ARIMA STATISTICAL MODELS OF TRANSIENT HEAT CONDUCTION ERRORS FOR AITKEN'S CONFIDENCE REGION

By

Irwin Perry Schisler

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Mechanical Engineering

1979

ABSTRACT

ARIMA STATISTICAL MODELS OF TRANSIENT
HEAT CONDUCTION ERRORS FOR
AIKEN'S CONFIDENCE REGION

By

Irwin Perry Schisler

Parameter estimation is central to the scientific analysis of data. When thermal properties are found from transient temperature measurements, the thermal parameters are usually determined by a parameter estimation procedure based on standard least squares. However, at the high sampling rates associated with minicomputers, the model for the errors is important because the temperature residuals display serial correlation and possible signatures. Hence, the usual model for the errors of $N(0,\sigma^2\ I)$ is inadequate and the model $N(\mu,\sigma^2\ V)$ is required.

The V matrix in the $N(0,\sigma^2\ V)$ model was represented by an ARIMA model having zero mean. An ARIMA (1,0,1) model fit the residuals better than other ARIMA (p,d,q) models. The best fitting ARIMA model was determined by applying Akaike's information criterion to the standard least squares residuals. The coefficients required in Akaike's criterion were estimated from a rapid solution of the Yule-Walker equations. A small-dominant parameter analysis was used to bound the change in the estimated thermal parameters caused by the

non-zero value of the μ vector. A one small parameter model that approximately fit the observed signatures in the residuals was used to predict changes in the thermal parameters that could be neglected.

The confidence region for the thermal parameters in the non-linear parameter estimation problem was approximated by an ellipse typically used for linear estimation problems. The size of this ellipse is related to the fractile. The fractile is computed from the central F-distribution when the V matrix is known. It was found that the fractile should be computed from the noncentral F-distribution when the V matrix is estimated and the sample size is small. The presence of ARIMA (1,0,1) errors is important in the statement of the correct confidence ellipse. When the $N(0,\sigma^2\ V)$ model is correct, the incorrect $N(0,\sigma^2\ I)$ model yielded a 95 percent confidence ellipse that in reality is only a 15 percent confidence ellipse.

When the $N(0,\sigma^2\ V)$ model is true, the thermal parameters should be estimated by Aitken's least squares. An iterative Aitken's least squares estimator was implemented within an existing standard least squares computer program for the heat conduction problem. The accuracy of this modified computer code was verified in a simulation study.

It was found that the effects of correlated errors for the available dynamic measurements and associated models can be treated without great difficulty. The parameter estimates using standard least squares tend to be very close to the values computed by the more powerful Aitken's least squares method. Hence the parameters required in the covariance matrix, V, can be estimated accurately and easily from the standard least squares residuals. The covariance

matrix for the thermal parameters can be evaluated by a simple recursive difference scheme rather than using the covariance matrix V directly in the matrix product. Thus, Aitken's confidence region and standard errors can be computed easily.

DEDICATION

My entry into a doctoral program evolved from my high school days where I was impressed by what proved to be the farsighted and sophisticated acquaintance with engineering attained by Mr. Wendell Martin, an omnibus and resourceful high school science teacher.

ACKNOWLEDGMENTS

Computer time for the Monte Carlo study was funded in part under National Science Foundation Grant GK-41495. Also used was subsidized computer time allocated at Michigan State University to doctoral research. An eight month graduate research assistantship was obtained through the Division of Engineering Research, J. W. Hoffman, Director.

I appreciate the efforts made by the original doctoral committee: K. J. Arnold, F. W. Bakker-Arkema, J. V. Beck (Chairman), and M. C. Potter. Professor Beck discussed and provided insights into his published research procedures for the correlated error problem. Professor Bakker encouraged entry into a doctoral program by financial support to participate in a low-grade reject heat project. Professors Arnold and Potter made suggestions on the proposed research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF IMPORTANT SYMBOLS

Symbol		Equation
a	Vector of normal random variates	(C.8)
a	Axis length of ellipse	(3.44)
a^{++}	Akaike's Information Criterion	(2.15)
A	Covariance matrix of the a vector	(2.10.A)
b	Estimate of the parameter β vector	(3.4)
c	Product of specific heat and density	(2.1)
d	Number of difference operations	
e	Vector of residuals, e = y - X b	(3.20.A)
F	Fisher-Snedecor distribution, $F(k,n-k,1-\alpha)$	(B.3)
F'	Noncentral F distribution, $F'(k,n-k,\lambda,1-\alpha)$	(3.13.B)
G	Fractile for confidence coefficient	(3.5)
k	Thermal conductivity	(2.1)
k	Number of thermal parameters	(2.1)
L	Autoregressive transformation matrix	(3.32)
m	Number of sensors in multisensor data	(3.19.A)
n	Number of measurements per sensor	(3.19.A)
p	Number of autoregressive parameters, ARIMA(p,d,q)	(A.1)
P	Transformation matrix, w = P a	(3.34)
q	Number of moving-average parameters, ARIMA(p,d,q)	(A.1)
Q	Matrix for quadratic form	(3.6)
s^2	Estimate of white noise variance, σ^2	(A.21)

Symbol		Equation
t	Student's t-distribution, $t(n-k,1-\alpha)$	(1.11)
u	Confidence region statistic	(3.15)
V	Matrix determined by ARIMA model, $W = \sigma^2 V$	(3.2)
w	Vector of additive errors, $y = X \beta + w$	(3.33)
W	Covariance matrix of the w vector	(3.2)
y	Vector of measurements	(3.22)
α	Confidence coefficient	(3.44)
β	Parameter vector in linear model	(3.22)
θ	Moving-average coefficient	(A.19)
λ	Noncentrality coefficient	(3.13.B)
μ	Mean value of residuals	(4.4)
ν	Angle of ellipse with k-axis	(3.44)
σ^2	White noise variance	(B.2)
$\emptyset$	Autoregressive coefficient	(A.18)
$\approx$	Is distributed as	(1.2)

CHAPTER I

DESCRIPTION OF THE PROBLEM

In Section 1.1 the importance to parameter estimation of modeling the errors is discussed. In Section 1.2 three features are isolated in modeling errors and in using the correct covariance matrix. The literature on these three factors is reviewed in Section 1.3 for simple physical models. This is extended in Section 1.4 to more complicated physical models that more closely resemble nonlinear estimation of the heat conduction problem. In Section 1.5 the problem investigated in this dissertation is defined as being only one aspect of ongoing parameter estimation studies at Michigan State University, and in Section 1.6 the topics presented in subsequent chapters are outlined.

1.1 Introduction

Mathematical modeling of physical phenomena is an important part of engineering and science. In many cases either the physical model is unknown or many of the associated constants are not known accurately enough. In such cases it is imperative to perform experiments and to take measurements. Parameter estimation is central to the scientific analysis of such data to determine parameter values and accuracies and also to provide insight for improving upon the mathematical model of the physical phenomena.

In addition to modeling the physical phenomena it is necessary to model the measurement errors to provide (1) the most accurate parameter estimates, (2) accurate confidence regions, and (3) optimal experiment designs. In dynamic experiments in which the variables are being repeatedly measured, the common assumption of uncorrelated measurements may not be valid. Heat conduction data can be used to test the development of these worthwhile parameter estimation techniques.

Currently optimal design in heat conduction is determined by selecting boundary conditions, by determining the duration of the experiments, and by choosing locations for the measurement sensors. In addition to these conditions the optimal design depends on the covariance matrix of the errors, denoted W. Different covariance matrices are obtained as the time between measurements is changed and there may be a best sampling rate. The W matrix can also be used in discriminating between rival mathematical models and in maximum a posteriori estimation.

Several examples can be given. Van Fossen (1973, Figure 4.2.9) found that the presence of correlated errors prevented effective discrimination between rival models describing the heating of a Bismuth-Lead alloy, because no technique was available to determine the error covariance matrix W. Beck (1975) cited as examples three heat conduction papers where correlated errors are undoubtedly present and modeling these errors could improve the parameter estimates or confidence regions. Clarke (1973) suggested using correlated error models to improve the fit of hydrologic groundwater equations.

1.2 Nomenclature for Correlated Error Studies

The specific techniques considered in this dissertation involve modeling errors in order to improve the estimation of thermal parameters and the statement of confidence regions for these parameters. Although finding thermal parameters requires nonlinear parameter estimation it is easier to define the techniques by the following linear model having additive errors; namely,

$$y = X \beta + w \tag{1.1}$$

where y is an n by 1 vector of measurements, X is a known matrix, β is the estimated k by 1 parameter vector, and w is an n by 1 vector of measurement errors. The standard assumption for w is

$$w \approx N(0,\sigma^2 I) \tag{1.2}$$

which means that the errors have zero mean, constant identical variance, and a normal distribution. The variance σ^2 is assumed to be estimated, throughout this dissertation. A $100(1-\alpha)$ percent confidence ellipse for the parameter vector b is

$$(b - \beta)^t X^t X (b - \beta) \leq k s^2 F(k,n-k,1-\alpha) \tag{1.3}$$

where b is the estimate of β, s^2 is an estimate of σ^2, and F is the fractile yielding probability α in the right tail of the F-distribution. The confidence interval for the j-th parameter is given by

$$b_j \pm t(n-k,1-\alpha) (s^2 (X^t X)^{-1}_{jj})^{\frac{1}{2}} \tag{1.4}$$

where t is the fractile for the Student's t-distribution, and $(X^t X)^{-1}_{jj}$ is the j-th diagonal entry in the inverse of the $X^t X$ matrix. Some researchers such as Draper and Smith (1966, p. 65) denote the estimated standard errors as

$$\text{est. s. e. } (b_j) = (s^2 \, (X^t \, X)^{-1}_{jj})^{\frac{1}{2}} \tag{1.5}$$

The confidence interval in Equation (1.4) depends on two factors: (1) the estimated standard errors, and (2) the fractile of the Student's t-distribution.

Beck (1975) indicated that errors in heat conduction experiments conducted at Michigan State University were not independent. Thus, the model that should be used for the errors is

$$w \approx N(0,\sigma^2 \, V) \tag{1.6}$$

where V is not equal to the identity matrix I. The presence of the correlated errors was noticed when a minicomputer was used for data acquisition. The correlation was noticed by plotting the residuals, e = y - X b, and testing whether the number of sign changes was abnormally small. In the statistical literature this sign test is also called a run test or a Geary test. The analysis by the sign test indicated that the errors are more correlated as the sampling rate on the minicomputer is increased. Accurate data were collected by Farnia (1976) using an IBM-1800 minicomputer data acquisition system using a rate of three-tenths of a second between measurements. I used these data to study the problems associated with including the V matrix in the estimation and confidence region equations. The data are undoubtedly representative of data from dynamic experiments in other engineering problems so that the analysis techniques developed also apply to these engineering problems.

When the covariance matrix V in Equation (1.6) is known the data and model can be transformed so that the errors satisfy Equation (1.2). Assuming that the V matrix can be factored as

$$V = L\,L^{t} \tag{1.7}$$

then the transformations

$$F = L^{-1}\,y \quad \text{and} \quad Z = L^{-1}\,X \tag{1.8}$$

make the linear physical model in Equation (1.1) become

$$F = Z\,\beta + L^{-1}\,w \tag{1.9}$$

where

$$L^{-1}\,w \approx N(0,\sigma^{2}\,I) \tag{1.10}$$

Thus the transformed errors are distributed according to the error model given by Equation (1.2). Hence the confidence interval for the j-th parameter in Equation (1.9) is

$$b_j \pm t(n-k,1-\alpha)\ \ (s^{2}\,(Z^{t}\,Z)^{-1}_{jj}\,)^{\frac{1}{2}} \tag{1.11}$$

where the estimated standard errors are given by

$$\text{est. s. e. } (b_j) = (s^{2}\,(Z^{t}\,Z)^{-1}_{jj}\,)^{\frac{1}{2}} \tag{1.12.A}$$

Using Equation (1.8) it is clear that Equation (1.12.A) can also be written as

$$\text{est. s. e. } (b_j) = (s^{2}\,(X^{t}\,V^{-1}\,X)^{-1}_{jj}\,)^{\frac{1}{2}} \tag{1.12.B}$$

which clearly shows that Equation (1.12) depends on the V matrix. When the V matrix is known, the standard errors are given by Equation (1.12) rather than the incorrect Equation (1.5).

When the V matrix is estimated, Theil (1971, p. 246) used the expression asymptotic standard errors since the asymptotic

estimates of the standard errors are based on the asymptotic approximation to the sampling distribution. Thus the confidence interval given by Equation (1.11) should be written as

$$b_j \pm G(1-\alpha) \quad (\text{est. s. e. } (b_j)) \tag{1.13}$$

where Equation (1.12) is used to estimate the standard errors. The more accurate the estimate of the V matrix, the better is the approximation that $G(1-\alpha)$ is $t(n-k,1-\alpha)$.

In terms of the above presentation the need for an error model given by Equation (1.6) instead of by Equation (1.2) has three features. First, the coefficients in the V matrix in Equation (1.6) must be estimated. Second, although less obvious than the first feature, the structure of the V matrix in terms of the L matrix in Equation (1.10) must be modelled. Third, an adequate approximation of $G(1-\alpha)$ in Equation (1.13) must be proposed.

1.3 Review of Correlated Error Studies

Three features of the correlated error problems were isolated in Section 1.2. In this section these three features will be discussed for simple physical models.

1.3.1 Estimation Using a Covariance Matrix

The first feature of the correlated error problems is estimating the standard errors. In order to compute the standard errors a computation scheme must be available to estimate the ARIMA parameters in the V matrix in Equation (1.7).

Maximum likelihood estimators are preferred when the physical parameters are estimated and the errors are correlated. The maximum

likelihood procedure uses as the cost function the logarithm of the probability density f_{ML} given by

$$f_{ML}(\beta) = (2\pi)^{-\frac{1}{2}n} (\det W)^{-\frac{1}{2}} \exp(- \tfrac{1}{2} S)$$

where

$$S = (y - X \beta)^t W^{-1} (y - X \beta) \qquad (1.14.A)$$

with the error covariance matrix given by

$$W = \sigma^2 V.$$

The statistical assumptions for using f_{ML} are normally distributed additive errors that have zero mean and are correlated. The covariance matrix W is estimated but X is errorless and there is no prior information. Beck and Arnold (1977) used a concise notation where a string of eight numbers referred to tabulated assumptions so that the above problem is denoted 11-01211 when the V matrix is estimated.

When the error covariance matrix is assumed to be known the maximum likelihood estimator reduces to the Aitken's least squares estimator which has the cost function

$$f_{ALS}(\beta) = (y - X \beta)^t W^{-1} (y - X \beta) \qquad (1.14.B)$$

Aitken (1935) proved that the cost function f_{ALS} yields the minimum variance estimator when W is known. The prefix A in ALS is used throughout this dissertation to denote Aitken with LS used to denote least squares.

For data acquired by a minicomputer the sample size can be large, say n = 1000, or even larger. Thus, it could be difficult to find the matrix W^{-1} even if the error covariance matrix W were known. In previous research at MSU by Van Fossen (1973) solving the problem

of inverting the W matrix seemed to be impractical. Beck (1974) approached the problem of inverting W by assuming the errors could be modelled by certain classes of autoregressive integrated moving-average (ARIMA) models, and found that the inverse for these models was related to certain differences of the data and model values. This occurs because the inverse itself is not needed; what is needed is only the evaluation of a quadratic form in Equation (1.14) containing the inverse of the error covariance matrix W.

For an autoregressive process with known coefficients, Beck (1974) proposed using the differences

$$Z_j(i) = e_j(i) - \phi_{1j}\, e_j(i-1) - \ldots - \phi_{pj}\, e_j(i-p)$$

where the residuals are

$$e_j(i) = y_j(i) - T_j(i)$$

and the subscripts refer to the j-th thermocouple at the i-th time for the data vector y and the temperature T predicted from the model. The Aitken's least squares cost function for the differenced residuals is

$$f_{ALS}(\beta) = \sum_{j=1}^{m} \sum_{i=1}^{n} Z_j(i)\, Z_j(i) \qquad (1.14.C)$$

where the limits on the summation is for m sensors and n measurements per sensor.

The implementation of parameter estimation using the cost function given by Equation (1.14.C) is considered in this dissertation when the ϕ_j's are estimated. Deutsch (1965, p. 64) stated the exact values for the elements of the error covariance matrix W are seldom

known. When the error covariance matrix is estimated, Dhrymes (1971), extending the doctoral thesis at MSU by Ruble, showed that a particular computational scheme for iterative Aitken's least squares yielded the same estimates as from the more complicated maximum likelihood cost function. The iterative Aitken's least square procedure uses a two step scheme. In the first step estimates for the physical parameters are found using an error covariance matrix that is regarded as the true error covariance matrix; in the second step estimates for the error covariance matrix are found using the residuals from the first step. The two step scheme is repeated until the parameter estimates converge.

1.3.2 Developing ARIMA Error Models

A second feature of the correlated error problem is finding an ARIMA model that is an accurate approximation to the errors. The ARIMA (AutoRegressive Integrated Moving-Average) model determines the V matrix in Equation (1.7) and Equation (1.6). Beck (1974) and Beck (1975) have suggested that the V matrix in heat conduction is not the identity matrix. Bard (1974, p. 248) stated that the estimation of the V matrix when serial correlation is unknown is relatively difficult. Apparently because replicated data are rarely available, Bard (1974, pp. 63-66) only considered contemporaneous correlation models, see Equation (2.10), when the V matrix is estimated.

It can be more effective to use a deterministic model rather than to find an autocorrelated model for the V matrix. Carr (1972) showed in a Cobb-Douglass model of Bell Telephone of Canada that the need

for a nondiagonal V matrix can be eliminated by changing the specification of the physical model. The residuals in the initial specification of the model had one sign before the introduction of automatic switching equipment and had the other sign after that time. Thus, the residuals had a pattern or signature. By introducing a binary variable for the effect of the introduction of the switching equipment, Carr (1972) obtained residuals in the revised model that had more sign changes and passed tests for independent errors. For these independent errors the error covariance matrix depends on the identity matrix I rather than on the nondiagonal matrix V. In engineering problems the V matrix may be associated with small parameter physical processes. Bard (1974, p. 202) stated:

> "particularly in cases where data are very accurate neglected effects outweigh random errors in measurement, and consequently nonrandomness of residuals is the rule, rather than the exception when models are fitted to good data."

Hence, it may be necessary to find a model for the V matrix.

Since the book by Box and Jenkins (1970) has become well-known, ARIMA(p,d,q) models for errors have become popular. The three order parameters indicate there are p autoregressive coefficients, d differences, and q moving-average coefficients that must be determined in order to find the ARIMA(p,d,q) model that is the best fit. Pandit (1973, pp. 17-49) gave an excellent historical review of the work in 1938 by Wold that is the basis of the ARIMA model, and Pandit (1973, pp. 170-184) interpreted ARIMA models in current mathematical terminology.

A discrimination criterion is required to find the order parameters: p, d, and q. Beck and Arnold (1977, p. 473) discussed discrimination procedures for the physical model but did not indicate a discrimination procedure for the ARIMA order. Gallant and Goebel (1976) stated that the referees of their paper recommended the final prediction error procedure developed in 1969 by Akaike to specify the order parameters. An improved version of this criterion was developed by Akaike (1972) and seems to be the criterion accepted by investigators in the area for specifying the order parameters. The criterion by Akaike (1972) was used in this dissertation.

To study the situation expected in real problems where close but not exact identification is made for the orders p and q, Gallant and Goebel (1976) estimated an ARIMA(2,0,0) model when the simulated real model was either white noise, ARIMA(0,0,4), or ARIMA(1,0,0). Schmidt (1970) also considered the sensitivity of the assumed order to the actual order. Both a linear and distributed lag model were considered by Schmidt (1970) with the linear model being

$$y_i = \beta + \alpha x_k + w_i. \tag{1.15}$$

Schmidt (1970) was able to study both the effect of assuming autocorrelation when it is absent, the effect of assuming independent errors (white noise) when autocorrelation is present, and the effect of assuming an insufficient number of terms in the autocorrelation model. The three error models used were:

(1) independent errors or white noise

$$w_i = a_i \tag{1.16}$$

(2) first order autoregressive errors

$$w_i = \emptyset\, w_{i-1} + a_i \ , \tag{1.17}$$

and (3) second order autoregressive errors

$$w_i = \emptyset_1\, w_{i-1} + \emptyset_2\, w_{i-2} + a_i \ . \tag{1.18}$$

The definition is introduced that the white noise vector satisfies

$$a \approx N(0,\sigma^2\, I) \ . \tag{1.19}$$

The system of equations based on the physical model in Equation (1.15) with α set equal to zero, and the statistical model given by Equation (1.17) will be used in example problems throughout this dissertation.

The sampling statistics reported by Schmidt (1971) were the variance of β and the bias of σ^2 because these quantities both occur in the confidence region. The best choice was to assume second order autoregressive errors when any of the three error models was true. Schmidt (1970, p. 18) stated this is true for infinite samples since any coefficient $\emptyset_j$ that is zero is estimated consistently as being zero, but this need not be true in finite (small) samples.

1.3.3 Generalized Standard Errors: Monte Carlo Studies

This section discusses recent Monte Carlo studies that explored the estimation of the standard errors and the accuracy of using the usual t-distribution. Hence, these studies investigated two features indicated in Section 1.2 for the correlated error problem.

Beck (1974) conducted and Beck and Arnold (1977, p. 318) interpreted a Monte Carlo study for correlated errors in a particular

linear model given by Equation (1.15) with errors given by Equation (1.17): parameter values used were $\alpha = 100$, $\beta = 0.1$, $\sigma^2 = 1$, and a range of values of $\emptyset = -1, -0.5, 0, 0.5, 1$. The sample size was $n = 60$ and $t = 34$ trials were used. Their results are presented in Table 1.1 in terms of two sampling statistics: the bias (rows 3 and 5) and the estimated standard errors (rows 4 and 6). The results of the study indicated that for positive values of $\emptyset$ the estimated standard error is underestimated because the average of the maximum likelihood estimate of $\emptyset$ is biased toward zero. The estimates of the physical parameters (α and β) are good.

Goebel (1974) investigated the distribution $G(1-\alpha)$where $G(1-\alpha)$ is contained in Equation (1.13). They considered the non-linear physical model that can be written as

$$y_i = \beta_1 \exp(\beta_2 x_i) + w_i \ , \tag{1.20}$$

and conducted a Monte Carlo study with sample size $n = 60$ and $t = 200$ trials. Three models were used for the errors: Equation (1.16) with $\sigma^2 = 0.25$, Equation (1.17) with $\sigma^2 = 0.25$ and $\emptyset = 0.735$, and the fourth order moving-average model denoted ARIMA(0,0,4)

$$w_i = 1.5\ a_i + a_{i-1} + 0.85\ a_{i-2} + 0.33\ a_{i-3} + 0.5\ a_{i-4} \ .$$

Goebel (1974) computed the t-distribution confidence coefficients from a fifteen cell histogram and applied a chi-square goodness of fit test to show that the distribution was not a t-distribution. This is evident from the empirical fractile points for the five ($\alpha = 95$) and ninety-five ($\alpha = 0.05$) percent confidence intervals that were -2.28 and 1.99 compared to the true values of -1.67 and

Table 1.1

Linear Model Monte Carlo Study*

Estimates	-1	-0.5	0	0.5	1
$\emptyset$	-0.969	-0.485	-0.024	0.430	0.881
s^2	0.958	0.966	0.966	0.963	0.933
$(a - \alpha)/\alpha$	0.00013	0.00016	0.00020	0.00014	0.00012
est. s. e.(a)/s. e. (α)	1.024	1.029	1.005	0.938	0.912
$(b - \beta)/\beta$	-0.00048	-0.00060	-0.00080	-0.00093	-0.00048
est. s. e. (b)/s. e. (β)	0.997	1.004	0.981	0.917	0.353

*Source: Beck and Arnold (1977, Table 6.18).

1.67, respectively. This flatness in $G(1-\alpha)$ at nine fractile points was confirmed by Gallant and Goebel (1976) in a Monte Carlo study involving Equation (1.20) with errors described by Equation (1.18).

The Monte Carlo studies described above used small sample sizes while it is conventional wisdom in econometrics that for large sample sizes the distribution is the same whether the statistical parameters are known or estimated consistently. There are several papers that allegedly show that the distribution is the same; e.g., Kmenta (1971, p. 507), Maddala (1971), Kmenta (1971, p. 529), and Schmidt (1970, p. 5). However, Schmidt (1976, p. 69) constructed a counter-example. This is for Equation (1.15) with α equal zero and the particular V matrix given by

$$V^{-1} = \text{diag}(1, \gamma, \gamma^2, \ldots, \gamma^{n-1}),$$

and the resulting distribution for the physical parameter is

$$n^{\frac{1}{2}} (b - \beta) \approx N(0,\sigma^2\delta). \tag{1.21}$$

From Equation (1.21) it is clear that the estimate of β is the same but the variance is increased by δ, when V is estimated. When γ is unity, the asymptotic variance is increased by about eight percent, in this counter-example. Schmidt (1976, p. 71) suggested that his experience indicated the large sample variance has δ equal to unity when the errors are described by ARIMA models.

The standard errors can also be increased if Equation (1.5) is used instead of Equation (1.12); i.e., if standard least squares is used but is not valid. Magness and McGuire (1962) showed that the increase in the standard errors depends on the particular combination of X matrix from Equation (1.1) and V matrix from

Equation (1.6). This dependency affects conclusions from a Monte Carlo study since only specific combinations of X and V are used.

1.4 Review of Simulation Studies

The purpose of this section is to review the use of small trial procedures to understand complicated physical models in contrast to the procedures used for the simple physical models presented in Section 1.3.

1.4.1 Standard Errors in Heat Conduction

In a Monte Carlo study, Beck and Arnold (1977, p. 401) made an assessment of the effect of correlated errors on a heat conduction problem. The physical model was heat conduction in a semi-infinite solid subject to a step change in surface temperature. The temperature in the solid was given as

$$T(x,t) = T_0 + (T_\infty - T_0)\ \mathrm{erfc}(x\ (4\ \alpha\ t)^{-\frac{1}{2}}\ )$$

where the physical parameter is α the thermal diffusivity, the temperature at the surface is maintained at value T_0 and the temperature infinitely far into the solid is T_∞, t is time, and T(x,t) is the temperature at position x and time t. In order to simulate actual heat conduction measurement conditions two sensors were used (m = 2) and they were located at 0.125 and 0.250 inches from the surface, and n = 28 measurements were generated for each sensor. The error model was ARIMA(1,0,0) given in Equation (1.17) and three values of the coefficient $\emptyset$ were used: 0, 0.5 and 0.9. The estimated standard error was given as

$$\text{est. s. e.}(\alpha) = (s^2 (X_1^t V^{-1} X_1 + X_2^t V^{-1} X_2)^{-1})^{\frac{1}{2}}$$

where the subscripts 1 and 2 on the X vector indicate the two measurement locations. The X vector is the partial derivative of T with respect to α for both locations and for all the experimental times. In Table 1.2 Z^tZ is defined by

$$Z^t Z = X_1^t V^{-1} X_1 + X_2^t V^{-1} X_2.$$

The quantities calculated for Table 1.2 are related to the problems for correlated errors isolated in Section 1.2. The results for estimating the V matrix are that the standard errors are sensitive to the true value of the autoregressive coefficient $\emptyset$ with the standard errors increasing by a factor of about seven as the true autoregressive coefficient increased from $\emptyset = 0$ to $\emptyset = 0.9$. The results for approximating $G(1-\alpha)$ are that in three out of five cases the estimated value is within the standard deviation so that the distribution of the standard errors is consistent with a t-distribution.

1.4.2 Confidence Regions for Nonlinear Parameter Estimates

In real engineering problems the models have nonlinear parameters and a simulation study using these models can be expensive. Although the distribution of $G(1-\alpha)$ in Equation (1.13) when the errors satisfy white noise as in Equation (1.2) was shown by Ivanov (1972) and Chambers and Ertel (1975) to be consistent and normal when the sample size is large, a procedure is required for small samples. This was investigated by Tierney (1971) for the standard least squares procedure where β is estimated from the equation

Table 1.2

One Parameter Heat Conduction Simulation Study*

Case	$\emptyset$	α X 10^4 (ft^2/hr)	s^2	$Z^t Z$ X 10^{-6}	est. s. e. (α) X 10^4
1	0	99.71	1.080	197	0.740
2	0	100.64	0.985	192	0.716
3	0	100.17	1.067	195	0.740
4	0	99.29	0.955	199	0.693
5	0	98.56	1.191	202	0.768
6	0.5	99.49	1.106	52.4	1.45
7	0.5	101.88	0.977	49.7	1.40
8	0.5	100.39	1.076	51.3	1.45
9	0.5	98.69	0.958	53.4	1.34
10	0.5	97.22	1.024	55.3	1.36
11	0.9	98.6	1.176	4.275	5.25
12	0.9	115.3	0.971	3.09	5.61
13	0.9	101.7	1.079	4.00	5.19
14	0.9	97.2	1.003	4.40	4.97
15	0.9	91.0	1.125	5.06	4.72

*Source: Beck and Arnold (1977, Table 7.13).

$$(y - T)^t T_\beta = 0$$

which Tierney (1971, p. 115) approximates by a seven term series

$$\sum_{i=0}^{6} A_i (\beta - b) = 0 \qquad (1.22)$$

that is solved for $(\beta - b)$ in terms of the A_i's by SNOBOL, a symbolic manipulation computer language. From $(\beta - b)$ the moments of the distribution was found for the nonlinear parameters and hence the mean and variance of the probability distribution was calculated. It is beyond the scope of this dissertation to investigate the probability distribution in this much detail, and therefore normal and F distributions will be used.

Not only is it difficult to determine the probability distribution for nonlinear parameters using Equation (1.22) but also it is difficult to obtain moments from a simulation study. A relevant example is the nonlinear solid state physics model investigated by both Pfeiffer and Lichtenwalner (1974) and by Chambers and Ertel (1975) which is

$$y_i = \beta_1 - \frac{\beta_2}{1 + 4\,\beta_3^{-2}\,(x_i - \beta_4)^2} + w_i$$

where the additive errors w_i are white noise as in Equation (1.2). Pfeiffer and Lichtenwalner (1974) used t = 400 trials to plot ordered estimates of β_3 on quantile paper to see whether the distribution is consistent with a normal distribution. The t = 400 trial study was "extremely expensive of computer time." Chambers and Ertel (1975) repeated this study with t = 100 trials in order to find an analysis

procedure adequate for small trials and they investigated: the quantiles from the plotted ordered values, a linearized estimate as Equation (1.5) and a quadratic estimate. Again it is beyond the scope of this dissertation to investigate the probability distribution in this much detail.

1.5 Problem Statement

The general emphasis of this research stems from Beck and Arnold (1977, p. 289) who stated:

> "In order to present meaningful confidence regions it is necessary that the underlying assumptions are valid. Two assumptions frequently violated in scientific work are that the errors have zero mean and that the errors are uncorrelated. Erroneously taking these assumptions to be true has led many to present overly small confidence regions."

The implication of this statement is that the model considered for the errors should be

$$w \approx N(\mu, \sigma^2 V). \tag{1.23}$$

This dissertation concentrates on the error model given by Equation (1.6) which differs from Equation (1.23) by the assumption that $\mu = 0$. The effect of a non-zero μ created by a signature in the residuals is not considered in detail.

Based on the three features of the correlated error problem outlined in Section 1.2 and the review of the literature, the topics that need to be investigated in the correlated error problem are as follows. First, it is necessary to find simultaneous estimates of both the physical and statistical (ARIMA) parameters; this will be implemented by modifying an existing computer code (called PROPTY) so that it computes iterative Aitken's least squares estimates.

Second, the ARIMA model for the autocorrelated errors is found from real data; the residuals from standard least squares estimation with computer program PROPTY will be used to find the proper ARIMA model. Third, the Aitken's least squares and standard least squares confidence regions are constructed from real data and the accuracy of using Equation (1.12) can be compared to using the incorrect Equation (1.5); a Monte Carlo study is also conducted with t = 20 trials to gain insight into the effect of estimating the error covariance matrix V.

The originality in this dissertation is (1) to isolate relevant procedures primarily from the econometric literature in order to present confidence regions when autocorrelated errors are present, (2) to make improvements in the procedures so that they can be implemented in a problem where heat conduction parameters are estimated, and (3) to apply the procedures to available heat conduction data in order to show how the procedures can be used.

The illustrative heat conduction problem is for the estimation of thermal conductivity and specific heat in an Armco iron disk using transient temperature and heat flux data. This analysis procedure is undoubtedly characteristic of many other related parameter estimation problems in mechanical, agricultural, civil, and chemical engineering.

In Section 1.1 and Section 1.2 the general importance of modeling the covariance matrix is discussed. A more specific listing of factors in parameter estimation that are affected by the presence of a covariance matrix is outlined in National Science Foundation (NSF)

proposal GE-41495. This dissertation solves some but not all of the problems suggested in NSF proposal GE-41495. In this proposal to NSF five modifications in computer program PROPTY were indicated: (1) to extremize a maximum likelihood cost function that contains an error covariance matrix, (2) to test the resulting computer routine for numerical stability since the additional ARIMA coefficients are estimated, (3) to incorporate sequential estimation, (4) to utilize prior information about the error covariance matrix from previous experiments, and (5) to replace the estimate of the standard errors from PROPTY with an estimate of Aitken's (or maximum likelihood) standard errors.

These five aspects are motivated in part by the desire to have an estimator usable on a minicomputer data acquisition and analysis system. It is reasonable to expect researchers to try using ARIMA models for autocorrelated errors in data acquired by laboratory minicomputers, because procedures were developed in the early 1960's to handle ARIMA errors associated with tracking data processed by spacecraft minicomputers. An explicit algorithm for spacecraft minicomputer data was developed by Blum (1961) for sequential estimation when the coefficients in the autoregressive model are known.

The third and fourth aspects are considered by Beck and Arnold (1977) in computer program NLINA, but are not considered in this dissertation for several reasons. Beck and Arnold (1977, p. 276) gave equations for sequential estimation when the correlated errors are known. However, there is a question of the usefulness of sequential estimates when the goal is to construct a confidence region with

unknown ARIMA parameters. Odell and Lewis (1971) found that although autoregressive parameters could be estimated as part of a recursive algorithm, the statistical properties of the estimated parameters could not. The fourth aspect, which is including prior information, creates two difficulties: as stated for aspect three the residuals may not have the same statistical properties, and the additional terms in the quadratic form for the standard errors alters the statistical distribution of the quadratic form. Hence aspects three and four are not considered in this dissertation.

This dissertation was written to explore the fifth aspect of the research proposed in GK-41495 and indirectly the first and second aspects are considered. The first aspect is achieved by using iterated Aitken's least squares estimation which Dhrymes (1971) showed converged to the maximum likelihood estimates. The second aspect is not expected to be a problem because Goldstein and Swerling (1970) found that two iterations of iterated Aitken's least squares yielded estimates close to those obtained with a known covariance matrix V. The fifth aspect (in its simplest form) merely requires using the estimates from Aitken's least squares in Equation (1.12.B).

1.6 Plan of Investigation

The solution of the problem outlined in Section 1.5 proceeds as follows. Chapter II describes some experimental data obtained at Michigan State University that is analyzed to identify the proper ARIMA model. Chapter III is a development of models for the confidence regions and also a Monte Carlo study of the confidence region

when the V matrix is estimated. Also in Chapter III confidence regions for real heat conduction data are presented. Chapter IV discusses additional aspects of the fitted ARIMA model. Chapter V gives the major conclusions for the correlated error problem.

CHAPTER II

STATISTICAL ANALYSIS OF HEAT CONDUCTION DATA

This chapter develops autoregressive integrated moving-average (ARIMA) models for the errors. In Section 2.1 a procedure is outlined to obtain temperature residuals by estimating thermal parameters from the experimental transient temperature data using standard least squares. In Section 2.2 the residuals are plotted and a general model for the errors is outlined. In Section 2.3 it is shown that the residuals are not correlated contemporaneously, but are correlated serially. In Section 2.4 an ARIMA(1,0,1) model for the residuals is shown to have the best fit based on computed values of Akaike's Information Criterion, and by the agreement of estimated and predicted coefficients in the ARIMA(1,0,1) model when the sampling rate changes from 0.3 to 0.6 seconds between measurements. In Section 2.5 additional comments about specifying the ARIMA model as ARIMA(1,0,1) are given.

2.1 Description of Farnia's Transient Temperature Data

2.1.1 Heat Conduction Model

The conditions in the heat conduction experiments performed by Farnia (1976) are described in this section. The specimen was Armco Magnetic Ingot Iron for DC Applications. The shape of the specimen is a disk 3 inches in diameter and 1 inch in thickness.

This disk is insulated on one face (and on the edge) and is heated on the other face. The total duration of a test was about 40 seconds. A known constant heat flux was applied during a nominal 15.3 second interval at the start of the experiment.

The physical model to describe these experimental conditions is the heat conduction equation in one dimension

$$c \frac{\partial T}{\partial t} = \frac{\partial}{\partial x} \left(k \frac{\partial T}{\partial x} \right) \tag{2.1}$$

subject to the conditions

$$T(x,0) = T_o \tag{2.2.A}$$

$$k \frac{\partial T(0,t)}{\partial x} = 0 \tag{2.2.B}$$

$$k \frac{\partial T(L,t)}{\partial x} = q_o H(t) H(s - t) \tag{2.3}$$

The thermal properties of Armco iron at T_o for a small temperature rise are two parameters in the model: k the thermal conductivity, and c the product of the density and the specific heat at constant pressure. The reader is referred to Table 2.2 for typical numerical values of k and c for the range of conditions covered in the ten experiments conducted by Farnia (1976). The density is constant and is taken by Farnia (1976, p. 37) as 490.71 lbm ft^{-3}.

The estimates of k and c in this section of the dissertation are obtained by minimizing the standard least squares cost function

$$f(k,c) = \sum_{j=1}^{m} \sum_{i=1}^{n} (T(x_j,t_i) - T_e(x_j,t_i))^2 \tag{2.4}$$

where T_e is the experimental temperature, T is the calculated temperature, and there are m sensors at locations denoted by x_j,

and there are n measurements per sensor taken at times denoted by t_i. The procedure used to establish T_e and q_0 are discussed in the following sections. The method used by Farnia (1976) to obtain his estimates of k and c did not involve Equation (2.4).

2.1.2 Experimental Conditions

Thermocouple millivoltages were processed by Farnia (1976) in the following manner. The millivoltages from each thermocouple were amplified and sent to an IBM-1800 computer. The computer program to convert voltages to temperatures was developed by Van Fossen (1973), who removed 60 cycle noise by averaging a voltage at a given time with another voltage one-hundred-twenty-th (1/120) of a second later. This voltage and a similarly processed voltage from a thermocouple 180° away on the same face of the disk provided the average voltage of two thermocouples and is called the output from one sensor. The sensor voltage is converted to a temperature value using regression equations for a J-type thermocouple. Farnia (1976, p. 48) used four thermocouples per face on each of the two symmetrically placed disks; that is, four thermocouples on each of the four faces. By averaging these sixteen thermocouple readings, eight sensor readings were available for analysis: four from sensors on the heated surface and four from sensors on the insulated surface.

The experiments by Farnia (1976) can be thought of as a single disk with four sensors on its heated face and four sensors on its insulated face. A three-tenths of a second sampling rate was selected for the IBM-1800, and 144 temperature values were recorded

during the 43 seconds of the experiment and pre-experiment and post-experiment times. The temperature of the disk was approximately uniform at both the start and at the end of the 43 second time period. Using thermocouple response values by the IBM-1800 and these values were punched on cards by the IBM-1800. The data on these cards were processed by computer program COND to yield values that were aligned at the start and end of the experiment; these values were used as the experimental temperatures.

