SEEMINGLY UNRELATED NONLINEAR REGRESSIONS

by

A. Ronald Gallant

Institute of Statistics
Mimeograph Series No. 900
Raleigh - November 1973

SEEMINGLY UNRELATED NONLINEAR REGRESSIONS

by

A. Ronald Gallant*

ABSTRACT

The paper considers the estimation of the parameters of a set of nonlinear regression equations when the responses are contemporaneously but not serially correlated. Conditions are set forth such that the estimator obtained is strongly consistent, asymptotically normally distributed, and asymptotically more efficient than the single-equation least squares estimator. The methods presented allow estimation of the parameters subject to nonlinear restrictions across equations. The paper includes a discussion of methods to perform the computations, a worked example, and a Monte Carlo simulation.

* Assistant Professor of Statistics and Economics, Institute of Statistics, North Carolina State University, Raleigh, North Carolina 27607.

1. INTRODUCTION

This paper may be viewed as a generalization of Zellner's (1962) paper in the following respects. The response functions are allowed to be nonlinear in the parameters as well as linear. The parameters of the model may be estimated subject to nonlinear restrictions across equations as well as linear restrictions across equations.

We rely on Zellner's paper to provide the general economic considerations which motivate the study of the Seemingly Unrelated Regressions situation. The extension to the nonlinear case is motivated by concern for the specification error due to requiring that economic theory accommodate the assumptions of linear regression. The argument based on Taylor's Theorem used to justify the linear model as an adequate approximation of the economic model often leaves the reader of an econometric study less than satisfied. The case when the approximation is inadequate provides the primary motivation for this paper. It attempts to give the practioner the freedom to let his econometric model more adequately represent his economic model.

The estimation procedure set forth in Section 2 is the same as Zellner's (1962) Seemingly Unrelated Regressions technique with the exception that nonlinear response functions and nonlinear parametric constraints are permitted. The variances and covariances of the disturbances are estimated from the residuals derived from an equation-by-equation application of least squares. All parameters are then estimated simultaneously by applying Aitken's (1935) generalized least squares to the whole system of equations using these estimated variances and covariances. Nonlinear constraints on the parameters may be imposed in the Aitken phase of the procedure.

Conditions are set forth in Section 3 under which this procedure yields estimators which are strongly consistent (almost sure convergence) and asymptotically normally distributed. These assertions are proved in Section 4 using a set of lemmas stated and proved in the Appendix. The conditions used to obtain these results are patterned after the assumptions used by Malinvaud (1970) in a nonlinear regression context.

Either Hartley's (1961) Modified Gauss-Newton Method or Marquardt's (1963) Algorithm may be used to obtain the restricted or unrestricted estimates of the parameters in the Aitken phase of the procedure as well as in the equation-by-equation phase. The details and a worked example are in Section 5.

The unrestricted estimator obtained according to the Zellner procedure is shown to be asymptotically more efficient than the equation-by-equation least squares estimator except in two special cases. The first case would be more apt to occur in a designed multivariate experiment than in an econometric investigation. It is the case when each of the contemporaneous responses has the same response function and same independent variables. This special situation is entirely analogous with the one discussed by Zellner (1962, p. 170). A second case where the equation-by-equation least squares estimator is fully efficient is when there are no contemporaneous correlations among the responses.

The reader who is only interested in the statistical and numerical methods proposed in this paper need only read Section 2 for a description of the method and Section 5 for a discussion of how available computer programs may be used to carry out the computations.

Section 7 reports the results of a Monte-Carlo simulation.

2. ESTIMATION PROCEDURE

The estimation procedure is entirely analogous to Zellner's (1962) Seemingly Unrelated Regressions estimation method except that (possibly) nonlinear response functions are substituted for linear response functions.

We have a set of M nonlinear regression equations

$$y_{t\alpha} = f_\alpha(\underset{\sim}{x}_{t\alpha}, \underset{\sim}{\theta}^\circ_\alpha) + e_{t\alpha} \qquad (\alpha = 1, 2, \ldots, M) \qquad (t = 1, 2, \ldots, n)$$

where the inputs $\underset{\sim}{x}_{t\alpha}$ are k_α by 1 vectors and the unknown parameters $\underset{\sim}{\theta}^\circ_\alpha$ are p_α by 1 vectors known to be contained in the sets $\underset{\sim}{\Theta}_\alpha$. The errors

$$\underset{\sim}{u}_t = (e_{t1}, e_{t2}, \ldots, e_{tM})' \qquad (M \times 1)$$

are assumed to be independent, each having mean $\underset{\sim}{0}$, the same distribution function, and positive definite variance-covariance matrix $\underset{\sim}{\Sigma}$.

Each of the M regression equations may be written in a convenient vector form as

$$\underset{\sim}{y}_\alpha = \underset{\sim}{f}_\alpha(\underset{\sim}{\theta}^\circ_\alpha) + \underset{\sim}{e}_\alpha \qquad (\alpha = 1, 2, \ldots, M)$$

where

$$\underset{\sim}{y}_\alpha = (y_{1\alpha}, y_{2\alpha}, \ldots, y_{n\alpha})' \qquad (n \times 1),$$

$$\underset{\sim}{f}_\alpha(\underset{\sim}{\theta}_\alpha) = (f_\alpha(\underset{\sim}{x}_{1\alpha}, \underset{\sim}{\theta}_\alpha), f_\alpha(\underset{\sim}{x}_{2\alpha}, \underset{\sim}{\theta}_\alpha), \ldots, f_\alpha(\underset{\sim}{x}_{n\alpha}, \underset{\sim}{\theta}_\alpha))' \qquad (n \times 1),$$

$$\underset{\sim}{e}_\alpha = (e_{1\alpha}, e_{2\alpha}, \ldots, e_{n\alpha})' \qquad (n \times 1).$$

The first step of the procedure is to obtain the least squares estimators $\hat{\theta}_\alpha$ by minimizing

$$Q_\alpha(\theta_\alpha) = \frac{1}{n}(y_\alpha - f_\alpha(\theta_\alpha))'(y_\alpha - f_\alpha(\theta_\alpha))$$

over Θ_α equation by equation.

The second step is to form the residual vectors

$$\hat{e}_\alpha = y_\alpha - f_\alpha(\hat{\theta}_\alpha) \qquad (\alpha = 1, \ldots, M)$$

and estimate the elements $\sigma_{\alpha\beta}$ of the variance-covariance matrix Σ by

$$\hat{\sigma}_{\alpha\beta} = \frac{1}{n}\hat{e}_\alpha{}'\hat{e}_\beta \qquad (\alpha = 1, 2, \ldots, M) \quad (\beta = 1, 2, \ldots, M)$$

to obtain the estimate $\hat{\Sigma}$ of Σ.

To carry out the next step, the set of M regressions are arranged in a single regression

$$y = f(\theta^\circ) + e$$

where

$$y = (y_1', y_2', \ldots, y_M')' \qquad (nM \times 1),$$

$$f(\theta) = (f_1'(\theta_1), f_2'(\theta_2), \ldots, f_M'(\theta_M))' \qquad (nM \times 1),$$

$$\theta = (\theta_1', \theta_2', \ldots, \theta_M')' \qquad (\Sigma_{\alpha=1}^M p_\alpha \times 1),$$

$$e = (e_1', e_2', \ldots, e_M') \qquad (nM \times 1) .$$

The variance-covariance matrix for this regression is

$$\mathcal{E}(\underset{\sim}{e}\,\underset{\sim}{e}') = \underset{\sim}{\Sigma} \otimes \underset{\sim}{I} = \begin{pmatrix} \sigma_{11}\underset{\sim}{I} & \sigma_{12}\underset{\sim}{I} & \cdots & \sigma_{1M}\underset{\sim}{I} \\ \sigma_{21}\underset{\sim}{I} & \sigma_{22}\underset{\sim}{I} & \cdots & \sigma_{2M}\underset{\sim}{I} \\ \vdots & \vdots & & \vdots \\ \sigma_{M1}\underset{\sim}{I} & \sigma_{M2}\underset{\sim}{I} & \cdots & \sigma_{MM}\underset{\sim}{I} \end{pmatrix} \quad (nM \times nM)$$

where $\underset{\sim}{I}$ is the n by n identity matrix. This variance-covariance matrix is estimated by $\hat{\underset{\sim}{\Sigma}} \otimes \underset{\sim}{I}$ obtained in the second step of the procedure.

