1978

Estimation Of Linear Structural And Functional Relationships

Tak Kwan Mak

ESTIMATION OF LINEAR STRUCTURAL AND FUNCTIONAL RELATIONSHIPS

by

Tak Kwan Mak

Department of Mathematics

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario

September, 1978

ABSTRACT

Models, estimation problems and some well known solution procedures of linear structural and functional relationships are described. Results of this thesis are summarized. Connections with other areas such as covariance structure analysis and simultaneous-equation models are indicated. Recently developed methods such as estimation using serial correlations, matrix attenuation and Bayesian approach are reviewed. Further problems in theory and applications are then proposed.

Maximum likelihood estimation of the five parameters of a linear structural relationship $y = \alpha + \beta x$ when α is known is considered. The parameters are β, the two variances of observation errors on x and y, the mean and variance of x. When the estimates cannot be obtained by solving a simple system of five equations, they are found by maximizing the likelihood function directly.

Maximum likelihood estimation of the parameters of a linear structural relationship $y = \alpha + \beta x$ when r repeated observations are made on each (x,y) is considered. The estimate of β is found to be a root of a fourth degree polynomial and to be consistent as r increases. Estimates of other parameters can then be easily obtained. The asymptotic variances and covariances of the estimates of the parameters are computed through a simplified procedure.

Two adaptive procedures of reducing the finite sample mean square errors of consistent estimates of β in a linear structural relationship $y = \alpha + \beta x$ model are proposed. They are based on the

idea of constructing estimates through inspecting the sample estimates of the asymptotic variances of the original estimates. These two procedures are applied to the estimates of Geary, Wolfowitz and a modified Scott's estimate, which is obtained from a proposed method of constructing conjugate estimates. Monte Carlo experiments show that the procedures yield much higher precision in finite samples and in general are more efficient than the ordinary least squares estimate, and the modified Scott's estimate is superior to the estimates of Geary and Wolfowitz. Extension to more than one independent variable is considered.

By considering a model similar to a factor analysis model, the maximum likelihood estimate of the slope parameter β of the linear structural relationship when the errors of observations are correlated with covariance matrix known to within a proportionality factor, is obtained. It is the same as the maximum likelihood estimate of β when the covariance matrix is known completely and is also identical to that of β in the linear functional relationship. These results are generalized to multivariate case when there is one independent variable. In the functional form of the model without normality assumption in the error terms, an estimate being consistent under some mild conditions is obtained by maximizing certain quadratic forms. This estimate coincides with the maximum likelihood estimate under normality assumption. Simple methods of computing the estimate are given for some special cases.

Regularity conditions under which the maximum likelihood estimate of the parameter $\underset{\sim}{\theta}$ in the presence of incidental parameters is asymptotically normal are given. Although the probability limit of such estimate of θ is not necessarily equal to the true parameter, it is seen that in some situations a consistent estimate of θ, which is a function of the estimate, can be constructed. The results are applied to the estimation of a linear functional relationship to obtain conditions under which the maximum likelihood estimates of the intercept and slope parameters are consistent and asymptotically normal. The method is different from the usual approach which relies on the explicit form of the maximum likelihood estimates.

ACKNOWLEDGEMENTS

I wish to express my deep gratitude to my supervisor, Professor L. K. Chan for suggesting this topic, for his extremely helpful guidance and continuous encouragement throughout my Ph.D. program of study. Thanks also go to Dr. E. R. Mead for reading my thesis and making many helpful suggestions.

I am also grateful to Ms. Janet Williams for her excellent typing and to the University of Western Ontario for financial support provided during my period of study.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

The problem of estimating linear structural and functional relationships has a long history and occurs frequently in the behavioral sciences, economics, education and the natural sciences, particularly in biology. An early inspiration to the subject was given by Lindley (1947) who first showed that a consistent estimate of the slope parameter does not exist in a structural relationship model and its "maximum" likelihood estimate is inconsistent in the corresponding functional relationship. Lindley's results therefore imply that additional information is required (he assumed that ratio of the error variances is known) in the estimation problems involved, in order to obtain satisfactory results. Much of the further work done was concentrated on consistent estimation under additional information or under different assumptions and was summarized in the review by Madansky (1959). More recently, comprehensive reviews were given by Kendall and Stuart (1973, ch. 29) and Moran (1971). Moran gave particular insight into various aspects of the subject and stimulated further research by clearly indicating the underlying principles and rationale involved. In this section we define the problem through the case of one linear relationship between two variables. In the next section, we recapitulate and comment on some fundamental results in the estimation of linear structural and functional relationships, leaving the details to the reviews mentioned above and the references of original work cited there. In section 3, we describe the main results obtained in this thesis, and in section 4 we review recent alternative approaches and development not included in the above reviews. This chapter is concluded with some proposed problems.

Consider two unobservable variables x and y linearly related by $y = \alpha + \beta x$. To estimate the slope parameter β, a sample of size n is taken. The $(x_1,y_1),\ldots,(x_n,y_n)$ are observed through $(\xi_1,\eta_1),\ldots,(\xi_n,\eta_n)$, respectively, where

$$\begin{aligned} y_i &= \alpha + \beta x_i, \\ \xi_i &= x_i + \delta_i, \quad \eta_i = y_i + \epsilon_i, \quad i = 1,\ldots,n. \end{aligned} \tag{1.1}$$

The (δ_i,ϵ_i) are independent and identically distributed (i.i.d.) as $N(\underset{\sim}{0},\underset{\sim}{\Sigma})$ and

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \sigma_\delta^2 & \sigma_{\delta\epsilon} \\ \sigma_{\delta\epsilon} & \sigma_\epsilon^2 \end{bmatrix}.$$

Two situations arise: the x_i can either be fixed constants or i.i.d. random variables independent of the (δ_j,ϵ_j). In the former case the relationship $y = \alpha + \beta x$ in (1.1) is usually referred to as functional relationship and in the latter case as structural relationship. In structural relationship, it is assumed that each x_i has finite mean μ and variance σ^2. Also $\underset{\sim}{\Sigma}$ may or may not be diagonal, depending on whether the errors δ_i and ϵ_i are uncorrelated or correlated. Thus four different models can be derived from (1.1).

2. THE ESTIMATION PROBLEMS

2.1. Linear Structural Relationship with Uncorrelated Errors

In the present case, $E(\delta_i\epsilon_i) = \sigma_{\delta\epsilon} = 0$ and the x_i are i.i.d. It is also assumed that the x_i are normally distributed with $E(x_i) = \mu$ and $Var(x_i) = \sigma^2$. Thus each of the (ξ_i,η_i) have a common bivariate normal distribution completely specified by

$$
\begin{aligned}
E(\xi_i) &= \mu, \\
E(\eta_i) &= \alpha + \beta\mu, \\
\mathrm{Var}(\xi_i) &= \sigma^2 + \sigma_\delta^2, \qquad (2.1.1)\\
\mathrm{Var}(\eta_i) &= \beta^2\sigma^2 + \sigma_\epsilon^2, \\
\mathrm{Cov}(\xi_i, \eta_i) &= \beta\sigma^2.
\end{aligned}
$$

If all of the μ, α, β, σ^2, σ_δ^2 and σ_ϵ^2 are unknown, there are six parameters in five equations and only μ is completely determined by $E(\xi_i)$. Since it is possible to choose (cf. Moran, 1971) different sets of $(\alpha, \beta, \sigma^2, \sigma_\delta^2, \sigma_\epsilon^2)$ which together with $E(\xi_i) = \mu$, give the same first two moments, and hence the same distribution of the (ξ_i, η_i), the parameters β, $\sigma^2, \sigma_\delta{}^2$ and σ_ϵ^2 are unidentifiable (cf. Reiersøl, 1950). To avoid this difficulty, additional information is therefore required. We list below different conditions under which β is identifiable and can be consistently estimated. Details of the methods used to solve the problems can be found in Kendall and Stuart (1973, ch. 29), Malinvaud (1970, ch. 10), Moran (1971), and also in the references listed there.

A. Information concerning the variances σ_δ^2 and/or σ_ϵ^2.

Maximum likelihood estimates (MLE) of β have been obtained in the following cases:

A1. σ_δ^2 (or σ_ϵ^2) is known,

A2. $\lambda = \sigma_\epsilon^2/\sigma_\delta^2$ is known,

A3. σ_δ^2 and σ_ϵ^2 are both known.

B. Replications are available.

In this case, corresponding to each (x_i, y_i), r independent replicates (ξ_{ij}, η_{ij}), $j = 1,2,\ldots r$, are observed. The maximum likelihood solution is given in chapter 3.

C. α is known and $\mu \neq 0$.

A complete maximum likelihood solution is given in chapter 2.

D. Grouping of observations.

Suppose $n = 2m$ and that it is possible to divide the (ξ_i, η_i) into two groups of size m according to a certain criterion which is unaffected by the errors δ_i and ϵ_i. Then, if with probability tending to 1 $\underline{\lim} |\bar{x}^{(1)} - \bar{x}^{(2)}| > 0$, where the $\bar{x}^{(j)}$ are the means of the x_i in the j^{th} group, $j = 1,2$, β can be estimated consistently.

E. Existence of instrumental variables.

Suppose there exist i.i.d. random variables z_i, $i = 1,\ldots, n$, such that the z_i are independent of the (δ_j, ϵ_j) but the $\mathrm{Cov}(z_i, x_i)$ are non-zero. Then β can be estimated consistently.

The problem of unidentifiability of β occurs when we assume that the $(x_i, \delta_i, \epsilon_i)$ are normal. If we assume only that the (δ_i, ϵ_i) are normal but the x_i are not, then β can be estimated consistently by the method of moments and cumulants.

2.2. Linear Structural Relationship with Correlated Errors

For this situation, investigations in the literature are sparse. We now make comments on conditions 2.1A to 2.1E, and assume that the $(x_i, \delta_i, \epsilon_i)$ are normal. Similar to the discussion in section 2.1, we have five equations as in (2.1.1) except that the last one is replaced by $\mathrm{Cov}(\xi, \eta) = \beta\sigma^2 + \sigma_{\delta\epsilon}$ and again all the parameters except μ are not identifiable.

It is clear that none of the conditions in 2.1A makes β identifiable although they do make some of the other parameters identifiable (identifiable parameters are those which can be solved in terms of the first two moments and the known parameters in the system of equations). But as generalizations to 2.1A2 and 2.1A3, one can consider the situation when $\underset{\sim}{\Sigma} = c\underset{\sim}{A}$ and $\underset{\sim}{A}$ is known. When c is an unknown scalar, we say $\underset{\sim}{\Sigma}$ is known to within a proportionality factor. This is studied in chapter 4.

When replication is possible (2.1B), we can estimate $\underset{\sim}{\Sigma}$ by $(n-1)^{-1}\sum(\xi_i - \bar{\xi}.,\ \eta_i - \bar{\eta}.)'(\xi_i - \bar{\xi}.,\ \eta_i - \bar{\eta}.)$, where $\sum$ denotes $\sum_{i=1}^{n}$, $\bar{\xi}. = \sum \xi_i/n$, $\bar{\eta}. = \sum \eta_i/n$ and use this estimated $\underset{\sim}{\Sigma}$ in the MLE of β when $\underset{\sim}{\Sigma}$ is known. However, it would be interesting to consider the direct MLE of β although the algebra would be complicated and an explicit solution might not exist.

When α is known (2.1C), β is identifiable but σ^2, σ_δ^2, σ_ϵ^2 and $\sigma_{\delta\epsilon}$ are not.

When 2.1D or 2.1E is satisfied, it can be easily seen that the same estimates would also estimate β consistently in the present case.

Finally, it should be pointed out that, for estimating β, there seems to be no theory existing in the literature using the method of moments or cumulants (if it is possible).

2.3. Linear Functional Relationship

Here the difficulty of unidentifiability corresponding to structural relationship is reflected in that roots of the likelihood equations (in the case $\sigma_{\delta\epsilon} = 0$) satisfy $\beta^2 = \sigma_\epsilon^2/\sigma_\delta^2$ and that the MLE of β is inconsistent (as first shown by Lindley, 1947). Solari (1969) then showed that the roots are saddle points and do not give a local maximum. To settle the problem of finding consistent estimates of β, conditions similar to those in structural relationship are usually considered.

When σ_δ^2 is known and $\sigma_{\delta\epsilon} = 0$, the MLE of β for structural relationship is still consistent in functional relationship, but it is no longer the MLE of β which is inconsistent (Moberg and Sundberg, 1978). The MLE of β when $\underset{\sim}{\Sigma}$ (which can be correlated or uncorrelated) is known to within a proportionality factor or known completely can be found in Kendall and Stuart (1973, ch. 29) and Sprent (1969).

When replication is possible, the MLE of β for the correlated errors model was obtained by Anderson (1958), and the uncorrelated errors model was discussed by Barnett (1970), and Dolby and Lipton (1972).

When α is known and $\sum x_i/n \to \mu \neq 0$ as $n \to \infty$, the estimate in structural relationship is still consistent for β in the functional relationship with correlated or uncorrelated errors. The same can be said about the estimates constructed in the structural relationship based on the method of grouping and instrumental variables (but the conditions in 2.1E should be reformulated as: there exist independent z_i, $i = 1, \ldots, n$, such that $\underline{\lim} \sum_n (x_i - \bar{x}.) E(z_i)/n$ is positive).

When the (δ_i, ϵ_i) are not necessarily normal and possibly correlated, Sprent (1966) proposed a generalized least squares method of estimating β. Dolby (1972) then showed that under normality of the (δ_i, ϵ_i), the procedure is the same as for the MLE of β.

2.4. Multivariate Generalization

A multivariate generalization of (1.1) is

$$\begin{aligned} \eta_i &= \alpha + \beta x_i + \epsilon_i, \\ \xi_i &= x_i + \delta_i, \end{aligned} \tag{2.4.1}$$

where β is a $q \times p$ matrix to be estimated, η_i, α, ϵ_i, ξ_i, x_i and δ_i are vectors of either q or p components, and the (δ_i, ϵ_i) are i.i.d. as $N(0, \Sigma)$. Only the (ξ_i, η_i) are observable, and as before, we have either a structural relationship or a functional relationship depending on whether the x_i are i.i.d. random vectors independent of the (δ_i, ϵ_i) or fixed constants. Results for the particular case $p = 1$ and $q = 1$ which has already been discussed have fairly natural generalizations to the multivariate case. More details, discussions and references can be found in Anderson (1958, 1976), Gleser and Watson (1973), Kendall and Stuart (1973, ch. 29), Malinvaud (1970, ch. 10), Moran (1971), Schneeweiß (1976) and Sprent (1969). See Robinson (1977) for a different approach.

3. MAIN RESULTS

3.1 Introduction

This thesis is partly concerned with the estimation of unknown parameters in linear structural and functional relationships under various assumptions which have been discussed in the literature but no satisfactory or asymptotically optimal procedures had been attained. We also establish conditions under which the MLE of β in (2.4.1) in both structural and functional relationships are the same when Σ is known. Then one does not have to worry whether the $\underset{\sim}{x}_i$ should be considered as being generated from a superpopulation or as fixed constants. The theory of maximum likelihood estimation when the number of unknown parameters increases with sample size is considered and is applied to linear functional relationship.

The results are contained in chapters 2 to 6. The presentations of the five chapters are self-contained so that they can be read independently.

3.2. Chapter 2

Suppose α is known and μ is known to be non-zero in the structural relationship model (1.1) with uncorrelated errors (cf. 2.1C). Although $(\bar{\eta}. - \alpha)/\bar{\xi}.$ is a consistent estimate of β, as Zellner (1971) pointed out, the maximum likelihood estimates obtained by equating the first two sample moments of (ξ, η) to their corresponding expected values (cf. (2.1.1)), may give negative estimates of σ^2, σ_δ^2 and σ_ϵ^2 and therefore are not admissible. Moran (1971) also discussed the situation intuitively and included the problem of finding the complete maximum likelihood solution in his list of unsolved problems. We solve this problem in chapter 2.

3.3. Chapter 3

Consider the same model again but with α unknown. In conditions 2.1A we assume some or all of the σ_δ^2 and σ_ϵ^2 are known. Frequently the information on σ_δ^2 and σ_ϵ^2 is gained through replication and 2.1B would be more interesting. As Moran (1971) pointed out, although the procedure of using the estimates $\hat{\sigma}_\delta^2 = \sum (\xi_i - \bar{\xi}.)^2/(n-1)$ and $\hat{\sigma}_\epsilon^2 = \sum (\eta_i - \bar{\eta}.)/(n-1)$ in the estimate of β when both σ_δ^2 and σ_ϵ^2 are known usually gives better results than the method of using variance components, it is still not the optimal procedure since $\bar{\xi}_{i\cdot}$ and $\bar{\eta}_{i\cdot}$, where $\bar{\xi}_{i\cdot} = \sum_{j=1}^{r} \xi_{ij}/r$ and $\bar{\eta}_{i\cdot} = \sum_{j=1}^{r} \eta_{ij}/r$ themselves contribute some information about σ_{δ}^2 and σ_{ϵ}^2. He included the problem of finding the MLE of β in his list of unsolved problems. In this chapter we show that the MLE of β is given by a root of a fourth degree polynomial and the MLE of other parameters can be found easily once $\hat{\beta}$ is computed. Thus the problem of solving a system of likelihood equations by iterative methods is avoided. The asymptotic variances and covariances of the estimates of the parameters are computed through a simplified procedure.

3.4. Chapter 4

Suppose now we assume that the x_i are non-normal and have non-zero third central moments. In this case, consistent estimates of β had been constructed by Geary (1942), Scott (1950) and Wolfowitz (1952) based on the method of moments and cumulants; however, Madansky (1959) and Malinvaud (1970, ch. 10) observed that estimates using higher moments are not very precise. Madansky gave an example where the approximate mean square error of Geary's estimate is so large that it is useless. Quite often these estimates even perform much worse than the biased ordinary least squares (OLS) estimate. In this chapter we propose two adaptive procedures to increase the finite sample efficiencies of the estimates of Geary, Wolfowitz and a modified Scott's estimate based on the proposed idea of conjugate estimates. Monte Carlo experiments are used to demonstrate that the procedures yield much higher precision in finite samples and in general these are more efficient than the OLS estimate. The modified Scott's estimate is also seen to dominate the estimates of Geary and Wolfowitz.

3.5. Chapter 5

Consider 2.1A2 (linear structural relationship with $\sigma_{\delta\epsilon} = 0$). Since $\sigma_\epsilon^2 = \lambda\sigma_\delta^2$ and λ is known, we now have only five unknown parameters in the five equations in (2.1.1) and a consistent estimate of β can be obtained by solving the equations with the left hand sides replaced by the corresponding sample estimates. In fact this gives the MLE of β. When both σ_δ^2 and σ_ϵ^2 are known (2.1A3), we have only four unknowns in five equations and it is easily seen that by choosing different subsets of the system of equations in (2.1.1), we get different consistent estimates of β. This problem of "overidentification", as noted by Madansky (1959), was solved by Barnett (1967) and Birch (1964) by solving the likelihood equations directly but the algebra involved is quite complicated as indicated by Dolby (1976). The same difficulty arises in the correlated errors case when $\underset{\sim}{\Sigma}$ is known and we are not aware of any published results on the MLE of β in this case. By specializing the results obtained for a general model discussed in this chapter to the model (2.4.1) with structural relationship and a general $\underset{\sim}{\Sigma}$, we are able to obtain the maximum likelihood solution to $\underset{\sim}{\beta}$ when $p = 1$ and $\underset{\sim}{\Sigma}$ is known to within a proportionality factor or known completely (when $q = 1$, this is the case in section 2.2). The MLE of $\underset{\sim}{\beta}$ in both cases are the same and are also identical to the MLE in the corresponding functional relationship model. The model with $p = 1$ is important in econometrics (see comments by Robinson, 1977) and in many practical situations. Real examples were given by Barnett (1969) and Taylor (1973) where different instruments measuring a certain lung function had to be compared with a more expensive and hard to operate standard instrument.

In this chapter we also consider as a particular case of a general model, the functional relationship model of (2.4.1) when $\underset{\sim}{\Sigma}$ is known, and the normality assumption on the (δ_i, ϵ_i) is relaxed. We propose an estimate of $\underset{\sim}{\beta}$ that maximizes a certain quadratic form in the observations. Easy computational methods that do not require iteration are also given.

3.6. Chapter 6

When we have functional relationship in (1.1) or (2.4.1), the number of unknown parameters increases as $n \to \infty$ since each time we are introducing an additional x_i. Thus the asymptotic theory of the MLE in the i.i.d. case does not apply here. Neyman and Scott (1948) called an unknown parameter which appears only a finite number of times in the probability distributions of the observed variables an incidental parameter and called the others structural parameters. In the present case, the x_i are incidental parameters while the μ, α, β, σ^2, σ_δ^2 and σ_ϵ^2 are structural parameters. Neyman and Scott considered the general problem of estimation in the presence of incidental parameters and in particular demonstrated that the MLE of structural parameters might not be consistent. Patefield (1977) also pointed out that the asymptotic covariance matrix of the MLE of structural parameters is not necessarily given by the inverse of the information matrix. Thus the asymptotic theory is quite different from the usual i.i.d. case and needs special treatment. In chapter 6, regularity conditions are given in order for the MLE of structural parameters to be convergent (not necessarily to the true parameters) and asymptotically normally distributed. Although the MLE might not be consistent, it is seen that quite often a function of the MLE is consistent. These results are applied to the estimation problem in 2.1A2.

4. SOME RECENT APPROACHES AND CONTRIBUTIONS

4.1. Estimation When the Serial Correlation of the True x_i is Non-Zero

A method was proposed by Karni and Weissman (1974) for the functional relationship with uncorrelated errors model of (1.1) (when $\alpha = 0$ and $\sum x_i/n \to 0$). Assuming that the limit ρ_1 of the serial correlation of lag 1 of the x_i

$$\lim_{n\to\infty} \sum_{i=2}^{n} x_i x_{i-1} / \sum_{i=1}^{n} x_i^2$$

is non-zero, they showed that asymptotically,

$$\begin{aligned} s_{\eta\eta} &= S_x^2 + \sigma_\delta^2, \\ s_{\xi\eta} &= \beta S_x^2, \\ s_{\Delta\xi\Delta\xi} &= 2S_x^2(1-\rho_1) + 2\sigma_\delta^2, \\ s_{\Delta\xi\Delta\eta} &= 2\beta S_x^2(1-\rho_1), \end{aligned}$$

where s denotes the sample product moment of its subscripts, $s_x^2 = \sum x_i^2/n$ (limit exists as $n \to \infty$), and $\Delta a_i = a_i - a_{i-1}$ for every sequence $\{a_i\}_{i=1}^{\infty}$. Solving these equations, one gets a consistent estimate of β given by $\hat{\beta} = (s_{\xi\eta} - 2^{-1}s_{\Delta\xi\Delta\eta})/(s_{\xi\xi} - 2^{-1}s_{\Delta\xi\Delta\xi})$.

They also gave the asymptotic variance of $\hat{\beta}$. The accuracy of this method depends on the strength of the serial correlation. Smaller values of ρ_1 would result in larger MSE of $\hat{\beta}$.

4.2. Estimation When the Reliability of Each of the Independent Variables is Known

Consider the structural relationship model of (2.4.1) with uncorrelated errors (i.e., $\underset{\sim}{\Sigma}$ is diagonal) when $q = 1$. Let $\underset{\sim}{\xi}_i = (\xi_{i1},\ldots,\xi_{ip})'$ and $\underset{\sim}{\delta}_i = (\delta_{i1},\ldots,\delta_{ip})'$. Then the reliability of the j^{th} variable ξ_{ij}, $j = 1,\ldots, p$, is defined to be $1 - \lambda_j$, with $\lambda_j = \mathrm{Var}(\delta_{ij})/\mathrm{Var}(\xi_{ij})$. Suppose we know the λ_j. To illustrate the underlying principle, consider the particular case when p also equals one (model (1.1) of section 2.1), so that we know $\lambda_1 = \sigma_\delta^2/(\sigma^2 + \sigma_\delta^2)$. Putting $\sigma_\delta^2 = \lambda_1\sigma^2(1 - \lambda_1)^{-1}$ in (2.1.1) and solving the equations, we obtain a consistent estimate $\hat{\beta} = s_{\xi\eta}/(s_{\xi\xi} - \lambda s_{\xi\xi})$ of β. This can be viewed as an "adjusted" OLS estimate (OLS estimate $= s_{\xi\eta}/s_{\xi\xi}$) obtained by replacing $s_{\xi\xi}$ with a consistent estimate $s_{\xi\xi} - \lambda_1 s_{\xi\xi}$ of the variance σ^2 of the x_i. Returning to the general situation, we first fit a multiple regression of η on $\underset{\sim}{\xi}$. Then the vector of regression coefficients is estimated by $(n^{-1}\underset{\sim}{\psi}'\underset{\sim}{\psi})^{-1}(n^{-1}\underset{\sim}{\psi}'\underset{\sim}{\eta})$, where $\underset{\sim}{\eta} = (\eta_1,\ldots,\eta_n)'$ and $\underset{\sim}{\psi} = [\underset{\sim}{\xi}_1 - \bar{\underset{\sim}{\xi}}, \ldots, \underset{\sim}{\xi}_n - \bar{\underset{\sim}{\xi}}]'$. We thus replace $n^{-1}\underset{\sim}{\psi}'\underset{\sim}{\psi}$ by the consistent estimate $\underset{\sim}{H} = n^{-1}\underset{\sim}{\psi}'\underset{\sim}{\psi} - \underset{\sim}{D}\underset{\sim}{\Lambda}\underset{\sim}{D}$ of the true covariance matrix of the $\underset{\sim}{x}_i$ where $\underset{\sim}{D}^2$ is a diagonal matrix whose diagonal is that of $n^{-1}\underset{\sim}{\psi}'\underset{\sim}{\psi}$, and $\underset{\sim}{\Lambda}$ is also a diagonal matrix with the (j,j) element equal to λ_j. The matrix $\underset{\sim}{H}$ is then said to be a matrix corrected for attenuation (cf. Bock and Peterson, 1975). It can then be shown that $\underset{\sim}{H}^{-1}n^{-1}\underset{\sim}{\psi}'\underset{\sim}{\eta}$ is a consistent estimate of $\underset{\sim}{\beta}$. However, since $\underset{\sim}{H}$ might not be positive definite, some slight adjustments are required (Fuller and Hidiroglou, 1978). The method was discussed by Warren, White and Fuller (1974). Fuller and Hidiroglou (1978) investigated the asymptotic properties of the estimate and extended them to a general $\underset{\sim}{\Sigma}$.

4.3. Covariance Structure Analysis

In the analysis of covariance structure, Jöreskog (1970, 1971) considered n mutually independent r-dimensional random vectors $\underset{\sim}{z}_1, \ldots, \underset{\sim}{z}_n$ each having a multivariate normal distribution with the same covariance matrix

$$\underset{\sim}{V} = \underset{\sim}{B}(\underset{\sim}{\Lambda}\underset{\sim}{\phi}\underset{\sim}{\Lambda}' + \underset{\sim}{\psi}^2)\underset{\sim}{B}' + \underset{\sim}{\theta}^2,$$

and

$$E[\underset{\sim}{z}_1 \ \ldots \ \underset{\sim}{z}_n]' = \underset{\sim}{A}\underset{\sim}{Z}\underset{\sim}{P},$$

where $\underset{\sim}{\psi}$ and $\underset{\sim}{\theta}$ are diagonal matrices, $\underset{\sim}{\phi}$ is symmetric, and $\underset{\sim}{A}$ and $\underset{\sim}{P}$ are known $n \times q$ and $h \times r$ matrices, respectively, with rank $(\underset{\sim}{A}) = q \le n$ and rank $(\underset{\sim}{P}) = h \le r$. The elements of the matrices $\underset{\sim}{B}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\phi}$, $\underset{\sim}{\psi}$, $\underset{\sim}{\theta}$ and $\underset{\sim}{Z}$ are either known constants or unknown parameters which can be

(i) free parameters that are not constrained to be equal to any other parameters, or

(ii) constrained parameters that are unknown but equal to one or more other parameters.

He outlined a computational procedure for obtaining the MLE of the unknown parameters when they are identifiable.

The theory of analysis of covariance structure can be applied to the problem of estimating linear structural relationship (Jöreskog, 1970) which we now illustrate by the following example. Consider the structural relationship model of (2.4.1) when for each $(\underset{\sim}{x}_i, \underset{\sim}{y}_i)$, r independent repeated observations $(\underset{\sim}{\xi}_{ij}, \underset{\sim}{\eta}_{ij})$ are available. Then the model becomes

$$\begin{aligned} \underset{\sim}{\eta}_{ij} &= \underset{\sim}{\alpha} + \underset{\sim}{B}\underset{\sim}{x}_i + \underset{\sim}{\epsilon}_{ij}, \\ \underset{\sim}{\xi}_{ij} &= \underset{\sim}{x}_i + \underset{\sim}{\delta}_{ij}, \end{aligned} \tag{4.3.1}$$

where the $\underset{\sim}{x}_i$ are i.i.d. as $N(\underset{\sim}{0}, \underset{\sim}{\Sigma}_x)$ and are independent of the $(\underset{\sim}{\delta}_{ij}, \underset{\sim}{\epsilon}_{ij})$, which are i.i.d. as $N(\underset{\sim}{0}, \underset{\sim}{\Sigma})$, where

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\Sigma}_\delta & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\Sigma}_\epsilon \end{bmatrix}, \qquad \underset{\sim}{\Sigma}_\delta = \text{Cov}(\underset{\sim}{\delta}_{ij}), \qquad \underset{\sim}{\Sigma}_\epsilon = \text{Cov}(\underset{\sim}{\epsilon}_{ij}),$$

and $\underset{\sim}{\Sigma}_\delta$ and $\underset{\sim}{\Sigma}_\epsilon$ are diagonal. Since $\hat{\underset{\sim}{\mu}} = \sum \sum_r \underset{\sim}{\xi}_{ij}/nr$ and $\hat{\underset{\sim}{\alpha}} = \sum \sum_r \underset{\sim}{\eta}_{ij}/nr - \hat{\underset{\sim}{\beta}}\hat{\underset{\sim}{\mu}}$, where "^" denotes MLE, we assume $\underset{\sim}{\alpha} = \underset{\sim}{\mu} = \underset{\sim}{0}$ (cf. chapter 3, section 1). Now let $\underset{\sim}{z}_i' = (\underset{\sim}{\xi}_{i1}', \ldots, \underset{\sim}{\xi}_{ir}', \underset{\sim}{\eta}_{i1}', \ldots, \underset{\sim}{\eta}_{ir}')$. Then the $\underset{\sim}{z}_i$ are independent and

$$\underset{\sim}{z}_i = \underset{\sim}{B}\,\underset{\sim}{x}_i + (\underset{\sim}{\delta}_{i1}', \ldots, \underset{\sim}{\delta}_{ir}', \underset{\sim}{\epsilon}_{i1}', \ldots, \underset{\sim}{\epsilon}_{ir}')', \tag{4.3.2}$$

where $\underset{\sim}{B}' = [\underset{\sim}{I} \ldots \underset{\sim}{I}\ \underset{\sim}{B}' \ldots \underset{\sim}{B}']$ is $2r \times p$. Thus the $\underset{\sim}{z}_i$ have the same covariance matrix $\underset{\sim}{B}\,\underset{\sim}{\Sigma}_x\,\underset{\sim}{B}' + \underset{\sim}{D}$, where

$$\underset{\sim}{D} = \begin{bmatrix} \underset{\sim}{\Sigma}_\delta & & & & \underset{\sim}{0} \\ & \ddots & & & \\ & & \underset{\sim}{\Sigma}_\delta & & \\ & & & \underset{\sim}{\Sigma}_\epsilon & \\ & & & & \ddots \\ \underset{\sim}{0} & & & & \underset{\sim}{\Sigma}_\epsilon \end{bmatrix}$$

is 2rp × 2rp. It is now clear that by setting $\underset{\sim}{R} = \underset{\sim}{\beta}$, $\underset{\sim}{\Lambda} = \underset{\sim}{I}$, $\underset{\sim}{\Phi} = \underset{\sim}{\Sigma}_x$, $\underset{\sim}{\Psi} = \underset{\sim}{0}$ and $\underset{\sim}{\theta}^2 = \underset{\sim}{\Omega}$, the model (4.3.1) is expressed in the form of Jöreskog's model and hence his computational procedure is applicable here. Note that (2.4.2) looks like a factor analysis model which is also a particular case of Jöreskog's (1970) model.