Farnia (1976) found it necessary to correct the temperature for each sensor by computer program COND because the same temperature was not measured by thermocouples at the same surface despite the accuracy of the calibration. The regression model he used to line up the temperature at the beginning and end of the experiment was:

$$T^*_I = \alpha_1 + \beta_1 T_I$$

$$T^*_H = \alpha_2 + \beta_2 T_H$$

where T* is the corrected temperature, T is the temperature recorded by the IBM-1800, and the subscripts are I for insulated surface and H for heated surface. The regression constants α_1, β_1, α_2, and β_2 are solutions of the system of four equations

$$\tfrac{1}{2}(T_{IB} + T_{HB}) = \alpha_1 + \beta_1 T_{IB}$$

$$\tfrac{1}{2}(T_{IB} + T_{HB}) = \alpha_2 + \beta_2 T_{HB}$$

$$\tfrac{1}{2}(T_{IF} + T_{HF}) = \alpha_1 + \beta_1 T_{IF}$$

$$\tfrac{1}{2}(T_{IF} + T_{HF}) = \alpha_2 + \beta_2 T_{HF}$$

where the subscript B denotes initial time and F denotes final time.

Computer program COND determines four sets of α's and β's with one set for each pair of thermocouples. Because they are in adjacent columns of the IBM-1800 punched data cards, the pairs are taken in the order with thermocouple numbers (1,5), (2,6), (3,7), and (4,8). Each pair has one thermocouple at the heated surface and one thermocouple at the insulated surface. This arrangement facilitates the estimation of k and c by the integral method to be described in the next section since one estimate of k and c is obtained from each pair of sensors with the average value used as the final estimate. This alignment of the sensors affects the preheat-flux residuals in Section 2.3.4.

The data obtained by Farnia (1976) are summarized in Table 2.1. There were ten cases with starting temperatures ranging from 80 to 360 degrees Fahrenheit, and two levels of the applied heat flux. These two levels of the applied heat flux produced respective temperature rises of 15 and 30 degrees Fahrenheit.

2.1.3 Farnia's Estimates of the Heat Flux

The heat flux q(t) was not recorded in detail by Farnia (1976) because the parameters were estimated with a method requiring only the total heat flux. The estimation method was developed by Beck and Al-Araji (1974). It is useful to briefly indicate the equations used in this method. Assume that the surface of the disk at x = L is insulated and the surface at x = 0 has a heat flux q(t). The total heat input Q is given by

$$Q = \int_0^\infty q(t)\, dt$$

is assumed to be known but the detailed flux q(t) is not necessarily known. The known value of Q is computed from a scalar equation written as

$$Q = V^2 \, (2 \, R \, A)^{-1}$$

where V is the measured voltage drop and A is the known area of the face of the disk. The resistance of the heater is modelled as a linear regression on the average of the initial and final temperature

$$R = \alpha_3 + \beta_3(T_{IB} + T_{HB} + T_{IF} + T_{HF}) \, .$$

The conductivity and specific heat are estimated from

$$k = \frac{L \, Q}{2 \int_0^{\infty} (T(0,t) - T(L,t))dt} \qquad (2.5.A)$$

$$c = \frac{Q}{L \, (T(x,\infty) - T(x,0))} \, . \qquad (2.5.B)$$

Farnia (1976) applied the heat flux, q(t), by a thin electrical heater confined between the two identical Armco iron disks. Each disk was three inches in diameter and one inch thick. Half the heat generated by the heater went into each disk. After some initial sensor temperatures were recorded, a constant voltage was applied by a D.C. power supply to the heater for approximately 15.3 seconds. The resulting nearly constant heat flux can be described mathematically as

$$q(t) = q_o \, H(t - t_B) \, H(t_E - t) \qquad (2.6)$$

where q_o is a constant, t is time, H is the unit step function, and t_E and t_B are the times the heat flux ends and begins, respectively.

The constant q_o is obtained from the known value of Q by the expression

$$q_o = (t_E - t_B)^{-1} Q$$

The duration of this heat flux was determined by an electric timer. However, neither t_B nor t_E was recorded, and these times do not necessarily coincide with times that the temperatures were recorded. In Farnia's analysis these times were not needed while in computer program PROPTY, the times t_B and t_E are needed. An estimate of the time the heat flux ends is obtained by noting the time when the temperature at the heated surface ceases to increase. An estimate of the starting time, t_B, is obtained from t_E and the nominal heater time of 15.3 seconds. A closer estimate of t_B is obtained by using the average initial temperature and the fact that initially the temperature increases as the square root of time.

2.1.4 Estimates of the Thermal Parameters

Table 2.1 displays estimates of the thermal properties k and c from computer program COND. Approximately two levels of the applied heat flux were used in the experiments. The symbol T_{min} indicates the initial temperature of the specimen before the heat flux was applied. A range of initial temperatures was used so that the data are not replicated data. T_{rise} indicates the temperature rise caused by the heat flux; notice that it is proportional to the heat flux.

A nonlinear estimation computer program with a standard least squares cost function was also used to estimate the physical (thermal) properties. The estimates of the physical parameters are tabulated

Table 2.1

Parameters Estimated by Computer Program COND

Case	q Btu $sec^{-1} ft^{-2}$	T_{Min} Fahrenheit	T_{Rise} Fahrenheit	k Btu $hr^{-1} F^{-1}$	c Btu $ft^{-3} F^{-1}$
1	5.17	88.82	29.6	42.5635	55.2654
2	2.77	150.15	16.5	39.5384	55.7381
3	5.12	157.32	31.0	39.8046	56.4074
4	2.65	192.27	16.5	36.6699	55.9789
5	5.09	200.04	32.1	38.5519	57.0874
6	5.07	253.55	31.1	36.5940	57.8671
7	2.64	300.90	16.8	36.4586	59.2804
8	5.20	306.00	32.6	35.9029	60.0224
9	2.63	357.73	16.7	35.8161	62.0250

in Table 2.2, with the ten cases in this table arranged in the order of increasing initial temperature, T_{min}. Table 2.3 gives the starting time t_B and ending time t_E for the heat flux that were not recorded by Farnia (1976) and had to be assigned. The thermal conductivity k and the specific heat c were estimated using computer program PROPTY developed by Beck at MSU. Also given in Table 2.2 are the differences between the estimates from COND and PROPTY divided by the estimate from PROPTY. The differences are small, being less than 0.5 percent for k and 2 percent for c. Hence if these levels of accuracy for the method of analysis are satisfactory, the equations used in COND (Equations [2.5.A] and [2.5.B]) might be preferred over the least squares method. Note, however, that the method used in COND does not have an established statistical basis so that confidence intervals can not be found.

2.2 Description of the Covariance Matrix Problem

In Section 2.2.1 the residuals are presented. In Section 2.2.2 a model that accounts for both contemporaneous and serial correlation is outlined. Contemporaneous correlation is correlation between residuals from different sensors at a specified time. Serial correlation is correlation between residuals at a given sensor at adjacent times.

2.2.1 Computed Temperature Residuals

The temperature residuals computed from the ten sets of experimental data collected by Farnia (1976) are plotted in Figures 2.1 through 2.10 for the insulated surface and in Figures 2.11 through 2.20

Table 2.2

Parameters Estimated by Computer Program PROPTY

Case	k Btu $hr^{-1} ft^{-1} F^{-1}$	c Btu $ft^{-3} F^{-1}$	rms F	$\delta k/k$	$\delta c/c$
0	43.35722	55.66701	0.32399	N.A.	N.A.
1	42.83348	56.19796	0.56289	-0.0063	-0.0166
2	40.16488	56.65857	0.26681	-0.0156	-0.0162
3	40.13055	56.97504	0.68491	-0.0081	-0.0100
4	37.62633	57.96610	0.32132	-0.0254	-0.0342
5	38.60237	57.79679	0.86569	-0.0013	-0.0123
6	36.71381	58.35350	0.72629	-0.0033	-0.0083
7	36.42638	59.61019	0.46212	0.0009	-0.0055
8	35.86396	60.46524	0.83699	0.0011	-0.0073
9	35.37538	61.62738	0.47600	0.0125	0.0065
				Average	
				-0.0051	-0.0188

for the heated surface. Specific features of these plots are discussed when the residuals are modelled. Table 2.3 also indicates the quality of the residuals that are displayed in Figures 2.1 through 2.20. The magnitudes of the residuals are not of the same magnitude in all ten cases with the spread in the magnitudes of the residuals falling roughly into three groups: 0.8, 1.6, and 3.2 degrees Fahrenheit. These magnitudes are the basis of the quality index denoted A, B, and C. The residuals at the insulated surface are smaller in general than those at the heated surface. The smaller of the two heat fluxes is associated with the smaller residuals. The quality of the residuals is used as a factor in selecting the ARIMA model in a well-executed experiment in Table 2.8.

Table 2.3

Parameters Assigned by Studying the Residuals

Case	t_B sec	t_E sec	t_F sec	Quality
0	3.45	18.6	30.0	B A = A
1	2.7	18.0	27.9	B B = B
2	6.3	21.6	31.5	A A = A
3	4.8	20.1	30.0	B C = C
4	5.1	20.4	30.3	A A = A
5	5.1	20.4	30.3	A C = C
6	4.62	19.96	29.7	B C = C
7	6.17	21.53	31.2	A B = A
8	5.26	20.51	30.3	B C = C
9	5.7	21.0	30.9	B B = B