The third step of the procedure is to obtain the Aitken type estimator $\hat{\hat{\underset{\sim}{\theta}}}$ by minimizing

$$Q(\underset{\sim}{\theta}) = \frac{1}{n}(\underset{\sim}{y} - \underset{\sim}{f}(\underset{\sim}{\theta}))'(\hat{\underset{\sim}{\Sigma}}^{-1} \otimes \underset{\sim}{I})(\underset{\sim}{y} - \underset{\sim}{f}(\underset{\sim}{\theta}))$$

over $\underset{\sim}{\Theta} = \times_{\alpha=1}^{M} \underset{\sim}{\Theta}_{\alpha}$.

Define

$\underset{\sim}{\nabla}_{\alpha} f_{\alpha}(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}_{\alpha})$ = the p_{α} by 1 vector whose i^{th} element is

$$\frac{\partial}{\partial \theta_{i\alpha}} f_{\alpha}(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}_{\alpha}) ,$$

$\underset{\sim}{F}_{\alpha}(\underset{\sim}{\theta}_{\alpha})$ = the n by p_{α} matrix whose t^{th} row is $\underset{\sim}{\nabla}'_{\alpha} f_{\alpha}(\underset{\sim}{x}_{t\alpha}, \underset{\sim}{\theta}_{\alpha})$,

$$\underset{\sim}{F}(\underset{\sim}{\theta}) = \begin{pmatrix} \underset{\sim}{F}_1(\underset{\sim}{\theta}_1) & \underset{\sim}{0} & & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{F}_2(\underset{\sim}{\theta}_2) & \cdots & \underset{\sim}{0} \\ \vdots & \vdots & & \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{F}_M(\underset{\sim}{\theta}_M) \end{pmatrix} \quad (nM \times \Sigma_{\alpha=1}^{M} p_{\alpha}) .$$

The fourth and final step is to obtain the inverse of the matrix

$$\hat{\hat{\Omega}} = \frac{1}{n} F'(\hat{\hat{\theta}})(\hat{\Sigma}^{-1} \otimes I) F(\hat{\hat{\theta}}) \; .$$

In Section 4 it is shown that $\sqrt{n}\,(\hat{\hat{\theta}} - \theta^{\circ})$ is asymptotically normally distributed with a variance-covariance matrix for which $(\hat{\hat{\Omega}})^{-1}$ is a strongly consistent estimator.

One may wish to impose restrictions across equations in the Aitken improvement phase of the estimation. For our purposes, the most convenient way to represent these restrictions is by reparameterization. Let ρ be an r by 1 vector of new parameters and let g_α be a p_α by 1 vector valued function relating the original parameters to the new parameters according to

$$\theta_\alpha = g_\alpha(\rho) \; .$$

We assume that $r \leq p = \Sigma_{\alpha=1}^{M} p_\alpha$ and that ρ is contained in the parameter space P .

Define

$$g(\rho) = (g_1'(\rho),\ g_2'(\rho),\ \ldots,\ g_M'(\rho))' \qquad (p = \Sigma_{\alpha=1}^{M} p_\alpha \times 1)$$

$\nabla_\rho\, g_{i\alpha}(\rho)$ = the r by 1 vector whose j^{th} element is $\frac{\partial}{\partial \rho_j}\, g_{i\alpha}(\rho)$,

$G_\alpha(\rho)$ = the p_α by r matrix whose i^{th} row is $\nabla_\rho'\, g_{i\alpha}(\rho)$,

$$G(\rho) = \begin{pmatrix} G_1(\rho) \\ G_2(\rho) \\ \vdots \\ G_M(\rho) \end{pmatrix} \qquad (p = \Sigma_{\alpha=1}^{M} p_\alpha \times r) \; .$$

The third step of the procedure is modified to read: Obtain the estimator $\hat{\hat{\rho}}$ by minimizing

$$Q(g(\rho)) = \frac{1}{n} (y - f(g(\rho)))'(\hat{\Sigma}^{-1} \otimes I)(y - f(g(\rho)))$$

over the parameter space P.

The fourth step of the procedure is modified to read: Obtain the inverse of the matrix

$$\hat{\hat{G}}'\hat{\hat{\Omega}}\hat{\hat{G}} = G'(\hat{\hat{\rho}})[\frac{1}{n} F'(g(\hat{\hat{\rho}}))(\hat{\Sigma}^{-1} \otimes I) F(g(\hat{\hat{\rho}}))] G(\rho) .$$

In Section 4 it is shown that $\sqrt{n}(\hat{\hat{\rho}} - \rho^{\circ})$ is asymptotically normally distributed with a variance-covariance matrix for which $(\hat{\hat{G}}'\hat{\hat{\Omega}}\hat{\hat{G}})^{-1}$ is a strongly consistent estimator.

Either Hartley's (1961) Modified Gauss-Newton Method or Marquardt's (1963) Algorithm may be used to find θ_{α} minimizing $Q_{\alpha}(\theta_{\alpha})$. Using a suitable transformation, either of these methods may be applied to find $\hat{\hat{\theta}}$ minimizing $Q(\theta)$ and $(\hat{\Omega})^{-1}$ or $\hat{\hat{\rho}}$ minimizing $Q(g(\rho))$ and $(\hat{\hat{G}}\hat{\hat{\Omega}}\hat{\hat{G}})^{-1}$ as the case may be. The details are discussed in Section 5.

One may wish to estimate the parameters of the model subject to the restrictions $g(\rho) = \theta$ but present the results in terms of the original parameters θ. This is done by setting $\hat{\hat{\theta}} = g(\hat{\hat{\rho}})$. As shown in Section 4, $\hat{\hat{\theta}}$ is strongly consistent for θ° and $\sqrt{n}(\hat{\hat{\theta}} - \theta^{\circ})$ is asymptotically normally distributed with a variance-covariance matrix for which $\hat{\hat{G}}(\rho)(\hat{\hat{G}}'\hat{\hat{\Omega}}\hat{\hat{G}})^{-1}G'(\hat{\hat{\rho}})$ is a strongly consistent estimator. This matrix will be singular when $r < p$.

3. ASSUMPTIONS

In order to obtain asymptotic results, it is necessary to specify the behavior of the inputs $x_{t\alpha}$ as n becomes large. A general way of specifying the limiting behavior of nonlinear regression inputs is due to Malinvaud (1970). Two of his definitions are repeated below for the reader's convenience; a more complete discussion and examples are contained in his paper. In reading the two definitions, it will help if the reader keeps in mind two situations where the inputs satisfy Definitions 3.1 and 3.2. Let $\mathcal{X}_\alpha$ be the subset of R^{k_α} from which the inputs $x_{t\alpha}$ are to be chosen, $k = \Sigma_{\alpha=1}^M k_\alpha$,

$$x = (x_1', x_2', \ldots, x_M')' \qquad (k \times 1),$$

and let $\mathcal{X} = \times_{\alpha=1}^M \mathcal{X}_\alpha$. The first is to choose some set of points $\{x_t^*\}_{t=1}^T$ and choose inputs according to the scheme

$$x_1 = x_1^*,\ x_2 = x_2^*,\ \ldots,\ x_T = x_T^*,\ x_{T+1} = x_1^*,\ x_{T+2} = x_2^*,\ \ldots,$$

that is, "constant in repeated samples" (Theil, 1971, p. 364). The second is to choose inputs by randomly sampling from a distribution function defined over the set $\mathcal{X}$; for example, the k dimensional uniform distribution.