4.4. The Use of Instrumental Variables and the Connections of Linear Functional Relationship with Simultaneous-Equation Models

In econometrics, instrumental variables are commonly used in the estimation of linear structural and functional relationships (also known as errors-in-variables estimation in econometrics). There are various statistical models involving the use of instrumental variables. In this section, the connection of one such model with simultaneous-equation models is illustrated. Goldberger (1972) and Zellner (1970, 1971) considered the linear functional relationship model of (2.4.1) when the unobservable constant $\underset{\sim}{x}_i$ can be expressed as $\underset{\sim}{x}_i = \underset{\sim}{\pi}_0 + \underset{\sim}{\Pi}\, \underset{\sim}{z}_i$, where the $\underset{\sim}{z}_i$ are non-stochastic and observable k-dimensional ($p \le k$) vectors which play the roles of instrumental variables. $\underset{\sim}{\Pi}_0$ is a p-component vector, and $\underset{\sim}{\Pi}$ is a $p \times k$ unknown matrix. For convenience, we assume $\underset{\sim}{\alpha} = \underset{\sim}{0}$ in (2.4.1) and $\underset{\sim}{\pi}_0 = \underset{\sim}{0}$. Then the model can be written in the form:

$$\begin{aligned} \underset{\sim}{\eta}_i &= \underset{\sim}{B}\, \underset{\sim}{\xi}_i + \underset{\sim}{\gamma}_i, \quad \underset{\sim}{\gamma}_i = \underset{\sim}{\varepsilon}_i - \underset{\sim}{B}\, \underset{\sim}{\delta}_i, \\ \underset{\sim}{\xi}_i &= \underset{\sim}{\Pi}\, \underset{\sim}{z}_i + \underset{\sim}{\delta}_i, \quad i = 1, \ldots, n. \end{aligned} \tag{4.4.1}$$

Each $\underset{\sim}{\gamma}_i$ has zero mean and is correlated with $\underset{\sim}{\xi}_i$ and $\underset{\sim}{\eta}_i$. This is the "structural" form of a system of simultaneous equations with the $\underset{\sim}{z}_i$ identified as "exogeneous" variables (in econometrics, exogeneous variables are variables with values determined outside the model); the ξ_i and η_i are "endogeneous" variables both correlated with the errors γ_i and δ_i, and the model is in "structural" form because relationships are expressed directly between the endogenous variables (note the different usage of "structural" here). The "reduced" form of the model (4.4.1) is

$$\eta_i = B \Pi z_i + \epsilon_i,$$
$$\xi_i = \Pi z_i + \delta_i, \quad i = 1,\dots n.$$

To estimate B, let $B' = [\beta_1 \dots \beta_q]$, $\Pi' = [\pi_1 \dots \pi_p]$, $Z = [z_{ij}] = [g_1 \dots g_k]$, $\Xi = [\xi_{ij}] = [f_1 \dots f_p]$, $N = [n_{ij}] = [n_1 \dots n_q]$, $\Gamma = [\gamma_{ij}] = [t_1 \dots t_q]$ and $\Delta = [\delta_{ij}] = [d_1 \dots d_p]$, where t_{ij} denotes the j^{th} component of the i^{th} vector t_i, and all the matrices are partitioned into columns. Then (4.4.1) can be rewritten as

$$[n_1 \dots n_q] = \Xi[\beta_1 \dots \beta_q] + [t_1 \dots t_q],$$
$$[f_1 \dots f_p] = Z[\pi_1 \dots \pi_p] + [d_1 \dots d_p].$$

Thus for each $\ell = 1,\dots, p$, $f_\ell = Z \pi_\ell + d_\ell$ is an ordinary multiple regression model and the OLS procedure can be applied to obtain an estimate $\hat{\pi}_\ell$ of π_ℓ. Now for each $m = 1,\dots q$, replace Ξ by $\hat{\Xi} = [\hat{f}_1 \dots \hat{f}_p]$, where the $\hat{f}_\ell = Z \hat{\pi}_\ell$ are fitted values, in $n_m = \Xi\beta_m + t_m$ and estimate β_m by the OLS procedure. The consistent estimate obtained in this way is known as the two stage least squares (SLS) estimate. The covariance matrix of (γ_i, δ_i) can also be estimated by the usual residual mean sum of squares and products. Another method which estimates $\underset{\sim}{B}$ and $\underset{\sim}{\Pi}$ simultaneously is the three SLS. To obtain the three SLS estimate, consider the linear model:

$$\begin{bmatrix} \underset{\sim}{Z}'\underset{\sim}{y}_1 \\ \vdots \\ \underset{\sim}{Z}'\underset{\sim}{y}_q \\ \underset{\sim}{Z}'\underset{\sim}{f}_1 \\ \vdots \\ \underset{\sim}{Z}'\underset{\sim}{f}_p \end{bmatrix} = \begin{bmatrix} \underset{\sim}{Z}'\underset{\sim}{X} & & & & & \\ & \ddots & & & \underset{\sim}{0} & \\ & & \underset{\sim}{Z}'\underset{\sim}{X} & & & \\ & & & \underset{\sim}{Z}'\underset{\sim}{Z} & & \\ & \underset{\sim}{0} & & & \ddots & \\ & & & & & \underset{\sim}{Z}'\underset{\sim}{Z} \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta}_1 \\ \vdots \\ \underset{\sim}{\beta}_q \\ \underset{\sim}{\pi}_1 \\ \vdots \\ \underset{\sim}{\pi}_p \end{bmatrix} + \begin{bmatrix} \underset{\sim}{Z}'\underset{\sim}{\epsilon}_1 \\ \vdots \\ \underset{\sim}{Z}'\underset{\sim}{\epsilon}_q \\ \underset{\sim}{Z}'\underset{\sim}{d}_1 \\ \vdots \\ \underset{\sim}{Z}'\underset{\sim}{d}_p \end{bmatrix} . \qquad (4.4.2)$$

Note that $\underset{\sim}{Z}'$ is first multiplied by each of the $\underset{\sim}{f}_\ell = \underset{\sim}{Z}\,\underset{\sim}{\pi}_\ell + \underset{\sim}{d}_\ell$ and $\underset{\sim}{y}_m = \underset{\sim}{X}\underset{\sim}{\beta}_m + \underset{\sim}{\epsilon}_m$ to construct (4.4.2). A two SLS is first carried out to obtain an estimate of the covariance matrix $\underset{\sim}{V}$ of $(\underset{\sim}{\epsilon}_i, \underset{\sim}{d}_i)$, which is then used in Aitken's generalized least squares estimates of $\underset{\sim}{\beta}_1, \ldots, \underset{\sim}{\beta}_q$ and $\underset{\sim}{\pi}_1, \ldots, \underset{\sim}{\pi}_p$ in (4.4.2) assuming $\underset{\sim}{V}$ is known (so that the covariance matrix of $\underset{\sim}{Z}'\underset{\sim}{\epsilon}_1, \ldots, \underset{\sim}{Z}'\underset{\sim}{\epsilon}_q, \underset{\sim}{Z}'\underset{\sim}{d}_1, \ldots, \underset{\sim}{Z}'\underset{\sim}{d}_p$ is known). Under a normality assumption on the $(\underset{\sim}{\epsilon}_i, \underset{\sim}{d}_i)$, another computationally more complicated method is to obtain the "full information" maximum likelihood (FIML) estimates of $\underset{\sim}{B}$ and $\underset{\sim}{\pi}$ by solving the likelihood equations of the model (4.4.1) directly. However, the asymptotic covariance matrix of the three SLS estimates is the same as that of the FIML estimates (Rothenberg and Leenders, 1964).

For greater details of estimation in model (4.4.1), see Goldberger (1972) and Zellner (1970, 1971). Carlson, Sobel and Watson (1966) also discussed the use of econometric methods in a biological example.

Anderson (1976) also showed how the estimation of a coefficient in one equation of a simultaneous system of stochastic equations is related to the estimate of the slope paramater β in the linear functional relationship model of (1.1) with uncorrelated errors. Based on this connection, he pointed out that in many applications in econometrics, it is more relevant to consider asymptotic properties as $S_x^2 = n^{-1} \sum (x_i - \bar{x}.)^2/n \to \infty$ while n is fixed. In this case, as $S_x^2 \to \infty$, the OLS estimate is also consistent and has the same limiting distribution as the MLE of β when λ is known. Comparisons of the two estimates have to be made based on the asymptotic expansions of their distributions.

4.5. Bayesian Approach

Lindley and El-Sayyad (1968) considered Bayesian estimation of the functional relationship model of (1.1) with uncorrelated errors. They assumed that the incidental parameters $x_1, \ldots, x_n$ have a common prior distribution $N(0,\tau^2)$, where τ^2 is unknown, and are independent, and an arbitrary prior distribution $\pi(\beta,\sigma_\delta^2,\sigma_\epsilon^2,\tau^2)$ (assuming $\alpha = 0$) was considered. After making the transformation $(\beta,\sigma_\delta^2,\sigma_\epsilon^2,\tau^2) \to (\beta,\theta_{11},\theta_{22},\theta_{12})$, where $\theta_{11} = \tau^2 + \sigma_\delta^2$, $\theta_{22} = \beta^2\tau^2 + \sigma_\epsilon^2$ and $\theta_{12} = \beta\tau^2$ (cf. (2.1.1)), they showed that for large samples, the marginal posterior distribution of $(\theta_{11},\theta_{22},\theta_{12})$ concentrates around the point $(s_{\xi\xi},s_{\eta\eta},s_{\xi\eta})$ (s denotes the central product moment of its subscripts). Thus as $n \to \infty$, with certainty we know the θ_{ij}. However, the posterior distribution of β does not concentrate around any value as $n \to \infty$ and its variance does not tend to zero. This means that whatever the size of sample, the true value of β is never known. This is a phenomenon inherited from the difficulty of unidentifiability as previously discussed. While in the maximum likelihood approach the likelihood equation solution $(s_{\eta\eta}/s_{\xi\xi})^{1/2}$ for β does converge (to $\sigma_\epsilon/\sigma_\delta$), but in general does not converge to the true parameter (Lindley, 1947). However in the Bayesian approach, we do learn something about β from the posterior distribution which incorporates both our prior knowledge (not only about β, but σ_δ^2, σ_ϵ^2 and τ also) and the information contained in the data collected, although the posterior distribution does not have zero dispersion. Of course, the choice of the prior distribution π is important. When prior knowledge about the parameters is available, the Bayesian approach is a good one.

Zellner (1971) also considered a Bayesian approach to the functional relationship model of (1.1) with uncorrelated errors, but with different assumptions on the prior distributions. In particular, he assumed that the prior distribution of $(\alpha,\beta,\sigma_\delta^2,\sigma_\epsilon^2,x_1,\ldots, x_n)$ is proportional to $1/(\sigma_\delta^2\sigma_\epsilon^2)$ and found that the posterior distribution of (α,β) has a bivariate student-t form with mean $(\hat\alpha,\hat\beta)$, the OLS estimate of (α,β).

For further discussion see Florens, Mouchart and Richard (1974).

5. SOME PROPOSED PROBLEMS

We conclude this chapter with some proposed problems.

(1) Consider a simple regression model

$$y = \alpha + \beta x + e, \tag{5.1}$$

where the presence of a normally distributed term e with zero mean indicates that the variables x and y are not exactly linear related and e is not interpreted as error of measurement (Malinvaud (1970 , p. 201) call e the error in the equation). If the "independent" variable x (which can be deterministic or stochastic) and the "dependent" variable y are observed with errors δ and ϵ, respectively, the model becomes

$$\begin{aligned} \eta_i &= y_i + \epsilon_i = \alpha + \beta x_i + e_i + \epsilon_i, \\ \xi_i &= x_i + \delta_i, \qquad i = 1, \ldots, n, \end{aligned} \tag{5.1}$$

where the $(e_i, \epsilon_i, \delta_i)$ are i.i.d. as $N(\underset{\sim}{0}, \underset{\sim}{V})$. The model (1.1) can be considered as the particular case when $e_i = 0$. Schneeweiß (1976) discussed estimation in a model similar to (5.1) but with more than one independent variable. However, since he considered $e_i + \epsilon_i$ as a whole, no new estimation problem different from that of (1.1) arises. The fact that methods of estimation in (5.1) are not always the same as those in (1.1) can be seen as follows. Consider the case when replication is possible, i.e., for each (x_i, y_i), we have repeated measurements (ξ_{ij}, η_{ij}), $j = 1, \ldots, r$. Given x_i, ξ_{ij}, $j = 1, \ldots, r$, are independent but η_{ij}, $j = 1, \ldots, r$, are correlated through e_i, in contrast to model (1.1) with repeated measurements, and the maximum likelihood solution should therefore be different from that of chapter 3. A complete solution does not seem to have been attained. Model (5.2) is appropriate, for instance, when x and y are two different kinds of measurement both describing the same phenomenon (say lung function in physiology) so that (5.1) may be more realistic than an exact linear relation. Because of fluctuation due to operations, we observe ξ and η as defined in (5.2).

(2) Consider (5.2) again. Now suppose the measurement ξ is hard and expensive to make, while the measurement η is easy and cheap to make. After the parameters in model (5.2) have been estimated, one may only want to make repeated measurements η_{ij} on y_i and try to predict the corresponding x_i based on the model (5.2) using the estimated parameters. Lawley and Maxwell (1973) and Chan (1977) considered a similar prediction problem in a factor analysis model and introduced the concept of unbiasedness. In analogy, we call a predictor ψ of x_i unbiased if $E(\psi \mid x_i) = x_i$. It would be interesting to find a minimum mean square error unbiased predictor of x_i and to find out its relationship with the corresponding predictor in factor analysis.

(3) When replication is possible in model (1.1) or (2.4.1), as previously indicated, one estimation procedure is to estimate $\underset{\sim}{\Sigma}$ first by the usual within group sum of squares and products and then use this estimate in the MLE of $\underset{\sim}{\beta}$ (or β) assuming $\underset{\sim}{\Sigma}$ is known. Another better procedure is to solve the likelihood equations directly but the computation is much more involved. Thus it would be interesting to look at the efficiency of the former procedure relative to the latter.

(4) Lindley and El-Sayyad (1968) proposed a Bayesian approach of estimating the functional relationship model of (1.1) with uncorrelated errors in which the x_i have common prior distribution $N(0,\tau^2)$ and are independent. They found that even for large

samples the posterior distribution does not concentrate around any value. A Bayesian approach using additional information such as instrumental variables, grouping with criterion independent of errors, and repeated observations would intuitively give better results and seems worth further investigation. Zellner's (1970, 1971) Bayesian discussion of model (4.4.1) with $p = q = 1$ can be considered as one such approach.

(5) In the presence of infinitely many incidental parameters, Wald (1948) gave necessary and sufficient conditions for the existence of uniformly consistent estimates of structural parameters. It is seen in chapter 6 that although the MLE of a structural paramater is not necessarily consistent, there are situations when a function of it is consistent. Under what further assumptions do the conditions given by Wald also imply the existence of a consistent estimate which is a function of the MLE?

REFERENCES

ANDERSON, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann. Math. Statist., 22, 327-351.

ANDERSON, T. W. (1976). Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics. J. Roy. Statist. Soc., Ser. B, 38, 1-36.

BARNETT, V. D. (1967). A note on linear structural relationships when both residual variances are known. Biometrika, 54, 670-672.

BARNETT, V. D. (1969). Simultaneous pairwise linear structural relationships. Biometrics, 25, 129-142.

BARNETT, V. D. (1970). Fitting straight lines—the linear functional relationship with replicated observations. Appl. Statist., 19, 135-144.

BIRCH, M. W. (1964). A note on maximum likelihood estimation of a linear structural relationship. J. Amer. Statist. Assoc., 59, 1175-1178.

BOCK, R. D., and PETERSON, A. C. (1975). A multivariate correction for attenuation. Biometrika, 62, 673-678.

CARLSON, E. D., SOBEL, D. and WATSON, G. S. (1966). Linear relationships between variables affected by errors. Biometrics, 22, 252-267.

CHAN, N. N. (1977). On an unbiased predictor in factor analysis. Biometrika, 64, 642-644.

*DAVIES, R. B. and HUTTON, B. (1975). The effect of errors in the independent variables in linear regression. Biometrika, 62, 383-391.

DOLBY, G. R. (1976). The ultrastructural relation: a synthesis of the functional and structural relations. Biometrika, 63, 39-50.

DOLBY, G. R. and LIPTON, S. (1972). Maximum likelihood estimation of a general non-linear functional relationship with replicated observations and correlated errors. Biometrika, 59, 121-129.

*DORFF, M. and GURLAND, J. (1961a). Estimation of the parameters of a linear functional relationship, J. Roy. Statist. Soc., Ser. B, 23, 160-170.

*DORFF, M. and GURLAND, J. (1961b). Small sample behaviour of slope estimators in a linear functional relation. Biometrics, 17, 283-298.

DRION, E. F. (1951). Estimation of the parameters of a straight line and of the variances of the variables, if they are both subject to error. Indag. Math., 13, 256-260.

*DURBIN, J. (1954). Errors in variables. Rev. Inst. Internat. Statist., 22, 23-32.

*FELDSTEIN, M. (1974). Errors in variables: a consistent estimator with smaller MSE in finite samples. J. Amer. Statist. Assoc., 69, 990-996.

FLORENS, J., MOUCHART, M., and RICHARD, J. F. (1974). Bayesian inference in error-in-variables models. J. Mult. Anal., 4, 419-452.

FULLER, W. and HIDIROGLOU, M. (1978). Regression estimation after correction for attenuation. J. Amer. Statist. Assoc., 73, 99-104.

GEARY, R. C. (1942). Inherent relations between random variables. Proc. Roy. Irish. Acad., Ser. A, 47, 63-76.

GEARY, R. C. (1949). Determination of linear relations between systematic parts of variables with errors of observation, the variances of which are unknown. Econometrica, 17, 30-59.

GLESER, L. J. and WATSON, G. S. (1973). Estimation of a linear transformation. Biometrika, 60, 525-534.

GOLDBERGER, A. S. (1972). Maximum likelihood estimation of regressions containing unobservable independent variables. Int. Econ. Rev., 13, 1-15.

JÖRESKOG, K. G. (1970). A general method for analysis of covariance structures. Biometrika, 57, 239-251.

JÖRESKOG, K. G. (1971). Statistical analysis of sets of congeneric tests. Psychometrika, 109-133.

KARNI, E. and WESSIMAN, I. (1974). A consistent estimate of the slope in a regression model with errors in the variables. J. Amer. Statist. Assoc., 69, 211-213.

KEIFER, J. and WOLFOWITZ, J. (1956). Consistence of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Statist., 27, 887-906.

KENDALL, M. G. and STUART, A. (1973). The Advanced Theory of Statistics, Vol. 2, 3rd edn. London: Griffin.

LAWLEY, D. N. and MAXWELL, A. E. (1973). Regression and factor analysis. Biometrika, 60, 331-338.

LINDLEY, D. V. (1947). Regression lines and the linear functional relationship. Suppl. J. Roy. Statist. Soc., 9, 218-244.

*LINDLEY, D. V. (1953). Estimation of a functional relationship. Biometrika, 40, 47-49.

LINDLEY, D. V. and EL-SAYYAD, G. M. (1968). The Bayesian estimation of a linear functional relationship. J. Roy. Statist. Soc., Ser. B, 30, 190-202.

MALINVAUD, E. (1970). Statistical Methods of Econometrics, 2nd ed. Amsterdam: North Holland.

MOBERG, L. and SUNDBERG, R. (1978). Maximum likelihood estimation of a linear functional relationship when one of the departure variacnes is known. Scand. J. Statist., 5, 61-64.

MORAN, P. A. P. (1971). Estimating structural and functional relationships. J. Mult. Anal., 1, 232-255.

*NEYMAN, J. (1951). Existence of consistent estimates of the directional parameters in a linear structural relation between two variables. Ann. Math. Statist., 22, 497-511.

NEYMAN, J. and SCOTT, E. (1948). Consistent estimates based on partially consistent observations. Econometrica, 16, 1-32.

*NEYMAN, J. and SCOTT, E. (1951). On certain methods of estimating the linear structural relationship. Ann. Math. Statist., 22, 352-361.

PATEFIELD, W. M. (1977). On the information matrix in the linear functional relationship problem. Appl. Statist., 26, 69-70.

REIERSØL, O. (1945). Confluence analysis by means of instrumental sets of variables. Ark. Mat. Astron. Fys., 32, 1-119.

REIERSØL, O. (1950). Identifiability of a linear relation between variables which are subject to error. Econometrica, 18, 375-389.

ROBINSON, P. M. (1977). The estimation of a multivariate linear relation. J. Mult. Anal., 7, 409-423.

ROTHENBERG, T. J. and LEENDERS, C. T. (1964). Efficient estimation of simultaneous equation systems. Econometrica, 32, 57-76.

SCHNEEWEIß, H. (1976). Consistent estimation of a regression with errors in the variables. Metrika, 23, 101-115.

SCOTT, E. L. (1950). Note on consistent estimates of the linear structural relation between two variables. Ann. Math. Statist., 21, 284-288.

SOLARI, M. E. (1959). The maximum likelihood solution of the problem of estimating a linear functional relationship. J. Roy. Statist. Soc., Ser. B, 31, 372-375.

SPRENT, P. (1966). A generalized least-squares approach to linear functional relationships. J. Roy. Statist. Soc., Ser. B, 28, 278-297.

SPRENT, P. (1969). Models in Regression and Related Topics. London: Methuen.

TAYLOR, J. (1973). A method of fitting several linear functional relations and of testing for differences between them. Appl. Statist., 22, 239-248.

*TUKEY, J. W. (1951). Components in regression. Biometrics, 7, 33-69.

VILLEGAS, C. (1961). Maximum likelihood estimation of a linear functional relationship. Ann. Math. Statist., 32, 1048-1062.

*WALD, A. (1940). The fitting of straight lines if both variables are subject to error. Ann. Math. Statist., 11, 284-300.

WALD, A. (1948). Estimation of a parameter when the number of unknown parameters increases indefinitely with the number of observations. Ann. Math. Statist., 19, 220-227.

WARREN, R., WHITE, J., and FULLER, W. (1974). An errors in variables analysis of managerial role performance. J. Amer. Statist. Assoc., 69, 886-893.

WOLFOWITZ, J. (1952). Consistent estimators of the parameters of a linear structural relation. Skand. Aktuarietidskr., 35, 132-151.

* References not mentioned in this chapter but from which original work in section 2 can be found.

CHAPTER 2

MAXIMUM LIKELIHOOD ESTIMATION OF A LINEAR STRUCTURAL RELATIONSHIP WHEN THE INTERCEPT IS KNOWN

1. INTRODUCTION

Consider a bivariate random variable (x,y) satisfying the linear relation $y = \alpha + \beta x$, β being unknown and to be estimated. Suppose x and y cannot be observed exactly, but instead we observe $\xi = x + \delta$ and $\eta = y + \epsilon$, where the errors δ and ϵ have zero means and unknown variances σ_δ^2 and σ_ϵ^2, respectively.

If α is unknown, x, δ and ϵ are independent and normally distributed, and x has unknown mean μ and variance σ^2, then β is not identifiable and cannot be estimated consistently from n independent observations (ξ_i, η_i), $i = 1, \ldots, n$ (cf. Kendall and Stuart, 1973, ch. 29; Moran, 1971). When σ_δ^2 (or σ_ϵ^2) or $\sigma_\epsilon^2/\sigma_\delta^2$ is known, β becomes identifiable and Maximum Likelihood (ML) estimates in these cases have been obtained (Lindley, 1947; Birch, 1964).

If α is known and μ is only known to be non-zero, then β also becomes identifiable and can be estimated consistently by $(\bar{\eta}. - \alpha)/\bar{\xi}.$, where $\bar{\eta}. = \sum_{i=1}^{n} \eta_i/n$, $\bar{\xi}. = \sum_{i=1}^{n} \xi_i/n$. Without loss of generality, let α be zero. The model then becomes

$$\xi = x + \delta,$$
$$\eta = \beta x + \epsilon.$$

The ML estimate $(\hat{\mu}, \hat{\beta}, \hat{\sigma}^2, \hat{\sigma}_\delta^2, \hat{\sigma}_\epsilon^2)$ of $(\mu, \beta, \sigma^2, \sigma_\delta^2, \sigma_\epsilon^2)$ is well-known when $\hat{\sigma}^2, \hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ are non-negative. However, when one of the variance estimates is negative no full solution to the estimation of $(\mu, \beta, \sigma^2, \sigma_\delta^2, \sigma_\epsilon^2)$ was available, as pointed out by Moran (1971, p. 252) and Zellner (1971, p. 130).

2. MAXIMUM LIKELIHOOD SOLUTION

The model is

$$\xi_i = x_i + \delta_i,$$
$$\eta_i = \beta x_i + \epsilon_i, \qquad i = 1, \ldots, n, \qquad (2.1)$$

where x_i, $i = 1, \ldots, n$, are i.i.d. as $N(\mu, \sigma^2)$, $\mu \neq 0$, δ_i, $i = 1, \ldots, n$, are i.i.d. as $N(0, \sigma_\delta^2)$, ϵ_i, $i = 1, \ldots, n$, are i.i.d. as $N(0, \sigma_\epsilon^2)$, and for each i, x_i, δ_i and ϵ_i are independent.

We further assume that (ξ_i, η_i) is non-singular. The model then becomes that each (ξ_i, η_i) has the bivariate normal distribution with mean $(\mu, \beta\mu)$ and positive definite covariance matrix

$$\underset{\sim}{V} = \begin{bmatrix} \sigma^2 + \sigma_\delta^2 & \beta\sigma^2 \\ \beta\sigma^2 & \beta^2\sigma^2 + \sigma_\epsilon^2 \end{bmatrix}.$$

The positive definitness of $\underset{\sim}{V}$ is equivalent to the condition that at most one of σ^2, σ_δ^2 and σ_ϵ^2 is zero and $\beta \neq 0$ if $\sigma_\epsilon^2 = 0$. The likelihood function L for (ξ_i, η_i), $i = 1, \ldots, n$, is thus the product of the bivariate normal probability functions.

Let $m_{\xi\eta} = \sum_{i=1}^n \xi_i\eta_i/n$, $m_{\xi\xi} = \sum_{i=1}^n \xi_i^2/n$, and $m_{\eta\eta} = \sum_{i=1}^n \eta_i^2/n$.

Then $(\bar{\xi}., \bar{\eta}., m_{\xi\xi} - \bar{\xi}.^2, m_{\eta\eta} - \bar{\eta}.^2, m_{\xi\eta} - \bar{\xi}.\bar{\eta}.)$ is the unique ML estimate of the transformed parameter $E(\xi) = \mu$, $E(\eta) = \beta\mu$,

$\mathrm{Var}(\xi) = \sigma^2 + \sigma_\delta^2$, $\mathrm{Var}(\eta) = \beta^2\sigma^2 + \sigma_\epsilon^2$, $\mathrm{Cov}(\xi,\eta) = \beta\sigma^2$) when L is considered as a function of the transformed parameter. The transformation is one-one if $\beta \neq 0$. Consider first the ML estimation with the restriction that $\beta \neq 0$. It will be shown later that the probability that the MLE of β being zero is zero. By lemma 3.2.3 of Anderson (1958), the solution $(\hat{\mu},\hat{\beta},\hat{\sigma}^2,\hat{\sigma}_\delta^2,\hat{\sigma}_\epsilon^2)$ for the equations

$$\bar{\xi}. = \mu,$$

$$\bar{\eta}. = \beta\mu,$$

$$m_{\xi\xi} - \bar{\xi}.^2 = \sigma^2 + \sigma_\delta^2,$$

$$m_{\eta\eta} - \bar{\eta}. = \beta^2\sigma^2 + \sigma_\epsilon^2,$$

$$m_{\xi\eta} - \bar{\xi}.\bar{\eta}. = \beta\sigma^2$$

maximizes L on $\Omega = \{(\mu,\beta,\sigma^2,\sigma_\delta^2,\sigma_\epsilon^2)$: $\mu \neq 0$, $\beta \neq 0$, $\underset{\sim}{V}$ is positive definite$\}$. Hence it is the ML estimate of $(\mu,\beta,\sigma^2,\sigma_\delta^2,\sigma_\epsilon^2)$ provided that $\hat{\sigma}^2 \geq 0$, $\hat{\sigma}_\delta^2 \geq 0$ and $\hat{\sigma}_\epsilon^2 \geq 0$. In this case, we have

$$\hat{\mu} = \bar{\xi}.,$$

$$\hat{\beta} = \bar{\eta}./\bar{\xi}.,$$

$$\hat{\sigma}^2 = m_{\xi\eta}(\bar{\xi}./\bar{\eta}.) - \bar{\xi}.^2,$$

$$\hat{\sigma}_\delta^2 = m_{\xi\xi} - m_{\xi\eta}(\bar{\xi}./\bar{\eta}.),$$

$$\hat{\sigma}_\epsilon^2 = m_{\eta\eta} - m_{\xi\eta}(\bar{\eta}./\bar{\xi}.) .$$

However, complication arises when one of the $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ is less than zero. Then the likelihood function L has to be maximized directly. L has only one local maximum on the open set

Ω at $(\hat{\mu},\hat{\beta},\hat{\sigma}^2,\hat{\sigma}_\delta^2,\hat{\sigma}_\epsilon^2)$. If one of the $\hat{\sigma}^2,\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ is negative, then when restricted to the set of all admissible values:
$\omega = \{(\mu,\beta,\sigma^2,\sigma_\delta^2,\sigma_\epsilon^2)$: $\mu \neq 0$, $\beta \neq 0$, $\sigma^2 \geq 0$, $\sigma_\delta^2 \geq 0$, $\sigma_\epsilon^2 \geq 0$, $\underset{\sim}{V}$ is positive definite$\}$, L cannot have a local maximum at a point such that all of σ^2, σ_δ^2 and σ_ϵ^2 are positive. Thus the problem reduces to maximizing L in each case when $\sigma^2 = 0$, $\sigma_\delta^2 = 0$ or $\sigma_\epsilon^2 = 0$ and take the one which gives the largest value of . as our ML solution.