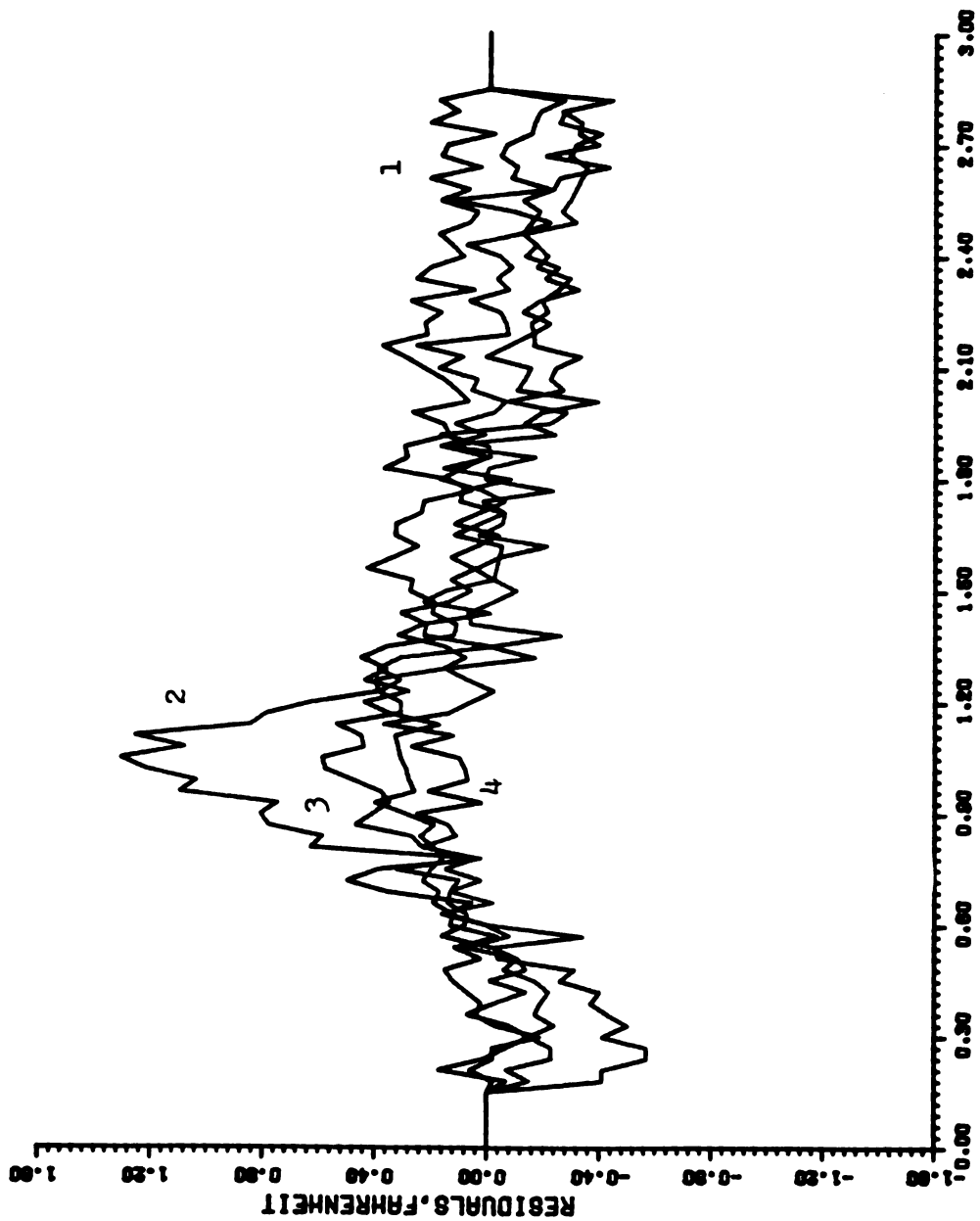

Figure 2.1. Insulated surface residuals: Case-0

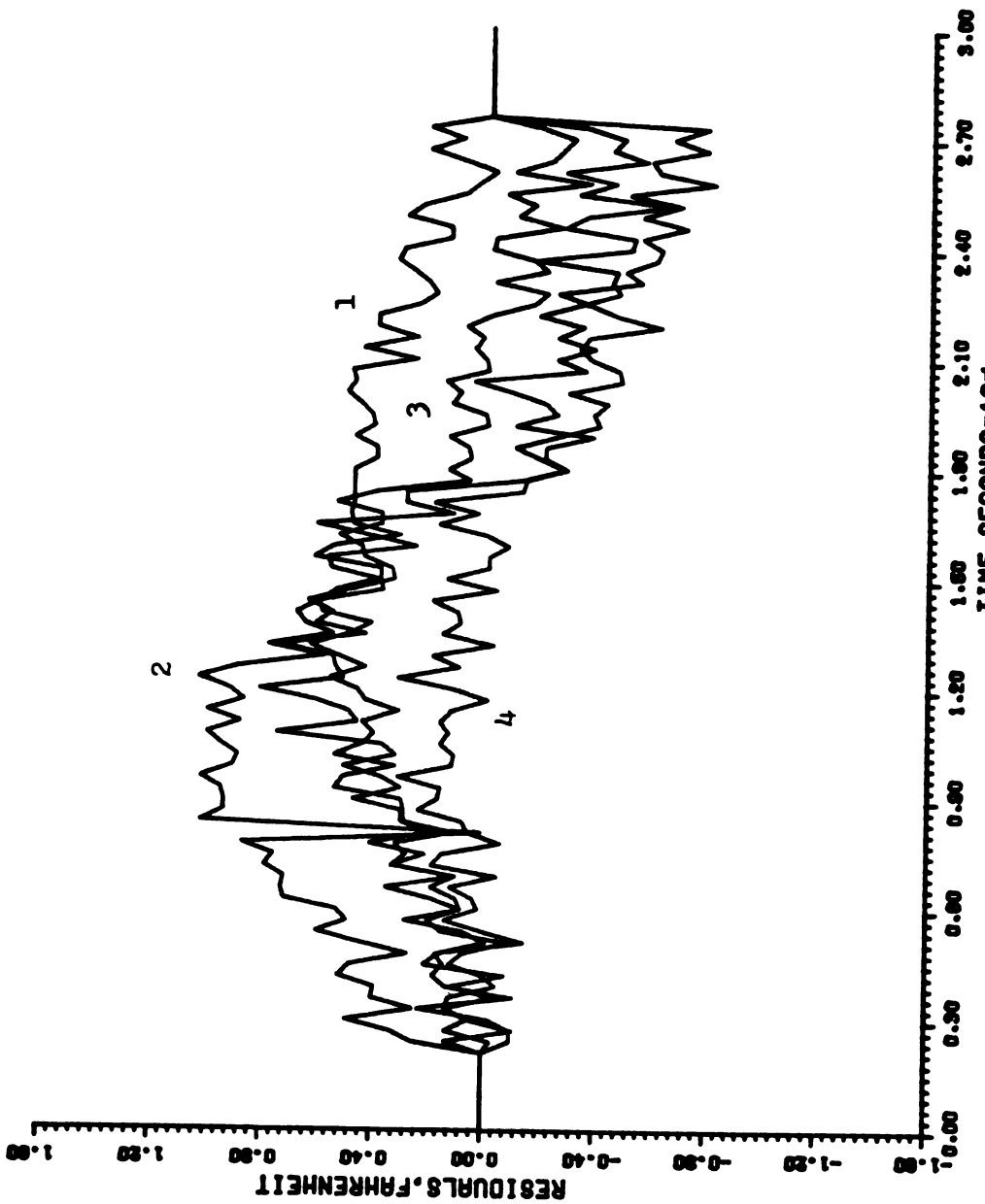

Figure 2.2. Insulated surface residuals: Case-1

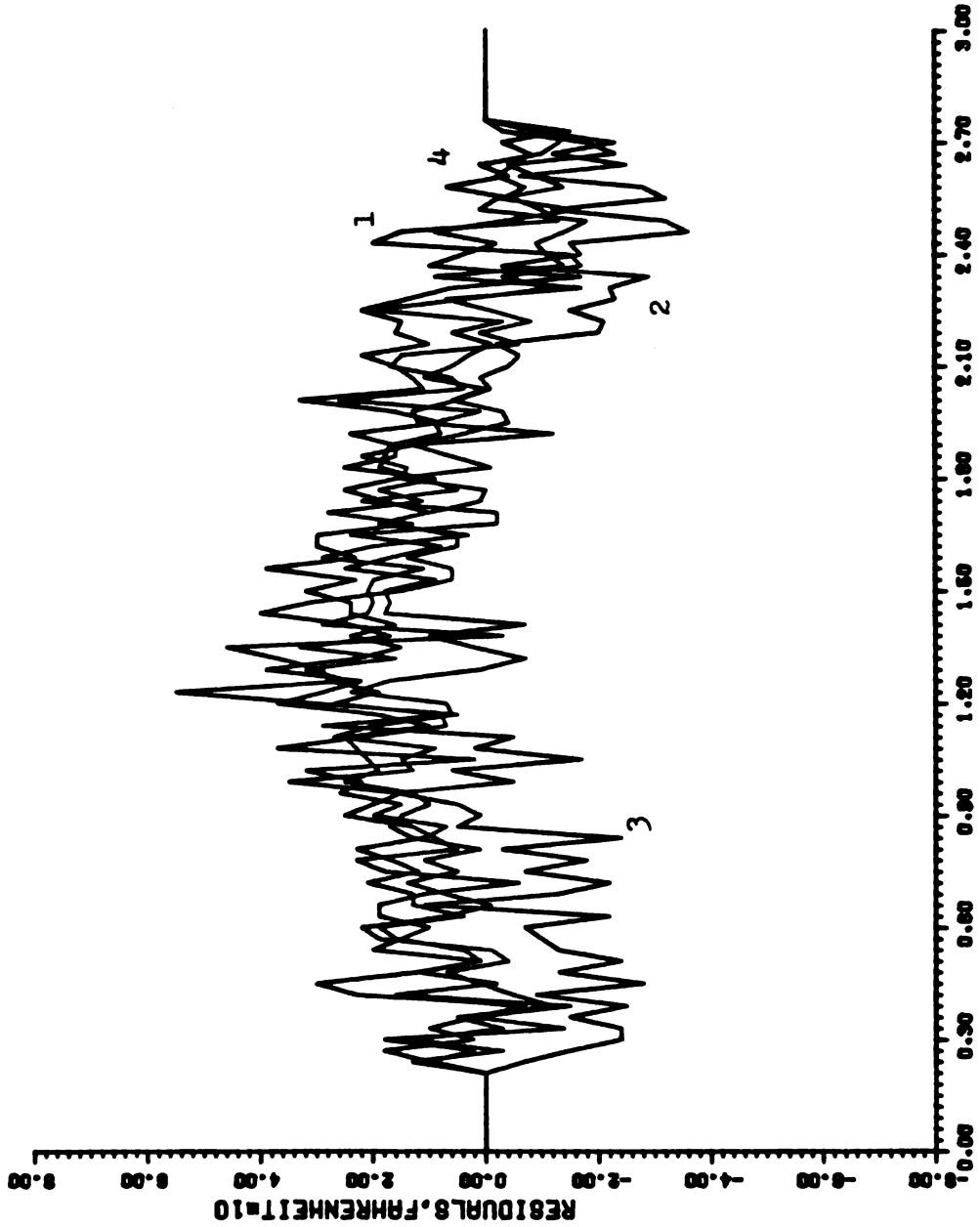

Figure 2.3. Insulated surface residuals: Case-2

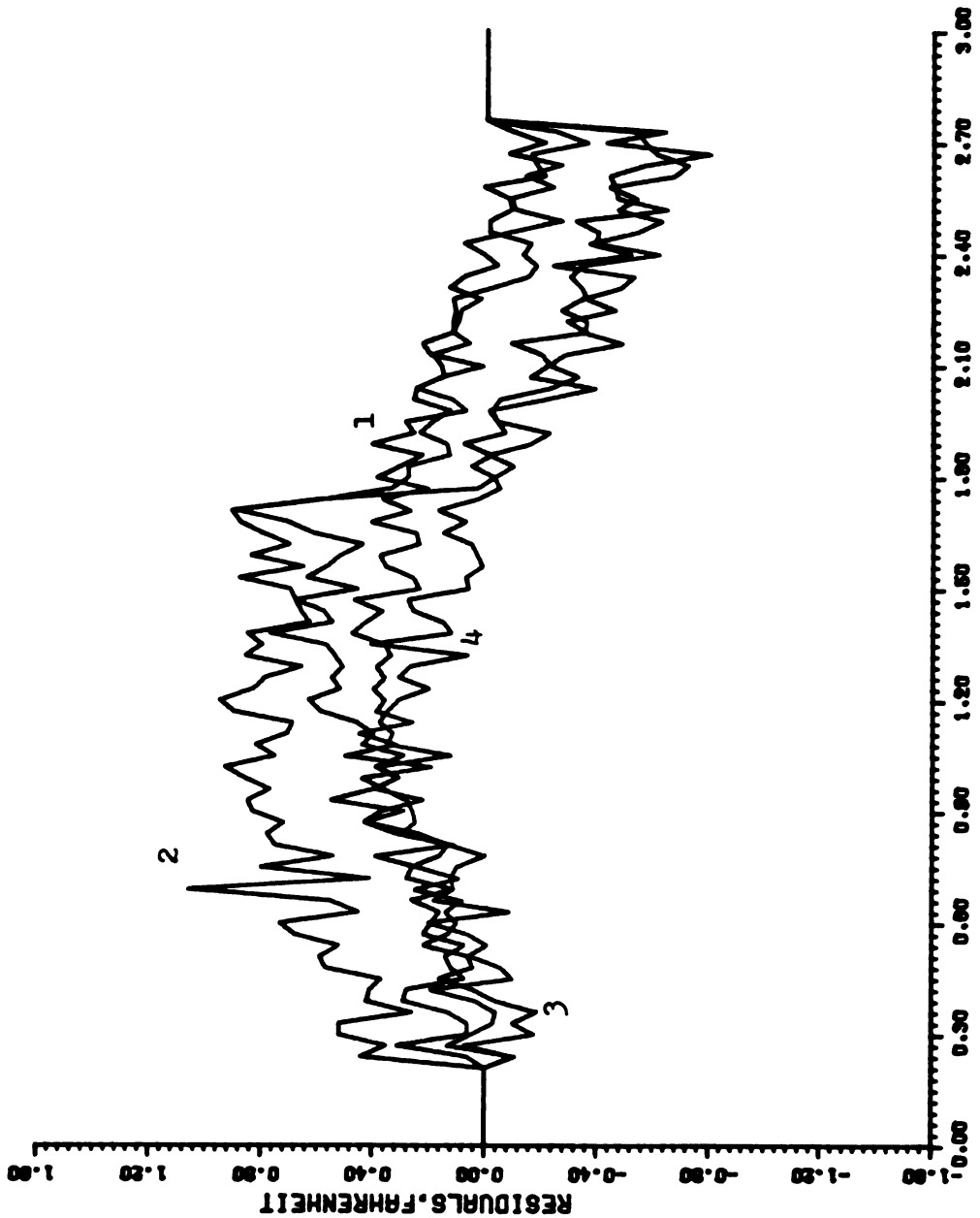

Figure 2.4. Insulated surface residuals: Case-3

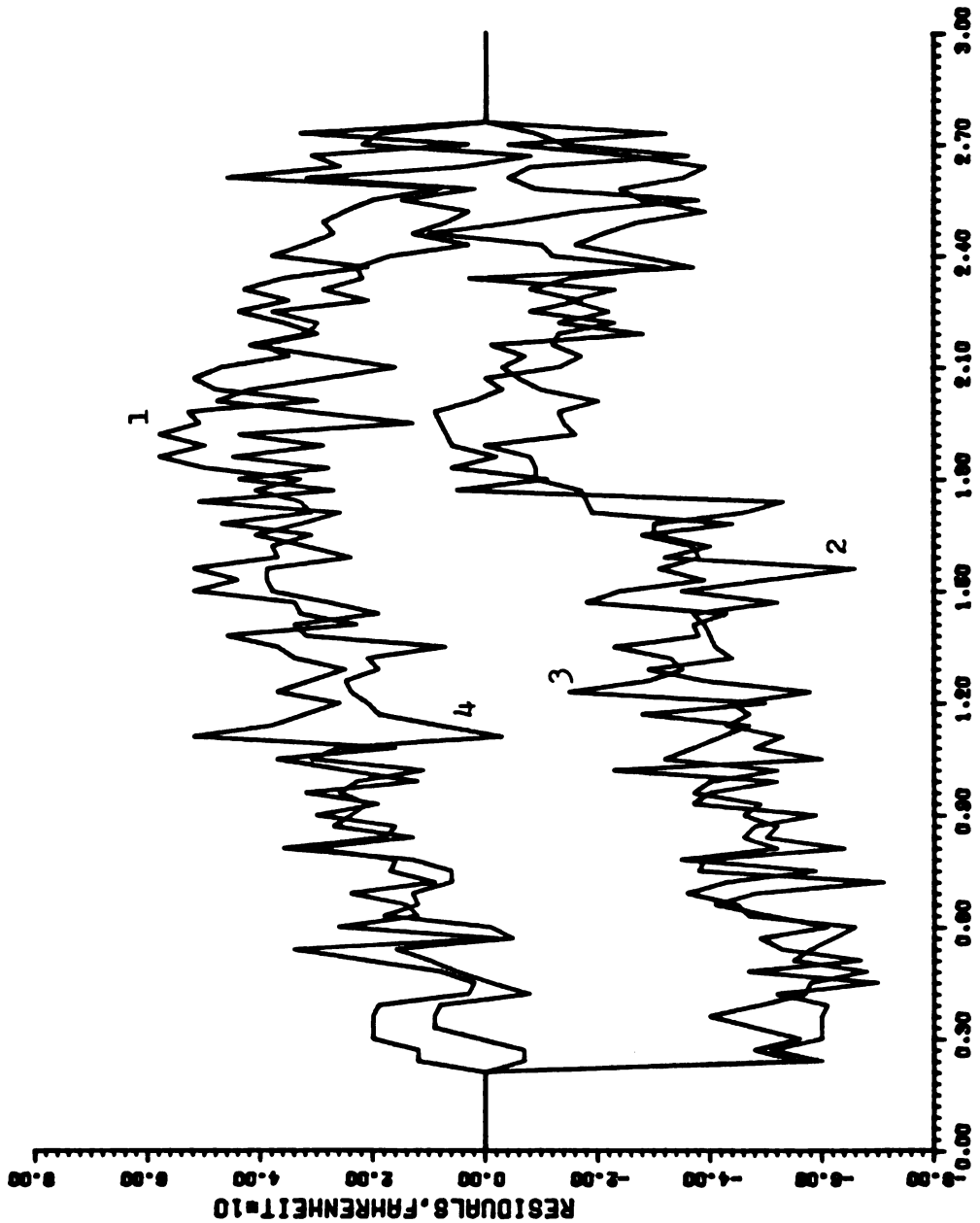

Figure 2.5. Insulated surface residuals: Case-4

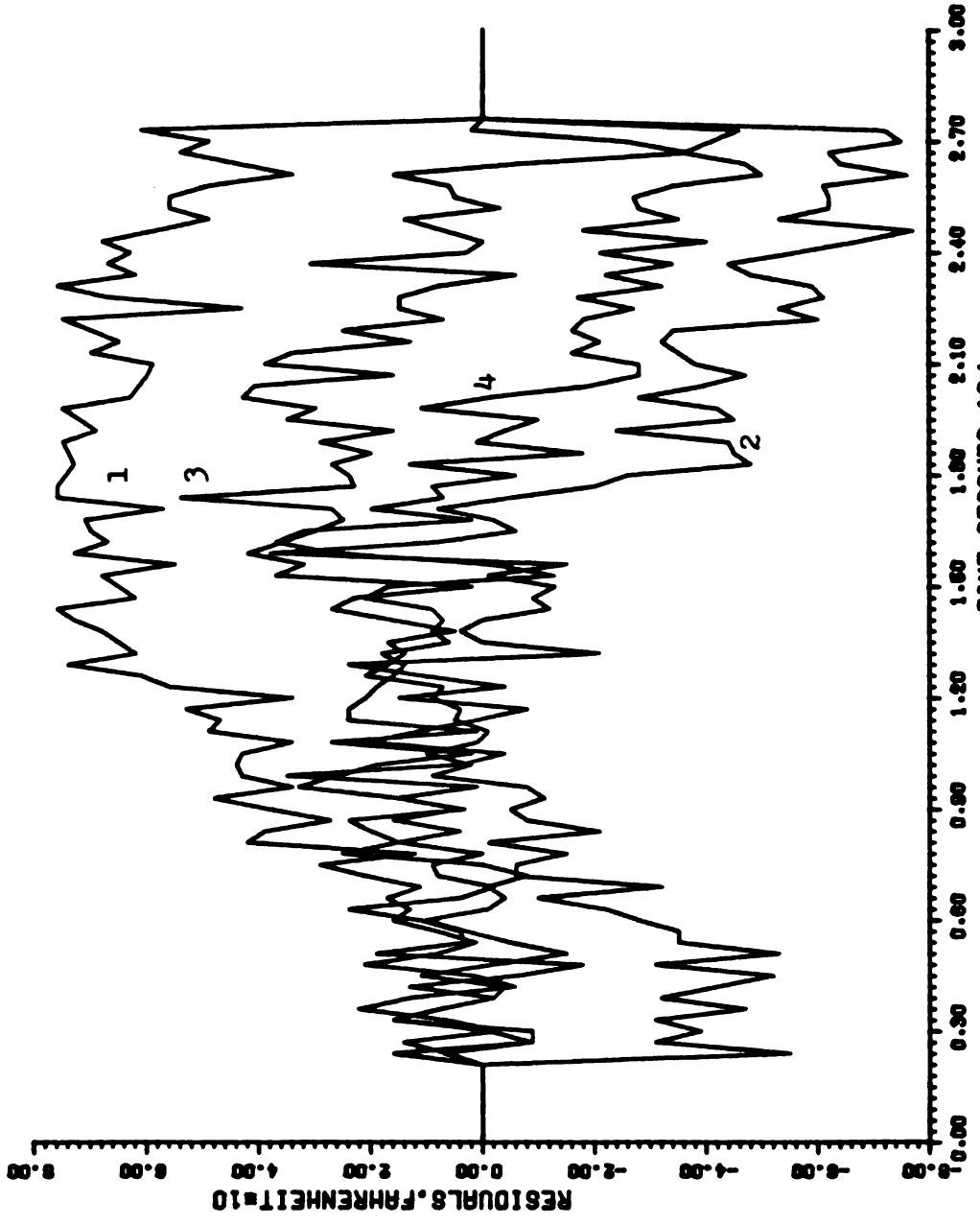

Figure 2.6. Insulated surface residuals: Case-5

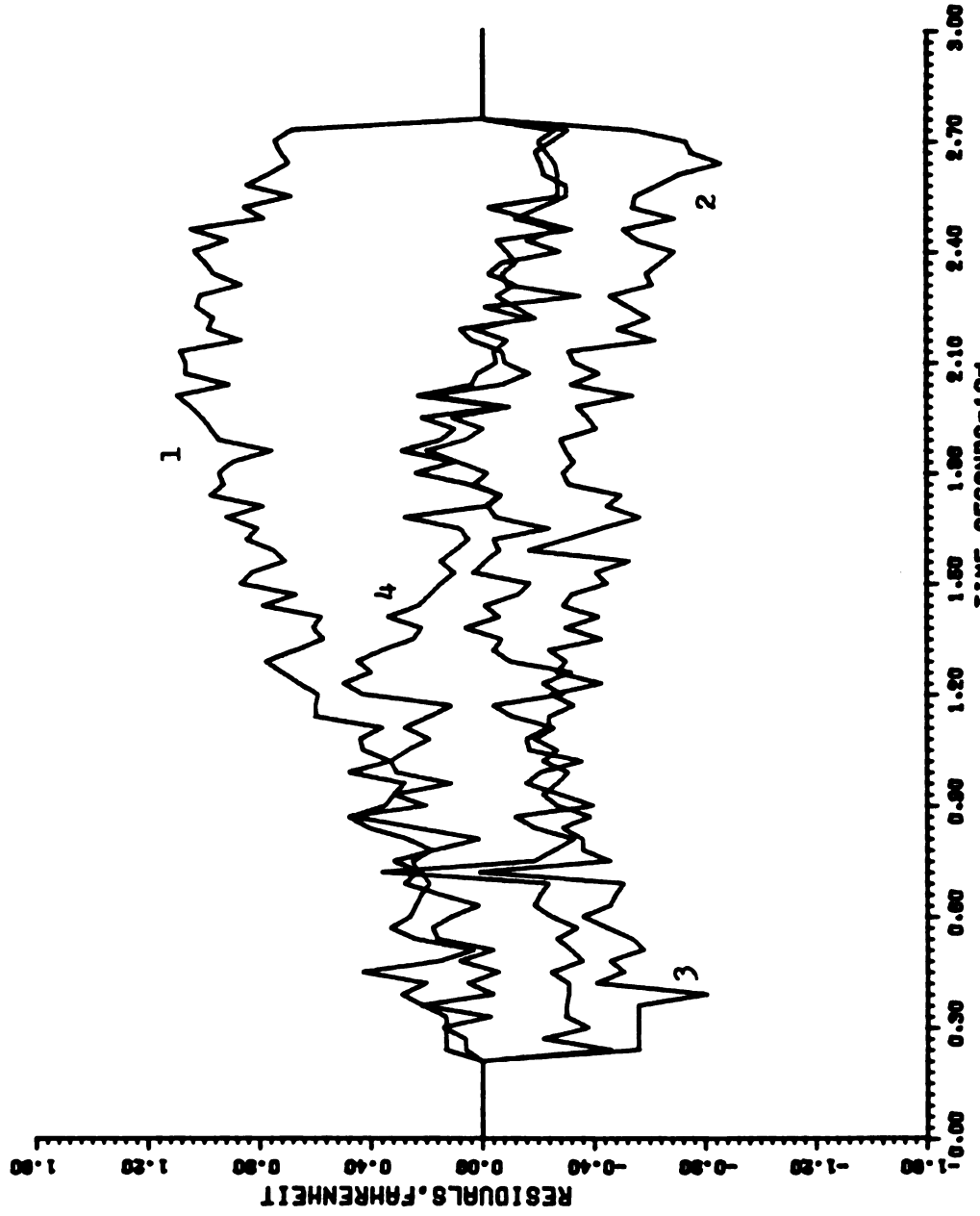

Figure 2.7. Insulated surface residuals: Case-6

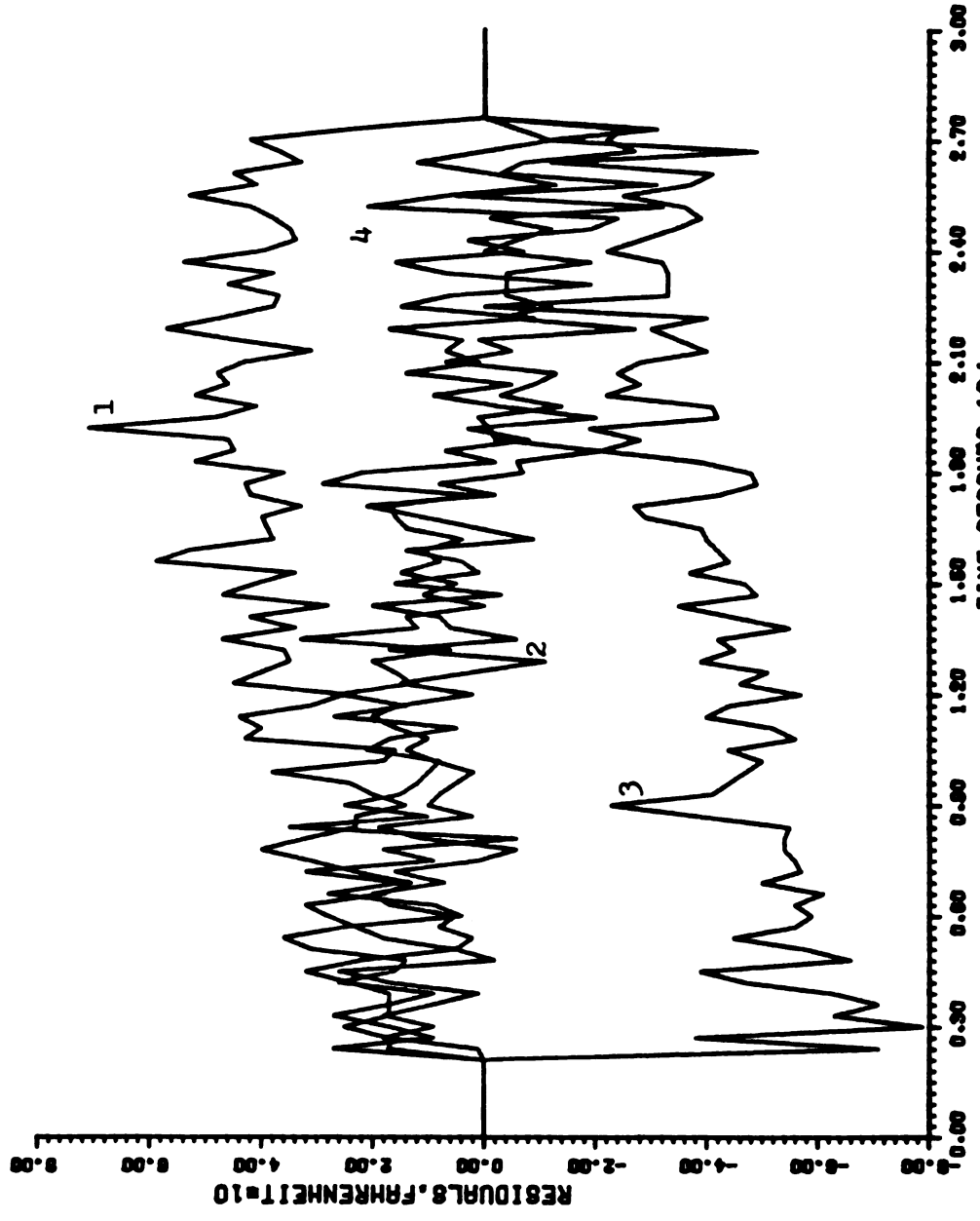

Figure 2.8. Insulated surface residuals: Case-7

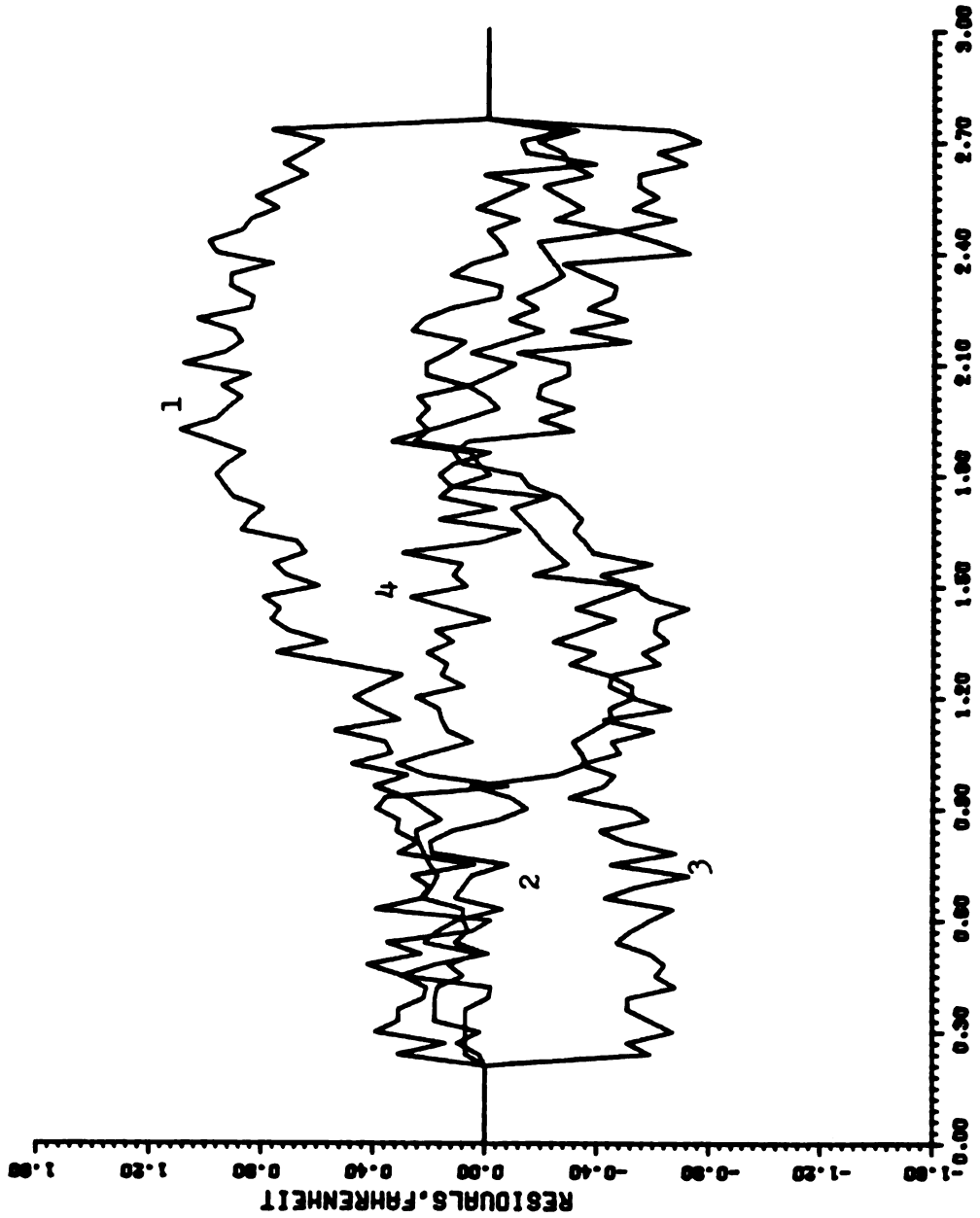

Figure 2.9. Insulated surface residuals: Case-8

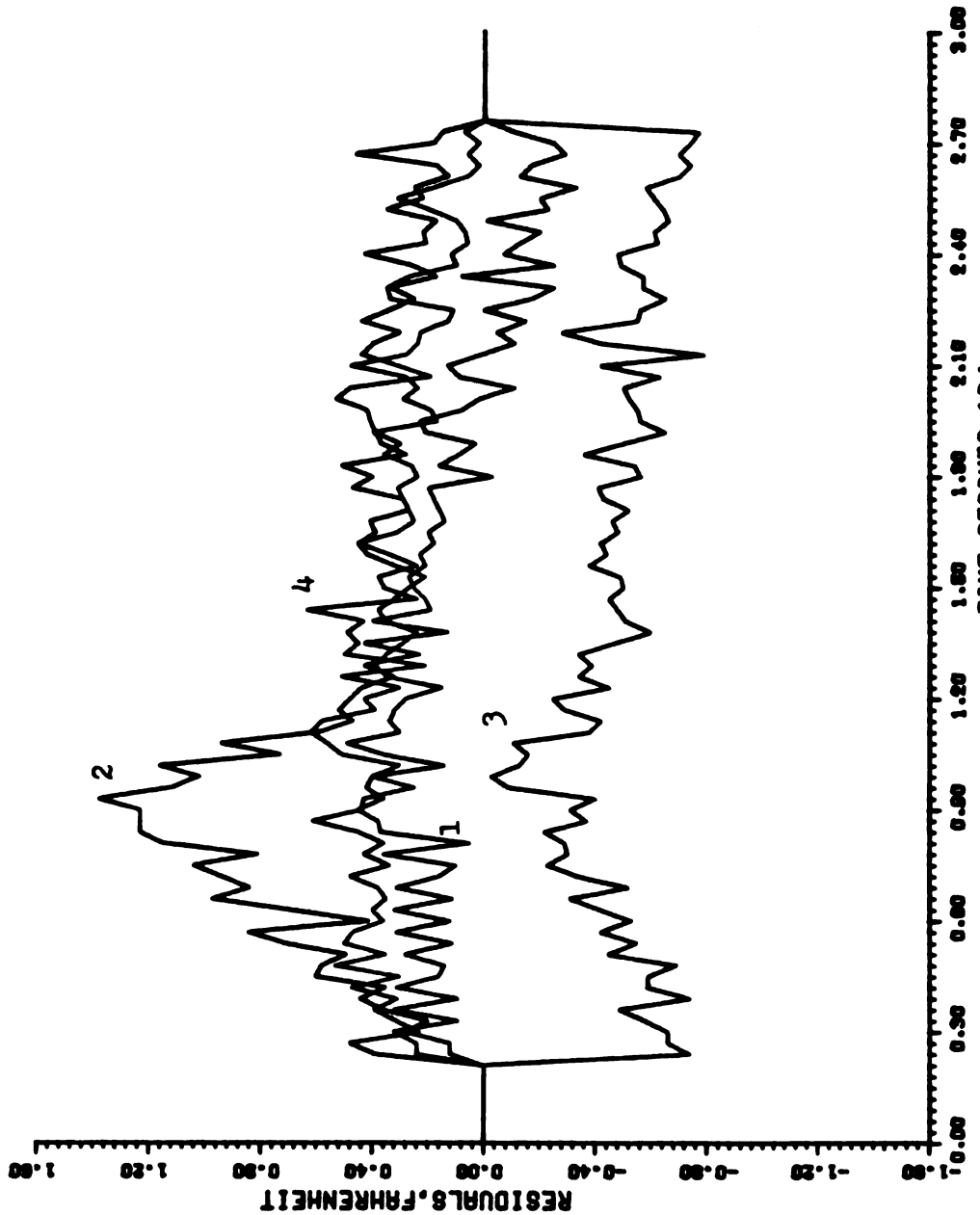

Figure 2.10. Insulated surface residuals: Case-9

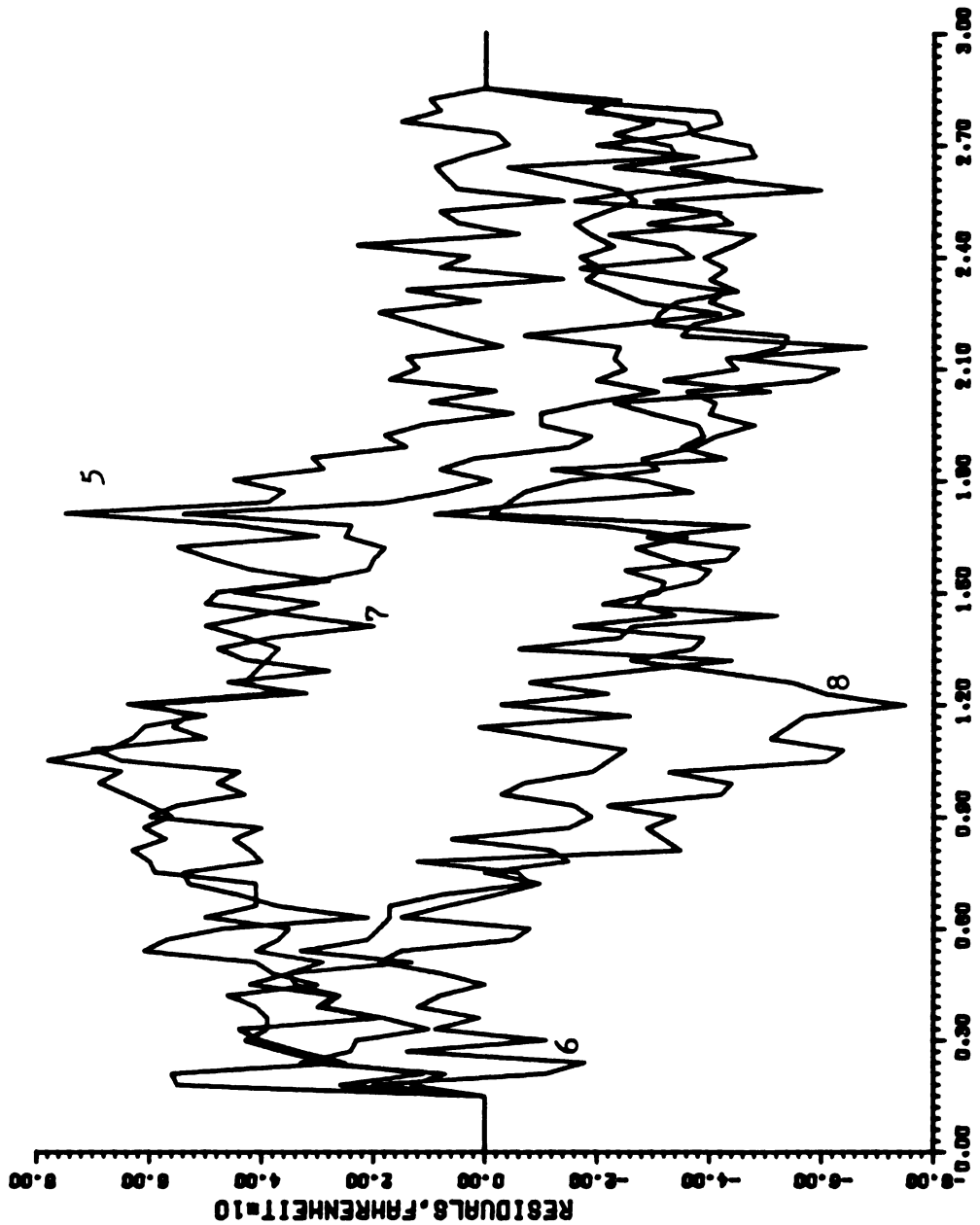

Figure 2.11. Heated surface residuals: Case-0

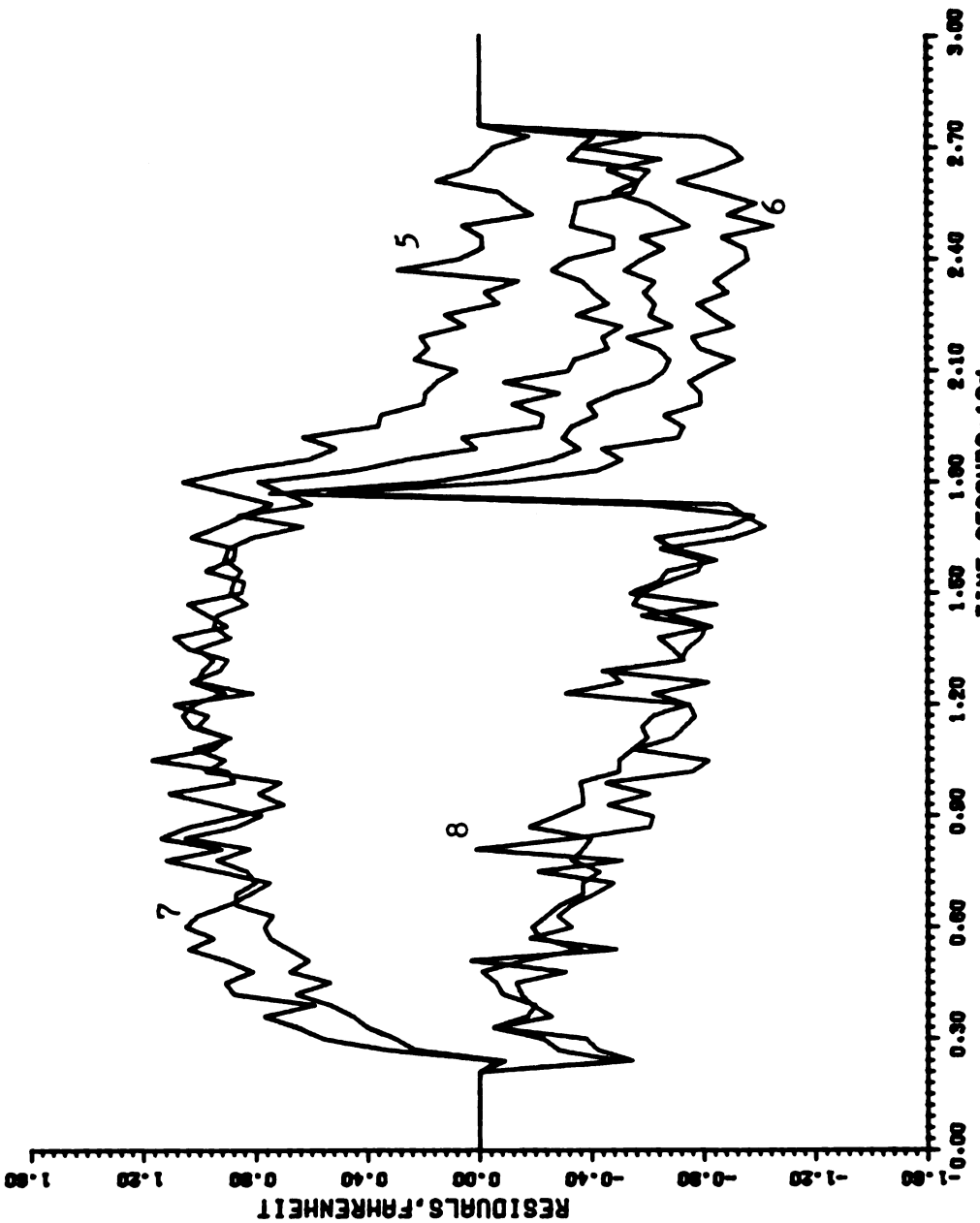

Figure 2.12. Heated surface residuals: Case-1

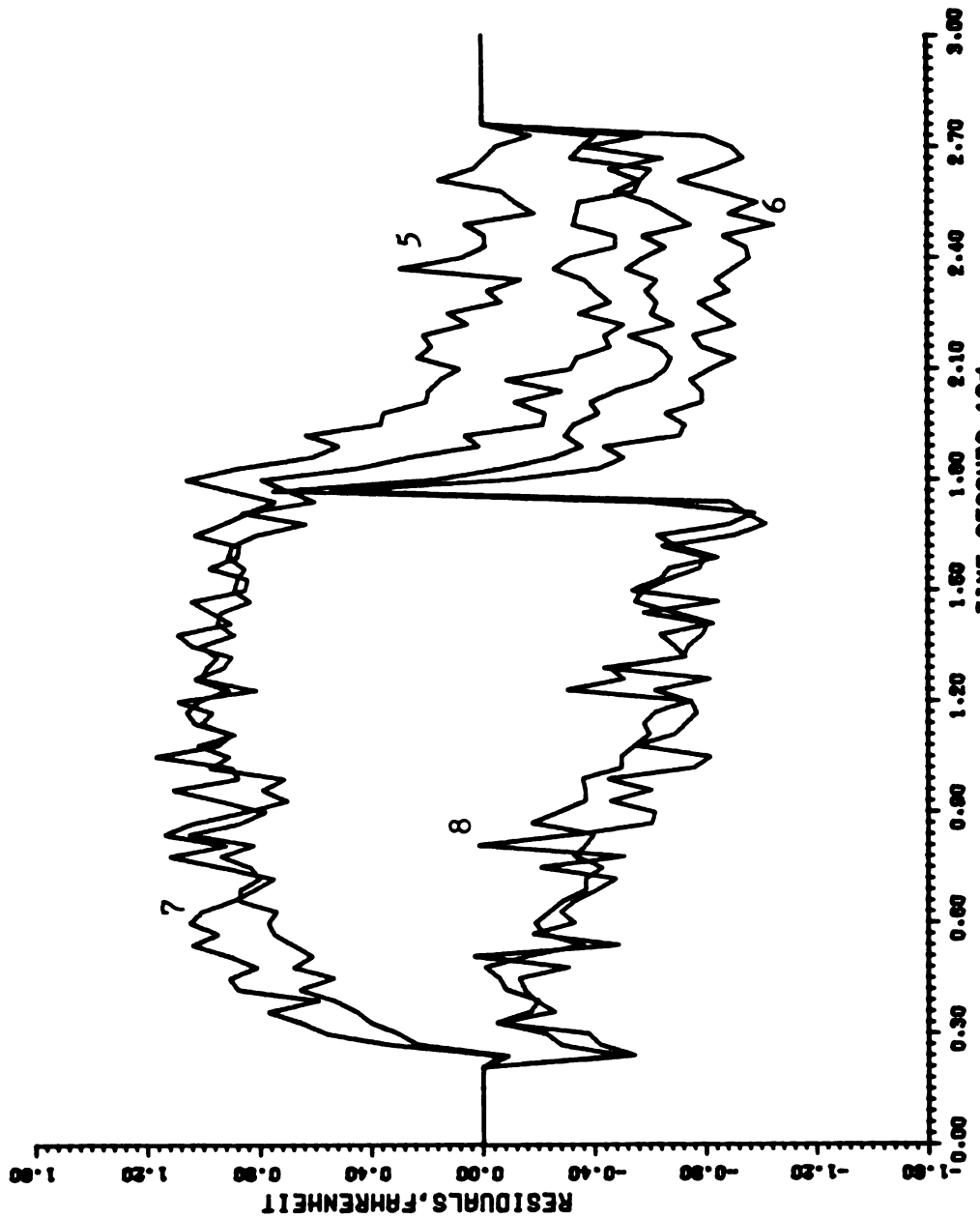

Figure 2.12. Heated surface residuals: Case-1

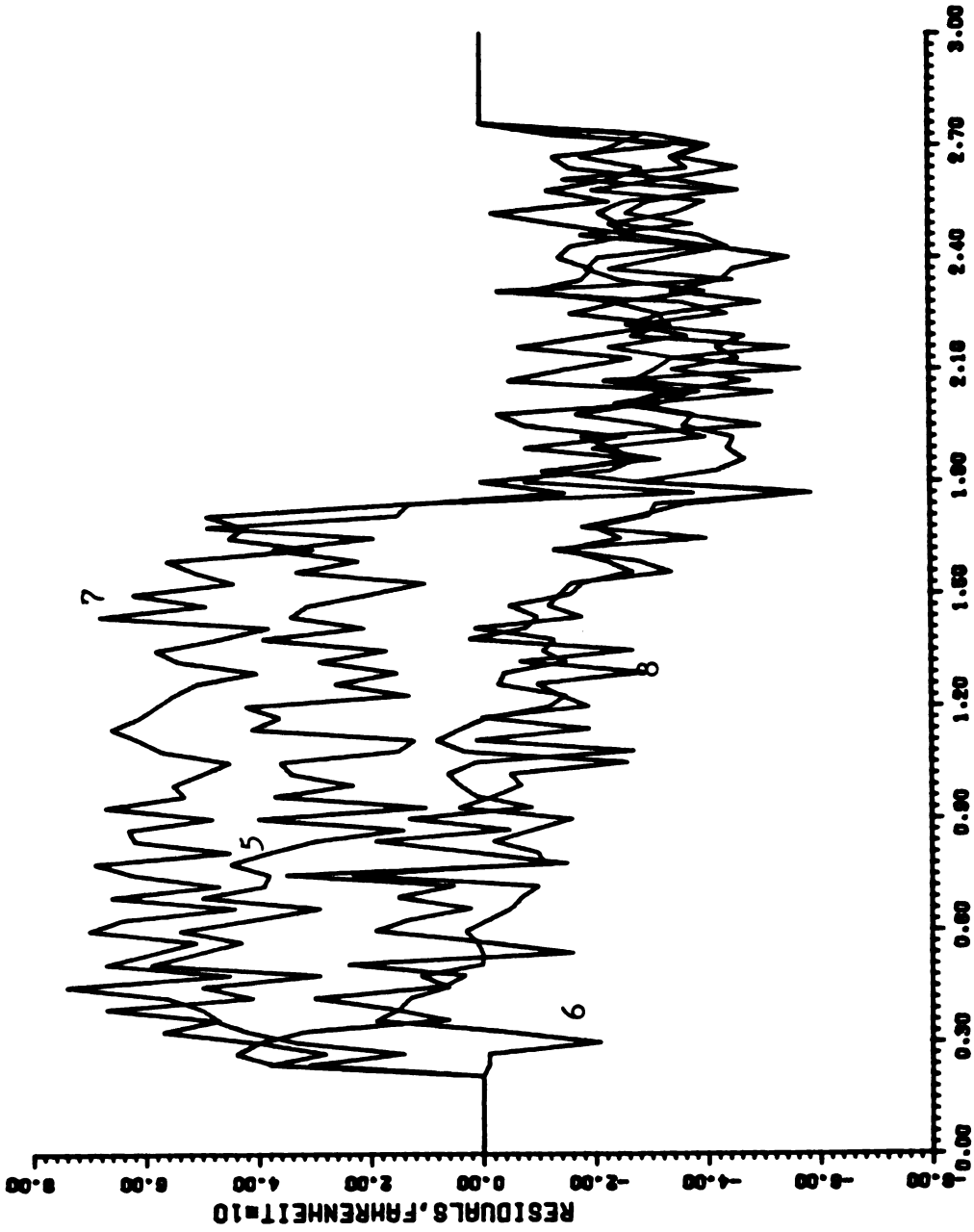

Figure 2.13. Heated surface residuals: Case-2

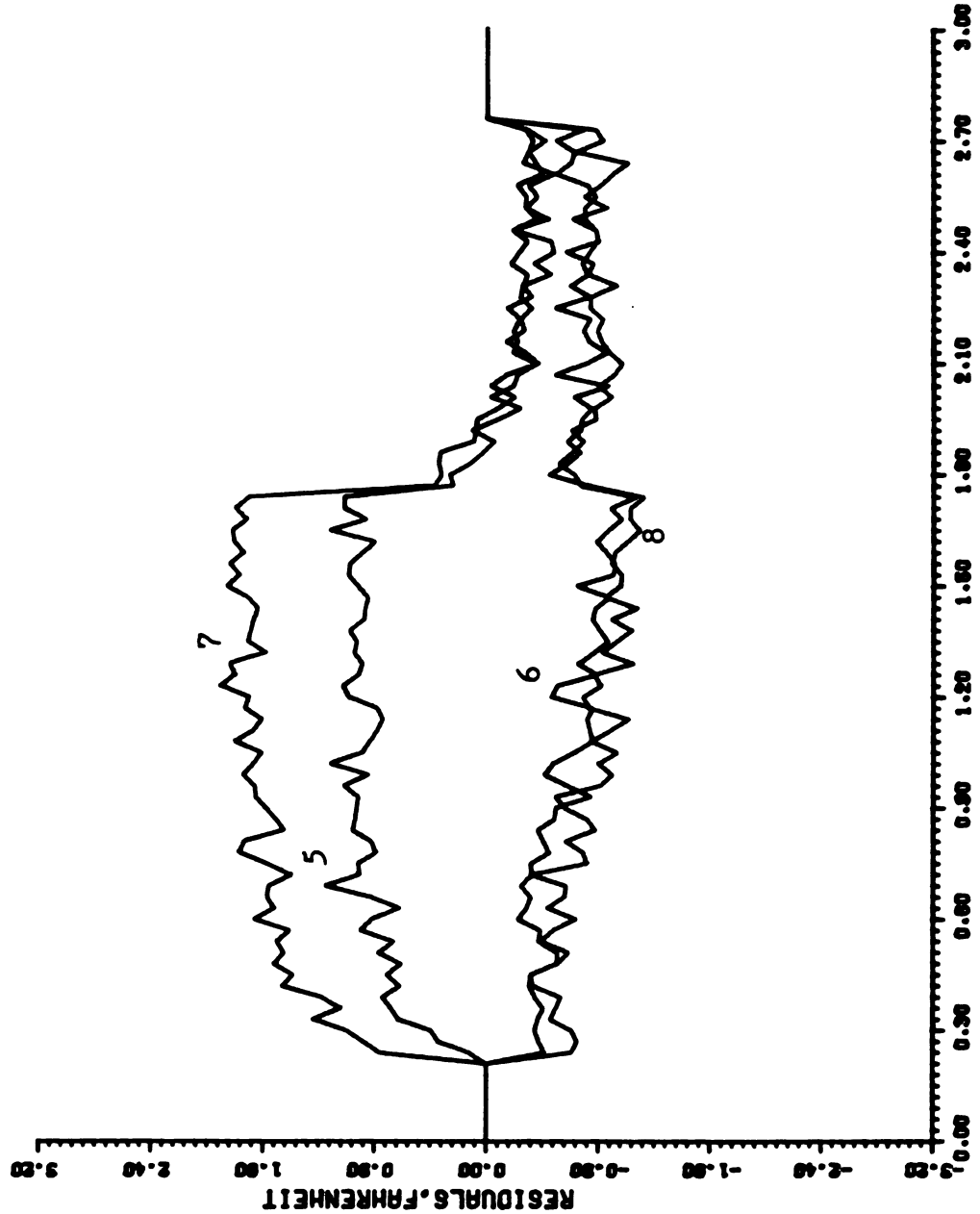

Figure 2.14. Heated surface residuals: Case-3

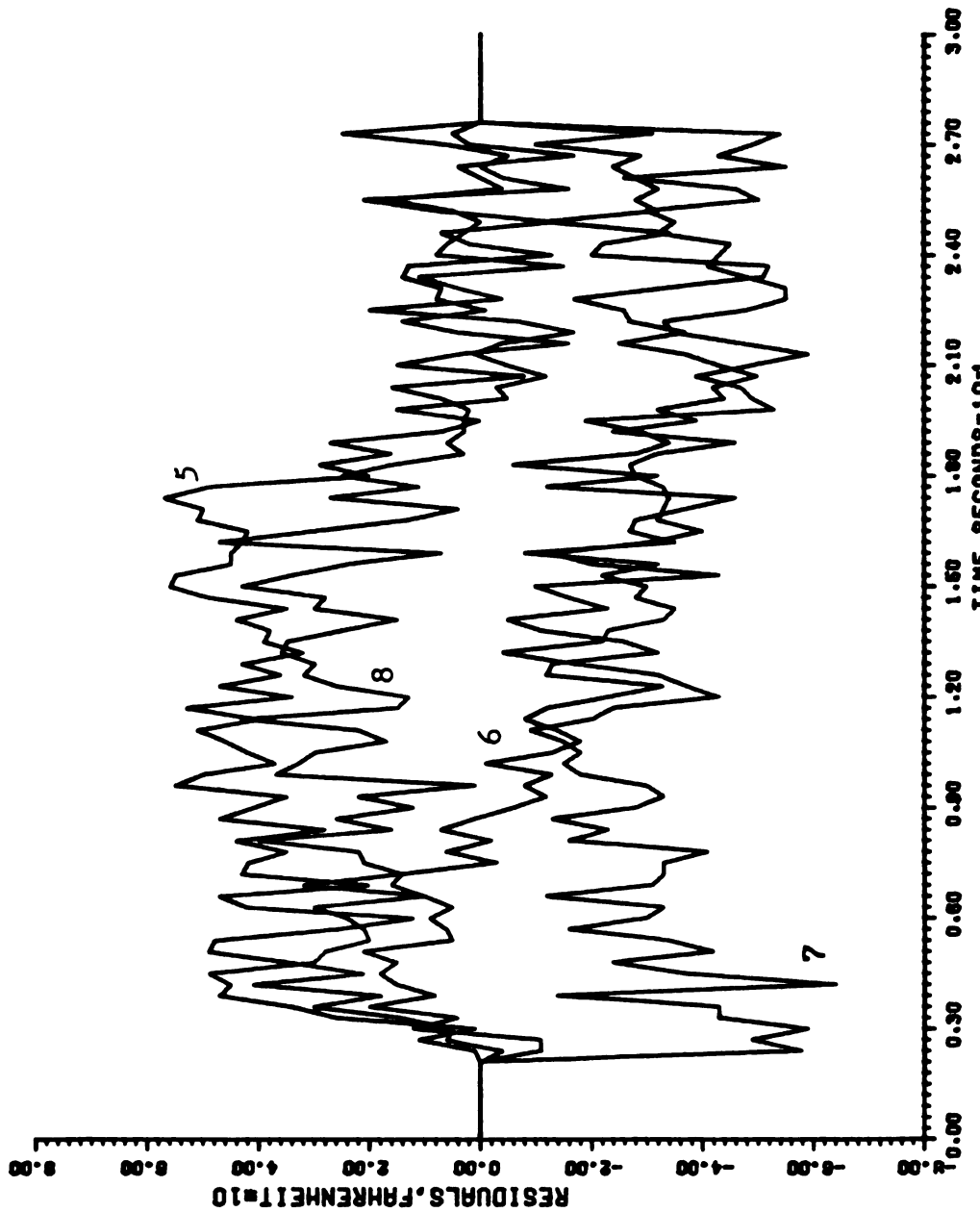

Figure 2.15. Heated surface residuals: Case-4

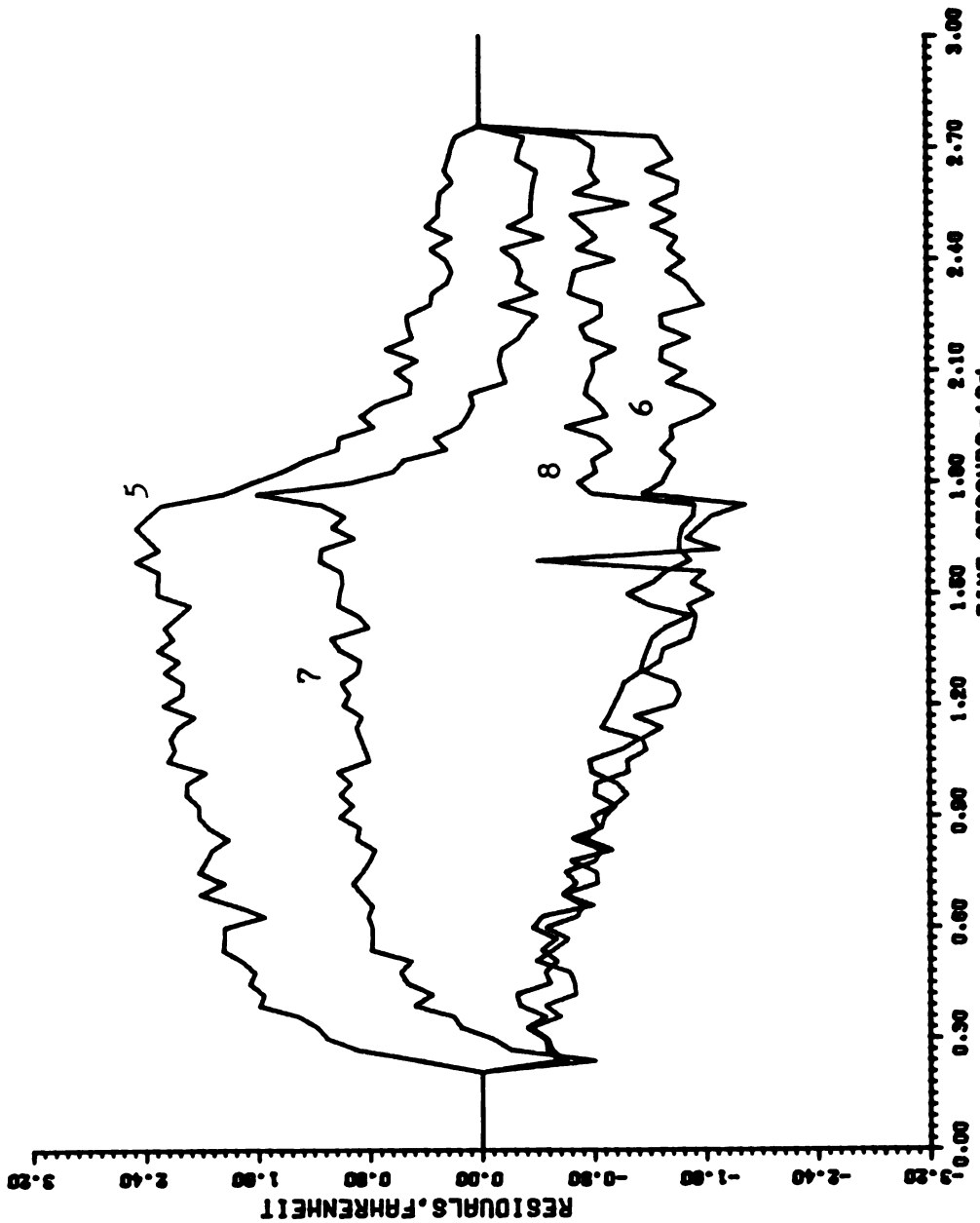

Figure 2.16. Heated surface residuals: Case-5

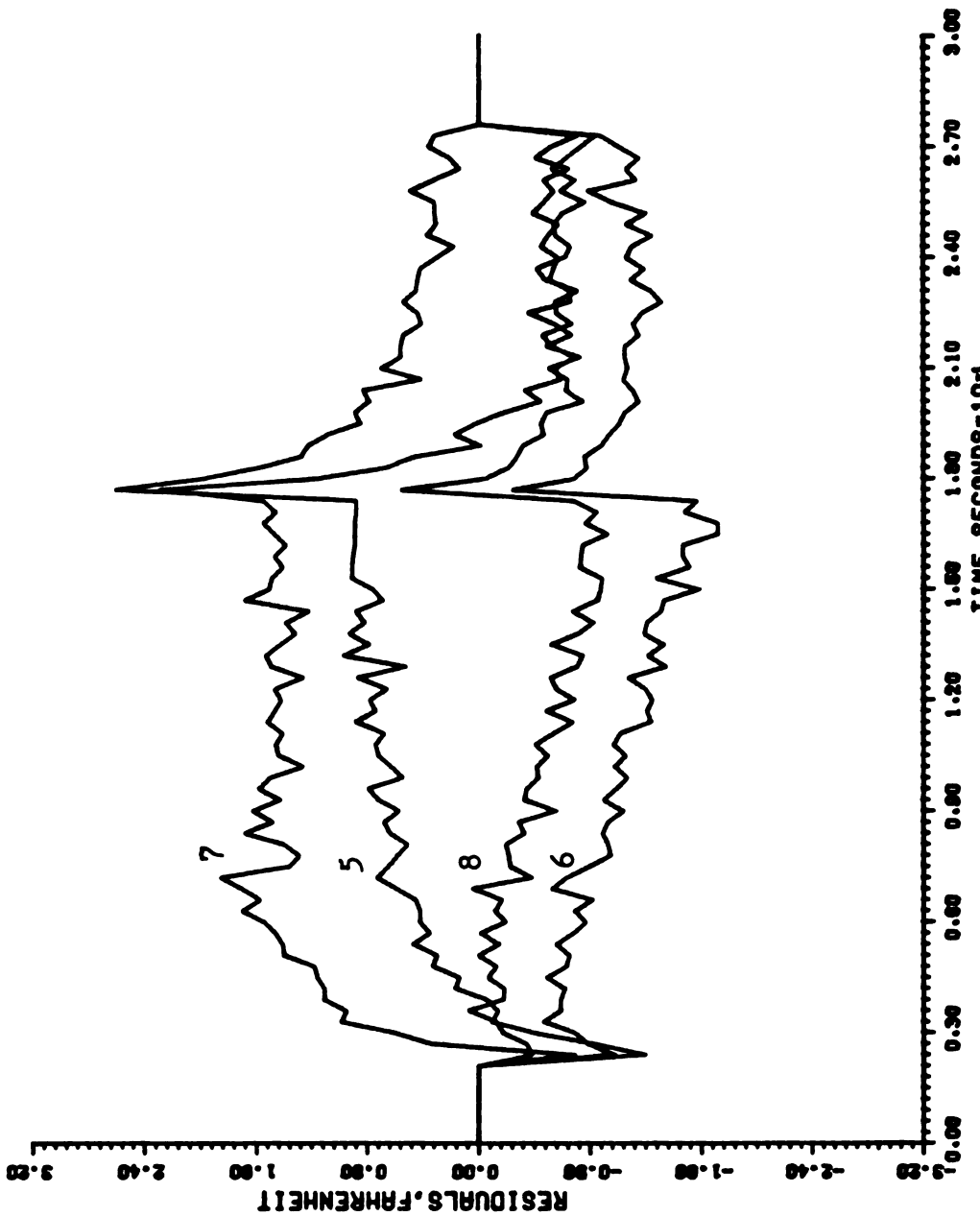

Figure 2.17. Heated surface residuals: Case-6

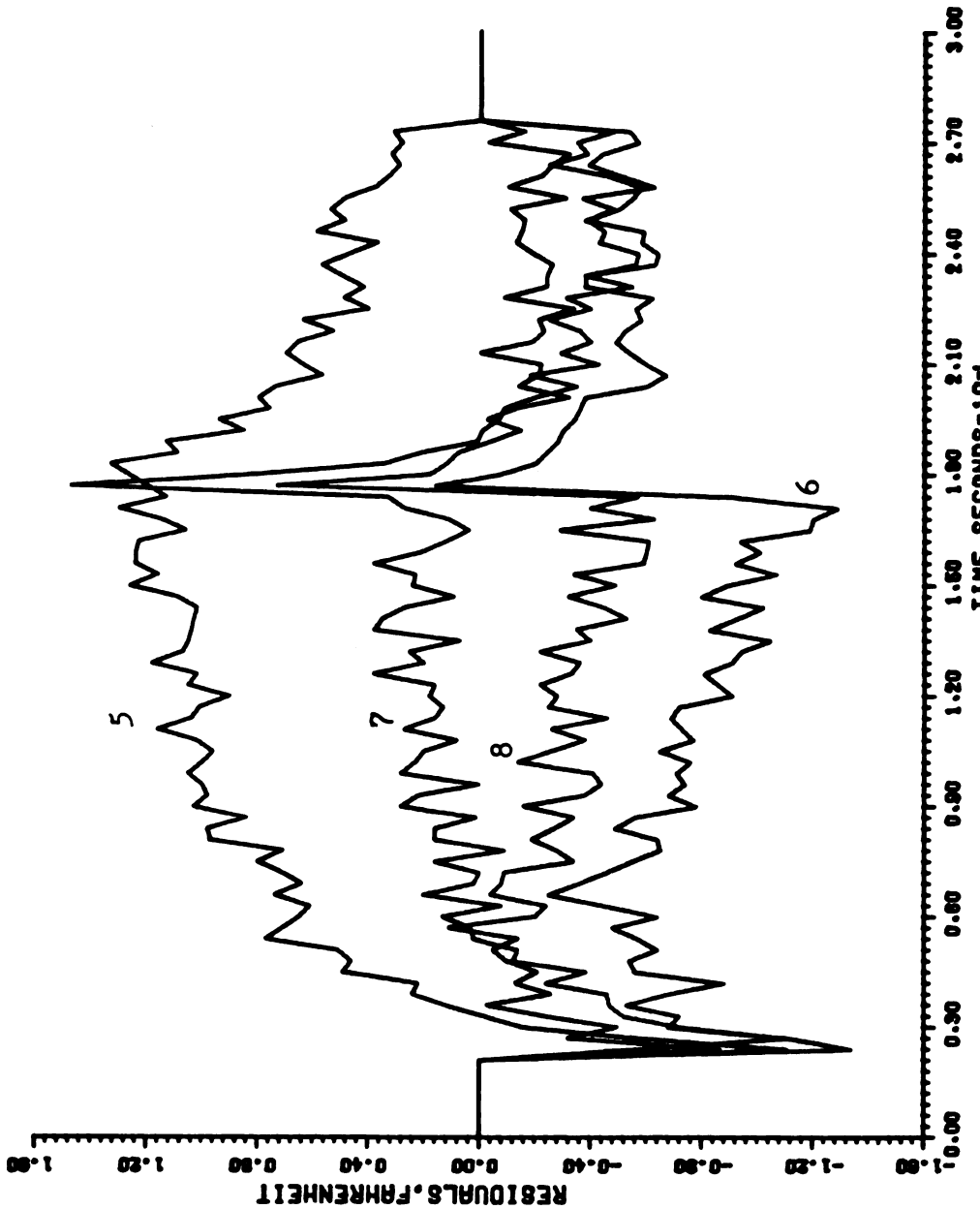

Figure 2.18. Heated surface residuals: Case-7

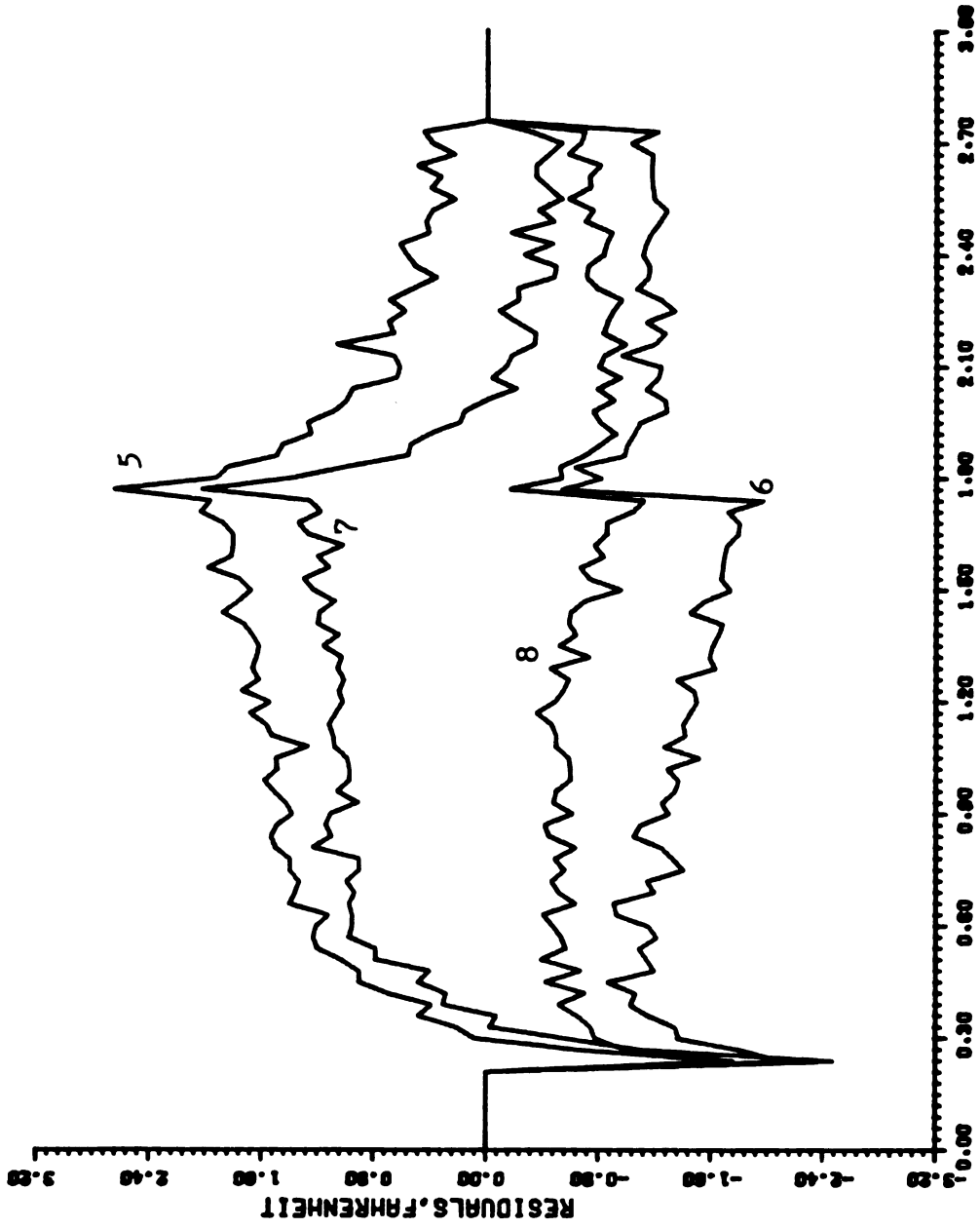

Figure 2.19. Heated surface residuals: Case-8

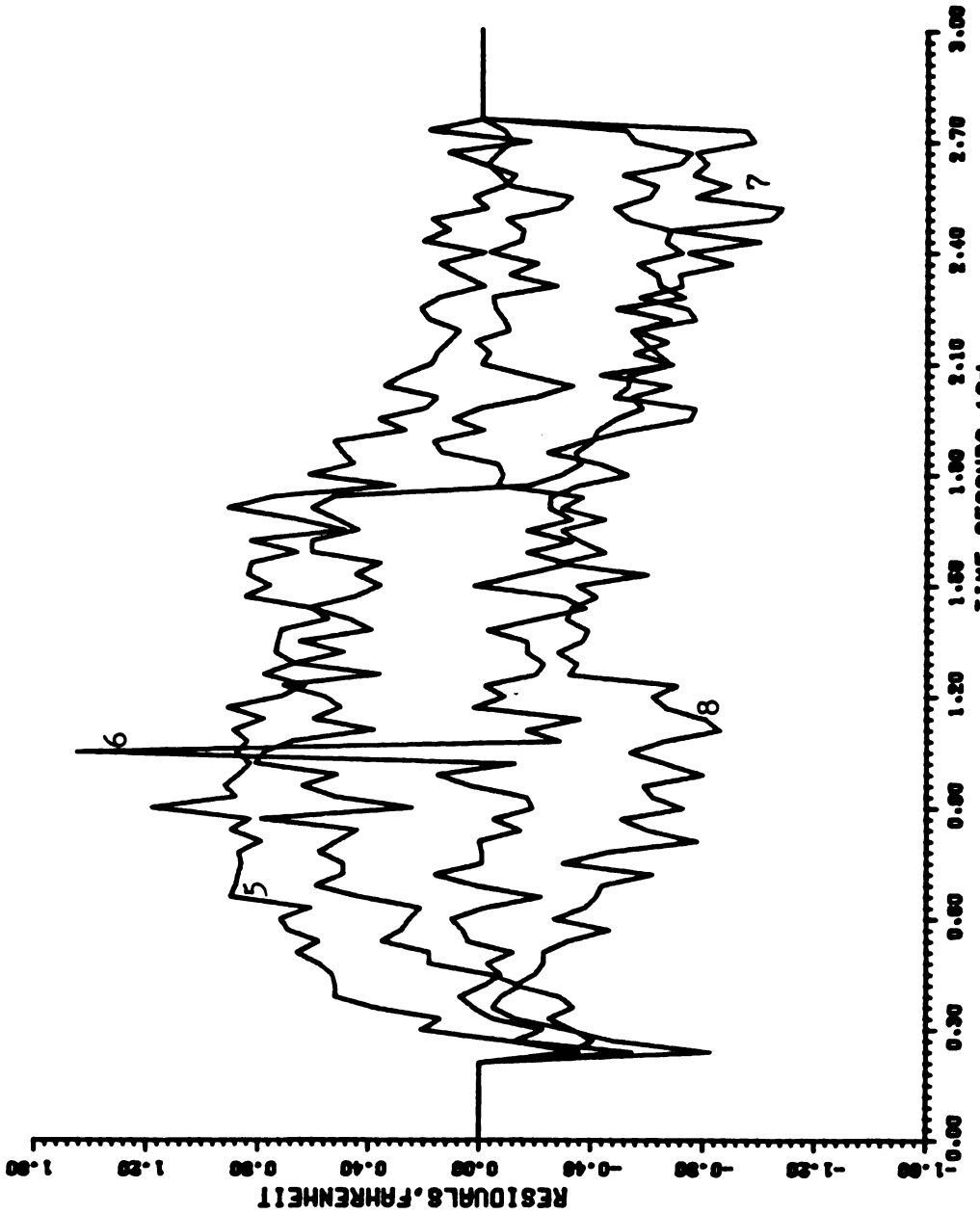

Figure 2.20. Heated surface residuals: Case-9

2.2.2 General Structure of the Error Covariance Matrix

The purpose of this section is to identify the general form of the matrix V*, the covariance matrix for the errors. In real problems the error covariance is unknown. One way of proceeding involves using the residuals produced using standard least squares. Draper and Smith (1966, p. 79) have stated concisely that "in practical problems it is often difficult to obtain specific information on the form of V* at first, and for this reason it is sometimes necessary to make the (known to be erroneous) assumption V* = I and then attempt to discover something about the form of V* by examining the residuals." The assumption V* = I implies that a standard least squares cost function, such as Equation (2.4), is used when the residuals are calculated.

The structure of the V* matrix is related to a P_m matrix and an A* matrix defined below. For the j-th sensor an additive error at time i is denoted $w_i^{(j)}$ and for each sensor can be written as an n component column vector $w^{(j)}$. The white noise consists of normal identically independently distributed errors denoted $a_i^{(j)}$ for the j-th sensor at the i-th time, and the white noise for each sensor can be written as an n component column vector $a^{(j)}$. A practical ARIMA error model for serial correlation relates the additive error vector $w^{(j)}$ to the white noise vector $a^{(j)}$ as

$$w^{(j)} = P^{(j)} a^{(j)} \tag{2.7}$$

where $P^{(j)}$ is a square matrix of dimension n.

Equation (2.7) is an essential equation used elsewhere in the dissertation. Equation (1.10) given previously assumes that $P^{(j)}$ can

be defined, although the matrix $P^{(j)}$ was denoted L. Equation (3.22) given later is a computational $P^{(j)}$ matrix for ARIMA errors, although the matrix $P^{(j)}$ will be denoted $L\ M^{-1}$.

The particular ARIMA(p,d,g) model appropriate for the heat conduction residuals determines the particular form of the $P^{(j)}$ matrix. When there is more than one sensor there can be contemporaneous correlations. For multisensor data with m sensors the convention used is to stack the $w^{(j)}$ vectors by sensor as

$$w^* = \begin{bmatrix} w^{(1)} \\ w^{(2)} \\ \vdots \\ w^{(m)} \end{bmatrix} \qquad \text{and} \qquad a^* = \begin{bmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(m)} \end{bmatrix}$$

The model for serially correlated multisensor data is

$$w^* = P_m\ a^* \tag{2.8}$$

where P_m is a square matrix of dimension mn with the matrices $P^{(j)}$ as its diagonal components

$$P_m = \begin{bmatrix} P^{(1)} & 0 & . & 0 \\ 0 & P^{(2)} & . & 0 \\ . & . & . & 0 \\ 0 & 0 & . & P^{(m)} \end{bmatrix}$$

The error covariance matrix is denoted W*, equals cov(w*), and satisfies

$$W^* = P_m\ A^*\ P_m^t \tag{2.9}$$

The contemporaneous correlation defined as the correlation between different sensors is written as

$$A^* = E(a^* a^{*t}) \tag{2.10.A}$$

The matrix A* is square and symmetric and has dimension mn.

The matrix A* has the general form

$$A^* = \begin{bmatrix} A_{11} & A_{12} & \cdot & A_{1m} \\ A_{21} & A_{22} & \cdot & A_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ A_{m1} & A_{m2} & \cdot & A_{mm} \end{bmatrix} \tag{2.10.B}$$

with each component matrix A_{ij} being a square matrix with dimension n. Each component matrix A_{ij} is diagonal because by definition white noise is not serially correlated. It is convenient to assume that the contemporaneous correlation has homoscedasticity defined as constant variance for A_{ii}. This is discussed further in Section 4.5. Thus, in a given A_{ij} the diagonal entries are all assumed to equal the same constant.

In order to be practical, the combined model for serial and contemporaneous correlation must only involve a few parameters. Assuming that each sensor obeys the same order ARIMA(p,d,q) process then there are a total of $(p + q)\, m + m + \frac{1}{2}(m^2 - m)$ coefficients from the ARIMA process in the matrix W*. This can be bounded by two considerations. First, Box and Jenkins (1970) suggested that usually the sum p + q is two or less. Second, Farnia's data only have eight sensors, m = 8, although sixteen thermocouples were used. Thus there can be as many as 52 different parameters in the matrix W*.

2.3 Observed Structure of the Covariance Matrix

A reasonable approximation to the matrix W* was investigated as follows. Whether the serial correlation coefficients are zero is tested in Section 2.3.1 and whether they are unity is tested in Section 2.3.2. Contemporaneous correlations are represented and estimated for the pre-heat-flux data in Section 2.3.3. It is concluded that the matrix W* is dominated by serial correlation and that instead of the 52 different coefficients suggested in Section 2.2.2 only 3 different coefficients are needed.

In Section 2.3.4 a calibration problem related to the randomness of the white noise is discussed.

2.3.1 Durbin-Watson and Schmidt Tests

The Durbin-Watson test statistic for the presence of serial correlation discussed by Kmenta (1971) is

$$d = 2\ (1 - \emptyset_1)$$

where $\emptyset_1$ is the coefficient in the ARIMA(1,0,0) model computed from the Yule-Walker equations given as Equation (A.7). In a test of the hypothesis of independent errors versus the alternative of positive first order autocorrelation (with n = 90 and the 0.01 level) the rule is

(1) reject if $d < d_L = 1.47$,

(2) do not reject if $d > d_U = 1.56$,

(3) do neither when $d_L < d < d_U$.

From Table 2.4 the average value of $\emptyset_1$ was estimated to be 0.772. Thus the statistic is

$$d = 2\ (1 - 0.772) = 0.456.$$

Since this value is less than 1.47 the hypothesis of independent errors is rejected in favor of the hypothesis of first order autocorrelated errors. The hypothesis is also rejected for each of the individual sensors since the largest value of d is 0.82.

A Durbin-Watson type procedure for second order autocorrelation--i.e., ARIMA(2,0,0)--was developed by Schmidt (1972). The test uses two statistics; one is

$$d_1 = 2(1 - \phi_1)$$

where ϕ_1 is the Yule-Walker estimator for the coefficient in the ARIMA(1,0,0) process, and the other statistic is

$$d_2 = 2(1 - \phi_2)$$

where ϕ_2 is the Yule-Walker estimator for the second coefficient in the ARIMA(2,0,0) process. The test statistic used by Schmidt (1972) implicitly assumes the coefficients have equal order of magnitudes and it is

$$\delta = d_1 + d_2$$

In a test of the hypothesis of independent errors versus the alternative of second order autocorrelation (with n = 90 and the 0.01 level of significance) the rules are

(1) reject if $\delta < d_L = 3.18$,

(2) do not reject if $\delta > d_U = 3.35$,

(3) do neither when $d_L < \delta < d_U$.

From Table 2.4 the estimates of ϕ_1 and ϕ_2 are 0.772 and 0.306, respectively. Hence the test statistic is

$\delta = 2(1 - 0.772) + 2(1 - 0.306) = 1.844.$

Thus, the hypothesis of independent errors is rejected and the alternate hypothesis of second order autocorrelation is accepted. The hypothesis of independent errors is also rejected for each sensor because the largest value of δ is 2.192.

Table 2.4

Yule-Walker Estimates for Case-0 Heat Conduction Residuals

ARIMA(1,0,0) Process		ARIMA(2,0,0) Process			Sensor
$\emptyset_1$	σ^2	$\emptyset_1$	$\emptyset_2$	σ^2	
0.613	0.0098	0.435	0.291	0.0089	1
0.834	0.0161	0.498	0.403	0.0135	2
0.918	0.0318	0.681	0.258	0.0297	3
0.696	0.0210	0.446	0.359	0.0183	4
0.782	0.0168	0.531	0.321	0.0151	5
0.919	0.0157	0.774	0.157	0.0153	6
0.590	0.0180	0.324	0.451	0.0144	7
0.831	0.0282	0.659	0.208	0.0270	8
Average		Average			
0.773		0.544	0.306		

2.3.2 Cumulative Error Test

Cumulative errors are the special case of first order autoregressive errors with $\emptyset$ having the value of unity. When $\emptyset$ is close to unity the ARIMA(1,0,0) errors may be approximated as being cumulative errors.

Some authors advocate describing errors by cumulative error models from considerations of the nature of the physical process and/or the measuring device. Mandel (1957) postulated that an extent of reaction for surcose measured on a fixed specimen has errors that are cumulative.

Heuvel et al. (1976) postulated that the integral number of counts in an x-ray diffraction experiment are cumulative and used the procedure proposed by Mandel (1957) and improved by Beck (1974).

Mandel (1964) suggested there are two types of errors: cumulative errors associated with "process" errors and other "analytical" errors. An examination of the least squares residuals could be used to decide whether the process errors dominate so that a cumulative error model is appropriate. However, the analysis by Anderson (1975, p. 142) showed that when white noise analytical errors and cumulative process errors are both present from more than one source the sum of these errors is an ARIMA(p,d,q) process. Thus, unless process errors are the only errors it is unlikely that the observed errors will be simply cumulative. This is discussed further in Section 2.5.

A simple screening method to determine if cumulative errors dominate is to use a hard limit estimator of the coefficient $\emptyset$ in an ARIMA(1,0,0) process. A hard limited process Z has two states that are assigned numerical values 0 and 1. These are related to a continuous valued w_i-series by the rule:

(1) if w_i is non-negative then $Z_i = 1$,

(2) if w_i is negative then $Z_i = 0$.

For the ARIMA(1,0,0) process in Equation (1.17) Kedem (1976) derived the hard limit estimator of $\emptyset$ as

$$\emptyset = \sin(\pi(c - \tfrac{1}{2})) \tag{2.11}$$

where

$$c = \frac{2 \sum_{i=2}^{n} Z_i Z_{i-1} - 2 \sum_{i=1}^{n} Z_i + Z_1 + Z_n + (n-1)}{n - 1}$$

Let an excursion of the w-series above the axis be called a hump. For example, there are two humps in the series (01110110) and for this series the first two terms in the numerator of the estimator of c equals four. In general, the first two terms in the numerator is twice the number of humps; thus, the parameter c can be estimated by inspection of the plotted w-series. This evaluation procedure was applied to the heat conduction residuals in Figures 2.1 and 2.11 from Beck and Arnold (1977, p. 408). Because the sample size in these plots is relatively large (n = 85), the hard limit estimator is in close agreement with the Yule-Walker estimates from a numerical version of the same data. The estimate of $\emptyset$ is about 0.86 for both estimators in Table 2.5. For this value of $\emptyset$ the residuals can be said to be neither independent nor cumulative. The estimates of $\emptyset$ in Table 2.5 are somewhat larger than those in Table 2.4 because the mean was assumed to be zero for the estimates in Table 2.5 while in Table 2.4 it was estimated.

2.3.3 Contemporaneous Correlation Test

The general structure of the contemporaneous correlation matrix A* is defined in Equation (2.10). In order to simplify the identification process several assumptions are made. First, in order to obtain a simple model only the pre-heat-flux residuals are modelled. Thus, the thermal model is the constant mean temperature which is

Table 2.5

Hard Limit Estimates for Case-0 Heat Conduction Residuals

Sensor	ø Yule-Walker	Number of Humps	ø Hard-Limit
1	0.8602	4	0.9488
2	0.9291	6	0.9096
3	0.8033	11	0.7071
4	0.6748	12	0.6548
5	0.9421	8	0.8005
6	0.8987	7	0.8600
7	0.9425	2	0.9841
8	0.9230	3	0.9689
	Average 0.8717		Average 0.8543

described by Equation (1.15) with α equal to zero. Second, because the sensors are identical, a common set of coefficients is used in the contemporaneous correlation model. These common coefficients are in a matrix denoted R that permits the contemporaneous correlation matrix to be written as the following kronecker matrix product

$$A^* = \sigma^2 R_m \# I_n. \tag{2.12}$$

In Equation (2.12) the subscript on the matrix R indicates there are m sensors, the subscript on the identity matrix I indicates there are n measurements per sensor, and the special symbol # is used to denote the kronecker matrix product whose properties are discussed by Theil (1971).

Two forms for the R_m matrix will be considered. First, the matrix is assumed to depend on five coefficients with the matrix written as

$$R_8 = \begin{bmatrix} a & & & & & & & \\ c & a & & & \text{symmetric} & & & \\ c & c & a & & & & & \\ c & c & c & a & & & & \\ d & d & d & d & b & & & \\ d & d & d & d & e & b & & \\ d & d & d & d & e & e & b & \\ d & d & d & d & e & e & e & b \end{bmatrix} \qquad (2.13.A)$$

For each experimental case, the R_8 matrix represents the correlation between the residuals of different sensors. The entries in the R_8 matrix are ordered so that the first four are for thermocouples at one face and the last four are at the other surface. Thus, the coefficients have the following expected values

$$a = E(w_i^{(j)} w_i^{(j)}) \quad \text{for } j = 1,2,3,4$$

$$b = E(w_i^{(j)} w_i^{(j)}) \quad \text{for } j = 5,6,7,8$$

$$c = E(w_i^{(j)} w_i^{(k)}) \quad \text{for } j,k = 1,2,3,4 \qquad (2.13.B)$$

$$d = E(w_i^{(j)} w_i^{(k)}) \quad \text{for } j = 1,2,3,4 \text{ and } k = 5,6,7,8$$

$$e = E(w_i^{(j)} w_i^{(k)}) \quad \text{for } j,k = 5,6,7,8$$

The sample means are used as estimators. Note that unlike serial correlation they have expected values that only depend on the current time index i. The ratios of these average values can be used to define the contemporaneous correlation coefficients which can be used to assess the importance of contemporaneous correlations.

Average values can be found from the values displayed in Table 2.6. These values are a = 0.062, b = 0.064, c = -0.016, d = -0.006, and e = -0.017. Because $r_d = d\ (a\ b)^{-\frac{1}{2}}$ this yields $r_d = -0.0095$ for the correlation between sensors on opposite surfaces. Because $r_c = c\ a^{-1}$ this yields $r_c = -0.25$ for sensors

Table 2.6

