambda \, \dot{y}^{\lambda-1}, \tag{6.2}$$

is considered where $\dot{y}$ is the geometric mean of the observations and λ is the power to be estimated. Delehanty (1983) estimates the power λ by maximizing the generalized likelihood $L(c)$, over a range of λ values. The value of λ was found to be about -0.75. For a more meaningful transformation, the value of $\lambda = -1$ was chosen; this transforms survival times into death rates. An additive model is fit to the reciprocal of the survival times. The values of our test statistic, for this model, are 0.586×10^{-3} and 0.214×10^{-3} for $c = 0.01$ and 0.02, respectively. Using Table 5.1, we see that both values are considerably less than the 0.75 percentage points. The model-critical weights for $c = 0.4$ are shown in Table 6.4 and indicate that the additive model of the transformed data is reasonable. Figures 6.11a and 6.11b are probability plots of the residuals for $c = 0$ and 0.4, respectively; Figure 6.12 is a plot of the residuals ($c = 0$) versus fitted value. These plots confirm the normality of the residuals.

TABLE 6.4

Model-Critical Weights for Death Rates with
the Additive Model, c = 0.4

		Treatment			
		1	2	3	4
Poison	1	0.65	0.99	0.99	0.45
		0.88	0.96	1.0	0.99
		0.85	1.0	0.68	0.97
		0.94	0.92	0.45	0.93
	2	0.97	0.86	0.91	0.99
		0.77	0.97	0.92	0.59
		0.82	0.72	0.65	0.92
		0.14	0.65	0.99	0.42
	3	1.0	0.93	0.97	1.0
		0.95	0.88	0.97	0.76
		0.35	0.83	1.0	0.99
		0.97	0.86	0.86	0.93

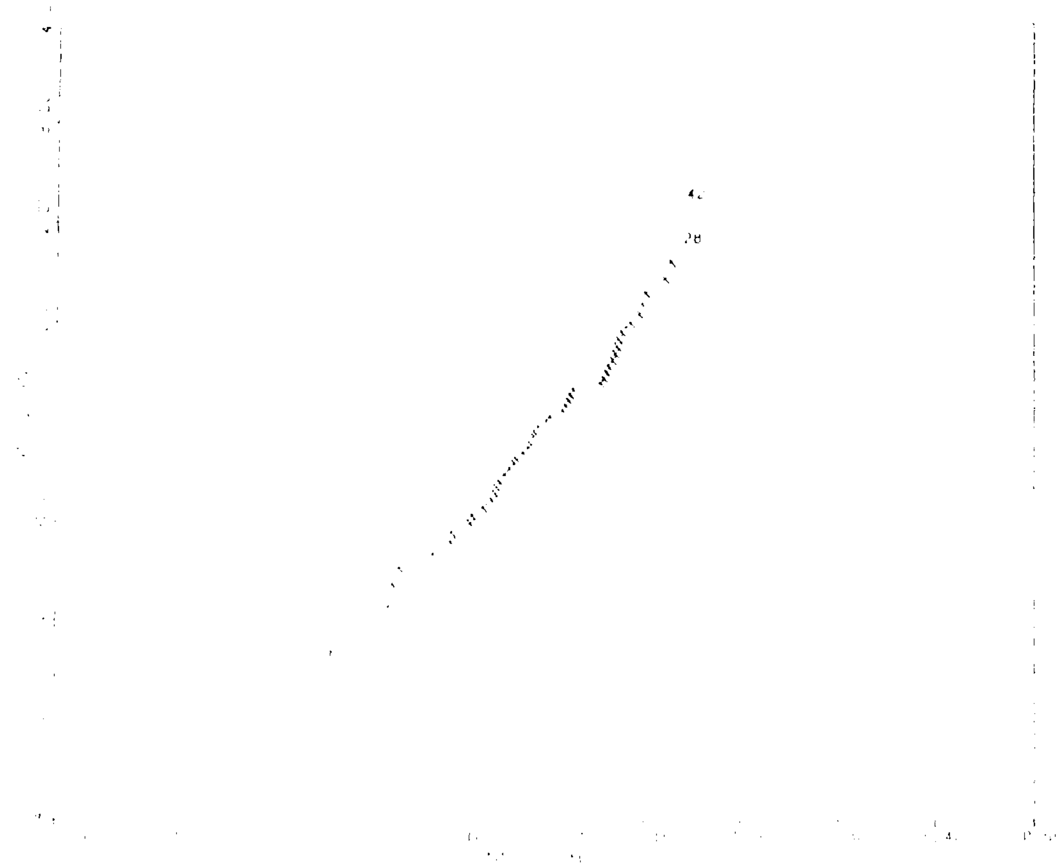

FIGURE 6.11a. Probability Plot of the Residuals from the Additive Model Fit to the Reciprocal of the Survival Times; $c = 0$