Case 1: $\sigma^2 = 0$. After some algebraic manipulation one can express

$$L = (1/((2\pi)^n\sigma_\delta^n\sigma_\epsilon^n))\exp\{-(1/2)[\sum_{i=1}^n(\xi_i - \mu)^2/\sigma_\delta^2 + \sum_{i=1}^n(\eta_i - \beta\mu)^2/\sigma_\epsilon^2]\}.$$

Hence it is clear that L is maximized when

$$\begin{aligned}\mu &= \bar{\xi}.,\\ \beta &= \bar{\eta}./\bar{\xi}.,\\ \sigma_\delta^2 &= m_{\xi\xi} - \bar{\xi}.^2,\\ \sigma_\epsilon^2 &= m_{\eta\eta} - \bar{\eta}.^2,\end{aligned} \qquad (2.2)$$

and at this point

$$\ln L = -n\ln(2\pi) - (n/2)\ln[(m_{\xi\xi} - \bar{\xi}.^2)(m_{\eta\eta} - \bar{\eta}.^2)] - n. \qquad (2.3)$$

Case 2: $\sigma_\delta^2 = 0$. After some algebraic manipulation one can express

$$L = (1/((2\pi)^n\sigma^n\sigma_\epsilon^n))\exp\{-(1/2)[\sum_{i=1}^n(\xi_i - \mu)^2/\sigma^2 + \sum_{i=1}^n(\eta_i - \beta\xi_i)^2/\sigma_\epsilon^2]\}.$$

Hence L is maximized when

$$\mu = \bar{\xi}.,$$

$$\beta = m_{\xi\eta}/m_{\xi\xi},$$

$$\sigma^2 = m_{\xi\xi} - \bar{\xi}^2.,$$

$$\sigma_\epsilon^2 = m_{\eta\eta} - m_{\xi\eta}^2/m_{\xi\xi}, \qquad (2.4)$$

and at this point

$$\ln L = -n \ln (2\pi) - (n/2)\ln[(m_{\xi\xi} - \bar{\xi}^2.)(m_{\eta\eta} - m_{\xi\eta}^2/m_{\xi\xi})] - n. \qquad (2.5)$$

Case 3: $\sigma_\epsilon^2 = 0$. After some algebraic manipulation one can express

$$L = (1/((2\pi)^n\beta^n\sigma^n\sigma_\delta^n))\exp\{-(1/2)[\sum_{i=1}^{n}(\eta_i - \beta\mu)^2/\beta^2\sigma^2 + \sum_{i=1}^{n}(\beta\xi_i - \eta_i)^2/\beta^2\sigma_\delta^2]\}.$$

Maximization through solving likelihood equations yields

$$\mu = \bar{\eta}.(m_{\xi\eta}/m_{\eta\eta}),$$

$$\beta = m_{\eta\eta}/m_{\xi\eta},$$

$$\sigma^2 = (m_{\xi\eta}/m_{\eta\eta})^2(m_{\eta\eta} - \bar{\eta}^2.),$$

$$\sigma_\delta^2 = m_{\xi\xi} - m_{\xi\eta}^2/m_{\eta\eta}, \qquad (2.6)$$

and at this point

$$\ln L = -n \ln (2\pi) - (n/2) \ln [(m_{\eta\eta} - \bar{\eta}^2.)(m_{\xi\xi} - m_{\xi\eta}^2/m_{\eta\eta})] - n. \qquad (2.7)$$

Thus if one of $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ is negative, the ML estimate is given by either (2.2), (2.4) or (2.6) depending on which of (2.3), (2.5) and (2.7) gives the largest value. It is not difficult to see that if $\hat{\sigma}_\delta^2 < 0$, then (2.5) is greater than (2.7) and (2.3). Hence (2.4) gives the ML solution. Similarly if $\hat{\sigma}_\epsilon^2 < 0$, (2.6) gives the ML solution.

Now let us remove the restriction that $\beta \neq 0$. Suppose that L attains its maximum at a point $(\mu', \beta', \sigma'^2, \sigma_\delta'^2, \sigma_\epsilon'^2)$ with $\beta' = 0$. At this point $\underset{\sim}{V}$ becomes a diagonal matrix with elements $\sigma'^2 + \sigma_\delta'^2$ and $\sigma_\epsilon'^2$. However the point $(\mu', 0, \sigma'^2 + \sigma_\delta'^2, 0, \sigma_\epsilon'^2)$ also gives the same maximum. From case 2 we notice that this point is given by (2.4). Hence $m_{\xi\eta}/m_{\xi\xi} = 0$ whose occurrence has probability zero because ξ and η are continuous random variables.

<u>REMARK</u>: $\hat{\sigma}_\delta^2 < 0$ implies $m_{\xi\eta}/m_{\xi\xi} > \bar{\eta}./\bar{\xi}. = \hat{\beta}$ and $\hat{\sigma}_\epsilon^2 < 0$ implies $m_{\eta\eta}/m_{\xi\eta} < \bar{\eta}./\bar{\xi}.$. These situations correspond to the cases when the estimate $\bar{\eta}./\bar{\xi}.$ lies outside the bounds formed by the least square regression $m_{\xi\eta}/m_{\xi\xi}$ of η on ξ and the reciprocal of the least square regression $m_{\eta\eta}/m_{\xi\eta}$ (since $\alpha = 0$, we require the regression lines pass through the origin) of ξ on η. Moran (1971) discussed these situations intuitively and pointed out that in these cases, the sample variances and covariance of ξ and η should give some information on the slope parameter β. Estimates of β in (2.4) and (2.6) therefore give the necessary adjustment when $\bar{\eta}./\bar{\xi}.$ lies outside the bounds. If none of the true values of $\sigma^2, \sigma_\delta^2$ and σ_ϵ^2 is zero, with probability tending to one when sample size increases, $\hat{\beta}$ will give the ML estimate of β and is consistent.

Since $\hat{\beta} = \bar{\eta}./\bar{\xi}.$ and $E(\bar{\xi}.) = \mu$, if μ is near to zero, $\hat{\beta}$ will fluctuate wildly and have large mean square error in finite sample size although it is consistent. On the other hand, the estimate of β in (2.4) is asymptotically biased but has small variance. Thus it would be interesting to know if there is a reduction of mean square error by combining $\hat{\beta}$ linearly with the estimate of β in (2.4), the weights being determined based on sample information (Feldstein (1974) applied this technique to the case of the use of instrumental variable).

The asymptotic variance of $\hat{\beta}$ is given by (cf. Kendall and Stuart, 1973, equation 10.17):

$$(E(\bar{\eta}.)/E(\bar{\xi}.))^2(\mathrm{var}(\bar{\eta}.)/E^2(\bar{\eta}.) + \mathrm{var}(\bar{\xi}.)/E^2(\bar{\xi}.)$$

$$- 2\,\mathrm{cov}(\bar{\eta}.,\bar{\xi}.)/E(\bar{\xi}.)E(\bar{\eta}.)) = (\sigma_\epsilon^2 + \beta^2\sigma_\delta^2)/(n\mu^2) .$$

REFERENCES

ANDERSON, T. W. (1958). An Introduction to Multivariate Analysis. New York: Wiley.

BIRCH, M. W. (1964). A note on the maximum likelihood estimation of a linear structural relationship. J. Amer. Statist. Assoc., 59, 1175-1178.

FELDSTEIN, Martin (1974). Errors in variables: A consistent estimator with smaller MSE in finite samples. J. Amer. Statist. Assoc., 69, 990-996.

KENDALL, M. G. and STUART, A. (1973). The Advanced Theory of Statistics, Vols. I and II, 3rd Ed. London: Griffin.

LINDLEY, D. V. (1947). Regresion lines and the linear functional relationships. J. Roy. Statist. Soc. Suppl., 9, 218-244.

MORAN, P. A. P. (1971). Estimating structural and functional relationships. J. Multivar. Anal., 1, 232-255.

ZELLNER, A. (1971). An Introduction to Bayesian Inference in Econometrics. New York: Wiley.

CHAPTER 3

MAXIMUM LIKELIHOOD ESTIMATION OF A LINEAR STRUCTURAL RELATIONSHIP WITH REPLICATION

1. INTRODUCTION

Consider a bivariate random variable (x,y) satisfying the linear relationship $y = \alpha + \beta x$ with unknown α and β to be estimated. x and y cannot be observed directly. Instead we observe the values of $\xi = x + \delta$ and $\eta = y + \epsilon$ with errors δ and ϵ, respectively. The relationship is usually called structural relationship. For each (x_i, y_i), r repeated observations ξ_{ij} and η_{ij}, $j = 1,\ldots,r$, are obtained. The model considered here is

$$y_i = \alpha + \beta x_i, \qquad\qquad (1.1)$$
$$\xi_{ij} = x_i + \delta_{ij}, \quad \eta_{ij} = y_i + \epsilon_{ij}, \quad (i = 1,\ldots,n;\ j = 1,\ldots,r),$$

where x_i, δ_{ij} and ϵ_{ij} are mutually independent, $x_i \sim N(\mu,\sigma^2)$, $\delta_{ij} \sim N(0,\sigma_\delta^2)$, $\epsilon_{ij} \sim N(0,\sigma_\epsilon^2)$, and μ, σ^2, σ_δ^2 and σ_ϵ^2 are unknown. When $r = 1$, it is well known that β is unidentifiable. Here we assume that $r > 1$. (1.1) was first considered by Tukey (1951) and Madansky (1959) and estimates of β were obtained using variance components. As Madansky (1959) and Moran (1971) both pointed out, a maximum likelihood solution has not been obtained for this model. (A related problem, which is not considered here, is that x is assumed to be non-stochastic. The relationship is called functional relationship and was considered by Barnett (1970), Dorff and Gurland (1961a, 1961b), Housner and Brennan (1948) and Villegas (1961).) Dolby (1976) obtained the maximum likelihood solution for a general model of which the functional relationship with replication is a special case. The concept of replication in (1.1) is different from that of Dolby since the within group replicates in (1.1) are correlated through x_i while those in Dolby's model are independent. Anderson (1951) also considered a general estimation problem of which (1.1) is not a special case. In this chapter, it is found that the maximum likelihood estimate of β is a root of a fourth degree polynomial and the maximum likelihood estimates of μ, α, σ^2, σ_δ^2 and σ_ϵ^2 can be obtained subsequently (sections 2 and 3). The information matrix for the six parameters is obtained. Simplified formulas for inverting this matrix and for the asymptotic variance of the maximum likelihood estimate of β are derived (section 4). It is also shown that as the number of replicates increases, the polynomial mentioned above has a root which converges in probability to β (section 5). A numerical example is given to illustrate the computation of the estimates and their asymptotic variances (section 6).

2. THE LIKELIHOOD FUNCTION

Let $\underset{\sim}{\xi}_i = (\xi_{i1}, \ldots, \xi_{ir})'$, $\underset{\sim}{\eta}_i = (\eta_{i1}, \ldots, \eta_{ir})'$, and $\underset{\sim}{z}_i' = (\underset{\sim}{\xi}_i', \underset{\sim}{\eta}_i')$, $i = 1, \ldots, n$, then $\underset{\sim}{z}_i \sim N(\underset{\sim}{m}, \underset{\sim}{V})$, where

$$\underset{\sim}{m}' = (\mu \underset{\sim}{1}_{r1}', \ (\alpha + \beta\mu)\underset{\sim}{1}_{r1}'), \qquad \underset{\sim}{V} = \begin{bmatrix} \sigma^2 \underset{\sim}{1}_{rr} + \sigma_\delta^2 \underset{\sim}{I}_r & \beta\sigma^2 \underset{\sim}{1}_{rr} \\ \beta\sigma^2 \underset{\sim}{1}_{rr} & \beta^2\sigma^2 \underset{\sim}{1}_{rr} + \sigma_\epsilon^2 \underset{\sim}{I}_r \end{bmatrix}, \quad (2.1)$$

$\underset{\sim}{1}_{rs}$ denotes the $r \times s$ matrix with all entries 1, $\underset{\sim}{I}_r$ the $r \times r$ identity matrix. The log likelihood is

$$\ln L = \text{constant} - \tfrac{1}{2} n \ln |\underset{\sim}{V}| - \tfrac{1}{2} \sum_{i=1}^{n} \underset{\sim}{d}_i' \underset{\sim}{V}^{-1} \underset{\sim}{d}_i ,$$

where $\underset{\sim}{d}_i = \underset{\sim}{z}_i - \underset{\sim}{m}$. Let $\underset{\sim}{b}' = (\underset{\sim}{1}_{r1}', \beta \underset{\sim}{1}_{r1}')$,

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \sigma_\delta^2 \underset{\sim}{I}_r & 0 \\ 0 & \sigma_\epsilon^2 \underset{\sim}{I}_r \end{bmatrix},$$

then $\underset{\sim}{V} = \underset{\sim}{\Sigma} + \sigma^2 \underset{\sim}{b}\underset{\sim}{b}'$. By the Binomial Inverse Theorem,

$$\begin{aligned} \underset{\sim}{V}^{-1} &= \underset{\sim}{\Sigma}^{-1} - \underset{\sim}{\Sigma}^{-1} \underset{\sim}{b} \left(\sigma^{-2} + \underset{\sim}{b}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{b}\right)^{-1} \underset{\sim}{b}' \underset{\sim}{\Sigma}^{-1} \\ &= \begin{bmatrix} \sigma_\delta^{-2} \underset{\sim}{I}_r & 0 \\ 0 & \sigma_\epsilon^{-2} \underset{\sim}{I}_r \end{bmatrix} - a^{-1} \begin{bmatrix} \sigma_\delta^{-4} \underset{\sim}{1}_{rr} & \beta \sigma_\delta^{-2} \sigma_\epsilon^{-2} \underset{\sim}{1}_{rr} \\ \beta \sigma_\delta^{-2} \sigma_\epsilon^{-2} \underset{\sim}{1}_{rr} & \beta^2 \sigma_\epsilon^{-4} \underset{\sim}{1}_{rr} \end{bmatrix}, \end{aligned} \tag{2.2}$$

where $a = \underset{\sim}{b}' \underset{\sim}{\Sigma}^{-1} \underset{\sim}{b} + \sigma^{-2} = r\sigma_\delta^{-2} + r\beta^2 \sigma_\epsilon^{-2} + \sigma^{-2}$.

<u>LEMMA 2.1</u>

$$\begin{aligned} \ln L = \text{constant} &- \tfrac{1}{2} n \ln \sigma^2 - \tfrac{1}{2} rn \ln \sigma_\delta^2 - \tfrac{1}{2} rn \ln \sigma_\epsilon^2 \\ &- \tfrac{1}{2} n \ln a - \tfrac{1}{2} \textstyle\sum_n (h_i - c_i^2 a^{-1}), \end{aligned} \tag{2.3}$$

where

$$h_i = \sigma_\delta^{-2} \textstyle\sum_r (\xi_{ij} - \mu)^2 + \sigma_\epsilon^{-2} \sum_r (\eta_{ij} - \alpha - \beta\mu)^2,$$

$$c_i = \sigma_\delta^{-2} \textstyle\sum_r (\xi_{ij} - \mu) + \beta \sigma_\epsilon^{-2} \sum_r (\eta_{ij} - \alpha - \beta\mu),$$

$\sum_n$ and $\sum_r$ denote $\sum_{i=1}^{n}$ and $\sum_{j=1}^{r}$, respectively.

<u>Proof</u>: See Appendix A.

3. THE MAXIMUM LIKELIHOOD SOLUTION

<u>LEMMA 3.1</u>. The maximum likelihood estimates $\hat{\alpha}$ and $\hat{\mu}$ of α and μ when $\beta, \sigma^2, \sigma_\delta^2$ and σ_ϵ^2 are fixed satisfy

$$\hat{\mu} = \bar{\xi}_{..} ,$$
$$\hat{\alpha} + \beta\hat{\mu} = \bar{\eta}_{..} ,$$

where $\bar{\xi}_{..} = \sum_n \sum_r \xi_{ij}/nr$, $\bar{\eta}_{..} = \sum_n \sum_r \eta_{ij}/nr$. Thus to maximize ln L, it suffices to maximize (2.3) with μ, $\alpha + \beta\mu$ replaced by $\bar{\xi}_{..}$, $\bar{\eta}_{..}$ (cf. Richards, 1961) .

<u>Proof</u>: See Appendix B.

Let L_1 denote the L of (2.3) when μ and $\alpha + \beta\mu$ are replaced by $\bar{\xi}_{..}$ and $\bar{\eta}_{..}$. Because of lemma 3.1, from now on for simplicity in the process of finding the maximum likelihood estimates of $\beta, \sigma^2, \sigma_\delta^2$ and σ_ϵ^2, we write ξ_{ij} for $\xi_{ij} - \bar{\xi}_{..}$ and η_{ij} for $\eta_{ij} - \bar{\eta}_{..}$

Let

$$\bar{\xi}_{i\cdot} = \sum_r \xi_{ij}/r , \qquad \bar{\eta}_{i\cdot} = \sum_r \eta_{ij}/r ,$$

$$t_{\xi\xi} = \sum_n \sum_r \xi_{ij}^2/nr , \qquad t_{\eta\eta} = \sum_n \sum_r \eta_{ij}^2/nr,$$

$$w_{\xi\xi} = \sum_n \sum_r (\xi_{ij} - \bar{\xi}_{i\cdot})^2/rn , \qquad w_{\eta\eta} = \sum_n \sum_r (\eta_{ij} - \bar{\eta}_{i\cdot})^2/rn,$$

$$s_{\xi\xi} = \sum_n \bar{\xi}_{i\cdot}^2/n , \quad s_{\eta\eta} = \sum_n \bar{\eta}_{i\cdot}^2/n , \quad s_{\xi\eta} = \sum_n \bar{\xi}_{i\cdot}\bar{\eta}_{i\cdot}/n .$$

By differentiating $\ln L_1$ with respect to the parameters $\beta, \sigma^2, \sigma_\delta^2$ and σ_ϵ^2 and equating to zero, we get the following likelihood equations:

$$-n\beta + \textstyle\sum_n c_i \bar{\eta}_{i\cdot} - \sum_n c_i^2 \beta / a = 0\,, \qquad (3.1)$$

$$-n + n/(a\sigma^2) + \textstyle\sum_n c_i^2/(a^2\sigma^2) = 0\,, \qquad (3.2)$$

$$-n + n/(a\sigma_\delta^2) + nt_{\xi\xi}/\sigma_\delta^2 - 2\textstyle\sum_n c_i \bar{\xi}_{i\cdot}/(a\sigma_\delta^2) + \sum_n c_i^2/(a^2\sigma_\delta^2) = 0\,, \qquad (3.3)$$

$$-n + n\beta^2/(a\sigma_\epsilon^2) + nt_{\eta\eta}/\sigma_\epsilon^2 - 2\beta\textstyle\sum_n c_i \bar{\eta}_{i\cdot}/(a\sigma_\epsilon^2) + \beta^2\sum_n c_i^2/(a^2\sigma_\epsilon^2) = 0\,. \qquad (3.4)$$

After the algebraic manipulations in Appendix C, we have

$$\beta^2 s_{\xi\eta} + \beta(\lambda s_{\xi\xi} - s_{\eta\eta}) - \lambda s_{\xi\eta} = 0\,, \qquad (3.5)$$

$$\sigma_\epsilon^2 + \beta^2\sigma^2 = t_{\eta\eta}\,, \qquad (3.6)$$

$$\sigma_\delta^2 + \sigma^2 = t_{\xi\xi}\,, \qquad (3.7)$$

$$(2r-1)\sigma_\epsilon^2 = r(t_{\xi\xi}\lambda + t_{\eta\eta}) - r(\beta s_{\xi\eta} + \lambda s_{\xi\xi})\,, \qquad (3.8)$$

where $\lambda = \sigma_\epsilon^2/\sigma_\delta^2$. (3.5) is the familiar equation in the linear structural relationship model when $\sigma_\epsilon^2/\sigma_\delta^2$ is known. It is also shown in Appendix C that

$$p_{nr}(\beta) = \frac{1}{r}(k_0\beta^4 + k_1\beta^3 + k_2\beta^2 + k_3\beta + k_4) = 0\,, \qquad (3.9)$$

where

$$k_0 = (r-1)s_{\xi\xi}s_{\xi\eta}t_{\xi\xi},$$

$$k_1 = rs^2_{\xi\xi}w_{\eta\eta} - (r-1)s^2_{\xi\eta}t_{\xi\xi} - (r-1)s_{\xi\xi}s_{\eta\eta}t_{\xi\xi} - rs^2_{\xi\eta}w_{\xi\xi},$$

$$k_2 = (3r-1)(s_{\xi\eta}s_{\eta\eta}w_{\xi\xi} - s_{\xi\eta}s_{\xi\xi}w_{\eta\eta}),$$

$$k_3 = rs^2_{\xi\eta}w_{\eta\eta} + (r-1)s^2_{\xi\eta}t_{\eta\eta} + (r-1)s_{\xi\xi}s_{\eta\eta}t_{\eta\eta} - rs^2_{\eta\eta}w_{\xi\xi},$$

$$k_4 = -(r-1)s_{\xi\eta}s_{\eta\eta}t_{\eta\eta}.$$

Let "^" signify a maximum likelihood estimate. We therefore see that $\hat{\beta}$ can be obtained by solving (3.9) and then picking up the root which gives the largest value of $\ln L_1$. Thus we have proved the following theorem.

THEOREM 3.2. The maximum likelihood estimate $\hat{\beta}$ of β is a solution of the equation (3.9) if a real solution exists. Then we obtain $\hat{\sigma}^2_\epsilon$ from (3.8) through the $\hat{\lambda}$ in (3.5), $\hat{\sigma}^2_\delta = \hat{\sigma}^2_\epsilon/\hat{\lambda}$, $\hat{\sigma}^2 = t_{\xi\xi} - \hat{\sigma}^2_\delta$ from (2.7), and $\hat{\alpha}$ and $\hat{\mu}$ by lemma 3.1.

4. ACCURACY OF THE MAXIMUM LIKELIHOOD ESTIMATES

To derive Fisher's information matrix, we use the following formula proved by Dolby (1976):

$$-E\left(\frac{\partial^2 \ln L}{\partial\psi\partial\phi}\right) = n\{\tfrac{1}{2}\mathrm{tr}\ (\underset{\sim}{V}^{-1}\underset{\sim}{V}_\psi\underset{\sim}{V}^{-1}\underset{\sim}{V}_\phi) + \underset{\sim}{d}'_\psi\underset{\sim}{V}^{-1}\underset{\sim}{d}_\phi\}, \tag{4.1}$$

where $\psi,\phi = \alpha,\mu,\beta,\sigma^2,\sigma_\delta^2,\sigma_\epsilon^2$, $\underset{\sim}{V}_\psi = \partial\underset{\sim}{V}/\partial\psi$, $\underset{\sim}{d}_\psi = \partial\underset{\sim}{d}/\partial\psi$, with similar meanings for $\underset{\sim}{V}_\phi$, $\underset{\sim}{d}_\phi$ and $\underset{\sim}{d} = \sum_n \underset{\sim}{d}_i/n$.

First restrict ψ and ϕ to the sequence $\beta,\sigma^2,\sigma_\delta^2,\sigma_\epsilon^2$ and consider the 4×4 symmetric matrix $\underset{\sim}{P}$ whose elements are

$$\frac{1}{2}\ \mathrm{tr}\ (\underset{\sim}{V}^{-1}\underset{\sim}{V}_\psi\underset{\sim}{V}^{-1}\underset{\sim}{V}_\phi) \quad .$$

To simplify notations, only the upper triangular elements of P are given here. By using (2.1) and (2.2), we find that

$$\underset{\sim}{P} = 1/(2\sigma_\delta^2\sigma_\epsilon^2 a^2)\begin{bmatrix} 2r^2(\sigma^2 a+2\beta^2/\lambda) & 2r^2\beta(1+\beta^2/\lambda)/\sigma^2 & -2r^2\beta\sigma_\delta^{-2} & (2r\beta/\lambda)(a-r\beta^2/\sigma_\epsilon^2) \\ & \sigma_\delta^2\sigma_\epsilon^2\{(a-1/\sigma^2)/\sigma^2\}^2 & r\lambda/\sigma^4 & (r/\lambda)(\beta/\sigma^2)^2 \\ \cdot & & r\lambda(a^2-2a/\sigma_\delta^2+r/\sigma_\delta^4) & r^2\beta^2/(\sigma_\delta^2\sigma_\epsilon^2) \\ & & & (r/\lambda)(a^2-2a\beta^2/\sigma_\epsilon^2+r\beta^4/\sigma_\epsilon^4) \end{bmatrix}$$

$$= \begin{bmatrix} p_{\beta\beta} & \underset{\sim}{t}' \\ \underset{\sim}{t} & \underset{\sim}{T} \end{bmatrix} ,$$

where $p_{\beta\beta}$ is the $(1,1)^{th}$ element. By (2.6.7) of Press (1972) we have

$$p_{\beta\beta} - \underset{\sim}{t}'\underset{\sim}{T}^{-1}\underset{\sim}{t} = |\underset{\sim}{P}|/|\underset{\sim}{T}| \quad . \qquad (4.2)$$

Similarly, with $\psi,\phi = \beta,\alpha,\mu$ we can consider the 3×3 matrix $\underset{\sim}{D} = [\underset{\sim}{d}_\psi'\underset{\sim}{V}^{-1}\underset{\sim}{d}_\phi]$. Direct algebraic manipulation shows that

$$\mathbf{D} = \begin{bmatrix} \mu^2 v & \mu v & \mu g \\ \mu v & v & g \\ \mu g & g & (a-\sigma^{-2})/(a\sigma^2) \end{bmatrix} = \begin{bmatrix} \mu^2 v & \mathbf{g}' \\ \mathbf{g} & \mathbf{Q} \end{bmatrix},$$

where $v = r/\sigma_\epsilon^2 - r^2\beta/(a\sigma_\epsilon^4)$, $g = r\beta/(\sigma_\epsilon^2\sigma^2 a)$. The first column of $\mathbf{D}$ can be obtained from multiplying the second column by the scalar μ and hence $\mathbf{D}$ is singular, i.e., $|\mathbf{D}| = 0$. Thus again by (2.6.7) of Press (1972) we have

$$\mu^2 v - \mathbf{g}'\mathbf{Q}^{-1}\mathbf{g} = |\mathbf{D}|/|\mathbf{Q}| = 0. \qquad (4.3)$$

If we sequence the parameters as $\beta,\alpha,\mu,\sigma^2,\sigma_\delta^2,\sigma_\epsilon^2$, by (4.1) Fisher's information matrix can now be expressed as

$$\mathbf{F} = n\begin{bmatrix} P_{\beta\beta} + \mu^2 v & \mathbf{g}' & \mathbf{t}' \\ \mathbf{g} & \mathbf{Q} & \mathbf{O}' \\ \mathbf{t} & \mathbf{O} & \mathbf{T} \end{bmatrix} = n\begin{bmatrix} \mathbf{K} & \mathbf{M}' \\ \mathbf{M} & \mathbf{T} \end{bmatrix}, \qquad (4.4)$$

where $\mathbf{K}$ is a 3×3 matrix. $\mathbf{F}^{-1}$ gives the asymptotic covariance matrix of the maximum likelihood estimates of $\beta,\alpha,\mu,\sigma^2,\sigma_\delta^2$ and σ_ϵ^2. It can be obtained from first calculating $\mathbf{F}$ using (4.4) and then inverting the resulting matrix. However some simplification is possible and if one is only interested in the asymptotic variance $Av(\hat{\beta})$ of $\hat{\beta}$, no matrix inversion is required. To see this, let

$$n\mathbf{F}^{-1} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{21}' \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{bmatrix}, \quad \mathbf{K}^{-1} = \begin{bmatrix} \mathbf{k}_{11} & \mathbf{k}_{21}' \\ \mathbf{k}_{21} & \mathbf{k}_{22} \end{bmatrix}.$$

where $\mathbf{F}_{22}$ and $\mathbf{k}_{22}$ are 3×3 and 2×2 matrices respectively, and $\mathbf{F}_{22}/n$ is the asymptotic covariance matrix of $\sigma^2, \sigma_\delta^2$ and σ_ϵ^2. Then by (4.3) and (2.6.3) to (2.6.5) of Press (1972), we have

$$k_{11}^{-1} = p_{\beta\beta} + \mu^2 v - \mathbf{g}'\mathbf{Q}^{-1}\mathbf{g} = p_{\beta\beta},$$

$$\mathbf{F}_{11} = (\mathbf{K} - \mathbf{M}'\mathbf{T}^{-1}\mathbf{M})^{-1}$$

$$= \begin{bmatrix} p_{\beta\beta} + \mu^2 v - \mathbf{t}'\mathbf{T}^{-1}\mathbf{t} & \mathbf{g}' \\ \mathbf{g} & \mathbf{Q} \end{bmatrix}^{-1}$$

$$\mathbf{F}_{22} = (\mathbf{T} - \mathbf{M}\mathbf{K}^{-1}\mathbf{M}')^{-1}$$

$$= (\mathbf{T} - k_{11}\mathbf{t}\mathbf{t}')^{-1}$$

$$\mathbf{F}_{21} = -\mathbf{T}^{-1}\mathbf{M}\,\mathbf{F}_{11}$$

$$= [-\mathbf{T}^{-1}\mathbf{t}, \mathbf{O}]\,\mathbf{F}_{11}.$$

The (1,1) element of F_{11} when divided by n gives the asymptotic variance $AV(\hat{\beta})$, and by (4.2), (4.3) and (2.6.3) of Press (1972) we have

$$AV(\hat{\beta}) = 1/\{n(p_{\beta\beta} + \mu^2 v - \mathbf{t}'\mathbf{T}^{-1}\mathbf{t} - \mathbf{g}'\mathbf{Q}^{-1}\mathbf{g})\}$$

$$= |\mathbf{T}|/(n|\mathbf{P}|).$$

Thus no matrix inversion is required. Now to determine the whole 6×6 matrix $\mathbf{F}^{-1}$, it is only necessary to invert four 3×3 matrices, namely $\mathbf{K}$, $\mathbf{T} - k_{11}\mathbf{t}\mathbf{t}'$, $\mathbf{T}$ and

$$\begin{bmatrix} \mu^2 v + |\underset{\sim}{P}|/|\underset{\sim}{T}| & \underset{\sim}{g}' \\ \underset{\sim}{g} & \underset{\sim}{Q} \end{bmatrix} .$$

5. CONVERGENCE IN PROBABILITY AS THE NUMBER OF REPLICATES INCREASES

Now we proceed to show that the polynomial $p_{nr}(b)$ in (3.9) has a root which converges in probability to β as the number of replicates r tends to infinity.