Estimate of Contemporaneous Correlation

Cases	n	a	b	c	d	e	f
0	10	0.0186	0.0195	-0.00102	-0.00378	-0.00054	-0.00312
1	8	0.0107	0.0122	0.00122	-0.00001	0.00145	0.00108
2	20	0.0322	0.0200	-0.00720	-0.00066	-0.00421	0.00830
3	15	0.0253	0.0288	-0.00599	-0.00010	-0.00755	0.01260
4	16	0.0393	0.0517	-0.00931	-0.00069	-0.01367	0.02049
5	16	0.0522	0.0450	-0.00626	0.00044	-0.01058	0.02092
6	14	0.0700	0.0658	-0.02084	0.00154	-0.01921	0.03353
7	19	0.0987	0.1150	-0.02918	0.00046	-0.03578	0.06034
8	16	0.0866	0.1190	-0.02698	-0.00030	-0.03573	0.05970
9	18	0.1840	0.1647	-0.05306	-0.00289	-0.04725	0.09253
Mean		0.0618	0.0642	-0.01586	-0.000598	-0.01731	0.03064
Variance		0.0027	0.0027	0.00028	0.000003	0.00028	0.00096

located at the insulated surface. Because $r_e = e\ b^{-1}$ this yields $r_e = -0.27$ for sensor located at the heated surface. If there were no contemporaneous correlation these r values would be zero. If there were significant contemporaneous correlation these r values would have absolute values near unity. Because the r values are relatively small the contemporaneous correlation is not significant and can be ignored.

The estimates of c, d, and e in Equation (2.13.A) may be small because average values are computed and the elements are not necessarily of the same sign. The proper signs can be found by considering the cross product estimator

$$S_{ij} = n^{-1} \sum_{k=1}^{n} w_k^{(i)}\ w_k^{(j)}$$

where S_{ij} is the element in the i-th row and the j-th column of the matrix S. The matrix S has the same dimension as the matrix R_m. The matrix S is called a biased but good estimator by Theil (1971, p. 321). Using this information about the signs, the second R_8 matrix was written as

$$R_8 = \begin{bmatrix} a & & & & & & & \\ f & a & & & \text{symmetric} & & & \\ -f & -f & a & & & & & \\ f & f & -f & a & & & & \\ 0 & 0 & 0 & 0 & b & & & \\ 0 & 0 & 0 & 0 & f & b & & \\ 0 & 0 & 0 & 0 & -f & -f & b & \\ 0 & 0 & 0 & 0 & f & f & -f & b \end{bmatrix} \qquad (2.13.B)$$

Table 2.6 displays the estimates of the element f for the ten cases. Because $r_f = f\ (a\ b)^{-\frac{1}{2}}$ this yields $r_f = 0.486$ for an upper bound on the contemporaneous correlation. Since r_f is small

the R_8 matrix in Equation (2.12) is assumed to be the identity matrix of dimension m; i.e.,

$$R_m = I \qquad (2.13.C)$$

2.3.4 Normality Test

Not all numbers occur in the less significant digits of the temperature measured by a given sensor. Thus, the initial residuals do not appear to be from a normal distribution. It is customary to have the vector a used in Equation (2.7) and in Equation (2.8) to have a normal distribution.

An examination of the pre-heat-flux residuals, for instance, for case 3 data indicated a displacement and alignment by thermocouple pairs by computer COND and a step-like pattern with occasional flat spots in the residuals. This step-like pattern probably occurs because the ten volt signal was digitized to an accuracy that yielded temperature values of approximately 0.02 degrees Fahrenheit. Thus, for a given sensor the processed data tend to have only either odd or even digits in the hundredths place. This truncation occurred in the analog to digital conversion of the measured temperature; i.e., as the thermocouple milli-voltage is converted to a numerical value by the IBM-1800 computer for storage in the disk memory. Barballa (1970) analyzed the mapping of an analog signal to a binary coded word using a ladder transfer curve. Because there are limited digits available in the ladder transfer curve, truncation occurs.

The residuals were tested for normality by the Kuiper test. The Kuiper test statistics for the sample size twenty, n = 20,

are given in Table 2.7. Using Louter and Koerts (1970), the critical point is 0.359 at the 0.01 level of significance. Thus the hypothesis of normality is rejected for four thermocouples and accepted for four thermocouples. The cases where the hypothesis is rejected seem to have flat spots in the residuals. Hence the non-normality might not exist if the temperatures were read more accurately than 0.02 degrees Fahrenheit. The non-normality will be considered as a calibration problem rather than as a problem of the distribution being non-normal.

2.4 Serial ARIMA Models for the Covariance Matrix

In Section 2.3 it was shown that the model for the covariance matrix of the errors, W*, can be approximated by considering only the serial correlations. This section determines the values of the order parameters p, d, and q in the ARIMA(p,d,q) model for the serial correlation. These values are determined by a criterion denoted AIC that has emerged by general consensus of the workers in the field as the accepted identification criterion. The AIC criterion has replaced the procedure of examining the pattern of the estimated autocorrelation and partial autocorrelation and the associated Q-statistic used by Box and Jenkins (1970, p. 291).

2.4.1 ARIMA Model Specification by Akaike's Criterion

The criterion AIC was introduced by Akaike (1972) and is commonly referred to as Akaike's Information Criterion. The AIC criterion is based on Kullback discrimination that is advocated by Beck and

Table 2.7

Kuiper Statistics for Pre-Heat-Flux Residuals

Sensor	0	2	3	4	5	6	7	8
Statistic	0.3204	0.3500	0.5000	0.4020	0.5000	0.3000	0.2880	0.4000

Arnold (1977, p. 467). The form of the criterion depends on the number of parameters that are estimated in the ARIMA model. When the mean is estimated the criterion is written as

$$a^{+} = n \ln(\sigma^2) + 2 n (n - d)^{-1} (p + q + 2) \qquad (2.14)$$

where n is the sample size, σ^2 is the white noise variance, and the number of estimated parameters in the ARIMA(p,d,q) process is p + q + 2. The numerical value two in the number of estimated parameters is used because both the mean and variance are estimated. The criterion a+ was used successfully by Akaike (1972) to identify the order of four time series in the literature that are considered to be test problems, and was also successfully used by Ozaki (1975) for the test series given in Box and Jenkins (1970) as series-A through series-F.

Because the criterion a+ is used for statistical time series, the correct order is not identified for every realization of a series generated from a known ARIMA(p,d,q) process. Shibata (1976) investigated the random behavior of the values estimated for a+ in finite samples by a Monte Carlo study. The results for a true ARIMA(1,0,0) process with candidate processes having values of p from zero through ten were that the order identified most often is the true order. Thus, for the heat conduction data it is reasonable to expect that the correct ARIMA process can be identified by analyzing several realizations but might not be correctly identified from one realization.

Two adjustments were made in Equation (2.14) so that it would perform well for the heat conduction residuals. First, the mean

is zero in the ARIMA(p,d,q) model for the heat conduction residuals specified by matrix Equation (2.8); therefore, the quantity p + q + 2 is changed to the quantity p + q + 1. Second, Jones (1975) showed by a Monte Carlo study that unless the estimator of σ^2 is unbiased too high an order of the ARIMA(p,0,0) process is selected as best; therefore, an adjustment factor is included since the estimator of σ^2 in Appendix A is biased. Hence the AIC criterion used in this dissertation is

$$a^{++} = n \ln(n \varepsilon \sigma^2) + 2 n (n - d)^{-1} (p + d + 1) \qquad (2.15)$$

where

$$\varepsilon^{-1} = n - 1 - p - q$$

and where σ^2 is a biased estimator computed by the procedures in Appendix A. Appendix A also contains algorithms for computing the coefficients (the $\emptyset$'s and θ's) in the ARIMA models in order to insure that the models ranked using Equation (2.15) are both stationary and invertible. The tests for stationarity and invertibility are based on seldom-used equations developed by Wise (1956). When the candidate process did not have a moving-average component, the coefficients were solved with an algorithm developed by Pagano (1972) that reduces the round-off and truncation error. When the candidate process had a moving-average component an analytical solution developed by the author as Equations (A.21) and (A.22) was used. An analytical solution is better than a numerical solution procedure because the nonlinear equations can have more than one

root and the roots are not necessarily real valued when the candidate process does not fit the data.

The AIC criterion was used on several groupings of the residuals: by experiment, by surface, and by sensor. Table 2.8 displays estimates of a^{++} computed to find the best ARIMA model by experiment; no model is consistently best and the values of a^{++} are approximately the same. Although the a^{++} criterion does not clearly identify the best model, it does eliminate from consideration most of the forty-five models considered as candidate models. Because the highest order process considered is ARIMA(4,2,2) there are forty-five candidate models. The elimination from consideration of most of the models is illustrated by comparing the estimates of a^{++} in Table 2.8 with the typical value for white noise of 600, and noting that several candidate models had a value of the a^{++} criterion less than that for white noise. A best model does not emerge in the models displayed in Table 2.9 for the 336 residuals at the insulated surface, nor does one emerge in the models displayed in Table 2.10 for the heated surface.

In the data gathered by Farnia (1976) a single predetermined sampling rate is used and for this single sampling rate the lack of uniqueness in the best model in Tables 2.8 through 2.10 is good since the most easily computed model in the top six can be used as the correct model. The estimates of the coefficients in the top six models are not equally stable, as shown in Table 2.11. The standard deviations (s.d.) for the coefficients of $\emptyset$ and θ are smallest for the ARIMA(1,0,1) and ARIMA(0,1,1) models. When the residuals are

Table 2.8

Top Six ARIMA Models by Experiment

Case 0	
a^{++}	ARIMA
-2901.22	1,1,1
-2895.90	1,0,2
-2891.48	0,0,2
-2887.10	1,0,1
-2864.93	2,0,0
-2860.40	2,1,0

Case 1	
a^{++}	ARIMA
-2348.13	2,0,2
-2233.82	1,0,1
-2228.29	1,0,2
-2218.82	2,0,0
-2188.26	3,1,0
-2180.33	0,1,1

Case 2	
a^{++}	ARIMA
-2834.74	1,1,1
-2830.29	0,1,2
-2829.53	1,0,1
-2816.78	2,1,1
-2798.33	1,0,2
-2764.39	4,1,1

Case 3	
a^{++}	ARIMA
-2595.86	3,1,1
-2534.87	1,1,1
-2534.36	0,1,2
-2530.40	2,1,0
-2528.53	0,1,1
-2527.40	3,1,0

Case 4	
a^{++}	ARIMA
-2879.95	1,1,1
-2874.30	0,1,2
-2849.02	1,1,2
-2825.30	3,1,0
-2819.04	0,1,1
-2805.18	3,1,1

Case 5	
a^{++}	ARIMA
-2422.12	0,1,1
-2418.25	2,1,1
-2414.03	1,1,1
-2413.95	0,1,2
-2412.57	2,1,0
-2409.84	3,1,0

Case 6	
a^{++}	ARIMA
-2189.23	3,1,1
-2185.90	0,1,1
-2182.38	1,1,1
-2182.37	0,1,2
-2181.97	2,1,0
-2180.55	1,1,0

Case 7	
a^{++}	ARIMA
-2361.11	2,1,1
-2290.40	0,1,1
-2276.13	1,1,1
-2276.02	0,1,2
-2276.00	2,1,0
-2275.86	1,1,0

Case 8	
a^{++}	ARIMA
-2208.29	2,1,2
-2203.82	0,1,2
-2203.76	1,1,1
-2202.92	2,1,0
-2202.84	0,1,1
-2201.83	1,1,0

Case 9	
a^{++}	ARIMA
-2546.83	2,1,1
-2531.60	4,1,1
-2511.28	2,1,2
-2503.65	0,1,1
-2475.27	3,1,1
-2460.07	1,1,1

Table 2.9

Top Six ARIMA Models at Insulated Surface

Case 0	
a^{++}	ARIMA
-1458.21	0,1,1
-1453.33	1,0,1
-1443.21	1,0,2
-1437.30	2,0,0
-1435.36	1,1,1
-1434.75	0,1,2

Case 1	
a^{++}	ARIMA
-1362.03	2,1,1
-1333.33	0,1,1
-1305.08	1,1,1
-1304.71	0,1,2
-1294.40	1,1,2
-1285.85	3,1,0

Case 2	
a^{++}	ARIMA
-1507.39	1,0,2
-1505.95	1,0,1
-1450.74	2,0,0
-1446.96	1,1,1
-1446.29	0,1,2
-1430.47	3,1,0

Case 3	
a^{++}	ARIMA
-1338.37	0,1,1
-1330.85	3,1,0
-1326.76	1,1,1
-1326.54	0,1,2
-1324.61	2,1,0
-1317.72	1,1,0

Case 4	
a^{++}	ARIMA
-1467.84	1,1,1
-1467.76	0,1,2
-1460.96	0,1,1
-1458.55	2,1,1
-1433.80	3,1,0
-1429.38	4,1,1

Case 5	
a^{++}	ARIMA
-1416.34	1,1,2
-1399.50	0,1,2
-1372.19	3,1,0
-1357.93	2,1,0
-1350.26	0,1,1
-1339.49	1,0,2

Case 6	
a^{++}	ARIMA
-1413.91	1,1,1
-1410.75	0,1,2
-1396.49	3,1,0
-1386.11	0,1,1
-1381.93	2,1,0
-1366.69	4,1,1

Case 7	
a^{++}	ARIMA
-1472.11	1,1,2
-1452.45	1,1,1
-1439.72	0,1,2
-1429.16	3,1,0
-1410.68	2,1,0
-1399.76	1,0,2

Case 8	
a^{++}	ARIMA
-1404.26	1,1,2
-1399.99	2,1,2
-1390.18	3,1,0
-1383.12	2,1,0
-1374.64	0,1,1
-1350.80	1,1,0

Case 9	
a^{++}	ARIMA
-1447.61	0,1,1
-1428.80	2,1,1
-1405.34	1,1,2
-1403.15	1,1,1
-1401.98	0,1,2
-1400.83	2,1,0

Table 2.10

Top Six ARIMA Models at Heated Surface

Case 0	
a^{++}	ARIMA
-1481.96	1,1,1
-1469.60	0,1,2
-1462.10	1,1,2
-1456.11	3,1,0
-1461.04	2,1,0
-1450.51	2,1,2

Case 1	
a^{++}	ARIMA
-1033.38	3,1,0
-1032.95	0,1,1
-1029.27	1,1,1
-1029.27	0,1,2
-1028.69	2,1,0
-1027.48	1,1,0

Case 2	
a^{++}	ARIMA
-1369.05	1,1,1
-1367.61	0,1,2
-1357.00	2,1,1
-1340.66	1,0,1
-1338.24	3,1,0
-1336.93	2,1,0

Case 3	
a^{++}	ARIMA
-1229.18	1,1,1
-1228.72	0,1,2
-1227.89	0,1,1
-1227.85	2,1,0
-1226.19	1,1,0
-1225.95	3,1,0

Case 4	
a^{++}	ARIMA
-1447.07	1,1,1
-1440.01	0,1,2
-1420.67	1,1,2
-1412.25	3,1,0
-1407.41	0,1,1
-1407.10	2,1,0

Case 5	
a^{++}	ARIMA
-1139.92	0,1,1
-1133.38	1,1,0
-1130.89	0,1,2
-1130.67	1,1,1
-1130.67	2,1,0
-1129.90	2,1,2

Case 6	
a^{++}	ARIMA
-981.27	0,1,1
-980.45	1,1,0
-977.65	1,1,1
-977.65	0,1,2
-977.65	2,1,0
-974.87	1,1,2

Case 7	
a^{++}	ARIMA
-1051.50	2,1,1
-1051.08	0,1,1
-1050.02	1,1,0
-1047.30	1,1,1
-1047.29	0,1,2
-1047.29	2,1,0

Case 8	
a^{++}	ARIMA
-1013.41	2,1,2
-997.73	0,1,2
-997.16	2,1,0
-994.38	1,1,2
-991.30	0,1,1
-991.27	1,1,0

Case 9	
a^{++}	ARIMA
-1151.89	0,1,1
-1123.10	1,1,1
-1122.52	0,1,2
-1122.31	2,1,0
-1121.49	1,1,2
-1121.19	3,1,0

Table 2.11

Stability of ARIMA Coefficients by Experiment for 0.3 Second Data

	ARIMA(1,0,1)				ARIMA(0,1,1)	
Case	ϕ	θ	σ^2	Case	θ	σ^2
0	0.973	0.459	0.0172	0	0.473	0.0174
1	0.979	0.413	0.0357	1	0.438	0.0373
2	0.992	0.742	0.0147	2	N.V.	N.V.
3	0.985	0.292	0.0262	3	0.365	0.0221
4	0.985	0.587	0.0159	4	0.632	0.0143
5	0.990	0.343	0.0320	5	0.416	0.0259
6	0.978	0.283	0.0421	6	0.307	0.0370
7	0.965	0.353	0.0376	7	0.400	0.0316
8	0.978	0.289	0.0589	8	0.199	0.0361
9	0.984	0.529	0.0284	9	0.609	0.0229
avg.	0.981	0.429	0.0309	avg.	0.4266	0.0272
s.d.	0.008	0.152	0.0137	s.d.	0.1362	0.0087

	ARIMA(1,1,1)				ARIMA(1,1,0)	
Case	ϕ	θ	σ^2	Case	ϕ	σ^2
0	0.157	0.723	0.0161	0	-0.387	0.0183
1	-0.005	0.431	0.0374	1	-0.367	0.0385
2	-0.045	0.827	0.0139	2	-0.515	0.0180
3	0.965	0.481	0.0218	3	-0.322	0.0224
4	0.070	0.803	0.0130	4	-0.452	0.0159
5	-0.050	0.355	0.0261	5	-0.355	0.0266
6	-0.013	0.293	0.0370	6	-0.280	0.0373
7	-0.170	0.200	0.0322	7	-0.345	0.0323
8	-0.586	-0.412	0.0359	8	-0.192	0.0361
9	-0.138	0.394	0.0244	9	-0.444	0.0252
avg.	-0.019	0.0410	0.0258	avg.	-0.366	0.0271
s.d.	0.388	0.3603	0.0095	s.d.	0.092	0.0085

	ARIMA(0,1,2)				ARIMA(1,0,0)	
Case	θ_1	θ_2	σ^2	Case	ϕ	σ^2
0	0.550	0.080	0.0163	0	0.898	0.0202
1	0.436	-0.002	0.0374	1	0.933	0.0411
2	0.866	-0.041	0.0140	2	0.838	0.0211
3	0.384	0.036	0.0218	3	0.969	0.0283
4	0.725	0.048	0.0131	4	0.895	0.0205
5	0.404	-0.021	0.0261	5	0.976	0.0356
6	0.305	-0.004	0.0370	6	0.956	0.0452
7	0.369	-0.067	0.0322	7	0.910	0.0415
8	0.179	-0.117	0.0359	8	0.953	0.0633
9	0.529	-0.079	0.0244	9	0.918	0.0354
avg.	0.475	-0.017	0.0258	avg.	0.925	0.0352
s.d.	0.202	0.061	0.0095	s.d.	0.041	0.0135

considered by sensor the number of trials is increased from ten to eighty but the sample size per trial is reduced to eighty-four. For the ten experiments with eight sensors per experiment the number of trials is large enough to gain information by tabulating the frequency that a particular ARIMA model is identified as being best. In Table 2.12 the ARIMA(0,1,1) model occurs most frequently and is the best model, although the estimates of the coefficients are somewhat inaccurate because there are only eighty-four residuals per sensor.