<u>Definition 3.1</u>: Let $\mathcal{A}$ be the Borel subsets of $\mathcal{X}$ and let $\{x_t\}_{t=1}^\infty$ be a sequence of inputs chosen from $\mathcal{X}$. For a set A in $\mathcal{A}$ let $I_A(x)$ be its indicator function. The probability measure μ_n on $(\mathcal{X}, \mathcal{A})$ is defined by

$$\mu_n(A) = \frac{1}{n} \Sigma_{t=1}^n I_A(x_t) .$$

Definition 3.2: A sequence of measures $\{\mu_n\}_{n=1}^{\infty}$ on $(\underset{\sim}{X}, \mathcal{A})$ is said to converge weakly to a measure μ on $(\underset{\sim}{X}, \mathcal{A})$ if for every real valued, bounded, and continuous function g with domain $\underset{\sim}{X}$,

$$\lim_{n \to \infty} \int g(\underset{\sim}{x})\, d\mu_n(\underset{\sim}{x}) = \int g(\underset{\sim}{x})\, d\mu(\underset{\sim}{x}) .$$

The following Assumptions are sufficient to obtain the asymptotic properties of the estimator $\hat{\hat{\underset{\sim}{\theta}}}$ in the case when no restrictions are imposed on the parameters $\underset{\sim}{\theta}$.

Assumptions: The parameter spaces $\underset{\sim}{\Theta}_\alpha$ and sets $\underset{\sim}{X}_\alpha$ are compact. The response function $f_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha)$, the first partial derivatives in $\underset{\sim}{\theta}_\alpha$, and the second partial derivatives in $\underset{\sim}{\theta}_\alpha$ are continuous on $\underset{\sim}{X}_\alpha \times \underset{\sim}{\Theta}_\alpha$ for each $\alpha = 1, 2, \ldots, M$. The sequence of inputs $\{\underset{\sim}{x}_t\}_{t=1}^{\infty}$ are chosen such that the sequence of measures $\{\mu_n\}_{n=1}^{\infty}$ converges weakly to a measure μ defined on $(\underset{\sim}{X}, \mathcal{A})$. The true parameter value $\underset{\sim}{\theta}_\alpha^\circ$ is contained in an open spehre $\underset{\sim}{S}_\alpha$ which is, in turn, contained in $\underset{\sim}{\Theta}_\alpha$ for each $\alpha = 1, 2, \ldots, M$. If $f_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha) = f_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha^\circ)$ except for $\underset{\sim}{x}$ in A where $\mu(A) = 0$, it is assumed that $\underset{\sim}{\theta}_\alpha = \underset{\sim}{\theta}_\alpha^\circ$. The p by p matrix

$$\underset{\sim}{\Omega} = \begin{pmatrix} \sigma^{11}\underset{\sim}{V}_{11} & \sigma^{12}\underset{\sim}{V}_{12} & \cdots & \sigma^{1M}\underset{\sim}{V}_{1M} \\ \sigma^{21}\underset{\sim}{V}_{21} & \sigma^{22}\underset{\sim}{V}_{22} & \cdots & \sigma^{2M}\underset{\sim}{V}_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma^{M1}\underset{\sim}{V}_{M1} & \sigma^{M2}\underset{\sim}{V}_{M2} & \cdots & \sigma^{MM}\underset{\sim}{V}_{MM} \end{pmatrix}$$

is non-singular where the $\sigma^{\alpha\beta}$ are the elements of $\underset{\sim}{\Sigma}^{-1}$ and the p_α by p_β matrices $\underset{\sim}{V}_{\alpha\beta}$ have the ij^{th} element

$$v_{ij\alpha\beta} = \int \frac{\partial}{\partial\theta_{i\alpha}} f_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}^\circ_\alpha) \frac{\partial}{\partial\theta_{j\beta}} f_\beta(\underset{\sim}{x}_\beta, \underset{\sim}{\theta}^\circ_\beta)\, d\mu(\underset{\sim}{x}) .$$

As mentioned earlier, the errors $\underset{\sim}{\mu}_t$ are independently and identically distributed with mean $\underset{\sim}{0}$ and unknown positive-definite variance-covariance matrix $\underset{\sim}{\Sigma}$.

In the case where the restrictions $\underset{\sim}{\theta} = \underset{\sim}{g}(\underset{\sim}{\rho})$ are imposed on the parameters, the following additional assumptions are required.

<u>Assumptions</u>: (Continued) The parameter space $\underset{\sim}{P}$ is compact. The range of $\underset{\sim}{g}$ is a subset of $\underset{\sim}{\Theta} = X_{\alpha=1}^{M} \underset{\sim}{\Theta}_\alpha$. The function $\underset{\sim}{g}(\underset{\sim}{\rho})$ and its first and second partial derivatives in $\underset{\sim}{\rho}$ are continuous. The true parameter value $\underset{\sim}{\rho}^\circ$ satisfies $\underset{\sim}{g}(\underset{\sim}{\rho}^\circ) = \underset{\sim}{\theta}^\circ$ and $\underset{\sim}{\rho}^\circ$ is contained in an open sphere $\underset{\sim}{S}$ which is, in turn, contained in $\underset{\sim}{P}$. If $\underset{\sim}{g}(\underset{\sim}{\rho}) = \underset{\sim}{\theta}^\circ$ it is assumed that $\underset{\sim}{\rho} = \underset{\sim}{\rho}^\circ$. The matrix $\underset{\sim}{G}'\underset{\sim}{\Omega}\,\underset{\sim}{G}$ where $\underset{\sim}{G} = \underset{\sim}{G}(\underset{\sim}{\rho}^\circ)$ is non-singular.

4. STRONG CONSISTENCY AND ASYMPTOTIC NORMALITY

Two theorems establishing the strong consistency and asymptotic normality of the estimator $\hat{\hat{\rho}}$ are proved in this section; the strong consistency and asymptotic normality of $\hat{\hat{\theta}}$ follow as corollaries to these two theorems. In order to simplify the notation in this section and the remainder of the paper we will adopt the convention of writing f_α for the function $f_\alpha(\theta_\alpha)$ when it is evaluated at the true value of the parameter $\theta_\alpha = \theta_\alpha^\circ$. Similarly, we will write F_α for $F_\alpha(\theta_\alpha^\circ)$, f for $f(\theta^\circ)$, F for $F(\theta^\circ)$, and G for $G(\rho^\circ)$.

Theorem 1: The estimator $\hat{\hat{\rho}}$ converges almost surely to ρ° under the Assumptions listed in Section 3.