THEOREM 5.1. Given $n, \delta, \epsilon > 0$, there exists an $r_0 > 0$ independent of $\underset{\sim}{x}$ such that for $r > r_0$,

$$\Pr\{p_{nr}(b) \text{ has a root in } (\beta - \epsilon, \beta + \epsilon)\} > 1 - \delta.$$

Proof: Given n and $\delta > 0$, there exists a closed bounded set $\bar{B} \subset R^n$ such that $\Pr\{\underset{\sim}{x} = (x_1, \ldots, x_n)' \in \bar{B}\} > 1 - \delta/2$. Given fixed $n, \underset{\sim}{x} \in \bar{B}$ and b, since $p_{nr}(b, \underset{\sim}{x})$ (where $p_{nr}(b, \underset{\sim}{x})$ denotes the $p_{nr}(b)$ when $\underset{\sim}{x}$ is fixed) is a polynomial in sample moments of ξ and η, as $r \to \infty$ it is asymptotically normal with mean $f_n(b, \underset{\sim}{x})$ and variance $V_n(\underset{\sim}{x})$ (cf. Cramér (1946, §28.4)). So using Tchebychev's inequality, it can be proved that given $\epsilon > 0$,

$$\Pr(|p_{nr}(b, \underset{\sim}{x}) - f_n(b, \underset{\sim}{x})| \leq \epsilon | \underset{\sim}{x}) \geq 1 - V_n(\underset{\sim}{x})/\epsilon^2 \geq 1 - M_n/\epsilon^2 \qquad (5.1)$$

for every $\underset{\sim}{x} \in \bar{B}$, where M_n is independent of $\underset{\sim}{x}$ and is of order $1/r$ since $V_n(\underset{\sim}{x})$ is of order $1/r$ and is a continuous function of $\underset{\sim}{x}$ on $\bar{B}$. It can be derived that $f_n(b, \underset{\sim}{x})$ is a polynomial in b involving β $\bar{\ }$, σ_ϵ^2 and σ_δ^2, and at $b = \beta$, f_n and its derivatives f_n' and f_n'' with respect to b are zero and $f_n''' > 0$. Expand $p_{nr}(b, \underset{\sim}{x})$ and $f_n(b, \underset{\sim}{x})$ at β to the third order and consider the change of sign in a small neighbourhood of β. It follows from (5.1) that there exists r_0

independent of x such that for every $\underset{\sim}{x} \in \overline{B}$ we can find $b_1(\underset{\sim}{x})$, $b_2(\underset{\sim}{x})$ with $\beta - \epsilon < b_1(\underset{\sim}{x}) < \beta < b_2(\underset{\sim}{x}) < \beta + \epsilon$ and for $r > r_0$

$$\Pr(p_{nr}(b_1(\underset{\sim}{x}),\underset{\sim}{x}) \cdot p_{nr}(b_2(\underset{\sim}{x}),\underset{\sim}{x}) < 0 | \underset{\sim}{x}) > 1 - \delta/2.$$

Then

$$\Pr\{p_{nr}(b) \text{ has a root in } (\beta - \epsilon, \beta + \epsilon)\}$$

$$\geq \Pr\{p_{nr}(b) \text{ has a root in } (\beta - \epsilon, \beta + \epsilon) \text{ and } \underset{\sim}{x} \in \overline{B}\}$$

$$= \int_{\overline{\overline{B}}} \Pr\{p_{nr}(b) \text{ has a root in } (\beta - \epsilon, \beta + \epsilon) \mid \underset{\sim}{x}\} f_x(\underset{\sim}{x})\, d\underset{\sim}{x}$$

$$\geq \int_{\overline{\overline{B}}} (1 - \delta/2) f_x(\underset{\sim}{x})\, dx > 1 - \delta.$$

6. A NUMERICAL EXAMPLE

Consider the simulated data in Table 1 with $n = 12$ and $r = 3$. From the data, we find

$$\bar{\xi}_{..} = -0.417, \quad \bar{\eta}_{..} = 0.549,$$

$$t_{\xi\xi} = 18.002, \quad t_{\eta\eta} = 38.815,$$

$$w_{\xi\xi} = 0.515, \quad w_{\eta\eta} = 0.766,$$

$$s_{\xi\xi} = 17.487, \quad s_{\eta\eta} = 38.049, \quad s_{\xi\eta} = 25.453$$

and the polynomial $p_{nr}(\beta)$ is

$$-75180.9 + 101194.2\beta + 1265.4\beta^2 - 47580\beta^3 - 16025.2\beta^4.$$

-1.458 and 1.479 are the two real roots of this polynomial. The latter together with the values of the other estimates

$$\hat{\alpha} = 1.166, \quad \hat{\mu} = -0.417, \quad \hat{\sigma}^2 = 17.211,$$

$$\hat{\sigma}_\delta^2 = 0.791, \quad \sigma_\epsilon^2 = 1.152,$$

obtained through theorem 3.2 give the larger value of ln L. Hence they are the maximum likelihood estimates of the parameters. A glance at the data also tells us the positive root should be taken.

The data was actually simulated from model (1.1) with $\alpha = 1$, $\beta = 1.5$, $\mu = 0$, $\sigma^2 = 10$ and $\sigma_\delta^2 = \sigma_\epsilon^2 = 1$. Using these true parameter values in (4.4), we compute Fisher's information matrix F and by inverting F, we find that the asymptotic standard errors of $\hat{\alpha}$, $\hat{\beta}$, $\hat{\mu}$, $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ are 0.228, 0.096, 0.923, 4.217, 0.260 and 0.283, respectively. The large variances of $\hat{\mu}$ and $\hat{\sigma}^2$ are not surprising; even if $\mu = 0$, $\sigma^2 = 10$ are estimatd from 12 independent observations from $N(\mu,\sigma^2)$, the standard errors of sample mean and sample variance are 0.910 and 3.910 respectively. In fact the sample variance based on the x values in Table 1 is 17.168.

In general, the true parameters are unknown and estimated values have to be used to estimate the asymptotic variances of the maximum likelihood estimates.

TABLE 1

Data for Model (1.1), n = 12, r = 3

	i = 1		i = 2		i = 3		i = 4	
unobserved x_i, y_i	-1.180,	-0.770	3.814,	6.721	-4.993,	-6.489	-7.274,	-9.910
observed ξ_{ij}, η_{ij} j=1	-1.879,	-0.138	3.494,	7.517	-4.915,	-6.674	-8.842,	-9.734
j=2	-1.869,	-1.198	3.594,	8.136	-4.442,	-6.753	-8.163,	-8.378
j=3	-0.603,	-1.571	5.429,	4.702	-3.989,	-6.992	-7.977,	-9.684
	i = 5		**i = 6**		**i = 7**		**i = 8**	
unobserved x_i, y_i	-2.824,	-3.236	-6.161,	-8.241	4.651,	7.976	5.708,	9.562
observed ξ_{ij}, η_{ij} j=1	-3.138,	-1.537	-6.486,	-8.251	3.935,	8.032	4.326,	9.285
j=2	-4.218,	-4.083	-4.809,	-8.558	5.686,	9.45	3.813,	10.241
j=3	-3.560,	-4.245	-6.220,	-9.289	5.002,	7.646	6.139,	8.999
	i = 9		**i = 10**		**i = 11**		**i = 12**	
unobserved x_i, y_i	2.110,	4.165	0.653,	1.979	2.798,	5.197	-0.014,	0.979
observed ξ_{ij}, η_{ij} j=1	2.577,	3.139	1.194,	2.37	1.444,	3.045	-1.278,	2.633
j=2	0.503,	1.833	0.487,	2.923	3.180,	4.780	0.784,	-0.913
j=3	2.526,	3.582	0.148,	3.296	2.525,	4.251	0.588,	1.914

APPENDIX A

Lemma 2.1 can be proved algebraically using (2.2) and establishing the identity $|\underset{\sim}{V}| = \sigma_\delta^{2r}\sigma_\epsilon^{2r}a$. However, we give a simple and direct proof here. Let $f(z_i)$, $g(x_i)$, and $f(z_i \mid x_i)$ be the density functions of z_i, x_i and z_i given x_i, respectively. Then,

$$f(z_i) = \int_{-\infty}^{\infty} f(z_i|x_i)g(x_i)\, dx_i$$

$$= \int_{-\infty}^{\infty} (2\pi)^{-(\frac{1}{2}+r)} \sigma_\delta^{-r}\sigma_\epsilon^{-r}\sigma^{-1} \exp\{-\frac{1}{2}\sigma_\delta^{-2} \textstyle\sum_r(\xi_{ij}-x_i)^2$$

$$-\frac{1}{2}\sigma_\epsilon^{-2} \textstyle\sum_r(\eta_{ij}-\alpha-\beta x_i)^2 - \frac{1}{2}\sigma^{-2}(x_i-\mu)^2\}\, dx_i .$$

Expanding and regrouping terms of the expression inside the exp, we find

$$f(z_i) = (2\pi)^{-(\frac{1}{2}+r)} \sigma_\delta^{-r}\sigma_\epsilon^{-r}\sigma^{-1} \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}(ax_i^2 + 2c_i'x_i + h_i')\}\, dx_i ,$$

where

$$h_i' = \frac{rm_{i\xi}}{\sigma_\delta^2} + \frac{r\alpha^2}{\sigma_\epsilon^2} - \frac{2r\alpha\bar{\eta}_{i\cdot}}{\sigma_\epsilon^2} + \frac{rm_{i\eta}}{\sigma_\epsilon^2} + \frac{\mu^2}{\sigma^2},$$

$$c_i' = \frac{r\alpha\beta}{\sigma_\epsilon^2} - \frac{r\bar{\xi}_{i\cdot}}{\sigma_\delta^2} - \frac{r\beta\bar{\eta}_{i\cdot}}{\sigma_\epsilon^2} - \frac{\mu}{\sigma^2},$$

$$m_{i\xi} = \textstyle\sum_r\xi_{ij}^2/r, \quad m_{i\eta} = \sum_r\eta_{ij}^2/r, \quad \bar{\xi}_{i\cdot} = \sum_r\xi_{ij}/r, \quad \bar{\eta}_{i\cdot} = \sum_r\eta_{ij}/r .$$

On completing the square, we have

$$f(z_i) = (2\pi)^{-r}\sigma_\delta^{-r}\sigma_\epsilon^{-r}\sigma^{-1}a^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left\{h_i' - \frac{(c_i')^2}{a}\right\}\right] .$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\ a^{-\frac{1}{2}}} \exp\left[-\frac{1}{2}\ \frac{1}{a^{-1}}\left\{x_i - (-\frac{c_i'}{a})\right\}^2\right] dx_i$$

$$= (2\pi)^{-r}\sigma_\delta^{-r}\sigma_\epsilon^{-r}\sigma^{-1}a^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left\{h_i' - \frac{(c_i')^2}{a}\right\}\right] .$$

The proof is completed by taking products of $f(z_i)$, $i = 1,\ldots, n$, and using the algebraic identity

$$\sum_n\left\{h_i' - \frac{(c_i')^2}{a}\right\} = \sum_n\left\{h_i - \frac{c_i^2}{a}\right\} .$$

APPENDIX B

From $\partial \ln L/\partial\alpha = 0$, $\partial \ln L/\partial\mu = 0$, we have

$$nr(\bar{\eta}_{..} - \alpha - \beta\mu) - \frac{r\beta}{a} \sum_n c_i = 0 , \tag{B1}$$

$$\frac{nr}{\sigma_\delta^2} (\bar{\xi}_{..} - \mu) + \frac{nr\beta}{\sigma_\epsilon^2} (\bar{\eta}_{..} - \alpha - \beta\mu) - \frac{1}{a} \left(\frac{r}{\sigma_\delta^2} + \frac{r\beta^2}{\sigma_\epsilon^2}\right) \sum_n c_i = 0 . \tag{B2}$$

(B1) and (B2) imply that

$$n(\bar{\xi}_{..} - \mu) - \frac{1}{a} \sum_n c_i = 0 , \tag{B3}$$

which together with (B1) give

$$\bar{\eta}_{..} - \alpha - \beta\mu = \beta(\bar{\xi}_{..} - \mu) . \tag{B4}$$

Hence

$$\sum_n c_i = \frac{nr}{\sigma_\delta^2} (\bar{\xi}_{..} - \mu) + \frac{nr\beta}{\sigma_\epsilon^2} (\bar{\eta}_{..} - \alpha - \beta\mu)$$

$$= n \left(a - \frac{1}{\sigma^2}\right) (\bar{\xi}_{..} - \mu) .$$

(B3) then gives

$$n(\bar{\xi}_{..} - \mu)\left\{1 - \frac{1}{a}\left(a - \frac{1}{\sigma^2}\right)\right\} = 0,$$

or

$$\frac{n}{\sigma^2} (\bar{\xi}_{..} - \mu) = 0 .$$

Since $\sigma^2 > 0$, we have $\mu = \bar{\xi}_{..}$ and also $\alpha + \beta\mu = \bar{\eta}_{..}$ from (B4).

APPENDIX C

From (3.1), one finds

$$(-n\beta a + \frac{1}{\sigma^2}\sum_n c_i\bar{\eta}_{i\cdot}) = \frac{nr^2}{\sigma_\delta^2\sigma_\epsilon^2}\{\beta^2 s_{\xi\eta} + \beta(\lambda s_{\xi\xi} - s_{\eta\eta}) - \lambda s_{\xi\eta}\} = 0, \qquad (C1)$$

where $\lambda = \sigma_\epsilon^2/\sigma_\delta^2$. Equations (3.1) and (3.2) together give

$$-n\beta + \frac{1}{a\sigma^2}\sum_n c_i\bar{\eta}_{i\cdot} = 0\,, \qquad (C2)$$

so that the first bracketed term in (C1) is zero and hence (3.5) is obtained. From (3.1) and (3.4) we obtain

$$-\frac{\beta}{a\sigma_\epsilon^2}\sum_n c_i\bar{\eta}_{i\cdot} - n + \frac{nt_{\eta\eta}}{\sigma_\epsilon^2} = 0\,.$$

(3.6) then follows from (C2). Adding (3.3) and (3.4) and using the definition of 'a', we have

$$-2n + \frac{n}{ar}(a - \frac{1}{\sigma^2}) + n(\frac{t_{\xi\xi}}{\sigma_\delta^2} + \frac{t_{\eta\eta}}{\sigma_\epsilon^2}) - \frac{2\beta}{a\sigma_\epsilon^2}\sum_n c_i\bar{\eta}_{i\cdot}$$
$$- \frac{2}{a\sigma_\delta^2}\sum_n c_i\bar{\xi}_{i\cdot} + \frac{1}{a^2 r}(a - \frac{1}{\sigma^2})\sum_n c_i^2 = 0\,,$$

and by (3.2),

$$-2n + \frac{1}{ar}\sum_n c_i^2 + n(\frac{t_{\xi\xi}}{\sigma_\delta^2} + \frac{t_{\eta\eta}}{\sigma_\epsilon^2}) - \frac{2\beta}{a\sigma_\epsilon^2}\sum_n c_i\bar{\eta}_{i\cdot} - \frac{2}{a\sigma_\delta^2}\sum_n c_i\bar{\xi}_{i\cdot} = 0\,. \qquad (C3)$$

It can be verified that

$$\frac{1}{r}\sum_n c_i^2 - \frac{\beta}{\sigma_\epsilon^2}\sum_n c_i\bar{\bar{\eta}}_{i\cdot} - \frac{1}{\sigma_\delta^2}\sum_n c_i\bar{\bar{\xi}}_{i\cdot} = 0,$$

and hence (C3) is reduced to

$$-2nr - \frac{1}{a}\sum_n c_i^2 + nr\left(\frac{t_{\xi\xi}}{\sigma_\delta^2} + \frac{t_{\eta\eta}}{\sigma_\epsilon^2}\right) = 0\,. \qquad \text{(C4)}$$

Using (3.1) and (3.5), we can obtain equation (3.8) from (C4). Also by (3.2), (3.6) and (C4) we have equation (3.7). From (C2), using the definitions of 'a' and 'c_i' we have

$$\sigma^2 = (r\lambda s_{\xi\eta} + r\beta s_{\eta\eta} - \beta\sigma_\epsilon^2)/\{\beta r(\lambda + \beta^2)\}.$$

Substituting this into (3.6) yields

$$\sigma_\epsilon^2(\beta^2 - r\beta^2 - r\lambda) + r(\lambda + \beta^2)t_{\eta\eta} - \beta r(\lambda s_{\xi\eta} + \beta s_{\eta\eta}) = 0\,. \qquad \text{(C5)}$$

Eliminating σ_ϵ^2 from (3.8) and (C5), we obtain

$$\{(t_{\xi\xi}\lambda + t_{\eta\eta}) - (\beta s_{\xi\eta} + \lambda s_{\xi\xi})\}\{-(r - 1)\beta^2 - r\lambda\} + (2r - 1)(\lambda + \beta^2)t_{\eta\eta}$$
$$- \beta(2r - 1)(\lambda s_{\xi\eta} + \beta s_{\eta\eta}) = 0,$$

which can be simplified, using (3.5) and the relations $t_{\xi\xi} = w_{\xi\xi} + s_{\xi\xi}$, $t_{\eta\eta} = w_{\eta\eta} + s_{\eta\eta}$, to

$$\beta^2 r w_{\eta\eta} - \beta^2 (r - 1)\lambda t_{\xi\xi} + (r - 1)\lambda t_{\eta\eta} - r\lambda^2 w_{\xi\xi} = 0 . \quad (C6)$$

Substituting $\lambda = (\beta^2 s_{\xi\eta} - \beta s_{\eta\eta})/(s_{\xi\eta} - \beta s_{\xi\xi})$

of (3.5) into (C6), we finally obtain

$$k_0\beta^4 + k_1\beta^3 + k_2\beta^2 + k_3\beta + k_4 = 0.$$

REFERENCES

ANDERSON, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann. Math. Statist., 22, 327-351.

BARNETT, V. D. (1970). Fitting straight lines - the linear functional relationships with replicated observations. Appl. Statist., 19, 135-44.

CRAMÉR, H. (1946). Mathematical Methods of Statistics, Princeton: Princeton University Press.

DOLBY, G. R. (1976). The ultrastructural relation: A synthesis of the functional and structural relations. Biometrika, 63, 39-50.

DORFF, M. and GURLAND, J. (1961a). Estimation of the parameters of a linear functional relationship. J. R. Statist. Soc. B, 23, 160-70.

DORFF, M. and GURLAND, J. (1961b). Small sample behaviour of slope estimators in a linear functional relation. Biometrics, 17, 283-98.

HOUSNER, G. W. and BRENNAN, J. F. (1948). The estimation of linear trends. Ann. Math. Statist., 19, 380-88.

KENDALL, M. G. and STUART, A. (1973). The Advanced Theory of Statistics, Vol. 2, 3rd edition. London: Griffin.

MADANSKY, A. (1959), The fitting of straight lines when both variables are subject to error. J. Amer. Statist. Assoc., 54, 173-205.

MORAN, P. A. P. (1971). Estimating structural and functional relationships, J. Mult. Anal., 1, 232-55.

PRESS, J. (1972). Applied Multivariate Analysis. New York: Holt, Rinehart and Winston.

RICHARDS, F. S. G. (1961). A method of maximum-likelihood estimation. J. R. Statist. Soc. B, 23, 469-75.

TUKEY, J. W. (1951). Components in regression. Biometrics, 7, 33-69.

VILLEGAS, C. (1961). Maximum likelihood estimation of a linear functional relationship. Ann. Math. Statist., 32, 1048-62.

CHAPTER 4

TWO ADAPTIVE PROCEDURES FOR ESTIMATION OF A LINEAR STRUCTURAL RELATIONSHIP

1. INTRODUCTION

In some practical situations such as estimation of the relation between income and some component of consumption as pointed out by Malinvaud (1970, p. 374), the assumption on regression analysis that the independent variable can be observed exactly may be unrealistic. A more appropriate model would be

$$y = \alpha + \beta x,$$
$$\xi = x + \delta, \qquad \eta = y + \epsilon, \tag{1.1}$$

where α and β are unknown and to be estimated, δ and ϵ are uncorrelated error terms with zero means and variances σ_δ^2 and σ_ϵ^2, respectively, and only ξ and η are observable. The relationship between x and y is usually called a functional relationship if x is nonstochastic and a structural relationship if x is a random variable independent of δ and ϵ. Comprehensive reviews of the problem of estimating linear functional and structural relationships were given by Kendall and Stuart (1973, Chapter 29), Madansky (1959), Malinvaud (1970, Chapter 10), and Moran (1971).

Here we are concerned with the estimation of β in the linear structural relationship. After an estimate $\hat{\beta}$ is obtained, α can be estimated by $\bar{\eta}. - \hat{\beta}\bar{\xi}.$, where $\bar{\eta}.$ and $\bar{\xi}.$ are sample means.

The ordinary least squares (OLS) estimate of β is known to be inconsistent, and what is worse, if x, δ and ϵ are independent and normally distributed with unknown variances, β becomes unidentifiable (see Kendall and Stuart (1973 , Chapter 29); Moran (1971)). Different approaches have been used to estimate β consistently and to overcome the problem of unidentifiability. In one approach, additional information about the variances of the error terms is assumed to be available, e.g., σ_δ^2 (or σ_ϵ^2) or $\sigma_\epsilon^2/\sigma_\delta^2$ is known, or both σ_δ^2 and σ_ϵ^2 are known (Birch (1964), Lindley (1947)). In econometrics, a common procedure is to use instrumental variables which are correlated with x but not with δ and ϵ (Geary, 1949; Reiersøl, 1945). In other approaches, the true ordering of x is assumed to be known (Dorff and Gurland 1961a, 1961b), or grouping of observations is assumed to be possible (Wald, 1940). If x has a non-normal distribution, then β is identifiable and consistent estimates were proposed by Drion (1951), Geary (1942), Scott (1950), and Wolfowitz (1952). Although estimates based on instrumental variables, true ordering of x, grouping and non-normality of x are consistent, they are usually unreliable in finite samples (cf. Madansky (1959)), and quite often perform much worse than the OLS estimate. When the variance of δ is large, Feldstein (1974) combined the instrumental variables estimate linearly with the OLS estimate and achieved a reduction of mean squared error (MSE) in finite samples while preserving consistency. Although his procedure does not dominate the OLS estimate, the loss in efficiency when the latter is superior seems to be outweighed by the substantial gain in efficiency when the former is superior. This is a good practical procedure. However, the algebra involved in applying Feldstein's procedure to other cases is usually formidable, especially in structural relationship. The extension of his procedure to multivariate cases would be a more complicated problem.

In section 3 we propose two adaptive procedures for the estimation of β, the Pre-test Procedure (PP) and Estimated Ratio Procedure (ERP) which are based on the idea of constructing an estimate by comparing the sample asymptotic variances of various estimates. When they are applied to the consistent estimates by Geary, Wolfowitz and Scott (section 2), consistency is preserved and the MSE's estimated by Monte Carlo experiments are improved. The adaptive procedures in general also yield smaller estimated MSE's than that of the OLS estimate (section 4).

An extension of the ERP to more than one independent variable x is also proposed (section 5).

If we have n independent observations (ξ_i, η_i) each generated from (1.1), the model can be written as

$$\xi_i = x_i + \delta_i, \quad \eta_i = y_i + \epsilon_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.2}$$

where $(x_i, \delta_i, \epsilon_i)$, $i = 1, \ldots, n$, are independent and identically distributed and x_i, δ_i and ϵ_i are mutually independent with unknown variances $\sigma^2, \sigma_\delta^2$ and σ_ϵ^2, respectively. Also $\beta \neq 0$, $E(x_i) = \mu$, $E(\delta_i) = E(\epsilon_i) = 0$ and $\mu_3 = E(x_i^3) \neq 0$ for all i. Assume that the sixth moments of x_i, δ_i and ϵ_i exist so that the sixth product moments of ξ_i and η_i exist. Let $\mu_k = E(x_i - \mu)^k$,

$\mu_{k\ell} = E(\xi_i - \mu)^k(\eta_i - \alpha - \beta\mu)^\ell$ and let $s_{k\ell} = \sum_n (\xi_i - \bar{\xi}.)^k(\eta_i - \bar{\eta}.)^\ell/n$ be the corresponding sample estimate, where $\sum_n$ denotes $\sum_{i=1}^{n}$, $\bar{\xi}. = \sum_n \xi_i/n$, and $\bar{\eta}. = \sum_n \eta_i/n$.

2. SOME CONSISTENT ESTIMATES

The following consistent estimates will be used in the two adaptive procedures in section 3.

1. Geary's (1942) estimate.

$$\hat{\beta}_G = s_{12}/s_{21}.$$

It is consistent if $\beta \neq 0$.

2. Wolfowitz's (1952) estimate.

$$\hat{\beta}_W = (s_{03}/s_{30})^{\frac{1}{3}}.$$

It is consistent.

3. Modified Scott's (1950) estimate.

$$\hat{\beta}_S = s_{21}/s_{30} \quad \text{if } \hat{AV}(s_{21}/s_{30}) \le \hat{AV}(s_{03}/s_{12}),$$

$$= s_{03}/s_{12} \quad \text{otherwise},$$

where $\hat{AV}(s_{21}/s_{30})$ and $\hat{AV}(s_{03}/s_{12})$ are estimates of the asymptotic variances of s_{21}/s_{30} and s_{03}/s_{12}, respectively, and they can be constructed by the method described in section 3. Since s_{21}/s_{30} and s_{03}/s_{12} are both consistent when $\beta \neq 0$, $\hat{\beta}_S$ is consistent.

The estimate s_{21}/s_{30} is related to Scott's (1950) consistent estimate which is a root of

$$f_n(b) = \sum_n [(\eta_i - \bar{\eta}.) - b(\xi_i - \bar{\xi}.)]^3/n = 0.$$

But since $f_n'(\beta) \xrightarrow{p} 0$, $f_n''(\beta) \xrightarrow{p} 0$ and are of order $n^{-\frac{1}{2}}$, where $\xrightarrow{p}$ denotes convergence in probability, a Taylor's expansion of f_n at β shows that $\hat{\beta}_n - \beta$ is of order greater than $n^{-\frac{1}{2}}$. Hence asymptotically Scott's estimate has zero efficiency relative to $\hat{\beta}_G$ and $\hat{\beta}_W$. Thus a reasonable substitute for Scott's estimate would be a root of $f_n''(b) = 0$, since $f_x'''(\beta) \xrightarrow{p} -6\mu_3 \neq 0$. Obviously, s_{21}/s_{30} is the unique root of $f_n''(b) = 0$.

The rationale of choosing adaptively between s_{21}/s_{30} and s_{03}/s_{12} for $\hat{\beta}_S$ is as follows. Let $\hat{\beta}$ be a consistent estimate of β in the model (1.2). Now rewrite (1.2) as

$$\eta_i = y_i + \epsilon_i, \quad \xi_i = -\alpha\beta^{-1} + \beta^{-1} y_i + \delta_i, \quad i = 1, \ldots, n \tag{2.1}$$

which is called the dual model of (1.2) and has the same form as (1.2) with the roles of x and y, and hence ξ and η, interchanged. The method used in constructing $\hat{\beta}$ can be used to construct a consistent estimate $\widehat{\beta^{-1}}$ of the β^{-1} in (2.1). The reciprocal of $\widehat{\beta^{-1}}$, denoted by $\hat{\beta}^c$ and called the conjugate of $\hat{\beta}$, is a consistent estimate of β. It can be proved that $\hat{\beta}$ is the conjugate estimate of $\hat{\beta}^c$, i.e., $\hat{\beta}^{cc} = \hat{\beta}$. If $\hat{\beta}^c = \hat{\beta}$, we call $\hat{\beta}$ a self-conjugate estimate. We note that

(i) $\hat{\beta}_G$ and $\hat{\beta}_W$ are self-conjugate estimates;

(ii) s_{21}/s_{30} and s_{03}/s_{12} are conjugates of each other.

If $\hat{\beta}^c \neq \hat{\beta}$, define the completion $c(\hat{\beta}, \hat{\beta}^c)$ of the consistent estimates $\hat{\beta}$ and $\hat{\beta}^c$ by

$$\begin{aligned} c(\hat{\beta}, \hat{\beta}^c) &= \hat{\beta} \quad \text{if } AV(\hat{\beta}) \leq AV(\hat{\beta}^c) \\ &= \hat{\beta}^c \quad \text{otherwise.} \end{aligned}$$

$\hat{\beta}_S$ is simply $c(s_{21}/s_{30}, s_{03}/s_{12})$. It can be shown that $c(\hat{\beta},\hat{\beta}^c)$ is self-conjugate and hence no new conjugate estimate can be constructed. The precision of a consistent estimate $\hat{\beta}$ of β usually depends on β. It may be good for large values of β but not for small values (or vice versa). Thus it is quite natural to ask whether accuracy can be gained by estimating β^{-1} in (2.1) first and taking the reciprocal. This is the basic motivation for proposing conjugate estimates. The idea of completion therefore gives a guideline to decide whether to apply the original estimating procedure $\hat{\beta}$ to the model (1.2) or to estimate β through model (2.1) using the conjugate estimate $\hat{\beta}^c$ of $\hat{\beta}$. The method of conjugate estimation and completion does not apply only to the construction of the modified Scott's estimate; it can be applied to any other consistent estimates of β in (1.2). From preliminary Monte Carlo experiments we found that s_{21}/s_{30}, s_{03}/s_{12} are usually no better than $\hat{\beta}_G$, $\hat{\beta}_W$, but in section 4 it will be seen that $\hat{\beta}_S$ has much higher precision than $\hat{\beta}_G$ and $\hat{\beta}_W$.

3. TWO ADAPTIVE PROCEDURES IN FINITE SAMPLES

Let

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix} = \begin{bmatrix} \sigma^2 + \sigma_\delta^2 & \beta\sigma^2 \\ \beta\sigma^2 & \beta^2\sigma^2 + \sigma_\epsilon^2 \end{bmatrix} \tag{3.1}$$

be the covariance matrix of (ξ,η). Thus

$$0 \le \sigma_\delta^2 = \mu_{20} - \mu_{11}\beta^{-1}, \quad 0 \le \sigma_\epsilon^2 = \mu_{02} - \mu_{11}\beta, \tag{3.2}$$

so that when $\beta > 0$ ($\beta < 0$)

$$\mu_{11}/\mu_{20} \le \beta \le \mu_{02}/\mu_{11} \quad (\mu_{02}/\mu_{11} \le \beta \le \mu_{11}/\mu_{20}). \tag{3.3}$$

Let $\hat{\beta}_L = s_{11}/s_{20}$ be the OLS estimate and $\hat{\beta}_U = s_{02}/s_{11}$ be the reciprocal of the least squares regression of ξ on η. Then $|\hat{\beta}_L| \le |\hat{\beta}_U|$. Since with probability going to one, $\hat{\beta}_L \to \mu_{11}/\mu_{20}$ and $\hat{\beta}_U \to \mu_{02}/\mu_{11}$, we have asymptotically, when $\beta > 0$ ($\beta < 0$), $\hat{\beta}_L \le \beta \le \hat{\beta}_U$ ($\hat{\beta}_U \le \beta \le \hat{\beta}_L$). Let $\hat{\beta}$ be a consistent estimate of β. The asymptotic bias (AB) of $\hat{\beta}_L$ can be estimated by $\hat{\beta}_L - \hat{\beta}$. However, when $\beta > 0$, asymptotically $\hat{\beta}_L - \beta < 0$ and $\hat{\beta}_U - \beta > 0$. To preserve these inequalities when $\beta > 0$ for finite samples, estimate β by the consistent estimate

$$\begin{aligned} \hat{\beta}_M &= \hat{\beta}_L \text{ if } \hat{\beta} < \hat{\beta}_L, \\ &= \hat{\beta} \text{ if } \hat{\beta}_L \le \hat{\beta} \le \hat{\beta}_U, \\ &= \hat{\beta}_U \text{ if } \hat{\beta}_U < \hat{\beta}. \end{aligned}$$

Since in general it is unknown whether $\beta > 0$, replace this condition by $\hat{\beta}_L > 0$. This seems to be reasonable since the s_{11} in $\hat{\beta}_L = s_{11}/s_{20}$ is a consistent estimate of $\beta\sigma^2$ and hence asymptotically β and $\hat{\beta}_L$ have the same sign. When $\beta < 0$, all the inequalities are reversed. $\hat{\beta}_L$, $\hat{\beta}_U$ and the three estimates in section 2 are functions g of t_1/t_2, where t_1 and t_2 are sample statistics. The asymptotic MSE is

$$\text{AMSE}(g(t_1/t_2)) = \text{AV}(g(t_1/t_2)) + [\text{AB}(g(t_1/t_2))]^2.$$

We have described how to estimate the AB of $\hat{\beta}_L$ and $\hat{\beta}_U$ when there is a consistent estimate. $\text{AV}(g(t_1/t_2))$ is a function of $\text{AV}(t_1)$, $\text{AV}(t_2)$ and $\text{ACov}(t_1, t_2)$ (cf. Kendall and Stuart (1977, Eq. (10.12))).