Thus, to select the ARIMA(1,0,1) model as best involved not only the AIC criterion but also the stability of the estimated coefficients and the frequency the models are identified as best when there are many time series. Combining the results from Tables 2.8 through 2.12 yields the best model as ARIMA(1,0,1) with the following values for the coefficients: $\phi = 0.981$, $\theta = 0.429$, and $\sigma^2 = 0.031$.

2.4.2 Sampling Rate and ARIMA Model Specification

The order of an ARIMA process varies with the sampling rate. Wei (1977) and MacGregor (1977) showed that an ARIMA(p,d,q) model becomes an ARIMA(p,d,q*) model when only every h-th value of the original ARIMA(p,d,q) process is used. The change in the order of the moving-average component is given by the integer part of a combination of the orders in the original model. This combination is written as

$$q^* = \text{int}(p + d + (q - p - d)\, h^{-1}) \qquad (2.16)$$

If the original process, for example, is ARIMA(1,0,1) then the process when every h-th point in the series is used remains ARIMA(1,0,1).

It is important to investigate the change in order of the ARIMA(p,d,q) model with sampling rate because sampling rates other than 0.3 seconds between measurements can be used on the IBM-1800 computer and data acquisition system. The best sampling rate depends on the covariance matrix for the physical parameters which contains the error covariance matrix W*. The matrix W* is determined by the ARIMA(p,d,q) model and its coefficients.

The ARIMA models identified as best at a sampling rate of 0.3 seconds between measurements are shown in Table 2.12, and are based on eighty-four measurements per sensor. The ARIMA models identified as best at a sampling rate of 0.6 seconds between measurements are shown in Table 2.13, and are based on forty-two measurements per sensor. In both cases the residuals are from the standard least squares estimates of the physical parameters k and c. The same time period was analyzed in both studies with every other residual point used in the 0.6 second study. As expected when the sampling rate increases from 0.3 seconds to 0.6 seconds, more white noise models are identified by Equation (2.15) as being best. In both tables the most frequently selected model is ARIMA(0,1,1), which occurs nineteen times out of eighty cases in Table 2.12 and thirty-one times out of eighty cases in Table 2.13. Because an ARIMA(0,1,1) model remains an ARIMA(0,1,1) model when the sampling rate changes, the ARIMA(0,1,1) model is identified as having the best representation of the data. Because the sample size is small for the analysis in both Table 2.12

Table 2.12

Frequency of Best ARIMA Models for 0.3 Second Data

	q = 0				q = 1				q = 2			
	p=0	1	2	3	p=0	1	2	3	p=0	1	2	3
d = 0	8	0	6	0	0	9	0	0	0	3	0	0
d = 1	0	1	0	9	19	10	3	1	2	8	0	1
d = 2	0	0	0	0	0	0	0	0	0	0	0	0

Table 2.13

Frequency of Best ARIMA Models for 0.6 Second Data

	q = 0				q = 1				q = 2			
	p=0	1	2	3	p=0	1	2	3	p=0	1	2	3
d = 0	15	4	2	0	1	8	0	0	1	5	0	0
d = 1	0	3	4	3	31	2	1	0	0	0	0	0
d = 2	0	0	0	0	0	0	0	0	0	0	0	0

and Table 2.13, the ARIMA(0,1,1) model could be an ARIMA(1,0,1) model with a value of the autoregressive coefficient ϕ close to unity.

2.4.3 Sampling Rate and ARIMA Coefficients

At both sampling rates considered in Section 2.4.2 the best model was ARIMA(1,0,1). In this section the sampling rate is changed and the associated change in the coefficients in the ARIMA(1,0,1) model is investigated. The change with sampling rate of the ARIMA(1,0,1) coefficients is predicted by Equation (2.18).

Equation (2.18) follows directly from Equations (A.6) and (A.7) in MacGregor (1976) and Equation (3.4.7) in Box and Jenkins (1970). The notation for the ARIMA(1,0,1) series when every point is used ($h = 1$) is

$$w_i = \phi\, w_{i-1} + a_i - \theta\, a_{i-1} \quad \text{where } a_i \approx N(0,\sigma^2)$$

while for the series sampled at every h-th point the notation used for the ARIMA(1,0,1) model is

$$w_j = \phi^h\, w_{j-1} + a_j - \alpha\, a_{j-1} \quad \text{where } a_j \approx N(0,\beta^2)$$

with h an exponent. The moving-average parameter and white noise variance using every h-th value are found by equating the theoretical autocovariance and are determined by the pair of equations

$$\frac{1 - 2\alpha\,\phi^h + \alpha^2}{1 - \phi^{2h}}\ \beta 2 = \frac{1 - 2\,\phi\,\theta + \theta^2}{1 - \phi^2}\ \sigma^2 \qquad (2.18.A)$$

$$\frac{\phi^h + \alpha^2\,\phi^h - \alpha\,\phi^{2h} - \alpha}{1 - \phi^{2h}}\ \beta^2 = \frac{\phi + \phi\,\theta^2 - \theta\,\phi^2 - \theta}{1 - \phi^2}\ \sigma^2 \qquad (2.18.B)$$

The reference values of θ and σ^2 become the α and β^2 values predicted by Equation (2.18) while the reference value ϕ becomes the predicted value ϕ^h.

Table 2.14 contains estimates of the coefficients in the ARIMA(1,0,1) and ARIMA(0,1,1) models when the sampling rate is 0.6 seconds between measurements. Table 2.15 contains predicted coefficients for the ARIMA(1,0,1) model for 0.6 seconds between measurements that are obtained by using Equation (2.18) on the estimates given in Table 2.11. The agreement between the estimated values in Table 2.14 and the predicted values in Table 2.15 is good. This gives additional support to the conclusion in Section 2.4.2 that the best fitting model is ARIMA(1,0,1).

2.5 Investigation of the ARIMA(1,0,1) Model and Residuals

The standard deviations in Table 2.11 are high for the ARIMA(1,0,1) model, especially for the moving-average coefficient and the white noise variance. In another problem, Beck and Arnold (1977, p. 263) suggested that the variance may have a magnitude related to the magnitude of the heat flux. Hence the ARIMA(1,0,1) coefficients were averaged using the two levels of the heat flux recorded in Table 2.1. At the low heat flux the parameter estimates are ϕ = 0.980, θ = 0.534 and σ^2 = 0.023 while at the high heat flux the parameter estimates are ϕ = 0.982, θ = 0.324, and σ^2 = 0.039. Although there are only a few experiments on which to base a conclusion, the magnitude of θ and σ^2 seems to depend on the applied heat flux while ϕ does not.

Table 2.14

Stability of ARIMA Coefficients by Experiment at 0.6 Seconds

	ARIMA(1,0,1)				ARIMA(0,1,1)	
Case	$\emptyset$	θ	σ^2	Case	θ	σ^2
0	0.940	0.303	0.0229	0	0.402	0.0225
1	0.962	0.404	0.0599	1	0.507	0.0632
2	0.951	0.449	0.0191	2	0.503	0.0193
3	0.967	0.139	0.0408	3	0.252	0.0323
4	0.975	0.521	0.0198	4	0.667	0.0153
5	0.978	0.212	0.0508	5	0.319	0.0395
6	0.959	0.220	0.0677	6	0.273	0.0638
7	0.945	0.315	0.0510	7	0.336	0.0480
8	0.963	0.159	0.0678	8	0.687	0.0456
9	0.959	0.358	0.0409	9	0.433	0.0346
avg.	0.960	0.308	0.0437	avg.	0.438	0.0384

Table 2.15

Predicted ARIMA(1,0,1) Coefficients at 0.6 Seconds

Case	$\emptyset$	θ	σ^2
0	0.947	0.273	0.0191
1	0.958	0.211	0.0396
2	0.984	0.649	0.0159
3	0.970	0.036	0.0283
4	0.970	0.445	0.0176
5	0.980	0.114	0.0352
6	0.957	0.021	0.0452
7	0.931	0.124	0.0410
8	0.957	0.031	0.0634
9	0.968	0.369	0.0316
avg.*	0.962	0.234	0.0343

*The average is the predicted average.

Anderson (1975) investigated whether the fitted ARIMA model can be realized from a linear combination of ARIMA processes. This is called the interpretation step in ARIMA modeling. If Equation (2.19) is satisfied then an ARIMA(1,0,1) model with particular coefficients can be viewed as being generated as the sum of values generated by an ARIMA(1,0,0) process and the values generated by a white noise process. The white noise process is an ARIMA(0,0,0) process. The condition that all the variances in the component processes and the sum of the component processes are positive is used to obtain the inequality given by Anderson (1975) as

$$0 \leq \theta \, \phi^{-1} \leq (1 + \theta^2) \, (1 + \phi^2)^{-1} \tag{2.19}$$

Equation (2.19) is satisfied at both the low heat flux ARIMA(1,0,1) model where $\phi = 0.980$ and $\theta = 0.534$ and at the high heat flux ARIMA (1,0,1) model where $\phi = 0.982$ and $\theta = 0.324$. Therefore, one can view the ARIMA(1,0,1) process as the sum of two processes that have a physically reasonable interpretation. Before the heat flux is applied, the measured temperature is a random fluctuation about the constant mean temperature; hence, the residuals are white noise. After the heat flux is applied, the temperature is sampled faster than the duration of small factors "not-fully-accounted-for" that can be described as ARIMA(1,0,0) processes. The use of an ARIMA (0,1,0) model for the "not-fully-accounted-for" factors is discussed in Section 2.3.2.

By analyzing the residuals from standard least squares, an ARIMA(1,0,1) model was selected in Section 2.4 to represent the

residuals. A short discussion is required for the fact that the residuals almost always show serial correlation even when the true errors possess none; see for example Bard (1974, p. 248) and Draper and Smith (1966, p. 94).

It is easily shown that for the linear model with independent errors

$$y = X \beta + w \tag{2.20}$$

that the residuals

$$e = y - X b \tag{2.21}$$

are related to the errors, w, by

$$e = M w \tag{2.22}$$

where M is the idempotent matrix

$$M = I - X (X^t X)^{-1} X^t.$$

In Equation (2.21) b is the least squares estimator for β in Equation (2.20). Because M is both idempotent and not the identity matrix, M is singular. Therefore w can not be computed from a knowledge of M and e. However, Theil (1971, p. 271) stated that M is its own Moore-Penrose generalized inverse. Hence, to gain insight on the relationship between the errors and the residuals for the heat conduction problem a mapped residual vector $w^{\#}$ is defined as

$$w^{\#} = M e. \tag{2.23}$$

If the residuals e are not changed greatly under the idempotent mapping in Equation (2.23), then it is reasonable to assume that the errors mapped by Equation (2.22) are not changed greatly either.

Typical residuals and mapped residuals from case-0 data are shown in Figures 2.22 and 2.23. A subjective judgment is required on whether the mapping has greatly changed the residuals. The shape of the curve does seem to be shifted toward the zero axis. The fluctuations along the curve seem to be unchanged. Based on Equation (2.11), these two observations indicate that if the errors are ARIMA(1,0,1) then the residuals are also ARIMA(1,0,1) with the coefficient $\emptyset$ being smaller than its value for the errors. The effect of mapping by Equation (2.23) and hence implicitly by Equation (2.22) seems to be small enough, however, that the identification of the proper ARIMA model is not affected by using residuals rather than by using a more complicated procedure involving an estimate of the errors.

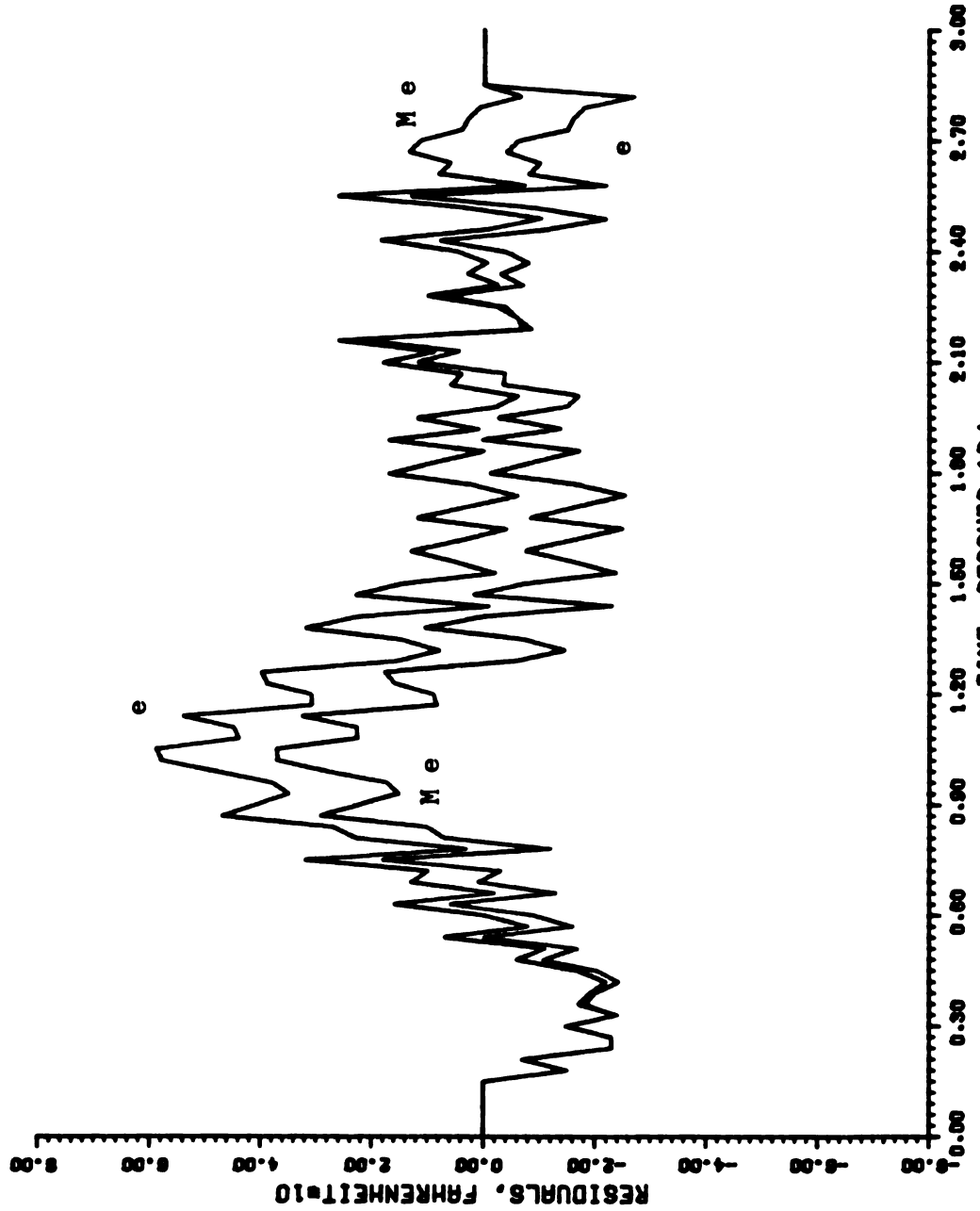

Figure 2.21. Insulated surface, e and M e: Case-0 Sensor-3

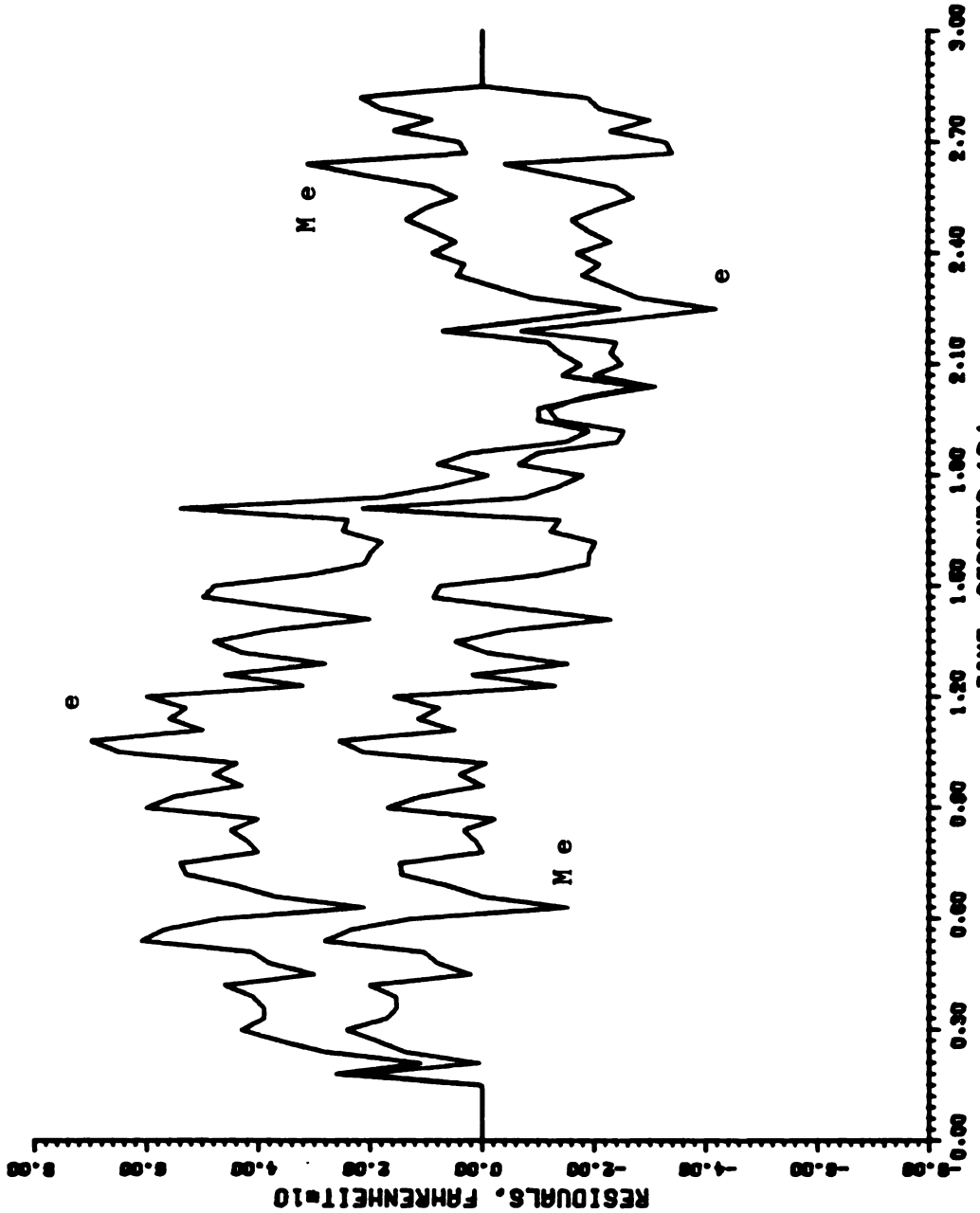

Figure 2.22. Heated surface, e and M e: Case-0 Sensor-7

CHAPTER III

ESTABLISHING THE CONFIDENCE REGION

In this chapter statistically valid confidence regions are established and Aitken's least squares estimators are obtained for both the thermal parameters and for the confidence ellipses. In Section 3.1 the confidence region is defined as a confidence ellipse with particular forms given that are necessarily statistically valid. In Section 3.2 the statistical validity of the central F distribution is shown when the ARIMA coefficients are known. In Section 3.3 the noncentral F distribution is proposed as a statistically valid approximation when the ARIMA coefficients are estimated. In Section 3.4 Aitken's point estimators of the thermal and ARIMA parameters are given and tested. In Section 3.5 the accuracy of the point estimates of the thermal and ARIMA(1,0,1) parameters is investigated by a Monte Carlo study. In Section 3.6 the several forms of the confidence ellipse given in Section 3.1 are plotted and compared using estimates from Section 3.5.

3.1 Estimators and Confidence Regions Considered

This section investigates three confidence regions. The linear (or linearized) physical model under consideration is

$$y = X \beta + w \qquad (3.1.A)$$

where the additive correlated errors have the distribution

$$w \approx N(0,W) \tag{3.1.B}$$

The eight assumptions made in using this error model are stated following Equation (1.14.A). It is assumed that the error covariance matrix is generated from an ARIMA model and can be written as

$$W = \sigma^2 V \tag{3.2}$$

where σ^2 is estimated and V is not the identity matrix.

Two estimators are considered for the physical parameter vector β. Equation (3.3) is the standard least squares estimator that can be written as

$$b_{SLS} = (X^t X)^{-1} X^t y. \tag{3.3}$$

Equation (3.4) is the Aitken's least squares estimator that can be written as

$$b_{ALS} = (X^t V^{-1} X)^{-1} X^t V^{-1} y. \tag{3.4}$$

In Equation (3.3) and (3.4) the subscripts on the vector b are used to distinguish the standard least squares (SLS) from Aitken's least squares (ALS).

The confidence region is assumed to be represented by an ellipse. The test in Draper and Shaw (1974) to justify the use of this linearized confidence region for a nonlinear parameter estimation problem is beyond the scope of this dissertation. The ellipse can be written as

$$(b - \beta)^t Q^{-1} (b - \beta) \leq k s^2 G(1 - \alpha) \tag{3.5}$$

where

$$\text{cov}(b) = \sigma^2 Q. \tag{3.6}$$

Note that the left hand side of Equation (3.5) is a quadratic form involving the inverse of the covariance matrix for the physical parameters given in Equation (3.6). The size of the ellipse in Equation (3.5) is determined by the fractile $G(1 - \alpha)$ for a given confidence coefficient $1-\alpha$. The value for the fractile is determined by the probability distribution G. The meaning of Equation (3.5) is that the probability is α that the ellipse centered at the true value β will not contain a point estimate of b.

Three forms are considered for the Q matrix in Equation (3.5). The three cases are distinguished by subscripts SLS, ALS, and MLS where MLS denotes mixed least squares. When it is incorrectly assumed that V equals I in Equation (3.2) and Equation (3.3) is used to estimate β then the Q matrix is written as

$$Q_{SLS} = (X^t X)^{-1} \tag{3.7}$$

which is incorrect. When Equation (3.3) is used to estimate β but the correct error covariance matrix is used to compute the parameter covariance matrix then the Q matrix is written as

$$Q_{MLS} = (X^t X)^{-1} X^t V X(X^t X)^{-1} \tag{3.8}$$

which is correct. When Equation (3.4) is used to estimate the vector β then the Q matrix is written as

$$Q_{ALS} = (X^t V^{-1} X)^{-1} \tag{3.9}$$

which is correct provided V is known.

The distribution appropriate for $G(1-\alpha)$ in Equation (3.5) is established in Sections 3.2 and 3.3. It is necessary to

establish a statistically valid distribution in order to use Equation (3.5) in a correct manner.

3.2 Confidence Region When the V Matrix Is Known

For the V matrix in Equation (3.2) two cases are considered: V matrix known, and V matrix estimated. The case where the error covariance matrix V is known is much more straightforward than the case of an estimated V matrix discussed in the next section. However, even for this case, standard textbooks such as Theil (1971) do not derive the probability distribution of G in Equation (3.5) for all three covariance matrices given in Equations (3.7) through (3.9). These three covariance matrices are used in Equation (3.5) to determine three confidence regions: Aitken's least squares (ALS), mixed least squares (MLS), and standard least squares (SLS).

The steps in the derivation of the probability distributions for these three confidence regions are given in Appendix B. The derivation for the MLS problem was not found in the literature. The assumptions needed for the ALS problem are stated more explicitly in Appendix B than they are in derivations in the literature.

The left hand side of Equation (3.5) involves the ratio of two quadratic forms in normal variates. For sake of discussion the forms are denoted w^t A w and w^t B w and the probability distribution of w is N(0,C). The derivation in Appendix B depends on three theorems. A theorem by Shanbhag (1970) is used to show that w^t A w is distributed as chi-square if and only if

$$\mathrm{tr}(A\,C)^2 = \mathrm{tr}(A\,C)^3 = \mathrm{tr}(A\,C)^4 \qquad (3.10.A)$$

where tr denotes the trace of the matrix. (Khatri [1978] reviewed other equivalent necessary and sufficient conditions.) A second theorem by Shanbhag (1970) is used to show that the degrees of freedom of the chi-square distribution of w^t A w is the rank of A which can be computed from the trace as

$$\text{rank}(A) = \text{tr}(A\ C)^2. \qquad (3.10.B)$$

The procedure represented by Equations (3.10) is used to determine whether w^t B w is chi-square also. The independence of the two quadratic forms w^t A W and w^t B w is needed to show that their ratio has an F distribution. The independence is tested by a theorem stated by Carpenter (1950) that statistical independence holds if and only if

$$A\ C\ B = 0 \qquad (3.11)$$

The results obtained in Appendix B are that G in Equation (3.5) has a central F distribution when either Equation (3.7) or Equation (3.9) is used. The use of Equation (3.7) is still incorrect, however, when V is not the identity matrix because Q is incorrect. When Equation (3.8) is used the distribution of G in Equation (3.5) could not be determined.

3.3 Confidence Region When the V Matrix Is Estimated

Few papers have been written on the problem of confidence regions when the error covariance matrix V is estimated. A confidence region was suggested by Glesser and Olkin (1972) for an estimated error covariance matrix that has a completely unknown and arbitrary structure. The error covariance matrix was estimated from

replicated data with t replicas of the data vector denoted $y^{(1)}$ through $y^{(t)}$, and the sample cross product matrix S used. The matrix S is computed from

$$S = \sum_{i=1}^{t} (y^{(i)} - \bar{y})\ (y^{(i)} - \bar{y})$$

The matrix S is related to the matrix W by $E(S) = (t - 1)\ W$. One confidence region suggested by Glesser and Olkin (1972) has the form

$$t^2\ (b - \beta)^t\ X^t\ S^{-1}\ X\ (b - \beta) \leq (1 + g)\ k\ F(k,t-n,1-\alpha) \qquad (3.12)$$

Thus in Equation (3.12) the effect of estimating the error covariance matrix is to multiply the usual fractile by a factor (1 + g) where

$$g = t\ (\bar{y} - X\ b)^t\ S^{-1}\ (\bar{y} - X\ b). \qquad (3.13.A)$$

It appears that no model is given in the literature for the distribution of the error covariance matrix when it is estimated from an ARIMA model. Hence, a reasonable approach is to use the confidence region in Equation (3.12) but to treat the random statistic g in Equation (3.13.A) as an empirical parameter.

One approach to take in finding the empirical parameter g given in Equation (3.13.A) is as follows. Patnaik's approximation to a non-central F distribution is

$$F'(n_1,n_2,\lambda,1-\alpha) = (1 + \lambda\ n_1^{-1})\ F(m_1,n_2,1-\alpha)$$

where

$$m_1 = (n_1 + \lambda)^2\ (n_1 + 2\ \lambda)^{-1}.$$

If m_1 is approximately equal to n_1 then the Patnaik's approximation merely multiplies the central F fractile by a factor $(1 + \lambda\ n_1^{-1})$. Alternately for Equation (3.12) a factor times a central F fractile can be viewed as a non-central F which is useful because the non-centrality parameter in the non-central F can be estimated. The postulate to be tested is that the random parameter g in Equation (3.12) can be related to the noncentrality parameter when the errors come from an ARIMA model; i.e.,

$$g = \lambda\ n_1^{-1} \ . \qquad (3.13.B)$$

3.3.1 Confidence Coefficient Estimator of Noncentrality Coefficient

In Section 3.3 a non-central F distribution is proposed for $G(1-\alpha)$ in Equation (3.5). Some data are available in the literature to test the fit of the distribution, specifically the data by Gallant and Goebel (1976). I analyzed these data in order to estimate λ from the empirical confidence coefficients at the percentiles of α = 0.01, 0.05, 0.10, and 0.5 for the central F distribution. The estimates of λ at α = 0.1 were based on a fit in which the non-central F was computed using the transformation suggested by Severo and Zelen (1960) and reviewed by Tiku (1966).

The two parameter model used by Gallant and Goebel (1976) is

$$y_i = \beta_1 \exp(\beta_2\ x_i) + w_i \qquad (3.14)$$

where w_i was assigned either second order autoregressive errors, ARIMA(2,0,0), or white noise errors, ARIMA(0,0,0). For the white noise errors, the results are shown in Table 3.1. The observed

confidence coefficients agree with the assigned values from the F distribution for the white noise errors which indicates that t = 2000 trials is adequate to establish the empirical distribution and that Equation (3.5) is valid for Equation (3.14) which is a non-linear parameter estimation problem.

Table 3.1

Confidence Coefficients for ARIMA(0,0,0) Compared to Central F Values

Assigned Fractile	Observed Confidence Coefficient		Value for Assigned Fractile, $\lambda = 0$
$F(\alpha)$	β_1	β_2	$1-\alpha$
7.09	0.990	0.990	0.99
4.00	0.948	0.951	0.95
2.79	0.892	0.892	0.90
0.46	0.490	0.491	0.50

In the case when w_i represents ARIMA(2,0,0) errors, the estimation of the autoregressive coefficients causes the empirical distribution of the β's to have thicker tail regions than when the autoregressive coefficients are not estimated. This empirical distribution is postulated to be represented by a non-central F distribution. The empirical percentiles at the ninety-five percent level in the two tailed t-distribution were found by Gallant and Goebel (1976, Table 2). The corresponding one tail central F-distribution value is $\alpha = 0.1$ which has the fractile value 5.712 for β_1 and empirical fractile value 5.593 for β_2. The noncentrality coefficient that

yields a computed value of $\alpha = 0.1$ are $\lambda = 1.242$ for β_1 and $\lambda = 1.188$ for β_2. The agreement is good between the empirical and computed confidence coefficients, as shown in Table 3.2. In Table 3.2 the average value was used for the noncentrality coefficient ($\lambda = 1.215$), and the Severo and Zeren (1960) approximation was used in computing the non-central F distribution.

Table 3.2

Confidence Coefficients for ARIMA(2,0,0)
Compared to Non-Central F Values

Assigned Fractile	Observed Confidence Coefficient		Value for Assigned Fractile, $\lambda = 1.215$
$F(\alpha)$	β_1	β_2	$1-\alpha$
7.09	0.924	0.928	0.9383
5.712	0.90	N.V.	0.9020
5.593	N.V.	0.90	0.8979
4.00	0.843	0.842	0.8207
2.79	0.763	0.774	0.7185
0.46	0.389	0.397	0.2600

3.3.2 Trial Statistic Estimator of the Noncentrality Coefficient

Because there were $t = 2000$ trials, the estimates of the parameters were not presented in either Goebel (1974) or Gallant and Goebel (1976). If they had been presented then an estimator of λ could be used that does not require the construction of an empirical distribution used in Section 3.3.1. Pandey and Rahman (1971) derived a

maximum likelihood estimator of λ based on the statistic u. The statistic u is defined as

$$u = (b - \beta)^t X^t V^{-1} X (b - \beta) s^{-2} k^{-1} \qquad (3.15)$$

where k is the number of physical parameters. The statistic in Equation (3.15) can be obtained by substituting Equation (3.9) into Equation (3.5) and solving for $G(1-\alpha)$. In a Monte Carlo study all the terms in Equation (3.15) are either known or estimated so that the value of u at the t-th trial is available and will be denoted u_t.

For Monte Carlo studies on a one parameter linear model and on a two parameter heat conduction model we need the following results from Pandey and Rahman (1971) valid for large sample size. For a one parameter model the estimator of λ is

$$t = \sum_{i=1}^{t} u_i \, (2 \lambda u_i)^{-\frac{1}{2}} \tanh((2 \lambda u_i)^{\frac{1}{2}}) \qquad (3.16.A)$$

For a model with two parameters an estimator for λ is

$$t = \sum_{i=1}^{t} \frac{2 u_i I_1((2 \lambda u_i)^{\frac{1}{2}})}{(2 \lambda u_i)^{\frac{1}{2}} I_0((2 \lambda u_i)^{\frac{1}{2}})} \qquad (3.16.B)$$

Equation (3.16.A) is used in Section 3.3.3 and Equation (3.16.B) is used in Section 3.5.3.

3.3.3 Monte Carlo Study of a Linear Model

In this section parameters are estimated for the following simple linear model

$$y_i = \beta + w_i \tag{3.17}$$

Equation (3.17) is Equation (1.15) with the coefficient α set to zero. A Monte Carlo study was used by both Kmenta (1971, Table 8-5) and by Beck (1974) to study the relative efficiency of maximum likelihood to standard least squares estimator where relative efficiency is defined in Section 1.3.3. The purpose of this section is (1) to become familiar with the Monte Carlo method, (2) to determine how accurately ARIMA parameters can be estimated, and (3) to test the estimators of the noncentrality parameter discussed in Section 3.3.1 and Section 3.3.2.

Maximum likelihood estimates of the parameters in Equation (3.17) at several combinations of parameter values are presented in Table 3.3. For a general parameter p the bias is denoted δp and the sample variance is denoted γp. The bias is computed from the estimator

$$\delta p = \sum_{i=1}^{t} t^{-1} (p_i - p^{\#})$$

and the sample variance is computed from the estimator

$$\gamma p = \sum_{i=1}^{t} t^{-1} (p_i - p^{\#})^2$$

where the true value of the parameter p is $p^{\#}$ and the estimate or parameter p at the i-th trial is p_i, and there are t trials. The results in Table 3.3 indicate that the variance $\gamma\beta$ and the bias $\delta\beta$ both increase as the autoregressive coefficient $\emptyset$ approaches unity.

The Monte Carlo study in Table 3.3 contains 4600 trials and the computation of the estimates for the table required 830 central

Table 3.3

Sample Mean and Variance for Equation (3.17) by Monte Carlo Study

t	n	β	ϕ	σ^2	$\delta\beta$	$\delta\phi$	$\delta\sigma^2$	$\gamma\beta$	$\gamma\phi$	$\gamma\sigma^2$
600	60	0.0	0.9	0.01	0.0029	-0.0534	-0.0003	0.0080	0.0111	N.A.
600	60	0.0	0.9	1.0	0.0286	-0.0534	-0.0310	0.8005	0.0111	0.0331
600	60	0.0	0.9	100.0	0.2864	-0.0534	-3.1040	80.052	0.0111	331.09
600	60	100.0	0.9	0.01	0.0029	-0.0536	-0.0003	0.0080	0.0112	N.A.
600	60	100.0	0.9	1.0	0.0286	-0.0536	-0.0313	0.8016	0.0112	0.0331
600	60	100.0	0.9	100.0	0.2863	-0.0535	-3.1132	80.100	0.0111	331.07
50	60	0.0	0.0	1.0	0.0241	-0.0025	-0.0320	0.0151	0.0125	0.0281
50	60	0.0	0.6	1.0	0.0502	-0.0334	-0.0330	0.0899	0.0101	0.0288
50	60	0.0	0.7	1.0	0.0574	-0.0405	-0.0344	0.1508	0.0095	0.0289
50	60	0.0	0.8	1.0	0.0597	-0.0472	-0.0362	0.2877	0.0087	0.0290
50	60	0.0	0.9	1.0	0.1057	-0.0502	-0.0401	0.7238	0.0085	0.0294
50	60	0.0	0.95	1.0	0.1284	-0.0507	-0.0422	1.211	0.0085	0.0292
50	60	0.0	0.99	1.0	0.0398	-0.0531	-0.0409	1.048	0.0061	0.0287
50	60	0.0	0.999	1.0	-0.0093	-0.0392	-0.0415	0.9667	0.0058	0.0285
50	120	0.0	0.0	1.0	0.0164	-0.0155	0.0094	0.0071	0.0051	0.0193
50	120	0.0	0.6	1.0	0.0438	-0.0318	0.0078	0.0427	0.0050	0.0194
50	120	0.0	0.7	1.0	0.0492	-0.0346	0.0074	0.0734	0.0046	0.0195
50	120	0.0	0.8	1.0	0.0877	-0.0365	0.0067	0.1508	0.0040	0.0197
50	120	0.0	0.9	1.0	0.1850	-0.0346	0.0054	0.4666	0.0033	0.0198
50	120	0.0	0.95	1.0	0.2246	-0.0312	0.0039	0.8872	0.0026	0.0193
50	120	0.0	0.99	1.0	0.2896	-0.0210	0.0037	1.4058	0.0016	0.0188
50	120	0.0	0.999	1.0	0.2217	-0.0158	0.0043	1.2441	0.0012	0.0188
50	180	0.0	0.0	1.0	0.0084	-0.0269	0.0132	0.0058	0.0051	0.0113
50	180	0.0	0.7	1.0	0.0307	-0.0327	0.0117	0.0635	0.0040	0.0111
50	180	0.0	0.9	1.0	0.0924	-0.0282	0.0109	0.4328	0.0024	0.0114
50	180	0.0	0.99	1.0	0.1063	-0.0118	0.0127	1.3361	0.0006	0.0114

processor unit seconds on the Control Data Corporation 6500 computer system.