Proof: Consider the quadratic form

$$Q(\theta) = \frac{1}{n}(y - f(\theta))'(\hat{\Sigma}^{-1} \otimes I)(y - f(\theta))$$

$$= \Sigma_{\alpha=1}^{M} \Sigma_{\beta=1}^{M} \hat{\sigma}^{\alpha\beta} \frac{1}{n}(y_\alpha - f_\alpha(\theta_\alpha))'(y_\beta - f_\beta(\theta_\beta)) .$$

The random variables $\hat{\sigma}^{\alpha\beta}$ converge almost surely to the elements $\sigma^{\alpha\beta}$ of Σ^{-1} by Lemma A.5. Using arguments similar to those employed in the proof of Lemma A.5, the terms $\frac{1}{n}(y_\alpha - f_\alpha(\theta_\alpha))'(y_\beta - f_\beta(\theta_\beta))$ converge almost surely and uniformly for θ in Θ to

$$\sigma_{\alpha\beta} + \int \delta_\alpha(x_\alpha, \theta_\alpha)\, \delta_\beta(x_\beta, \theta_\beta)\, d\mu(x)$$

where $\delta_\alpha(x_\alpha, \theta_\alpha) = f_\alpha(x_\alpha, \theta_\alpha^\circ) - f_\alpha(x_\alpha, \theta_\alpha)$. Thus, $Q(\theta)$ converges almost surely to

$$Q^*(\underset{\sim}{\theta}) = \Sigma_{\alpha=1}^{M} \Sigma_{\beta=1}^{M} \sigma^{\alpha\beta}\sigma_{\alpha\beta} + \int \Sigma_{\alpha=1}^{M} \Sigma_{\beta=1}^{M} \sigma^{\alpha\beta}\delta_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha)\delta_\beta(\underset{\sim}{x}_\beta, \underset{\sim}{\theta}_\beta)\, d\mu(\underset{\sim}{x})$$

$$= M + \int v(\underset{\sim}{x}, \underset{\sim}{\theta})\, d\mu(\underset{\sim}{x})$$

uniformly for $\underset{\sim}{\theta}$ in $\underset{\sim}{\Theta}$. The matrix Σ^{-1} is positive definite so the integrand $v(\underset{\sim}{x}, \underset{\sim}{\theta})$ is positive for all $(\underset{\sim}{x}', \underset{\sim}{\theta}')'$ in $\underset{\sim}{X} \times \underset{\sim}{\Theta}$. Consequently, if $Q^*(\underset{\sim}{\theta}) = M$ then $v(\underset{\sim}{x}, \underset{\sim}{\theta}) = 0$ except for $\underset{\sim}{x}$ in A where $\mu(A) = 0$. This implies $\delta_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha) = 0$ except for $\underset{\sim}{x}$ in A which implies $\underset{\sim}{\theta}_\alpha = \underset{\sim}{\theta}_\alpha^\circ$ by Assumption. Thus, $Q^*(\underset{\sim}{g}(\underset{\sim}{\rho})) = M$ implies $\underset{\sim}{g}(\underset{\sim}{\rho}) = \underset{\sim}{\theta}^\circ$ which implies $\underset{\sim}{\rho} = \underset{\sim}{\rho}^\circ$.

Consider a sequence $\{\hat{\underset{\sim}{\rho}}_n\}$ of points minimizing $Q(\underset{\sim}{g}(\underset{\sim}{\rho}))$ over $\underset{\sim}{P}$ corresponding to a realization of the errors $\{\underset{\sim}{u}_t\}_{t=1}^{\infty}$. Since $\underset{\sim}{P}$ is compact there is at least one limit point $\underset{\sim}{\rho}^*$ and one subsequence $\{\hat{\underset{\sim}{\rho}}_{n_m}\}_{m=1}^{\infty}$ such that $\lim_{m\to\infty} \hat{\underset{\sim}{\rho}}_{n_m} = \underset{\sim}{\rho}^*$. Unless this realization belongs to the exceptional E, $Q(\underset{\sim}{g}(\underset{\sim}{\rho}))$ converges to $Q^*(\underset{\sim}{g}(\underset{\sim}{\rho}))$ uniformly for $\underset{\sim}{\rho}$ in $\underset{\sim}{P}$ whence

$$M \le Q^*(\underset{\sim}{g}(\underset{\sim}{\rho}^*)) = \lim_{m\to\infty} Q(\underset{\sim}{g}(\hat{\underset{\sim}{\rho}}_{n_m})) \le \lim_{m\to\infty} Q(\underset{\sim}{g}(\underset{\sim}{\rho}^\circ)) = M .$$

This implies $Q^*(\underset{\sim}{g}(\underset{\sim}{\rho}^*)) = M$ which implies $\underset{\sim}{\rho}^* = \underset{\sim}{\rho}^\circ$. Thus, the sequence $\{\hat{\underset{\sim}{\rho}}_n\}_{n=1}^{\infty}$ has only one limit point $\underset{\sim}{\rho}^\circ$ except for realizations of the errors $\{\underset{\sim}{u}_t\}_{t=1}^{\infty}$ contained in E where $P(E) = 0$. ▯

<u>Corollary 1</u>: The estimator $\hat{\hat{\underset{\sim}{\theta}}}$ converges almost surely to $\underset{\sim}{\theta}^\circ$ under the Assumptions listed in Section 3.

<u>Proof</u>: Let $\underset{\sim}{g}(\underset{\sim}{\rho})$ be the identity mapping of $\underset{\sim}{\Theta}$ onto $\underset{\sim}{\Theta}$. The set $X_{\alpha=1}^{M} \underset{\sim}{S}_\alpha$ is an open set containing $\underset{\sim}{\theta}^\circ$ which is, in turn, contained

in Θ. Thus, there exists an open sphere S containing θ° which is, in turn, contained in Θ. The matrix $G(\rho) = I$ so $G'\Omega G$ is nonsingular. Consequently, the identity mapping satisfies the full set of assumptions. ∎

Theorem 2: If the Assumptions listed in Section 3 are satisfied then $\sqrt{n}\,(\hat{\hat{\rho}} - \rho^{\circ})$ converges in distribution to an r-variate normal with mean 0 and variance-covariance matrix $(G'\,\Omega\,G)^{-1}$. The matrix

$$\hat{G}'\hat{\Omega}\,\hat{G} = G'(\hat{\rho})[\,\tfrac{1}{n}\,F'(g(\hat{\rho}))(\hat{\Sigma}^{-1}\otimes I)\;F(g(\hat{\rho}))\,]G(\hat{\rho})$$

converges almost surely to $G'\Omega\,G$.

Proof: Define $\dot{\rho} = \hat{\hat{\rho}}$ if $\hat{\hat{\rho}}$ is in S and $= \rho^{\circ}$ if $\hat{\hat{\rho}}$ is not in S. Set $\dot{\theta} = g(\dot{\rho})$. Since $\sqrt{n}(\dot{\rho} - \hat{\hat{\rho}})$ converges almost surely to 0 by Theorem 1, it will suffice to prove the theorem for $\dot{\rho}$.

The first order Taylor series expansion of $f(\dot{\theta})$ may be written as

$$f(\dot{\theta}) = f + F\,G(\dot{\rho} - \rho^{\circ}) + H(\dot{\rho} - \rho^{\circ})$$

where H is the nM by r matrix

$$H = (H_1',\ H_2',\ \ldots,\ H_M')'\ ;$$

the t^{th} row of the n by r submatrix H_{α} is $\frac{1}{2}(\dot{\rho} - \rho^{\circ})'\nabla_{\rho}^{2} f_{\alpha}(x_{t\alpha}, g_{\alpha}(\bar{\rho}))$ where $\bar{\rho}$ is on the line segment joining $\dot{\rho}$ to ρ°. Consider the matrix $\frac{1}{n}\,F'(\dot{\theta})(\Sigma^{-1}\otimes I)H$ which is composed of the p_{α} by r submatrices

$$\Sigma_{\beta=1}^{M}\,\frac{\sigma^{\alpha\beta}}{2}\,\frac{1}{n}\,\Sigma_{t=1}^{n}\,\nabla_{\alpha} f_{\alpha}(x_{t\alpha},\ \dot{\theta}_{\alpha})(\dot{\rho} - \rho^{\circ})\nabla_{\rho}^{2}\, f_{\beta}(x_{t\beta},\ g_{\beta}(\bar{\rho}))\ .$$

The elements of this submatrix converge almost surely to zero by the continuity of the first and second order partial derivatives of $f_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha)$ and $\underset{\sim}{g}_\alpha(\underset{\sim}{\rho})$, the compactness of the set $\underset{\sim}{X} \times \underset{\sim}{\Theta} \times \underset{\sim}{P}$, and the almost sure convergence of $\dot{\underset{\sim}{\rho}}$ to $\underset{\sim}{\rho}^\circ$. In consequence, the matrix $\frac{1}{n}\underset{\sim}{G}'(\dot{\underset{\sim}{\rho}}) \times \underset{\sim}{F}'(\dot{\underset{\sim}{\theta}})(\underset{\sim}{\Sigma}^{-1} \otimes \underset{\sim}{I})\underset{\sim}{H}$ converges almost surely to $\underset{\sim}{0}$.