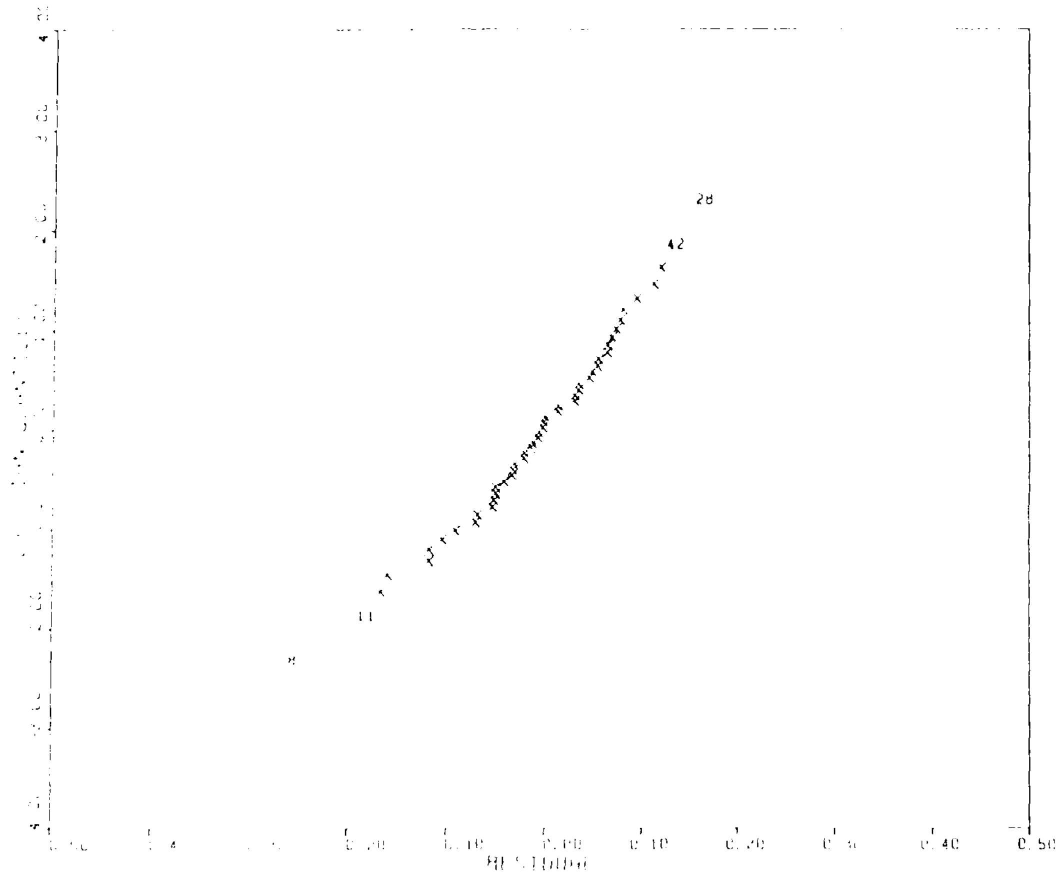

FIGURE 6.11b. Probability Plot of the Residuals from the Additive Model Fit to the Reciprocal of the Survival Times; c = 0.4