So, for example when $g(t_1/t_2) = \hat{\beta}_L = s_{11}/s_{20}$,

$$AV(s_{11}/s_{20}) = (\mu_{11}/\mu_{20})^2\{AV(s_{11})\mu_{11}^{-2} + AV(s_{20})\mu_{20}^{-2} - 2ACov(s_{11},s_{20})\mu_{11}^{-1}\mu_{20}^{-1}\},$$

where $AV(s_{11})$, $AV(s_{20})$ and $ACov(s_{12},s_{20})$ are functions of μ_{ij} (cf. Kendall and Stuart (1977, Eq. (10.23), (10.24))). Since some or all of the μ_{ij} are unknown, the $AV(s_{11}/s_{20})$ is estimated by $\hat{A}V(s_{11}/s_{20})$ obtained from estimating each unknown μ_{ij} by s_{ij}.

In this section, we propose two adaptive procedures which would lead to possible improvement of efficiency in finite samples. A stimulating description of the principle of adaptive procedures through robust estimation was given by Hogg (1974). When applying the two procedures to a consistent estimate $\hat{\beta}$ such as $\hat{\beta}_G$, $\hat{\beta}_W$ and $\hat{\beta}_S$, we need estimates of the asymptotic MSE's of these estimates (which have zero AB) and of $\hat{\beta}_L$ and $\hat{\beta}_U$ which can be constructed using the methods just described.

1. The Pre-test Procedure (PP). The PP chooses the estimate among $\hat{\beta}_L$, $\hat{\beta}_U$ and $\hat{\beta}$ with the smallest estimated asymptotic MSE. Although $\hat{\beta}_L$ and $\hat{\beta}_U$ are in general inconsistent, in finite samples, the consistency of $\hat{\beta}$ does not guarantee that it is superior to $\hat{\beta}_L$ and $\hat{\beta}_U$. Thus in the PP, the estimate $\hat{\beta}_P$ is defined adaptively as one of $\hat{\beta}_L$, $\hat{\beta}_U$ and $\hat{\beta}$ (cf. Feldstein (1972) in the case of instrumental variables).

2. The Estimated Ratio Procedure (ERP). Define a class of estimates of β indexed by λ, $0 \le \lambda \le \infty$, by

$$\hat{\beta}(\lambda) = \{(s_{02} - \lambda s_{20}) + [(s_{02} - \lambda s_{20})^2 + 4\lambda s_{11}^2]^{\frac{1}{2}}\}/2s_{11}. \tag{3.4}$$

$\hat{\beta}(\lambda)$ is consistent if and only if $\lambda = \sigma_\epsilon^2/\sigma_\delta^2$, and it is the generalized least squares estimate proposed by Sprent (1966), which was proposed originally for functional relationship when $\sigma_\epsilon^2/\sigma_\delta^2$ is known. Clearly, $\hat\beta(0) = \hat\beta_U$, $\hat\beta(\infty) = \hat\beta_L$ and $\hat\beta(\lambda)$ is a strictly monotone function of λ (decreasing if $\hat\beta_L > 0$, increasing otherwise). When the value of $\sigma_\epsilon^2/\sigma_\delta^2$ is unknown and estimated by $\hat\lambda' = \hat\sigma_\epsilon^2/\hat\sigma_\delta^2$, where $\hat\sigma_\epsilon^2 = s_{02} - s_{11}\hat\beta$ and $\hat\sigma_\delta^2 = s_{20} - s_{11}(\hat\beta)^{-1}$ (cf. (3.2)), then $\hat\beta(\hat\lambda') = \hat\beta$ provided that $2\, s_{11}\hat\beta \geq s_{02} - \hat\lambda s_{20}$ (cf. eq. (3.4)). To propose a different estimate, consider $h(\beta) = (s_{02} - s_{11}\beta)/(s_{20} - s_{11}\beta^{-1})$. Expanding it at $\hat\beta$ to the second order term and replacing $\beta - \hat\beta$ and $(\beta - \hat\beta)^2$ by $AE(\beta - \hat\beta) = 0$ and a consistent estimate $\hat{A}V(\hat\beta)$ of $AV(\hat\beta)$, respectively, we obtain

$$\hat\lambda' + \hat\beta^{-4}(\hat\beta s_{20} - s_{11}(\hat\beta)^{-1})^{-3}(\hat\beta s_{11})(s_{20}s_{02} - s_{11}^2)\hat{A}V(\hat\beta). \qquad (3.5)$$

Now let $\hat\lambda = \infty$ if $s_{20} - s_{11}(\hat\beta)^{-1} \leq 0$ or $\hat\beta s_{11} < 0$, let $\hat\lambda = 0$ if (3.5) is negative, and let $\hat\lambda$ be (3.5) otherwise. In the ERP, the estimate $\hat\beta_R$ is defined to be $\hat\beta(\hat\lambda)$. It is easily seen from the definition of $\hat\beta_R$ that if $\hat\lambda$ is not given by (3.5), then $\hat\beta_R$ equals the truncated estimate $\hat\beta_M$. Otherwise, $\hat\lambda > \hat\lambda'$ since the second term of (3.5) becomes non-negative. Hence the ERP pulls $\hat\beta(\lambda)$ towards $\hat\beta_L$. The second term of (3.5) can be expressed as

$$(\hat\beta\hat\sigma_\delta^2)^{-3}s_{11}(s_{20}s_{02} - s_{11}^2)\hat{A}V(\hat\beta)$$

which increases with $\hat{A}V(\hat\beta)$ and decreases with $|\hat\beta|$ and $\hat\sigma_\delta^2$. The pull towards $\hat\beta_L$ is large when $\hat{A}V(\hat\beta)$ is large and the pull is small when $|\hat\beta|$ and $\hat\sigma_\delta^2$ are large. This is intuitively reasonable since if $\hat\beta$ has large asymptotic variance, we are in favour of $\hat\beta_L$. On the other hand if $|\beta|$ and σ_δ^2 are large, the bias of $\hat\beta_L$ becomes severe (the asymptotic bias of $\hat\beta_L$ is $\beta\sigma_\delta^2/(\sigma^2 + \sigma_\delta^2)$) and therefore less weight should be placed on $\hat\beta_L$. Since $\hat\beta$ is consistent, $AV(\hat\beta) \to 0$ as $n \to \infty$, hence the pull becomes small and we are in favour of $\hat\beta$ in large samples. This is desirable since $\hat{\beta}$ is consistent and $\hat{\beta}_L$ is asymptotically biased (cf. Feldstein (1974)). In fact, by using the definition of convergence in probability and the Taylor's expansion of $\hat{\beta}(\lambda)$ at $\hat{\lambda} = \hat{\sigma}_\epsilon^2/\hat{\sigma}_\delta^2$ we can prove the following theorem.

<u>THEOREM 3.1</u>. Let $\hat{\beta}$ be a consistent estimate of β in (1.2); then (a) $\hat{\beta}_P$ is consistent and $n^r(\hat{\beta}_P - \hat{\beta}) \xrightarrow{P} 0$ as $n \to \infty$, for any $r > 0$;

(b) $\hat{\beta}_R$ is consistent, and if $AV(\hat{\beta}) = o(n^{-r})$, then $n^r(\hat{\beta}_R - \hat{\beta}) \xrightarrow{P} 0$ as $n \to \infty$.

Theorem 3.1 says that asymptotically all the estimates $\hat{\beta}_P$, $\hat{\beta}_R$ and $\hat{\beta}$ are equivalent. The advantages of using $\hat{\beta}_P$ and $\hat{\beta}_R$ therefore lie in finite samples. We apply PP and ERP to $\hat{\beta} = \hat{\beta}_G, \hat{\beta}_S, \hat{\beta}_W$ (and use $\hat{\beta}_{GP}$ to denote the PP applied to $\hat{\beta}_G$, and so on) and their efficiences are evaluated by Monte Carlo experiments in the next section.

4. RESULTS ON MONTE CARLO STUDIES

Given the covariance matrix $\underset{\sim}{\Sigma}$ of (ξ,η) (which can be estimated consistently by the sample covariance matrix) in (3.1), one cannot determine uniquely β, σ^2, σ_δ^2 and σ_ϵ^2. In fact for every $\beta \in [\mu_{11}\mu_{20}^{-1}, \mu_{02}\mu_{11}^{-1}]$ ($[\mu_{02}\mu_{11}^{-1}, \mu_{11}\mu_{20}^{-1}]$ if $\mu_{11} < 0$; without loss of generality, we assume subsequently $\mu_{11} > 0$), the set of parameter values

$$\beta, \quad \sigma^2 = \mu_{11}/\beta, \quad \sigma_\delta^2 = \mu_{20} - \sigma^2, \quad \sigma_\epsilon^2 = \mu_{02} - \beta\mu_{11}, \tag{4.1}$$

would give the same $\underset{\sim}{\Sigma}$. For β lying outside that interval, (4.1) does not give admissible values since one of σ^2, σ_δ^2 and σ_ϵ^2 is negative. The precision of an estimate of β depends on the set of parameters which actually generates $\underset{\sim}{\Sigma}$. For this reason, it is interesting to look at, for a given $\underset{\sim}{\Sigma}$, the "average (overall admissible values of β) MSE" defined as follows.

Let $\underset{\sim}{\Sigma}$ be the positive definite matrix in (3.1) and let f be a density function. For $\beta \in [\mu_{11}\mu_{20}^{-1}, \mu_{02}\mu_{11}^{-1}]$, let the parameters σ^2, σ_δ^2 and σ_ϵ^2 in (1.2) be given by (4.1). Furthermore we specify the distribution of x as that of $z/\{\sigma_z(\beta\mu_{11}^{-1})^{1/2}\}$, where z has density function f, and σ_z^2 is the variance of z. The distribution of (ξ,η) in (1.2) is then completely determined (setting $\alpha = 0$). Let $\hat{\beta}$ be any estimate of β. We call

$$(\mu_{02}\mu_{11}^{-1} - \mu_{11}\mu_{20}^{-1})^{-1} \int_{\mu_{11}\mu_{20}^{-1}}^{\mu_{02}\mu_{11}^{-1}} E_\beta(\hat{\beta} - \beta)^2 \, d\beta \tag{4.2}$$

the $(\underset{\sim}{\Sigma},f)$-square error (s.e.) of $\hat{\beta}$, where $E_\beta(\hat{\beta} - \beta)^2$ is the expected value of $(\hat{\beta} - \beta)^2$ when the distribution of (ξ,η) is specified by $(\underset{\sim}{\Sigma},f)$ and a given β. (4.2) is an average of $E_\beta(\hat{\beta} - \beta)^2$.

In finite samples, the (original and adaptive) consistent estimates have infinite second moments and hence their MSE's may be infinite. This may be due to heavy tails at the extremes of their density functions although their probability mass within a wide interval is much higher than that of $\hat{\beta}_L$. This leads to the consideration of MSE's based on truncated distributions of the estimates, i.e., $E_\beta[(\hat{\beta} - \beta)^2 \mid \hat{\beta} \in (-a,b)]$, where a and b are large positive numbers.

The MSE's used for computing the efficiencies in the table are these finite "truncated" MSE's estimated by the Monte Carlo experiments.

Similar consideration to the difficulty of "estimating "infinite variances was also given by Feldstein (1974).

Given $\underset{\sim}{\Sigma}$, let $\beta(\theta) = \mu_{11}\mu_{20}^{-1} + \theta[\mu_{02}\mu_{11}^{-1} - \mu_{11}\mu_{20}^{-1}]$, where $\theta \in [0,1]$. The $(\underset{\sim}{\Sigma}, f)$ - s.e. gives an indication of the average performance of $\hat{\beta}$ over the range of $\beta(\theta)$. It is also interesting to know the efficiency of an estimate for different values of θ. In the Monte Carlo experiments, $\underset{\sim}{\Sigma}$ was set at two levels:

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \qquad \mu_{11}\mu_{20}^{-1} = 1, \qquad \mu_{02}\mu_{11}^{-1} = 2$$

$$\underset{\sim}{\Sigma}_2 = \begin{bmatrix} 1 & 5 \\ 5 & 27 \end{bmatrix}, \qquad \mu_{11}\mu_{20}^{-1} = 5, \qquad \mu_{02}\mu_{11}^{-1} = 5.4.$$

f was set at f_1 = density of $\Gamma(1.5, 1)$ and f_2 = density of $B(0.5, 0.2)$. The sample size was fixed at n = 20 and n = 50. The values of $(\underset{\sim}{\Sigma}_i, f_j)$ - s.e., $i,j = 1,2$, and the MSE's for several values of θ were estimated from 1000 simulated samples. The efficiency is defined to be the ratio of the MSE of the OLS estimate $\hat{\beta}_L$ to the MSE of the estimate considered. Results of the Monte Carlo experiments are given in the table from which we draw the following conclusions.

(i) $\hat{\beta}_S$ dominates $\hat{\beta}_G$ and $\hat{\beta}_W$ and there are cases where $\hat{\beta}_S$ is substantially more efficient than $\hat{\beta}_G$ and $\hat{\beta}_W$. $\hat{\beta}_G$ and $\hat{\beta}_W$ have very low precision and for most θ-values remain much worse than $\hat{\beta}_L$ even when the sample size is as large as 50.

(ii) The PP and the ERP dominate the original estimates and the gain in efficiencies from applying these procedures is substantial. This is true even when the original estimate is more efficient than the OLS estimate $\hat{\beta}_L$. The efficiencies using the PP and the ERP increase as the sample size n and θ increase.

(iii) The ERP dominates the PP in $(\sum, f)$-s.e. except the $\hat{\beta}_{GP}$ and $\hat{\beta}_{GR}$ for $(\sum_1, f_1)$ and n = 20. The ERP is also superior to the PP for most θ-values and the percentage gain in efficiency when PP is superior is less than the percentage gain in efficiency when the ERP is superior. In this sense, the ERP is preferable to the PP.

(iv) Although the ERP does not dominate $\hat{\beta}_L$ for every value of θ, it is superior to the latter in $(\sum_i, f_j)$-s.e. when n = 50 and even in some cases when n = 20. When $\hat{\beta}_L$ is substantially biased, the gain in efficiency when the ERP is used may be substantial. Generally speaking, for $\hat{\beta}_\gamma$, γ = G, W and S, the estimate $\hat{\beta}_{\gamma R}$ has at least an efficiency of 75% relative to the optimum estimate in the class $\{\hat{\beta}_L, \hat{\beta}_U, \hat{\beta}_{\gamma P}, \hat{\beta}_{\gamma R}\}$ and in many cases is itself the optimum estimate even when the sample size is as small as 20.

(v) Conclusions (ii) to (iv) therefore suggest that the ERP should be used in general. The question is to which estimate it should be applied. The ERP when applied to $\hat{\beta}_W$ seems to give slightly better results than when it is applied to $\hat{\beta}_G$ and $\hat{\beta}_S$. However, conclusion (i) suggests that it is quite possible that $\hat{\beta}_S$ is superior to $\hat{\beta}_G$ and $\hat{\beta}_W$ in large samples. Since the ERP converges asymptotically to the original estimate, it seems to be reasonable to use the estimate $\hat{\beta}_{SR}$ in estimating β when the experimenter has no additional information about the errors and σ_δ^2 is not negligible.

REMARK. The MSE used in computing the efficiency of an estimate $\hat{\beta}$ for each θ in the table was calculated based on all the 1000 samples simulated. This MSE thus represents an estimate of the conditional MSE $E[(\hat{\beta} - \beta)^2 \mid \hat{\beta} \in (-a,a)]$ where $Pr(\hat{\beta} \in (-a,a)) > 1-\gamma$ with γ very small (say 0.00001) so that the expected number of the 1000 simulated $\hat{\beta}$ falling outside $(-a,a)$ is less than 1. A disadvantage of choosing γ too small is that for an estimate $\hat{\beta}$ the distribution of which has thick tails, the estimated MSE's could be substantially inflated due to the presence of a few extreme values although $\hat{\beta}$ may highly center around β. Since in practice β is estimated based on one sample, an estimate $\hat{\beta}$ with distribution highly centering around β is preferred even though there is very small probability of getting a value far from β. Thus in our Monte Carlo experiments, the MSE's of the $\hat{\beta}$'s were also estimated based on the simulated $\hat{\beta}$ values retained after a total of γ (= 1, 5 and 10)% of the 1000 simulated $\hat{\beta}$ were deleted from both ends in such a way that the estimated MSE is the minimum among all truncations with the same γ. The simulated MSE's are then estimates of the quantity $\min_{(a,b)} \{E[(\hat{\beta} - \beta)^2 \mid \hat{\beta} \in (a,b)]: Pr(a,b) = 1 - 100^{-1}\gamma\}$. The efficiencies (not presented here) relative to $\hat{\beta}_L$ (also truncated with the same γ) were then calculated.

It was then observed that as γ increases, the efficiences of all the estimates increase for every θ except when $\theta = 0$ the relative efficiencies decrease slightly as γ increases. Even when $\gamma = 1$, the estimated MSE's of all the estimates discussed (except $\hat{\beta}_L$) are substantially less than the MSE's estimated based on all 1000 simulated values. This indicated that the estimates may have a few extreme values. For the untruncated simulation, in general $\hat{\beta}_W$ is better than $\hat{\beta}_G$. However, after truncation, $\hat{\beta}_G$ improves substantially and performs much better than $\hat{\beta}_W$ (this is true even for $\gamma = 1$), but remains inferior to $\hat{\beta}_S$. The relative efficiences of $\hat{\beta}_{GR}$, $\hat{\beta}_{WR}$ and $\hat{\beta}_{SR}$ become very close. The ERP remains better than the PP and the increases of efficiences over the original estimates remain substantial.

5. SOME GENERALIZATIONS

In this section, we extend the ERP to the estimation of $\underset{\sim}{\beta}$ in the model

$$\begin{aligned} \eta_i &= \alpha + \underset{\sim}{\beta}'\underset{\sim}{x}_i + \epsilon_i, \\ \underset{\sim}{\xi}_i &= \underset{\sim}{x}_i + \underset{\sim}{\delta}_i, \end{aligned} \qquad i = 1,\ldots, n,$$

where $\underset{\sim}{\beta}$, $\underset{\sim}{x}_i$, $\underset{\sim}{\xi}_i$ and $\underset{\sim}{\delta}_i$ are p-component column vectors, $(\underset{\sim}{x}_i, \underset{\sim}{\delta}_i, \epsilon_i)$, $i = 1,\ldots, n$, are i.i.d., $\underset{\sim}{x}_i$, $\underset{\sim}{\delta}_i$ and ϵ_i are independent and

$$\text{Var}(\underset{\sim}{x}_i) = \underset{\sim}{\Sigma}, \qquad \text{Var}(\underset{\sim}{\delta}_i) = \begin{bmatrix} \sigma^2_{\delta_1} & & \underset{\sim}{0} \\ & \ddots & \\ \underset{\sim}{0} & & \sigma^2_{\delta_p} \end{bmatrix} = \underset{\sim}{\Sigma}_\delta, \qquad \text{Var}(\epsilon_i) = \sigma^2_\epsilon .$$

Let

$$\underset{\sim}{S} = \begin{bmatrix} \underset{\sim}{S}_{\xi\xi} & \underset{\sim}{S}_{\xi\eta} \\ \underset{\sim}{S}'_{\xi\eta} & s_{\eta\eta} \end{bmatrix}$$

be the sample covariance matrix of $(\underset{\sim}{\xi}, \eta)$, then

$$\underset{\sim}{S} \overset{p}{\to} \begin{bmatrix} \underset{\sim}{\Sigma} + \underset{\sim}{\Sigma}_\delta & \underset{\sim}{\Sigma}\underset{\sim}{\beta} \\ \underset{\sim}{\beta}'\underset{\sim}{\Sigma} & \underset{\sim}{\beta}'\underset{\sim}{\Sigma}\underset{\sim}{\beta} + \sigma^2_\epsilon \end{bmatrix} \qquad \text{as } n \to \infty .$$

Sprent (1966) proposed a generalized least squares estimate of $\underset{\sim}{\beta}$, which is a function of $\sigma^2_\epsilon \underset{\sim}{\Sigma}^{-1}_\delta$, for functional relationship. It can be shown that it is consistent for structural relationship.

Replace the $\sigma_\epsilon^2 \Sigma_\delta^{-1}$ by a general positive definite matrix Λ and denote the new estimate by $\hat{\beta}(\Lambda)$. Let $\hat{\beta}$ be a consistent estimate of β.

To construct p functions similar to $h(\beta)$ of section 3 when $\sigma_\epsilon^2 \Sigma_\delta^{-1}$ is unknown, let

$$\ell(\beta) = s_{\eta\eta} - \beta' s_{\xi\eta}, \qquad \ell_j(\beta) = \beta_j^{-1}(\nu_j'\beta - s_{j\eta}), \tag{5.1}$$

where $s_{\xi\xi} = (\nu_1, \ldots, \nu_p)'$ and $s_{\xi\eta} = (s_{1\eta}, \ldots, s_{p\eta})'$. Also let H_j be the matrix whose (g,k) element is $\partial^2[\ell(\beta)/\ell_j(\beta)]/\partial\beta_g\partial\beta_k$ evaluated at $\hat{\beta}$. Then based on the same arguments as in section 3, we estimate

$\lambda_j = \sigma_\epsilon^2/\sigma_{\delta_j}^2$ by $\hat{\lambda}_j = \max\{0, \hat{\lambda}_j'\}$, $j = 1, \ldots, p$, where

$$\hat{\lambda}_j' = \infty \qquad \text{if } \ell_j(\hat{\beta}) \le 0 \quad \text{or } s_{jj} - \ell_j(\hat{\beta}) \le 0$$

$$= \ell(\hat{\beta})/\ell_j(\hat{\beta}) + \frac{1}{2}\,\mathrm{tr}(H_j \hat{A}V(\hat{\beta})) \qquad \text{otherwise,}$$

and $\hat{A}V(\hat{\beta})$ is a consistent estimate of the asymptotic covariance matrix of $\hat{\beta}$. $\hat{\beta}_R$ is then defined to be $\hat{\beta}(\hat{\Lambda})$.

(ξ, f)-s.e. and Efficiencies Relative to the OLS Estimate $\hat{\beta}_L$

ESTIMATE	(ξ_1, f_1) $\theta=0.25$ $\beta(\theta)=1.25$	(ξ_1, f_1) $\theta=0.5$ $\beta(\theta)=1.5$	(ξ_1, f_1) $\theta=0.75$ $\beta(\theta)=1.75$	(ξ_1, f_1) (ξ, f)-s.e.	(ξ_1, f_2) $\theta=0.25$ $\beta(\theta)=1.25$	(ξ_1, f_2) $\theta=0.5$ $\beta(\theta)=1.5$	(ξ_1, f_2) $\theta=0.75$ $\beta(\theta)=1.75$	(ξ_1, f_2) (ξ, f)-s.e.
	<u>n = 20</u>				<u>n = 20</u>			
$\hat{\beta}_L^a$	.142	.369	.691	.435	.120	.290	.639	.379
$\hat{\beta}_U$	.008	.002	.786	2.798	.033	.390	.340	1.923
$\hat{\beta}_G$	.007	.010	.002		.013	.000	.001	
$\hat{\beta}_{GP}$	.396	.779	1.340	.481	.510	.778	1.299	.360
$\hat{\beta}_{GR}$	.677	.972	1.497	.833	.806	1.000	1.558	.284
$\hat{\beta}_W$	.065	.123	.131	3.054	.064	.094	.116	3.339
$\hat{\beta}_{WP}$	.515	.904	1.548	.390	.638	.944	1.658	.284
$\hat{\beta}_{WR}$	.979	1.381	1.669	.297	1.029	1.412	1.776	.230
$\hat{\beta}_S$	.164	.385	.692	.947	.262	.018	.992	.653
$\hat{\beta}_{SP}$	.593	.973	1.495	.354	.620	1.055	1.420	.290
$\hat{\beta}_{SR}$	.619	1.020	1.538	.313	.984	1.190	1.439	.310
	<u>n = 50</u>				<u>n = 50</u>			
$\hat{\beta}_L^a$	.096	.302	.630	.344	.081	.261	.583	.337
$\hat{\beta}_U$	.100	.602	2.326	.615	.119	.658	3.247	.494
$\hat{\beta}_G$	.015	.014	.058	4.974	.078	.307	.127	3.504
$\hat{\beta}_{GP}$	.664	1.372	2.899	.175	.745	1.548	3.412	.146
$\hat{\beta}_{GR}$	.972	1.957	2.950	.137	1.167	2.008	3.774	.118
$\hat{\beta}_W$	.181	.257	.225	1.711	.299	.336	.253	1.500
$\hat{\beta}_{WP}$	.736	1.563	3.257	.174	.841	1.686	3.497	.138
$\hat{\beta}_{WR}$	1.157	2.160	3.289	.140	1.321	2.320	3.731	.122
$\hat{\beta}_S$	.528	1.497	2.740	.864	.791	1.672	2.967	.159
$\hat{\beta}_{SP}$	.828	1.575	3.125	.178	.938	1.595	3.021	.137
$\hat{\beta}_{SR}$	1.133	1.869	3.058	.157	1.330	1.890	3.077	.131

ESTIMATE	(Σ_2, f_1) $\theta=0.00$ $\beta(\theta)=5.00$	$\theta=0.50$ $\beta(\theta)=5.20$	$\theta=1.00$ $\beta(\theta)=5.4$	(Σ, f)-s.e.	(Σ_2, f_2) $\theta=0.00$ $\beta(\theta)=5.00$	$\theta=0.50$ $\beta(\theta)=5.2$	$\theta=1.00$ $\beta(\theta)=5.4$	(Σ, f)-s.e.
	n = 20				n = 20			
$\hat{\beta}_L^a$	.128	.179	.365	.219	.120	.161	.269	.177
$\hat{\beta}_U$	.289	.729	2.421	.308	.367	.745	1.951	.215
$\hat{\beta}_G$	.010	.014	.003	6.125	.000	.159	.068	3.102
$\hat{\beta}_{GP}$	.663	.918	1.681	.242	.775	.952	1.656	.162
$\hat{\beta}_{GR}$	.792	.981	1.634	.221	.940	1.047	1.616	.148
$\hat{\beta}_W$	.033	.062	.042	5.640	.099	.052	.146	2.878
$\hat{\beta}_{WP}$	.679	.867	1.712	.246	.776	.944	1.658	.161
$\hat{\beta}_{WR}$	.808	.996	1.618	.213	.942	1.035	1.615	.149
$\hat{\beta}_S$	.189	.232	.338	1.722	.315	.484	.734	.439
$\hat{\beta}_{SP}$	.619	.938	1.575	.241	.822	.961	1.650	.155
$\hat{\beta}_{SR}$	.792	1.009	1.536	.212	.957	1.044	1.582	.147
	n = 50				n = 50			
$\hat{\beta}_L^a$	.045	.092	.233	.110	.040	.081	.201	.088
$\hat{\beta}_U$	.168	.871	4.329	.130	.185	.900	3.891	.104
$\hat{\beta}_G$	.349	.705	1.337	.159	.524	.989	2.020	.084
$\hat{\beta}_{GP}$	.587	1.042	2.817	.092	.731	1.072	2.857	.066
$\hat{\beta}_{GR}$	.767	1.214	2.494	.083	.938	1.241	2.513	.058
$\hat{\beta}_W$	.333	.691	1.263	.170	.524	1.001	1.919	.085
$\hat{\beta}_{WP}$	.585	1.060	2.833	.092	.737	1.071	2.857	.066
$\hat{\beta}_{WR}$	.764	1.215	2.500	.083	.945	1.238	2.513	.058
$\hat{\beta}_S$	.380	.744	1.534	.131	.599	1.005	2.342	.079
$\hat{\beta}_{SP}$	.550	1.078	2.688	.091	.733	1.101	2.865	.067
$\hat{\beta}_{SR}$	.726	1.229	2.358	.083	.953	1.230	2.404	.061

[a] The rows corresponding to $\hat{\beta}_L$ give the MSE of $\hat{\beta}_L$.

REFERENCES

BIRCH, M. W. (1964). A note on the maximum likelihood estimation of a linear structural relationship. J. Amer. Statist. Assoc., 59, 1175-1178.

DORFF, M. and GURLAND, J. (1961a). Estimation of the parameters of a linear functional relationship. J. Roy. Statist. Soc., B, 23, 160-170.

DORFF, M. and GURLAND, J. (1961b). Small sample behaviour of slope estimators in a linear functional relation. Biometrics, 17, 283-298.

DRION, E. F. (1951). Estimation of the parameters of a straight line and of the variances of the variables, if they are both subject to error. Indag. Math., 13, 256-260.

FELDSTEIN, M. (1974). Errors in variables: a consistent estimator with smaller MSE in finite samples. J. Amer. Statist. Assoc., 69, 990-996.

GEARY, R. C. (1948). Studies in relations between economic time series. J. Roy. Statist. Soc., B, 10, 140-158.

HOGG, R. V. (1974). Adaptive robust procedures: a partial review and some suggestions for future applications and theory. J. Amer. Statist. Assoc., 69, 909-927.

KENDALL, M. G. and STUART, A. (1977, 1973). The Advanced Theory of Statistics, Vol. 1, 4th ed., Vol. 2, 3rd ed. London: Griffin.

LINDLEY, D. V. (1947). Regression lines and the linear functional relationship. J. Roy. Statist. Soc. Suppl., 9, 218-244.

MADANSKY, A. (1959). The fitting of straight lines when both variables are subject to error. J. Amer. Statist. Assoc., 54, 173-205.

MALINVAUD, E. (1970). *Statistical Methods of Econometrics*, 2nd ed. Amsterdam: North Holland.

MORAN, P. A. P. (1971). Estimating structural and functional relationships. *J. Mult. Anal.*, 1, 232-255.

REIERSØL, O. (1945). Confluence analysis by means of instrumental sets of variables. *Ark. Mat. Astron. Fys.*, 32, 1-119.

SCOTT, E. L. (1950). Note on consistent estimates of the linear structural relation between two variables. *Ann. Math. Statist.*, 21, 284-288.