Using Taylor's Theorem, we may write

$$\underset{\sim}{G}'(\dot{\underset{\sim}{\rho}})\underset{\sim}{F}'(\dot{\underset{\sim}{\theta}})(\underset{\sim}{\Sigma}^{-1} \otimes \underset{\sim}{I})\,\underset{\sim}{e} = \underset{\sim}{G}'\underset{\sim}{F}'(\underset{\sim}{\Sigma}^{-1} \otimes \underset{\sim}{I})\,\underset{\sim}{e} + \underset{\sim}{D}(\dot{\underset{\sim}{\rho}} - \underset{\sim}{\rho}^\circ)$$

where $\underset{\sim}{D}$ is the r by r matrix with i,j^{th} element

$$d_{ij} = \Sigma_{\alpha=1}^{M} \Sigma_{\beta=1}^{M} \sigma^{\alpha\beta} \Sigma_{t=1}^{n} \frac{\partial^2}{\partial\rho_j \partial\rho_i} f_\alpha(\underset{\sim}{x}_{t\alpha}, \underset{\sim}{g}_\alpha(\bar{\underset{\sim}{\rho}}))\, e_{t\beta}\,.$$

Note that $\frac{1}{n} d_{ij}$ converges almost surely to 0 by Lemma A.4.

Using the Taylor series expansions obtained in the two previous paragraphs, we may write the r by 1 vector of partial derivatives in $\underset{\sim}{\rho}$ of $(-\sqrt{n}/2)Q(\underset{\sim}{g}(\dot{\underset{\sim}{\rho}}))$ as

$$(-\sqrt{n}/2)\underset{\sim}{\nabla}_\rho Q(\underset{\sim}{g}(\dot{\underset{\sim}{\rho}})) = \frac{1}{\sqrt{n}}\underset{\sim}{G}'\underset{\sim}{F}'(\underset{\sim}{\Sigma}^{-1} \otimes \underset{\sim}{I})\,\underset{\sim}{e}$$

$$- [\tfrac{1}{n}\underset{\sim}{G}'(\dot{\underset{\sim}{\rho}})\underset{\sim}{F}'(\dot{\underset{\sim}{\theta}})(\underset{\sim}{\Sigma}^{-1} \otimes \underset{\sim}{I})(\underset{\sim}{F}\,\underset{\sim}{G} + \underset{\sim}{H}) - \tfrac{1}{n}\underset{\sim}{D}]\,\sqrt{n}\,(\dot{\underset{\sim}{\rho}} - \underset{\sim}{\rho}^\circ)$$

$$+ \tfrac{1}{n}\underset{\sim}{G}'(\dot{\underset{\sim}{\rho}})\underset{\sim}{F}'(\dot{\underset{\sim}{\theta}})[\sqrt{n}\,(\hat{\underset{\sim}{\Sigma}}^{-1} - \underset{\sim}{\Sigma}^{-1}) \otimes \underset{\sim}{I}]\,\underset{\sim}{e}\,.$$

$$+ \tfrac{1}{n}\underset{\sim}{G}'(\dot{\underset{\sim}{\rho}})\underset{\sim}{F}'(\dot{\underset{\sim}{\theta}})[\sqrt{n}\,(\hat{\underset{\sim}{\Sigma}}^{-1} - \underset{\sim}{\Sigma}^{-1}) \otimes \underset{\sim}{I}]\,(\underset{\sim}{f} - \underset{\sim}{f}(\dot{\underset{\sim}{\theta}}))$$

The estimator $\hat{\hat{\underset{\sim}{\rho}}}$ is contained in $\underset{\sim}{S}$ for n sufficiently large except on an exceptional event E occurring with probability zero by Theorem 1; thus $(-\sqrt{n}/2)Q(\underset{\sim}{g}(\dot{\underset{\sim}{\rho}}))$ converges to $\underset{\sim}{0}$ almost surely as $\dot{\underset{\sim}{\rho}}$ is a stationary point of $Q(g(\rho))$ when $\hat{\hat{\rho}}$ is in S. In consequence, $\sqrt{n}\,(\dot{\rho} - \rho^{\circ})$ converges in distribution to a r-variate normal with mean 0 and variance-covariance matrix $(G'\Omega G)^{-1}$ because:

A. The first term on the right converges in distribution to a r-variate normal with mean 0 and variance-covariance matrix $G'\Omega G'$ by Lemma A.6.

B. The matrix in brackets appearing in the second term on the right converges almost surely to $G'\Omega G$. This follows because $\frac{1}{n} F'(\dot{\theta})(\hat{\Sigma}^{-1} \otimes I) F$ converges almost surely to Ω by Lemma A.3, $G(\dot{\rho})$ converges almost surely to G, and the remaining terms converge almost surely to 0 by our preceeding remarks.

C. The third term on the right converges in probability to 0. This follows because the subvectors $\sum_{\beta=1}^{M} \sqrt{n}\,(\hat{\sigma}^{\alpha\beta} - \sigma^{\alpha\beta}) \frac{1}{n} F'_{\alpha}(\dot{\theta}_{\alpha}) e_{\beta}$ converge in probability to 0 by Lemma A.5 and Lemma A.4.

D. The fourth term on the right converges in probability to 0. This follows because the subvectors $\sum_{\beta=1}^{M} \sqrt{n}\,(\hat{\sigma}^{\alpha\beta} - \sigma^{\alpha\beta}) \frac{1}{n} F'_{\alpha}(\dot{\theta}_{\alpha}) \times (f_{\beta} - f_{\beta}(\dot{\theta}_{\beta}))$ converge in probability to 0 by Lemma A.5, Lemma A.1, and Theorem 1.

The last sentence of the theorem follows by applying Lemma A.5 to obtain the almost sure convergence of $\hat{\sigma}^{\alpha\beta}$ to $\sigma^{\alpha\beta}$, by applying Theorem 1 to obtain the almost sure convergence of $g(\hat{\hat{\rho}})$ to θ° and $G(\hat{\hat{\rho}})$ to G, and by applying Lemma A.3 to obtain the almost sure convergence of the submatrices $\frac{1}{n} F'_{\alpha}(g_{\alpha}(\hat{\hat{\rho}}))\, F_{\beta}(g_{\beta}(\hat{\hat{\rho}}))$ to $V_{\alpha\beta}$. ▯

Corollary 2: If the Assumptions listed in Section 3 are satisfied then $\sqrt{n}(\hat{\hat{\theta}} - \theta^\circ)$ converges in distribution to a p-variate normal with mean 0 and variance-covariance matrix Ω^{-1}. The matrix

$$\hat{\hat{\Omega}} = \frac{1}{n} F'(\hat{\hat{\theta}})(\hat{\Sigma}^{-1} \otimes I)\, F(\hat{\hat{\theta}})$$

converges almost surely to Ω.

Proof: The proof is the same as the proof of Corollary 1. ▯

Theorem 3: Let the Assumptions of Section 3 be satisfied and set $\hat{\hat{\hat{\theta}}} = g(\hat{\rho})$. Then $\hat{\hat{\hat{\theta}}}$ converges almost surely to θ° and $\sqrt{n}(\hat{\hat{\hat{\theta}}} - \theta^\circ)$ is asymptotically normally distributed with mean 0 and (possibly) singular variance-covariance matrix $G(G'\Omega G)^{-1}G'$. The matrix $\hat{G}(\hat{\rho})(\hat{G}'\hat{\Omega}\hat{G})^{-1}G'(\hat{\rho})$ is strongly consistent for $G(G'\Omega G)^{-1}G'$.

