Multivariate Weighted Least Squares as a Preferable Alternative to the Determinant Criterion for Multiresponse Parameter Estimation

by

Paul William Oxby

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemical Engineering

Waterloo, Ontario, Canada, 1997

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

0-612-22225-X

Canada

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

Box and Draper's (1965) determinant criterion for multiresponse parameter estimation is commonly used in preference to ordinary least squares when the measurement error covariance matrix is unknown. Phillips (1976) has shown that the determinant criterion is numerically equivalent to an iterated generalized least squares scheme. From this equivalence, it is shown that, of all such weighting schemes, the determinant criterion in a certain sense minimizes the estimated parameter variances. However, when the number of sets of measurements is not large relative to the number of responses, Monte-Carlo simulation reveals that a multivariate weighted least squares scheme can give parameter variances that are smaller than those given by the determinant criterion. This suggests that the optimality property of the determinant criterion cited above is only asymptotically valid. Monte-Carlo simulation also reveals that, in contrast to multivariate weighted least squares, the determinant criterion can yield parameter estimates whose frequency distribution is very far from normal in the tails. Multivariate weighted least squares (MWLS) is therefore recommended as a robust alternative to the determinant criterion for multiresponse parameter estimation.

Acknowledgments

The loving support of my family, the patient support of my thesis advisors, Professors Tom Duever and Park Reilly, and financial support from the Dow Chemical Company and NSERC are gratefully acknowledged.

Table of Contents

Chapter 1 - Introduction

Chemical engineers routinely deal with mathematical models of physio-chemical processes. These models frequently contain parameters that must be estimated by a statistical analysis of experimental data. The statistical techniques trace their history to the development of the method of least squares by Gauss and Legendre at the beginning of the nineteenth century. The method of least squares is commonly applied where one variable, the response or dependent variable, is considered as a function of one or more independent variables. But mathematical models of interest to chemical engineers often consist of systems of equations where several response or dependent variables are considered as functions of one or more independent variables. This multiresponse parameter estimation problem is complicated by the fact that different equations may share common parameters. However, the complication of shared parameters offers an opportunity in that the model parameters can, in principle, be better estimated from multiple responses than from single responses.

Two approaches to multiresponse parameter estimation are currently popular. A two-step generalized least squares method, based on Zellner's (1962) work, is popular with econometricians (see for example Davidson & MacKinnon (1993)). Econometricians often use Zellner's term "seemingly unrelated regressions" to refer to the multiresponse parameter estimation problem. The determinant criterion for multiresponse parameter estimation, popular with engineers, was developed by Box and Draper (1965). Phillips (1976) has shown that the determinant criterion is equivalent to an iterated generalized least squares scheme. The two-step generalized least squares method can be regarded as the first iteration of this scheme with the appropriate initialization. Because of this close relationship between the determinant criterion and Zellner's two-step generalized least squares method, the latter method will not be considered in this thesis.

Since Box and Draper's (1965) introduction of the determinant criterion, other investigators have shown that the determinant criterion gives unreliable parameter estimates in cases where the measurement error covariance matrix is inherently singular. In this thesis, I extend this work by showing that the determinant criterion can give unreliable parameter estimates even when the error covariance matrix is not inherently singular. For cases where the number of sets of measurements is not large or the measurement error structure does not satisfy certain idealized assumptions, I am proposing the use of an alternative multiresponse parameter estimation method, multivariate weighted least squares (MWLS), that appears to be relatively robust in the sense that it gives reliable parameter estimates in cases where the determinant criterion does not.

One very important point of terminology will be clarified here to avoid confusion. The terms 'weighted least squares' and 'generalized least squares' do not have universally accepted definitions in the literature. These terms will be defined here with reference to the quadratic form (1.0.1), where z is a vector of deviations and A is a symmetric matrix:

$$z^T A z \qquad (1.0.1)$$

In this thesis, ***weighted*** least squares refers to cases where A is a diagonal matrix and ***generalized*** least squares refers to cases where A is not a diagonal matrix. When A is a diagonal matrix it will be referred to as a weight matrix and when A is not diagonal it will be referred to as a generalized weight matrix. (For the sake of completeness, it is a commonly accepted definition that ***ordinary*** least squares refers to cases where A is an identity matrix and the quadratic form reduces to $z^T z$.)

1.1 Literature Review

A review article by Stewart et al. (1992) attributes the first treatment of multiresponse parameter estimation to Gauss early in the nineteenth century. Modern treatments of the subject originate with Aitken's (1935) generalized least squares method. This method is applicable to multiresponse parameter estimation when the measurement error covariance matrix, Σ_ε, is known and the measurement error structure satisfies certain assumptions. In this case, the parameter estimates minimize the generalized sum of squares of deviations:

$$\mathrm{tr}\left[Z(\theta)^T \Sigma_\varepsilon^{-1} Z(\theta) \right] \qquad (1.1.1)$$

where Z is the m by n matrix of deviations:

$$Z(\theta) = y - f(x, \theta) \qquad (1.1.2)$$

m is the number of response variables, n is the number of sets of measurements, θ is the vector of model parameters, y is the m by n matrix of measurements, f is the matrix function representing the model, and x is the vector of independent variables.

In most practical applications, the error covariance matrix is unknown and expression (1.1.1) cannot be evaluated. An alternative to expression (1.1.1) is to use an estimate, $\hat{\Sigma}_\varepsilon$, of the error covariance matrix giving:

$$\mathrm{tr}\left[Z(\theta)^T \hat{\Sigma}_\varepsilon^{-1} Z(\theta) \right] \qquad (1.1.3)$$

The error covariance matrix can be estimated by fitting the model parameters to each of the individual response variables using ordinary least squares. A vector of residuals is calculated for each fit to each response variable. The residual covariance matrix is calculated from all of the vectors of residuals. This matrix is taken as an estimate, $\hat{\Sigma}_\varepsilon$, of

the measurement error covariance matrix. Note that although each separate fit to the individual response variables will give a different set of parameter estimates, it is the residuals that are of interest, not the parameter estimates. Once the error covariance matrix is estimated, the second step of the method applies Aitken's generalized least squares in the usual manner. This approach was first proposed by Zellner (1962) for linear models and extended by Beauchamp and Cornell (1966) to nonlinear models. Gallant (1975) showed that Beauchamp and Cornell's (1966) multivariate approach could be transformed to a univariate one.

The two-step generalized least squares method suggests an extension. The parameters estimated from the generalized least squares step can be used to calculate residuals from which the error covariance matrix is reestimated:

$$\hat{\Sigma}_c = Z(\hat{\theta})\, Z(\hat{\theta})^T / n \qquad (1.1.4)$$

With this equation, the estimate of the error covariance matrix depends on the parameter estimates and the parameter estimates depend on the estimate of the error covariance matrix through expression (1.1.3). Phillips (1976) has shown that the mutual satisfaction of these two conditions, which in practice is achieved through iteration to convergence, is numerically equivalent to Box and Draper's (1965) determinant criterion.

Box and Draper (1965) used a Bayesian argument to derive the determinant criterion for multiresponse parameter estimation. Their derivation assumes that the response measurement errors are normally distributed with zero mean. Errors of different measured responses within a measurement set are assumed to be correlated. (By 'measurement set' I mean a vector of measured responses for a fixed value of the independent variable.) The unknown error covariance matrix is assumed to be constant. Errors between measurement sets are assumed to be independent. With these assumptions and a noninformative (Jeffreys) prior on the error covariance matrix, the

elements of the error covariance matrix can be integrated out of the posterior probability density function giving the following marginal posterior probability density function for the model parameters:

$$p(\theta|y) \propto \left|Z(\theta)\,Z(\theta)^T\right|^{-n/2} \qquad (1.1.5)$$

Point estimates of the parameters can be obtained by maximizing this function with respect to the parameters. This is equivalent to minimizing the determinant of the deviation covariance matrix, $\left|Z(\theta)\,Z(\theta)^T/n\right|$, with respect to the parameters, θ. Bard (1974) has shown that this determinant criterion is also a maximum likelihood criterion under the assumption of normality in the measurement error distribution.

Box et al. (1973) assert that application of the determinant criterion will lead to serious difficulties if the elements of the measurement vectors are linearly dependent. They claim that such a dependency will cause the deviation covariance matrix to be singular. McLean et al. (1979) clarify this issue by pointing out that a linear dependency in the data must also be implicit in the model to cause a singularity due to a linear dependency in the deviations. The argument is straightforward:

$$\text{given} \quad a y_j = b \qquad (1.1.6a)$$

$$\text{then} \quad a\left[y_j - f_j(x,\theta)\right] = 0 \qquad (1.1.6b)$$

$$\text{if and only if} \quad a f_j(x,\theta) = b \qquad (1.1.6c)$$

To remedy the singularity problem, Box et al. propose a method based on an eigenvalue analysis of the data matrix. Khuri (1990) shows that this remedy is scale dependent and gives an method for handling the scaling problem. The singularity problem in the context of the two-step generalized least squares method has been discussed by Takada et al. (1995).

The method of multivariate weighted least squares (MWLS) introduced in this thesis consists of two steps iterated to convergence. In the first step, a weighted sum of squares is minimized with respect to the model parameters:

$$\mathrm{tr}\left[Z(\theta)^T W Z(\theta) \right] \tag{1.1.7}$$

Here W is a diagonal matrix where each element of the diagonal corresponds to a response. In the second step, the reciprocals of the diagonal elements of the residual covariance matrix are assigned to the diagonal elements of W:

$$W = \left[\mathrm{diag}\left(Z(\hat{\theta}) Z(\hat{\theta})^T / n \right) \right]^{-1} \tag{1.1.8}$$

Carroll and Ruppert (1988) discuss an iteratively weighted least squares scheme in the context of uniresponse models where the measurement errors are assumed to be independent but heteroscedastic. In this case, the weights are also determined from the residuals. Of course, the standard multiresponse parameter estimation problem can be thought of in terms of a heteroscedastic error model. However, no literature reference to an iteratively weighted least squares scheme applied to the multiresponse parameter estimation problem has been found.

A few complete examples of applications to data from chemical kinetics can be found in the literature. Box and Draper (1965) discuss an example with three responses. The purpose of this example is to show that multiresponse parameter estimation applied to all of the responses gives much better parameter estimates than does uniresponse parameter estimation applied to any of the individual responses. Box et al. (1973) consider an example with five responses based on data from Fuguitt & Hawkins (1947). In their textbooks, Bates and Watts (1988) and Seber and Wild (1989) repeat the analysis of Box et al. (1973) apparently without appreciating the relevance of the paper by McLean et al. (1979). This example will be considered in this thesis in some detail.

A particularly challenging four-response problem that originated with the Dow Chemical Company is described by Biegler et al. (1986). What makes this problem challenging is the fact that the model is made up of a stiff system of mixed differential and algebraic equations. Unfortunately, the computational burden associated with solving this system of equations makes this problem unsuitable as the basis for a simulation study. In all of the examples cited thus far the independent variable is time. In her recent PhD research, Burke (1994) did copolymerization experiments yielding six responses where the independent variable is feed composition. This example will also be considered in this thesis in some detail.

In this thesis, the issue of the number of degrees of freedom (defined in §3.6) associated with multiresponse parameter estimation will be considered. If the number of response variables is m, the number of sets of measurements is n, and the number of model parameters is p, then Bard (1974) takes the number of degrees of freedom, ν, to be:

$$\nu = n - p/m \qquad (1.1.9)$$

Bates and Watts (1985) and Kang and Bates (1990) argue for the traditional equation:

$$\nu = n - p \qquad (1.1.10)$$

I will show that neither equation can be generally applicable but that equation (1.1.9) can be considered to be a reasonable approximation.

Some of the results given in this thesis require a fair amount of matrix algebra. Searle's (1982) book on the subject was very useful. Selected routines from the collection of numerical recipes by Press et al. (1992) were used in the coding of the simulation programs.

Chapter 2 - Research Motivation

The research leading to this thesis began with an investigation of chemical engineering applications of the Gibbs sampler, a technique for randomly sampling from probability density functions. In the course of this investigation, the determinant criterion for multiresponse parameter estimation was applied to an estimation problem originating with a PhD thesis by Burke (1994). In this examination of Burke's estimation problem, an anomaly in the results from the determinant criterion was observed. This anomaly is described in §2.1.

To check if this anomaly was specific to this particular problem, another multiresponse estimation problem described by Box et al. (1973) was examined. This examination revealed another anomaly of a quite different character, as described in §2.2. These two anomalous results associated with the determinant method motivated the research described in this thesis.

2.1 Burke's Multiresponse Parameter Estimation Problem

In her thesis, Burke (1994) considered copolymerization reactions. In the modeling of these reactions, the two reactants can be denoted by subscripts 1 and 2. $R_{n,i}$ denotes a chain of n monomer units ending in the free radical monomer unit i where i is 1 or 2. M_j denotes the monomer j. The four polymer chain propagation reactions are:

$$R_{n,1} + M_1 \xrightarrow{k_{11}} R_{n+1,1} \qquad (2.1.1a)$$

$$R_{n,1} + M_2 \xrightarrow{k_{12}} R_{n+1,2} \qquad (2.1.1b)$$

$$R_{n,2} + M_1 \xrightarrow{k_{21}} R_{n+1,1} \qquad (2.1.1c)$$

$$R_{n,2} + M_2 \xrightarrow{k_{22}} R_{n+1,2} \qquad (2.1.1d)$$

where the k_{ij} are kinetic rate constants for the addition of monomer j to a chain ending in radical monomer unit i. Of interest are the reactivity ratios r_1 and r_2 defined as:

$$r_1 = \frac{k_{11}}{k_{12}} \tag{2.1.2a}$$

$$r_2 = \frac{k_{22}}{k_{21}} \tag{2.1.2b}$$

The reactivity ratios can be determined from triad fractions, the fractions of sequences of three consecutive monomer units in the copolymer chain. These fractions are denoted by A_{ijk} where i, j and k are monomer units 1 or 2. This gives eight triad fractions but A_{211} and A_{112} are indistinguishable as are A_{221} and A_{122}, leaving six distinct triad fractions. The relationships between the reactivity ratio r_1 and the three monomer-1 centered triad fractions are expressed in the following equations given by Koenig (1980):

$$d_1 = r_1^2 f_1^2 + 2 r_1 f_1 f_2 + f_2^2 \tag{2.1.3a}$$

$$A_{111} = r_1^2 f_1^2 / d_1 \tag{2.1.3b}$$

$$A_{112} = 2 r_1 f_1 f_2 / d_1 \tag{2.1.3c}$$

$$A_{212} = f_2^2 / d_1 \tag{2.1.3d}$$

where f_1 and f_2 are the feed fractions of monomer reactants 1 and 2 so $f_1 + f_2 = 1$. Similarly, for reactivity ratio r_2 and the three monomer-2 centered triad fractions we have:

$$d_2 = r_2^2 f_2^2 + 2 r_2 f_2 f_1 + f_1^2 \tag{2.1.4a}$$

$$A_{222} = r_2^2 f_2^2 / d_2 \tag{2.1.4b}$$

$$A_{122} = 2 r_2 f_2 f_1 / d_2 \tag{2.1.4c}$$

$$A_{121} = f_1^2 / d_2 \tag{2.1.4d}$$

Triad fractions cannot be measured directly but, for the copolymer styrene-methylmethacrylate they are linearly related to C^{13}-NMR spectral data through the following equations given by Aerdts (1993):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0 & 0 & (1-\sigma_{12})^2 \\ 1 & 1-\sigma_{12} & 2\sigma_{12}(1-\sigma_{12}) \\ 0 & \sigma_{12} & \sigma_{12}^2 \end{bmatrix} \begin{bmatrix} A_{111} \\ A_{112} \\ A_{212} \end{bmatrix} \qquad (2.1.5a)$$

$$\begin{bmatrix} A \\ B+C \\ D \end{bmatrix} = \begin{bmatrix} \sigma_{12}^2 & \sigma_{22}\sigma_{12} & \sigma_{22}^2 \\ 1-\sigma_{12}^2 & 1-\sigma_{22}\sigma_{12} & 2\sigma_{22}(1-\sigma_{22}) \\ 0 & 0 & (1-\sigma_{22})^2 \end{bmatrix} \begin{bmatrix} A_{121} \\ A_{122} \\ A_{222} \end{bmatrix} \qquad (2.1.5b)$$

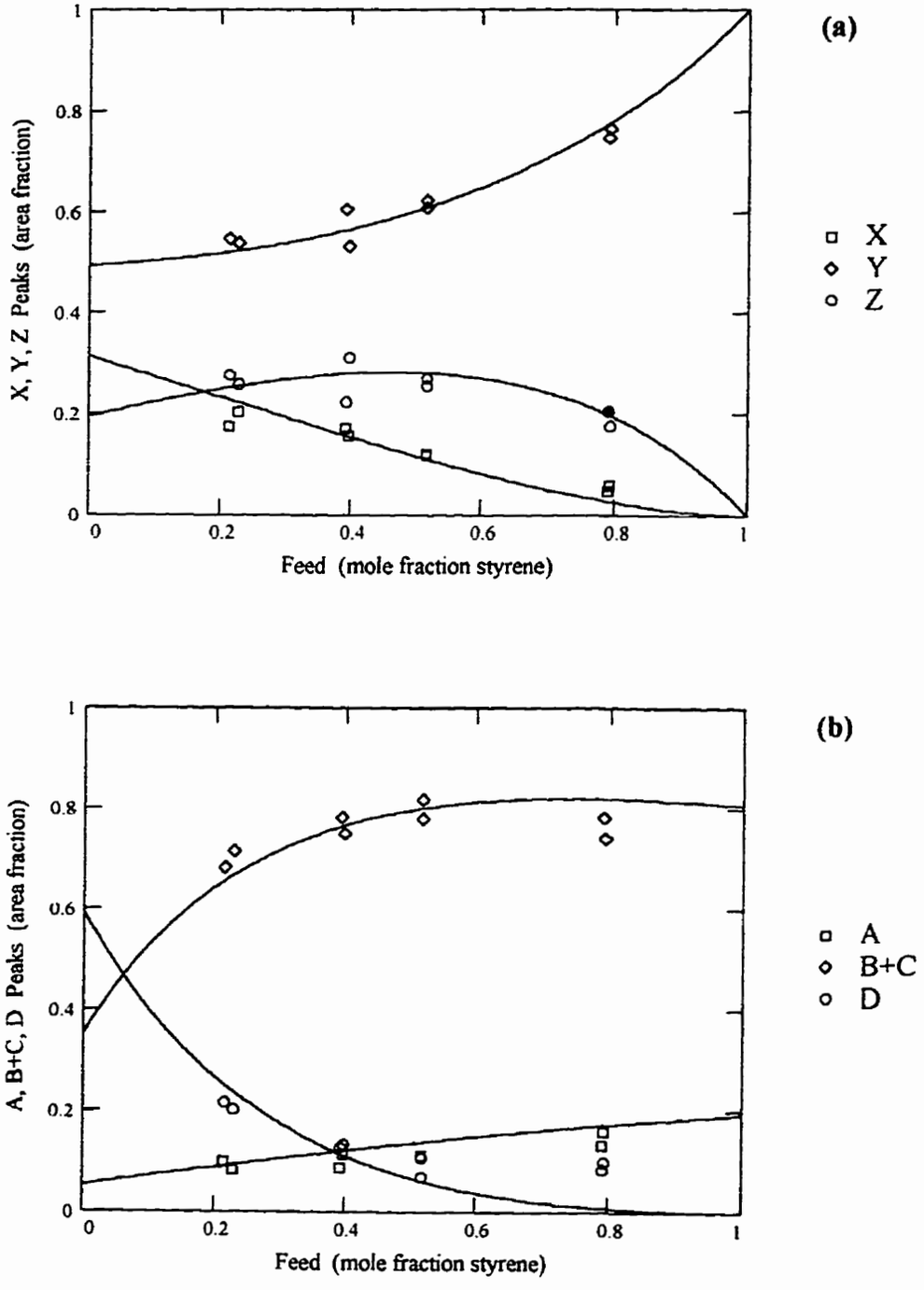
where X, Y, Z, A, B, C, and D are relative peak areas in the C^{13}-NMR spectrum. The fixed parameters $\sigma_{12} = 0.44$ and $\sigma_{22} = 0.23$ relate to the C^{13}-NMR spectral characteristics of styrene-methylmethacrylate and are not standard deviations of any kind. It can be seen from the structure of the coefficient matrices that normalized triad fractions will lead to normalized peak areas because the matrix columns sum to unity.