WALD, A. (1940). The fitting of straight lines if both variables are subject to errors. *Ann. Math. Statist.*, 11, 284-300.

WOLFOWITZ, J. (1952). Consistent estimators of the parameters of a linear structural relation. *Skand. Aktuarietidskr.*, 35, 132-151.

CHAPTER 5

SOME GENERAL PROCEDURES OF ESTIMATING BOTH LINEAR STRUCTURAL AND FUNCTIONAL RELATIONSHIPS

1. INTRODUCTION

Consider two p and q dimensional variables $\underset{\sim}{x}$ and $\underset{\sim}{y}$ linearly related by $\underset{\sim}{y} = \underset{\sim}{\alpha} + B\underset{\sim}{x}$, where $\underset{\sim}{\alpha}$ and B are unknown q-component vector and q×p matrix, respectively. (To simplify notations, in this chapter matrices denoted by capital letters will not be underlined by "∿".) The variables $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are unobservable, and instead, we observe $\underset{\sim}{\eta}_1 = \underset{\sim}{x} + \underset{\sim}{\varepsilon}_1$ and $\underset{\sim}{\eta}_2 = \underset{\sim}{y} + \underset{\sim}{\varepsilon}_2$, where $(\underset{\sim}{\varepsilon}_1, \underset{\sim}{\varepsilon}_2)$ is distributed as $N(\underset{\sim}{0}, \Sigma)$. With n such $(\underset{\sim}{x}, \underset{\sim}{y})$ and the corresponding $(\underset{\sim}{\eta}_1, \underset{\sim}{\eta}_2)$, the model becomes

$$\begin{aligned} \underset{\sim}{y}_i &= \underset{\sim}{\alpha} + B\underset{\sim}{x}_i, \\ \underset{\sim}{\eta}_{i1} &= \underset{\sim}{x}_i + \underset{\sim}{\varepsilon}_{i1}, \\ \underset{\sim}{\eta}_{i2} &= \underset{\sim}{y}_i + \underset{\sim}{\varepsilon}_{i2}, \qquad i = 1, \ldots, n, \end{aligned} \tag{1.1}$$

where the vector of "errors" of observations $(\underset{\sim}{\varepsilon}_{i1}, \underset{\sim}{\varepsilon}_{i2})$ are i.i.d. as $N(\underset{\sim}{0}, \Sigma)$. $\Sigma = [\sigma_{ij}]$ may be or may not be diagonal. We want to estimate B based on the $(\underset{\sim}{\eta}_{i1}, \underset{\sim}{\eta}_{i2})$. The $\underset{\sim}{x}_i$ can either be constants or generated independently from a superpopulation. The relationship $\underset{\sim}{y}_i = \underset{\sim}{\alpha} + B\underset{\sim}{x}_i$ in (1.1) for the former is usually referred to as a functional relationship and for the latter as a structural relationship. In the structural relationship we assume that the $\underset{\sim}{x}_i$ are i.i.d. as $N(\underset{\sim}{\mu}, \Sigma_x)$, where $\underset{\sim}{\mu}$ and Σ_x are unknown, and are independent of the $(\underset{\sim}{\varepsilon}_{j1}, \underset{\sim}{\varepsilon}_{j2})$. The problem of estimating parameters in linear structural and functional relationships was comprehensively reviewed by Kendall and Stuart (1973, chapter 29) and Moran (1971).

For the structural relationship model of (1.1), the simplest case $p = q = 1$ and Σ is diagonal had been extensively studied in the literature. It is well known that $\beta = B$ (we write β for B when

B is 1 × 1) is unidentifiable if σ_{11} and σ_{22} are unknown (Reiersøl, 1950). Unidentifiability results from the fact that (n_1, n_2) has a bivariate normal distribution and is completely specified by its five first two moments which are determined by six unknown parameters. To avoid this difficulty, additional information is required. Two commonly studied cases are (i) both σ_{11} and σ_{22} are known, and (ii) σ_{22}/σ_{11} is known (which is equivalent to knowing $\sum$ to within a proportionality factor (t.w.p.f.), i.e., knowing the A in $\sum = cA$, where the (1,1) element of A is 1 and c is an unknown non-zero scalar (Lindley, 1947; Moran, 1971)). In the former case, the maximum likelihood estimate (MLE) of β is obtained from solving the five unknown parameters in the five equations formed by equating the first two sample moments of the (n_{i1}, n_{i2}) to their corresponding expected values. In the latter case this cannot be done since there are only four unknowns in five equations. This difficulty of "overidentification", as noted by Madansky (1959), had aroused considerable discussion and was resolved by Barnett (1967) and Birch (1964) who solved the likelihood equations and the algebra involved is quite complicated as indicated by Dolby (1976). It is interesting to note that the MLE of β in both cases (i) and (ii) are algebraically the same. The problem of finding the MLE of B for the general multidimensional structural relationship model (p and q are not both one) of (1.1) when $\sum$ (not necessarily diagonal) is known or known t.w.p.f. is more complicated and seems to have not been solved.

In the functional relationship model of (1.1), no particular difficulty arises in obtaining the MLE of B when $\sum$ is known or known t.w.p.f. and the solutions can be found in Kendall and

Stuart (1973, chapter 2.) and Sprent (1969). It should be noted that the MLE of B in the case Σ is known and the case Σ is known t.w.p.f. are algebraically the same.

For both structural and functional relationships, when Σ is not known t.w.p.f., consistent estimates can be obtained if for each $(\underset{\sim}{x}_i, \underset{\sim}{y}_i)$, s replicated observations are available. The model (1.1) then becomes

$$\begin{aligned} \underset{\sim}{y}_i &= \underset{\sim}{a} + B\underset{\sim}{x}_i, \\ \underset{\sim}{\eta}_{ij1} &= \underset{\sim}{x}_i + \underset{\sim}{\epsilon}_{ij1}, \\ \underset{\sim}{\eta}_{ij2} &= \underset{\sim}{y}_i + \underset{\sim}{\epsilon}_{ij2}, \qquad i = 1, \ldots, n; \quad j = 1, \ldots, s, \end{aligned} \tag{1.2}$$

where the $(\underset{\sim}{\epsilon}_{ij1}, \underset{\sim}{\epsilon}_{ij2})$ are i.i.d as $N(\underset{\sim}{0}, \Sigma)$. One procedure is to estimate Σ by $\sum_{i=1}^{n} (\underset{\sim}{\eta}_{ij1} - \bar{\underset{\sim}{\eta}}_{i\cdot 1}, \underset{\sim}{\eta}_{ij2} - \bar{\underset{\sim}{\eta}}_{i\cdot 2})' (\underset{\sim}{\eta}_{ij1} - \bar{\underset{\sim}{\eta}}_{i\cdot 1}, \underset{\sim}{\eta}_{ij2} - \bar{\underset{\sim}{\eta}}_{i\cdot 2})/n$, where $\bar{\underset{\sim}{\eta}}_{i\cdot k} = \sum_{j=1}^{s} \underset{\sim}{\eta}_{ijk}/s$, $k = 1, 2$, and then use the MLE of B assuming Σ is known. Another procedure is to solve the likelihood equations directly. In the functional relationship model of (1.2), this was studied by Anderson (1951), Barnett (1969), Dolby and Lipton (1972), and Villegas (1961). The computation is in general complicated and can only be solved by iterative method. The structural relationship model when $p = q = 1$ is studied in chapter 3. A different estimation procedure for model (1.1) under different assumptions was given by Robinson (1977).

In this chapter, we are mainly concerned with the estimation of $\underset{\sim}{\theta}_\beta$ in the following model:

$$\begin{aligned} \eta_i &= \alpha + B\xi_i + \varepsilon_i, \qquad i = 1,\ldots, n, \\ \alpha &= \alpha(\theta_\alpha), \\ B &= B(\theta_\beta), \end{aligned} \tag{1.3}$$

where η_i, α and ε_i are r-component vectors, ξ_i is a p-component vector and B is a $r \times p$ matrix. α and B are one to one differentiable functions of the parameter vectors θ_α and θ_β, respectively. Also the ε_i are i.i.d. as $N(0,\Sigma)$, where Σ is either known completely or a known one to one differentiable function of an unknown parameter vector θ_σ. Only the η_i are observable and the ξ_i are either unknown constants (functional relationship) or random vectors (structural relationship) which are independent of the ε_j and i.i.d. as $N(\mu,\Sigma_x)$ with unknown μ and Σ_x. The models (1.1) and (1.2) are particular cases of (1.3) which is quite similar to the factor analysis model. Jöreskog (1970) in an analysis of covariance structure considered a very general model which includes a wide range of models as particular cases of his. By imposing various specifications on the parametric structure of his general model, he specialized his model to the multivariate linear structural relationship when Σ is assumed to be diagonal; but he did not arrive at any explicit estimate except suggesting a numerical procedure which was outlined in his general model. By considering the less general but simpler model (1.3), the likelihood function can be expressed in a convenient form. From it we are able to prove, when specialized to the structural relationship model of (1.1) with $p = 1$, that the MLE of B in the case Σ is known and the case Σ is known t.w.p.f. are algebraically the same and are also identical to the MLE of B when we have the functional relationship. This is useful since we do not have to determine whether the $\underset{\sim}{x}_i$ should be taken as constants or i.i.d. random vectors in our design. Simpler methods of computing the MLE of B when p = 1 is given in section 4. In section 4, we relax the normality assumption on $\underset{\sim}{\varepsilon}_i$ in (1.3) and propose an estimate of $\underset{\sim}{\beta}_\beta$ in the functional relationship when Σ is known or known t.w.p.f. This estimate coincides with the MLE of $\underset{\sim}{\beta}_\beta$ under a normality assumption. The rationale of constructing the present estimate enables us to establish consistency under some mild assumptions on the asymptotic behaviour of the $\underset{\sim}{x}_i$. This implies that the MLE of $\underset{\sim}{\beta}_\beta$ under a normality assumption is also consistent, a result which is not always true in maximum likelihood estimation with infinitely many incidental parameters as pointed out by Neyman and Scott (1948). In section 4, we give methods of computing the proposed estimate which is a value maximizing certain quadratic form in the $\underset{\sim}{\eta}_i$.

2. MAXIMUM LIKELIHOOD ESTIMATION IN STRUCTURAL RELATIONSHIP

LEMMA 2.1. Let Σ_1, Σ_2 be non-singular matrices and $\underset{\sim}{B}$ be any matrix such that the matrix multiplications below are compatible; then

$$\begin{bmatrix} \Sigma_1 & \Sigma_1 B' \\ B\Sigma_1 & B\Sigma_1 B' + \Sigma_2 \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_1^{-1} + B'\Sigma_2^{-1}B & -B'\Sigma_2^{-1} \\ -\Sigma_2^{-1}B & \Sigma_2^{-1} \end{bmatrix} = P, \tag{2.1}$$

$$|B\Sigma_1 B' + \Sigma_2| = |\Sigma_1^{-1} + B'\Sigma_2^{-1}B|\,|\Sigma_1|\,|\Sigma_2| \,. \tag{2.2}$$

Proof: (2.1) can be proved by direct matrix multiplications. To Prove (2.2), we use the Binomial Inverse Theorem

$$(A + UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1} \quad (2.3)$$

to conclude that

$$(B\Sigma_1 B' + \Sigma_2)^{-1} = \Sigma_2^{-1} - \Sigma_2^{-1}B(\Sigma_1^{-1} + B'\Sigma_2^{-1}B)^{-1}B'\Sigma_2^{-1} . \quad (2.4)$$

Now from the theory of partitioned matrices, we have

$$\begin{aligned} |P| &= |\Sigma_1^{-1} + B'\Sigma_2^{-1}B| \, |\Sigma_2^{-1} - \Sigma_2^{-1}B(\Sigma_1^{-1} + B'\Sigma_2^{-1}B)^{-1}B'\Sigma_2^{-1}| \\ &= |\Sigma_1^{-1} + B'\Sigma_2^{-1}B| \, |B\Sigma_1 B' + \Sigma_2|^{-1} . \end{aligned}$$

But

$$P = \begin{bmatrix} I & -B' \\ \underset{\sim}{0} & I \end{bmatrix} \begin{bmatrix} \Sigma_1^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \Sigma_2^{-1} \end{bmatrix} \begin{bmatrix} I & \underset{\sim}{0} \\ -B & I \end{bmatrix}$$

and

$$\det\begin{bmatrix} I & -B' \\ \underset{\sim}{0} & I \end{bmatrix} = \det\begin{bmatrix} I & \underset{\sim}{0} \\ -B & I \end{bmatrix} = 1 .$$

Hence

$$|P| = \det\begin{bmatrix} \Sigma_1^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \Sigma_2^{-1} \end{bmatrix} = |\Sigma_1|^{-1}|\Sigma_2|^{-1}$$

and (2.2) is obtained.

It is interesting to observe that (2.4) can also be deduced from (2.1) and the theory of partitioned matrices without using (2.3).

Consider the model (1.3), since $\underset{\sim}{x}_i$ and $\underset{\sim}{\varepsilon}_i$ are assumed to be normally distributed, $\underset{\sim}{\eta}_i$, $i = 1,\ldots, n$, are i.i.d. as N $(\underset{\sim}{\alpha} + B\underset{\sim}{\mu}, B\Sigma_x B' + \Sigma)$. It follows from (2.2) and (2.4) that the log likelihood for model (1.3) is

$$\ln L = \text{constant} - \frac{n}{2}\ln|\Sigma_x| - \frac{n}{2}\ln|\Sigma| - \frac{n}{2}\ln|\Sigma_x^{-1} + B'\Sigma^{-1}B|$$

$$- \frac{1}{2}\sum_{i=1}^{n} (\underset{\sim}{\eta}_i - \underset{\sim}{\alpha} - B\underset{\sim}{\mu})'[\Sigma^{-1} - \Sigma^{-1}B(\Sigma_x^{-1} + B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}]$$

$$(\underset{\sim}{\eta}_i - \underset{\sim}{\alpha} - B\underset{\sim}{\mu}) . \tag{2.5}$$

To get the likelihood equations, we use the theory of differentiating a scalar value function of a matrix variable (cf. Dwyer, 1967). Following his notations we use $\langle X\rangle_{ij}$ to denote the $(i,j)^{th}$ element of the matrix X and if $y = y(X)$ is a scalar function of X, then we use $\partial y/\partial X$ to denote the matrix whose $(i,j)^{th}$ element is the partial derivative of y with respect to the $(i,j)^{th}$ element of X.

We further assume that $p \le r$ and rank $(B) = p$, so that $B'\Sigma^{-1}B$ is positive definite.

Differentiating (2.5) with respect to $\underset{\sim}{\mu}$, $\underset{\sim}{\theta}_\alpha = (\theta_{\alpha 1}, \ldots, \theta_{\alpha a})'$ and equating to zero, we have

$$B'F((\underset{\sim}{\alpha} - \frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{\eta}_i) + B\underset{\sim}{\mu}) = \underset{\sim}{0}, \tag{2.6}$$

$$\left[\frac{\partial \underset{\sim}{\alpha}}{\partial \theta_{\alpha k}}\right]' F((\underset{\sim}{\alpha} - \frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{\eta}_i) + B\underset{\sim}{\mu}) = \underset{\sim}{0}, \quad k = 1, \ldots, a, \tag{2.7}$$

where

$$F = \Sigma^{-1} - \Sigma^{-1}B(\Sigma_x^{-1} + B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1} . \tag{2.8}$$

ASSUMPTION I: Let $\tilde{\theta}_\alpha$ and $\tilde{\mu}$ be the MLE of θ_α and μ given that θ_β, θ_α and Σ_x are fixed. Then

$$c(\tilde{\theta}_\alpha) + B(\theta_\beta)\tilde{\mu} \tag{2.9}$$

is independent of θ_β, θ_σ and Σ_x.

Assumption I holds for models (1.1) and (1.2) in section 1. For example, in

$$\eta_i = c + Bx_i + \epsilon_i, \quad i = 1,\ldots, n,$$

where

$$\eta_i = \begin{bmatrix} n_{i11} \\ \vdots \\ n_{is1} \\ n_{i12} \\ \vdots \\ n_{is2} \end{bmatrix}, \quad c = \begin{pmatrix} 0 \\ \alpha 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ \beta 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11}I & \sigma_{12}I \\ \sigma_{12}I & \sigma_{22}I \end{pmatrix},$$

and 0 and 1 are s-component vectors and I is s×s, (i.e., the model (1.2) with p = q = 1 when it is expressed in the form of (1.3)), (2.9) is equal to

$$\begin{bmatrix} a_1 1 \\ a_2 1 \end{bmatrix},$$

where $a_1 = \sum_{i=1}^{n} \sum_{j=1}^{s} n_{ij1}/ns$, $a_2 = \sum_{i=1}^{n} \sum_{j=1}^{s} n_{ij2}/ns$.

Now consider the log likelihood $\ln L_1$ obtained by letting $c + B\mu = c(\tilde{\theta}_\alpha) + B(\theta_\beta)\tilde{\mu}$ in ln L. Under assumption I, to maximize ln L, we only have to maximize $\ln L_1$ with respect to θ_β, Σ_x, and θ_σ

if $\sum$ is unknown (cf. Richards, 1961). Also for simplicity we can write $\underset{\sim}{n}_i$ for $\underset{\sim}{n}_i - (\underset{\sim}{\alpha}(\hat{\underset{\sim}{\theta}}_\alpha) + B(\underset{\sim}{\theta}_\beta)\tilde{\underset{\sim}{x}})$ in the course of maximizing $\ln L_1$. Letting

$$C = B'\textstyle\sum^{-1}B, \quad D = (\textstyle\sum_x^{-1} + C)^{-1}, \quad \underset{\sim}{p}_i = B'\textstyle\sum^{-1}\underset{\sim}{n}_i, \quad S = \sum_{i=1}^{n} \underset{\sim}{n}_i\underset{\sim}{n}_i'/n,$$

the likelihood equations are given by

$$\underset{\sim}{0} = \frac{\partial \ln L_1}{\partial \sum_x} = -n\textstyle\sum_x^{-1} + n\sum_{u,v} \langle D\rangle_{uv}\textstyle\sum_x^{-1}K_{uv}\textstyle\sum_x^{-1} + \sum_{i=1}^{n}\sum_{u,v} \langle D\underset{\sim}{p}_i\underset{\sim}{p}_i'D\rangle_{uv}\textstyle\sum_x^{-1}K_{uv}\textstyle\sum_x^{-1}, \qquad (2.10)$$

$$0 = \frac{\partial \ln L_1}{\partial \theta_{\beta k}} = \mathrm{tr}\left(-nD\frac{\partial C}{\partial \theta_{\beta k}} - \sum_{i=1}^{n} D\underset{\sim}{p}_i\underset{\sim}{p}_i'D\frac{\partial C}{\partial \theta_{\beta k}}\right) + 2\sum_{i=1}^{n}\underset{\sim}{p}_i'D\frac{\partial p_i}{\partial \theta_{\beta k}}, \; k = 1,\ldots,b, \qquad (2.11)$$

where $\underset{\sim}{\theta}_\beta = (\theta_{\beta 1},\ldots,\theta_{\beta b})$, and K_{uv} is the matrix with the $(u,v)^{th}$ element 1 and all other elements 0. If $\sum$ is unknown, we have also the likelihood equations

$$0 = \frac{\partial \ln L_1}{\partial \theta_{\sigma k}} = \mathrm{tr}\left\{-\frac{n}{2}D\frac{\partial C}{\partial \theta_{\sigma k}} + \left[\frac{1}{2}\sum_{i=1}^{n}\textstyle\sum^{-1}\underset{\sim}{n}_i\underset{\sim}{n}_i'\textstyle\sum^{-1} + \sum_{i=1}^{n}\textstyle\sum^{-1}\underset{\sim}{n}_i\underset{\sim}{n}_i'\textstyle\sum^{-1}BDB'\textstyle\sum^{-1}\right.\right.$$

$$\left.\left. + \sum_{i=1}^{n}\textstyle\sum^{-1}BD\underset{\sim}{p}_i\underset{\sim}{p}_i'DB'\textstyle\sum^{-1}\right]\frac{\partial \sum}{\partial \theta_{\sigma k}}\right\}, \; k = 1,\ldots c, \qquad (2.12)$$

where $\underset{\sim}{\theta}_\sigma = (\theta_{\sigma 1},\ldots,\theta_{\sigma c})$.

Pre- and Post-multiplying (2.10) by $\sum_x$, we have

$$-n\textstyle\sum_x + nD + \sum_{i=1}^{n} D\underset{\sim}{p}_i\underset{\sim}{p}_i'D = 0, \qquad (2.13)$$

which implies that

$$\textstyle\sum_x = C^{-1}B'\textstyle\sum^{-1}S\textstyle\sum^{-1}BC^{-1} - C^{-1}. \qquad (2.14)$$

From (2.11) and (2.13), we have

$$\mathrm{tr}(-n\Sigma_x \frac{\partial C}{\partial \theta_{\beta k}}) + 2 \sum_{i=1}^{n} \underset{\sim}{p}_i' D \frac{\partial \underset{\sim}{p}_i}{\partial \theta_{\beta k}} = 0, \quad k = 1, \ldots, b. \tag{2.15}$$

Using (2.4), it can be shown that

$$D = C^{-1} - (B'\Sigma^{-1}S\Sigma^{-1}B)^{-1} . \tag{2.16}$$

Now (2.14), (2.15) and (2.16) together give

$$\mathrm{tr}\left\{[C^{-1} - (B'\Sigma^{-1}S\Sigma^{-1}B)^{-1}][-nB'\Sigma^{-1}S\Sigma^{-1}BC^{-1} \frac{\partial C}{\partial \theta_{\beta k}} + 2 \sum_{i=1}^{n} \frac{\partial \underset{\sim}{p}_i}{\partial \theta_{\beta k}} \underset{\sim}{p}_i']\right\}$$

$$= 0, \qquad k = 1, \ldots, b, \tag{2.17}$$

which does not involve Σ_x.

When $p = 1$, (2.17) reduces to

$$-nB'\Sigma^{-1}S\Sigma^{-1}BC^{-1} \frac{\partial C}{\partial \theta_{\beta k}} + 2 \sum_{i=1}^{n} \frac{\partial \underset{\sim}{p}_i}{\partial \theta_{\beta k}} \underset{\sim}{p}_i' = 0, \ k = 1, \ldots b. \tag{2.18}$$

Now (2.18) also holds with Σ replaced by A throughout. Thus we have proved that

<u>THEOREM 2.2</u>. Under assumption I, the MLE of $\underset{\sim}{\theta}_\beta$ for the structural model of (1.3) with $p = 1$ are algebraically the same in the case Σ is known completely and the case known t.w.p.f.

<u>EXAMPLE 2.1</u>. (2.18) can be used to find the MLE of $\beta = B$ in the structural relationship model of (1.1) with $p = q = 1$ and a general Σ when Σ is known or known t.w.p.f. By theorem 2.2, it suffices to find the MLE of β when A is known since this will also be the MLE of β when Σ is known. Now $\underset{\sim}{\alpha} = (0,\alpha)'$, $B = (1,\beta)'$, $x_i \sim N(\mu,\sigma^2)$ and

$$\Sigma = \sigma_{11}\begin{bmatrix} 1 & \sigma_{12}/\sigma_{11} \\ \sigma_{12}/\sigma_{11} & \sigma_{22}/\sigma_{11} \end{bmatrix} = \sigma_{11}\begin{bmatrix} 1 & a_{12} \\ a_{12} & a_{22} \end{bmatrix} = \sigma_{11}A$$

with A known. Since

$$\Sigma^{-1} = [\sigma_{11}(a_{22} - a_{12}^2)]^{-1}F, \quad F = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{12} & 1 \end{bmatrix}.$$

(2.18) then becomes

$$(1,\beta)FSF(1,\beta)'(0,1)F(1,\beta)' - (1,\beta)F(1,\beta)'(0,1)FSF(1,\beta)' = 0.$$

Solving for β, we have

$$\hat{\beta} = \{(a_{22}r_{22} - r_{11}) + [(a_{22}r_{22} - r_{11})^2 + 4(r_{12} + a_{12}r_{22})(a_{22}r_{12} + a_{12}r_{11})]^{\frac{1}{2}}\}/ [2(r_{12} + a_{12}r_{22})],$$

where

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} n_{i1}^2 & n_{i1}n_{i2} \\ n_{i1}n_{i2} & n_{i2}^2 \end{bmatrix},$$

$$FSF = \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix}.$$

Thus the same MLE of β is obtained whether we know only σ_{22}/σ_{11} and σ_{12}/σ_{11}, or all of σ_{11}, σ_{22} and σ_{12}. This is a generalization of the case when Σ is diagonal, in which the MLE of β when σ_{11}, σ_{22} are both known is only a function of σ_{22}/σ_{11} (cf. Kendall and Stuart, 1973, chapter 29).

For the multivariate generalization of example 2.1 with $p=1$, a method of finding the MLE of Σ is given in section 5.

3. MAXIMUM LIKELIHOOD ESTIMATION IN FUNCTIONAL RELATIONSHIP

In the present case, the $\underset{\sim}{x}_i$ are unknown constants, the log likelihood is

$$\ln L = \text{constant} - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{\eta}_i - \underset{\sim}{\alpha} - B\underset{\sim}{x}_i)'\Sigma^{-1}(\underset{\sim}{\eta}_i - \underset{\sim}{\alpha} - B\underset{\sim}{x}_i).$$

Hence

$$\frac{\partial \ln L}{\partial \theta_{\alpha k}} = \frac{\partial \underset{\sim}{\alpha}'}{\partial \theta_{\alpha k}}\Sigma^{-1}(\underset{\sim}{\alpha} + B\bar{\underset{\sim}{x}}. - \bar{\underset{\sim}{\eta}}.), \qquad k = 1, \ldots a,$$

where $\bar{\underset{\sim}{x}}. = \sum_{i=1}^{n} \underset{\sim}{x}_i/n$, $\bar{\underset{\sim}{\eta}}. = \sum_{i=1}^{n} \underset{\sim}{\eta}_i/n$.

<u>ASSUMPTION II</u>: Let $\tilde{\underset{\sim}{\theta}}_\alpha$ be the MLE of $\underset{\sim}{\theta}_\alpha$ given that the $\underset{\sim}{x}_i$, $\underset{\sim}{\theta}_\sigma$ and $\underset{\sim}{\theta}_\beta$ are fixed. Then there exist $\underset{\sim}{f}$, $\underset{\sim}{g}$ independent of the $\underset{\sim}{x}_i$, $\underset{\sim}{\theta}_\sigma$ and $\underset{\sim}{\theta}_\beta$ and dependent only on $\underset{\sim}{\eta}_1, \ldots, \underset{\sim}{\eta}_n$ such that

$$\underset{\sim}{\alpha}(\tilde{\underset{\sim}{\theta}}_\alpha) + B(\underset{\sim}{\theta}_\beta)\underset{\sim}{f} = \underset{\sim}{g}. \qquad (3.1)$$

Assumption II holds also for model (1.1) and (1.2) in section 1. Now consider the log likelihood ln L_1 obtained from replacing $\underset{\sim}{\alpha}$ by $\underset{\sim}{\alpha}(\underset{\sim}{\gamma}_\alpha)$ in ln L. Under assumption II, to maximize ln L we only have to maximize ln L_1 with respect to $\underset{\sim}{x}_i$, $\underset{\sim}{\theta}_\beta$ and also $\underset{\sim}{\theta}_\alpha$ if Σ is unknown. Also for simplificity we can write $\underset{\sim}{\eta}_i$ for $\underset{\sim}{\eta}_i - \underset{\sim}{g}$ and $\underset{\sim}{x}_i$ for $\underset{\sim}{x}_i - \underset{\sim}{f}$. ln L_1 becomes

$$\text{constant} - \frac{n}{2} \ln |\Sigma| - \frac{1}{2} \sum_{i=1}^{n} (\underset{\sim}{\eta}_i - B\underset{\sim}{x}_i)' \Sigma^{-1} (\underset{\sim}{\eta}_i - B\underset{\sim}{x}_i).$$

Equating $\partial \ln L_1/\partial \underset{\sim}{x}_i$ to zero, we have

$$B'\Sigma^{-1}(B\underset{\sim}{x}_i - \underset{\sim}{\eta}_i) = \underset{\sim}{0}, \quad i = 1,\ldots, n,$$

so that $\underset{\sim}{x}_i = C^{-1}\underset{\sim}{p}_i$ (C and $\underset{\sim}{p}_i$ are defined in section 2). So from

$$0 = \frac{\partial \ln L_1}{\partial \theta_{\beta k}} = 2 \sum_{i=1}^{n} \underset{\sim}{x}_i' \frac{\partial \underset{\sim}{p}_i}{\partial \theta_{\beta k}} - \sum_{i=1}^{n} \underset{\sim}{x}_i' \frac{\partial C}{\partial \theta_{\beta k}} \underset{\sim}{x}_i, \quad k = 1,\ldots, b,$$

we obtain

$$-\sum_{i=1}^{n} \underset{\sim}{p}_i' C^{-1} \frac{\partial C}{\partial \theta_{\beta k}} C^{-1} \underset{\sim}{p}_i + 2\sum_{i=1}^{n} \underset{\sim}{p}_i' C^{-1} \frac{\partial \underset{\sim}{p}_i}{\partial \theta_{\beta k}} = 0,$$

or

$$\operatorname{tr}\left\{C^{-1}[-nB'\Sigma^{-1}S\Sigma^{-1}BC^{-1} \frac{\partial C}{\partial \theta_{\beta k}} + 2\sum_{i=1}^{n} \frac{\partial \underset{\sim}{p}_i}{\partial \theta_{\beta k}} \underset{\sim}{p}_i']\right\} = 0. \qquad (3.2)$$

Comparing (2.17) with (3.2), we see that when $p > 1$ and $\sum$ is known or known t.w.p.f., the MLE of θ_β for structural and functional relationships in (1.3) are in general different. However, when $p = 1$ we have

<u>THEOREM 3.1</u>. Consider $p = 1$ in model (1.3). Suppose assumption I holds when (1.3) is structural, and assumption II holds when (1.3) is functional with g of (3.1) are algebraically being the $\alpha(\tilde{\theta}_\alpha) + B(\theta_\beta)\tilde{\mu}$ of (2.9). Then the MLE of θ_β are algebraically the same whether (1.3) is taken to be structural or functional, and $\sum$ is known completely or known t.w.p.f.

In particular, theorem 3.1 holds for model (1.1) with $p = 1$, which is the most commonly studied model.

4. ESTIMATION WITHOUT NORMALITY ASSUMPTION ON THE ERRORS

We shall restrict ourselves to the functional relationship model of (1.3) and assume that $\alpha = 0$. Suppose $\sum$ is known or known t.w.p.f. Let θ_β° be the true parameter vector, and B_0 be $B(\theta_\beta^\circ)$. For any i, the expected value of the squre of the length (the norm) of the vector $(B'\sum^{-1}B)^{-1/2}B'\sum^{-1}\eta_i$ is

$$E(\underset{\sim}{n}_i'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\underset{\sim}{n}_i) = \operatorname{tr}[B'\Sigma^{-1}E(\underset{\sim}{n}_i\underset{\sim}{n}_i')\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}]$$

$$= \operatorname{tr}(B'\Sigma^{-1}(\Sigma + B_0\underset{\sim}{x}_i\underset{\sim}{x}_i'B_0')\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}) = p + \phi_i(\underset{\sim}{\theta}_B),$$

where $\phi_i(\underset{\sim}{\theta}_\beta) = \underset{\sim}{x}_i'B_0'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}B_0\underset{\sim}{x}_i$.