The Monte Carlo study reported in Table 3.3 required nonlinear parameter estimation because $\emptyset$ was estimated simultaneously with β using maximum likelihood estimation. The low cost in terms of central processor time for this study was achieved by the simplicity of Equation (3.17) and the effectiveness of the Newton-Raphson estimator. The Newton-Raphson estimator used in the study is

$$\mu^{(i+1)} = \mu^{(i)} - h\, R\, g$$

where i is the iteration index, μ is the vector of parameters, h is the step size, R is a positive definite approximation to the inverse of the Hessian matrix, and g is the gradient vector. The matrix R was obtained by finding the eigenvalues and vectors of the Hessian matrix by the QL method in Martin et al. (1968) and Dubrulle (1970), and a modified R matrix formed by scaled decomposition and Newton-Greenstadt replacement of the eigenvalues as suggested by Bard (1974, p. 307). The QL method solved the eigenproblem in Rosser et al. (1951) while the Jacobi method in Boothroyd (1968) that was used initially did not. The step size adjustment used was one additional try version-b in Bard (1970). Bard (1974, p. 114) stated a convergence criterion as

$$\text{abs}(\mu^{(i+1)} - \mu^{(i)}) \leq \gamma\, \text{abs}(\mu^{(i)} + \delta) \qquad (3.18)$$

which insures that the search terminates when each parameter ceases to change value. The convergence constants are $\gamma = 10^{-5}$ and $\delta = 10^{-4}$.

The estimation of the noncentrality parameter associated with Equation (3.17) was considered for the parameter values $\beta = 0$, $\emptyset = 0.94$, $\sigma^2 = 1$, $n = 180$, and $t = 1000$ trials. The estimator used is Equation (3.16.A) and the estimate is $\lambda = 0.3$. Although only 1000 trials were available instead of the 10,000 trials suggested by Dickey and Fuller (1976), empirical percentiles were computed by ordering the estimates of u_i computed by Equation (3.15). These percentiles yielded good agreement with the confidence coefficients when $\lambda = 0.985$ as shown in Table 3.4. Perhaps an estimator of λ given in Pandey and Rahman (1971) valid for small sample sizes should have been used.

Table 3.4

Confidence Coefficients for Equation (3.17)

Confidence Coefficient of Interest	Empirical Percentile	Predicted Confidence Coefficient at Empirical Percentile	
$1-\alpha$	$F(1-\alpha)$	$\lambda = 0.985$	$\lambda = 0$
0.1	0.0329	0.079	0.168
0.5	0.662	0.367	0.578
0.9	4.304	0.868	0.963
0.95	6.879	0.950	0.991
0.975	9.68	0.981	0.998
0.99	14.19	0.995	0.9997

3.4 Aitken's Least Squares Estimator

The purpose of this section is to demonstrate that Aitken's least squares can be implemented within a computer code written to compute standard least squares estimates.

There are advantages in using an existing standard least squares (denoted SLS) computer code rather than programming a new computer code for Aitken's least squares (denoted ALS). First, an existing SLS called PROPTY was available at Michigan State University that estimates thermal parameters in the heat conduction problem. Second, PROPTY was already programmed to provide information that is useful to understand the parameter estimation problem; namely, at each measurement time it prints the two sensitivity coefficients evaluated at the estimated physical parameters, the running difference between the estimates of the physical parameters based on all the measurements and their value using only data up to the current measurement time, and the value of a design criterion in Equation (4.1). Third, PROPTY uses a solution technique and minimization routine that is effective for the heat conduction problem; e.g., it uses special rules for the Gauss step size.

3.4.1 Adjustment of SLS Computer Code to Obtain ALS Estimates

The essential features of computer code PROPTY are the SLS cost function and the Gauss procedure for minimizing the cost function.

The SLS cost function for computer program PROPTY is given in Beck (1964) as

$$f(k,c) = \sum_{j=1}^{m} \sum_{i=1}^{n} A_j \left(T(x_j,t_i) - T_e(x_j,t_i)\right)^2 \qquad (3.19.A)$$

In Equation (3.19.A), the temperature computed by the Crank-Nicolson finite difference scheme is denoted T(x,t), the measured temperature

discussed in Section 2.1.2 is denoted $T_e(x,t)$, and the temperatures are both computed and measured at locations x_j and at times t_i. The residuals are defined by

$$e_i^{(j)} = T(x_j,t_i) - T_e(x_j,t_i) \tag{3.20.A}$$

where the vector e is obtained by stacking as was done for Equation (2.8). An abbreviation is introduced for the double summation used in Equation (3.19.A) so that Equation (3.19.A) can be written compactly as

$$f(k,c) = (e,e) \tag{3.19.B}$$

which implies that the product of the two quantities enclosed in the parentheses on the right hand side of Equation (3.19.B) are multiplied at corresponding times and sensors and the sum computed.

The Gauss procedure used in PROPTY is well-known. In the Gauss procedure Equation (3.19) is minimized by calculating the thermal parameters by the iterative scheme written as

$$\begin{bmatrix} k \\ c \end{bmatrix}^{(i+1)} = \begin{bmatrix} k \\ c \end{bmatrix}^{(i)} + h \begin{bmatrix} (T_k,T_k) & (T_k,T_c) \\ (T_k,T_k) & (T_c,T_c) \end{bmatrix}^{-1} \begin{bmatrix} (T_k,e) \\ (T_c,e) \end{bmatrix} \tag{3.21}$$

In Equation (3.20) the notation defined in Equation (3.19.B) is used. The superscripts denote the i-th and (i+1)-th iteration. The two partial derivatives of the temperature are called sensitivity coefficients. The rules for the step size h are as follows. Rules on the step size h were introduced to improve the convergence rate of Equation (3.21). The step size is unity unless the percentage change

in the parameter value is too large. The two alternate step sizes when the change in the parameter values in an iteration is too large are: (1) $h = 0.6\ k\ (\text{abs}\ \delta k)^{-1}$ when either $k^{-1}\ \delta k$ exceeds 1.6 with δk positive or $k^{-1}\ \delta k$ exceeds 0.6 with δk negative, and (2) $h = 0.6\ c\ (\text{abs}\ \delta c)^{-1}$ when $c^{-1}\ \delta c$ exceeds 1.6 and δc is positive. The starting values of k and c are computed from a table of temperature versus thermal properties supplied by the user of computer code PROPTY. The convergence criterion is Equation (3.18) with $\gamma = 0.01$ and $\delta = 0$. After the convergence criterion is satisfied one more iteration is performed so that the useful information discussed in Section 3.2 can be printed.

In order to use the ALS procedure two computational procedures are required: (1) a procedure to convert estimation with a SLS cost function given by Equation (3.19.A) to one with an ALS cost function given by Equation (1.14.B), and (2) a procedure to estimate the transformation matrix P. Estimation of the P matrix will be discussed in Section 3.4.2 for a test problem. The conversion of the SLS code in PROPTY to an ALS code will be considered next.

For convenience in notation consider a linear model written as

$$y = X\ \beta + P\ a \tag{3.22}$$

where y is the vector of measurements, X is a known matrix, β is the estimated parameter vector, a is a vector of white noise, and P is a matrix presented in Equation (2.7). Comparing Equation (3.22) with Equation (3.1.A) we make the association

$$w = P\ a. \tag{3.23}$$

By definition

$$a \approx N(0,\sigma^2 I) \tag{3.24}$$

which implies

$$\text{cov } (w) = E(P\ a\ a^t\ P^t) = \sigma^2\ P\ P^t. \tag{3.25.A}$$

By defining a matrix V, Equation (3.25.A) can be written as

$$\text{cov } (w) = \sigma^2\ V \tag{3.25.B}$$

where

$$V = P\ P^t. \tag{3.26}$$

Hence, the covariance matrix of the additive errors can be written as

$$W = \sigma^2\ P\ P^t\ . \tag{3.25.C}$$

Equation (3.25.C) contains the covariance matrix that should be used in Equation (1.14.B) to yield Aitken's least squares estimates.

Hence, the quadratic form that should be minimized is

$$f_{ALS}(\beta) = \sigma^{-2}\ (y - X\ \beta)^t\ (P\ P^t)^{-1}\ (y - X\ \beta)\ . \tag{3.26.A}$$

But Equation (3.26.A) can be factored and written as

$$f_{ALS}(\beta) = \sigma^{-2}\ (P^{-1}\ (y - X\ \beta)\ )^t\ (P^{-1}\ (y - X\ \beta)\ ) \tag{3.26.B}$$

For the linear model the quantity corresponding to the residuals in Equation (3.20) is the residual written as

$$e = y - X\ b \tag{3.20.B}$$

Hence, Equation (3.26.B) can be written as

$$f_{ALS}(\beta) = \sigma^{-2}\ (P^{-1}\ e)^t\ (P^{-1}\ e), \tag{3.26.C}$$

while it is clear that for a linear model like Equation (3.22) it is possible to have Equation (3.19.A) written as

$$f(\beta) = A_1 \, e^t \, e. \qquad (3.19.C)$$

Thus, the cost function in computer code PROPTY given by Equation (3.19.C) can be used to compute the Aitken's least squares cost function given by Equation (3.26.C) by merely replacing the residual vector e in PROPTY by the transformed residual P^{-1} e. This transformation was indicated but not stated explicitly in obtaining Equation (1.14.C).

In converting computer code PROPTY into a procedure to obtain Aitken's least squares estimates, the changes in the computer code are made for the Gauss minimization procedure given as Equation (3.21) rather than for the cost function given as Equation (3.19). Using Equation (3.20.B), the transformed residuals can be written as

$$P^{-1} \, e = P^{-1} \, y - P^{-1} \, X \, b. \qquad (3.27)$$

For the linear model in Equation (3.22), the sensitivity coefficient is the X matrix. In Equation (3.27) three quantities are transformed: the residuals (e), the measurements (y), and the sensitivity coefficient (X). However, for the Gauss scheme in Equation (3.21) that is used by computer code PROPTY it is only necessary to adjust two quantities: the residuals (e) and the sensitivity coefficients for each parameter (T_k and T_c).

3.4.2 Demonstration of ALS Estimation for ARIMA(1,0,0) Errors

For real problems the error covariance matrix is unknown; thus, a procedure is needed for estimating the P matrix used in Section 3.4.1. This section considers Aitken's estimation with a particular P matrix.

The P matrix for ARIMA(1,0,0) errors can be written as

$$P^{-1} = I - \phi K \tag{3.28}$$

where the non-zero entries of the matrix K satisfy

$$(K)_{ij} = 1 \quad \text{if} \quad i = j + 1, \tag{3.29}$$

where ϕ is the autoregressive coefficient, and I is the identity matrix. For this P matrix the Yule-Walker estimator given as Equation (A.23) can be written as

$$\phi = ((K\, e)^t K\, e)^{-1} e^t K\, e \tag{3.30.A}$$

$$\sigma^2 = (m\ n)^{-1} (P^{-1} e)^t P^{-1} e. \tag{3.30.B}$$

Because the matrix P depends on the coefficient ϕ, in each iteration of the ALS procedure the estimate of P is updated.

The feasibility of converting computer code PROPTY to a code yielding ALS estimates was investigated for ARIMA(1,0,0) errors. A subroutine was written to estimate the coefficients in the ARIMA (1,0,0) model using Equations (3.30). Within the subroutine a convergence criterion based on Equation (3.18) was used so that overall convergence in PROPTY can only occur after the estimates of the ARIMA(1,0,0) coefficients ϕ and σ^2 have converged. Call statements were inserted within PROPTY so that this subroutine was called

immediately after either the residuals (e) or the sensitivity coefficients (T_k) and (T_c) are computed. The subroutine resets these quantities according to Equation (3.27) and returns the reset value to PROPTY; the subroutine also uses the residuals before they are reset to estimate the ARIMA(1,0,0) coefficients using Equation (3.30).

The estimates obtained by the above scheme for case-0 data are shown in Table 3.5. In three iterations the thermal parameters k and c satisfied the convergence criterion set at $\gamma = 0.01$ in Equation (3.18). However, there were additional iterations because the parameter ϕ has not converged. The convergence criterion for ϕ was set at $\gamma = 0.0001$ in Equation (3.18) and this relatively tight criterion was used in order to examine the parameter estimation routine. Clearly, a larger value of γ can be used for ϕ in the estimation routine. The convergence shown in Table 3.5 clearly demonstrated the feasibility of an ALS estimation procedure.

Table 3.5

Convergence of Aitken's Estimator for Case-0 Data

Iteration	k Btu $hr^{-1}ft^{-1}$ F	c Btu ft^{-3} F^{-1}	ϕ	σ^2 F^2
0	42.00000	55.00000	0.0000	N.A.
1	43.31882	55.65817	0.9233	0.01962
2	42.97964	55.73655	0.9057	0.01958
3	43.05785	55.79758	0.9061	0.01959
4	43.05245	55.84707	0.9058	0.01959
5	43.05034	55.88694	0.9059	0.01959
6	43.04751	55.91910	0.9060	0.01955
SLS	43.35722	55.66701	0.8982	0.02022

The last row of Table 3.5 gives the standard least squares estimates for this same case-0 data. The parameter values are in close agreement: the difference in estimates divided by the ALS estimates are $k^{-1}\ \delta k = 0.0072$ and $c^{-1}\ \delta c = 0.0045$. Also the estimates for the ARIMA(1,0,0) coefficient are given. These were calculated from the SLS residuals and also are in close agreement with $\phi^{-1}\ \delta\phi = 0.0086$.

3.4.3 Aitken's Least Squares with ARIMA Errors

The estimation procedure demonstrated in Section 3.4.2 for an ARIMA(1,0,0) model can be generalized to estimation with an ARIMA (p,d,q) model. The basic requirement is to define the P matrix in the representation of the errors in Equation (2.7); i.e., $w = P\ a$. Basically the same equations for this representation were obtained by Beck and Arnold (1974) and Pagan (1974). The equations in Pagan (1974) were attributed to an unpublished paper in 1966 by Phillips.

The general form of the transformation matrix P will be considered for an ARIMA(s,0,0) and ARIMA(0,0,s) process in order to show the symmetry between the autoregressive and the moving-average process. Also by considering both models together the space required to present the notation is also shortened. After the notation is defined the transformation matrix P for the ARIMA(p,0,q) model will be stated.

Consider the ARIMA(s,0,0) process which can be written as

$$w_i = \gamma_1 w_{i-1} + \gamma_2 w_{i-2} + \cdots + \gamma_s w_{i-s} + a_i$$

and also consider the ARIMA(0,0,s) process that can be written as

$$w_i = a_i - \gamma_1 a_{i-1} - \ldots - \gamma_s a_{i-s} .$$

For a sample of size n, the following vectors are defined as

$$w = (w_1, w_2, \ldots, w_n)^t,$$

$$a = (a_1, a_2, \ldots, a_n)^t,$$

that have the initial values

$$w^* = (w_0^*, w_1^*, \ldots, w_s^*)^t, \text{ and}$$

$$a^* = (a_0^*, a_1^*, \ldots, a_s^*)^t$$

where in the initial values

$$w_{-1} = w_1^* \qquad \text{for } 0 \leq i, \text{ and}$$

$$a_{-1} = a_1^* \qquad \text{for } 0 \leq i.$$

In the matrix notation the ARIMA(s,0,0) process can be written

$$D^{-1} w - D^* w^* = a,$$

and for the ARIMA(0,0,s) process the model can be written

$$w = D^{-1} a - D^* a^*.$$

In both equations the matrix D^{-1} is an n by n lower band matrix written as

$$D^{-1} = \begin{bmatrix} 1 & & & & & & \\ -\gamma_1 & & & & & & \\ -\gamma_2 & -\gamma_1 & 1 & & & & \\ \ldots & & & & & & \\ -\gamma_s & -\gamma_{s-1} & \ldots & 1 & & & \\ 0 & -\gamma_s & \ldots & & 1 & & \\ \ldots & \ldots & \ldots & & & & \\ 0 & \ldots & -\gamma_s & \ldots & -\gamma_1 & 1 \end{bmatrix}$$

and D^* is an n by s matrix containing a band triangular structure written as

$$D^* = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 & \cdots & \gamma_s \\ \gamma_2 & \gamma_3 & & \gamma_s & 0 \\ \gamma_3 & & & & \\ \cdots & & & & \\ \gamma_s & & & & \\ 0 & & & & \\ \cdots & & & \cdots & \\ 0 & \cdots & & & 0 \end{bmatrix}$$

In indicial notation the n by n matrix D^{-1} has non-zero elements given by

$$(D^{-1})_{ij} = -\gamma_{i-j} \quad \text{for} \quad 0 \leq i - j \leq s,$$

where

$$\gamma_0 = -1,$$

while the n by s matrix D^* has non-zero elements given by

$$(D^*)_{ij} = \gamma_{i+j-1} \quad \text{for} \quad j + i - 1 \leq s.$$

The notation for the D and D^* matrices displays the symmetry between the ARIMA(s,0,0) and ARIMA(0,0,s) models.

A notation where the difference between the ARIMA(s,0,0) model and the ARIMA(0,0,s) models is as follows. For the ARIMA(p,0,0) process replace γ by $\emptyset$, D by L, and D^* by L^* so that the process is written as

$$L^{-1} w = L^* w^* + a.$$

For the ARIMA(0,0,q) process replace γ by θ, D by M, and D^* by M^* so that the process is written as

$$w = M^{-1} a - M^* a^*.$$

It is easy to show that the ARIMA(p,0,q) process is written as

$$L^{-1} w - L^* w^* = M^{-1} a - M^* a^* \tag{3.31}$$

Equation (3.31) is the matrix definition of the ARIMA(p,0,q) process.

Two forms of Equation (3.31) will be considered. The remainder of this section will discuss the case when both w^* and a^* are set to zero. Section 3.4.4 will discuss the form when both w^* and a^* are estimated.

When w^* and a^* are zero, Equation (3.31) can be written as

$$L^{-1} w = M^{-1} a \tag{3.32}$$

Equation (3.32) can be rearranged and written as

$$w = L M^{-1} a \tag{3.33}$$

By introducing the definition

$$P = L M^{-1} \tag{3.34}$$

Equation (3.33) has the form of Equation (3.23) as required for adjustment of PROPTY to obtain ALS estimates as was shown in Section 3.4.1. Equation (3.33) was stated in both Pagan (1974) and Beck and Arnold (1974).

In order to obtain a numerical algorithm, recall that both transformed residuals and sensitivity coefficients are used in the ALS computer code. Consider the residuals being transformed into a vector denoted Z according to the relation

$$Z = P^{-1} e \tag{3.35.A}$$

Using Equation (3.34), Equation (3.35.A) can be written as

$$Z = M L^{-1} e. \tag{3.35.B}$$

For evaluation by a computer, it was convenient to convert Equation (3.35) to a two stage recursive relation. This relation can be written as

$$(M^{-1} Z)_i = e_i - \phi_1 e_{i-1} - \ldots - \phi_p e_{i-p} \tag{3.36.A}$$

$$Z_i = (M^{-1}Z)_i + \theta_1(M^{-1} Z)_{i-1} - \ldots - \theta_q (M^{-1}Z)_{i-q}. \tag{3.36.B}$$

Equation (3.36) clearly reduces to the transformation used in Equation (1.14.C) for ARIMA(p,0,0) errors when M becomes I and all the θ's are zero.

3.4.4 Approximating Maximum Likelihood by Aitken's Estimator

It is more difficult to implement the maximum likelihood estimator for which w^* and a^* are estimated rather than being set to zero. Newbold (1974) presented a clear computational procedure for the ARIMA(1,0,1) process and maximum likelihood estimation.

The positive log-likelihood function can be written as

$$F(\phi,\theta,\sigma^2,w_0,a_0) = \tfrac{1}{2} n \ln(\sigma^2) + \tfrac{1}{2} \det(I + B^t B) + \varepsilon$$

where

$$\varepsilon = \tfrac{1}{2} \sigma^{-2} a^t a + \tfrac{1}{2} \sigma^{-2} \delta$$

The initial values of the w_i-series is w_0, and the initial value of the a_i-series is a_0. The n by 2 matrix B has components written as

$$(B)_{i1} = (\theta - \phi)\ \theta^{i-1}$$

$$(B)_{i2} = -\ \phi\ (\theta - \phi)\ (1 - \phi^2)^{-\frac{1}{2}}\ \theta^{i-1}$$

The n components of the vector a can be written as

$$a_i = A_i + (a_0\ \theta - \phi\ w_0)\ \theta^{i-1}$$

where Z_i is computed from a two stage recursive relation of the form of Equation (3.36.A) and (3.36.B). This relation can be written as

$$(M^{-1}\ Z)_i = w_i - \phi\ w_{i-1} \qquad \text{for } 2 \leq i \tag{3.37.A}$$

$$Z_i = (M^{-1}\ Z)_i + \theta\ (M^{-1}\ Z)_{i-1} \tag{3.37.B}$$

with the initial values

$$Z_1 = w_1$$

$$(M^{-1}\ Z)_1 = w_1.$$

The scalar δ arising from the initial conditions is

$$\delta = a_0^2 + (1 - \phi^2)\ (\theta - \phi)^2\ (w_0 - a_0)^2\ .$$

The values of a_0 and w_0 are estimated in terms of a two component vector c that can be written as

$$c = (I + B^t\ B)^{-1}\ B^t\ M\ L^{-1}\ w \tag{3.38}$$

with the estimates written as

$$a_0 = -\ c_1, \text{ and} \tag{3.39.A}$$

$$w_0 = c_1 - c_2\ (\theta - \phi)\ (1 - \phi^2)^{-\frac{1}{2}}. \tag{3.39.B}$$

This model reduces correctly since a_0 and w_0 equal to zero implies both δ and c are zero and hence B is zero.

Equation (3.39) was computed for case-0 data in order to gain insight into the magnitudes of a_0 and w_0. The vector c in Equation (3.38) used w from the vector of standard least squares residuals, and used matrices B, M, and L evaluated using the parameter estimates obtained in Table 2.11. The use of the standard least squares residuals for the w vector is a reasonable first approximation. The resulting estimates for the eight sensors in the case-0 heat conduction data are shown in Table 3.6. The initial values a_0 and w_0 can be set to zero, because each w_0 is close to the first value of w_i for the corresponding sensor plotted in Figures 2.1 and 2.11 and because the value of a_0 is small with respect to the estimate of σ equal 0.13 given in Table 2.13. Hence Aitken's least squares can be used instead of maximum likelihood.

Table 3.6

Estimates of ARIMA(1,0,1) Starting Values

Sensor	w_0	a_0
1	-0.0032	-0.000291
2	-0.3996	-0.036076
3	-0.1319	-0.011910
4	-0.0029	-0.000266
5	0.1329	-0.000291
6	0.0644	0.005811
7	0.2163	0.019524
8	0.4605	0.041574

3.5 Aitken's Least Squares Simulation Study

The purpose of this section is to display and interpret the results of a twenty trial Monte Carlo study the author made on Aitken's estimates for simulated heat conduction data. This study investigates the first and fifth aspect of the correlated error problem stated in Section 1.5; namely, estimation of the ARIMA coefficients and the modeling of the probability distribution of the confidence ellipse.

With good starting values for k and c, computer program PROPTY converges in usually two iterations and at most three iterations. The central processor time per iteration is quite large since the heat conduction equation is solved numerically. A close approximation to the central processor time in seconds as a function of the number of iterations I is

$$t = 13.9 + 6.3\ I.$$

The minimum central processor time is for one iteration or twenty seconds per trial instead of the 0.18 seconds per trial required in the simple linear study reported in Section 3.3.2. Hence it is impractical to have a detailed study of the statistical properties because there is a one hundred-fold increase in the computer time.

The simulation study that was conducted is as follows. The simulation of the experimental data is discussed in Appendix C. The conditions used in the study approximated the conditions estimated for case-0 data in Table 2.1. The conditions are: $k = 43$ Btu hr^{-1} ft^{-1} F^{-1}, $c = 55$ Btu ft^{-3} F^{-1}, $\phi = 0.98$, $\theta = 0.45$, $\sigma^2 = 0.016\ F^2$, $n = 89$, $m = 8$, $q = 2.67$ Btu ft^{-2} sec^{-1}, and $t_E - t_B = 15.3$ seconds.

Although only twenty trials were used, approximately 727 central processor seconds were required to estimate the parameters on the Control Data Corporation 6500 computer system. A Monte Carlo study with ten times as many trials would have been prohibitively expensive. Yet the state-of-the-art is such that Chambers and Ertel (1975) used one hundred trials and Pfeiffer and Lichtenwalner (1974) used four hundred trials. Hence the essence of the Monte Carlo problem is to determine whether small trial procedures can yield information on the estimates of the coefficients and the probability distribution for the fractile that determines the confidence ellipse.

3.5.1 Estimates of Thermal and ARIMA Parameters

The quantities computed for Table 3.7 are obvious. In Table 3.16 the averages of the physical parameters estimates are both remarkably close to the true values of the thermal parameters. The estimate of $\emptyset$ is only slightly biased downward. The estimate of θ is also biased toward zero, but is approximately thirteen percent in error. The agreement of the estimated parameters displayed in Table 3.7 is acceptable.

The last column in Table 3.7 is discussed in Section 3.5.3.

3.5.2 Estimates of the Information Matrix

Equation (1.21) showed that in some cases the estimation of the V matrix can increase the estimates of the variance of the physical parameters even for large sample sizes. Kenward (1976) proved that the asymptotic variance matrix has the information matrix as a limit

Table 3.7

Heat Conduction Simulation Study with ARIMA(1,0,1) Errors

Trial	k Btu $hr^{-1}ft^{-1}F^{-1}$	c Btu $ft^{-3}F^{-1}$	ϕ	θ	σ^2 F^2	u
1	43.20201	54.42676	0.9663	0.4605	0.01881	0.4637
2	42.43270	55.39035	0.9536	0.4124	0.01909	0.3287
3	41.75756	55.10096	0.9520	0.3801	0.02215	1.026
4	41.79732	55.69786	0.9356	0.4348	0.01694	3.078
5	43.52259	55.20628	0.9397	0.4093	0.01627	0.5483
6	43.55222	54.69688	0.9747	0.4143	0.01879	0.1744
7	42.72592	54.98493	0.9721	0.5196	0.01982	0.04155
8	42.12791	54.98691	0.9469	0.2837	0.02056	0.5472
9	43.62433	54.24719	0.9568	0.4241	0.01927	1.149
10	42.18312	54.53713	0.9620	0.4688	0.01875	1.015
11	43.56209	55.33860	0.9721	0.3427	0.02259	0.2379
12	42.80859	54.12595	0.9812	0.5287	0.01906	0.7095
13	44.14770	54.06201	0.9628	0.3820	0.02111	1.484
14	44.70050	55.10287	0.9266	0.3021	0.02129	2.753
15	42.03968	55.66083	0.9602	0.3540	0.01863	0.9640
16	44.80360	54.56801	0.9487	0.3209	0.01941	2.050
17	43.64540	56.02346	0.9454	0.3471	0.02168	2.629
18	42.59165	54.70477	0.9630	0.4684	0.01843	0.3116
19	43.45775	55.05160	0.9547	0.2377	0.01998	0.1184
20	45.17092	54.56628	0.9522	0.3455	0.02110	2.473
Mean	43.19268	54.91898	0.9563	0.3918	0.01969	1.105
s.d.	1.008	0.5337	0.0138	0.0768	0.00164	0.974
True	43.	55.	0.98	0.45	0.016	

when the errors are ARIMA. The information matrix for the positive log-likelihood function $f(\beta,\emptyset,\theta,\sigma^2)$ denoted $f(\mu)$ can be written as

$$J = n^{-1} \frac{\partial^2 f(\mu)}{\partial \mu_i \partial \mu_j} .$$

Kenward (1976) used a central limit theorem for dependent random variables to establish that the distribution about the true value of μ denoted $\mu^{\#}$ is given as

$$n^{\frac{1}{2}} (\mu - \mu^{\#}) \approx N(0,J^{-1}). \tag{3.40}$$

The information matrix J can be written as

$$J = \begin{bmatrix} C & 0 & 0 & 0 \\ 0 & F & E & 0 \\ 0 & E & T & 0 \\ 0 & 0 & 0 & \frac{1}{2}\sigma^{-4} \end{bmatrix} \tag{3.41}$$

For the linear model with ARIMA(p,0,q) errors in Equation (3.22) and with the explicit form for the P matrix given in Equation (3.33) the C matrix in Equation (3.41) can be written as

$$C = n^{-1} \sigma^{-2} (L^{-1} X)^t M^t M L^{-1} X . \tag{3.42}$$

Pierce (1971) stated consistent estimators for the p by p matrix F, the q by q matrix T, and the q by p matrix E in Equation (3.41) in terms of autocovariances and cross covariances but they are not considered in this dissertation.

Estimates are shown in Table 3.8 for the component of the C matrix from the estimates at each trial based on Equation (3.42). The last row in Table 3.8 has the evaluation of Equation (3.42) at the true parameter values used in the Monte Carlo study. Note that

the components of the C matrix are used to define the standard errors in Equation (1.12). The average value of C_{22} is appreciably larger than the true value which implies a conservative estimate is obtained for the standard error of the thermal conductivity when ALS estimation is used.

The last column in Table 3.8 is discussed in Section 4.4.

Table 3.8

Heat Conduction Study Information Matrix and Sensor Weights

Trial	C_{11}	C_{22}	C_{12}	Range/Mean
1	0.001618	0.004136	0.0005291	0.731
2	0.002127	0.005618	0.0007133	0.523
3	0.001938	0.004937	0.0006442	0.766
4	0.003976	0.010637	0.0013622	0.270
5	0.003185	0.009945	0.0011455	0.707
6	0.001218	0.002679	0.0003817	0.421
7	0.001493	0.003222	0.0004634	0.496
8	0.001999	0.005361	0.0006813	0.572
9	0.001834	0.005490	0.0006367	0.764
10	0.001969	0.004906	0.0006419	0.867
11	0.000964	0.002159	0.0003051	0.749
12	0.001309	0.002374	0.0003904	0.663
13	0.001317	0.003872	0.0004518	0.474
14	0.002524	0.009031	0.0009665	0.631
15	0.001762	0.004130	0.0005677	0.777
16	0.001736	0.005922	0.0006376	0.505
17	0.001877	0.005474	0.0006569	0.809
18	0.001855	0.004735	0.0006152	0.329
19	0.001425	0.004062	0.0004930	0.643
20	0.001488	0.005079	0.0005435	0.637
Mean	0.001882	0.005188	0.0006414	0.617
true	0.001663	0.003443	0.0006559	
s.d.	0.000693	0.002300	0.000257	0.164

3.5.3 Estimates of the Noncentrality Coefficient

The equations used to study the noncentrality coefficient were developed in Section 3.3.

In Table 3.7 the value of the u statistic defined by Equation (3.15) was displayed. If the noncentrality coefficient were equal to zero in Equation (3.16.B), then the average value of u is unity. The estimated average value of 1.1 is close to unity. Support for the noncentrality coefficient being zero is also given from the estimates obtained using Equation (3.16.A) that are also displayed in Table 3.9 with sequential estimation used to find λ because the number of trials is small. Computer program NLINA was used for the sequential estimation.

Because the number of trials is small, the adequacy of the twenty trials is explored by testing for normality of the u_k and u_c values using Kuiper statistics given by Louter and Koerts (1970). Because the statistics in Table 3.9 are accepted as being normal and the estimates of λ are close to zero, the noncentrality coefficient λ is assumed to be zero. Hence, the central F distribution will be used for $G(1-\alpha)$ in Equation (3.5).

3.6 Calculation of the Confidence Ellipses

In Section 3.5, it was shown that the results in Section 3.2 for the V matrix known were a good approximation even though the V matrix was estimated. Specifically, the estimated ARIMA(1,0,1) parameters that define the V matrix are close to the true values and the distribution of $G(1-\alpha)$ in Equation (3.9) is approximated well by a

Table 3.9

Kuiper and Noncentrality Estimates

Trial	Thermal Conductivity			Specific Heat		
	u_k	Kuiper Statistic	Sequential λ	u_c	Kuiper Statistic	Sequential λ
1	0.0470			0.9675		
2	0.4875			0.3373		
3	2.130			0.0358		
4	4.094	0.5677-A*	-0.	3.689	0.6487-R*	-0.
5	0.6195	0.6146-R	-0.	0.3012	0.6530-R	-0.
6	0.2646	0.6429-R	-0.	0.1752	0.5813-R	-0.
7	0.0799	0.6387-R	-0.	0.00052	0.6043-R	-0.
8	1.083	0.5355-R	-0.	0.00065	0.5394-R	-0.
9	0.5092	0.5396-R	-0.	2.215	0.5487-R	-0.
10	0.9356	0.5304-R	-0.	0.7483	0.5385-R	-0.
11	0.2168	0.5145-R	-0.	0.1762	0.4567-R	-0.
12	0.0342	0.4387-R	-0.	1.292	0.3885-R	-0.
13	1.235	0.3901-R	-0.	2.426	0.3718-R	-0.
14	5.197	0.3790-R	-0.	0.0681	0.3255-A	-0.
15	1.157	0.3675-R	-0.	1.284	0.2636-A	-0.
16	4.020	0.3092-A	-0.	0.7867	0.2619-A	-0.
17	0.5569	0.2635-A	0.	4.083	0.2068-A	-0.
18	0.2238	0.2276-A	0.0001	0.2938	0.1577-A	-0.
19	0.2126	0.2040-A	0.1127	0.0077	0.1139-A	0.
20	4.995	0.1810-A	0.2152	0.6804	0.1041-A	0.00006

*A indicates accept normality of the normalized u value at the .01 level of significance using the Kuiper values in Louter and Koerts (1970) and R indicates rejection.

central F-distribution. In this section the matrix C in Equation (3.42) for case-0 data is used to compute confidence ellipses. The matrix C is computed in the modification to computer code PROPTY that makes it yield Aitken's estimates.

Confidence ellipses for the parameter covariance matrices in Equations (3.7) through (3.9) are calculated and plotted for case-0 data. The confidence regions are plotted using the ninety-five percent level. Typically, at ninety-five percent the advantage of being more definite about the parameter value counterbalances the advantage of being more sure about the parameter value.

3.6.1 ALS and SLS Confidence Ellipses

The standard least squares (SLS) and Aitken's least squares (ALS) ellipses at the ninety-five percent level are plotted in Figure 3.1. The confidence ellipse is based on an ARIMA(1,0,1) model for the errors. The numerical estimates for the case-0 data are statistical parameters $\emptyset = 0.9741$, $\theta = 0.4431$, $\sigma^2 = 0.01602$ and thermal parameters $k = 43.00712$ and $c = 55.73395$.

The values in Table 3.10 are obtained from the values of the Q matrix in Equation (3.9) computed by modified computer program PROPTY, and the values in Table 3.11 were obtained from values of the Q matrix in Equation (3.7) computed in the subroutine called by PROPTY that computed the statistical parameters. Specifically, d_1 and d_2 are computed from Equation (3.45), ν is the angle d_1 makes with the k-axis, s^2 is the estimate of σ^2, a_1 and a_2 are the

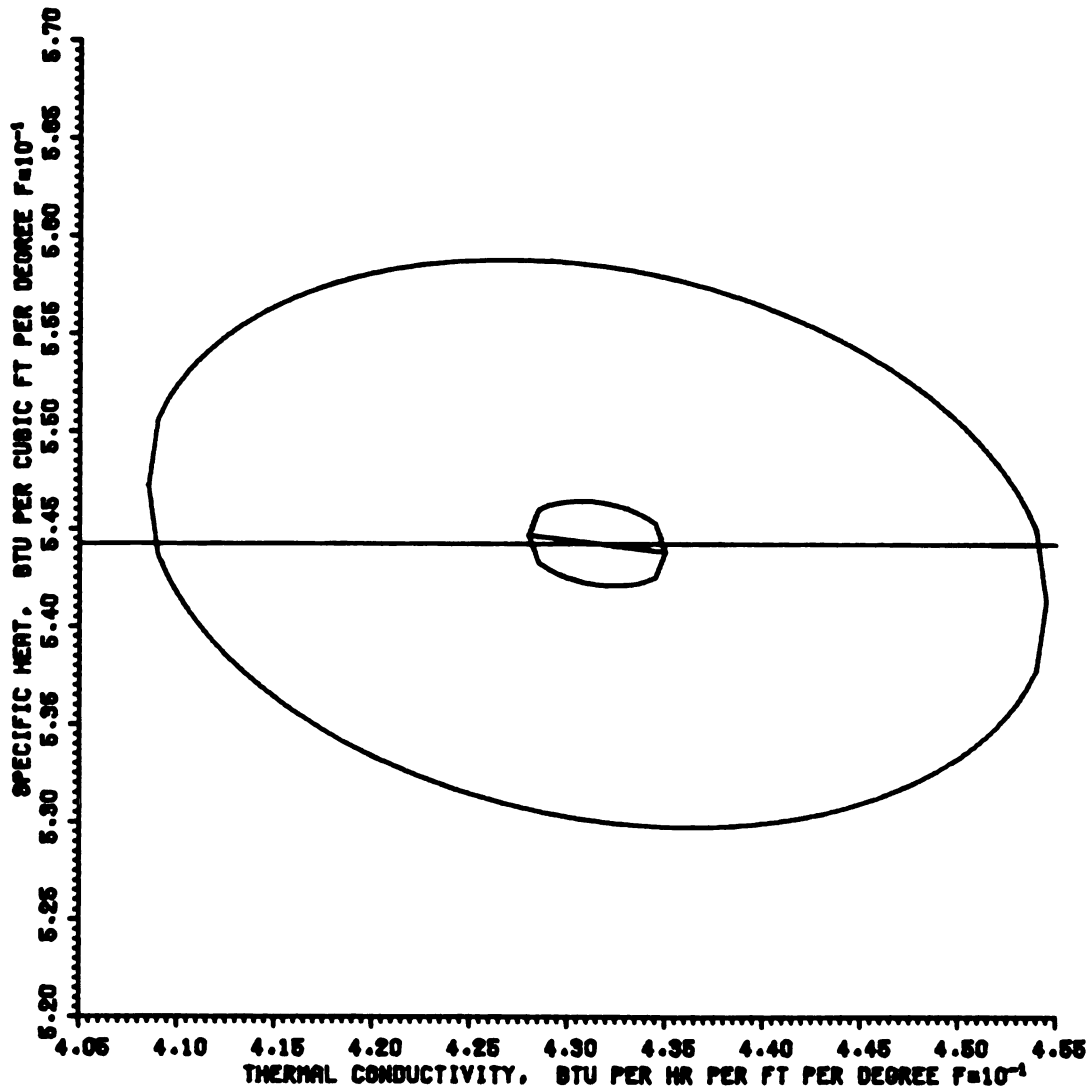

Figure 3.1. Ninety-five percent confidence ellipse for SLS and ALS

semi-length of the ellipse computed from Equation (3.44). The values in Table 3.10 and 3.11 are used to compute the values in Table 3.12.