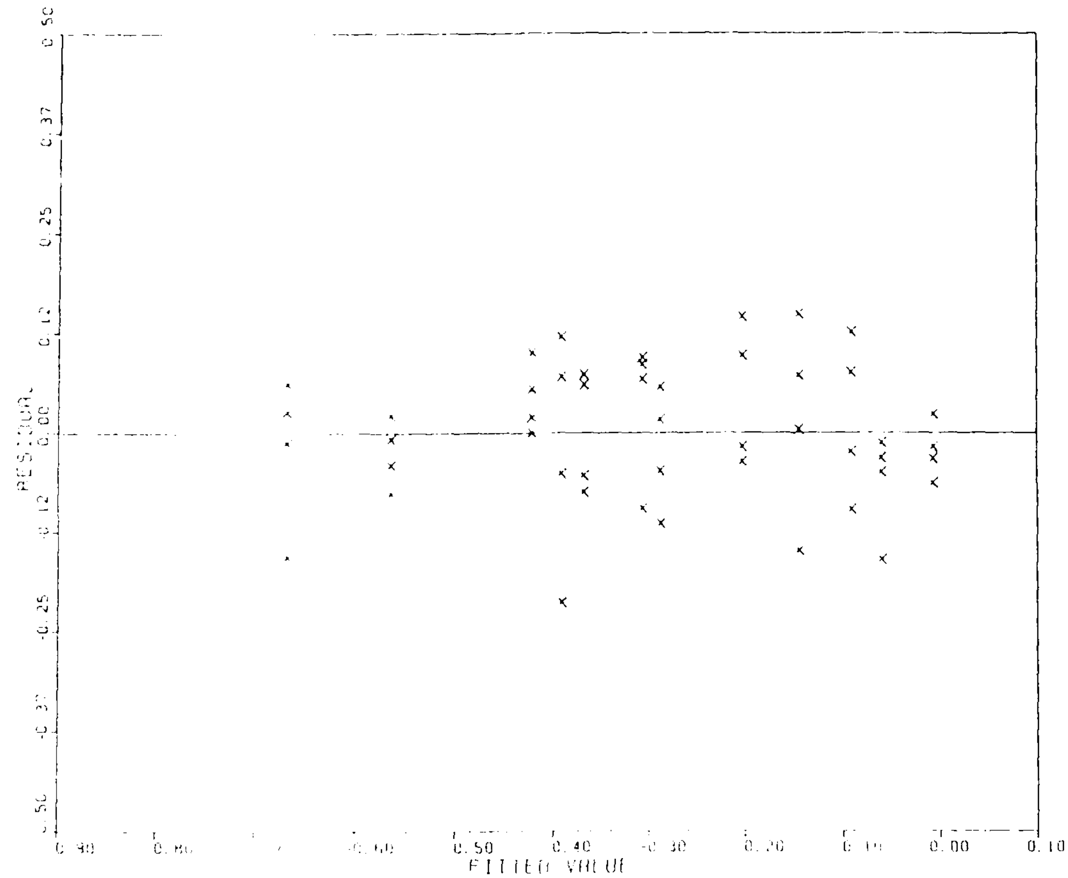

FIGURE 6.12. Residuals Versus Fitted Value for Death Rates with Additive Model; c = 0

6.5 Summary

In this part, we have illustrated the use of our model selection and test of fit procedures to analyze experimental data. The PSIC criterion can be useful for selecting regression and two-way layout models as well as ARMA models. As with the AIC, the PSIC criterion can be applied to a variety of parametric models. Our test of fit was shown to provide a measure of fit between the data and the selected model. The analysis of the examples yielded some surprising results.

For the linear regression example, the model-critical weights indicated the presence of potential outliers; however, the PSIC selection criterion did not select a smaller model with the outliers downweighted. For data with outliers downweighted, the model selected by the PSIC criterion is the one that best fits the remainder of the data. This model may have more or less parameters than the model selected without the outliers downweighted. Also the test of fit indicated that the quadratic model residuals are "supernormal" (Gentleman and Wilk, 1975). This can result when there are additional variables needed in the model or when a small sample size makes rejecting normality difficult. The above comments provide areas for further research.

The ARMA time series example illustrated that the analyst should use the largest possible value of c for the sample size, number of parameters and dimension of the data. A large value of c is required to adequately criticize the data and the model; however, the value of c must be kept small in order that the goodness of fit test remains conservative.

REFERENCES

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

Anderson, T. W. (1984). An Introduction to Multivariate Statistical Analysis, 2nd Ed., John Wiley and Sons, New York, NY.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). Robust Estimators of Location, Princeton University Press, Princeton, NJ.

Barnett, V., and Lewis, T. (1978). Outliers in Statistical Data, John Wiley and Sons, New York, NY.