Burke measured experimental data from which the reactivity ratios r_1 and r_2 can be estimated. The data, as given on page 160 of her thesis, are given in Table 2.1.1 below and plotted in Figure 2.1.1 on the next page.

Table 2.1.1 C^{13}-NMR Data from Burke's Thesis

i	f_1	X	Y	Z	A	B+C	D
1	0.21232	0.17477	0.54829	0.27695	0.09908	0.68361	0.21731
2	0.22490	0.20291	0.53759	0.25950	0.08317	0.71368	0.20315
3	0.39255	0.16945	0.60578	0.22477	0.08787	0.78201	0.13012
4	0.39715	0.15556	0.53333	0.31112	0.11592	0.75007	0.13401
5	0.51690	0.11864	0.62423	0.25713	0.11206	0.81746	0.07048
6	0.51701	0.12033	0.60967	0.27000	0.11255	0.77942	0.10803
7	0.79132	0.04592	0.74733	***0.21676***	0.13330	0.78146	0.08524
8	0.79186	0.05760	0.76525	0.17715	0.15984	0.74077	0.09938

Figure 2.1.1 Multivariate Weighted Least Squares Fit to C^{13}-NMR Data from Burke's Thesis



Because the peak areas are normalized we should have X+Y+Z = A+B+C+D = 1. The value of Z_7 is highlighted in Table 2.1.1 because the normalization check showed a discrepancy subsequently attributed to a typographical error in this entry. The correct value should be 0.20676 instead of 0.21676. The difference of 0.01 provides the basis for a sensitivity check on parameter estimates derived from the data.

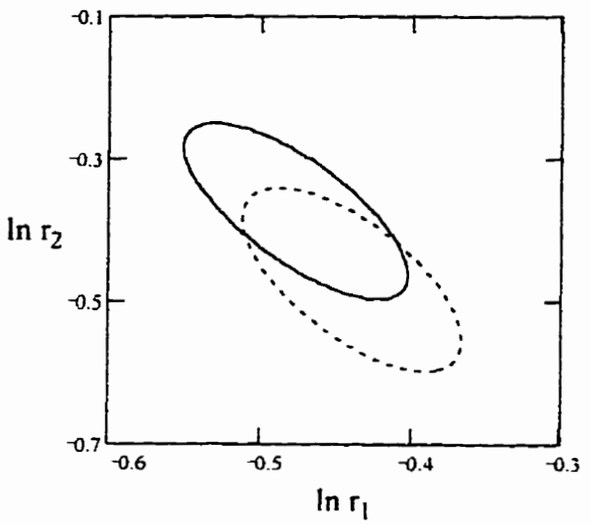
Table 2.1.2 below gives parameter estimates fitted to the data, for both values of Z_7, using the determinant criterion and multivariate weighted least squares (MWLS). The data and the fit from MWLS are graphically shown in Figures 2.1.1(a) and 2.1.1(b) on the previous page. In this figure Z_7 is the shaded circle in Figure 2.1.1(a).

Table 2.1.2 Point Estimates of Parameters ($\ln r_1$, $\ln r_2$)

Z_7	Determinant Criterion	Multivariate Weighted Least Squares (MWLS)
0.20676	(-0.477, -0.373)	(-0.439, -0.682)
0.21676	(-0.440, -0.469)	(-0.447, -0.682)

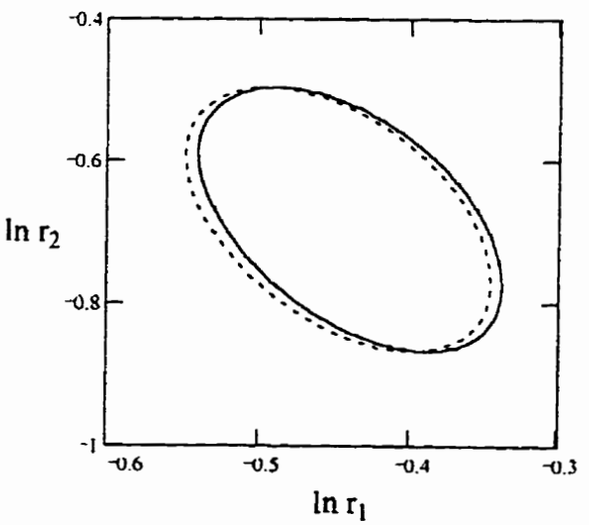
The standard deviation in Z is estimated from the MWLS residuals to be 0.027. This is almost three times the perturbation in Z_7 of 0.01, so one might expect that the perturbation in Z_7 would result in changes in the parameter estimates that are small relative to the parameter uncertainties. Estimates of the parameter uncertainties for the determinant criterion and MWLS are given by equations (3.3.8) and (3.3.6), respectively. These equations are based on a linearized model. Equations (2.1.3) and (2.1.4) are nonlinear in the parameters r_1 and r_2. It happens that the model is more nearly linear with respect to the logarithm of the parameters. Since reactivity ratios cannot be negative, this transformation gives parameter uncertainty estimates that are more realistic. The approximate 50% joint confidence regions for the parameter estimates is determined using equation 3.6.1.

Figure 2.1.2 Approximate 50% Joint Confidence Regions



(a)

Determinant Criterion



(b)

Multivariate Weighted Least Squares

——— $Z_7 = 0.20676$

- - - - - - - - $Z_7 = 0.21676$

Figures 2.1.2(a) and 2.1.2(b) on the previous page give the approximate 50% joint confidence regions for the parameter estimates from the determinant criterion and MWLS, respectively. The contours for the two values of Z_7 are given by the solid (Z_7 = 0.20676) and broken (Z_7 = 0.21676) ellipses. The scales of the two figures differ only by a translation in ln r_2. In comparing the two figures two things stand out. Firstly, the areas of the joint confidence regions from the determinant criterion are about half the areas of the regions from MWLS. Secondly, the shift between the solid and broken ellipses is much greater for the determinant criterion compared to MWLS. (A 50% joint condifence region was chosen instead of 95% to highlight this difference.)

These two facts present something of a paradox. It will be shown in §3.3 that the determinant criterion theoretically should give parameter estimates with a smaller variance than those given by MWLS. If so, then why are the parameter estimates from the determinant criterion much more sensitive to a perturbation in the data? The answer to this question is the subject of this thesis.

The results from the examination of Burke's (1994) multiresponse parameter estimation problem suggested that an examination of a textbook example might be a good idea. This example is discussed in §2.2. But before this example is discussed, a loose end should be tied up. Two linear dependencies in the data in Table 2.1.1, i.e. X+Y+Z = 1 and A+B+C+D = 1, also appear in the model. Box et al. (1973) and McLean et al. (1979) discuss why this will cause difficulty in applying the determinant criterion. The difficulty is avoided by dropping two of the redundant observations. If the roundoff error in the observations is small relative to the experimental error then it does not matter which observations are dropped. Burke dropped the X and D observations.

2.2 The Alpha-Pinene Parameter Estimation Problem

The alpha-pinene estimation problem has been discussed in a paper by Box et al. (1973) and in textbooks by Bates and Watts (1988) and Seber and Wild (1989). The data originated with Fuguitt and Hawkins (1947) who studied the thermal isomerization of alpha-pinene (y_1) to dipentene (y_2) and allo-ocimene (y_3). Allo-ocimene (y_3) in turn yields pyronene (y_4) and a dimer (y_5). The conversion of allo-ocimene (y_3) to the dimer (y_5) is reversible while the other conversions are irreversible. Thus we have the reactions:

$$y_1 \xrightarrow{k_1} y_2 \tag{2.2.1a}$$

$$y_1 \xrightarrow{k_2} y_3 \tag{2.2.1b}$$

$$y_3 \xrightarrow{k_3} y_4 \tag{2.2.1c}$$

$$y_3 \underset{k_5}{\overset{k_4}{\rightleftarrows}} y_5 \tag{2.2.1d}$$

Assuming first-order kinetics, the associated system of differential rate equations can be expressed in matrix format as:

$$\begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \\ \dot{y}_4 \\ \dot{y}_5 \end{bmatrix} = \begin{bmatrix} -k_1 - k_2 & 0 & 0 & 0 & 0 \\ k_1 & 0 & 0 & 0 & 0 \\ k_2 & 0 & -k_3 - k_4 & 0 & k_5 \\ 0 & 0 & k_3 & 0 & 0 \\ 0 & 0 & k_4 & 0 & -k_5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} \tag{2.2.2}$$

where italicized 'y's are used to represent concentration.

Box et al. give the analytical solution of this system of differential equations to be:

$$y_1(t) = y_1(0)\exp(-\phi t) \tag{2.2.3a}$$

$$y_2(t) = y_1(0)\frac{k_1}{\phi}[1-\exp(-\phi t)] \tag{2.2.3b}$$

$$y_3(t) = C_1\exp(-\phi t) + C_2\exp(\beta t) + C_3\exp(\gamma t) \tag{2.2.3c}$$

$$y_4(t) = k_3\left[\frac{C_1}{\phi}[1-\exp(-\phi t)] + \frac{C_2}{\beta}[\exp(\beta t)-1] + \frac{C_3}{\gamma}[\exp(\gamma t)-1]\right] \tag{2.2.3d}$$

$$y_5(t) = k_4\left[\frac{C_1}{k_5-\phi}\exp(-\phi t) + \frac{C_2}{k_5+\beta}\exp(\beta t) + \frac{C_3}{k_5+\gamma}\exp(\gamma t)\right] \tag{2.2.3e}$$

where:

$$\alpha = k_3 + k_4 + k_5 \tag{2.2.4a}$$

$$\beta = (-\alpha + \sqrt{\alpha^2 - 4k_3k_5})/2 \tag{2.2.4b}$$

$$\gamma = (-\alpha - \sqrt{\alpha^2 - 4k_3k_5})/2 \tag{2.2.4c}$$

$$\phi = k_1 + k_2 \tag{2.2.4d}$$

$$C_1 = y_1(0)\,k_2\,(k_5-\phi)/[(\phi+\beta)(\phi+\gamma)] \tag{2.2.4e}$$

$$C_2 = y_1(0)\,k_2\,(k_5+\beta)/[(\phi+\beta)(\beta-\gamma)] \tag{2.2.4f}$$

$$C_3 = y_1(0)\,k_2\,(k_5+\gamma)/[(\phi+\gamma)(\gamma-\beta)] \tag{2.2.4g}$$

and $y_1(0)$ is the initial value of y_1 (i.e. at time $t = 0$). It is assumed that the initial values of the other reactants are all zero (i.e. $y_2(0) = y_3(0) = y_4(0) = y_5(0) = 0$).

The experimental data measured by Fuguitt and Hawkins for the isothermal isomerization of alpha-pinene at 189.5 C are given in Table 2.2.1 below. The data are expressed as percentages so $y_1(0)$ would be 100. The difference between 100 and the sum of values in each row is attributable to roundoff error (except for the third row).

Table 2.2.1 Alpha-Pinene Isomerization Data

t (min)	y_1	y_2	y_3	y_4	y_5	sum
1230	88.35	7.3	2.3	0.4	1.75	100.1
3060	76.4	15.6	4.5	0.7	2.8	100.0
4920	65.1	23.1	5.3	1.1	5.8	100.4
7800	50.4	32.9	6.0	1.5	9.3	100.1
10680	37.5	42.7	6.0	1.9	12.0	100.1
15030	25.9	49.1	5.9	2.2	17.0	100.1
22620	14.0	57.4	5.1	2.6	21.0	100.1
36420	4.5	63.1	3.8	2.9	25.7	100.0

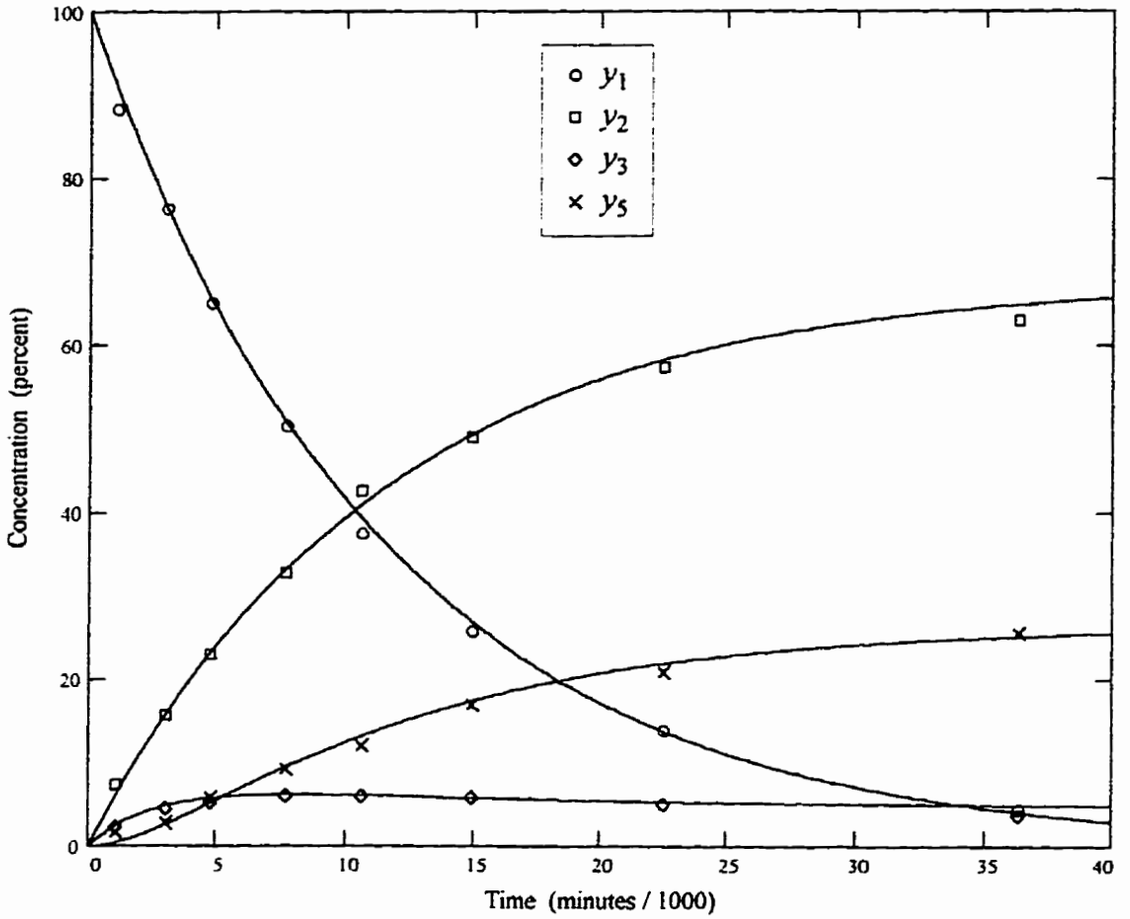
Table 2.2.2 below gives parameter estimates fitted to the data using two variants of the determinant criterion, designated as (3) and (4), and MWLS. The parameter estimates designated as determinant criterion (3) are the best estimates presented in the paper by Box et al (1973). The corresponding fit is shown in Figure 2.2.1 on the next page.

Table 2.2.2 Parameter Estimates for Alpha-Pinene Problem

	Determinant Criterion (3)	Determinant Criterion (4)	MWLS
$k_1 \times 10^5$	5.95	5.84	5.92
$k_2 \times 10^5$	2.84	2.95	2.95
$k_3 \times 10^5$	0.43	2.25	2.05
$k_4 \times 10^5$	31.3	51.3	30.3
$k_5 \times 10^5$	5.74	10.54	4.92

Figure 2.2.1 Determinant Criterion (3) Fit to Alpha-Pinene Data

(a)



(b)

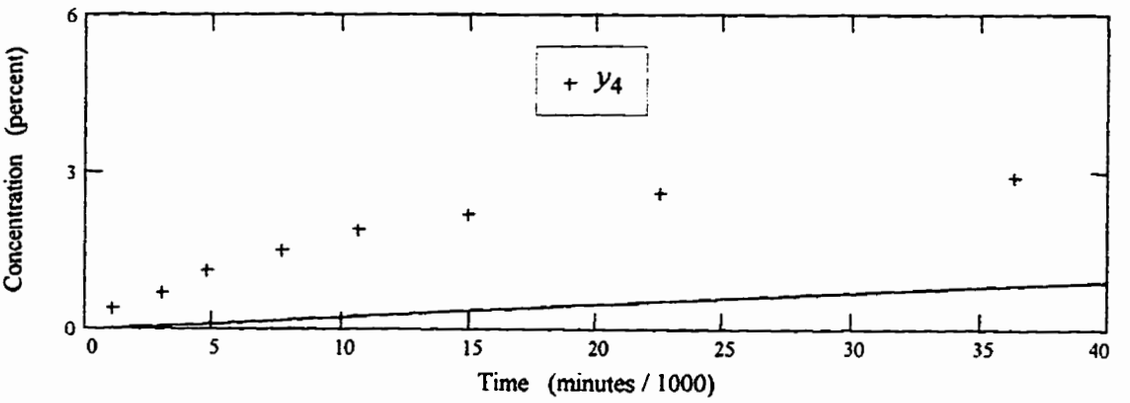


Figure 2.2.1(a) is a reproduction of Figure 2 on page 46 of Box et al. (1973). It shows their best fit to the data using the determinant criterion. This figure is also reproduced in Seber and Wild (1989). In both of these references, there is one common omission. The data and fit for y_4 is missing. This omission is corrected in Figure 2.2.1(b) and the fit for y_4 is terrible.

The reason for the terrible fit for y_4 can be traced to the way the investigators above handled two linear dependencies in the data. The first linear dependency is due to the fact that, in the absence of roundoff error, the measurements sum to $y_1(0)$:

$$y_1(t)+y_2(t)+y_3(t)+y_4(t)+y_5(t)=y_1(0) \qquad (2.2.5)$$

The second linear dependency is due to the fact that Fuguitt and Hawkins (1947) did not actually measure y_4. They synthesized the data for y_4 based on the assumption that y_4 is three percent of the conversion of y_1:

$$y_4(t)=0.03\left[y_1(0)-y_1(t)\right] \qquad (2.2.6)$$

Box et al. (1973), Bates and Watts (1988), and Seber and Wild (1989) all argued that these two linear dependencies in the data would reduce the rank of the measurement error covariance matrix by two from five to three. However, McLean et al. (1979) used the argument of §1.1 to show that a linear dependency in the data will reduce the rank of the error covariance matrix only if the same linear dependency is also implicit in the model. It happens that equation (2.2.5) is implicit in the model, equations (2.2.3) and (2.2.4), but equation (2.2.6) is not implicit in the model. For the first-order kinetics model described by equation (2.2.2) the relationship between y_4 and y_1 cannot be linear.

Box et al. (1973) describe how the determinant criterion should be applied when the rank of the error covariance matrix is less than the number of measured responses. In this case, taking the rank of the error covariance matrix to be three is a mistake which leads to loss of information and a terrible fit for y_4. The rank of the error covariance matrix should be four. In Table 2.2.2 of parameter estimates and Table 2.2.3 of mean squared residuals the designations (3) and (4) refer to the rank of the error covariance matrix used in the analysis.