Proof: The almost sure convergence of $\hat{\hat{\hat{\theta}}}$ to θ° follows from the continuity of $g(\rho)$ and Theorem 1.

Let $\dot{\rho}$ be as in the proof of Theorem 2. It will suffice to prove the theorem for $g(\dot{\rho})$ instead of $\hat{\hat{\hat{\theta}}}$ because $\sqrt{n}(\hat{\hat{\hat{\theta}}} - g(\dot{\rho}))$ converges almost surely to zero by Theorem 1. The first order Taylor series expansion of $g(\dot{\rho})$ is

$$g(\dot{\rho}) = \theta^\circ + G(\dot{\rho} - \rho^\circ) + H(\dot{\rho} - \rho^\circ)$$

where H is the p by r matrix $H = (H_1', H_2', \ldots, H_M')'$; the i^{th} row of the p_α by r submatrix H_α is $h_{i\alpha} = \frac{1}{2}(\dot{\rho} - \rho^\circ)'\nabla_\rho^2 g_{i\alpha}(\bar{\rho})$ where $\bar{\rho}$ lies on the line segment joining $\dot{\rho}$ to ρ°. The asymptotic normality of $\sqrt{n}(g(\dot{\rho}) - \theta^\circ)$ is due to the fact that $\sqrt{n}(\dot{\rho} - \rho^\circ)$ is asymptotically normally distributed as shown in the proof of Theorem 2, the elements of $\nabla_p^{\circ} g(\bar{\rho})$ are bounded by continuity and the compactness of P, and $(\hat{\rho} - \rho^{\circ})$ converges almost surely to 0 by Theorem 1.

The last sentence of the Theorem follows from Theorem 1, Theorem 2, and the continuity of $G(\rho)$. ▯

5. COMPUTATIONAL CONSIDERATIONS

The minimization of $Q_\alpha(\underset{\sim}{\theta}_\alpha)$ to obtain the ordinary least squares estimators $\underset{\sim}{\theta}_\alpha$ and residual vectors $\hat{\underset{\sim}{e}}_\alpha$ may be carried out using either Hartley's (1961) Modified Gauss-Newton Method or Marquardt's (1963) Algorithm. Either of these methods may be used to minimize $Q(\underset{\sim}{\theta})$ by proceeding as follows. Factor $\hat{\underset{\sim}{\Sigma}}^{-1}$ to obtain a matrix $\underset{\sim}{R}$ such that $\hat{\underset{\sim}{\Sigma}}^{-1} = \underset{\sim}{R}'\underset{\sim}{R}$. Let $r_{\alpha\beta}$ denote the element of $\underset{\sim}{R}$ with row index α and column index β. Define new input vectors,

$$\underset{\sim}{w}_{t\alpha} = (r_{\alpha 1}, r_{\alpha 2}, \ldots, r_{\alpha M}, \underset{\sim}{x}'_{t1}, \underset{\sim}{x}'_{t2}, \ldots, \underset{\sim}{x}'_{tM})' \qquad (M + k \times 1) ,$$

new responses,

$$z_{t\alpha} = \Sigma_{\beta=1}^{M} r_{\alpha\beta}\, y_{t\beta} ,$$

and a new response function,

$$h(\underset{\sim}{w}_{t\alpha}, \underset{\sim}{\theta}) = \Sigma_{\beta=1}^{M} w_{\beta t\alpha}\, f_\beta(\underset{\sim}{x}_{t\beta}, \underset{\sim}{\theta}_\beta) = \Sigma_{\beta=1}^{M} r_{\alpha\beta}\, f_\beta(\underset{\sim}{x}_{t\beta}, \underset{\sim}{\theta}_\beta) .$$

When the model is transformed in this fashion it is in a form permitting the application of either Hartley's or Marquardt's algorithm to minimize

$$\Sigma_{\alpha=1}^{M} \Sigma_{t=1}^{n} (z_{t\alpha} - h(\underset{\sim}{w}_{t\alpha}, \underset{\sim}{\theta}))^2 = n\, Q(\underset{\sim}{\theta})$$

and obtain $\hat{\hat{\underset{\sim}{\theta}}}$. The vector $\hat{\underset{\sim}{\theta}}$ is a good start value for the iterations. Most implementations of either algorithm will print the matrix

$$(\underset{\sim}{P}'\underset{\sim}{P})^{-1} = (\Sigma_{\alpha=1}^{M} \Sigma_{t=1}^{n} \underset{\sim}{\nabla}\, h(\underset{\sim}{w}_{t\alpha}, \hat{\hat{\underset{\sim}{\theta}}})\, \underset{\sim}{\nabla}'\, h(\underset{\sim}{w}_{t\alpha}, \hat{\hat{\underset{\sim}{\theta}}}))^{-1} .$$

Since $\hat{\hat{\underset{\sim}{\Omega}}}^{-1} = n(\underset{\sim}{P}'\underset{\sim}{P})^{-1}$ the estimated variance-covariance matrix of $\sqrt{n}\,(\hat{\hat{\underset{\sim}{\theta}}} - \underset{\sim}{\theta}^\circ)$ may be obtained without additional computations.

The same transformation of the inputs, responses, and response function may be used to obtain $\hat{\hat{\rho}}$ minimizing $Q(g(\rho))$ as was used to obtain $\hat{\hat{\theta}}$ minimizing $Q(\theta)$. The only modification required is to substitute $g(\rho)$ for θ in the response function $h(w_{t\alpha}, \theta)$ obtaining $h(w_{t\alpha}, g(\rho))$. Be careful that the number of parameters supplied to the program are reduced from p to r and that derivatives are taken with respect to ρ. Since

$$\nabla_{\rho} h(w_{t\alpha}, g(\hat{\hat{\rho}})) = G'(\hat{\hat{\rho}}) \nabla h(w_{t\alpha}, g(\hat{\hat{\rho}}))$$

the matrix printed by the program

$$(P'P)^{-1} = (\Sigma_{\alpha=1}^{M} \Sigma_{t=1}^{n} \nabla_{\rho} h(w_{t\alpha}, g(\hat{\hat{\rho}})) \nabla_{\rho}' h(w_{t\alpha}, g(\hat{\hat{\rho}})))^{-1}$$

will satisfy

$$(\hat{\hat{G}}'\hat{\Omega}\hat{G})^{-1} = n\ (P'P)^{-1} .$$

Thus it may be used to estimate the variance-covariance matrix of $\sqrt{n}\,(\hat{\hat{\rho}} - \rho^{\circ})$.

Start values for the iterations may be obtained by finding ρ which will satisfy at least r of the equations in the system $\hat{\theta} = g(\rho)$.