Box, G. E. P., and Cox, D.R. (1964). "An Analysis of Transformations," J. R. Statist. Soc. B, 26, 211-252.

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

Daniel, C., and Wood, F. S., (1980). Fitting Equations to Data, 2nd Ed., John Wiley and Sons, New York, NY.

Delaney, N. K. (1979). Robust Procedures for the Multivariate Normal Distribution, Unpublished Ph.D. Thesis, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY.

Delehanty, T. A. (1983). Sensitivity Analysis in Statistics Based on Self-Critical Estimation, Unpublished Ph.D Thesis, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY.

Denham, M. J. (1974). "Canonical Forms for the Identification of Multivariable Linear Systems," IEEE Transactions on Automatic Control, AC-19, 646-656.

Draper, N. R., and Smith, H. (1966). Applied Regression Analysis, John Wiley and Sons, New York, NY.

Duncan, D. B., and Horn, S. D. (1972). "Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis," J. Amer. Statist. Assoc., 67, 815-821.

Dunsmuir, W., and Hannan, E. J. (1976). "Vector Time Series Models," Advances in Applied Probability, 8, 339-364.

Durbin, J. (1960). "The Fitting of Time Series Models," Rev. Int. Inst. Stat., 28, 233-244.

Fox, A. J. (1972). "Outliers in Time Series," J. R. Statist. Soc. B, 34, 350-363.

Gelb, A., (ed.) (1974). Applied Optimal Estimations, The MIT Press, Cambridge, MA.

Gentleman, J.F., and Wilk, M. B. (1975). "Detecting Outliers in a Two-Way Table: J. Statistical Behavior of Residuals," Technometrics, 17, 1-14.

Golomb, S. W. (1966). "The Information Generating Function of a Probability Distribution," IEEE Transactions on Information Theory, 75-77.

Hannan, E. J. (1969). "The Identification of Vector Mixed Autoregressive-Moving Average Systems," Biometrika, 56, 223-225.

Hannan, E. J. (1981). "Estimating the Dimension of a Linear System," Journal of Multivariate Analysis, II, 459-473.

Hannan, E. J. (1982). "Fitting Multivariate ARMA Models," Statistics and Probability: Essays in Honor of C. R. Rao, G. Kallianpur, P. R. Krishnaiah, J. K. Ghosh, eds., North-Holland Publishing Co., 307-316.

Hannan, E. J., and Quinn, J. B. (1979). "The Determination of the Order of an Autoregression," J. R. Statistic. Soc. B, 41, 190-195.

Harvey, A. C., and Phillips, G. D. A. (1976). "The Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Kalman Filtering," Unpublished Technical Report, Department of Social Sciences, University of Kent, Canterbury.

Harvey, A. C., and Phillips, G. D. A. (1979). "Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances," Biometrika, 66, 49-58.

Heyde, C. C. (1973). "An Iterated Logarithm Result for Martingales and its Application in Estimation Theory for Autoregressive Processes," J. Appl. Prob., 10, 146-157.

Hogg, R. V. (1974). "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory," J. Amer. Statist. Assoc., 69, 909-927.

Hosking, J. R. M. (1980). "The Multivariate Portmanteau Statistic", J. Amer. Statist. Assoc., 75, 602-608.

Huber, P. J. (1981). Robust Statistics, John Wiley and Sons, New York, NY.

Hwang, H-L. (1984). Some Fourier Methods for Estimation and Inference, Unpublished Ph.D. Thesis, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY.

Jones, R. H. (1980). "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations," Technometrics, 22, 389-395.

Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, J. Basic Engineering, 82, 35-45.

Kleiner, B., Martin, R. D., and Thomson, D. J., (1979). "Robust Estimation of Power Spectra," J. R. Statist. Soc. B, 41, 313-351.

Kullback, S. (1959). Information Theory and Statistics, John Wiley and Sons, New York, NY.

Kupferschmid, M., and Ecker, J. G. (1984). "EA3: A Practical Implementation of an Ellipsoid Algorithm for Nonlinear Programming," Unpublished Technical Report, Mathematical Sciences Department, Rensselaer Polytechnic Institute, Troy, New York.

Lawrence, C. E., Paulson, A. S., and Swope, G. R. (1986). "Model-Critical or Generalized Likelihood Analysis of Variance with Application to Prothrombin Times," Submitted for publication to IEEE Transactions on Man, Systems, and Cybernetics.