Table 2.2.3 Mean Squared Residuals for Alpha-Pinene Problem

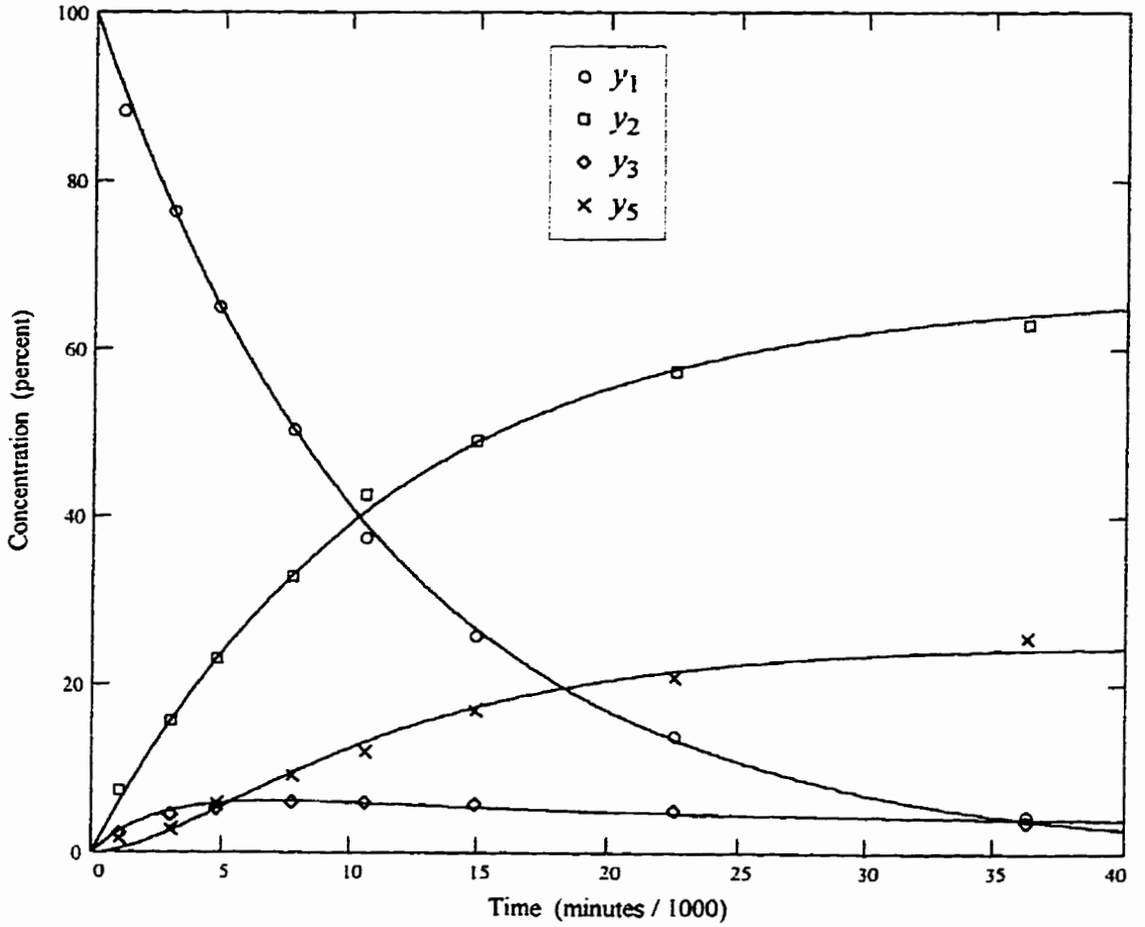
	Determinant Criterion (3)	Determinant Criterion (4)	MWLS
y_1	0.687	0.687	0.567
y_2	1.009	0.738	0.649
y_3	0.271	0.633	0.183
y_4	2.276	0.497	0.369
y_5	0.631	2.342	0.805

Table 2.2.3 shows that taking the rank of the error covariance matrix to be four improves the fit for y_4. But this improvement seems to come at the expense of the fits for y_3 and y_5. Table 2.2.3 also shows that the mean squared residuals for MWLS are all smaller than those for the determinant criterion (4). Since the determinant criterion is generally considered to be a standard technique for multiresponse parameter estimation, this result, taken with the anomalous result of §2.1, provided the motivation behind this thesis.

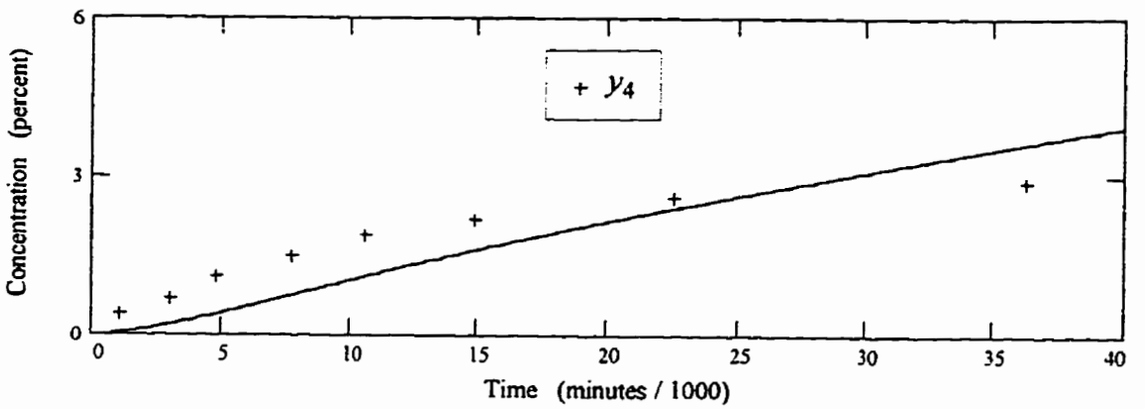
The fit for MWLS is shown in Figure 2.2.2 on the next page. There is still a mismatch between the data and model for y_4 and this simply reflects the fact that equation (2.2.6) used to synthesize the y_4 data cannot be consistent with the model and the rest of the data.

Figure 2.2.2 Multivariate Weighted Least Squares Fit to Alpha-Pinene Data

(a)



(b)



The discussion of the last two paragraphs raises an important issue about goodness of fit. For single response problems, relative goodness of fit is usually established by applying an F test to a ratio of residual sum of squares. The statistical validity of the test is based on the assumption that the measurement errors are independent and of constant variance. But in the multiresponse problem the residuals for different responses are assumed to be correlated and to have different variances. Therefore a simple F test is inappropriate for statistical inference in the multiresponse case.

Cox (1962) developed a generalization of the F test that, for multiresponse inference, uses the ratio of determinants of residual covariance matrices. This test will always favor the determinant criterion which minimizes the determinant of the residual covariance matrix. Given, then, the problematic nature of using goodness of fit as a means for assessing multiresponse parameter estimation methods, another means of assessment will have to be considered. It is usually considered desirable for an estimation method to generate good parameter estimates where 'good' means small biases and small norms of the covariance matrix of the parameter estimates. It is also desirable for the estimation method to yield a good estimate of the covariance matrix of the parameter estimates. Therefore, the quality of the parameter estimates, as determined in a frequentist sense using Monte-Carlo simulation, will be used in this thesis as the basis for assessing alternative multiresponse parameter estimation methods.

Chapter 3 - Theoretical Results

In this chapter a theoretical result that is the foundation of this thesis will be established. It is based on Phillips' (1976) result that the determinant criterion is numerically equivalent to an iterated generalized least squares scheme. This result will be used to show that, of all such iterated schemes, the determinant criterion is theoretically optimal in the sense that it generates parameter estimates with a minimum determinant of the estimated parameter covariance matrix. In addition to this result, two secondary theoretical results will also be established. The first is that multivariate weighted least squares (MWLS) has an equivalent counterpart resembling the determinant criterion. The second result concerns the number of degrees of freedom associated with multiresponse parameter estimation.

3.1 Multiresponse Estimation and the Determinant Criterion

In a multiresponse estimation problem, m dependent variables, $y_1, \ldots y_m$, are associated with one or more independent variables, x. For n measured values of the independent variable, $x_1, \ldots x_n$, there are mn measured values of the dependent variables, $y_{11}, \ldots y_{mn}$. Between the independent variable x and m dependent variables y there are m functional relationships, $f_1, \ldots f_m$, parameterized by a vector θ of p parameters. If the true parameter values are indicated by θ^* then:

$$y_{ij} = f_i(x_j, \theta^*) + \varepsilon_{ij} \qquad i = 1 \ldots m, \quad j = 1 \ldots n \qquad (3.1.1)$$

where ε_{ij} are the measurement errors in y. The measurement errors in x are assumed to be zero. The mn deviations, $y_{ij} - f_i(x_j, \theta)$, regarded as functions of θ, can be assembled into an m by n matrix $Z(\theta)$. (Some investigators define $Z(\theta)$ to be n by m, the transpose of the definition used here. Later the vec operator will be applied to $Z(\theta)$ and, for this purpose, it is convenient to define $Z(\theta)$ to be m by n.)

Box and Draper (1965) showed that if the measurement errors, ε, are random, normally distributed with zero mean, independent between measurement vectors, and correlated within measurement vectors with constant m by m covariance matrix Σ_ε, then the likelihood function for θ and Σ_ε is:

$$L(\theta, \Sigma_\varepsilon) \propto |\Sigma_\varepsilon|^{-n/2} \exp\left[-\tfrac{1}{2} \operatorname{tr}\left(Z(\theta)^T \Sigma_\varepsilon^{-1} Z(\theta) \right) \right] \qquad (3.1.2)$$

Bard (1974) showed that the likelihood function can be maximized with respect to the elements of θ and Σ_ε in two steps. In the first step, θ is held constant and the likelihood is maximized with respect to the elements of Σ_ε. This gives a maximum likelihood estimate of Σ_ε, $\hat{\Sigma}_\varepsilon(\theta)$, that is an implicit function of θ and is just the deviation covariance matrix:

$$\hat{\Sigma}_\varepsilon(\theta) = Z(\theta)\, Z(\theta)^T / n \qquad (3.1.3)$$

In the second step, the elements of Σ_ε are eliminated as parameters in the likelihood function by substituting equation (3.1.3) back into equation (3.1.2), giving the conditional likelihood function:

$$L\left[\theta | \hat{\Sigma}_\varepsilon(\theta)\right] \propto \left| Z(\theta)\, Z(\theta)^T \right|^{-n/2} \qquad (3.1.4)$$

Therefore, it follows that the value of θ that maximizes the likelihood is that which minimizes $\left| Z(\theta)\, Z(\theta)^T \right|$. Box and Draper (1965) first developed this determinant criterion using a Bayesian argument. Bard's derivation is presented here because, in the next section, it will be used to establish the equivalence of the determinant criterion with an iterated generalized least squares scheme. From equation (3.1.3), the maximum likelihood estimate of Σ_ε conditional on a parameter estimate, $\hat{\theta}$, is the residual covariance matrix:

$$\hat{\Sigma}_\varepsilon = Z(\hat{\theta})\, Z(\hat{\theta})^T / n \qquad (3.1.5)$$

3.2 The Determinant Criterion as an Iterated GLS Scheme

The likelihood function, (3.1.2), conditioned on a fixed estimate of the error covariance matrix, Σ_ε, is

$$L(\theta|\hat{\Sigma}_\varepsilon) \propto \exp\left[-\tfrac{1}{2}\,\mathrm{tr}\left(Z(\theta)^T \hat{\Sigma}_\varepsilon^{-1} Z(\theta)\right)\right] \qquad (3.2.1)$$

This can be written in terms of a more familiar quadratic form as

$$L(\theta|\hat{\Sigma}_\varepsilon) \propto \exp\left[-\tfrac{1}{2}\,\underline{z}(\theta)^T \hat{\Sigma}_{\varepsilon n}^{-1}\,\underline{z}(\theta)\right] \qquad (3.2.2)$$

where $\underline{z}(\theta) = \mathrm{vec}(Z(\theta))$ is a vector of length mn made up of the concatenated columns of $Z(\theta)$ and $\hat{\Sigma}_{\varepsilon n} = I_n \otimes \hat{\Sigma}_\varepsilon$ is a block diagonal matrix whose n blocks are $\hat{\Sigma}_\varepsilon$. Equation (3.2.2) has the form of a likelihood function for generalized least squares (GLS) which justifies the following Gauss-Newton equation for determining the value of θ that maximizes this (conditional) likelihood:

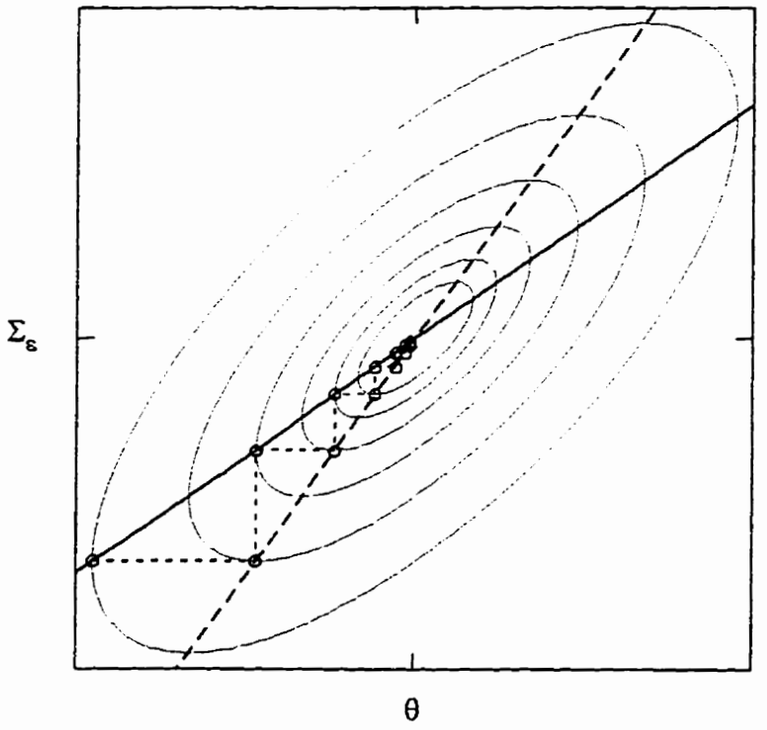
$$\Delta\hat{\theta}^{(k)} = (X^T \hat{\Sigma}_{\varepsilon n}^{-1} X)^{-1} X^T \hat{\Sigma}_{\varepsilon n}^{-1}\,\underline{z}(\hat{\theta}^{(k)}) \qquad (3.2.3)$$

where $X_{ij} = -\partial \underline{z}(\theta)_i / \partial \theta_j$ for $i = 1...mn$ and $j = 1...p$ and k denotes the k th Gauss-Newton iteration.

With the maximum likelihood estimate of θ conditioned on $\hat{\Sigma}_\varepsilon$ from (3.2.3), the maximum likelihood estimate of Σ_ε conditioned on $\hat{\theta}$ can be determined from (3.1.5). When iterated to convergence these two steps yield estimates of θ and Σ_ε that maximize the likelihood function, (3.1.2). This is because the ***partial*** derivatives of a function are zero at the maximum. This iterated GLS scheme is, then, numerically equivalent to the determinant criterion. This result is originally due to Phillips (1976).

The numerical equivalence of the determinant criterion and the iterated GLS scheme can also be explained with a graphical analogy. Figure 3.1.1 below shows the contours of a likelihood function for scalar values of θ and Σ_ε. The solid line is the locus of points that maximize the likelihood with respect to Σ_ε for constant values of θ. This locus is obtained from equation (3.1.5). The determinant criterion gives the value of θ on this locus that maximizes the likelihood with respect to θ and Σ_ε. The dashed line is the locus of points that maximize the likelihood with respect to θ for constant values of Σ_ε. This locus is obtained from equation (3.2.3). The maximum of the likelihood with respect to θ and Σ_ε is at the intersection of the two loci of points represented by the solid and dashed lines. Therefore the dotted 'staircase' of points obtained by iterating equations (3.2.3) and (3.1.5) will converge to the same optimum as the determinant criterion. (For multiple local optima this argument holds only in the neighborhood of a local optimum.)

Figure 3.1.1 Contours of a Likelihood Function



3.3 An Optimality Property of the Determinant Criterion

Equation (3.2.3) can be generalized by replacing the inverse of the estimate of the error covariance matrix, $\hat{\Sigma}_{\varepsilon n}^{-1}$, with an arbitrary symmetric positive definite matrix, W:

$$\Delta\hat{\theta}^{(k)} = (X^T W X)^{-1} X^T W \, z(\hat{\theta}^{(k)}) \qquad (3.3.1)$$

If the measurement errors are random with zero mean and known error covariance matrix Σ_ε, then it follows from equation (3.3.1) that the p by p parameter covariance matrix is estimated by:

$$\hat{\Sigma}_\theta = (X^T W X)^{-1} X^T W \Sigma_{\varepsilon n} W X (X^T W X)^{-1} \qquad (3.3.2)$$

where $\Sigma_{\varepsilon n} = I_n \otimes \Sigma_\varepsilon$. For linear models and constant W, this estimate of Σ_θ is exact. If the determinant of this estimate of the parameter covariance matrix, $\left|\hat{\Sigma}_\theta\right|$, is minimized with respect to the elements of W, the solution is independent of X and is simply:

$$W = \Sigma_{\varepsilon n}^{-1} \qquad (3.3.3)$$

Substitution of equation (3.3.3) into equation (3.3.1) gives a multiresponse version of Aitken's (1935) generalized least squares:

$$\Delta\theta^{(k)} = (X^T \Sigma_{\varepsilon n}^{-1} X)^{-1} X^T \Sigma_{\varepsilon n}^{-1} \, z(\theta^{(k)}) \qquad (3.3.4)$$

Substitution of equation (3.3.3) into equation (3.3.2) gives the estimate of the parameter covariance matrix:

$$\hat{\Sigma}_\theta = (X^T \Sigma_{\varepsilon n}^{-1} X)^{-1} \qquad (3.3.5)$$

Therefore, when the error covariance matrix is known, among all estimation schemes of the form of (3.3.1), equation (3.3.4) has the property of giving the minimum value of the determinant of the estimate of the parameter covariance matrix, $\left|\hat{\Sigma}_\theta\right|$.

If the error covariance matrix, Σ_ε, is unknown, then the next best alternative to equation (3.3.2) is to use an estimate of Σ_ε, $\hat{\Sigma}_\varepsilon$, in equation (3.3.2) giving:

$$\hat{\Sigma}_\theta = (X^T W X)^{-1} X^T W \hat{\Sigma}_{\varepsilon n} W X (X^T W X)^{-1} \qquad (3.3.6)$$

If the determinant of this estimate of the parameter covariance matrix is minimized with respect to the elements of W, the solution is:

$$W = \hat{\Sigma}_{\varepsilon n}^{-1} \qquad (3.3.7)$$

Substitution of equation (3.3.7) into equation (3.3.1) gives equation (3.2.3) which, when iterated with equation (3.1.5), is equivalent to the determinant criterion as shown in §3.2. Substitution of equation (3.3.7) into equation (3.3.6) gives the estimate of the parameter covariance matrix for the determinant criterion:

$$\hat{\Sigma}_\theta = (X^T \hat{\Sigma}_{\varepsilon n}^{-1} X)^{-1} \qquad (3.3.8)$$

Kang and Bates (1990) derived this equation using an argument based on likelihood inference.

Therefore, of all estimation schemes in the form of (3.3.1), it is equation (3.2.3), equivalent to the determinant criterion, that minimizes $\left|\hat{\Sigma}_\theta\right|$, the determinant of the estimate of the parameter covariance matrix, when the error covariance matrix is unknown. (Numerical experiments suggest that equation (3.3.7) minimizes all reasonable norms of $\hat{\Sigma}_\theta$.) This appears to be a rather strong result in support of the determinant criterion because $\left|\hat{\Sigma}_\theta\right|$ is proportional to the square of the estimate of the parameter confidence region hypervolume. It might also be noted that, unlike derivations based on likelihood, this result does not depend on the measurement errors being normally distributed because the validity of equations (3.3.2) and (3.3.6) does not depend on a normality assumption.

But this analysis exposes a serious weakness in the determinant criterion. Ideally one would want the generalized weight matrix, W, to be the inverse of the error covariance matrix, $\Sigma_{\varepsilon\varepsilon}$, as expressed by equation (3.3.3). But the determinant criterion, in effect, makes the compromise of substituting the residual covariance matrix, $Z(\hat{\theta})\,Z(\hat{\theta})^T/n$, for the error covariance matrix. The potential problem here is that, if the data set is not large, the residual covariance matrix may be a poor estimate of the error covariance matrix. A poor estimate of the error covariance matrix will likely lead, by equation (3.3.8), to a poor estimate of the parameter covariance matrix. Although the determinant criterion gives the minimum determinant of the estimate of the parameter covariance matrix, if this estimate is a poor, then the optimality property may be of little significance. This means that the optimality property of the determinant criterion is more relevant for large data sets than for small ones. This is equivalent to saying that the determinant criterion is only asymptotically optimal. The simulation studies of §4 will demonstrate that this is something of an Achilles heel for the determinant criterion.