The following lemma motivates the construction of our estimate.

LEMMA 4.1. $\underset{\sim}{x}_i'B_0'\Sigma^{-1}B_0\underset{\sim}{x}_i = \phi_i(\underset{\sim}{\theta}_\beta^o) \geq \phi_i(\underset{\sim}{\theta}_\beta)$.

Proof: Let

$$Q = \begin{bmatrix} B_0'\Sigma^{-1}B_0 & B_0'\Sigma^{-1}B \\ B'\Sigma^{-1}B_0 & B'\Sigma^{-1}B \end{bmatrix}.$$

Then, for any $\underset{\sim}{y} = (\underset{\sim}{y}_1', \underset{\sim}{y}_2')' \in R^{2p}$, we have

$$\underset{\sim}{y}'Q\underset{\sim}{y} = (\underset{\sim}{y}_1'B_0' + \underset{\sim}{y}_2'B')\Sigma^{-1}(B_0\underset{\sim}{y}_1 + B\underset{\sim}{y}_2) \geq 0 .$$

By setting

$$\underset{\sim}{y} = \begin{bmatrix} I \\ -B_0'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1} \end{bmatrix} \underset{\sim}{x}_i$$

in the above inequality, the lemma is proved.

LEMMA 4.2. Let Σ be any $r \times r$ positive definite matrix, and B and A be two $r \times p$ matrices of the form

$$B = \begin{pmatrix} I \\ R_B \end{pmatrix}, \qquad A = \begin{pmatrix} I \\ R_A \end{pmatrix},$$

where I is the $p \times p$ identity matrix. Then

$$B'\Sigma^{-1}A(A'\Sigma^{-1}A)^{-1}A'\Sigma^{-1}B = B'\Sigma^{-1}B$$

if and only if A = B.

Proof: The "if" part is obvious.

To prove the "only if" part, consider again

$$Q = \begin{bmatrix} B'\Sigma^{-1}B & B'\Sigma^{-1}A \\ A'\Sigma^{-1}B & A'\Sigma^{-1}A \end{bmatrix}.$$

For any $\underset{\sim}{x}_1 \in R^p$, let $\underset{\sim}{x}_2 = -(A'\Sigma^{-1}A)^{-1}(A'\Sigma^{-1}B)\underset{\sim}{x}_1 \in R^p$. Hence $A'\Sigma^{-1}B\underset{\sim}{x}_1 + A'\Sigma^{-1}A\underset{\sim}{x}_2 = \underset{\sim}{0}$. Now

$$Q\begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix} = \begin{bmatrix} I & -(B'\Sigma^{-1}A)(A'\Sigma^{-1}A)^{-1} \\ \underset{\sim}{0} & I \end{bmatrix}^{-1} \begin{bmatrix} I & -(B'\Sigma^{-1}A)(A'\Sigma^{-1}A)^{-1} \\ \underset{\sim}{0} & I \end{bmatrix} Q \begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix}$$

$$= \begin{bmatrix} I & -(B'\Sigma^{-1}A)(A'\Sigma^{-1}A)^{-1} \\ \underset{\sim}{0} & I \end{bmatrix}^{-1} \begin{bmatrix} \underset{\sim}{0} & \underset{\sim}{0} \\ A'\Sigma^{-1}B & A'\Sigma^{-1}A \end{bmatrix} \begin{pmatrix} \underset{\sim}{x}_1 \\ \underset{\sim}{x}_2 \end{pmatrix} = \underset{\sim}{0}.$$

So $$(\chi_1', \chi_2') Q \begin{bmatrix} \chi_1 \\ \chi_2 \end{bmatrix} = 0,$$ which implies that

$$(\chi_1' B' + \chi_2' A') \Sigma^{-1} (B\chi_1 + A\chi_2) = (\chi_1', \chi_2') Q \begin{bmatrix} \chi_1 \\ \chi_2 \end{bmatrix} = 0.$$

We therefore must have $\chi_1' B' + \chi_2' A' = 0$. From the form of B,A, we have

$$\chi_2 = -\chi_1, \quad (R_B - R_A)\,\chi_1 = 0 \;.$$

Since χ_1 can be arbitrarily chosen, $R_B = R_A$.

Lemma 4.1 says that the expected value of the norm of $(B'\Sigma^{-1}B)^{-\frac{1}{2}} B'\Sigma^{-1} \eta_i$ is maximized when $\theta_\beta = \theta_\beta^\circ$. This suggests that a possible consistent estimate of θ_β° would be the value of θ_β^* which maximizes the quadratic form

$$\frac{1}{n} \sum_{i=1}^{n} \eta_i' \Sigma^{-1} B (B'\Sigma^{-1}B)^{-1} B' \Sigma^{-1} \eta_i \;. \tag{4.1}$$

The following theorem gives conditions under which θ_β^* is consistent.

THEOREM 4.3. Suppose the following neighbourhood N of θ_β° exists. Given any $\delta > 0$ with $S_\delta = \{\theta_\beta : \|\theta_\beta - \theta_\beta^\circ\| = \delta\} \subseteq N$, there exists a $k_\delta > 0$ such that for every $\theta_\beta \in S_\delta$

$$\overline{\lim}\left[\frac{1}{n}\sum_{i=1}^{n}\phi_i(\theta_\beta)\Big/\frac{1}{n}\sum_{i=1}^{n}\phi_i(\theta_\beta^\circ)\right] < 1 - k_\delta \ . \tag{4.2}$$

Assume also that

$$0 < \underline{\lim}\ \frac{1}{n}\sum_{i=1}^{n}\phi_i(\theta_\beta^\circ) \ . \tag{4.3}$$

Then $\theta_\beta^* \xrightarrow{p} \theta_\beta^\circ$ as $n \to \infty$.

<u>Proof</u>: Let $\epsilon > 0$ and $\gamma > 0$ be given. Choose δ, $0 < \delta < \epsilon$, such that $S_\delta \subseteq N$. Then by (4.2) and (4.3) we can find $n_1 > 0$ such that for $n > n_1$, $\theta_\beta \in S_\delta$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\phi_i(\theta_\beta)\Big/\frac{1}{n}\sum_{i=1}^{n}\phi_i(\theta_\beta^\circ) < 1 - k_\delta' \tag{4.4a}$$

$$k_\delta' < \frac{1}{n}\sum_{i=1}^{n}[\phi_i(\theta_\beta^\circ) - \phi_i(\theta_\beta)] \ , \tag{4.4b}$$

where $1 > k_\delta' > 0$. Now, for any θ_β

$$\psi_n(\theta_\beta) = \frac{1}{n}\sum_{i=1}^{n}\eta_i'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\eta_i$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i'B_0' + \epsilon_i')\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}(B_0x_i + \epsilon_i)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\phi_i(\theta_\beta) + \frac{2}{n}\sum_{i=1}^{n}x_i'B_0'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\epsilon_i$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\epsilon_i'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\epsilon_i \ .$$

Let $d_n = \frac{1}{n}\sum_{i=1}^{n} [\phi_i(\theta^\circ_\beta) - \phi_i(\theta_\beta)]$. We therefore have

$$\psi_n(\theta^\circ_\beta) - \psi_n(\theta_\beta) = d_n[1 - d_n^{-\frac{1}{2}}(\frac{2}{n}\sum_{i=1}^{n} x_i'B_0'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\epsilon_i/d_n^{\frac{1}{2}})$$

$$+ d_n^{-\frac{1}{2}}(\frac{2}{n}\sum_{i=1}^{n} x_i'B_0'\Sigma^{-1}\epsilon_i/d_n^{\frac{1}{2}})] + \gamma_n, \qquad (4.5)$$

with $\gamma_n \overset{p}{\to} 0$ uniformly on S_δ because of the compactness of S_δ. Now

$$\text{Var}\left\{\frac{1}{n}\sum_{i=1}^{n} x_i'B_0'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\epsilon_i/d_n^{\frac{1}{2}}\right\}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} x_i'B_0'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}B_0x_i/d_n$$

$$\leq \frac{1}{n^2}\sum_{i=1}^{n} \phi_i(\theta^\circ_\beta)/[\frac{1}{n}\sum_{i=1}^{n} \phi_i(\theta^\circ_\beta) - \frac{1}{n}\sum_{i=1}^{n} \phi_i(\theta_\beta)]$$

$$= \frac{1}{n}[1 - \frac{1}{n}\sum_{i=1}^{n} \phi_i(\theta_\beta)/\frac{1}{n}\sum_{i=1}^{n} \phi_i(\theta^\circ_\beta)]^{-1} \leq \frac{1}{n}\frac{1}{k'_\delta} \to 0, \text{ as } n \to \infty.$$

The first inequality follows from Lemma 4.1 and the last from (4.4a). A similar argument shows that

$$\text{Var}[\sum_{i=1}^{n} x_i'B_0'\Sigma^{-1}\epsilon_i/(nd_n^{-\frac{1}{2}})] \leq \frac{1}{n}\frac{1}{k'_\delta} \to 0 \text{ as } n \to \infty.$$

So the bracketed term of (4.5) converges in probability uniformly on S_δ to 1. This together with the fact that $\gamma_n \to 0$ in probability uniformly on S_δ and (4.4b) show that one can find $n_2 > 0$ such that for $n > n_2$

$$\Pr(\psi_n(\underset{\sim}{\theta}^\circ_\beta) - \psi_n(\underset{\sim}{\theta}_\beta) > 0 \quad \text{for all } \underset{\sim}{\theta}_\beta \in S_\delta) > 1 - \gamma .$$

If the event inside the bracket is true, we have $\psi_n(\underset{\sim}{\theta}^\circ_\beta) > \psi_n(\underset{\sim}{\theta}_\beta)$ for all $\underset{\sim}{\theta}_\beta \in S_\delta$. This implies that a local maximum $\underset{\sim}{\theta}^*_\beta$ of ψ_n in $\{\underset{\sim}{\theta}_\beta: \|\underset{\sim}{\theta}_\beta - \underset{\sim}{\theta}^\circ_\beta\| \le \delta\}$ (exists by compactness) must be in the interior, i.e., satisfying $\|\underset{\sim}{\theta}^*_\beta - \underset{\sim}{\theta}^\circ_\beta\| < \delta$, completing the proof of the theorem.

Sometimes it is easier to check the consistency through the following theorem.

THEOREM 4.4. $\underset{\sim}{\theta}^*_\beta$ is consistent if B is of the form in Lemma 4.2 and

$$0 < \underline{\lim}\, \frac{1}{n} \sum_{i=1}^{n} x_{ij}^2 , \quad j = 1 \ldots p, \tag{4.6}$$

$$\overline{\lim}\, \frac{1}{n} \sum_{i=1}^{n} x_{ij}^2 < \infty, \quad j = 1,\ldots, p , \tag{4.7}$$

where $\underset{\sim}{x}_i = (x_{i1},\ldots, x_{ip})'$.

In particular if $S_n = \sum_{i=1}^{n} \underset{\sim}{x}_i \underset{\sim}{x}_i'/n$ converges to a non-singular matrix, then $\underset{\sim}{\theta}^*_\beta$ is consistent.

Proof Consider

$$\frac{1}{n} \sum_{i=1}^{n} \phi_i(\underset{\sim}{\theta}^\circ_\beta) = \frac{1}{n} \sum_{i=1}^{n} \underset{\sim}{x}_i' B_0' \Sigma^{-1} B_0 \underset{\sim}{x}_i$$

$$= \mathrm{tr}(B_0' \Sigma^{-1} B_0 S_n) = \mathrm{tr}(S_n^{\frac{1}{2}} B_0' \Sigma^{-1} B_0 S_n^{\frac{1}{2}}) . \tag{4.8}$$

$\underline{\lim} \frac{1}{n} \sum_{i=1}^{n} \phi_i(\theta^\circ_\beta) = 0$ would therefore contradict (4.6). Hence $\underline{\lim} \frac{1}{n} \sum_{i=1}^{n} \phi_i(\theta^\circ_\beta) > 0$. Furthermore, if $\overline{\lim} \frac{1}{n} \sum_{i=1}^{n} \phi_i(\theta^\circ_\beta) = \infty$, the second equality in (4.8) would imply that for at least one j, $\overline{\lim} \frac{1}{n} \sum_{i=1}^{n} x^2_{ij} = \infty$ which contradicts (4.7). Hence $\overline{\lim} \frac{1}{n} \sum_{i=1}^{n} \phi_i(\theta^\circ_\beta) < \infty$. By lemma 4.2 we have

$$B_0'\Sigma^{-1}B_0 - B_0'\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}B_0 = G > 0 .$$

Thus a similar argument as in (4.8) shows that

$$\underline{\lim} \frac{1}{n} \sum_{i=1}^{n} [\phi_i(\theta^\circ_\beta) - \phi_i(\theta_\beta)] > 0 . \qquad (4.9)$$

It can be shown that (4.9) is a continuous function of θ_β. Hence by the compactness of S_δ, there exists a $\ell_\delta > 0$ such that

$$\underline{\lim} \frac{1}{n} \sum_{i=1}^{n} [\phi_i(\theta^\circ_\beta) - \phi_i(\theta_\beta)] > \ell_\delta$$

for all $\theta_\beta \in \bar{S}_\delta$. This fact together with $\underline{\lim} \frac{1}{n} \sum_{i=1}^{n} \phi_i(\theta^\circ_\beta) > 0$, $\overline{\lim} \frac{1}{n} \sum_{i=1}^{n} \phi_i(\theta^\circ_\beta) < \infty$ implies that (4.2) and (4.3) are satisfied. The theorem then follows from theorem 4.3.

By differentiating (4.1) with respect to the $\theta_{\beta k}$ and equating to zero, we obtain equation (3.2). Thus under a normality assumption and with the A of $\Sigma = cA$ known, θ^*_β and the MLE coincide, and the conditions for consistency established here can be applied to the MLE. The consistency of the MLE of the functional relationship model of (1.1) in section I has not been thoroughly discussed in the literature, especially in the multi-variate cases (Lindley (1947), Kendall and Stuart (1973, Chapter 29) had sketched a proof of the consistency of the MLE of β for the case $p = q = 1$ and when Σ is diagonal.) For a multivariate model, the MLE does not have a closed form; the idea of their proof may not be applied without complications. Note that here we do not assume that S_n converges. Instead we assume that asymptotically the x_i should not be too close or too spread out. Moran (1971) pointed out that the conditions for consistency depend on the asymptotic behaviours of the x_i.

To maximize (4.1), we can differentiate the expression with respect to θ_β, equate it to zero and solve for θ_β. This in principle can be carried out by iterative methods, but the computation may be laborous. Here we give methods of maximizing (4.1) for some special forms of B.

Method 1: Suppose B is of the form

$$\begin{pmatrix} I \\ R \end{pmatrix},$$

where R is an unknown $(r - p) \times p$ matrix to be estimated. First, we shall maximize $tr(F'\Sigma^{-1}S_n\Sigma^{-1}F)$ subject to the condition that $F'\Sigma^{-1}F = I$, where F is a $r \times p$ matrix. The maximum value is $\sum_{i=1}^{p} \lambda_k$ and is attained when $F = \Sigma^{\frac{1}{2}}P$, where $P'P = I$, $P = (\underset{\sim}{p}_1, \ldots, \underset{\sim}{p}_p)$, and the $\underset{\sim}{p}_i$ are vectors satisfying

$$(\Sigma^{-\frac{1}{2}} S_n \Sigma^{-\frac{1}{2}} - \lambda_i I)\underset{\sim}{p}_i = \underset{\sim}{0}, \quad i = 1, \ldots, p, \tag{4.10}$$

λ_i is the ith largest root of $|S_n - \lambda\Sigma| = 0$ (Rao, 1973, p. 51). Now let

$$\Sigma^{\frac{1}{2}}P = \begin{pmatrix} F_1 \\ \\ F_2 \end{pmatrix},$$

where F_1 is $p \times p$. Then $\hat{R} = F_2F_1^{-1}$ maximizes (4.1) provided that F_1 is non-singular. To prove this, let

$$\hat{B} = \begin{pmatrix} I \\ \\ \hat{R} \end{pmatrix}.$$

Then

$$F_1'\hat{B}'\Sigma^{-1}\hat{B}F_1 = (\Sigma^{\frac{1}{2}}P)'\Sigma^{-1}(\Sigma^{\frac{1}{2}}P) = I$$

implies

$$\hat{B}'\Sigma^{-1}\hat{B} = F_1'^{-1}F_1^{-1}.$$

When $B = \hat{B}$, (4.1) reduces to

$$tr(F'\Sigma^{-1}S\Sigma^{-1}F) \geq tr(C'\Sigma^{-1}S\Sigma^{-1}C)$$

for C satisfying $C'\Sigma^{-1}C = I$. Now for any B, there exists a non-singular K such that

$$K'B'\Sigma^{-1}BK = I .$$

Let $C = BK$, then $C'\Sigma^{-1}C = I$. It is easily seen that (4.1) is

$$tr(\Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}S) = tr(C'\Sigma^{-1}S\Sigma^{-1}C) ,$$

which completes the proof of our assertion. It is easily seen that to find $\hat{R}$ it is only necessary to have P satisfying (4.10) without the orthogonality condition $P'P = I$.

Method 2: An alternative way of obtaining $\hat{R}$ is by using the equivalence of $\hat{\beta}_{\beta}$ to the MLE and the result of section 9 of Anderson (1976) (see also Geary (1948), Sprent (1969, p. 91)). Let $\omega_1, \ldots, \omega_{r-p}$ be vectors satisfying

$$(S_n - \lambda_j\Sigma)\omega_j = 0 , \qquad j = 1, \ldots, r-p,$$

where λ_j is the jth smallest root of

$$|S_n - \lambda\Sigma| = 0 .$$

Let $\Omega = [\omega_1, \ldots, \omega_{r-p}]'$. Partition Ω as $[\Omega_1, \Omega_2]$, where Ω_1 has p columns. Then $\hat{R} = -\Omega_2^{-1}\Omega_1$.

When Σ is a diagonal matrix and p is less than r − p, method 1 is less laborious. Otherwise method 2 is superior.

Method 3: Let λ_1 be the largest root of $|S_n - \lambda\Sigma| = 0$. If θ^*_β satisfies

$$B'(\theta_\beta)\, \Sigma^{-1}\, (S_n - \lambda_1\Sigma)\, \Sigma^{-1}\, B(\theta_\beta) = 0 \,, \tag{4.11}$$

then it maximizes (4.1). To see this let

$$\frac{1}{n}\sum_{i=1}^{n} x_i'\, \Sigma^{-1}B(B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}x_i = \lambda$$

which is equivalent to

$$\mathrm{tr}\{(B'\Sigma^{-1}B)^{-1}[B'\Sigma^{-1}S_n\Sigma^{-1}B - \frac{\lambda}{p}\,(B'\Sigma^{-1}B)]\} = 0 \,. \tag{4.12}$$

If $\frac{\lambda}{p} > \lambda_1$ then

$$B'\Sigma^{-1}S_n\Sigma^{-1}B - \frac{\lambda}{p}\,(B'\Sigma^{-1}B) = B'\Sigma^{-1}(S_n - \frac{\lambda}{p}\Sigma)\Sigma^{-1}B$$

would be negative definite, which implies that the left-hand side of (4.12) is negative. Hence $\lambda_1 p \geq \lambda$ and (4.1) has maximum value $\lambda_1 p$. This value is actually attained at θ^*_β since (4.12) is satisfied when $\lambda = \lambda_1 p$, by (4.11).

In the special case when B is of the form

$$\begin{pmatrix} 1 \\ R_B \end{pmatrix} ,$$

where R_B is $(r - 1) \times 1$, method 3 is less laborous than method 1. Thus method 3 should be applied to compute the MLE of B in the model (1.1) with $p = 1$ and $\sum$ is known or known t.w.p.f.

REFERENCES

ANDERSON, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann. Math. Statist., 22, 327-351.

ANDERSON, T. W. (1976). Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics. J. Roy. Statist. Soc., Ser. B, 38, 1-36.

BARNETT, V. D. (1967). A note on linear structural relationships when both residual variances are known. Biometrika, 54, 670-672.

BARNETT, V. D. (1970). Fitting straight lines—the linear functional relationship with replicated observations. Appl. Statist., 19, 135-144.

BIRCH, M. W. (1964). A note on maximum likelihood estimation of a linear structural relationship. J. Amer. Statist. Assoc., 59, 1175-1178.

DOLBY, G. R. (1976). The ultrastructural relation: A synthesis of the functional and structural relations. Biometrika, 63, 39-50

DOLBY, G. R. and LIPTON, S. (1972). Maximum likelihood estimation of a general non-linear functional relationship with replicated observations and correlated errors. Biometrika, 59, 121-129.

DORFF, M. and GURLAND, J. (1961a). Estimation of the parameters of a linear functional relationship, J. Roy. Statist. Soc., Ser. B, 23, 160-170.

DORFF, M. and GURLAND, J. (1961b). Small sample behaviour of slope estimators in a linear functional relation. Biometrics, 17, 283-298.

DWYER, P. S. (1967). Some applications of matrix derivatives in multivariate analysis. J. Amer. Statist. Assoc., 62, 607-625.

GEARY, R. C. (1948). Studies in relations between economic time series. J. Roy. Statist. Soc., Ser. B, 10, 140-158.

JÖRESKOG, K. G. (1970). A general method for analysis of covariance structures. Biometrika, 57, 239-251.

KENDALL, M. G. and STUART, A. (1973). The Advanced Theory of Statistics, Vol. 2, 3rd edn., London: Griffin.

LINDLEY, D. V. (1947). Regression lines and the linear functional relationship. J. Roy. Statist. Soc. Suppl., 9, 218-244.

MORAN, P. A. P. (1971). Estimating structural and functional relationships. J. Mult. Anal., 1, 232-255.

NEYMAN, J. and SCOTT, E. (1948). Consistent estimates based on partially consistent observations. Econometrica, 16, 1-32.

RAO, C. R. (1973). Linear Statistical Inference and its Applications, 2nd edn. New York: Wiley.

RICHARDS, F. S. G. (1961). A method of maximum likelihood estimation. J. Roy. Statist. Soc., B, 23, 469-475.

REIERSØL, O. (1950). Identifiability of a linear relation between variables which are subject to error. Econometrica, 18, 375-389.

ROBINSON, P. M. (1977). The estimation of a multivariate linear relation. J. Mult. Anal., 7, 409-423.

SOLARI, M. E. (1969). The maximum likelihood solution of the problem of estimating a linear functional relationship. J. Roy. Statist. Soc., Ser., B, 31, 372-375.

SPRENT, P. (1969). Models in Regression and Related Topics. London: Methuen.

VILLEGAS, C. (1961). Maximum likelihood estimation of a linear functional relationship. Ann. Math. Statist., 32, 1048-1062.

CHAPTER 6

MAXIMUM LIKELIHOOD ESTIMATION IN THE PRESENCE OF INCIDENTAL PARAMETERS

INTRODUCTION.

Regularity conditions under which the maximum likelihood estimate (MLE) is consistent and asymptotically normal when the observations are assumed to be independent and identically distributed (i.i.d.) have been extensively studied in the literature. A concise paper was written by Kulldorff (1957) and a review was written by Norden (1972, 1973). However in many practical situations, the basic assumption that the observations are i.i.d. does not hold. In a fundamental paper, Neyman and Scott (1948) considered the following more general problem of parametric estimation: Let $\{\underset{\sim}{x}_i\}_{i=1}^{\infty}$ be an infinite sequence of independent random vectors (they can have different dimensions). For each i, $\underset{\sim}{x}_i$ has a p.d.f. $f_i(\underset{\sim}{x}_i;\underset{\sim}{\theta},\underset{\sim}{\tau}_i)$, where $\underset{\sim}{\theta}$ and $\underset{\sim}{\tau}_i$ are p and r component vectors of parameters, respectively. $\underset{\sim}{\theta}$ is the same for each of i and is called a structural parameter, and the $\underset{\sim}{\tau}_i$ each of which appears only once in f_i are called incidental parameters. Here we are mainly concerned with estimating the structural parameter $\underset{\sim}{\theta}$ consistently. The well known problem of estimating linear functional relationship, also known as errors-in-variables estimation in econometrics, provides an important example belonging to this kind of estimation problem (see section 3). In the absence of incidental parameters, the usual properties of the MLE in the i.i.d. case, namely consistency, asymptotic normality and efficiency, have fairly natural generalizations to non-identically distributed random vectors. Regularity conditions under which incidental parameters is asymptotically normal. It is seen that the probability limit of the MLE of $\underset{\sim}{\theta}$ is not necessarily equal to the true parameter. However, it is shown in section 2. that in some situations a consistent estimate of $\underset{\sim}{\theta}$, which is a function of the MLE, can be constructed. In section 3, the results are applied to the estimation of linear functional relationship. This is different from the usual approaches which rely on the explicit form (when it exists) of the MLE (Barnett 1969, 1970; Patefield 1977). Consequently, we are able to give mild conditions under which the MLE of the intercept and slope parameters in a linear functional relationship are consistent and asymptotically normal. Our discussion is closely related to one of the problems raised by Moran (1971) in the conclusion of his review paper.

2. ASYMPTOTIC DISTRIBUTION

Let $\{\underset{\sim}{x}_i\}_{i=1}^{\infty}$, f_i, $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_p)^T$, $\underset{\sim}{\tau}_i$, $i = 1,2,\ldots,$ be as defined in section 1, and Ω and T_i be, respectively, the parametric spaces of $\underset{\sim}{\theta}$ and $\underset{\sim}{\tau}_i$ containing the true parameters $\underset{\sim}{\theta}^{\circ}$ and $\underset{\sim}{\tau}_i^{\circ}$, $i = 1,2,\ldots$ For a sample of size n, the log likelihood is therefore

$$L(\underset{\sim}{\theta},\underset{\sim}{\tau}_1,\ldots,\underset{\sim}{\tau}_n) = \sum \ell n\, f_i(\underset{\sim}{x}_i;\underset{\sim}{\theta},\underset{\sim}{\tau}_i),$$

where $\sum$ denotes $\sum_{i=1}^{n}$. The MLE $\hat{\underset{\sim}{\theta}}^n, \hat{\underset{\sim}{\tau}}_1^n, \ldots, \hat{\underset{\sim}{\tau}}_n^n$ of $\overset{\circ}{\underset{\sim}{\theta}}, \overset{\circ}{\underset{\sim}{\tau}}_1, \ldots, \overset{\circ}{\underset{\sim}{\tau}}_n$ are taken as roots of the equations $\partial L/\partial \underset{\sim}{\psi} = \underset{\sim}{0}$, where $\underset{\sim}{\psi} = \underset{\sim}{\theta}, \underset{\sim}{\tau}_1, \cdots \underset{\sim}{\tau}_n$.

Assume throughout that for each i, a unique solution $\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta})$ to the equation

$$\frac{\partial}{\partial \underset{\sim}{\tau}_i} \ln f_i\ (\underset{\sim}{x}_i;\underset{\sim}{\theta},\underset{\sim}{\tau}_i) = \underset{\sim}{0} \qquad (2.1)$$

exists when it is considered as a function in $\underset{\sim}{\tau}_i$. To simplify notation, let, whenever the derivatives exist,

$$\underset{\sim}{f}_{i\theta}(\underset{\sim}{x}_i,\underset{\sim}{\theta},\underset{\sim}{\tau}_i) = \frac{\partial}{\partial \underset{\sim}{\theta}} \ln f_i(\underset{\sim}{x}_i;\underset{\sim}{\theta},\underset{\sim}{\tau}_i), \quad \underset{\sim}{f}_{i\theta} = (f_{i\theta_1},\ldots, f_{i\theta_p})^T,$$

$$q_{ik\ell}(\underset{\sim}{x}_i,\underset{\sim}{\theta}) = \frac{\partial}{\partial \theta_\ell} f_{i\theta_k}(\underset{\sim}{x}_i,\underset{\sim}{\theta},\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta})).$$

Let $\underset{\sim}{A}_n(\underset{\sim}{\theta})$ be the symmetric random matrix whose $(k,\ell)^{th}$ element is $\sum q_{ik\ell}(\underset{\sim}{x}_i,\underset{\sim}{\theta})/n$. For every random vector $\underset{\sim}{Y}$ with distribution depending only on $\underset{\sim}{\theta}$ and $\underset{\sim}{\tau}_i$, we also write $E(\underset{\sim}{Y})$ for $E(\underset{\sim}{Y}|\underset{\sim}{\theta}^\circ,\underset{\sim}{\tau}_i^\circ)$, the expectation of $\underset{\sim}{Y}$ when the true parameters are $\underset{\sim}{\theta}_i^\circ$ and $\underset{\sim}{\tau}_i^\circ$.

<u>THEOREM 2.1A</u>. Let $\underset{\sim}{\theta}^1 \in \Omega^\circ$, where Ω° is the interior of Ω, and assume that the following regularity conditions are satisfied:

A1. For almost all $\underset{\sim}{x}_i$, $\frac{\partial}{\partial \theta_k} \ln f_i(\underset{\sim}{x}_i;\underset{\sim}{\theta},\underset{\sim}{\tau}_i)$, $\frac{\partial}{\partial \tau_{i\ell}} \ln f_i(\underset{\sim}{x}_i;\underset{\sim}{\theta},\underset{\sim}{\tau}_i)$, $\frac{\partial^2}{\partial \tau_{i\ell}\partial \theta_k} \ln f_i\ (\underset{\sim}{x}_i;\underset{\sim}{\theta},\underset{\sim}{\tau}_i)$ and $\frac{\partial^2}{\partial \theta_\ell \partial \theta_k} \ln f_i\ (\underset{\sim}{x}_i;\underset{\sim}{\theta},\underset{\sim}{\tau}_i)$ exist for every $(\underset{\sim}{\theta},\underset{\sim}{\tau}_i) \in \Omega^\circ \times T_i$.