Table 3.10

Aitken's ARIMA(1,0,1) Ellipse Statistics

Case	d_1	d_2	ν	s^2	a_1	a_2
0	0.0172	0.0413	-15.9	0.0172	2.44	1.58
1	0.0525	0.1111	-23.4	0.0357	2.02	1.39
2	0.0166	0.0323	-49.6	0.0147	2.31	1.65
3	0.0402	0.0792	-40.7	0.0262	1.98	1.41
4	0.0168	0.0323	-49.8	0.0159	2.38	1.72
5	0.0356	0.0774	-52.8	0.0320	2.32	1.57
6	0.0536	0.0992	-47.4	0.0421	2.17	1.60
7	0.0244	0.0423	-42.5	0.0376	3.04	2.31
8	0.0526	0.1032	-54.5	0.0589	2.59	1.85
9	0.0140	0.0319	-61.2	0.0284	3.49	2.31

Table 3.11

Standard Least Squares Ellipse Statistics

Case	d_1	d_2	ν	s^2	a_1	a_2
0	3.34	11.35	-10.3	0.1025	0.429	0.233
1	12.94	38.40	-12.7	0.3109	0.380	0.220
2	4.55	10.90	-17.1	0.0724	0.309	0.199
3	15.46	36.53	-17.6	0.4488	0.417	0.272
4	4.90	9.58	-26.0	0.1005	0.351	0.251
5	16.92	35.04	-22.6	0.7051	0.500	0.347
6	18.63	34.51	-30.7	0.5094	0.405	0.298
7	4.95	8.89	-35.4	0.2424	0.542	0.404
8	19.22	34.04	-40.5	0.7014	0.468	0.352
9	4.87	8.70	-45.8	0.2136	0.513	0.384

Table 3.12

Correct SLS Confidence Coefficients

Case	$F(2,670,1-\alpha^{\#})$	$1-\alpha^{\#}$
0	0.0778	0.083
1	0.0893	0.093
2	0.0484	0.056
3	0.122	0.119
4	0.0646	0.072
5	0.1429	0.136
6	0.1043	0.105
7	0.0935	0.096
8	0.1031	0.104
9	0.0733	0.079
avg.		0.094

A numerical comparison of the differences between the Aitken's and standard least squares confidence ellipses is shown in Table 3.12. The conclusion drawn from Table 3.12 is that an SLS confidence ellipse assumed to contain ninety-five percent of the estimates only contains at most fourteen percent and on the average nine percent of the estimates. The importance of the research suggested in Section 1.2 is confirmed since the overly small size of the confidence region when the standard assumptions in Equation (1.5) are used is a serious understatement of the parameter's accuracy.

The computational equations are as follows. Because the area of the confidence ellipse is proportional to the fractile, the correct confidence coefficient $\alpha^{\#}$ for the standard least squares ellipse is defined by

$$F(k,n-k,1-\alpha^{\#}) = (a_1\ a_2)_{SLS}\ (a_1\ a_2)^{-1}_{ALS}\ F(k,n-k,1-\alpha) \qquad (3.43)$$

where the area of each ellipse is proportional to the product of the lengths of the two semi-axes. In Equation (3.43) the length of each semi-axis is computed from the relation

$$a_i = (k\ s^2\ d_i^{-1}\ G(1-\alpha)\ )^{\frac{1}{2}} \tag{3.44}$$

where the i-th eigenvalue is denoted d_i and the corresponding i-th eigenvector used in Equation (3.45) is denoted e_i. The eigen decomposition of the Q matrix in Equations (3.7) through (3.9) is written as

$$Q = d_1\ e_1\ e_1^t + d_2\ e_2\ e_2^t \tag{3.45}$$

The distribution $G(1-\alpha) = F(k,n-k,1-\alpha)$ is used to compute the fractiles for the ellipse and the confidence coefficient used is $1-\alpha = 0.95$.

3.6.2 MLS and SLS Confidence Ellipses

The standard least squares (SLS) and mixed least squares (MLS) ellipses at the ninety-five percent level are plotted in Figure 3.2. The confidence ellipse in Figure 3.2 assumed the errors are for the ARIMA(1,0,0) model. This error model is simple to work with and is as appropriate as a more accurate ARIMA(p,d,q) model for the errors since Equation (B.3) showed that the distribution of the quadratic form for MLS is only approximately central F. A central F was used to compute fractile for the confidence ellipse.

Computational expressions developed by Beck and Arnold (1977, p. 306) facilitated the numerical evaluation of $X^t\ V\ X$ in Equation (3.8). For the MLS confidence ellipse in Figure 3.2 the coefficients

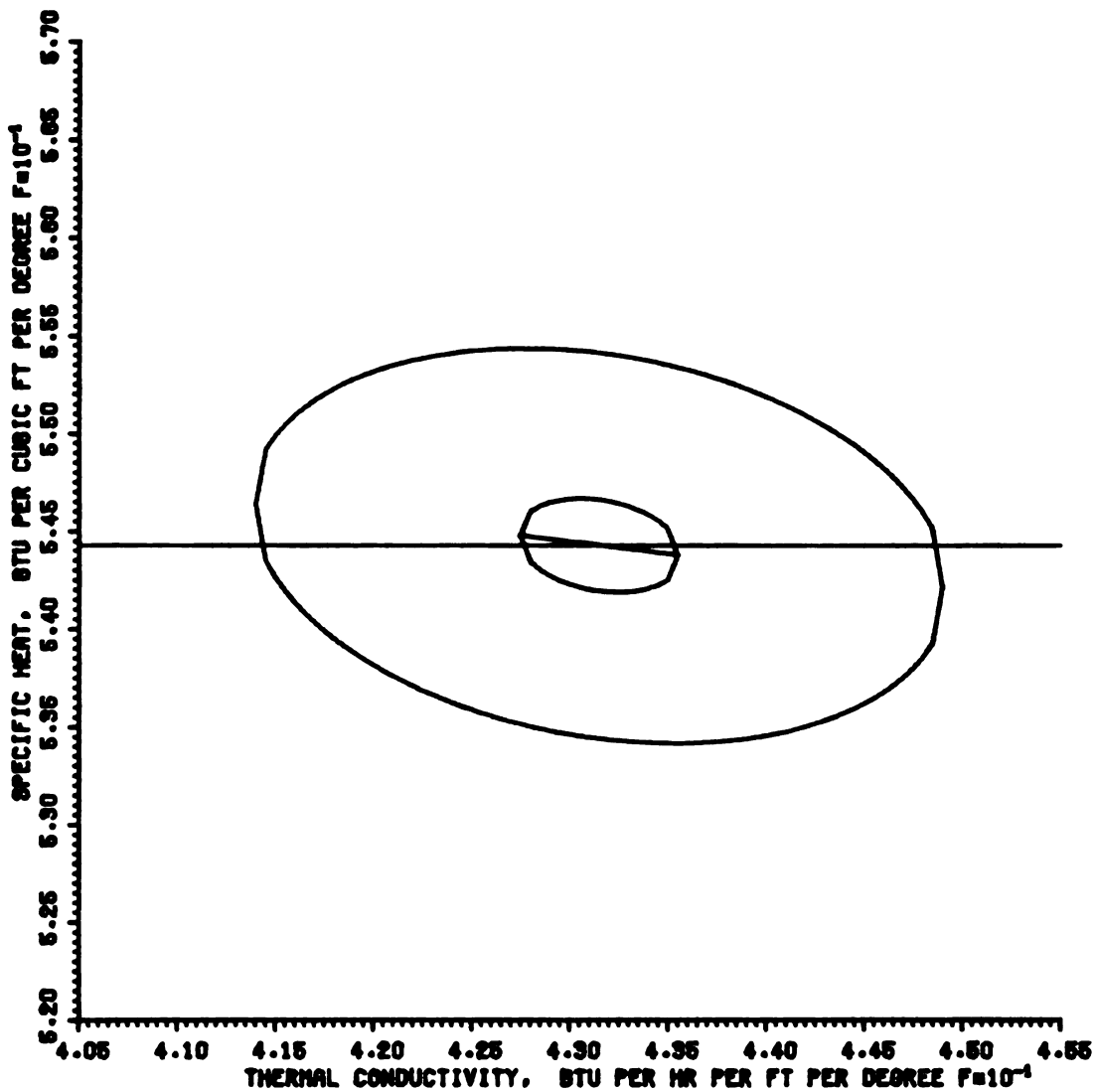

Figure 3.2. Ninety-five percent confidence ellipse for SLS and MLS

in the ARIMA(1,0,0) model are $\emptyset = 0.906$ and $\sigma^2 = 0.0186$ with sample size n = 712. The MLS confidence ellipse is clearly larger. While by comparing Figure 3.2 and Figure 3.1 it is clear that the Aitken's confidence region is only slightly larger than the MLS region. Thus, as a first approximation the standard least squares estimator with coefficients obtained from the residuals could be used. The reasonable nature of the approximation suggests that a central F distribution is reasonable for MLS in Equation (B.3).

A more detailed comparison can be made from the coefficients tabulated in Table 3.13 through Table 3.15 for the plotted ellipses and ellipses calculated using every h-th residual. The tabulated values imply the following behavior: (1) the orientation of all ellipses are independent of both the Q matrix and the sample size, (2) the only change in the relative shapes of the ellipses is that the SLS axes vary inversely with the square root of the sample size as expected from theory while the MLS axes vary inversely with what appears to be the eighth-root of the sample size. The inconsistent value at n = 232 in Table 3.14 is probably an artifact of the non-even integer value of the sampling interval h on the fitted ARIMA (1,0,0) model rather than being an optimum sampling rate.

Table 3.13

Confidence Ellipse Data SLS and MLS

d_1	d_2	ν	s^2	n
		SLS		
11.38	3.33	79.7	0.1026	712
5.64	1.66	79.6	0.0997	352
3.73	1.11	79.4	0.1030	232
2.86	0.84	79.8	0.0968	176
		MLS		
0.115	0.0338	80.2	0.0186	712
0.0994	0.0305	80.0	0.0221	352
0.0806	0.0260	79.4	0.0234	232
0.1027	0.0328	79.8	0.0270	176

Table 3.14

Ninety-Five Percent Ellipse Coefficients

n	SLS		MLS	
	Major Axis	Minor Axis	Major Axis	Minor Axis
712	0.430	0.232	1.817	0.985
352	0.600	0.326	2.085	1.155
232	0.746	0.407	2.324	1.319
176	0.834	0.451	2.222	1.256

Table 3.15

Comparison of Ellipse Coefficients

n	Ratio of Axes		Reduction of Axes	
	Major	Minor	SLS	MLS
712	4.226	4.245	0.515	0.805
352	3.475	3.543	0.720	0.932
232	3.115	3.240	0.897	1.048
176	2.664	2.780	1.000	1.000

CHAPTER IV

FURTHER INVESTIGATION OF THE COVARIANCE MATRIX

In Section 4.1 a brief analysis of real data is given that investigated the effect of the covariance matrix on the optimal duration of the experiment which was suggested in Section 1.1. Recall that in Chapter II the residuals were represented as ARIMA processes rather than as signatures. In Section 4.2 a temperature signature is used to investigate bias in estimates of thermal properties, while in Section 4.3 a heat flux signature is calculated from the residuals and this signature could be fit to an ARIMA model. In Section 4.4 calculations from computer program PROPTY are used to examine the use of the same coefficients in ARIMA models for all eight sensors. In Section 4.5 a discussion is given of possible alternatives to constant ARIMA coefficients. In Section 4.6 ways are suggested to improve the poor ARIMA model discrimination that was observed when Akaike's Information Criterion was used in Chapter II.

4.1 Optimum Experiment Design

The autocorrelation of the errors not only affects the confidence regions discussed in Section 3.6 but also affects the design of the optimum experiment. Beck and Arnold (1977, Equation 8.3.4) related the criterion for the design of the optimum experiment to the covariance matrix for the parameters. The author considered the situation

where the X matrix is specified and the optimum duration of the experiment is found. When standard least squares is incorrectly used the design criterion related to Equation (3.7) can be written as

$$D_{SLS} = i^{-2} X^t X \tag{4.1}$$

For Aitken's least squares the design criterion related to Equation (3.9) can be written as

$$D_{ALS} = i^{-2} X^t V^{-1} X \ . \tag{4.2}$$

In both Equation (4.1) and (4.2) the X matrix has variable row dimension i rather than fixed dimension n used previously. The column dimension of X is two because there are two parameters, k and c.

The optimum duration of the experiment is found by increasing the number of rows in the X matrix until the criterion D reaches a maximum. The value of i at this maximum and the known sampling rate in the experiment were used to compute the corresponding optimum duration of the experiment in seconds. The optimum durations for the conditions used in the simulation study are as follows: for Equation (4.1) the duration is 25.8 seconds, and for Equation (4.2) the duration is 16.8 seconds. Thus, for ARIMA(1,0,1) errors the experiment should end shortly after the imposed heat flux ends rather than having the experiment end at twice the duration of the imposed heat flux.

In Section 2.4.3 the coefficients in the ARIMA(1,0,1) model were shown to change with a change in sampling rate. Thus, the optimum duration of the experiment computed using Equation (4.2) depends

on the sampling rate. A sampling rate of 0.3 seconds between measurements was used in the simulation experiment reported in Section 3.5, although the IBM-1800 minicomputer system is capable of sampling at other rates.

4.2 Signature in the Residuals at the Heated Surface

As emphasized in Section 1.5, the errors in scientific work may have non-zero mean; i.e., the error model can be written as

$$w \approx N(\mu, W)$$

The assumption that μ equals zero yields a model for the error vector w that is convenient for use in constructing confidence regions. However, the residuals at the heated surface (e.g., Figure 2.19) seem to have a signature that can be represented by the following model

$$\mu_j(i) = \nu_j^{(1)} H(i - i_B) + \nu_j^{(2)} H(i - i_E) \qquad (4.3)$$

where $\mu_j(i)$ is the element in the μ vector for the j-th sensor at the i-th time, H is the unit step function, and $\nu_j^{(i)}$ and $\nu_j^{(2)}$ are coefficients appropriate for the j-th sensor.

Perhaps the signature represented by Equation (4.3) could be removed as Carr (1972) did for a problem in econometrics. Chen and Danh (1976) showed that with a step heat flux there is a cavity effect that produced a temperature bias that goes from zero to a steady value in dimensionless time $t^* = 0.5$, which is close to the dimensionless time between 0.35 and 0.45 that the imposed heat flux ends in Farnia's data. An alternate explanation was offered

by Van Fossen (1973, p. 96) that the correlated residuals can be explained as a non-uniform heat flux across the surface that causes three dimensional effects.

A possible bias in the estimate of the thermal parameters is undoubtedly the most serious consequence of assuming μ is zero when it is non-zero. The combined dominant-small parameter analysis by Beck (1970) can be used to estimate the bias in the theraml parameters caused by the non-zero value of μ. By inspecting the residuals plotted in Figures 2.1 through 2.20, it is clear that both $\nu_j^{(1)}$ and $\nu_j^{(2)}$ are zero when j corresponds to a thermocouple at the insulated surface and they have the values shown in Table 4.1 when j corresponds to a thermocouple at the heated surface. Note that the coefficients in Table 4.1 are dependent on the maximum absolute value of the residuals which is measured by the quality factor defined in Table 2.3.

Table 4.1

Initial Estimates of the Non-Zero Mean

Quality	$\nu_5^{(1)}$	$\nu_6^{(1)}$	$\nu_5^{(2)}$	$\nu_6^{(2)}$
A	0.4	0.0	0.0	-0.4
B	1.0	-0.2	-0.2	-0.6
C	1.2	-0.8	0.0	-0.4

An approximation to Equation (4.3) can be written as

$$\mu(i) = \tfrac{1}{2} \nu \, (e_5(i) + e_6(i)) \tag{4.4}$$

where e_5 and e_6 are the residuals at the heated surface at the i-th time. Equation (4.4) has only one parameter ν which is near unity. The temperature measured at the heated surface can be written as

$$T(k,c,\mu) = T(k,c) + \mu \tag{4.5}$$

where μ is given by Equation (4.4) and T(k,c) is the temperature calculated from Equation (2.1). The analysis in Beck (1970) for the two dominant thermal parameters k and c, and the one small parameter ν is as follows. The equations can be written as

$$k^{-1} \delta k = (c_{11} C_{22} - c_{12} C_{12}) (C_{11} C_{22} - C_{12} C_{12})^{-1}$$

$$c^{-1} \delta c = (c_{12} C_{11} - c_{11} C_{12}) (C_{11} C_{22} - C_{12} C_{12})^{-1}$$

where

$$c_{12} = (\mu, c\, T_c) \tag{4.6.A}$$

$$C_{11} = (k\, T_k, k\, T_k) \tag{4.6.B}$$

$$C_{22} = (c\, T_c, c\, T_c) \tag{4.6.C}$$

$$C_{12} = (k\, T_k, c\, T_c) \tag{4.6.D}$$

with the notation developed for Equation (3.19) used for Equations (4.6).

Case-3 data shown in Figure 2.14 have a pronounced signature at the heated surface. The results of a dominant-small parameter analysis for these data yielded the estimates $k^{-1} \delta k$ equals -0.016 while $c^{-1} \delta c$ equals 0.011. The maximum bias in the thermal parameters is

approximately one percent. This represents only a small shift in the ellipses plotted in Figure 3.1 and will be neglected.

4.3 Signature in the Heat Flux at the Insulated Surface

The residuals near the end of the experiment seem to exhibit a pattern of being negative. Negative residuals imply the calculated temperature is greater than the experimental temperature. One possible cause for this apparent pattern is a heat loss from the disk. This could occur because the insulated surface is insulated by an air layer. The signature in case-1 data shown in Figure 2.2 was analyzed by estimating the heat flux at the insulated surface. The estimated heat flux shown in Figure 4.2 has a maximum magnitude that is approximately ten percent of the applied heat flux. No heat flux signature was found.

The estimated heat flux in Figure 4.1 was modeled. The best fitting model was ARIMA(1,0,1) that can be written as

$$z_i = 0.879\, z_{i-1} + a_i - 0.747\, a_{i-1};\ z_i = w_i - w_{i-1}\ .$$

The temperature residuals at the same insulated surface had as the best fit the ARIMA(1,0,1) model given by

$$e_i = 0.974\, e_{i-1} + a_i - 0.513\, a_{i-1}\ .$$

The z_i component of the w_i series for the heat flux and the e_i series for the residuals have similar order and coefficients. This is an interesting apparent connection, because randomness was investigated previously for the heat flux by Ahamadi (1974) and Vakhaniya (1967).

The method used to estimate the heat flux will be outlined. Beck (1968) proposed an integral method for computing the heat flux

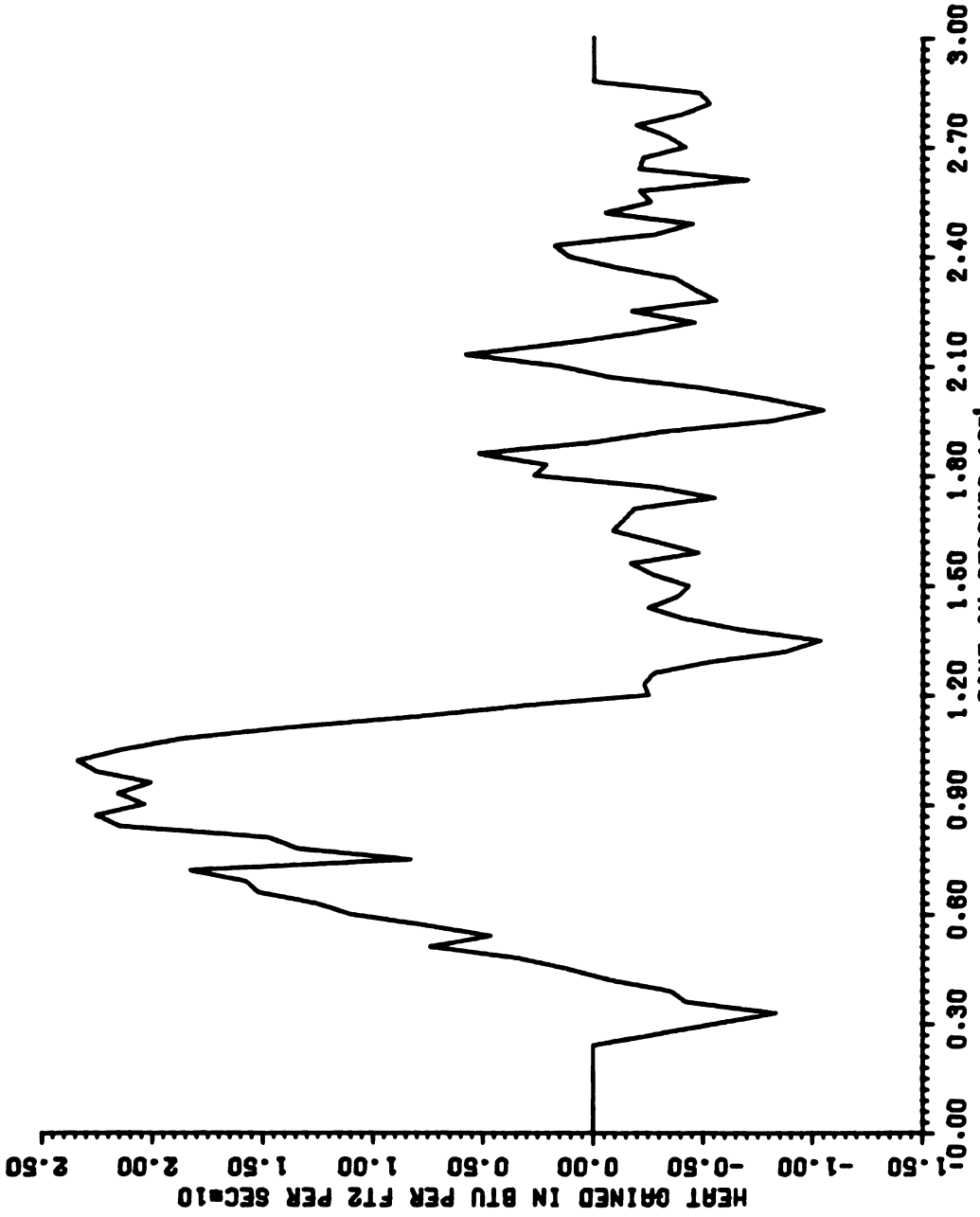

Figure 4.1. Heat flux at insulated surface: Case-1

so that small time steps could be used and noted that as the time steps are made smaller the calculated heat flux will oscillate due not to instability in the computation procedure but due to errors in the measured temperatures. Thus, the procedure is accurate enough to relate the random residuals to a random heat flux.

Computer program CONTL, developed by Van Fossen (1973), was used to estimate the heat flux at the alleged insulated surface. Program CONTL requires two surface temperatures and one interior temperature. In the input to CONTL the average measured temperature at the surface that is insulated in Farnia's data is used at the alleged heated surface. The average measured temperature at the surface that is heated in Farnia's data is used to create an insulated surface by assigning this temperature as the temperature for both the other surface and the adjacent interior temperature.

4.4 Same ARIMA Coefficients for All Sensors

The white noise variance and the other ARIMA(1,0,0) parameters are assumed to have the same numerical value for all sensor locations; namely, Equation (A.24) is used. The effect of unequal series variances can be investigated by examining the weights A_j computed by computer program PROPTY. The weights are computed from

$$A_j = \sigma_w^2 \ \sigma_j^{-2}$$

$$\sigma_j^2 = n^{-1} \sum_{i=1}^{n} e_j(i) \ e_j(i)$$

$$\sigma_w^2 = m^{-1} \sum_{j=1}^{m} \sigma_j^2$$

where $e_j(i)$ is the residual for the j-th sensor and i-th time defined in Equation (3.20.A).

The range divided by the mean for the weights A_j for case-0 data is 0.706 which is close to the average value of 0.617 computed in Table 3.8 where the coefficients are all the same because they are fixed in the simulation study. Thus, a common set of ARIMA coefficients for all the sensors is not unreasonable.

The change in the estimates of k and c when the weights A_j are used is also computed by computer program PROPTY. The correction for the thermal conductivity is δk equal 0.122 and δc equal -0.016. This uncertainty in the thermal parameters can be viewed as an insignificant shift in the ellipse plotted in Figure 3.1.

4.5 More Detailed ARIMA Coefficients

There are several models for non-constant values of the coefficient $\emptyset$ in the ARIMA(1,0,0) model. Beck and Arnold (1977, p. 320) displayed an ARIMA(2,0,0) process where the autoregressive coefficients can depend on the time index. This general form for the coefficients is computationally feasible when the coefficients $\emptyset$ are known. Computation with a variable $\emptyset$ was also considered by Blum (1961). Dutsch (1965, p. 125) cited spacecraft data as a case where the autoregressive coefficients are known functions of the time index.

Autoregressive coefficients that change with the time index would seem to be appropriate for the heat conduction residuals. In the heat conduction problem stated in Equation (2.3) the heat flux is only non-zero for a time period less than the time period that

the data are acquired. It is observed that the errors before the heat flux is applied are independent. It is expected that the errors are also independent after the solid reaches a new uniform temperature. While the heat flux is applied the errors are ARIMA(1,0,0). Thus the value of the parameter $\emptyset$ must be zero shortly before the heat flux begins and is again zero shortly after the experiment ends. Therefore, a model with a constant value of $\emptyset$ is only an approximation.

Only a few papers have considered estimating the $\emptyset$ coefficient when it is not constant. Subba-Rao (1969) estimated the ARIMA(1,0,0) parameter $\emptyset$ is known to have either an exponential or a gaussian dependency on time. Box and Jenkins (1970, Table 8.4) were early investigators of shifts in the coefficients, and showed that an ARIMA(0,1,1) model with shifts in the coefficients had a smaller Q-statistic and hence a better fit than a constant coefficient model for IBM stock prices. Ozaki and Tong (1975) used Akaike's information criterion to divide a series into intervals within which the coefficients are constant. Hsu (1973) used a block procedure with fifty points per block to detect changes of 0.1 in the autoregressive coefficient in the ARIMA(1,0,0) model.

Non-constant coefficients models are used in this dissertation for the white noise variance; see Equation (2.10.B) and Equation (A.24.B). The model assumed the following relation

$$A_{ij} = \sigma^2 \, I.$$

Beck (1974) listed three models for the A_{ij} matrix: homoscedastic, heteroscedastic, and steady-state. The homoscedastic model given above could be compared to the heteroscedastic model by using a range/mean plot suggested by Anderson (1976).

4.6 Improved ARIMA Order Specification

To establish the order of the ARIMA(p,d,q) model the author used the Akaike information criterion. There are other criteria which include the final prediction error, the criterion for auto-regressive transfer function. The study by Landers and Lacross (1977) showed that each criterion was comparable and each selected the same model. Chan et al. (1975) found these criteria were insensitive when the sample size is large. They developed a criterion that is useful for large samples, with the criterion depending on both the estimate of the white noise variance used in the previous criteria and on the Hessian matrix from the Newton-Raphson estimation procedure.

It is possible that the effect of sample size occurred in the poor discrimination in Tables 2.8 through 2.10 where the sample size is large and the better discrimination in Tables 2.12 through 2.13 where the sample size is small.

The insensitivity of the criterion in Equation (2.14) and in Equation (2.15) in large sample sizes is that the criterion can be written approximately as

$$a^{+} = n \ln(\sigma^2). \tag{4.7}$$

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

In Section 5.1 the major conclusions stated throughout the dissertation are collected and restated. Also some recommendations for implementing Aitken's least square are made. A subjective evaluation of the importance of these conclusions is given in Section 5.2 for engineering design and in Section 5.3 for parameter estimation.

5.1 Feasibility of Aitken's Least Squares Estimation

In Chapter II, the model of the error covariance matrix was shown to correspond to serial correlation. The procedure recommended in the literature to select the best ARIMA(p,d,q) model for the serial correlation is based on computation of a numerical criterion called Akaike's Information Criterion. However, it was concluded that additional considerations are required to select the best model because several different ARIMA(p,d,q) models gave a good fit to the residuals. Hence, three considerations were recommended which confirmed the choice of the ARIMA(1,0,1) model as best: (1) in Table 2.10 the ARIMA(1,0,1) model is identified as best more often than the other candidate models, (2) in Table 2.11 the coefficients in the ARIMA (1,0,1) model are more stable than those for the other models, and (3) in Table 2.15 good agreement is shown between the predicted and

observed change in coefficients in the ARIMA(1,0,1) model when the sampling rate changes.

In Chapter III, the parameter estimates were computed quite simply by the Aitken's least square procedure. The results displayed in Table 3.8 show that both the physical parameters (the thermal conductivity and specific heat) and the statistical parameters ($\emptyset$, θ, and σ^2) have estimated values close to their true values. The results displayed in Table 3.10 show that the estimated covariance matrix for the thermal parameters is reasonably close to its true value. I made a Monte Carlo study and presented the results in Table 3.5, and I analyzed data presented by Gallant and Goebel (1976) and presented the results in Table 3.2. The results of this analysis indicate that the distribution should be noncentral F when the ARIMA coefficients are estimated. The results of a simulation study on the heat conduction model indicated that the noncentrality parameter is close to zero. Hence, the use of the central F-distribution to compute the fractile for the confidence ellipse is recommended.

The plotted confidence regions in Figure 3.1 show that the effect of neglecting the presence of correlated errors is significant. However, developing all the necessary computer code to model and estimate the physical and statistical parameters by Aitken's least squares is time consuming. Hence, a recommended procedure to make a preliminary investigation of the effect of serially correlated errors on the estimated parameters and their standard errors is as follows. For the model given as Equation (1.1) the procedure is to make computations in six steps:

(1) Compute the estimate of b by Equation (5.1) where

$$b = (X^t X)^t X^t y \tag{5.1}$$

to obtain the residuals

$$e = y - X b, \tag{5.2}$$

(2) Estimate the coefficient $\emptyset$ for an ARIMA(1,0,0) model using Equation (2.11) which only involves examining the shape of the plotted residuals from Equation (5.2),

(3) Compute the standard least squares confidence interval using Equation (1.5),

(4) Define the vectors F and Z using Equation (1.8) where the matrix $L^{-1} = I - \emptyset K$ and K is defined in Equation (3.29),

(5) Compute the estimate of b using Equation (5.3) where

$$b = (Z^t Z)^{-1} Z^t F$$

and also obtain the residuals given by

$$e = F - Z b, \tag{5.4}$$

(6) Compute the modified confidence interval using Equation (1.11).

The significance of the correlated errors on the estimates of the physical parameters is the difference in the estimates from steps (1) and (5) while the significance of the correlated errors on the confidence interval is the difference in the estimates from steps (3) and (6). If the results are significant then the procedures in Section 2.4 may be necessary unless the estimate of $\emptyset$ is zero when step (2) is repeated on the residuals in Equation (5.4).

This six step procedure is essentially equivalent to the procedure used to obtain the confidence ellipses plotted as Figure 3.2 which is in good agreement with the statistically valid and more complicated to obtain ellipses plotted in Figure 3.1.

5.2 Engineering Design Implications

The presence of ARIMA(1,0,1) errors does not significantly affect engineering design calculations for Armco iron based on a solution of the heat conduction equation, Equation (2.1). Although the Aitken's least squares confidence ellipse displayed in Figure 3.1 is large, the values contained within the ellipse are still within the ±5 percent typically used in engineering design calculations. However, materials other than Armco iron are of interest to engineers since the heat conduction equation is used with a wide range of materials encountered in the academic disciplines of mechanical engineering, agricultural engineering, chemical engineering, petroleum engineering, and so forth. For some of these materials the size of the Aitken's confidence ellipse may be larger than the accuracy assumed for the physical properties in engineering design calculations.

5.3 Parameter Estimation Implications

It is shown in Figure 3.1 that the Aitken's confidence ellipse is larger than the standard least squares confidence ellipse. Therefore, the presence of ARIMA(1,0,1) errors should be accounted for by the Aitken's estimation procedure in order to make accurate statements of the confidence ellipse and of standard errors. The limited results

displayed in Table 4.1 suggest that the duration of the optimum experiment is less when there are ARIMA(1,0,1) errors. Thus, investigators in parameter estimation certainly need to use models of autocorrelated errors when designing experiments and stating confidence regions.

APPENDICES

APPENDIX A

COMPUTATION OF AKAIKE'S CRITERION

APPENDIX A

COMPUTATION OF AKAIKE'S CRITERION

The purpose of this section is to present computational methods to evaluate the Akaike's criterion in Equation (2.15).

A.1 Stationarity and Invertibility Tests

Because the sample size can be large, the computations are simplified when the ARIMA models are both stationary and invertible. Therefore, tests are given to insure that the models are simultaneously stationary and invertible.

The backward shift operator, B, is defined as $B\ z_i = z_{i-1}$. Using this shift operator notation the ARIMA(p,d,q) process can be written as

$$(1 - \phi_1 B - \ldots - \phi_p B^p)\ z_i = (1 - \theta_1 B - \ldots - \theta_q B^q)\ a_i \qquad (A.1)$$

The stationarity condition is that in the autoregressive representation of the ARIMA process; i.e., if the process is written as

$$z_i = \phi^{-1}(B)\ \theta(B)\ a_i$$

the variance, σ_z^2, of the z-series is finite. The invertibility condition is that in the moving-average representation; i.e., if the process is written as

$$a_i = \theta^{-1}(B)\ \phi(B)\ z_i$$

the present value of a_i only depends on z_i at prior times. The test is that the roots of the $\emptyset^{-1}(B)\ \theta(B)$ and $\theta^{-1}(B)\ \emptyset(B)$ polynomials lie within the unit circle. Any finite moving-average process is stationary and any finite autoregressive process is invertible. The reader is referred to Box and Jenkins (1970, p. 49) for additional information.

Recent researchers do not give tests for stationarity and invertibility when the values of p and q are greater than two. However, Wise (1956) developed conditions for orders up to ten and explicitly stated conditions for orders up to four. Either the autoregressive or the moving-average component of Equation (A.1) can be written as

$$x^n + \alpha_1 x^{n-1} + \ldots + \alpha_n = 0 \qquad \text{(A.2)}$$

where $x = B^{-1}$ and α is either $\emptyset$ or θ. A direct test on the roots would be to solve Equation (A.2) in order to verify that each x_i in Equation (A.3) is less than unity where

$$(x - x_1)\ (x - x_2) \ldots (x - x_n) = 0. \qquad \text{(A.3)}$$

An equivalent and more easily computed test is based on the α's in Equation (A.2). The conditions on the α's are necessary and sufficient.

For n equal one, the single condition is

$$0 \leq (1 + \alpha_1)\ (1 - \alpha_1)^{-1}. \qquad \text{(A.4)}$$

For n equal two, the two conditions are

$$0 \leq 2(1 - \alpha_2)\ (1 - \alpha_1 - \alpha_2)^{-1}, \text{ and} \qquad \text{(A.5.A)}$$

$$0 \leq (1 + \alpha_1 + \alpha_2)\ (1 - \alpha_1 + \alpha_2). \qquad \text{(A.5.B)}$$

For n equal three and n equal four, the reader is referred to the rather lengthy expressions given in Wise (1956). I verified the expressions given in Wise (1956) for orders up to and including four and I programmed them into my computer code.

A.2 Computation Method with MA Absent

The autoregressive process can be written as

$$z_i = \phi_1 z_{i-1} + \phi_2 z_{i-2} + \ldots + \phi_p z_{i-p} + a_i. \qquad \text{(A.6)}$$

The procedure used to estimate the ϕ's utilized the Yule-Walker equations. These equations are obtained by multiplying Equation (A.6) by z_{i-j} for the smallest j values satisfying $p \leq j$, taking expectations with $E(a_i\ z_{i-j}) = 0$, and using the estimates of c_j in Equation (A.8) for the non-zero expectations. The objective is to estimate the ϕ and σ^2 values in Equation (A.6) using the system of Equations (A.7) and (A.16). Matrix Equation (A.7) is the Yule-Walker equation

$$r = R\ \phi \qquad \text{(A.7)}$$

where

$$\phi^t = (\phi_1, \phi_2, \ldots, \phi_p)$$

$$r^t = (r_1, r_2, \ldots, r_p)$$

and

$$R = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ \cdots & & \cdots & & \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{bmatrix}$$

$$c_j = n^{-1} \sum_{i=1}^{n-j} a_i z_{i-j} \tag{A.8}$$

$$r_j = c_j / c_0 \tag{A.9}$$

Because the matrix R is symmetric, there are a number of well-known methods for solving Equation (A.7). A convenient method to solve the system of equations is the Choleski deconposition method stated as FORTRAN code in Healy (1968) with the code corrected by Farebrother and Berry (1974). An estimate of the white noise variance uses the r and $\emptyset$ values computed above and is given in Box and Jenkins (1970, Equation 3.2.8) as

$$\sigma^2 = (1 - \emptyset^t r)\, c_0.$$

However, to reduce the computational error, an algorithm suggested by Pagano (1972) was used to compute the estimate of σ^2.

The algorithm developed by Pagano (1972) computed σ^2 without computing the vector $\emptyset$. Define a square matrix R^+ as the R matrix in Equation (A.7) with dimension (p + 1) instead of p; this matrix can be written as

$$R^+ = \begin{bmatrix} R & J\, r \\ (J\, r)^t & 1 \end{bmatrix}$$

where the J matrix is a matrix with ones on the reverse diagonal. The matrix J is related to the usual identity matrix I by

$$J_{ik} = I_{ij} \quad \text{where } k = n + 1 - j. \tag{A.10}$$

The determinant of the R^+ matrix satisfies

$$\det(R^+) = (1 - r^t J R^{-1} J r) \det(R). \tag{A.11}$$

Because R^{-1} is centrosymmetric, it satisfies

$$J R^{-1} J = R^{-1}. \tag{A.12}$$

Substituting Equations (A.12) into Equation (A.11) yields

$$\det(R^+) = (1 - r^t R^{-1} r) \det(R). \tag{A.13}$$

The Yule-Walker equations (A.7) can be written as

$$R^{-1} r = \emptyset$$

and when substituted into Equation (A.13) yields

$$\det(R^+) = (1 - r^t \emptyset) \det(R). \tag{A.14}$$

Taking the ratio of Equation (A.14) and Equation (A.10) yields the more computationally efficient estimate of the white noise variance

$$\sigma^2 = c_0 \det(R^+) / \det(R). \tag{A.15}$$

In order to use Equation (A.15) we need to compute both R and R^+ and methods for doing this will now be considered. The R matrix has the Choleski decomposition

$$R = L D L^t$$

where L is lower triangular with ones on the diagonal and D is diagonal. Because the determinant of a product is the product of

the determinants, the determinant of the matrix L is one, and

$$\det R^{+} = \det D^{+} \quad \text{and} \quad \det R = \det D.$$

Therefore, the estimator in Equation (A.15) reduces to

$$\sigma^2 = c_0 \, d_{p+1} \tag{A.16}$$

where d_{p+1} is the entry in the (p+1) position in the D matrix which is diagonal.

A.3 Computation Method with MA Present

The estimates in Equations (A.7) and (A.16) are only valid when there is no moving-average component. Chow (1972) proposed a set of equations when there is a moving-average component. I developed a solution for the equations presented by Chow (1972) that involves a persymmetric transformation and symmetric decomposition to find the autoregressive coefficients, and an analytical solution to find the moving-average coefficients and the white noise.

The mixed ARIMA process can be written as

$$z_i = \phi_1 z_{i-1} + \ldots + \phi_p z_{i-p} + a_i - \theta a_{i-1} - \ldots - \theta_q a_{i-q}. \tag{A.17}$$

Chow (1972) used a slightly different notation for his ARIMA model but the conversion of the equations is quite straightforward. A set of equations for the autoregressive parameter vector ϕ is obtained by multiplying Equation (A.17) by z_{i-j} for the smallest j value satisfying $(q+1) \leq j$, noting that $E(a_i \, z_{i-j}) = 0$, and using the estimates of c_j in Equation (A.8) for the non-zero expectations.