In a simulation study there is an alternative way to estimate the parameter covariance matrix, Σ_θ. Equation (3.3.6) gives the estimate of the parameter covariance matrix from a single set of data. In a simulation study, q sets of simulated data may be used to estimate Σ_θ from q sets of estimated parameters:

$$\hat{\Sigma}_\theta = \frac{1}{q}\sum_{i=1}^{q}\left[\hat{\theta}_i - \theta^*\right]\left[\hat{\theta}_i - \theta^*\right]^T \qquad (3.3.9)$$

where θ^* is the true value of the parameter vector, known in simulation. If q is as large as 10^4, then equation (3.3.9) should give a better estimate of the parameter covariance matrix than does equation (3.3.6) because Σ_θ may be defined in a frequentist sense as:

$$\Sigma_\theta = \lim_{q\to\infty}\frac{1}{q}\sum_{i=1}^{q}\left[\hat{\theta}_i - \theta^*\right]\left[\hat{\theta}_i - \theta^*\right]^T \qquad (3.3.10)$$

3.4 Bias of the Parameter Estimates

For models that are linear with respect to the parameters, equation (3.3.1) converges in one iteration:

$$\hat{\theta} = (X^T W X)^{-1} X^T W y \tag{3.4.1}$$

If the generalized weight matrix W is constant, then taking expectations gives:

$$E(\hat{\theta}) = (X^T W X)^{-1} X^T W E(y) \tag{3.4.2}$$

The linear model can be written as:

$$y = X\theta^* + \varepsilon \tag{3.4.3}$$

where θ^* is the true value of the model parameters, θ. The expectation value of the measurement errors, $E(\varepsilon)$, is assumed to be zero so that:

$$E(y) = X\theta^* \tag{3.4.4}$$

Substitution of equation (3.4.4) back into equation (3.4.2) gives the result:

$$E(\hat{\theta}) = \theta^* \tag{3.4.5}$$

The expectation value of the parameter estimates is the true value of the parameter estimates. Therefore, for models linear with respect to the parameters, the parameter estimates from equation (3.4.1) are unbiased for any constant value of W for which $X^T W X$ is nonsingular. This derivation is not valid for models that are nonlinear with respect to the parameters or if W depends on the data as in equation (3.3.7). In the examples discussed in §4, both of these invalidating conditions hold and bias in the parameter estimates is, in fact, observed.

3.5 Multivariate Weighted Least Squares (MWLS)

As an alternative to the determinant criterion, a two-step iterated weighting scheme will be considered. In the first step, the model parameters are estimated by minimizing a weighted sum of squares of deviations with respect to the model parameters:

$$\hat{\theta} = \theta \text{ s.t. } \theta \text{ minimizes } \mathrm{tr}\left[Z(\theta)^T W Z(\theta) \right] \qquad (3.5.1)$$

Here W is a diagonal weight matrix which can be initialized to the identity matrix. In the second step the diagonal elements of the weight matrix are set to the inverse of the diagonal elements of the residual covariance matrix:

$$W = \left[\mathrm{diag}\left(Z(\hat{\theta})\, Z(\hat{\theta})^T / n \right) \right]^{-1} \qquad (3.5.2)$$

The m diagonal elements of W are the inverse of the estimated error variances of the m responses. Because $\hat{\theta}$ depends on W and W depends on $\hat{\theta}$, equations (3.5.1) and (3.5.2) are iterated to convergence.

Therefore this is a multivariate weighted least squares scheme where the weights are determined iteratively from the variance of the residuals. Carroll and Ruppert (1988) discuss a conceptually similar iteratively weighted least squares scheme in the context of uniresponse models where the measurement errors are assumed to be independent but heteroscedastic. However, no literature reference to an iteratively weighted least squares scheme applied to the multiresponse parameter estimation problem has been found.

The general Gauss-Newton equation (3.3.1) is applicable to MWLS. The estimate of the parameter covariance matrix for MWLS is given by equation (3.3.6) and, unlike the case for the determinant criterion, no further simplification is possible in this case.

Just as the determinant criterion has an equivalent iterative counterpart (§3.2), MWLS has an equivalent that resembles the determinant criterion. To demonstrate this, we go back to the discussion of the determinant criterion in §3.1 and the likelihood function for θ and Σ_ε for random and normally distributed measurement errors with zero mean:

$$L(\theta, \Sigma_\varepsilon) \propto |\Sigma_\varepsilon|^{-n/2} \exp\left[-\tfrac{1}{2}\operatorname{tr}\left(Z(\theta)^T \Sigma_\varepsilon^{-1} Z(\theta)\right)\right] \qquad (3.5.3)$$

For the sake of this argument it will be assumed that the measurement errors within a measurement vector are independent. This will make Σ_ε a diagonal matrix. Bard (1974) showed that under this assumption the maximum likelihood estimates of θ and Σ_ε are:

$$\hat{\theta} = \theta \text{ s.t. } \theta \text{ minimizes } \prod_{i=1}^{m}\left[Z(\theta)\, Z(\theta)^T\right]_{ii} \qquad (3.5.4)$$

$$\hat{\Sigma}_\varepsilon = \operatorname{diag}\left[Z(\hat{\theta})\, Z(\hat{\theta})^T / n\right] \qquad (3.5.5)$$

The likelihood functional conditional on an estimate of the error covariance matrix is:

$$L(\theta|\hat{\Sigma}_\varepsilon) \propto \exp\left[-\tfrac{1}{2}\operatorname{tr}\left(Z(\theta)^T \hat{\Sigma}_\varepsilon^{-1} Z(\theta)\right)\right] \qquad (3.5.6)$$

By comparing equations (3.5.1) and (3.5.2) with equations (3.5.5) and (3.5.6) it can be seen that equation (3.5.2) of MWLS corresponds to equation (3.5.5) and equation (3.5.1) of MWLS corresponds to equation (3.5.6). Under these conditions, the argument of §3.2 can be employed to show that MWLS is equivalent to equation (3.5.4). In other words, iterating equations (3.5.1) and (3.5.2) to convergence is equivalent to minimizing the product of the diagonal elements of the deviation covariance matrix with respect to the model parameters. This result is of computational significance because MWLS is first-order in its rate of convergence. But a second-order Newton method might be applied to the minimization problem posed by equation (3.5.4).

Equation (3.5.4) provides a maximum likelihood estimate of the model parameters when the measurement errors within a measurement vector are random, independent, and normally distributed with zero mean. Equation (3.5.4) is also equivalent to MWLS. However, this does not mean that the use of MWLS is being justified on likelihood grounds. Nor does it imply that the use of MWLS is justified only when the measurement errors within a measurement vector are independent. The simulation studies of §4 will demonstrate that, even when the measurement errors with a measurement vector are correlated, the use of MWLS can be justified on the grounds that the parameter estimates obtained using MWLS can have a smaller variance and are more nearly normally distributed than those obtained using the determinant criterion.

Before this section is concluded, a brief comment on the handling of redundant response variables will be made. The reason redundant response variables can and should be dropped when applying the determinant criterion is that the information in the redundant response variables is implicit in the deviation covariance matrix by equations (1.1.6). But in applying MWLS only the diagonal elements of this matrix are used as weights. Consequently, dropping redundant response variables can result in a loss of information for MWLS. Furthermore, retaining redundant response variables can do no harm because the weights will adjust themselves accordingly.

3.6 Degrees of Freedom for Multiresponse Estimation

Given an estimate of the parameter covariance matrix, $\hat{\Sigma}_\theta$, an equation for the approximate joint confidence region for the parameter estimates, as given for example by Draper and Smith (1981), can be written as:

$$(\theta - \hat{\theta})^T \hat{\Sigma}_\theta^{-1} (\theta - \hat{\theta}) \leq pF(p, \nu, \alpha) \qquad (3.6.1)$$

where $F(p, \nu, \alpha)$ is the upper α quantile of the F distribution with p and ν degrees of freedom. ν is the number of degrees of freedom associated with $\hat{\Sigma}_\theta$ which, in turn, is derived from $\hat{\Sigma}_\varepsilon$ through equation (3.3.6) for both MWLS and the determinant criterion. The maximum likelihood estimate of Σ_ε is given by equation (3.1.5) to be the residual covariance matrix $Z(\hat{\theta})\, Z(\hat{\theta})^T / n$. Now, ν would be equal to n if this estimate of Σ_ε was unbiased.

According to Bard (1974), ν is $n - p/m$ whereas Bates and Watts (1985) make a case for $n - p$. Either choice implies that the maximum likelihood estimate of Σ_ε, $Z(\hat{\theta})\, Z(\hat{\theta})^T / n$, is biased. It would be nice to resolve this issue because the estimate of the parameter covariance matrix, $\hat{\Sigma}_\theta$, depends on $\hat{\Sigma}_\varepsilon$ through equation (3.3.6). To make the mathematics tractable, I will consider the linear model:

$$y = X\theta^* + \varepsilon \qquad (3.6.2)$$

where θ^* is the true value of the parameters and $E(\varepsilon) = 0$. The least squares parameter estimate given a generalized weight matrix W is:

$$\hat{\theta} = (X^T W X)^{-1} X^T W y \qquad (3.6.3)$$

From equations (3.6.2) and (3.6.3) it follows that:

$$\hat{\theta} = \theta^* + (X^T W X)^{-1} X^T W \varepsilon \qquad (3.6.4)$$

Now a subset of the data, s, will be considered such that:

$$y_s = X_s \theta^* + \varepsilon_s \qquad (3.6.5)$$

In the multiresponse context, this data subset will be the data for a particular response. From equations (3.6.4) and (3.6.5), the residuals for the data subset are:

$$y_s - X_s \hat{\theta} = \varepsilon_s - X_s (X^T W X)^{-1} X^T W \varepsilon \qquad (3.6.6)$$

Given that the errors of the data subset, s, have the covariance matrix, $\Sigma_{\varepsilon s}$, the number of degrees of freedom, ν_s, associated with the generalized sum of squares of the subset residuals is, by definition:

$$\nu_s = E\left[(y_s - X_s \hat{\theta})^T \Sigma_{\varepsilon s}^{-1} (y_s - X_s \hat{\theta}) \right] \qquad (3.6.7)$$

Substituting equation (3.6.6) into equation (3.6.7) and assuming W to be constant gives:

$$\nu_s = n - 2\,\mathrm{tr}\left[X_s^T \Sigma_{\varepsilon s}^{-1} E(\varepsilon_s \varepsilon^T) W X (X^T W X)^{-1} \right] +$$

$$\mathrm{tr}\left[X_s^T \Sigma_{\varepsilon s}^{-1} X_s (X^T W X)^{-1} X^T W \Sigma_{\varepsilon n} W X (X^T W X)^{-1} \right] \qquad (3.6.8)$$

where n is the number of elements in the data subset. Because the set s is a subset of the whole data set, the elements of the matrices $\Sigma_{\varepsilon s}$ and $E(\varepsilon_s \varepsilon^T)$ are all elements of the error covariance matrix, Σ_ε. As before, $\Sigma_{\varepsilon n} = I_n \otimes \Sigma_\varepsilon$. If the data subset is the whole data set (i.e. $X_s = X$) and if W is the inverse of the error covariance matrix (i.e. $W = \Sigma_{\varepsilon n}^{-1}$) then equation (3.6.8) simplifies to the familiar result $\nu = n - p$.

For a multiresponse problem, the data subsets will correspond to m subsets of n elements per subset. Measurement errors between measurement vectors are assumed to be independent so, for the ith response, $\Sigma_{\varepsilon i} = I_n \cdot (\Sigma_\varepsilon)_{ii}$. For the special case where the error covariance matrix, Σ_ε, is known and $W = \Sigma_{\varepsilon n}^{-1}$, we have the result that:

$$E(\varepsilon_i \varepsilon^T) W X = E(\varepsilon_i \varepsilon^T) \Sigma_{\varepsilon n}^{-1} X = X_i \tag{3.6.9}$$

With this result equation (3.6.8) can be reduced to:

$$\nu_i = n - \mathrm{tr}\left[X_i^T X_i (X^T \Sigma_{\varepsilon n}^{-1} X)^{-1} \right] / (\Sigma_\varepsilon)_{ii} \tag{3.6.10}$$

For the special case where measurement errors within measurement vectors are independent, $\Sigma_{\varepsilon n}$ is a diagonal matrix. Because the trace operator is a linear operator, it follows from (3.6.10) that, for a diagonal $\Sigma_{\varepsilon n}$, the mean number of degrees of freedom per response is:

$$\frac{1}{m} \sum_{i=1}^{m} \nu_i = n - \frac{p}{m} \tag{3.6.11}$$

Equation (3.6.11) does not imply that each ν_i has the value $n - p/m$. Only the mean has this value and then only if $\Sigma_{\varepsilon n}$ is diagonal and $W = \Sigma_{\varepsilon n}^{-1}$.

A limiting case of equation (3.6.10) will now be considered. Suppose that, for response i, $(\Sigma_\varepsilon)_{ii} \to 0$. The number of degrees of freedom associated with response i can be determined by taking the limit of the right hand side of equation (3.6.10):

$$\lim_{(\Sigma_\varepsilon)_{ii} \to 0} \nu_i = n - p \tag{3.6.12}$$

This corresponds to the case where the measurement errors for one response, response i, are negligible relative to the errors for the other responses. In this case the parameter

estimates will, in effect, be solely determined by response i and the multiresponse problem becomes equivalent to a uniresponse problem with $n-p$ degrees of freedom.

Now, there are $n-p$ degrees of freedom for one of the responses in the special case just considered. But it cannot be that the number of degrees of freedom is $n-p$ in general for all responses. A general value of $n-p$ for all responses would be inconsistent with the special result of equation (3.6.11) (for $m > 1$). Bates and Watts (1985) consider the special case where $n-p$ and argue that $\nu = n-p$ applies in this case. They then suggest that $\nu = n-p$ is generally applicable. It has just been shown that this cannot be the case. Bard's (1974) suggestion that $\nu = n-p/m$ is generally applicable for all responses is, of course, consistent with equation (3.6.11) but it is not consistent with the special result of equation (3.6.12).

Equation (3.6.8) should dispel the notion that any simple expression for ν can be generally applicable to multiresponse estimation. Even equation (3.6.8) itself is only valid if W is constant (i.e. independent of the data). For both the determinant criterion and MWLS this is not the case. Therefore any simple expression for the number of degrees of freedom is going to have to be something of an approximation. In the light of equation (3.6.11), Bates and Watts (1985) suggestion of $n-p$ seems too conservative. In the absence of any other simple alternative to Bard's (1974) suggestion of $n-p/m$, I would recommend the use of this expression. This gives the following equation for an approximation to an unbiased estimate of the error covariance matrix:

$$\hat{\Sigma}_{\varepsilon} = \frac{Z(\theta)\,Z(\theta)^{T}}{n-p/m} \qquad (3.6.13)$$

Chapter 4 - Simulation Studies

Simulation studies for three multiresponse parameter estimation problems were done to show that the theoretical optimality property of the determinant criterion (§3.3) is of limited relevance to finite data sets. These studies were based on Burke's (1994) parameter estimation problem, the alpha-pinene problem of Box et al. (1973), and an example problem discussed by Box and Draper (1965).

4.1 Burke's Parameter Estimation Problem

Burke's (1994) parameter estimation problem was discussed in §2.1. The model described there is called the terminal model because the reaction rate is assumed to be influenced by the last, or terminal, monomer unit on the copolymer chain. Burke considered the case where the reaction rate is also assumed to be influenced by the second to last, or penultimate, monomer unit on the copolymer chain. The equations for this penultimate model are a slight modification of equations (2.1.3) and (2.1.4):

$$d_1 = r_{21}r_{11}f_1^2 + 2r_{21}f_1f_2 + f_2^2 \qquad (4.1.1a)$$

$$A_{111} = r_{21}r_{11}f_1^2 / d_1 \qquad (4.1.1b)$$

$$A_{112} = 2r_{21}f_1f_2 / d_1 \qquad (4.1.1c)$$

$$A_{212} = f_2^2 / d_1 \qquad (4.1.1d)$$

Equations (4.1.1) represent half the model. The equations for the monomer-2-centered triad fractions are obtained by interchanging the subscripts 1 and 2. In the penultimate model equations (2.1.5) remain unchanged. The penultimate model therefore contains four parameters: r_{11}, r_{21}, r_{22}, and r_{12}. Setting $r_{11} = r_{21}$ and $r_{22} = r_{12}$ restores the terminal model.

As part of a model discrimination study Burke performed an experiment that added a ninth set of measurements to the eight given in Table 2.1.1:

Table 4.1.1 Additional C^{13}-NMR Data from Burke's Thesis

i	f_1	X	Y	Z	A	B+C	D
9	0.56030	0.11975	0.56075	0.31950	0.17333	0.77779	0.04888

The parameters of the penultimate model were fitted to the nine sets of measurements using multivariate weighted least squares (MWLS). The point estimates are:

Table 4.1.2 Parameter Estimates

$\ln r_{11}$	-0.567
$\ln r_{21}$	-0.447
$\ln r_{22}$	-0.877
$\ln r_{12}$	0.041

The residuals yield the following estimates of the error standard deviations and correlation matrix:

$$\hat{\sigma}_\varepsilon = (\,2.41 \quad 3.02 \quad 3.03 \quad 1.99 \quad 4.04 \quad 4.30\,) \times 10^{-2} \tag{4.1.2}$$

$$\hat{\rho}_\varepsilon = \begin{bmatrix} 1 & -0.398 & -0.399 & -0.066 & -0.654 & 0.644 \\ -0.398 & 1 & -0.682 & -0.657 & 0.416 & -0.086 \\ -0.399 & -0.682 & 1 & 0.709 & 0.106 & -0.427 \\ -0.066 & -0.657 & 0.709 & 1 & -0.107 & -0.362 \\ -0.654 & 0.416 & 0.106 & -0.107 & 1 & -0.888 \\ 0.644 & -0.086 & -0.427 & -0.362 & -0.888 & 1 \end{bmatrix} \tag{4.1.3}$$

The correlation matrix shows fairly high correlation between the residuals for some of the response variables.

From the parameters of Table 4.1.2 one million data sets were synthesized with the simulated measurement errors multinormally distributed with zero mean and covariance matrix derived from equations (4.1.2) and (4.1.3). Each data set contained nine sets of measurements with nine values of the independent variable, f_1, taken from Tables 2.1.1 and 4.1.1. Parameters were estimated from each data set using both MWLS and the determinant criterion (Det) with two redundant response variables dropped for the latter method. The statistics of the sample distributions of the parameter estimates are tabulated below:

Table 4.1.3 Simulation Results from One Million Data Sets

Parameter	True Value	Means		Standard Deviations	
		MWLS	Det	MWLS	Det
$\ln r_{11}$	−0.567	−0.568	−0.565	0.156	0.201
$\ln r_{21}$	−0.447	−0.448	−0.450	0.133	0.167
$\ln r_{22}$	−0.877	−0.874	−0.877	0.179	0.242
$\ln r_{12}$	0.041	0.052	0.071	0.396	0.544

The mean values of the sample distributions of the parameter estimates do not suggest that bias is an issue here. But the standard deviations from MWLS are all smaller than those from the determinant criterion. Now, the theory of §3.3 showed that the determinant of the parameter covariance matrix should be smaller for the determinant criterion. Determinants, representing hypervolumes, greatly exaggerate small differences in spaces of even a moderate number of dimensions, p. Therefore, it is more realistic to deal with $|\Sigma_\theta|^{1/p}$ and $|\Sigma_\theta|^{1/2p}$ which will be referred to in this thesis as the generalized parameter variance and standard deviation, respectively. In this case, the generalized parameter standard deviation, as determined from equation (3.3.9), is 0.168 for MWLS and 0.219 for the determinant criterion. Clearly, the simulation result is not consistent with the theoretical result that $|\Sigma_\theta|$ should be smaller for the determinant criterion.

Figures 4.1.1(a) to 4.1.1(d) on the following pages show the sample distributions of the four model parameters estimated from the one million synthetic data sets. The frequency scale is logarithmic to highlight the tails of the distributions. On such a scale a normal distribution would take the form of a parabola. The distributions from MWLS (solid curves) are very nearly normal. But the tails of the distributions from the determinant criterion (broken curves) are heavier than those from a normal distribution.

With reference to equation (3.6.4), a multinormal measurement error distribution will give a multinormal distribution in parameter estimates if the model is linear in the parameters (i.e. matrix X is constant) and the weight matrix W is constant from sample to sample. The lack of normality in the sample distributions for the parameter estimates from the determinant criterion cannot be attributed to model nonlinearity which would affect the results from MWLS as well. So the difference in distributions must be associated with the weight matrix which is determined from the residual covariance matrix by equation (3.3.7) for the determinant criterion and equation (3.5.2) for MWLS.

To characterize the variation in the residual covariance matrix as a distribution, it is convenient to consider a scalar function of the residual covariance matrix, the ratio of the condition numbers of the residual covariance matrix and the ***sample*** error covariance matrix. For this discussion, the sample error covariance matrix is the covariance matrix of the synthetic errors for a given sample. Therefore,

$$\text{condition ratio} \equiv \frac{\text{condition number}\left[Z(\hat{\theta})\,Z(\hat{\theta})^T/n\right]}{\text{condition number}\left[Z(\theta^*)\,Z(\theta^*)^T/n\right]} \qquad (4.1.4)$$

where θ^* is the true value of the parameters. The sample error covariance matrix can only be known in a simulation and it is, of course, different from sample to sample. Taking the ratio of the condition numbers compensates for the effect of random variation in the condition numbers of the sample error covariance matrices.