Levinson, N. (1947). "The Wiener RMS Error Criterion in Filter Design and Prediction," J. Math. Phys., 25, 261-278.

Mallows, C. L. (1973). "Some Comments on C_p," Technometrics, 15, 661-675.

Martin, R. D. (1981). "Robust Methods for Time Series," in Findley, D. F., (ed.), Applied Time Series Analysis II, Academic Press, New York, NY.

Martin, R. D., and Thomson, P. J. (1982). "Robust-Resistant Spectrum Estimation," Proceeding of the IEEE, 70, 1097-1115.

Mayne, D. Q. (1972). "A Canonical Model for Identification of Multivariable Linear Systems," IEEE Transactions on Automatic Control, AC-17, 728-729.

Neudecker, H. (1969). "Some Theorems on Matrix Differentation with Special Reference to Kronecker Matrix Products," J. Amer. Statis. Assoc., 64, 953-963.

Newbold, P. (1974). "The Exact Likelihood Function for a Mixed Autoregressive-Moving Average Process," Biometrika, 61, 423-426.

Nicholls, D. F. (1976). "The Efficient Estimation of Vector Linear Time Series Models," Biometrika, 63, 381-390.

Osborn, D. (1977). "Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes," J. R. Statist. Soc. B, 39, 114-118.

Paulson, A. S., and Delaney, N. J. (1981). "Self-Critical, and Robust, Procedures for the Analysis of Multivariate Normal Data," Unpublished Technical Report, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY.

Paulson, A. S., and Delehanty, T. A. (1983). "Sensitivity Analysis in Experimental Design," Computer Science and Statistics: Proceedings of the 14th Symposium on the Interface," 52-57.

Paulson, A. S., Presser, M. A., and Lawrence, C. E. (1981). "Self-Critical, and Robust Procedures for the Analysis of Binary Data," Technical Report 37-81-P1, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY.

Paulson, A. S., Presser, M. A., and Nicklin, E. H. (1982). "Self-Critical and Robust Procedures for the Analysis of Univariate Complete Data," Unpublished Technical Report, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY.

Paulson, A. S., Roohan, P. J., Hwang, H. L., and Fuller, M. (1987). "A Test of Goodness of Fit for Multivariate Normality," Submitted for Publication.

Paulson, A. S., and Swope, G. R. (1987). "Autoregressive Order Selection via Model-Critical Methods," IEEE Journal of Oceanic Engineering, OE-12, 75-79.

Poskitt, D. S., and Tremayne, A. R. (1982). "Diagnostic Tests for Multiple Time Series Models," The Annals of Statistics, 10, 114-120.

Presser, M. A. (1980). "Self-Critical and Robust Estimation Procedures for a Wide Class of Probability Models," Unpublished Ph.D. Thesis, Dept. of Operations Research and Statistics, Rensselaer Polytechnic Institute, Troy, NY.

Shapiro, S. S. (1980). How to Test Normality and Other Distributional Assumptions, American Society for Quality Control, Milwaukee, WI.

Shapiro, S. S., and Wilk, M. B. (1965). "An Analysis of Variance Test for Normality (Complete Samples)," Biometrika, 52, 591-611.

Suich, R., and Derringer, G. C. (1977). "Is the Regression Equation Adequate - One Criterion," Technometrics, 19, 213-216.

Wilson, G. T. (1973). "The Estimation of Parameters in Multivariate Time Series Models," J. R. Statist. Soc. B, 35, 76-85.

INTIAL DISTRIBUTION LIST

Addressee	No. of Copies
NAVSEASYSCOM (SEA-53D: CDR L. Schneider, D. Early, Dr. Y. Yam, Dr. C. Walker; PMS-409: G. Humes, S. Lose, D. McNamara)	7
ONT (OCNR-231: CAPT R. Fitch, Dr. T. Warfield, Dr. N. Booth, Dr. A. J. Faulstich)	4
DARPA (Dr. R. H. Clark, Dr. C. Stuart, Dr. A. Ellinthrope, Naval Technology Office)	4
SPAWAR (PDW-124, PMW-180T, Dr. J. Sinsky)	3
CNA	1
DTIC	2