The computations will be illustrated using the data presented in Table 1. These observations were generated according to the model

$$y_{t1} = \theta_{11} + \theta_{21} x_{1t1} + \theta_{31} e^{\theta_{41} x_{2t1}} + e_{t1} ,$$

$$y_{t2} = \theta_{12} + \theta_{22} e^{\theta_{32} x_{1t2}} + e_{t2} .$$

Using the Modified-Gauss Newton Method, one finds that the least squares estimator for the first equation is

$$\hat{\underset{\sim}{\theta}}_1 = (1.0127, 1.0077, .9903, -1.0263)' ,$$

and for the second equation it is

$$\hat{\underset{\sim}{\theta}}_2 = (1.0450, .9709, -1.0979)'.$$

The estimated variance-covariance matrix is

$$\hat{\underset{\sim}{\Sigma}} = \begin{pmatrix} .0008856 & .0004516 \\ .0004516 & .0008258 \end{pmatrix}$$

and its inverse may be factored as $\hat{\underset{\sim}{\Sigma}}^{-1} = \underset{\sim}{R}'\underset{\sim}{R}$ where

$$\underset{\sim}{R} = \begin{pmatrix} 39.57 & -21.64 \\ 0 & 34.80 \end{pmatrix} .$$

The transformed inputs are, therefore,

$$\underset{\sim}{w}_{t1} = (39.57, -21.64, x_{1t1}, x_{2t1}, x_{1t2})'$$

$$\underset{\sim}{w}_{t2} = (0, 34.80, x_{1t1}, x_{2t1}, x_{1t2})' ,$$

Table 1. Inputs and Responses

t	y_{t1}	x_{1t1}	x_{2t1}	y_{t2}	x_{1t2}
1	1.98803	0	0	2.05446	0
2	1.74975	0	.25	1.73545	.25
3	1.56687	0	.5	1.58786	.5
4	1.48967	0	.75	1.49833	.75
5	1.40154	0	1	1.37966	1
6	2.28432	.25	0	2.02124	0
7	1.96907	.25	.25	1.74193	.25
8	1.83365	.25	.5	1.60170	.5
9	1.66293	.25	.75	1.43230	.75
10	1.61976	.25	1	1.34858	1
11	2.51247	.5	0	2.00361	0
12	2.31425	.5	.25	1.80896	.25
13	2.08653	.5	.5	1.55098	.5
14	2.05508	.5	.75	1.48081	.75
15	1.84513	.5	1	1.33590	1
16	2.74747	.75	0	1.96754	0
17	2.52438	.75	.25	1.76454	.25
18	2.37002	.75	.5	1.63567	.5
19	2.21222	.75	.75	1.41481	.75
20	2.09650	.75	1	1.36679	1
21	2.98463	1	0	1.98960	0
22	2.77650	1	.25	1.76581	.25
23	2.60803	1	.5	1.57198	.5
24	2.49681	1	.75	1.47238	.75
25	2.42138	1	1	1.40579	1
26	2.05712	0	0	2.04711	0
27	1.82886	0	.25	1.78453	.25
28	1.62770	0	.5	1.65160	.5
29	1.48741	0	.75	1.48088	.75
30	1.33268	0	1	1.38178	1
31	2.28643	.25	0	2.02916	0
32	2.08228	.25	.25	1.80538	.25
33	1.82655	.25	.5	1.60661	.5
34	1.68291	.25	.75	1.45213	.75
35	1.61076	.25	1	1.38909	1
36	2.47966	.5	0	1.96843	0
37	2.27283	.5	.25	1.75310	.25
38	2.09720	.5	.5	1.64080	.5
39	1.98836	.5	.75	1.48162	.75
40	1.85849	.5	1	1.37142	1
41	2.73098	.75	0	2.02424	0
42	2.54987	.75	.25	1.81938	.25
43	2.37716	.75	.5	1.62442	.5
44	2.25082	.75	.75	1.51076	.75
45	2.14113	.75	1	1.36210	1
46	2.99249	1	0	2.06032	0
47	2.78713	1	.25	1.82534	.25
48	2.63906	1	.5	1.60557	.5
49	2.48154	1	.75	1.48878	.75
50	2.36727	1	1	1.34132	1

the transformed responses are,

$$z_{t1} = 39.57\, y_{t1} - 21.64\, y_{t2}$$

$$z_{t2} = 34.80\, y_{t2} ,$$

and the new response function is

$$h(\underset{\sim}{w}_{t\alpha}, \underset{\sim}{\theta}) = w_{1t\alpha}(\theta_{11} + \theta_{21} w_{3t\alpha} + \theta_{31} e^{\theta_{41} w_{4t\alpha}})$$

$$+ w_{2t\alpha}(\theta_{12} + \theta_{22} e^{\theta_{32} w_{5t\alpha}}) .$$

This transformed model is in a suitable form for the use of the Modified Gauss-Newton Method. Using $\hat{\underset{\sim}{\theta}}$ as the start value, one obtains

$$\hat{\hat{\underset{\sim}{\theta}}} = (1.0119, 1.0096, .9902, -1.0266, 1.0450, .9709, -1.0979)' .$$

and

$$(\underset{\sim}{P}'\underset{\sim}{P})^{-1} = \begin{pmatrix} .006368 & .000051 & -.006047 & -.010380 & .002844 & -.002692 & -.005237 \\ .000051 & .000102 & 0 & 0 & 0 & 0 & 0 \\ -.006047 & 0 & .005833 & .009781 & -.002716 & .002606 & .004939 \\ -.010380 & 0 & .009781 & .017418 & -.004639 & .004332 & .008774 \\ .002844 & 0 & -.002716 & -.004639 & .004576 & -.004339 & -.008394 \\ -.002692 & 0 & .002606 & .004332 & -.004339 & .004180 & .007847 \\ -.005237 & 0 & .004939 & .008774 & -.008394 & .007847 & .015849 \end{pmatrix}$$

after four iterations.

Suppose, now, that we wish to impose the constraints

$$\theta_{11} = \theta_{12} ,$$

$$\theta_{31} = \theta_{22} ,$$

$$\theta_{41} = \theta_{32} .$$

The reparameterization corresponding to these restrictions is

$$g_1(\rho) = (\rho_1, \rho_2, \rho_3, \rho_4)'$$

$$g_2(\rho) = (\rho_1, \rho_3, \rho_4)' .$$

The matrix of partial derivatives is

$$G(\rho) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

Putting $g(\rho) = \hat{\theta}$ and solving the first four equations for ρ we obtain the start values for the iterations

$$\rho = (1.0127, 1.0077, .9903, -1.0263)' .$$

After five iterations of the Modified-Gauss Newton Method one obtains

$$\hat{\hat{\rho}} = (1.0315, 1.0040, .9792, -1.0650)$$

$$(P'P)^{-1} = \begin{pmatrix} .004143 & -.000009 & -.003936 & -.007214 \\ -.000009 & .000040 & 0 & 0 \\ -.003936 & 0 & .003794 & .006769 \\ -.007214 & 0 & .006769 & .012913 \end{pmatrix} .$$

6. ASYMPTOTIC EFFICIENCY

In Lemma A.2 it is shown that $\sqrt{n}\,(\hat{\theta} - \theta^{\circ})$ converges in distribution to a p-variate normal with mean 0 and variance-covariance matrix $V^{-1} T \, V^{-1}$ where

$$V = \begin{pmatrix} V_{11} & 0 & \cdots & 0 \\ 0 & V_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & V_{MM} \end{pmatrix}$$

and

$$T = \begin{pmatrix} \sigma_{11}V_{11} & \sigma_{12}V_{12} & \cdots & \sigma_{1M}V_{1M} \\ \sigma_{22}V_{21} & \sigma_{22}V_{22} & \cdots & \sigma_{2M}V_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma_{M1}V_{M1} & \sigma_{M2}V_{M2} & \cdots & \sigma_{MM}V_{MM} \end{pmatrix} .$$

We have seen in Theorem 2 that $\sqrt{n}\,(\hat{\hat{\theta}} - \theta^{\circ})$ converges in distribution to a p-variate normal with mean 0 and variance-covariance matrix Ω^{-1}.

The Aitken improved estimator $\hat{\hat{\theta}}$ is asymptotically more efficient than the equation by equation least squares estimator $\hat{\theta}$ in the sense that $V^{-1} T \, V^{-1} - \Omega^{-1}$ is a positive semi-definite matrix. This is proven as follows. By Lemma A.3 $V^{-1} T \, V^{-1}$ is the limit of $(\frac{1}{n} F'F)^{-1}(\frac{1}{n} F'(\Sigma \otimes I)F)(\frac{1}{n} F'F)^{-1}$ while Ω^{-1} is the limit of $(\frac{1}{n} F'(\Sigma^{-1} \otimes I)F)^{-1}$. The difference $n[(F'F)^{-1}(F'(\Sigma \otimes I)F)(F'F)^{-1} - (F'(\Sigma^{-1} \otimes I)F)^{-1}]$ is positive semi-definite by Aitken's Theorem

(Theil, 1971, p. 238) so the limit $\underset{\sim}{V}^{-1}\underset{\sim}{T}\,\underset{\sim}{V}^{-1} - \underset{\sim}{\Omega}^{-1}$ is positive semi-definite.