Figure 4.1.1 Sample Distributions of Parameter Estimates with 95% Confidence Intervals

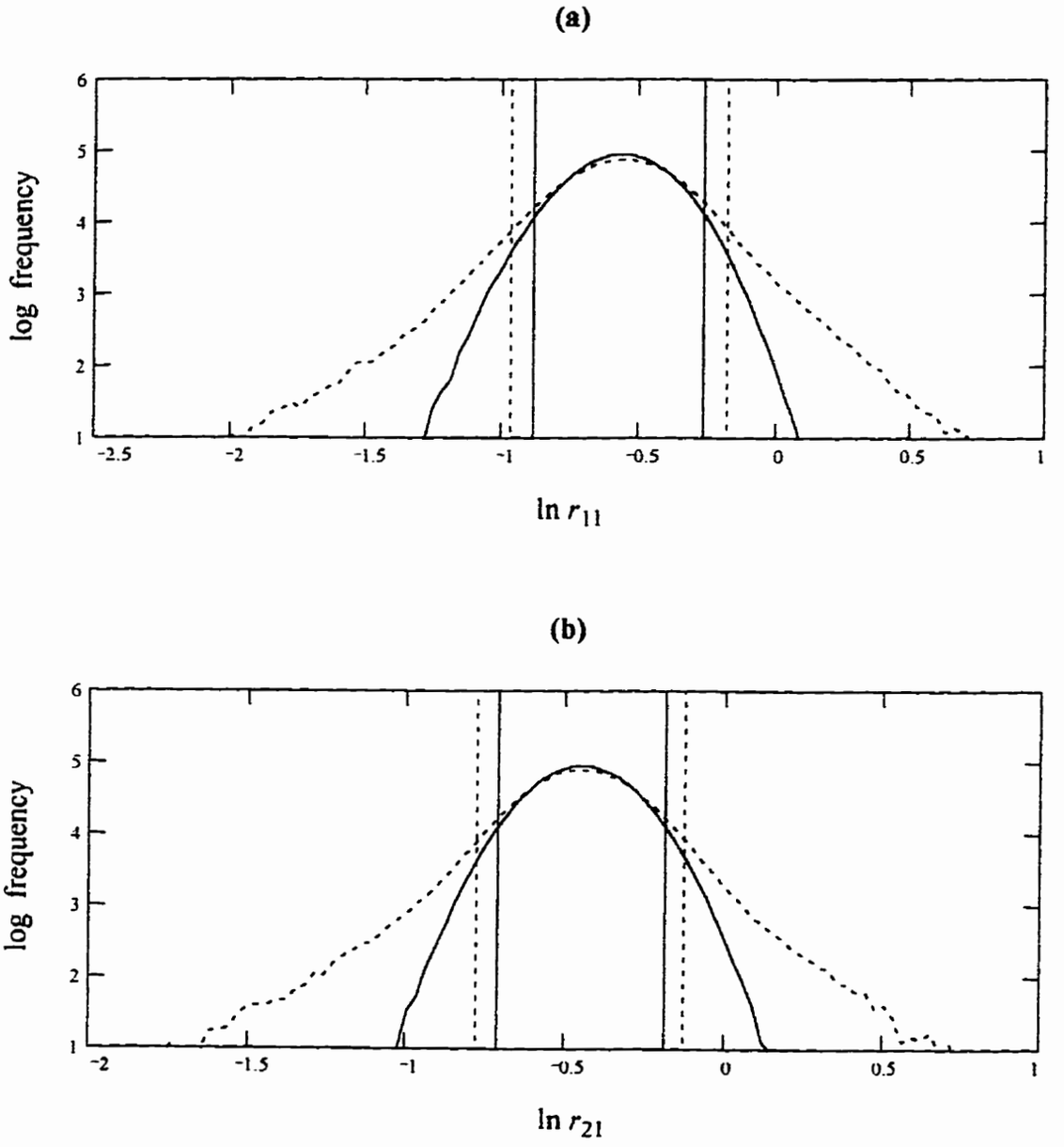
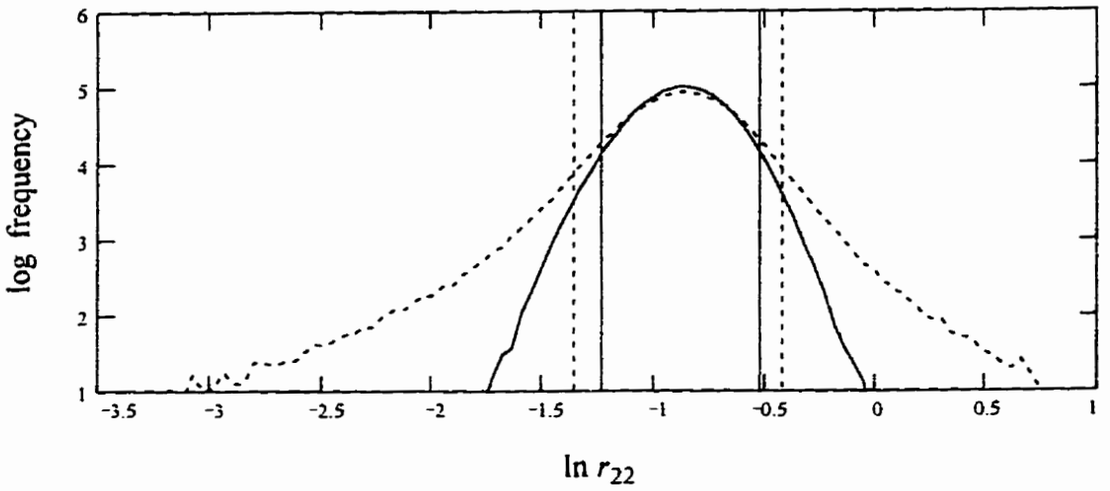
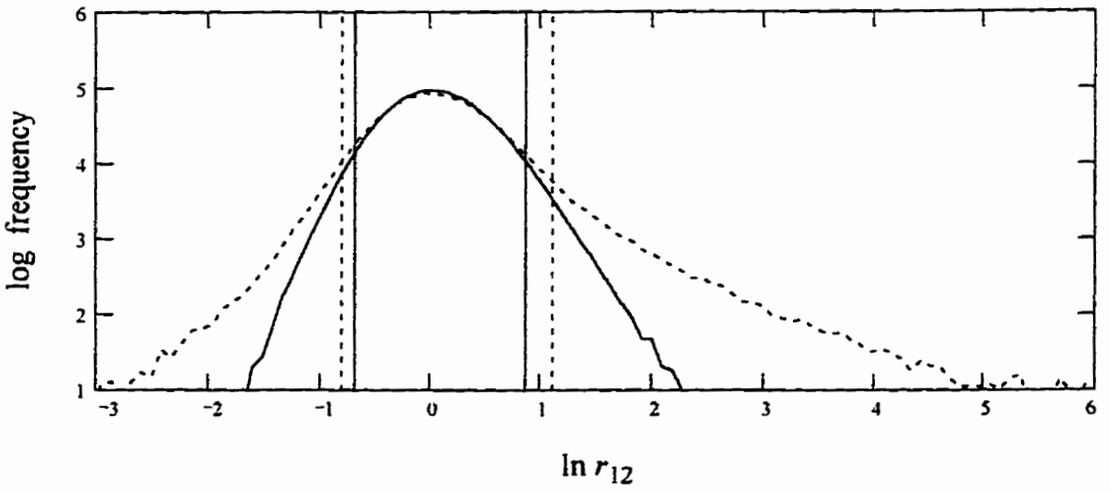


Figure 4.1.1 (Continued)

(c)



(d)

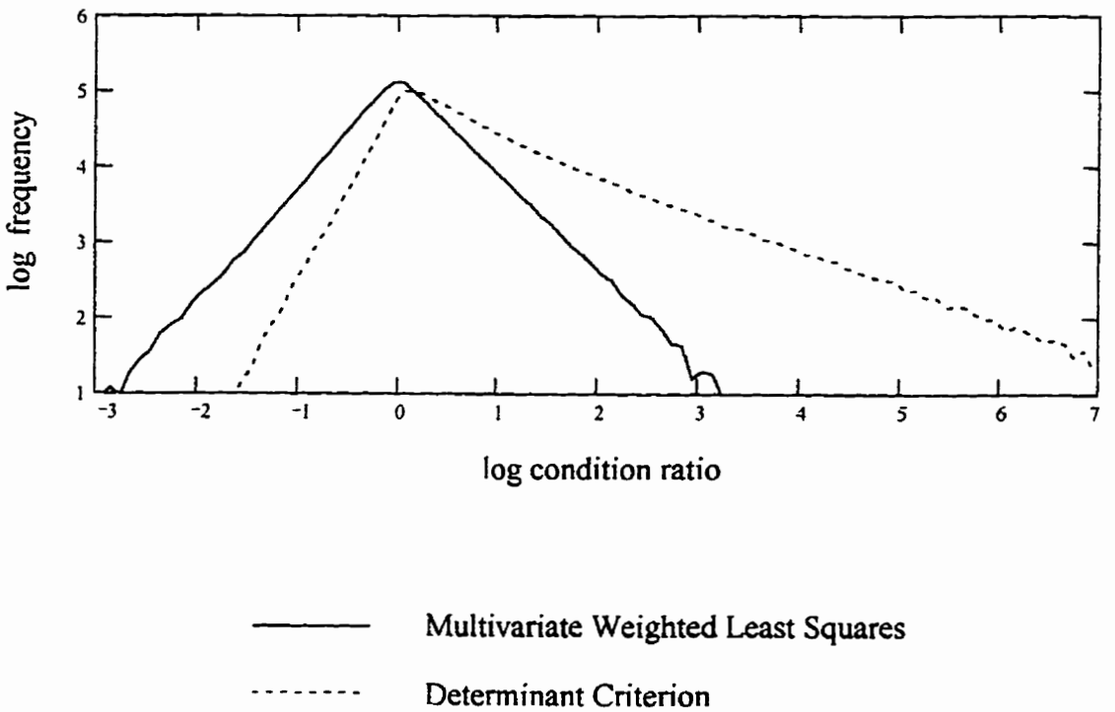


——— Multivariate Weighted Least Squares

- - - - - - - Determinant Criterion

The condition ratio is a measure of how much the residual covariance matrix is biased as an estimate of the sample error covariance matrix. By using condition number, the bias is expressed in terms of ill-conditioning. The higher the condition ratio, the more the residual covariance matrix is biased towards ill-conditioning. Figure 4.1.2 below shows a distribution nearly symmetrical about zero in the logarithm of condition ratio for MWLS. However, the distribution for the determinant criterion is skewed far to the right in the direction of ill-conditioning. (Note that the scale of the condition ratio in Figure 4.1.2 is log base *10*.)

Figure 4.1.2 Sample Distributions of Condition Ratio



The determinant criterion finds the point in parameter space that minimizes the determinant of the residual covariance matrix. The determinant is the product of the eigenvalues. The condition number is the ratio of largest to smallest eigenvalue. An examination of the distributions of the eigenvalues of the sample error and residual covariance matrices shows that the determinant criterion, in minimizing the determinant, tends to preferentially bias the smallest eigenvalue towards zero. This, in turn, biases the residual covariance matrix towards ill-conditioning. The lack of normality in the parameter distribution from the determinant criterion can therefore be attributed to the simple fact that the generalized weight matrix is the inverse of a residual covariance matrix that is biased towards ill-conditioning.

In the demonstration of §3.3 that the determinant criterion is optimal among weighting schemes where the error covariance matrix is unknown, a key assumption is that the residual covariance matrix is a good estimate of the error covariance matrix. This example shows that the determinant criterion can violate this assumption by biasing the residual covariance matrix as an estimate of the error covariance matrix. Bias towards ill-conditioning in the residual covariance matrix is not evident with MWLS. This explains why MWLS can give a smaller value of the determinant of the true parameter covariance matrix than the determinant criterion, contrary to the theory of §3.3 which deals with ***estimates*** of the parameter covariance matrix.

Figures 4.1.1(a) to 4.1.1(d) showed the distributions of point estimates of the model parameters from MWLS and the determinant criterion. But these two estimation methods also yield estimates of the uncertainty in the parameter estimates through equation (3.3.6) for the parameter covariance matrix, $\hat{\Sigma}_\theta$. So this raises the following question: How good are these estimates of the parameter uncertainty? This question can be addressed through the same Monte-Carlo procedure used to generate the distributions of the point estimates of the model parameters.

In this case, ten thousand synthetic data sets were generated using the same procedure described earlier. For each data set, point estimates of the parameters were obtained from MWLS and the determinant criterion and corresponding estimates of the parameter covariance matrix were obtained from equations (3.3.6) and (3.3.8), respectively. The determinant criterion gives the minimum value of $\left|\hat{\Sigma}_\theta\right|$. Since the frequency distribution of this quantity is very skewed, the related quantity, $\ln\left|\hat{\Sigma}_\theta\right|/p$, was used to generate more symmetric frequency distributions.

In simulation the error covariance matrix, Σ_ε, is known. By the argument of §3.3, if Σ_ε is known, then the optimal estimation method is generalized least squares with the generalized weight matrix set to the inverse of the error covariance matrix. Here, as in §3.3, optimal is in the sense of minimizing the determinant of the parameter covariance matrix. So, in simulation, generalized least squares provides a reference standard by which MWLS and the determinant criterion may be evaluated. If Σ_ε is known, then equation (3.3.5) can be used to estimate the parameter uncertainty for generalized least squares. However, the parameter uncertainty estimates for both MWLS and the determinant criterion are derived from equation (3.3.6). So, for the sake of a consistent comparison, equation (3.3.6) will be used for generalized least squares as well.

If the number of simulated data sets is very large, then equation (3.3.9) gives a much better estimate of the parameter covariance matrix than does equation (3.3.6). When the number of simulated data sets is as large as ten thousand then, for the purpose of comparison, equation (3.3.9) may be regarded as giving the 'true' value of Σ_θ.

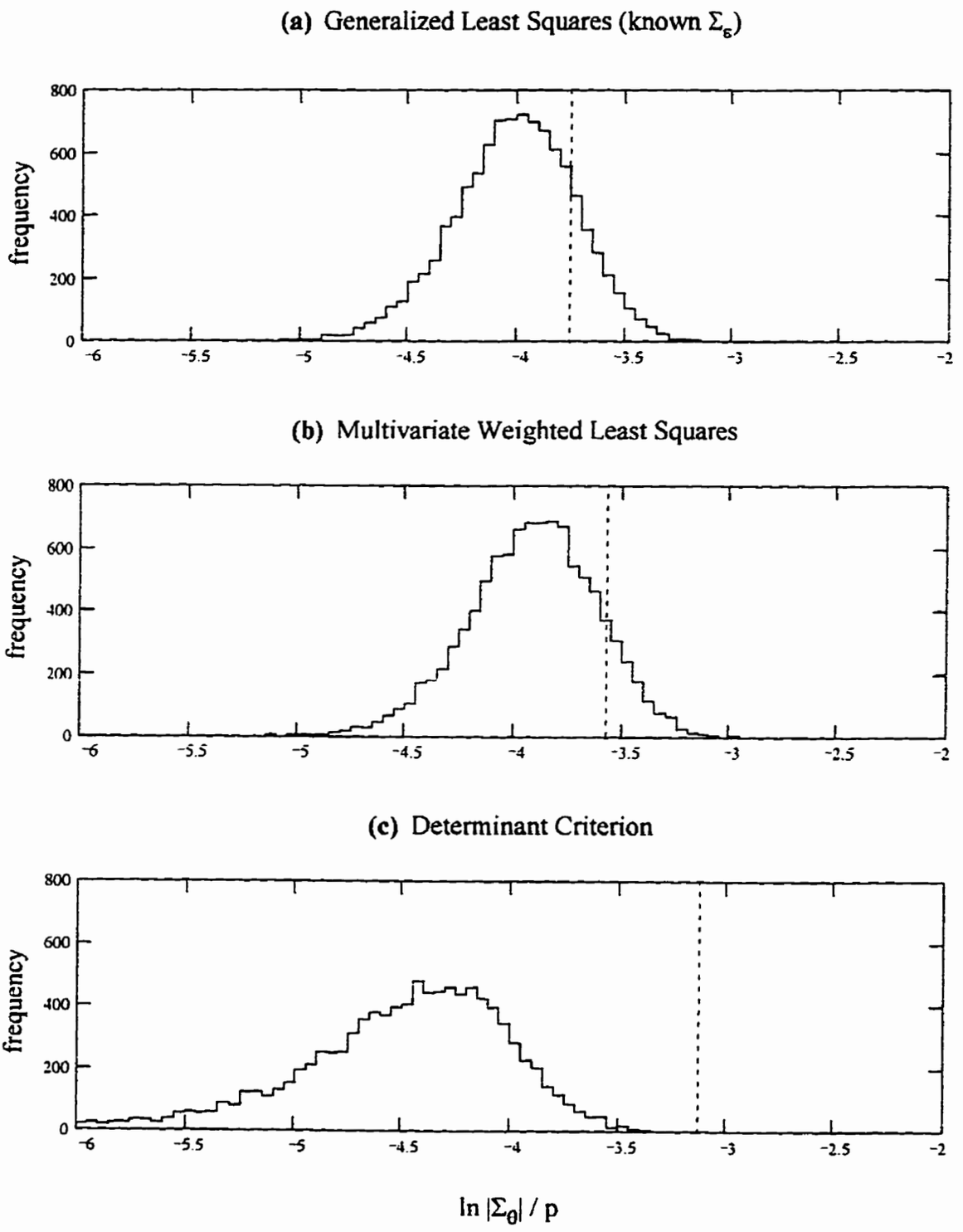
The results of the simulations are given in the histograms of Figures 4.1.3(a) to 4.1.3(c) on the next page. For each method, the 'true' value of $|\Sigma_\theta|$ from equation (3.3.9) is indicated by a vertical broken line. For generalized least squares, some bias in the estimates of $\ln|\Sigma_\theta|/p$ from equation (3.3.6) is evident because the quite symmetrical distribution is centered to the left of the 'true' value, as is also shown in Table 4.1.4 below. For MWLS both the distribution and the 'true' value shift to the right indicating more uncertainty in the parameter estimates. For the determinant criterion the distribution shifts to the left (and widens) while the 'true' value shifts even more to the right. The reason the distribution shifts to the left is that the determinant criterion, by the optimality property of §3.3, gives the smallest estimate of $|\Sigma_\theta|$. But, as has been discussed, minimizing the estimate of $|\Sigma_\theta|$ does not guarantee that $|\Sigma_\theta|$ itself will be minimized.

The main point here is that the distribution of estimates of the parameter uncertainty, as determined using equation (3.3.6), is much closer to the reference standard (generalized least squares) for MWLS than for the determinant criterion. This can be attributed to the fact that the determinant criterion biases the residual covariance matrix as an estimate of the error covariance matrix.