A2. Given $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\left|\frac{1}{n}\sum E[\sup_{\|\underset{\sim}{\theta}-\underset{\sim}{\theta}^1\|<\delta_0} q_{ik\ell}\ (\underset{\sim}{x}_i,\underset{\sim}{\theta})] - \frac{1}{n}\sum E[q_{ik\ell}\ (\underset{\sim}{x}_i,\underset{\sim}{\theta}^1)]\right| < \epsilon$$

for all $0 < \delta_0 < \delta$; the same is true when sup is replaced by inf.

$$\frac{\partial}{\partial \underset{\sim}{\tau}_i} \ell n f_i \, (\underset{\sim}{x}_i ; \underset{\sim}{\theta}, \underset{\sim}{\tau}_i) = \underset{\sim}{0} \tag{2.1}$$

exists when it is considered as a function in $\underset{\sim}{\tau}_i$. To simplify notation, let, whenever the derivatives exist,

$$\underset{\sim}{f}_{i\theta}(\underset{\sim}{x}_i, \underset{\sim}{\theta}, \underset{\sim}{\tau}_i) = \frac{\partial}{\partial \underset{\sim}{\theta}} \ell n f_i \, (\underset{\sim}{x}_i ; \underset{\sim}{\theta}, \underset{\sim}{\tau}_i), \quad \underset{\sim}{f}_{i\theta} = (f_{i\theta_1}, \ldots, f_{i\theta_p})^T,$$

$$q_{ik\ell}(\underset{\sim}{x}_i, \underset{\sim}{\theta}) = \frac{\partial}{\partial \theta_\ell} f_{i\theta_k}(\underset{\sim}{x}_i, \underset{\sim}{\theta}, \underset{\sim}{g}_i(\underset{\sim}{x}_i, \underset{\sim}{\theta})).$$

Let $\underset{\sim}{A}_n(\underset{\sim}{\theta})$ be the symmetric random matrix whose $(k,\ell)^{th}$ element is $\sum q_{ik\ell}(\underset{\sim}{x}_i, \underset{\sim}{\theta})/n$. For every random vector $\underset{\sim}{Y}$ with distribution depending only on $\underset{\sim}{\theta}$ and $\underset{\sim}{\tau}_i$, we also write $E(\underset{\sim}{Y})$ for $E(\underset{\sim}{Y} | \underset{\sim}{\theta}^\circ, \underset{\sim}{\tau}_i^\circ)$, the expectation of $\underset{\sim}{Y}$ when the true parameters are $\underset{\sim}{\theta}_i^\circ$ and $\underset{\sim}{\tau}_i^\circ$.

<u>THEOREM 2.1A</u>. Let $\underset{\sim}{\theta}^1 \in \Omega^\circ$, where Ω° is the interior of Ω, and assume that the following regularity conditions are satisfied:

A1. For almost all $\underset{\sim}{x}_i$, $\frac{\partial}{\partial \theta_k} \ell n f_i(\underset{\sim}{x}_i ; \underset{\sim}{\theta}, \underset{\sim}{\tau}_i)$, $\frac{\partial}{\partial \tau_{i\ell}} \ell n f_i(\underset{\sim}{x}_i ; \underset{\sim}{\theta}, \underset{\sim}{\tau}_i)$, $\frac{\partial^2}{\partial \tau_{i\ell} \partial \theta_k} \ell n f_i \, (\underset{\sim}{x}_i ; \underset{\sim}{\theta}, \underset{\sim}{\tau}_i)$ and $\frac{\partial^2}{\partial \theta_\ell \partial \theta_k} \ell n f_i \, (\underset{\sim}{x}_i ; \underset{\sim}{\theta}, \underset{\sim}{\tau}_i)$ exist for every $(\underset{\sim}{\theta}, \underset{\sim}{\tau}_i) \in \Omega^\circ \times T_i$.

A2. Given $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\left| \frac{1}{n} \sum E[\sup_{||\underset{\sim}{\theta} - \underset{\sim}{\theta}^1|| < \delta_0} q_{ik\ell} \, (\underset{\sim}{x}_i, \underset{\sim}{\theta})] - \frac{1}{n} \sum E[q_{ik\ell} \, (\underset{\sim}{x}_i, \underset{\sim}{\theta}^1)] \right| < \epsilon$$

for all $0 < \delta_0 < \delta$; the same is true when sup is replaced by inf.

A3. There exists a $\delta > 0$ and functions $h_{ik\ell}(\underset{\sim}{x}_i)$ such that $|q_{ik\ell}(\underset{\sim}{x}_i,\underset{\sim}{\theta})| \le h_{ik\ell}(\underset{\sim}{x}_i)$ for almost all $\underset{\sim}{x}_i$, $\underset{\sim}{\theta}$ with $\|\underset{\sim}{\theta} - \underset{\sim}{\theta}^1\| \le \delta$, and $\overline{\lim_n} \sum E(h_{ik\ell}(\underset{\sim}{x}_i))^2/n < \infty$.

Assume also that $\sum[\underset{\sim}{f}_{i\theta}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1)) - E\,\underset{\sim}{f}_{i\theta}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1))]/n \overset{P}{\to} \underset{\sim}{0}$, where $\overset{P}{\to}$ denotes convergence in probability as $n \to \infty$. Then a necessary condition for $\hat{\underset{\sim}{\theta}}^n \overset{P}{\to} \underset{\sim}{\theta}^1$ is

$$\frac{1}{n}\sum E[\underset{\sim}{f}_{i\theta}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1)) \mid \underset{\sim}{\theta}^\circ,\underset{\sim}{\xi}^\circ] \to 0 \quad \text{as } n \to \infty. \tag{2.2}$$

Conversely, we have

<u>THEOREM 2.1B</u>. If the assumptions A1 - A3 are satisfied and in addition the following condition holds:

A4. $\overline{\lim_n} \|[\underset{\sim}{A}_n(\underset{\sim}{\theta}^1)]^{-1}\| < \infty$, where $\|\underset{\sim}{A}\|$ of a matrix $\underset{\sim}{A}$ is defined as $\sup_{\|x\|\le 1} \|\underset{\sim}{A}\underset{\sim}{x}\|$.

Then (2.2) is sufficient for the following to hold: with probability going to one as $n \to \infty$, there exists a $\hat{\underset{\sim}{\theta}}^n$ satisfying the set of likelihood equations corresponding to $\underset{\sim}{\theta}$

$$\sum \underset{\sim}{f}_{i\theta}(\underset{\sim}{x}_i,\underset{\sim}{\theta},\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta})) = \underset{\sim}{0}$$

such that $\hat{\underset{\sim}{\theta}}^n \overset{P}{\to} \underset{\sim}{\theta}^1$. Furthermore, any other such sequence would equal $\hat{\underset{\sim}{\theta}}^n$ with probability going to one as $n \to \infty$.

The proofs of theorems 2.1A and 2.1B appear in Appendix A.

REMARK. A2 is implied by A3 together with the following condition:

A2'. $q_{ik\ell}(x_i, \theta)$ is a continuous function of θ uniformly in i. The proof of this is given in Appendix B. A2' is more easily verified than A2 in some circumstances.

It is thus seen that $\hat{\theta}^n$ converges in probability to θ^1 independently of the τ_i° if and only if (2.2) holds independently of the τ_i°. It is quite often that there exists a $\theta^1 = m(\theta^\circ)$ depending only on θ° such that (2.2) holds independently of the τ_i°, and we then have $\hat{\theta}^n \overset{p}{\to} m(\theta^\circ)$. This is precisely the situation in the linear functional relationship discussed in the next section. $m^{-1}(\hat{\theta}^n)$ would then be a consistent estimate of θ°. Thus although the MLE of θ may not be consistent, it is often possible to construct a consistent estimate which is a function of the MLE, and the problem reduces to searching for θ^1 such that (2.2) holds independently of the τ_i°.

We now give regularity conditions under which $\hat{\theta}^n$ is asymptotically normal. These conditions here do not involve any third derivatives.

A5. $E[f_{i\theta}(x_i, \theta^1, g_i(x_i, \theta^1))] = 0$, for all i.

A6. There exists a $\gamma > 0$ such that

$$\frac{1}{n^{1+\gamma/2}} \sum E|f_{i\theta_k}(x_i, \theta^1, g_i(x_i, \theta^1))|^{2+\gamma} \to 0 \text{ as } n\to\infty, \quad k = 1, \ldots, p.$$

A7. $$0 < \underline{\lim} \frac{1}{n} \sum \mathrm{Var}[f_{i\theta_k}(x_i, \theta^1, g_i(x_i, \theta^1))]$$

$$\leq \overline{\lim} \frac{1}{n} \sum \mathrm{Var}[f_{i\theta_k}(x_i, \theta^1, g_i(x_i, \theta^1))] < \infty, \quad k = 1, \ldots, p.$$

<u>REMARK</u>. A5 can be replaced by the weaker assumption that $\sum E[f_{i\theta_k}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1))]/n$ is of smaller order than $n^{-1/2}$.

The proof of the following theorem requires Appendix C.

<u>THEOREM 2.2</u>. Under assumptions A1 to A7, the MLE $\hat{\underset{\sim}{\theta}}^n$ is asymptotically normal with mean $\underset{\sim}{\theta}^1$ and covariance matrix $n^{-1/2}\{E[\underset{\sim}{A}_n(\underset{\sim}{\theta}^1)]\}^{-1}\ \underset{\sim}{V}_n\{E[\underset{\sim}{A}_n(\underset{\sim}{\theta}^1)]\}^{-1}$, where $\underset{\sim}{V}_n = \sum \mathrm{Var}[\underset{\sim}{f}_{i\theta}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1))]/n$, that is, $n^{1/2}\ \underset{\sim}{V}_n^{-1/2}\ E[\underset{\sim}{A}_n(\underset{\sim}{\theta}^1)](\hat{\underset{\sim}{\theta}}^n - \underset{\sim}{\theta}^1) \overset{d}{\to} N(\underset{\sim}{0},\underset{\sim}{I})$, where $\overset{d}{\to}$ means convergence in distribution as $n\to\infty$.

Note that in theorem 2.2, we do not assume that $\underset{\sim}{V}_n$ and $E[\underset{\sim}{A}_n(\underset{\sim}{\theta}^1)]$ converge to any limit. The convergences of $\underset{\sim}{V}_n$ and $E[\underset{\sim}{A}_n(\underset{\sim}{\theta}^1)]$ usually occur in the special case when the incidental parameters $\underset{\sim}{\tau}_i$ are generated from the same superpopulation.

<u>Proof</u>: By the Mean Value Theorem of a function from R^n to R^m, we have for any non-zero column vector $\underset{\sim}{\lambda} \in R^p$

$$-n^{1/2}\underset{\sim}{\lambda}^T\underset{\sim}{V}_n^{-1/2}[\frac{1}{n}\sum \underset{\sim}{f}_{i\theta}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1))]$$
$$= n^{1/2}\underset{\sim}{\lambda}^T\underset{\sim}{V}_n^{-1/2}\ \underset{\sim}{A}_n(\underset{\sim}{\theta}^n)(\hat{\underset{\sim}{\theta}}^n - \underset{\sim}{\theta}^1), \qquad (2.3)$$

where $\|\underset{\sim}{\theta}^n - \underset{\sim}{\theta}^1\| \le \|\hat{\underset{\sim}{\theta}}^n - \underset{\sim}{\theta}^1\|$. The left hand side of (2.3) when divided by $\sqrt{\underset{\sim}{\lambda}^T\underset{\sim}{\lambda}}$ has variance 1 and

$$\sum E \left| \frac{\lambda^T V_n^{-1/2} f_{i\theta}(x_i, \theta^1, g_i(x_i, \theta^1))}{\sqrt{n\lambda^T\lambda}} \right|^{2+\gamma}$$

$$\le \frac{1}{(\lambda^T\lambda)^{1+\frac{\gamma}{2}} n^{1+\frac{\gamma}{2}}} \sum \| \lambda^T V_n^{-\frac{1}{2}} \|^{2+\gamma} E \| f_{i\theta}(x_i, \theta^1, g_i(x_i, \theta^1)) \|^{2+\gamma}$$

$$\le \frac{K}{n^{1+\frac{\gamma}{2}}} \sum \sum_{k=1}^{p} E | f_{i\theta_k}(x_i, \theta^1, g_i(x_i, \theta^1)) |^{2+\gamma} \to 0$$

as $n \to \infty$ by A6. The last inequality is a consequence of A7 and the use of 8a of Rao (1973, p. 149). Thus by Liapounov's Theorem the left hand side of (2.3) $\xrightarrow{d} N(0, \lambda^T\lambda)$, or $\lambda^T V_n^{-1/2} A_n(\theta^n) n^{1/2}(\hat{\theta}^n - \theta^1) \xrightarrow{d} N(0, \lambda^T\lambda)$ for each $\lambda \in R^p$. By (iv) of Rao (1973, p. 128), $V_n^{-1/2} A_n(\theta^n) n^{1/2}(\hat{\theta}^n - \theta^1) \xrightarrow{d} N(0, I)$. Next we proceed to show that

$$E[A_n(\theta^1)] - A_n(\theta^n) \xrightarrow{p} 0. \tag{2.4}$$

Let $\epsilon > 0$ be given and let δ be as in A2 when ϵ is replaced by $\epsilon/2$. By A3 and Tchebychev's inequality, we have

$$\Pr\left(\left| \frac{1}{n} \sum \sup_{\|\theta - \theta^1\| < \delta} q_{ik\ell}(x_i, \theta) - \frac{1}{n} \sum E[\sup_{\|\theta - \theta^1\| < \delta} q_{ik\ell}(x_i, \theta)] \right| < \epsilon/2 \right) \to 1$$

as $n \to \infty$; the same is true when sup is replaced by inf. (2.4) follows immediately since $\hat{\theta}^n \xrightarrow{p} \theta^1$ (theorem 2.1B) and

$$\Pr\left(\left| \frac{1}{n} \sum E[q_{ik\ell}(x_i, \theta^1)] - \frac{1}{n} \sum q_{ik\ell}(x_i, \theta^n) \right| < \epsilon \right)$$

$$\geq \Pr(\left|\frac{1}{n}\sum \inf_{\|\theta-\theta^1\|<\delta} q_{ik\ell}(x_i,\theta) - \frac{1}{n}\sum E[\inf_{\|\theta-\theta^1\|<\delta} q_{ik\ell}(x_i,\theta)]\right| < \epsilon/2,$$

$$\left|\frac{1}{n}\sum \sup_{\|\theta-\theta^1\|<\delta} q_{ik\ell}(x_i,\theta) - \frac{1}{n}\sum E[\sup_{\|\theta-\theta^1\|<\delta} q_{ik\ell}(x_i,\theta)]\right| < \epsilon/2,$$

$$\|\hat{\theta}^n - \theta^1\| < \delta).$$

So, with probability going to 1 as $n \to \infty$, $[A_n(\theta^n)]^{-1}$ exists and

$$V_n^{-\frac{1}{2}} E[A_n(\theta^1)][A_n(\theta^n)]^{-1} V_n^{\frac{1}{2}} = I + V_n^{-\frac{1}{2}}\{E[A_n(\theta^1)]$$

$$- A_n(\theta^n)\}[A_n(\theta^n)]^{-1} V_n^{\frac{1}{2}} \xrightarrow{p} I \quad \text{as } n \to \infty$$

by (2.4), A4, A7 and Appendix C. Therefore

$$V_n^{-\frac{1}{2}} E[A_n(\theta^1)][n^{\frac{1}{2}}(\hat{\theta}^n - \theta^1)] = \{V_n^{-\frac{1}{2}} E[A_n(\theta^1)][A_n(\theta^n)]^{-1} V_n^{\frac{1}{2}}\}$$

$$\{V_n^{-\frac{1}{2}} A_n(\theta^n)[n^{\frac{1}{2}}(\hat{\theta}^n - \theta^1)]\} \xrightarrow{d} N(0, I).$$

In the case $\hat{\theta}^n$ is inconsistent but a function $m(\hat{\theta}^n)$ is consistent, we have that $m(\hat{\theta}^n)$ is asymptotically normal with mean $m(\theta^1) = \theta^o$ and covariance matrix $n^{-1}D\{E[A_n(\theta^1)]\}^{-1} V_n\{E[A_n(\theta^1)]\}^{-1} D^T$, where D^T is the matrix whose (i,j)th element is
$\left.\frac{\partial m_i}{\partial \theta_j}\right|_{\theta^1}$.

3. APPLICATION TO LINEAR FUNCTIONAL RELATIONSHIP

In this section, we apply the results in section 2 to discuss one of the various models in estimating linear functional relationship. Comprehensive reviews of the subject were given by Malinvaud (1970, Chapter 10), Moran (1971), and Kendall and Stuart (1973, Chapter 29). The following model is considered. Suppose two unobservable non-stochastic variables x and y are linearly related by $y = \alpha + \beta x$, where α and β are unknown and to be estimated. We observe $\xi = x + \delta$ and $\eta = y + \epsilon$, where δ and ϵ are independent and normal with zero means and variances $\sigma_{\delta\delta}$ and $\sigma_{\epsilon\epsilon}$, respectively. With a sample of size n, the model can be written as

$$\xi_i = x_i + \delta_i, \qquad \eta_i = \alpha + \beta x_i + \epsilon_i,$$

$$E(\delta_i \epsilon_j) = 0, \qquad E(\delta_i^2) = \sigma_{\delta\delta}, \qquad E(\epsilon_i^2) = \sigma_{\epsilon\epsilon}, \tag{3.1}$$

$$E(\delta_i \delta_j) = E(\epsilon_i \epsilon_j) = 0 \quad \text{when } i \neq j, \quad i, j = 1, \ldots, n.$$

Here we consider the case when $\lambda = \sigma_{\epsilon\epsilon}/\sigma_{\delta\delta}$ is assumed to be known (for unidentifiability difficulties arise when λ is unknown, cf. Kendall and Stuart, 1973, Chapter 29). The structural parameter is $\underset{\sim}{\theta} = (\alpha, \beta, \sigma_{\delta\delta})^T = (\theta_1, \theta_2, \theta_3)^T$ and the incidental parameters are the x_i.

Let $\underset{\sim}{\theta}^{\circ} = (\alpha^{\circ}, \beta^{\circ}, \sigma^{\circ}_{\delta\delta})^T$ be the true parameter. The explicit form of the MLE $\hat{\underset{\sim}{\theta}} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma}_{\delta\delta})^T$ can be found in Kendall and Stuart (1973, §29.16) and is a unique admissibile solution of the likelihood equations.

THEOREM 3.1. In model (3.1), $\hat{\underset{\sim}{\theta}} \overset{P}{\to} \underset{\sim}{\theta}^1 = (\alpha^{\circ}, \beta^{\circ}, \frac{1}{2}\sigma^{\circ}_{\delta\delta})^T$ provided that

$$0 < \underline{\lim}\, \frac{1}{n}\sum x_i^2 \le \overline{\lim}\, \frac{1}{n}\sum x_i^2 < \infty . \qquad (3.2)$$

If in addition, there exists a $\gamma > 0$ such that

$$\lim_{n\to\infty} \frac{1}{n^{1+\gamma/2}} \sum |x_i|^{2+\gamma} = 0, \qquad (3.3)$$

then $\hat{\underset{\sim}{\theta}}$ is asymptotically normal with mean $(\alpha^{\circ}, \beta^{\circ}, \frac{1}{2}\sigma^{\circ}_{\delta\delta})^T$ and covariance matrix

$$\frac{\lambda\sigma^{\circ}_{\delta\delta}}{S^2_{xx}} \begin{bmatrix} \Delta S^2_{xx} + k\,\bar{x}^2 & k\bar{x} & 0 \\ -k\bar{x} & k & 0 \\ 0 & 0 & \dfrac{S^2_{xx}\sigma^{\circ}_{\delta\delta}}{2\lambda} \end{bmatrix},$$

where $\Delta = (1 + (\beta^{\circ})^2/\lambda)$, $\bar{x} = \sum x_i/n$, $S_{xx} = \sum (x_i - \bar{x})^2/n$, and $k = (\Delta S_{xx} + \sigma^{\circ}_{\delta\delta})$.

Proof: We have

$$\ell n\ f_i((\xi_i,\eta_i);\underset{\sim}{\theta},x_i) = \text{constant} - \ell n\sigma_{\delta\delta} - \frac{1}{2}\ell n\ \lambda$$

$$+ \frac{1}{2\sigma_{\delta\delta}}\ [(\xi_i - x_i)^2 + \frac{1}{\lambda}\ (\eta_i - \alpha - \beta x_i)^2],$$

and A1 is obviously satisfied. Also after taking first partial derivates, we find $g_i((\xi_i,\eta_i),\underset{\sim}{\theta}) = (\lambda\xi_i + \beta\eta_i - \alpha\beta)/(\lambda + \beta^2)$,

$$f_{i\theta_1} = (\eta_i - \alpha - \beta\xi_i)/[\sigma_{\delta\delta}(\lambda + \beta^2)],$$

$$f_{i\theta_2} = [(\lambda\xi_i + \beta\eta_i - \alpha\beta)(\eta_i - \alpha - \beta\xi_i)]/[\sigma_{\delta\delta}(\lambda + \beta^2)^2], \quad (3.4)$$

$$f_{i\theta_3} = -\sigma_{\delta\delta}^{-1} + (\eta_i - \alpha - \beta\xi_i)^2/[2\sigma_{\delta\delta}^2(\lambda + \beta^2)],$$

and their expectations vanish at $\underset{\sim}{\theta}^1$. Differentiating (3.4) with respect to $\underset{\sim}{\theta}$, it can be seen that A2 and A3 are satisfied using (3.2). After some algebraic manipulation , it is found that

$$\{E[\underset{\sim}{A}_n(\underset{\sim}{\theta}^1)]\}^{-1} = -\frac{1}{2}\overset{\circ}{\sigma}_{\delta\delta}(\lambda + \beta^2)S_{xx}^{-1}\begin{bmatrix} 1 & \bar{x} & 0 \\ \bar{x} & \frac{1}{n}\sum x_i^2 & 0 \\ 0 & 0 & (\overset{\circ}{\sigma}_{\delta\delta})^{-1}(\lambda + \beta^2) \end{bmatrix}.$$

Since $\|\underset{\sim}{A}\| \le \sum_i \sum_j a_{ij}^2$, where $\underset{\sim}{A} = (a_{ij})$, it is clear that A4 is satisfied by (3.2). Hence $\hat{\underset{\sim}{\theta}} \xrightarrow{P} \underset{\sim}{\theta}^1$ by theorem 2.1B. Also we find

$$\underset{\sim}{V}_n = 2^2 \sigma^\circ_{\delta\delta}(\lambda + (\beta^\circ)^2)^{-1} \begin{bmatrix} 1 & \bar{x} & 0 \\ \bar{x} & \frac{1}{n}\sum x_i^2 + \frac{\lambda\sigma^\circ_{\delta\delta}}{\lambda + (\beta^\circ)^2} & 0 \\ 0 & 0 & \frac{2(\lambda + (\beta^\circ)^2)}{\sigma^\circ_{\delta\delta}} \end{bmatrix}.$$

Thus A6 and A7 hold because of (3.2) and (3.3). The proof is completed by using theorem 2.2.

We notice that the MLE of $\underset{\sim}{\theta}$ is consistent for α and β but not for $\sigma_{\delta\delta}$. As pointed out in section 1, the function $(\hat{\alpha},\hat{\beta},2\hat{\sigma}_{\delta\delta})$ of the MLE $(\hat{\alpha},\hat{\beta},\hat{\sigma}_{\delta\delta})$ is consistent. The inconsistency of $\sigma_{\delta\delta}$ had been observed in the literature (cf. Lindley 1947) and the usual unbiased correction is $\frac{2n}{n-2}\hat{\sigma}_{\delta\delta}$, which is asymptotically equivalent to $2\hat{\sigma}_{\delta\delta}$. The consistency of $\hat{\alpha}$ and of $\hat{\beta}$ has been demonstrated in the literature (cf. Lindley (1947); Kendall and Stuart (1973, Chapter 29)) but the method requires the convergence of $\bar{x}$ and S_{xx} to finite limits. Here we only require that the x_i should neither be too spread nor concentrated when $n \to \infty$ (see (3.2)). Asymptotic normality of the MLE of $\underset{\sim}{\theta}$ as $n \to \infty$ does not seem to have been investigated in the literature. The asymptotic covariance matrix of $(\hat{\alpha},\hat{\beta})$ was obtained by Patefield (1977) based on the explicit form of $(\hat{\alpha},\hat{\beta})$ (see also Barnett (1969, 1970); Robertson (1974).)

APPENDIX A

Proof of theorem 2.1A.

Applying the Mean Value Theorem we can write

$$-\frac{1}{n}\sum f_{i\theta_k}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1)) = \frac{1}{n}\sum f_{i\theta_k}(\underset{\sim}{x}_i,\hat{\underset{\sim}{\theta}}^n,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\hat{\underset{\sim}{\theta}}^n))$$

$$-\frac{1}{n}\sum f_{i\theta_k}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1)) = \underset{\sim}{A}_n(\underset{\sim}{\theta}^n)(\hat{\underset{\sim}{\theta}}^n-\underset{\sim}{\theta}^1),$$

where $\|\underset{\sim}{\theta}^n-\underset{\sim}{\theta}^1\| \le \|\hat{\underset{\sim}{\theta}}^n-\underset{\sim}{\theta}^1\|$. Since $\underset{\sim}{A}_n(\underset{\sim}{\theta}^n) - E[\underset{\sim}{A}_n(\underset{\sim}{\theta}^1)] \overset{p}{\to} \underset{\sim}{0}$ ((2.4) in the proof of theorem 2.2) and $\hat{\underset{\sim}{\theta}}^n-\underset{\sim}{\theta}^1 \overset{p}{\to} 0$, we therefore have $\sum f_{i\theta_k}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1))/n \overset{p}{\to} 0$ and hence $\sum E[f_{i\theta_k}(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1,\underset{\sim}{g}_i(\underset{\sim}{x}_i,\underset{\sim}{\theta}^1))]/n \to 0$ as $n \to \infty$.

Proof of Theorem 2.1B.

We first observe that using an argument similar to the proof of (2.4) in theorem 2.2, we have from A2 and A3, $\underset{\sim}{A}_n(\underset{\sim}{\theta}) - E[\underset{\sim}{A}_n(\underset{\sim}{\theta})] \overset{p}{\to} 0$ uniformly n a sufficiently small neighbourhood of $\underset{\sim}{\theta}^1$. It follows from A2 that $\sum E[\underset{\sim}{A}_n(\underset{\sim}{\theta})]/n$ is a equicontinuous function of $\underset{\sim}{\theta}$ at $\underset{\sim}{\theta}^1$ in n. A4 then ensures the existence of a λ such that $\lambda < \frac{1}{4}\|\{E[\underset{\sim}{A}_n(\underset{\sim}{\theta}^1)]\}^{-1}\|^{-1}$. The proof of theorem 2.1B can now be completed by applying the Inverse Function Theorem with an argument which is a suitable modification of those first used by Foutz (1977) in his proof of the existence and uniqueness of the MLE in the i.i.d. case.

APPENDIX B

A2' and A3 imply A2.

We prove the "sup" case. Suppose A2 is not true. Then we can find $\delta_n \to 0$ and k_n such that

$$\frac{1}{k_n}\sum_{i=1}^{k_n} E\Big[\sup_{\|\theta < \theta^1\| < \delta_n} q_{ik\ell}(x_i,\theta)\Big] - \frac{1}{k_n}\sum_{i=1}^{k_n} E[q_{ik\ell}(x_i,\theta^1)] \geq \epsilon .$$

Now by A2', we see that $W_n = k_n^{-1} \sum_{i=1}^{k_n} \sup_{\|\theta - \theta^1\| < \delta_n} q_{ik\ell}(x_i,\theta)$

$- k_n^{-1} \sum_{i=1}^{k_n} q_{ik\ell}(x_i,\theta^1)$ converges to 0 a.e. Also by A3 W_n is dominated uniformly in θ by a random variable Z_n in L^1. Hence by theorem 4.1.4 of Chung (1974), we have $E|W_n| \to 0$ as $n \to \infty$ and this is a contradiction.

APPENDIX C

Let $\{\underset{\sim}{A}_n\}_{n=1}^{\infty}$ be a sequence of $p \times p$ matrices and $\|\underset{\sim}{A}_n^{-1}\| < K$ for all n. If $\{\underset{\sim}{E}_n\}_{n=1}^{\infty}$ is a sequence of random matrices such that $\underset{\sim}{E}_n - \underset{\sim}{A}_n \overset{P}{\to} \underset{\sim}{0}$, then with probability going to 1 as $n \to \infty$, $\underset{\sim}{E}_n^{-1}$ exists and $(\underset{\sim}{E}_n - \underset{\sim}{A}_n)\underset{\sim}{E}_n^{-1} \overset{P}{\to} 0$.

__Proof__: Since $\underset{\sim}{E}_n - \underset{\sim}{A}_n \overset{P}{\to} 0$, with probability going to 1 as $n \to \infty$, $\|\underset{\sim}{E}_n - \underset{\sim}{A}_n\| < K^{-1} \le \|\underset{\sim}{A}_n^{-1}\|^{-1}$ which implies $\underset{\sim}{E}_n^{-1}$ exists by theorem 9.8(a) of Rudin (1964). Now it can be proved that with probability going to 1 as $n \to \infty$

$$\|\underset{\sim}{E}_n^{-1}\| \le (\|\underset{\sim}{A}_n^{-1}\|^{-1} - \|\underset{\sim}{E}_n - \underset{\sim}{A}_n\|)^{-1} < 2K^{-1}.$$

Hence

$$\|(\underset{\sim}{E}_n - \underset{\sim}{A}_n)\underset{\sim}{E}_n^{-1}\| \le \|\underset{\sim}{E}_n - \underset{\sim}{A}_n\| \, \|\underset{\sim}{E}_n^{-1}\| \overset{P}{\to} 0,$$

which implies $(\underset{\sim}{E}_n - \underset{\sim}{A}_n)\underset{\sim}{E}_n^{-1} \overset{P}{\to} \underset{\sim}{0}$.

REFERENCES

ANDERSEN, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. J. Roy. Statist. Soc., B, 32, 283-301.

BARNETT, V. D. (1969). Simultaneous pairwise linear structural relationships. Biometrics, 25, 129-142.

BARNETT, V. D. (1970). Fitting straight lines—the linear functional relationship with replicated observations. Appl. Statist., 19, 135-144.

CHUNG, K. L. (1974). A Course in Probability Theory, 2nd ed. New York: Academic Press.

FOUTZ, R. V. (1977). On the unique consistent solution to the likelihood equations. J. Amer. Statist. Assoc., 72, 147-148.

HOADLEY, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. Ann. Math. Statist., 42, 1977-1991.

KENDALL, M. G. and STUART, A. (1973). The Advanced Theory of Statistics, Vol. 2, 3rd ed. London: Griffin.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Statist., 27, 887-906.

KULLDORFF, G. (1957). On the consistency and asymptotic efficiency of maximum likelihood estimates. Skand. Aktuarietidskr., 40, 129-144.

LINDLEY, D. V. (1947). Regression lines and the linear functional relationship. J. Roy. Statist. Soc. Suppl., 9, 218-244.

MALINVAUD, E. (1970). _Statistical Methods of Econometrics_, 2nd ed. Amsterdam: North Holland.

MORAN, P. A. P. (1971). Estimating structural and functional relationships. _J. Mult. Anal._, 1, 232-255.

NEYMAN, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. _Econometrica_, 16, 1-32.

NORDEN, R. H. (1972). A survey of maximum likelihood estimation. _Rev. Int. Statist. Inst._, 40, 329-354.

NORDEN, R. H. (1973). A survey of maximum likelihood estimation–Part 2. _Rev. Int. Statist. Inst._, 41, 39-58.

PATEFIELD, W. M. (1977). On the information matrix in the linear functional relationship problem. _Appl. Statist._, 26, 69-70.

PHILIPPOU, A. N. and ROUSSAS, G. G. (1973). Asymptotic normality of the maximum likelihood estimate in the indpendent not identically distributed case. _Ann. Inst. Statist. Math._, 25, 45-56.

RAO, C. R. (1973). _Linear Statistical Inference and its Applications_, 2nd ed. New York: Wiley.

ROBERTSON, C. A. (1974). Large-sample theory for the linear structural relation. _Biometrika_, 61, 353-359.

RUDIN, W. (1964). _Principles of Mathematical Analysis_, 2nd ed. New York: McGraw-Hill.