The resulting set of equations can be written as

$$r = R \phi \tag{A.18}$$

where

$$r^t = (r_{q+1}, r_{q+2}, \ldots, r_{q+p})^t,$$

$$\phi^t = (\phi_1, \phi_2, \ldots, \phi_p)^t \text{ and}$$

$$R = \begin{bmatrix} r_q & r_{q-1} & r_{q-2} & \cdots & r_{q+1-p} \\ r_{q+1} & r_q & r_{q-1} & \cdots & r_{q+2-p} \\ & \cdots & & \cdots & \\ r_{q+p-1} & r_{q+p-2} & r_{q+p-3} & \cdots & r_q \end{bmatrix}$$

Because $1 \leq q$, the matrix R is not symmetric and the Choleski decomposition solution procedure can not be used directly. However, because R is persymmetric the matrix R J is symmetric. By using the identity I = J J, an algorithm by Rutishauser (1963) for symmetric but not necessarily positive definite matrices can be used to solve

$$r = R J J \phi$$

for J ϕ. The desired autoregressive parameter vector ϕ is then computed from the vector solution J ϕ by left multiplying by J since ϕ = J J ϕ.

Having found the autoregressive parameters, a system of equations for estimating both the moving-average parameters and the white noise variance is obtained by the three step procedure of writing the ARIMA model at two base times, multiplying corresponding sides, and

taking expectations. In matrix form the system of equations obtained by this procedure can be written as

$$\sigma^2 \, B \, \theta = - \, y \tag{A.19}$$

where the y vector depends on the estimated autoregressive parameter vector computed by Equation (A.18). The components of the system of equations in (A.19) are as follows:

$$\theta^t = (1, \, - \theta_1, \, - \theta_2, \, . \, . \, . \, , \, - \theta_q)$$

$$B = \begin{bmatrix} \theta_q & 0 & 0 & . \, . \, . & 0 \\ \theta_{q-1} & \theta_q & 0 & . \, . \, . & 0 \\ . \, . \, . & & & . \, . \, . & . \\ -1 & \theta_1 & \theta_2 & . \, . \, . & \theta_q \end{bmatrix}$$

and

$$y^t = (y_0, \, y_1, \, . \, . \, . \, , \, y_q)$$

The components of the y vector are given by

$$y_s = \sum_{i=0}^{p} \sum_{j=0}^{p} \phi_i \, \phi_j \, c_{q-s+i-j} \tag{A.20}$$

where the definition $\phi_0 = -1$ was used in Equation (A.20).

Chow (1972) does not indicate how Equation (A.19) is to be solved for the coefficient vector θ.

I found explicit solutions of Equation (A.19) when the order of the moving-average process is small. For q equal one the solution can be written as

$$\sigma^2 = \tfrac{1}{2}(y_1 + (y_1^2 - r\, y_0^2)^{\frac{1}{2}}), \text{ and} \qquad \text{(A.21.A)}$$

$$\theta_1 = - y_0\, \sigma^{-2}.$$

Because Equation (A.19) is nonlinear in θ, there are multiple solutions. The proper solution in Equation (A.21) has the positive radical; this choice was also used by Nelson (1974) because it makes σ^2 large and θ_1 small. For q equal two the solution can be written as

$$(\sigma\theta_1)^2 = \tfrac{1}{2}(y_2 + 2\, y_0 \pm (y_2^2 + 4y_0y_2 + 4y_0^2 - 4y_1^2)^{\frac{1}{2}}), \qquad \text{(A.22.A)}$$

$$\sigma = \tfrac{1}{2}(y_1 \pm (y_1^2 - 4y_0\, \sigma^2\, \theta_1^2)^{\frac{1}{2}})\ (\sigma\, \theta_1)^{-1}, \text{ and} \qquad \text{(A.22.B)}$$

$$\theta_2 = - y_0\, \sigma^{-2}. \qquad \text{(A.22.C)}$$

By evaluating Equations (A.22) in the order A through C, a non-iterative solution is obtained. There are multiple solutions with the correct solution obtained by testing the sign for θ_1 so that the following relation is satisfied

$$\sigma^2\, (\theta_2 - \theta_1) = y_1\ , \qquad \text{(A.22.D)}$$

and by selecting the sign on the radicals that makes σ^2 large and θ_1 and θ_2 small in absolute value.

A.4 Verification of the Pagano-Chow Algorithm

The solution procedure in Equations (A.7) and (A.15) is due to Pagano (1972) while the procedure in Equations (A.18) and (A.19) is based on Chow (1972). Hence, the resulting procedure will be called the Pagano-Chow algorithm. This algorithm was used because

it resulted in calculation, at reasonable cost, of the Akaike's criterion in Equation (2.15) for the forty-five candidate models shown in Tables 2.12 and 2.13. The accuracy of the Pagano-Chow algorithm was verified by comparing estimates with those reported by Ozaki (1975) for Series-A in Box and Jenkins (1970). The agreement shown in Table A.1 is good: there is a similar ranking of the top six models and similar coefficients. Both procedures selected the ARIMA(1,0,1) model as best which validates the Pagano-Chow algorithm.

Table A.1

ARIMA Model Discrimination for Series A

ARIMA Model	Estimated a^+	ARIMA Coefficients
	Estimates by Pagano-Chow Procedure	
1,0,1	-448.754	ϕ = 0.868; θ = 0.480
2,0,0	-445.127	ϕ = 0.427, 0.252
1,0,2	-444.079	ϕ = 0.803; θ = 0.395, -0.058
3,0,0	-444.049	ϕ = 0.637, 0.223, 0.068
1,0,0	-434.219	ϕ = 0.570
0,1,1	-432.639	θ = 0.528
	Ozaki's Estimates by Least Squares*	
1,0,1	-448.03	ϕ = 0.88; θ = 0.52
1,0,2	-446.79	ϕ = 0.92; θ = 0.54, 0.07
2,0,0	-445.13	ϕ = 0.43, 0.25
0,1,1	-444.71	θ = 0.70
3,0,0	-444.05	ϕ = 0.41, 0.22, 0.07
1,0,0	-434.22	ϕ = 0.57

*Source: Ozaki (1975).

A.5 Autocovariance for Multisensor Data

The heat conduction residuals (shown in Figures 2.1 through 2.20) came from experiments with four sensors at each of two surfaces of the disk. Each set of residuals yields a series which can be used to compute separate estimates of the coefficients of the ARIMA (1,0,0) model and other ARIMA(p,d,q) models. Also an overall estimate of ϕ could be computed using all the residuals at one time. The average of the individual estimates of ϕ is not the same as the overall estimate.

Consider an ARIMA(1,0,0) model where there are two coefficients ϕ and σ^2. Equation (A.7) yields an estimate for the j-th sensor that can be written as

$$\phi_j = \frac{(K\,w)^t\, w}{(K\,w)^t\, K\,w} \quad , \text{ and} \tag{A.23.A}$$

$$\sigma_j^2 = n^{-1}\,(P^{-1}\,w)^t\,P^{-1}\,w \tag{A.23.B}$$

where there are n measurements at the j-th sensor, w is the vector of residuals, and the matrices K and P are defined in Equations (3.29 and (3.28). The average of the individual estimates is obviously written, for m sensors, as

$$\phi = m^{-1} \sum_{j=1}^{m} \phi_j . \tag{A.24}$$

The overall estimates of the parameters in the ARIMA(1,0,0) model can be written, from Equation (A.7), as

$$\phi = \frac{((I \# K)\, w^*)^t\, w^*}{((I \# K)\, w^*)^t\, (I \# K)\, w^*} \qquad \text{(A.25.A)}$$

$$\sigma^2 = m^{-1}\, n^{-1}\, ((I \# P^{-1})\, w^*)^t\, (I \# P^{-1})\, w^* \qquad \text{(A.25.B)}$$

where w^* is a vector of dimension mn obtained by stacking the m residual vectors w, one from each sensor. The special symbol # denotes the kronecker matrix product discussed in Theil (1971).

In general the estimate from Equation (A.24) is not the same as the estimate obtained from Equation (A.25.A). It is fairly clear to see that the estimators in Equation (A.25) are based on the average of the autocovariances defined in Equation (A.8). The estimates in Tables 2.8 through 2.11 are those obtained by using the averaging of the autocovariances from the appropriate sensors.

APPENDIX B

DERIVATION OF DISTRIBUTION FOR PERCENTILES WHEN V IS KNOWN

APPENDIX B

DERIVATION OF DISTRIBUTION FOR PERCENTILES WHEN V IS KNOWN

In this appendix the percentiles for the confidence regions in Section 3.2 are derived. The probability distributions involved are expressed in terms of ratios of quadratic forms in normal variates. The derivation is for the physical model in Equation (1.1) and the statistical model in Equation (1.6) which satisfies the eight assumptions listed in connection with Equation (1.14.A). For notational convenience, the same symbols (b and s^2) are used, although they are defined differently in Sections B.1 through B.3. This is reasonable because only Equations (B.3) and (B.5) are used elsewhere in the dissertation.

B.1 ALS Confidence Ellipse

The Aitken's least squares (ALS) estimators of β and σ^2 can be written as

$$b = (X^t V^{-1} X)^{-1} X^t V^{-1} y, \text{ and}$$

$$s^2 = (n - k)^{-1} e^t V^{-1} e$$

where the residual vector e is written as

$$e = y - X b.$$

From Theil (1971, p. 238) the covariance matrix for the parameter vector b can be written as

$$\text{cov}(b) = \sigma^2 Q_{ALS}$$

where the following definition of Q_{ALS} is introduced

$$Q_{ALS} = (X^t V^{-1} X)^{-1}.$$

The quadratic form of interest can be written as

$$(b - \beta)^t (X^t W^{-1} X) (b - \beta) = \frac{(b - \beta)^t \sigma^{-2} Q_{ALS}^{-1} (b - \beta)}{(n - k)^{-1} \sigma^{-2} e^t V^{-1} e}.$$

Thus the quadratic form involves the ratios of two quadratic forms that can be expressed in normal variates.

The numerator quadratic form can be expressed in normal variates by substituting Equation (1.1) into the estimator of b to find $(b - \beta)$ and this substitution yields

$$(b - \beta)^t \sigma^{-2} Q_{ALS}^{-1} (b - \beta) = w^t N W^{-1} w$$

where the following definition of N has been introduced

$$N = V^{-1} X (X^t V^{-1} X)^{-1} X^t.$$

The conditions given by Theil (1971, p. 82) to establish whether this satisfies the conditions for the quadratic form to be chi-square do not apply. However, Shanbhag's theorem given as Equation (3.10.A) applies by noting that the variance of the vector w is the matrix W, and making the associations C = W and $A = N W^{-1}$. Because A C equals N and N is idempotent $(A C)^2 = (A C)^3 = (A C)^4$. Hence the quadratic form has a central chi-square distribution. The number of degrees of freedom in this distribution is found from Equation (3.10.B) as

$$\text{tr } N = \text{tr } V^{-1} X (X^t V^{-1} X)^{-1} X^t = \text{tr } (X^t V^{-1} X)(X^t V^{-1} X)^{-1}$$

Clearly tr N equals k the column dimension of the X matrix. Hence the numerator quadratic form is chi-square with k degrees of freedom.

The denominator quadratic form is related to the residual vector e which can be written as

$$e = y - X\,b = (I - X\,(X^t\,V^{-1}\,X)^{-1}\,X^t\,V^{-1})\,y.$$

Using the definition of the N matrix yields

$$e = (I - N^t)\,w$$

where w is a normal variate. The denominator quadratic form satisfies the following identity

$$e^t\,\sigma^{-2}\,V^{-1}\,e = e^t\,W^{-1}\,e$$

which can be expressed as a quadratic form in w as

$$e^t\,W^{-1}\,e = w^t\,(I - N)\,W^{-1}\,(I - N^t)\,w$$

or equivalently as

$$e^t\,W^{-1}\,e = w^t\,(W^{-1} - N\,W^{-1})\,w. \qquad (B.1)$$

Noting that the variance of the normal vector w is W, Shanbhag's theorem applies with the association $A = (I - N)\,W^{-1}$ and $C = W$. Because

$$A\,C = I - N$$

which is idempotent, Equation (3.10.A) can be used to show that the denominator quadratic form is central chi-square. The number of degrees of freedom is computed from Equation (3.10.B) as

$$\begin{aligned} \mathrm{tr}(I - N) &= \mathrm{tr}(I - W^{-1}\,X\,(X^t\,V^{-1}\,X)^{-1}\,X^t) \\ &= \mathrm{tr}\,I_n - \mathrm{tr}\,(X^t\,V^{-1}\,X)(X^t\,V^{-1}\,X)^{-1}. \end{aligned}$$

Hence the denominator quadratic form is distributed as σ^2 times a central chi-square with (n - k) degrees of freedom. The estimator of the white noise variance can be written as

$$s^2 = (n - k)^{-1} e^t V^{-1} e \qquad (B.2)$$

is unbiased since the expected value of the chi-square distribution that the quadratic form in Equation (B.1) satisfies is (n - k). The distribution of the quadratic form that is used for the ALS confidence ellipse can be written as

$$(b - \beta)^t s^{-2} Q_{ALS}^{-1} (b - \beta) = \frac{(n - k)\, \chi^2(k)}{\chi^2(n-k)}$$

where $\chi^2(p)$ denotes a chi-square with p degrees of freedom.

Because the appropriate covariance matrix is W, the test in Equation (3.11) for independence of these chi-square distributions is whether the following is zero,

$$(N W^{-1}) W (W^{-1} - N W^{-1}) = N W^{-1} - N N W^{-1}.$$

This is zero because N is idempotent. Thus the chi-square distributions in the numerator and in the denominator are independent. Hence, as stated by Beck and Arnold (1977, p. 300), the ALS confidence ellipse is described by a central F distribution; namely,

$$(b - \beta)^t Q_{ALS}^{-1} (b - \beta) \approx k\, s^2\, F(k,n-k,1-\alpha).$$

For comparison with $G(1-\alpha)$ used in Equation (3.5) note that the following identification can be made

$$G_{ALS}(1 - \alpha) = F(k,n-k,1-\alpha). \qquad (B.3)$$

B.2 MLS Confidence Ellipse

The mixed least squares estimators of β and σ^2 from Theil (1971, p. 240) can be written as

$$b = (X^t X)^{-1} X^t y, \text{ and}$$

$$s^2 = (n - k)^{-1} e^t V^{-1} e.$$

From Theil (1971, p. 247) the covariance matrix for the parameter vector b can be written as

$$\text{cov}(b) = \sigma^2 Q_{MLS}$$

where the following definition of Q_{MLS} is introduced

$$Q_{MLS} = (X^t X)^{-1} (X^t V X) (X^t X)^{-1}.$$

The quadratic form of interest can be written as

$$(b - \beta)^t X^t X (X^t W X)^{-1} X^t X (b - \beta) = \frac{(b - \beta)^t Q_{MLS}^{-1} (b - \beta)}{(n - k)^{-1} e^t V^{-1} e}.$$

The derivation proceeds as in Section B.1. Substituting Equation (1.1) into the estimator for b yields

$$b - \beta = (X^t X)^{-1} X^t w$$

so that the numerator quadratic form can be written as

$$(b - \beta)^t Q_{MLS}^{-1} (b - \beta) = \sigma^2 w^t X (X^t W X)^{-1} X^t w.$$

Noting that W is the covariance matrix of the normal vector w and making the associations $A = X (X^t W X)^{-1} X^t$ and $C = W$, Shanbhag's theorem in Equation (3.10.A) can be applied. By direct evaluation $(A C)^2 = A C$, so that the distribution appropriate for the numerator quadratic form is central chi-square. Because

$$A\ C = X\ (X^t\ W\ X)^{-1}\ X^t\ W$$

the number of degrees of freedom can be found using Equation (3.10.B) as

$$tr\ X\ (X^t\ W\ X)^{-1}\ X^t\ W = tr\ (X^t\ W\ X)^{-1}\ (X^t\ W\ X)\ .$$

Hence the numerator MLS quadratic form is distributed as central chi-square with k degrees of freedom.

The MLS denominator quadratic form is the same as the ALS denominator quadratic form given as Equation (B.1). This was shown to be chi-square with (n - k) degrees of freedom.

Therefore, the distribution of the quadratic form used for the MLS confidence ellipse can be written as

$$(b - \beta)^t\ s^{-2}\ Q_{MLS}^{-1}\ (b - \beta) \approx \frac{(n - k)\ \chi^2(k)}{\chi^2(n-k)}\ . \qquad (B.4)$$

Because the covariance matrix is W, the test for independence of the numerator and denominator chi-square distributions is that the following matrix product is zero; namely,

$$\sigma^2\ (X\ (X^t\ W\ X)^{-1}\ )\ W\ (W^{-1} - N\ W^{-1}).$$

This can be rearranged as

$$X\ (X^t\ V\ X)^{-1}\ X^t - X\ (X^t\ V\ X)^{-1}\ X^t\ X\ (X^t\ V^{-1}\ X)^{-1}\ X^t\ V^{-1}$$

which is orthogonal to X and is zero when V equals the identity matrix. Because the test in Equation (3.11) for independence was not passed, it is unclear what distribution $G_{MLS}(1-\alpha)$ has.

B.3 SLS Confidence Ellipse

In standard least squares the V matrix equals the identity matrix which is not correct when Equation (1.6) applies instead of Equation (1.2). The standard least squares estimators of β and σ^2 can be written as

$$b = (X^t X)^{-1} X^t y, \text{ and}$$

$$s^2 = (n - k)^{-1} e^t e.$$

From Theil (1971, p. 112) the covariance matrix for the parameter vector b can be written as

$$\text{cov}(b) = \sigma^2 Q_{SLS}$$

where the following definition of Q_{SLS} has been introduced

$$Q_{SLS} = (X^t X)^{-1}.$$

The quadratic form of interest can be written as

$$(b - \beta)^t \sigma^{-2} X^t X (B - \beta) = \frac{(b - \beta)^t Q_{SLS}^{-1} (b - \beta)}{(n - k)^{-1} e^t e}.$$

The derivation is given in Theil (1971, p. 129) and by others but will be sketched here briefly for completeness. The derivation proceeds as in Section B.1 and B.2.

The numerator quadratic form can be written as

$$(b - \beta)^t Q_{SLS}^{-1} (b - \beta) - w^t X (X^t X)^{-1} X^t W.$$

By making the associations $A = X (X^t X)^{-1} X^t$ and $C = I$, Shanbhag's theorem in Equation (3.10.A) can be applied to show that the distribution appropriate for the numerator quadratic form is chi-squared

B.3 SLS Confidence Ellipse

In standard least squares the V matrix equals the identity matrix which is not correct when Equation (1.6) applies instead of Equation (1.2). The standard least squares estimators of β and σ^2 can be written as

$$b = (X^t X)^{-1} X^t y, \text{ and}$$

$$s^2 = (n - k)^{-1} e^t e.$$

From Theil (1971, p. 112) the covariance matrix for the parameter vector b can be written as

$$\text{cov}(b) = \sigma^2 Q_{SLS}$$

where the following definition of Q_{SLS} has been introduced

$$Q_{SLS} = (X^t X)^{-1}.$$

The quadratic form of interest can be written as

$$(b - \beta)^t \sigma^{-2} X^t X (B - \beta) = \frac{(b - \beta)^t Q_{SLS}^{-1} (b - \beta)}{(n - k)^{-1} e^t e}.$$

The derivation is given in Theil (1971, p. 129) and by others but will be sketched here briefly for completeness. The derivation proceeds as in Section B.1 and B.2.

The numerator quadratic form can be written as

$$(b - \beta)^t Q_{SLS}^{-1} (b - \beta) - w^t X (X^t X)^{-1} X^t W.$$

By making the associations $A = X (X^t X)^{-1} X^t$ and $C = I$, Shanbhag's theorem in Equation (3.10.A) can be applied to show that the distribution appropriate for the numerator quadratic form is chi-squared

since by direct evaluation $(A\ C)^2 = A\ C$. The number of degrees of freedom is easily shown to be k. The denominator quadratic form can be written as

$$e^t\ e = w^t\ M\ w$$

where the following definition of the matrix M has been introduced

$$M = I - X\ (X^t\ X)^{-1}\ X^t.$$

Again Shanbhag's theorem can be applied to the denominator quadratic form to show that $e^t\ e$ is chi-square with (n-k) degrees of freedom. Because $M\ X = 0$, the chi-square distributions are independent and this can be shown by applying Equation (3.11). Hence, the SLS confidence ellipse is described by a central F distribution; namely,

$$(b - \beta)^t\ Q_{SLS}^{-1}\ (b - \beta) \approx k\ s^2\ F(k,n-k,1-\alpha)\ . \qquad (B.5)$$

Comparing Equation (B.5) with Equation (3.5) yields the following identification

$$G_{SLS}(1-\alpha) = F(k,n-k,1-\alpha). \qquad (B.6)$$

Equation (B.5) is correct only when the standard assumptions are valid including the assumption that the V matrix is the identity matrix.

B.4 Computation of Percentiles

A simple FORTRAN program by Selvin and Wong (1975) was used to compute the probabilities for the F distribution. In calculating the confidence ellipses, the eigenproblem was solved using the algorithm by Boothroyd (1968). The F percentiles were computed using the

Kelley-and-Arnold equation recommended by Sahai and Thompson (1974) with the required normal fractiles computed by an algorithm stated by Odeh and Evans (1974).

APPENDIX C

SIMULATION OF HEAT CONDUCTION DATA

APPENDIX C

SIMULATION OF HEAT CONDUCTION DATA

In this appendix the method used to generate the data analyzed in the simulation study in Section 3.5 is discussed.

C.1 Analytical Solution for Temperature

The theoretical temperature simulated is an analytical solution of the problem of a heat pulse of duration $t_E - t_B$ acting on one face of an insulated disk. The solution can be written as

$$T(x,t) = T_o + u(x,t-t_B)\, H(t-t_B) - u(x,t-t_E)\, H(t-t_E) \qquad (C.1)$$

where $u(x,t)$ is a solution of the following problem

$$c \frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2} \qquad (C.2)$$

subject to the conditions

$$u(x,0) = 0 \qquad (C.3)$$

$$k \frac{\partial u(0,t)}{\partial x} = 0 \qquad (C.4)$$

$$\frac{\partial u(L,t)}{\partial x} = q\, H(t). \qquad (C.5)$$

The analytical solutions to Equation (C.2) subject to the initial and boundary conditions in Equations (C.3) through (C.5) were given by Gautschi (1962) as follows. The solution at the heated surface can be written as

$$u(L,t) = q\ L\ k^{-1}\ (\alpha\ t\ L^{-2} + 3^{-1}\ D_3(\alpha\ t\ \pi^2\ L^{-2})\ ). \qquad (C.6)$$

The solution at the insulated surface can be written as

$$u(0,t) = q\ L\ k^{-1}\ (\alpha\ t\ L^{-2} - 6^{-1}\ D_4(\alpha\ t\ \pi^2\ L^{-2})\ ). \qquad (C.7)$$

As their argument increases from zero to infinity both the function D_3 and D_4 increase monotonically from zero to unity. Formal analytical solutions for either large times or for small times are also given in standard references, such as Carslaw and Jaeger (1947, p. 112).

However, the numerical evaluation of the formal analytical solutions is not stated as an algorithm in standard references. Frank (1962) tried to develop approximate expressions, valid for all times, that could be used in an algorithm; however, I found that these expansions of the Laplace integral transform solutions are not accurate at the insulated surface. Gautschi (1962) obtained approximate solutions to the theta function solutions given in Equations (C.6) and (C.7). I used these approximations and the series solutions also stated by Gautschi (1962) to obtain the set of solutions stated in Table C.1. The set of solutions in Table C.1 can be used as an algorithm to compute the solution of Equation (C.2).

C.2 Generation of Normal Variates

The simulated data used in Section 3.5 involved both the theoretical temperature computed using Equation (C.1) and additive errors. The additive errors were described by an ARIMA(1,0,1) process. The simulated temperature data, denoted $T_e(x,t)$, are computed from

Table C.1

Approximations for $u(0,t)$ and $u(L,t)$

Time Region		
	Insulated Surface $u(0,t)$, $t^* = 0.2$	
$\theta \leq t^*$	$u(0,t) = 0$	
$t^* \leq \theta$	$u(0,t) = q\, L\, k^{-1}\, \pi^{-2} (\theta - 6^{-1}\, \pi^2 - 2 \sum_{n=1}^{6} (-1)^n\, n^{-2} \exp(-\theta\, n^2))$	
	Heated Surface $u(L,t)$, $t^* = 0.8$	
$\theta \leq t^*$	$u(L,t) = 2\, q\, l\, k^{-1}\, \pi^{-1}\, \pi^{-\frac{1}{2}}\, \theta^{\frac{1}{2}}$	
$t^* \leq \theta$	$u(L,t) = 2\, q\, k^{-1}\, \pi^{-2} (\theta + 3^{-1}\, \pi^2 - 2 \sum_{n=1}^{3} n^{-2} \exp(-\theta\, n^2))$	

Note that $\theta = \alpha\, \pi^2\, t\, L^{-2}$ is dimensionless.

$$T_e(x,t_i) = T(x,t_i) + w_i$$

where the ARIMA(1,0,1) errors are computed from

$$w_i = \emptyset\, w_{i-1} + \sigma\, v_i - \sigma\, \theta\, v_{i-1}. \qquad (C.8)$$

The unit normal variates, v_i, are generated numerically.

Several methods are commonly used to generate the unit normal variates. Gallant and Goebel (1976) generated uniform variates with computer program RANDU and transformed these to normal variates by a Box-Muller procedure. The transformation in the Box-and-Muller procedure was stated by Muller (1959) in the following form

$$v_i = (-\,2\,\ln(u_1)\,)^{\frac{1}{2}} \cos(2\,\pi\,u_2), \text{ and} \qquad (C.9.A)$$

$$v_2 = (-\,2\,\ln(u_1)\,)^{\frac{1}{2}} \sin(2\,\pi\,u_2). \qquad (C.9.B)$$

The u-variates are uniform on the interval (0,1) and the v-variates are unit normal.

As discussed by Niederreiter (1976) pseudo-random uniform variates can be generated from a recursive integer equation

$$R_i = K\, R_{i-1} \quad \text{MODULO}(M)$$

which can be used to obtain uniform variates given by

$$u_i = M^{-1}\, R_i. \qquad (C.10)$$

In this generator I used the values $K = 5^{15}$ and $M = 2^{47}$. The values were generated on a Control Data Corporation 6500 series computer.

Although the transformation in Equation (C.9) is exact, the v-variates may not pass tests for normality when the u-variates are pseudo-random numbers. I computed the v-variates using a form of Equation (C.9) stated by Ahrens and Dieter (1972) that avoids the evaluation of the trigonometric terms. I found that the resulting generator of the v-variates had an overall goodness of fit of 954 compared to a value of 915 for the single generator used by Golder and Settle (1976); in this test the expected goodness of fit was set at 990.

REFERENCES

REFERENCES

Ahmadi, G. (1974) Heat conduction in solids with random initial conditions, J. Heat Transfer, 96 474-477.

Ahrens, J. H. and U. Dieter (1972) Computer methods for sampling from the exponential and normal distribution, Communications of ACM, 15 873-882.

Aitken, A. C. (1935) On least squares and linear combinations of observations, Proceedings of Royal Society of Edinburgh, 55 43-48.

Akaike, H. (1972) Use of an information theoretic quantity for statistical model identification, Fifth International Conference on Systems Science, Western Periodicals Publishers, 249-250.

Anderson, O. D. (1975) Time Series Analysis and Forecasting: The Box-Jenkins Approach, Buttersworth, 174 pp.

Anderson, O. D. (1976) On the transformation of raw time series data, Statistische Hefte, 17 285-289.

Barbella Jr., P. F. (1970) Errors in analog to digital converters, M. S. Thesis at Lowell Technological Institute, 63 pp.

Bard, Y. (1970) Comparison of gradient methods for the solution of nonlinear parameter estimation problems, SIAM J. Numerical Analysis, 7 157-186.

Bard, Y. (1974) Nonlinear Parameter Estimation, Academic Press, 341 pp.

Beck, J. V. (1964) The optimum analytical design of transient experiments for simultaneous determination of thermal conductivity and specific heat, Ph.D. Thesis at Michigan State University, 118 pp.

Beck, J. V. (1968) Surface heat flux determination using an integral method, Nuclear Engineering and Design, 7 170-178.

Beck, J. V. (1970) Sensitivity coefficients utilized in nonlinear estimation with small parameters in a heat transfer problem, J. Basic Engineering, 92 215-222.

Beck, J. V. and K. J. Arnold (1974) Parameter Estimation in Science and Engineering, Mechanical Engineering Department, Michigan State University.

Beck, J. V. (1974) Parameter estimation with cumulative errors, Technometrics, 16 85-92.

Beck, J. V. and S. Al-Araji (1974) Investigation of a new simple transient method of thermal property measurement, J. Heat Transfer, 96 59-64.

Beck, J. V. (1975) Discussion: Development and evaluation of a remote sensing technique for determining the temperature distribution in semi-transparent solids by R. E. Chupp and R. Viskanta, J. Heat Transfer, 97 158.

Beck, J. V. and K. J. Arnold (1977) Parameter Estimation in Science and Engineering, John Wiley, 553 pp.

Blum, M. (1961) A stagewise parameter estimation procedure for correlated data, Numerische Mathematik, 3 202-208.

Boothroyd, J. (1968) The symmetric matrix eigenproblem: Jacobi's method revisited, The Australian Computer J., 1 86-94.

Box, G. E. P. and G. M. Jenkins (1970) Time Series Analysis: Forecasting and Control, Holden-Day, 553 pp.

Carpenter, O. (1950) Note on the extension of Craig's theorem to non-central variates, Annals of Mathematical Statistics, 21 455-457.

C-rr, J. (1972) A suggestion for the treatment of serial correlation: A case in point, Canadian J. Economics, 5 301-306.

Carslaw, H. S. and J. C. Jaeger (1947) Conduction of Heat in Solids, Oxford Press.

Chambers, J. M. and J. E. Ertel (1975) Distribution of nonlinear estimates, American Statistical Association Proceedings on Statistical Computing, 51-56.

Chan, C. W., C. J. Harris and P. E. Wellstead (1975) Model order testing technique in parameter estimation, Proceedings IFAC Sixth World Congress, Paper 18.4.

Chen, C. J. and T. M. Danh (1976) Transient temperature distortion in a slab due to thermocouple cavity, AIAA Journal, 14 979-981.

Chow, J. C. (1972) On the estimation of the moving-average parameters, IEEE Transactions on Automatic Control, 17 268-269.

Clarke, R. T. (1973) A review of some mathematical models used in hydrology with observations on their calibration and use, J. Hydrology, 19 1-20.

Deutsch, R. (1965) Estimation Theory, Prentice-Hall, 261 pp.

Dhrymes, P. J. (1971) Equivalence of iterative Aitken and maximum likelihood estimators for a system of regression equations, Australian Economic Papers, 10 20-24.

Dickey, D. A. and F. A. Fuller (1976) Distribution of first order autoregressive estimators, American Statistical Association Proceedings on Business and Economic Statistics, 278-281.

Draper, N. R. and H. Smith (1968) Applied Regression Analysis, John Wiley, 407 pp.

Draper, N. R. and D. E. Shaw (1974) The effects of nonlinearity in regression models: A procedure for examination of the linear approximation, University of Wisconsin Mathematics Research Center Report 1408, 43 pp.

Dubrulle, A. (1970) A short note on the implicit QL algorithm for symmetric tridiagonal matrices, Numerishche Mathematik, 14 450.

Farebrother, R. W. and G. Berry (1974) A remark on algorithm AS-6: Triangular decomposition of a symmetric matrix, Applied Statistics, 23 477.

Farnia, K. (1976) Computer-assisted experimental and analytical study of time/temperature-dependent thermal properties of aluminum alloy 2024-T351, Ph.D. Thesis at Michigan State University, 183 pp.

Frank, I. (1962) An application of the Euler-MacLaurin sum formula to operational mathematics, Quarterly of Applied Mathematics, 20 89-91.

Gallant, A. R. and J. J. Goebel (1976) Nonlinear regression with autocorrelated errors, J. American Statistical Association, 71 961-967.

Gautschi, W. (1962) Diffusion functions for small arguments, SIAM Review, 4 227-229.

Gleser, L. J. and I. Oklin (1972) Estimation for a regression model with an unknown covariance matrix, Berkeley Symposium on Mathematical Statistics and Probability, 6 541-568.

Goebel, J. J. (1974) Nonlinear regression in the presence of autocorrelated errors, Ph.D. Thesis at Iowa State University, 213 pp.

Golder, E. R. and J. G. Settle (1976) The Box-Muller method for generating pseudo-random normal deviates, Applied Statistics, 25 12-20.

Goldstein, G. B. and P. Swerling (1970) Bootstrapping the generalized least squares estimator in colored gaussian noise with unknown covariance parameters, IEEE Transactions on Information Theory, 16 385-392.

Healy, M. J. R. (1968) Algorithm AS-6: Triangular decomposition of a symmetric matrix, Applied Statistics, 17 195-197.

Heuvel, H. M., R. Huisman and K. C. J. B. Lind (1976) Quantitative information from x-ray diffraction of nylon-6 yarns: Development of a model for the analytical description of equatorial x-ray profiles, J. Polymer Science: Polymer Physics, 14 921-940.

Hsu, D. (1973) Stochastic instability and stock prices, Ph. D. Thesis at University of Wisconsin, 191 pp.

Ivanov, A. V. (1972) Asymptotic behavior of least squares estimators in the case of nonlinear regression, SIAM Theory of Probability and Mathematical Statistics, 7 67-70.

Jones, R. H. (1975) Fitting autoregression, J. American Statistical Association, 70 590-592.

Kedem, B. (1976) Exact maximum likelihood estimation of the parameter in the AR(1) process after hard limiting, IEEE Transactions on Information Theory, 22 491-493.

Kenward, L. R. (1976) A new test for residual randomness in a class of dynamic autocorrelated econometric models, Canadian J. Statistics, 4C 41-64.

Khatri, C. G. (1978) A remark on the necessary and sufficient conditions for a quadratic form to be distributed as chi-squared, Biometrika 65 239-240.

Kmenta, J. (1971) Elements of Econometrics, MacMillan, 655 pp.

Landers, T. E. and R. T. Lacoss (1977) Some geophysical applications of autoregressive spectral estimates, IEEE Transactions on Geoscience Electronics, 15 26-32.

Louter, A. S. and J. Koerts (1970) On the Kuiper test for normality with mean and variance unknown, Statistica Neerlandica, 2 83-87.

MacGregor, J. F. (1976) Optimal choice of sampling interval for discrete process control, Technometrics, 18 151-160.

MacGregor, J. F. (1977) Erratum: Optimal choice of sampling interval for discrete process control, Technometrics, 19 224.

Maddala, G. S. (1971) Generalized least squares with an estimated variance covariance matrix, Econometrica, 39 23-33.

Magness, T. A. and J. B. McGuire (1962) Comparison of least squares and minimum variance estimates of regression parameters, Annals of Mathematical Statistics, 33 462-470.

Mandel, J. (1957) Fitting a straight line to certain types of cumulative data, J. American Statistical Association, 52 552-566.

Mandel, J. (1964) Answer to Query 3: Regression analysis of cumulative data, Technometrics, 6 225-227.

Martin, R. S. and J. H. Wilkinson (1968) The implicit QL method, Numerische Mathematik, 12 377-383.

Martin, R. S., C. Reinsch and J. H. Wilkinson (1968) Householder's tridiagonalization of a symmetric matrix, Numerische Mathematik, 11 181-185.

Muller, M. E. (1959) A comparison of methods for generating normal deviates on digital computers, ACM Journal, 6 376-383.

Nelson, C. R. (1974) The first-order moving average process: Identification estimation and prediction, J. Econometrics, 2 121-141.

Newbold, P. (1974) The exact likelihood function for a mixed autoregressive-moving average process, Biometrika, 61 423-426.

Nierderreiter, H. (1976) On the distribution of pseudo-random numbers by the linear congruential method, Mathematics of Computation, 30 571-597.

Odeh, R. E. and J. O. Evans (1974) The percentage points of the normal distribution, Applied Statistics, 23 96-97.

Odell, P. L. and T. O. Lewis (1971) Best linear recursive estimation, J. American Statistical Association, 66 893-896.

Ozaki, T. (1975) On ARIMA model identification, Eighth International Conference on Systems Science, Western Periodicals, 221-223.

Ozaki, T. and H. Tong (1975) On fitting of nonstationary autoregressive models in time series analysis, Eighth International Conference on Systems Science, 224-226.

Pagan, A. (1974) A generalized approach to the treatment of autocorrelation, Australian Economic Papers, 23 267-280.

Pagano, M. (1972) An algorithm for fitting autoregressive schemes, Applied Statistics, 21 274-281.

Pandey, J. N. and M. Rahman (1971) The maximum likelihood estimate of the noncentrality parameter of a noncentral F-variate, SIAM J. Mathematical Analysis, 2 269-276.

Pandit, S. (1973) Data dependent systems: Modeling analysis and optimal control via time series, Ph.D. Thesis at University of Wisconsin, 488 pp.

Pfeiffer, L. and C. P. Lichtenwalner (1974) Systematic error from nonlinear least squares computer fitting of Lorentzian lines, Review of Scientific Instruments, 45 803-808.

Pierce, D. A. (1971) Least squares estimation in regression models with autoregressive moving average errors, Biometrika, 58 299-321.

Rosser, J. B. C. Lanczos, M. R. Hestenes and W. Karush (1951) Separation of close eigenvalues of a real symmetric matrix, J. Research National Bureau of Standards, 47 291-297.

Rutishauser, H. (1963) Algorithm 150: SYMINV2, Communications of ACM, 6 67-68.

Sahai, H. and W. O. Thompson (1974) Comparison of approximations to the percentiles of the t, chi-square, and F distributions, J. Statistical Computation and Simulation, 3 81-93.

Schmidt, P. (1970) Regression analysis of a distributed lag model with second order autoregressive disturbances, Ph.D. Thesis at Michigan State University, 104 pp.

Schmidt, P. (1971) Estimation of a distributed lag model with second order autoregressive disturbances: A Monte Carlo experiment, International Economic Review, 12 372-380.

Schmidt, P. (1972) A generalization of the Durbin-Watson test, Australian Economic Papers, 11 203-209.

Selvin, S. and J. T. Wong (1975) An algorithm for calculating probabilities for the F distribution, Behavior Research Methods & Instrumentation, 7 482-483.

Severo, N. C. and M. Zelen (1960) Normal approximation to the chi-square and non-central F probability functions, Biometrika, 47 411-416.

Shanbhag, D. N. (1970) On the distribution of a quadratic form, Biometrika, 57 222-223.

Shibata, R. (1976) Selection of the order of an autoregressive model by Akaike's information criterion, Biometrika, 63 117-126.

Subba-Rao, T. (1969) On estimation of time dependent parameters of nonstationary time series models, International Statistical Institute Bulletin, 43 362-364.

Theil, H. (1971) Principles of Econometrics, John Wiley, 686 pp.

Tierney, D. (1971) Series expansions with applications to nonlinear models, Ph.D. Thesis at University of Wisconsin, 147 pp.

Tiku, M. L. (1966) A note on approximating the non-central F distribution, Biometrika, 53 606-610.

Vakhaniya, N. N. (1967) A probabilistic problem for the one dimensional heat equation, Theory of Probability and Its Applications, 12 666-667.

Van Fossen, Jr., G. J. (1973) Model building incorporating discrimination between rival mathematical models in heat transfer, PH.D. Thesis at Michigan State University, 164 pp.

Wei, W. W. S. (1977) Asymptotic behavior of the sub-series obtained through a systematic sampling from ARIMA processes, American Statistical Association Proceedings on Business and Economic Statistics, 335-336.

Wise, J. (1956) Stationary conditions for stochastic processes of the autoregressive and moving average type, Biometrika, 43 215-219.