Table 4.1.4 Simulation Results from Ten Thousand Data Sets

$\ln\|\Sigma_\theta\|/p$	Distribution Mean	'True' Value
Generalized least squares	−4.01	−3.75
MWLS	−3.91	−3.57
Determinant criterion	−4.51	−3.13

Figure 4.1.3 Sample Distributions of Parameter Uncertainty Estimates



In the next simulation study, the effect of replication in the data on the parameter variance was examined. The true values of the model parameters, θ, and the error covariance matrix, Σ_ε, were set to the estimates given in Table 4.1.2 and equations (4.1.2) and (4.1.3), respectively. The vector of independent variables was set to:

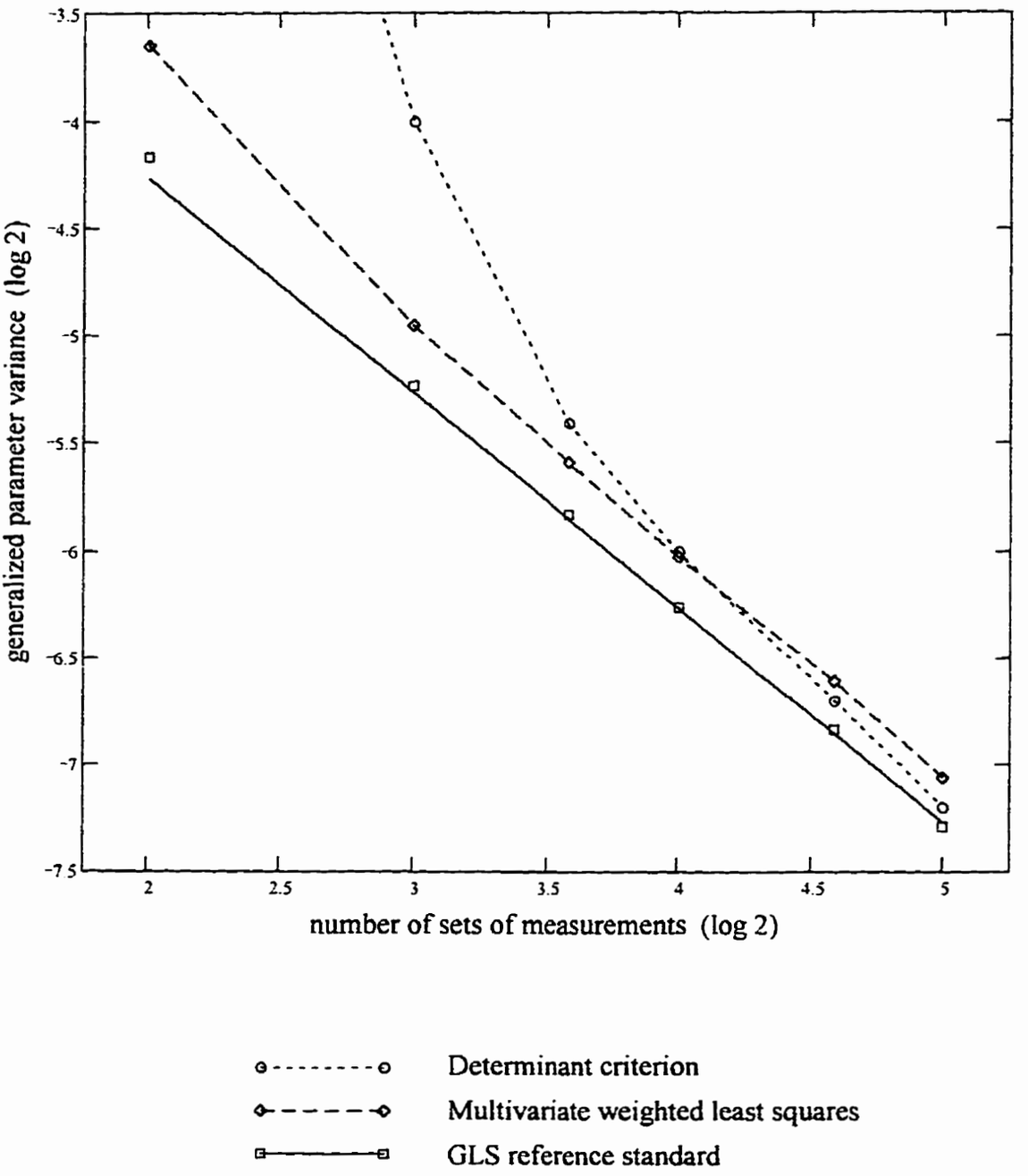
$$f_1 = (0.219 \;\; 0.395 \;\; 0.517 \;\; 0.792)^T \qquad (4.1.5)$$

Ten thousand synthetic data sets were generated with the simulated measurement errors being normally distributed with zero mean and covariance matrix as indicated above.

Model parameters were estimated from each data set using three methods: the determinant criterion, MWLS, and generalized least squares (GLS) where the generalized weight matrix is the inverse of the true error covariance matrix, Σ_ε. The latter method provides a reference that is idealized in the sense that the true error covariance matrix is usually known only in simulation. The determinant of the parameter covariance matrix, $|\Sigma_\theta|$, for each of the three methods was estimated from 10000 sets of fitted model parameters using equation (3.3.9).

To investigate the effect of the number of sets of measurements, n, on the results, the simulation study was repeated for six full levels of replication giving results for n = 4, 8, 12, 16, 24 and 32. These results, in the form of generalized parameter variance, $|\Sigma_\theta|^{1/p}$, plotted as a function of n on a log-log scale, are given in Figure 4.1.4 on the next page.

Figure 4.1.4 Effect of Sample Size on Parameter Variance



As expected, generalized least squares estimation using a generalized weight matrix derived from the true error covariance yields the smallest parameter variance. For this GLS reference standard, the relationship between generalized parameter variance and number of sets of measurements on a log-log scale is nearly linear with a slope of minus one (solid line), reflecting the inverse relationship between variance and sample size for simple estimation. As n increases, the results from the determinant criterion (dotted curve) converge to the ideal reference line because the residual covariance matrix converges to the true error covariance matrix with increasing n. The results from MWLS (dashed curve) show an offset from the reference line because the inverse of the diagonal weight matrix does not converge to the true error covariance matrix with increasing n. Therefore Figure 4.1.4 provides strong empirical support to the argument made in §3.3 that the determinant criterion is only asymptotically optimal.

The divergence of the curve for the determinant criterion from the GLS reference line with decreasing n leads to a limiting case for $n = m - 1$, where m is the number of response variables. For this limiting case the residual covariance matrix is singular and the parameter covariance for the determinant criterion is infinite. But a singular residual covariance matrix does not cause a problem for MWLS unless one of the diagonal elements is zero. Thus the crossing of the two curves for the determinant criterion and MWLS with decreasing n is not a peculiar feature of this example, as will be shown in §4.2 and §4.3.

4.2 The Alpha-Pinene Parameter Estimation Problem

The alpha-pinene parameter estimation problem was discussed in §2.2. Given the very small values of the reaction rate constants, of the order of 10^{-5}, the model will be reparameterized as follows:

$$\theta_i = k_i \times 10^5 \qquad i = 1...5 \qquad (4.2.1)$$

The parameters estimated using MWLS are:

Table 4.2.1 Parameter Estimates

θ_1	5.92
θ_2	2.95
θ_3	2.05
θ_4	30.3
θ_5	4.92

The residuals yield the following estimates of the error standard deviations and correlation matrix:

$$\hat{\sigma}_\varepsilon = (\ 0.753 \quad 0.806 \quad 0.428 \quad 0.607 \quad 0.897\ ) \qquad (4.2.2)$$

$$\hat{\rho}_\varepsilon = \begin{bmatrix} 1 & -0.830 & -0.015 & -0.378 & 0.163 \\ -0.830 & 1 & 0.228 & 0.366 & -0.568 \\ -0.015 & 0.228 & 1 & -0.161 & -0.632 \\ -0.378 & 0.366 & -0.161 & 1 & -0.487 \\ 0.163 & -0.568 & -0.632 & -0.487 & 1 \end{bmatrix} \qquad (4.2.3)$$

The residual correlation matrix shows fairly high correlation between the residuals for some of the response variables.

From the parameters of Table 4.2.1, one million data sets were synthesized with the simulated measurement errors being multinormally distributed with zero mean and covariance given by equations (4.2.2) and (4.2.3). Each data set contained eight sets of measurements with the independent variable, t, set to the eight values of Table 2.2.1. Next, the simulated values of y_4 were discarded and replaced with values determined using equation (2.2.6). This simulated the procedure used by Fuguitt and Hawkins in their original analysis of data that didn't include actual measurements of y_4. Therefore, the errors in the 'measurements' of y_4 are far from random and normally distributed.

Parameters were estimated from each data set using both MWLS and the determinant criterion (Det) with one redundant response variable dropped for the latter method. The statistics of the sample distributions of the parameter estimates are tabulated below:

Table 4.2.2 Simulation Results from One Million Data Sets

Parameter	True Value	Means		Standard Deviations	
		MWLS	Det	MWLS	Det
θ_1	5.92	5.94	5.93	0.074	0.27
θ_2	2.95	2.97	2.97	0.042	0.13
θ_3	2.05	2.11	2.16	0.085	0.28
θ_4	30.3	30.4	29.0	2.6	9.6
θ_5	4.92	4.89	3.97	0.88	2.6

In this case, the values of the generalized parameter standard deviation, $|\Sigma_\theta|^{1/2p}$, as determined by equation (3.3.9), are 0.176 for MWLS and 0.531 for the determinant criterion. In other words, the generalized parameter variance for MWLS is about an order of magnitude less than that for the determinant criterion. This result is, of course, wildly inconsistent with the optimality property of the determinant criterion derived in §3.3.

The sample distributions of the parameter estimates and the condition ratio, on a logarithmic frequency scale, are given in Figures 4.2.1(a) to 4.2.1(f) on the following pages. The nearly parabolic shapes of the parameter distributions for MWLS (solid curves) indicate near normal distributions in the parameter estimates. On the other hand, the shapes of the distributions for the determinant criterion (broken curves) are very far from the parabolic shape of a normal distribution. The heavy tails on these distributions greatly inflate the variance of the parameter estimates for the determinant criterion.

The sample distributions of the condition ratio, defined in §4.1 as the ratio of the condition numbers of the residual covariance matrix and the sample error covariance matrix, shows the same features of the condition ratio distributions in Figure 4.1.2. The distribution in the condition ratio for the determinant criterion is skewed to the right in the direction of ill-conditioning. Because the determinant criterion biases the residual covariance matrix as an estimate of the error covariance matrix, the validity of the theory supporting the optimality property of this criterion is compromised.

A word should be said here about the optimization procedure used to determine the point estimates of the parameters. The initial values of the parameter estimates were set to the true values of Table 4.2.1 used to generate the data set. Newton optimization steps could have been applied at first but this would have been risky, given that the objective function can have multiple local optima. So a conservative gradient method was used to get to the vicinity of the optimum, at which point Newton steps were applied to 'polish' the result. In the case of the determinant criterion, severe ill-conditioning of the deviation covariance matrix stalled the search in about 0.04% of the one million cases, but this would most likely result in the parameter variance for the determinant criterion to be underestimated because the searches were initialized with the true parameter values.

Figure 4.2.1 Sample Distributions of Parameter Estimates with 95% Confidence Intervals

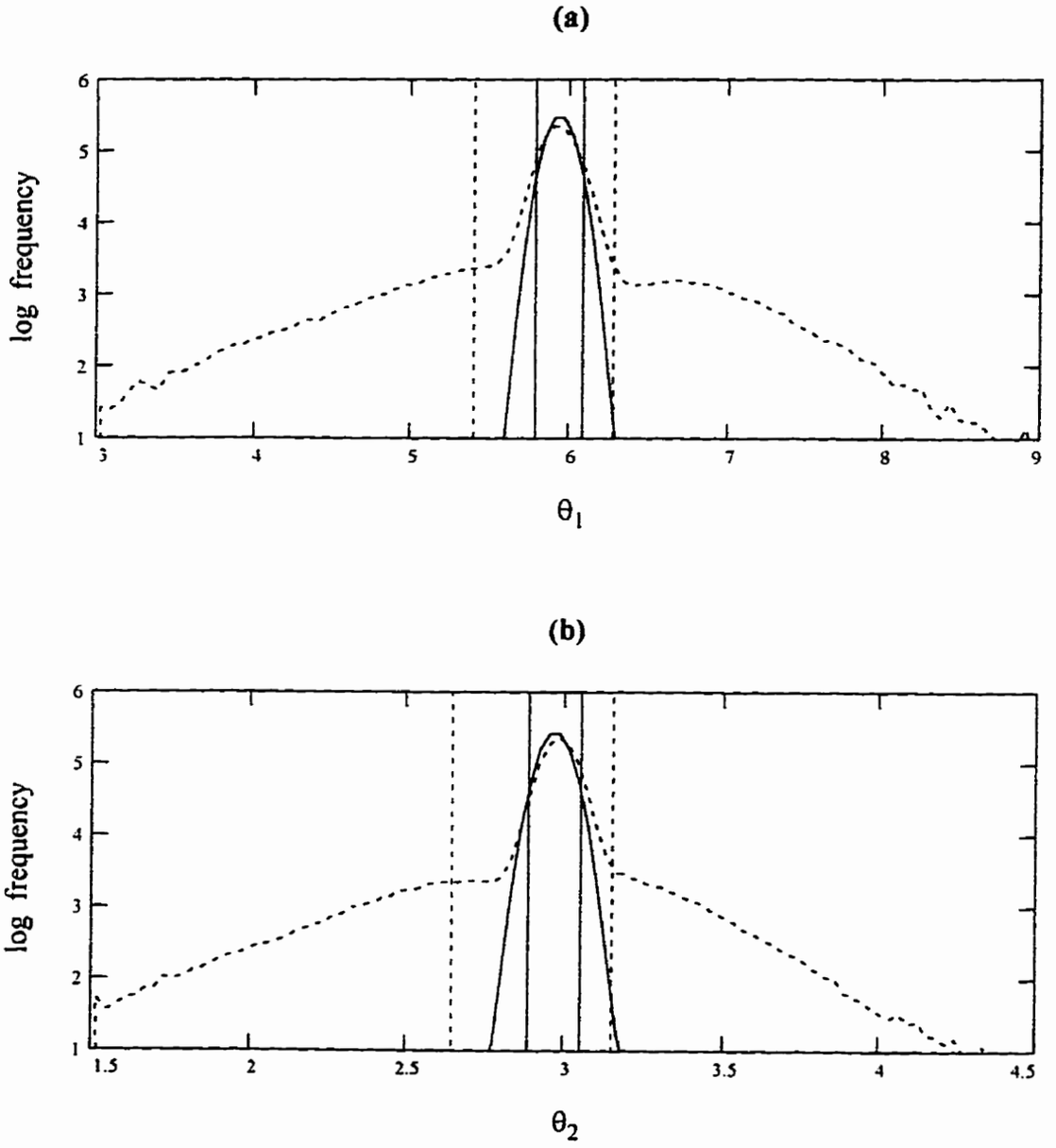
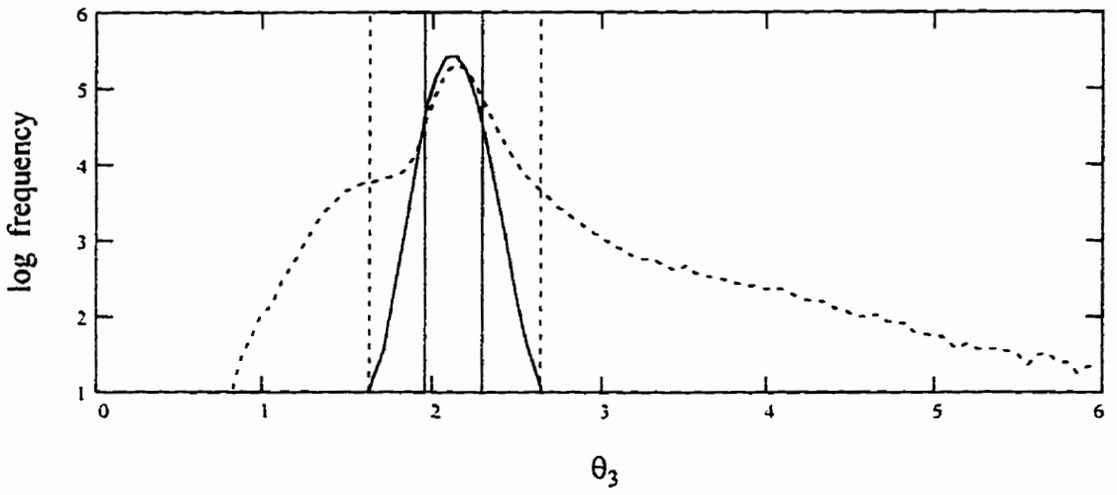
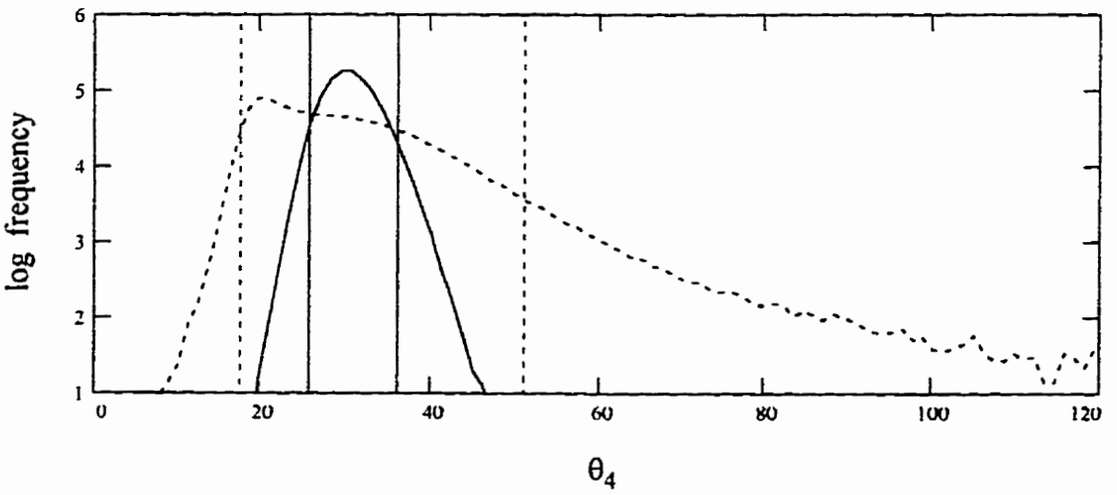


Figure 4.2.1 (Continued)

(c)



(d)

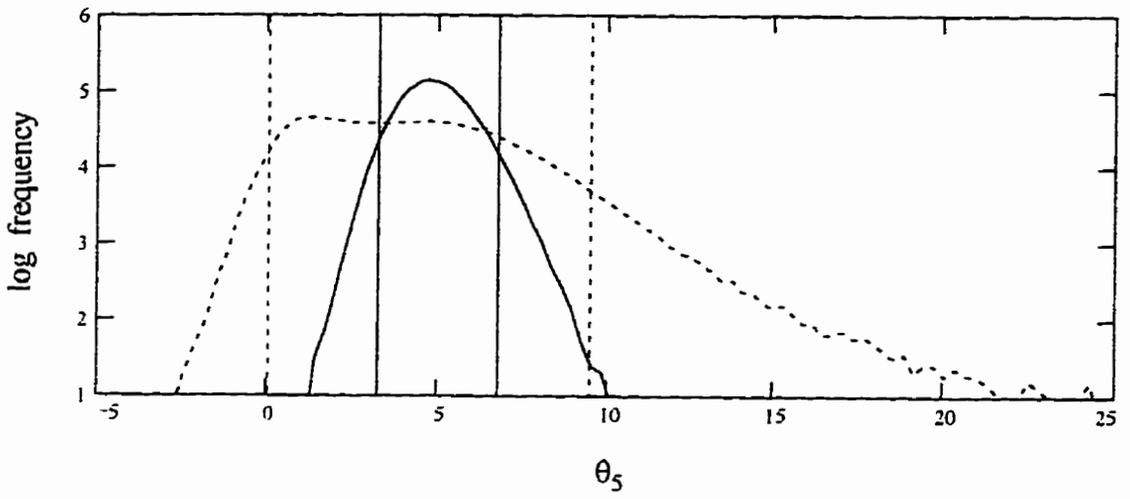


——— Multivariate Weighted Least Squares

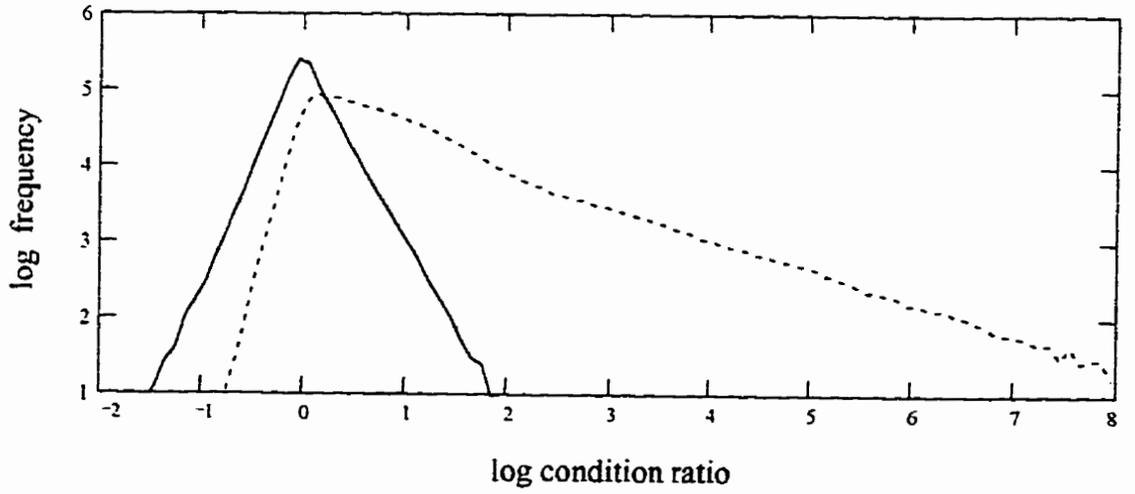
- - - - - - - Determinant Criterion

Figure 4.2.1 (Continued)

(e)



(f)



——— Multivariate Weighted Least Squares

- - - - - - - Determinant Criterion

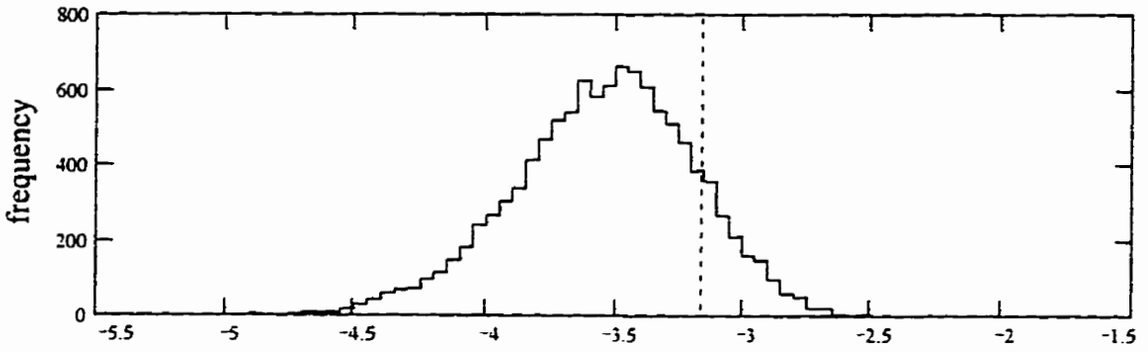
This example demonstrates that the determinant criterion can give very poor parameter estimates relative to MWLS when the measurement errors are not random with zero mean. In such a case, it is pointless to do a comparison between MWLS and the determinant criterion in terms of the quality of the estimates of parameter uncertainty from equation (3.3.6). This is because the validity of equation (3.3.6) depends on the measurement errors being random with zero mean. However, it is easy to remedy this by not building into the simulated data the defect that Fuguitt and Hawkins built into their data through equation (2.2.6). Thus, the simulated measurement errors will satisfy the assumptions that validate equation (3.3.6) as in the simulation study of §4.1.

A simulation study similar to the one described in §4.1 was conducted with the simulated measurement errors normally distributed with variances and correlation matrix given by equations (4.2.2) and (4.2.3). Generalized least squares (known Σ_ε), MWLS and the determinant criterion were applied to ten thousand simulated data sets and equation (3.3.6) was used to estimate the parameter covariance matrix, Σ_θ, for each method. The histograms of $\ln\left|\hat{\Sigma}_\theta\right|/p$ are given in Figures 4.2.2(a) to 4.2.2(c) on the next page. The 'true' values of the parameter uncertainty, as determined from equation (3.3.9), are indicated by vertical broken lines.

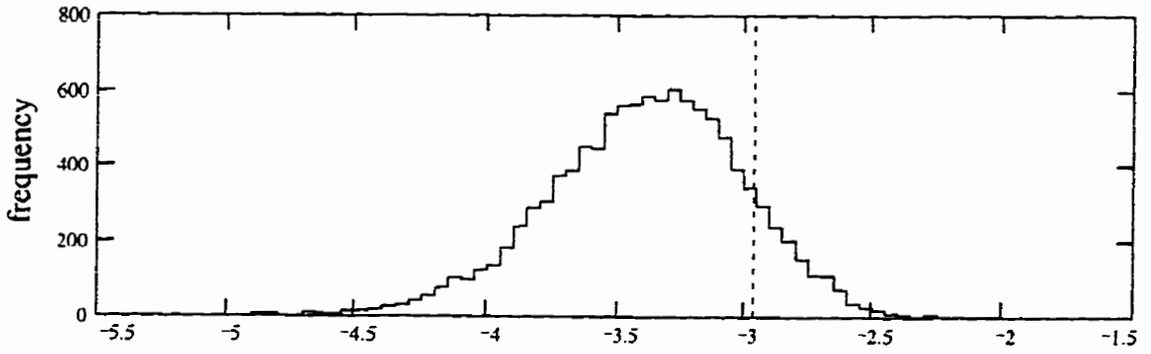
Figure 4.2.2 for the α-pinene problem shows the same characteristics as Figure 4.1.3 does for Burke's copolymerization problem. The distribution of estimates of $\ln|\Sigma_\theta|/p$, relative to the 'true' value, is about the same for generalized least squares and MWLS. However, the distribution of estimates for the determinant criterion is quite different in that it is much more biased towards underestimating the parameter uncertainty. The noteworthy point here is that, because Figures 4.1.3 and 4.2.2 are so similar, their characteristics are not likely due to some peculiar feature of the respective models or simulated data.

Figure 4.2.2 Sample Distributions of Parameter Uncertainty Estimates

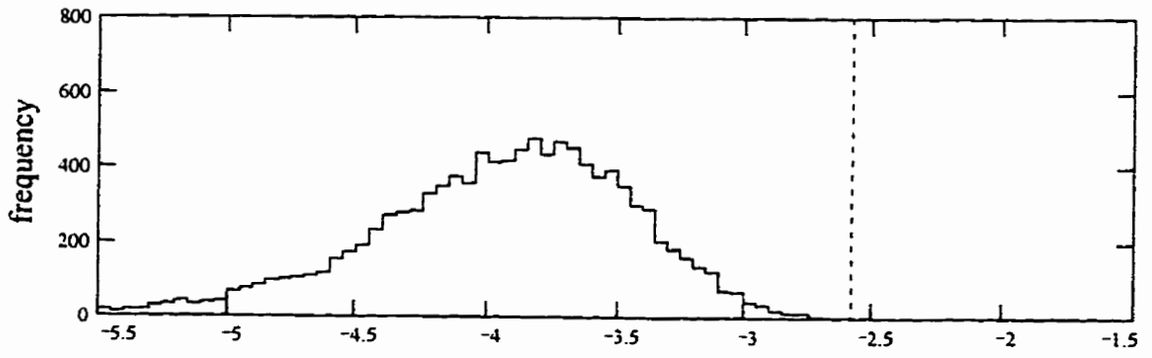
(a) Generalized Least Squares (known Σ_ε)



(b) Multivariate Weighted Least Squares



(c) Determinant Criterion



$\ln |\Sigma_0| / p$

In the next simulation study, the effect of replication in the data on the parameter variance was examined. The true values of the model parameters, θ, and the error covariance matrix, Σ_ε, were set to the estimates given in Table 4.2.1 and equations (4.2.2) and (4.2.3), respectively. The vector of independent variables was set to:

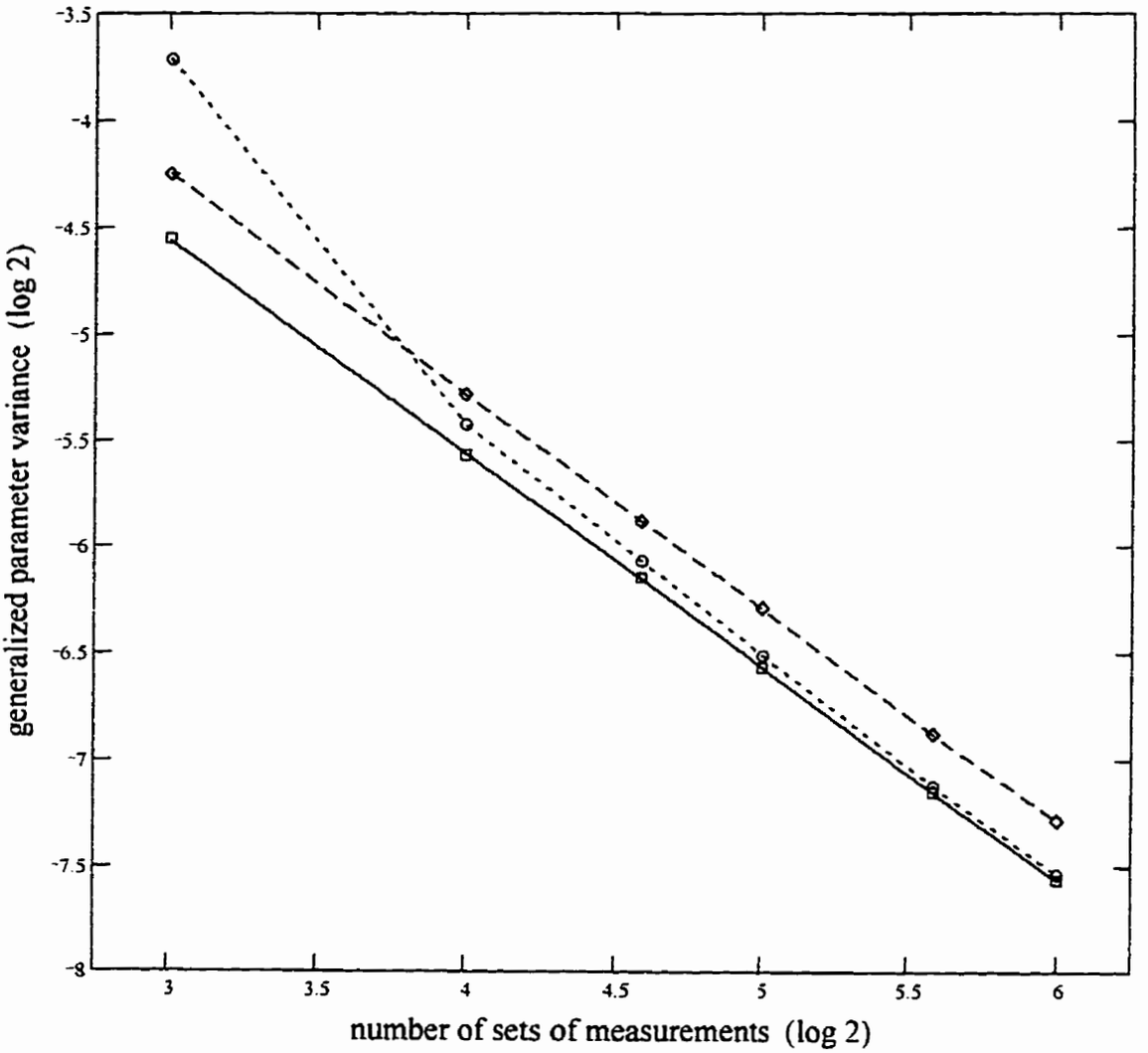
$$t = (\ 1230\ \ 3060\ \ 4920\ \ 7800\ \ 10680\ \ 15030\ \ 22620\ \ 36420\ )^T \qquad (4.2.4)$$

Ten thousand synthetic data sets were generated with the simulated measurement errors being normally distributed with zero mean and covariance matrix as indicated above.

Model parameters were estimated from each data set using three methods: the determinant criterion, MWLS, and generalized least squares (GLS) where the generalized weight matrix is the inverse of the true error covariance matrix, Σ_ε. The determinant of the parameter covariance matrix, $|\Sigma_\theta|$, for each of the three methods was estimated from 10000 sets of fitted model parameters using equation (3.3.9).

To investigate the effect of the number of sets of measurements, n, on the results, the simulation study was repeated for six full levels of replication giving results for n = 8, 16, 24, 32, 48 and 64. These results, in the form of generalized parameter variance, $|\Sigma_\theta|^{1/p}$, plotted as a function of n on a log-log scale, are given in Figure 4.2.3 on the next page. The features of Figure 4.2.3 are qualitatively similar to those of Figure 4.1.4. The crossing of the curves for MWLS and the determinant criterion supports the interpretation of Figure 4.1.4 discussed in §4.1.

Figure 4.2.3 Effect of Sample Size on Parameter Variance



4.3 Box and Draper's Parameter Estimation Problem

In introducing the determinant criterion, Box and Draper (1965) discussed an example from chemical reaction kinetics involving the reactions $A \rightarrow B \rightarrow C$. If the proportion of reactants A, B and C at time t are denoted by y_1, y_2 and y_3 then first-order reaction kinetics with rate constants ϕ_1 and ϕ_2 yields the system of ordinary differential equations:

$$dy_1/dt = -\phi_1 y_1 \qquad (4.3.1a)$$

$$dy_2/dt = \phi_1 y_1 - \phi_2 y_2 \qquad (4.3.1b)$$

$$dy_3/dt = \phi_2 y_2 \qquad (4.3.1c)$$

With initial conditions $y_1 = 1$ and $y_2 = y_3 = 0$ at $t = 0$, the solution to these ODEs is

$$y_1 = \exp(-\phi_1 t) \qquad (4.3.2a)$$

$$y_2 = [\exp(-\phi_1 t) - \exp(-\phi_2 t)]\,\phi_1/(\phi_2 - \phi_1) \qquad (4.3.2b)$$

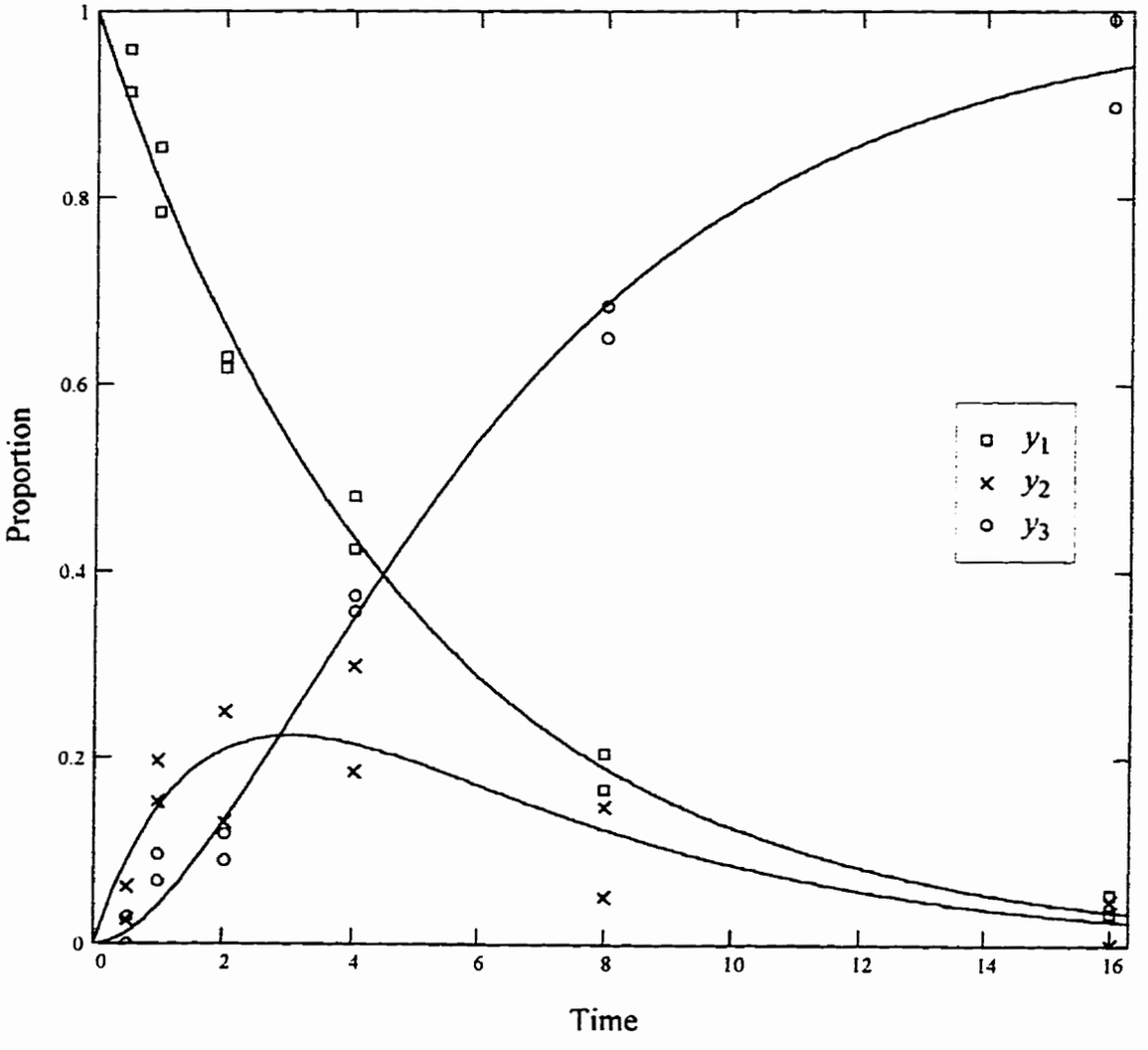
$$y_3 = 1 - y_1 - y_2 \qquad (4.3.2c)$$

The data Box and Draper used consisted of the twelve sets of measurements given in Table 4.3.1:

Table 4.3.1 Data for Box and Draper's Example

t	y_1	y_2	y_3
0.5	0.959	0.025	0.028
0.5	0.914	0.061	0.000
1	0.855	0.152	0.068
1	0.785	0.197	0.096
2	0.628	0.130	0.090
2	0.617	0.249	0.118
4	0.480	0.184	0.374
4	0.423	0.298	0.358
8	0.166	0.147	0.651
8	0.205	0.050	0.684
16	0.034	0.000	0.899
16	0.054	0.047	0.991

Figure 4.3.1 Determinant Criterion Fit to Data from Box and Draper's Example



Box and Draper parameterized the model in terms of $\theta = \ln(\phi)$. The determinant criterion applied to this data yields the parameter estimate:

$$\hat{\theta} = \ln(\hat{\phi}) = \begin{bmatrix} -1.572 \\ -0.702 \end{bmatrix} \tag{4.3.3}$$

The fit to the data is shown in Figure 4.3.1 on the previous page. The residuals yield the following estimates of the error standard deviations and correlation matrix:

$$\hat{\sigma}_\varepsilon = (\, 0.0319 \;\; 0.0498 \;\; 0.0321 \,)^T \tag{4.3.4}$$

$$\hat{\rho}_\varepsilon = \begin{bmatrix} 1 & -0.418 & 0.422 \\ -0.418 & 1 & 0.330 \\ 0.422 & 0.330 & 1 \end{bmatrix} \tag{4.3.5}$$

For the purpose of the simulation study, the true values of the model parameters, θ, and the error covariance matrix, Σ_ε, were set to the estimates from equations (4.3.3), (4.3.4) and (4.3.5). The vector of independent variables was set to:

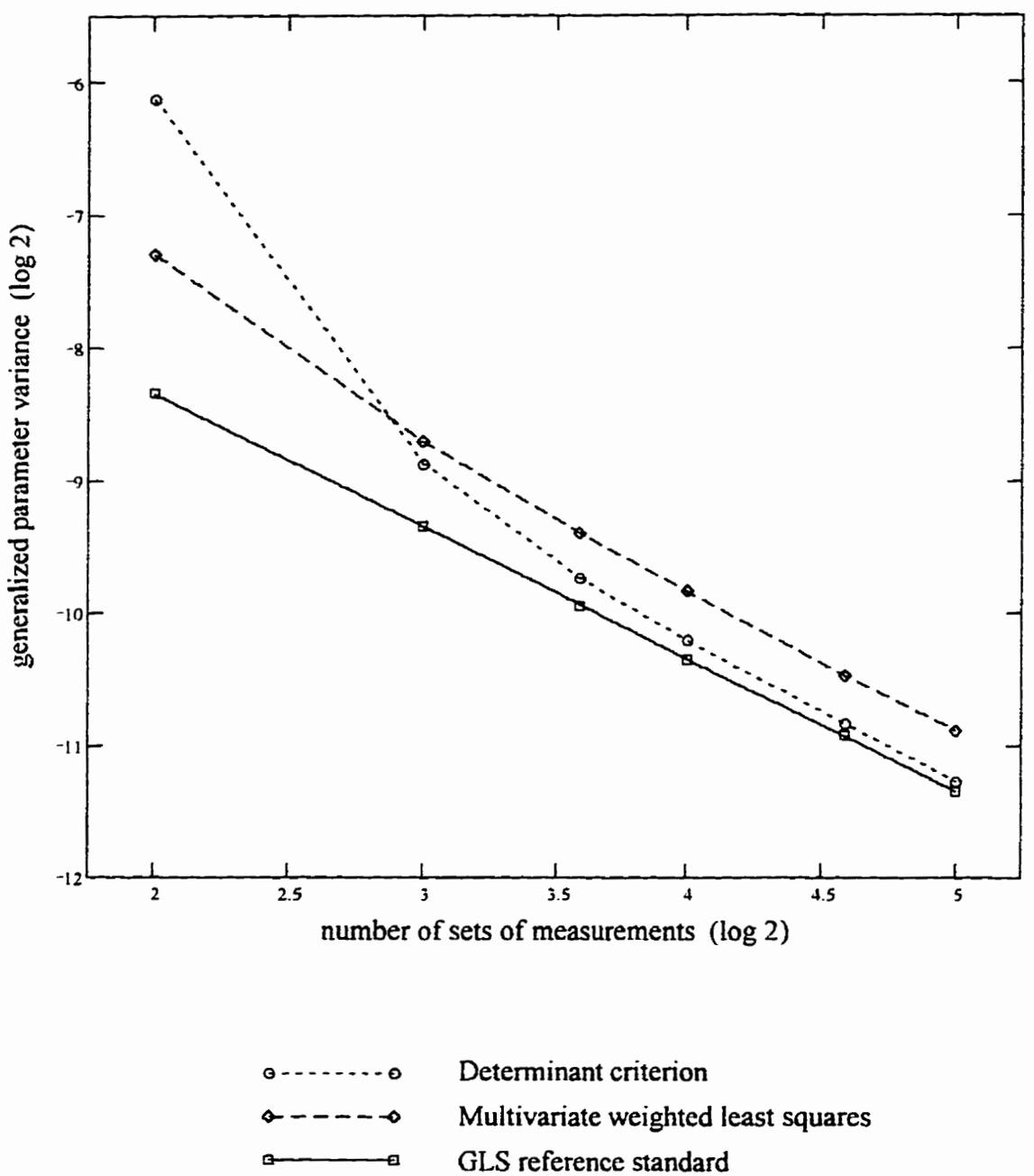
$$t = (\, 1 \;\; 2 \;\; 4 \;\; 8 \,)^T \tag{4.3.6}$$

Ten thousand synthetic data sets were generated with the simulated measurement errors being normally distributed with zero mean and covariance as given above.

Model parameters were estimated from each data set using three methods: the determinant criterion, MWLS, and generalized least squares (GLS) where the generalized weight matrix is the inverse of the true error covariance matrix, Σ_ε. The determinant of the parameter covariance matrix, $|\Sigma_\theta|$, for each of the three methods was estimated from 10000 sets of fitted model parameters using equation (3.3.9).

To investigate the effect of the number of sets of measurements, n, on the results, the simulation study was repeated for six full levels of replication giving results for $n = 4$, 8, 12, 16, 24 and 32 . These results, in the form of generalized parameter variance, $|\Sigma_\theta|^{1/p}$, plotted as a function of n on a log-log scale, are given in Figure 4.3.2 on the next page. The features of Figure 4.3.2 are qualitatively similar to those of Figures 4.1.4 and 4.2.3. The crossing of the curves for MWLS and the determinant criterion supports the interpretation of Figure 4.1.4 discussed in §4.1.

Figure 4.3.2 Effect of Sample Size on Parameter Variance



Chapter 5 - Discussion

The argument being made in this thesis can be broken down into the following components:

(a) In comparing multivariate weighted least squares (MWLS) with the determinant criterion for multiresponse parameter estimation, a goodness of fit test is problematic. As discussed in §2.2, Cox's (1962) generalization of the F test relies on the ratio of determinants of the residual covariance matrices. Therefore this test will always favor the determinant criterion. In §3.5, it was shown that MWLS was equivalent to minimizing the product of the diagonal elements of the deviation covariance matrix with respect to the model parameters. So a goodness of fit test based on the ratio of products of residual variances will always favor MWLS. Given, then, this problem with goodness of fit as an evaluation criterion in a multiresponse context, a frequentist evaluation approach was chosen.

(b) A frequentist approach to evaluating parameter estimation methods considers the frequency distribution of the parameter estimates. In a case study, the true values of the model parameters and measurement error structure are known. Therefore, Monte-Carlo simulation can be used to sample from the frequency distribution of the parameter estimates. As the sample size increases, the sample distribution of parameter estimates becomes more representative of the frequency distribution of parameter estimates. For evaluation purposes the sample distribution of parameter estimates is characterized by its mean and covariance matrix. The covariance matrix is further characterized by a suitable norm, its determinant, which is proportional to the square of the hypervolume of the parameter confidence region in parameter space.

(c) In §3.2 the numerical equivalence of the determinant criterion to an iterated generalized least squares scheme was discussed. In the iterated scheme, the parameters are estimated using generalized least squares where the weight matrix is the inverse of the residual covariance matrix. The residual covariance matrix, in turn, is determined from the parameter estimates, making for an iterated scheme. At convergence, the parameter estimates from this iterated scheme maximize the same likelihood function as the parameter estimates from the determinant criterion, thus establishing the numerical equivalence of the two methods. This result is originally due to Phillips (1976).

(d) In §3.3 it was shown that, when the measurement error covariance matrix is known, Aitken's (1935) generalized least squares estimation method minimizes the determinant of the estimate of the parameter covariance matrix. It does this by using a weight matrix that is the inverse of the error covariance matrix. When the error covariance matrix is unknown, of all weighting schemes, the determinant criterion minimizes the determinant of the estimate of the parameter covariance matrix. It does this by using a weight matrix that is the inverse of the residual covariance matrix. Thus, the determinant criterion, in effect, approximates the error covariance matrix by the residual covariance matrix.

(e) The three simulation studies of §4 clearly show that the determinant criterion's optimality property, giving the smallest determinant of the estimate of the parameter covariance matrix, is of limited practical significance. The simulation study of §4.1, based on Burke's (1994) parameter estimation problem, shows that even when the measurement errors satisfy the standard validating assumptions, MWLS can give a smaller determinant of the true parameter covariance matrix than the determinant criterion. The frequency distributions of the parameter estimates in Figure 4.1.1 show near normal distributions (parabolic on the log

scale used) from MWLS while the distributions from the determinant criterion show relatively heavy tails.

(f) In the simulation study of §4.2 based on the alpha-pinene parameter estimation problem, a defect was built into the measurement error structure. Figure 4.2.1 shows the effect of this defect on the frequency distributions of the parameter estimates. Again, MWLS yields near normal distributions of the parameter estimates while the determinant criterion yields distributions that are very far from normal. In fact, the distribution tails from the determinant criterion are so heavy that the parameter variances are an order of magnitude greater than those from MWLS.

(g) It was noted in (d) that the determinant criterion, in effect, approximates the error covariance matrix with the residual covariance matrix. Continuing with this perspective, MWLS approximates the error covariance matrix with the diagonal matrix of the diagonal elements of the residual covariance matrix. It would seem at first glance that the approximation used by the determinant criterion must be better than that used by MWLS. But the two estimation methods, applied to the same data, will not yield the same residual covariance matrices. Figures 4.1.2 and 4.2.1(f) show that the determinant criterion biases the residual covariance matrix as an estimate of the error covariance matrix in a way that MWLS does not. This bias is in the direction of ill-conditioning and it clearly can be severe enough to render the optimality property of the determinant criterion meaningless. This is because the validity of the optimality property depends on the residual covariance matrix being a good estimate of the error covariance matrix. (Davidson and MacKinnon (1993) note the tendency of the determinant criterion to bias the correlation between response residuals but the implication of this is not discussed.)

(h) Both MWLS and the determinant criterion yield estimates of the uncertainty in the point estimates of the parameters through equation (3.3.6). Figures 4.1.3 and 4.2.2 show that the estimates of the parameter uncertainty from MWLS can be substantially better than those from the determinant criterion. Here again, the difference can be attributed to the fact that the determinant criterion biases the residual covariance matrix as an estimate of the error covariance matrix to a greater degree than MWLS.

(i) In the simulation study of §4.3 based on Box and Draper's (1965) parameter estimation problem, the effect of the number of sets of measurements on the determinant of the parameter covariance matrix was examined. Figures 4.1.4, 4.2.3 and 4.3.2 show that with increasing replication in the data, the mean parameter variance for the determinant criterion converges to the mean parameter variance for generalized least squares where the weight matrix is the inverse of the known error covariance matrix. This is because, as the number of sets of measurements increases, the residual covariance matrix becomes a better estimate of the error covariance matrix. This illustrates the asymptotic nature of the determinant criterion's optimality property.

(j) This interpretation of the results from the simulation studies has been from a strictly frequentist perspective. The Bayesian context of §3.1 offers a different perspective. In this context, all of the elements of the measurement error covariance matrix are, strictly speaking, model parameters. They parameterize the error model. Although Box and Draper (1965) integrated them out in their Bayesian approach and Bard (1974) eliminated them as parameters by a maximization procedure in his maximum likelihood approach, they are model parameters nevertheless. A simple interpretation of the results is to regard the determinant criterion as being associated with an overparameterized measurement error model when the number of sets of measurements is not large. From a

Bayesian perspective, it might be argued that the error model is not overparameterized and that the posterior probability density function for the parameters faithfully represents the information contained in the data. However, the determinant criterion condenses the posterior probability density function into point estimates of the model parameters. There is no guarantee that these condensed point estimates will have desirable attributes like small bias and variance.

(k) These results show that while the determinant criterion is asymptotically optimal for certain measurement error structures, MWLS can give substantially better results when either the data set is not large or the measurement error structure does not conform to that which validates the determinant criterion (§3.1). The word that best describes this situation is 'robust'. The central point of this thesis is that MWLS is a more robust estimation method than the determinant criterion. Figures 4.1.4, 4.2.3 and 4.3.2 show that this robustness does come at a price. When there is a large amount of data whose measurement error structure conforms to that which validates the determinant criterion (§3.1), then MWLS will give parameter estimates with a larger variance than those given by the determinant criterion.

The argument presented here does depend on the empirical observation that the determinant criterion tends to bias the residual covariance matrix as an estimate of the error covariance matrix and that this bias is in the direction of ill-conditioning. This kind of bias is not nearly so evident with MWLS. The weakness with the general argument is that it is based on a limited number of case studies. While a larger number of case studies would strengthen this empirically based argument, a semi-quantitative explanation for the empirical observation regarding bias of the residual covariance matrix would be preferable.

Following is a semi-quantitative explanation for the mathematically tractable case of two response variables. In this case, the residual covariance matrix is a symmetric two by two matrix whose elements will be denoted by a, b, and c where a and b are the diagonal variance elements and c is the off-diagonal covariance element. The determinant criterion minimizes the determinant of the residual covariance matrix which is given by:

$$\begin{vmatrix} a & c \\ c & b \end{vmatrix} = ab - c^2 \qquad (5.0.1)$$

The eigenvalues of the residual covariance matrix are given by:

$$\text{eigenvalues}\begin{bmatrix} a & c \\ c & b \end{bmatrix} = \frac{1}{2}\left(a + b \pm \sqrt{a^2 - 2ab + b^2 + 4c^2}\right) \qquad (5.0.2)$$

The condition number of the residual covariance matrix is the ratio of the larger to the smaller eigenvalue. With some algebraic manipulation, the condition number can be expressed in terms of the determinant, $ab - c^2$, and the trace, $a + b$:

$$\text{cond. no.}\begin{bmatrix} a & c \\ c & b \end{bmatrix} = \frac{(a+b)^2}{4(ab-c^2)}\left[1 + \sqrt{1 - \frac{4(ab-c^2)}{(a+b)^2}}\right]^2 \qquad (5.0.3)$$

Equation (5.0.3) provides the incriminating link between minimizing the determinant and biasing the condition number. The parameter estimates from the determinant criterion will yield a residual covariance matrix determinant that is necessarily smaller than the residual covariance matrix determinant corresponding to the true parameters. Thus, the determinant criterion biases the determinant of the residual covariance matrix towards zero. By equation (5.0.3), this will bias the condition number of the residual covariance matrix in the direction of ill-conditioning which is exactly what is observed. (This assumes that the trace will be relatively unaffected by bias in the determinant. Given that the trace is the sum of variances this assumption seems reasonable.)

MWLS, on the other hand, does not minimize the determinant of the residual covariance matrix. In §3.5, it was shown that MWLS is equivalent to minimizing the product of the diagonal elements of the residual covariance matrix. For the two response case under consideration, MWLS will bias the product $a\,b$ downward. Now, equation (5.0.3) can be rewritten as:

$$\text{cond. no.}\begin{bmatrix} a & c \\ c & b \end{bmatrix} = \frac{2 + a/b + b/a}{4(1 - c^2/a\,b)}\left[1 + \sqrt{1 - \frac{4(1 - c^2/a\,b)}{2 + a/b + b/a}}\right]^2 \qquad (5.0.4)$$

The empirical observation that the condition ratio is relatively unbiased by minimizing $a\,b$ can be explained by the reasonable assumption that bias in the product $a\,b$ does not imply bias in the squared correlation term, $c^2/a\,b$, or the ratio of a to b. (Note that because a, b and c are mutually dependent on the parameter estimates, c is not independent of a and b. Therefore, minimizing $a\,b$ will not necessarily bias $c^2/a\,b$.) Furthermore, because a and b are variances, one would not expect a large bias in the product $a\,b$ (except when the number of sets of measurements, n, is not greater than the number of responses, m). Therefore, it follows that there can be a significant difference, in terms of the effect on condition number, between biasing just the product $a\,b$ (MWLS) and biasing the determinant, $a\,b - c^2$ (determinant criterion).

This argument can be summed up concisely, if somewhat loosely, using teleological language. In minimizing the determinant of the residual covariance matrix, the determinant criterion has an incentive to find a point in parameter space that will introduce a near linear dependency in this matrix even if the dependency is purely spurious due to the random nature of the measurement errors. This is what can make the distribution of the parameter estimates for the determinant criterion far from normal in the tails. On the other hand, MWLS has no incentive to search for spurious near linear dependencies in parameter space. Being well behaved in this regard, MWLS gives parameter estimate distributions that are much more nearly normal.

This discussion leads to a question of practical significance: For a particular parameter estimation problem, which method should be used, MWLS or the determinant criterion? This question can be answered definitively (through simulation) only when the true measurement error model is known. In this situation the question is moot because generalized least squares should be used to estimate the parameters. For practical estimation problems the true measurement error model is almost certainly unknown.

A purely pragmatic answer to the question can be offered by considering the two possible outcomes of applying MWLS and the determinant criterion to an estimation problem. Suppose that the parameter estimates from the two methods can be judged by some criterion to be either significantly different or not significantly different. If the estimates are significantly different, then the results presented in this thesis would call into question the estimates from the determinant criterion. In this case the estimates from MWLS should be chosen. If, on the other hand, the estimates are not significantly different, then the estimates from MWLS can still be chosen.

This question can also be looked at from a risk-benefit perspective. The alpha-pinene example shows that when the measurement error structure does not conform to that which validates the determinant criterion (§3.1), use of the determinant criterion can be very risky relative to MWLS. Figures 4.1.4, 4.2.3 and 4.3.2 all show that, even when the measurement error structure conforms to that which validates the determinant criterion (§3.1), the relative benefit from using the determinant criterion is small or nonexistant. This leads me to conclude that, in general, MWLS is a preferable alternative to the determinant criterion for multiresponse parameter estimation.

Chapter 6 - Summary and Conclusions

The use of Box & Draper's determinant criterion for multiresponse parameter estimation is justified on Bayesian grounds for the case where the measurement errors are multinormally distributed within measurement sets and independent between measurement sets. This study shows that even when these conditions hold, multivariate weighted least squares (MWLS) can give parameter estimates with a smaller variance when the number of sets of measurements is not large. This empirical result is seemingly at variance with a theoretical result to the effect that the determinant criterion is optimal in the sense of giving the smallest parameter variance. But the theoretical result is based on an approximation that is strictly valid only for infinite data sets. So, in this sense, the determinant criterion is only asymptotically optimal. Whether an optimal multiresponse estimation method for finite data sets exists or not is an interesting question.

This study also shows that when the validating assumptions of the determinant criterion do not hold, as in the alpha-pinene case, the variances of the parameter estimates from the determinant criterion can be as much as an order of magnitude greater than that for MWLS. This shows a notable lack of robustness in the determinant criterion that is attributable to an inherent tendency of the determinant criterion to bias the residual covariance matrix as an estimate of the error covariance matrix. The alpha-pinene case also highlights the problem of correctly identifying redundant response variables when applying the determinant criterion. Therefore, in the final analysis, it is the robust character of MWLS that leads me to conclude that its use is preferable to the determinant criterion for practical multiresponse parameter estimation applications.

The results described in this thesis suggest a direction for further work based on the key issue of the conditioning of the generalized weight matrix. MWLS makes this matrix well-conditioned by making it diagonal. This is simple and effective but it may not be the best way to make the matrix well-conditioned. Numerical methods such as Levenberg- Marquardt optimization often deal with ill-conditioned positive definite matrices by augmenting the diagonal elements. MWLS can be regarded as taking this approach to the limiting extreme where the diagonal elements are multiplied by an infinitely large constant. Further work might explore less extreme augmentation schemes or other approaches to avoid ill-conditioning of the generalized weight matrix.

References

Aerdts, A.M. (1993). Microstructure of Styrene Methyl Methacrylate Copolymers Grafted onto Polybutadiene Seeds. *PhD Thesis*. Technische Universiteit Eindhoven, Denmark

Aitken, A.C. (1935). On Least Squares and Linear Combinations of Observations. *Proc. Roy. Soc. Edinburgh* A, **55**, 42-47.

Bard, Y. (1974). *Nonlinear Parameter Estimation*. Academic Press.

Bates, D.M. & Watts, D.G. (1985). Multiresponse Estimation with Special Application to Linear Systems of Differential Equations. *Technometrics*, **27**, 329-339

Bates, D.M. & Watts, D.G. (1988). *Nonlinear Regression Analysis and its Applications*. Wiley

Beauchamp, J.J. & Cornell, R.G. (1966). Simultaneous Nonlinear Estimation. *Technometrics*, **8**, 319-326

Biegler, L.T., Damiano, J.J., & Blau, G.E. (1986). Nonlinear Parameter Estimation: a Case Study Comparison. *AIChE J*, **32**, 29-45

Box, G.E.P. & Draper, N.R. (1965). The Bayesian Estimation of Common Parameters from Several Responses. *Biometrika*, **52**, 355-365

Box, G.E.P., Hunter, W.G., MacGregor, J.F., & Erjavec, J. (1973). Some Problems Associated with the Analysis of Multiresponse Data. *Technometrics*, **15**, 33-51

Burke, A.L. (1994). Model Discrimination Techniques for the Modelling of Copolymerization Reactions. *PhD Thesis*, University of Waterloo, Waterloo

Carroll, R.J. & Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall

Cox, D.R. (1962). Further Results on Tests of Separate Families of Hypotheses. *J. Royal Stat. Soc. B*, **24**, 406-424

Davidson, R. & MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics.* Oxford University Press

Draper, N.R. & Smith, H. (1981). *Applied Regression Analysis, 2nd Ed.* Wiley

Fugguitt, R.E. & Hawkins, J.E. (1947). Rate of Thermal Isomerization of Alpha-Pinene in the Liquid Phase. *J. Am. Chem. Soc.*, **69**, 319

Gallant, R.A. (1975). Seemingly Unrelated Nonlinear Regressions. *J. Econometrics*, **3**, 35-50

Kang, G. & Bates, D.M. (1990). Approximate Inferences in Multiresponse Regression Analysis. *Biometrika*, **77**, 321-31

Khuri, A.I. (1990). The Effect of Response Scaling on the Detection of Linear Dependencies Among Multiresponse Data. *Metrika*, **37**, 217-231

Koenig, J.L. (1980). *Chemical Microstructure of Polymer Chains.* Wiley

McLean, D.D., Pritchard, D.J., Bacon, D.W. & Downie, J. (1979). Singularities in Multiresponse Modelling. *Technometrics*, **21**, 291-298

Phillips, P.C.B. (1976). The Iterated Minimum Distance Estimator and the Quasi-Maximum Likelihood Estimator. *Econometrica*, **44**, 449-460

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (1992). *Numerical Recipes in Fortran, 2nd Ed.* Cambridge University Press

Searle, S.R. (1982). *Matrix Algebra Useful for Statistics.* Wiley

Seber, G.A.F. & Wild, C.J. (1989). *Nonlinear Regression.* Wiley

Stewart, W.E., Caracotsios, M. & Sorensen, J.P. (1992). Parameter Estimation from Multiresponse Data. *AIChE J*, **38**, 641-650

Stewart, W.E., & Sorensen, J.P. (1981). Bayesian Estimation of Common Parameters from Multiresponse Data With Missing Observations. *Technometrics*, **23**, 131-141

Takada, H., Ullah, A. & Chen, Y.M. (1995). Estimation of the Seemingly Unrelated Regression Model When the Error Covariance Matrix is Singular. *J. Applied Statistics*, **22**, 517-530

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *J. Am. Stat. Assoc.*, **57**, 348-368