There is one situation, likely to occur in a designed experiment, where the Aitken improvement of $\hat{\underset{\sim}{\theta}}$ does not result in a gain in asymptotic efficiency. When the response functions $f_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha)$ all have the same functional form and the inputs $\underset{\sim}{x}_{t\alpha}$ are the same for $\alpha = 1, 2, \ldots, M$ the matrix $\underset{\sim}{\Omega}^{-1} = \underset{\sim}{V}^{-1}\underset{\sim}{T}\,\underset{\sim}{V}^{-1}$. This follows directly from the fact that, in this case, $\underset{\sim}{V}_{\alpha\beta}$ is the same matrix, say $\underset{\sim}{W}$, for $\alpha, \beta = 1, 2, \ldots, M$. Thus,

$$\underset{\sim}{V}^{-1}\underset{\sim}{T}\,\underset{\sim}{V}^{-1} = (\underset{\sim}{I} \otimes \underset{\sim}{W}^{-1})(\underset{\sim}{\Sigma} \otimes \underset{\sim}{W})(\underset{\sim}{I} \otimes \underset{\sim}{W}^{-1}) = (\underset{\sim}{\Sigma} \otimes \underset{\sim}{W}^{-1}) = \underset{\sim}{\Omega}^{-1} .$$

In the case when the response functions and inputs are the same, the recommended procedure is to skip step three and estimate the variance-covariance matrix of $\sqrt{n}\,(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}^{\circ})$ by $\hat{\underset{\sim}{\Omega}}^{-1} = (\frac{1}{n}\underset{\sim}{F}'(\hat{\underset{\sim}{\theta}})(\hat{\underset{\sim}{\Sigma}}^{-1} \otimes \underset{\sim}{I})\underset{\sim}{F}(\hat{\underset{\sim}{\theta}}))^{-1}$.

7. MONTE-CARLO SIMULATION

This section reports the results of a Monte-Carlo study which was undertaken to gain an indication of the adequacy of the asymptotic theory in applications. Briefly, the study indicates the need for some conservatism in setting confidence intervals and testing hypotheses using the standard error estimates obtained in step four of the procedure. This is primarily due to evidence that these estimates understate the actual standard deviation of the estimators $\hat{\hat{\theta}}_{i\alpha}$ and, to some extent, due to departures in the shape of the small sample distributions of the $\hat{\hat{\theta}}_{i\alpha}$ from the limiting normal distributions.

The details of the study are as follows. Bivariate normal errors were generated with

$$\underset{\sim}{\Sigma} = \begin{pmatrix} .001 & .0005 \\ .0005 & .001 \end{pmatrix}$$

and added to the response functions of Section 5

$$f_1(\underset{\sim}{x}_{t1}, \underset{\sim}{\theta}_1) = \theta_{11} + \theta_{21}x_{1t1} + \theta_{31}e^{\theta_{41}x_{2t1}}$$

$$f_2(\underset{\sim}{x}_{t2}, \underset{\sim}{\theta}_2) = \theta_{12} + \theta_{22}e^{\theta_{32}x_{1t2}}$$

with

$$\underset{\sim}{\theta}^{\circ} = (1, 1, 1, -1, 1, 1, -1)'$$

using the inputs in Table 1. The replication scheme is "constant in repeated samples" of size twenty-five. The sample size n was taken as fifty yielding two replicates of each input. The results of the simulation are summarized in Tables 2 and 3.

Table 2. Location and Shape Parameters of the Sampling and Asymptotic Distributions

Parameter	Monte-Carlo				Asymptotic			
	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis
				Ordinary Least Squares				
θ_{11}	.9789	.1082	-1.0160	4.7282	1	.0892	0	3
θ_{21}	1.0001	.0124	- .0212	2.7972	1	.0126	0	3
θ_{31}	1.0211	.1039	1.0325	4.8957	1	.0855	0	3
θ_{41}	- .9942	.1530	.0251	2.9165	-1	.1407	0	3
θ_{12}	.9831	.0973	-1.0827	5.5266	1	.0890	0	3
θ_{22}	1.0164	.0939	1.0728	5.5581	1	.0855	0	3
θ_{32}	- .9952	.1378	.0684	3.0470	-1	.1407	0	3
				Seemingly Unrelated Regressions				
θ_{11}	.9792	.1079	-1.0171	4.7110	1	.0892	0	3
θ_{21}	.9996	.0106	.1084	2.8214	1	.0110	0	3
θ_{31}	1.0210	.1038	1.0249	4.8379	1	.0855	0	3
θ_{41}	- .9943	.1530	.0258	2.9141	-1	.1407	0	3
θ_{12}	.9832	.0972	-1.0780	5.5006	1	.0890	0	3
θ_{22}	1.0164	.0938	1.0673	5.5267	1	.0855	0	3
θ_{32}	- .9952	.1377	.0650	3.0510	-1	.1407	0	3
				Constrained Seemingly Unrelated Regressions				
ρ_1	.9850	.0883	- .9445	4.6118	1	.0771	0	3
ρ_2	1.0000	.0066	.0558	2.8926	1	.0069	0	3
ρ_3	1.0147	.0849	.9386	4.6365	1	.0740	0	3
ρ_4	- .9950	.1275	.0770	3.1294	-1	.1219	0	3

Table 3. Estimated Standard Errors

Parameter	Mean Estimated Standard Errors	Monte-Carlo Estimate of the Standard Error	Ratio
	Seemingly Unrelated Regressions		
θ_{11}	.0922	.1079	1.17
θ_{21}	.0104	.0106	1.02
θ_{31}	.0887	.1038	1.17
θ_{41}	.1335	.1530	1.15
θ_{12}	.0912	.0972	1.07
θ_{22}	.0878	.0938	1.07
θ_{32}	.1346	.1377	1.02
	Constrained Seemingly Unrelated Regressions		
ρ_1	.0782	.0883	1.13
ρ_2	.0067	.0066	.99
ρ_3	.0753	.0849	1.13
ρ_4	.1164	.1275	1.10

Table 2 conveys the impression that the bias, skewness, and peakedness of the small sample distributions of the estimators $\hat{\theta}_{i\alpha}$ or $\hat{\rho}_i$ are set in the ordinary least squares phase of the estimation procedure and are not appreciably modified in the later unrestricted or restricted Aitken phase. No general observations are possible. Some parameters are estimated with positive bias and some with negative bias, some distributions are left skewed and some right skewed, some are leptokurtic and some platykurtic. The only effect of the Aitken phase of the estimation procedure is to reduce the standard errors of the estimators.

When each single equation model is linear, the variance reduction in the Aitken phase is entirely due to linear restrictions across equations. Usually these restrictions take the form of knowledge that some parameters entering the model are zero. (See Zellner, 1962, p. 353). In a sense this is true in the nonlinear situation. We have seen previously that if the inputs $\underset{\sim}{x}_{t\alpha}$ are the same and each of the response functions $f_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha)$ have the same functional form then there is no asymptotic variance reduction in the Aitken phase of the estimation procedure. Only when the functional forms differ or restrictions across equations are imposed does an opportunity for variance reduction arise. The Monte-Carlo study indicates that these same considerations remain valid in the small sample distributions of the parameter estimators. The slight variation in functional form between $f_1(\underset{\sim}{x}_1, \underset{\sim}{\theta}_1)$ and $f_2(\underset{\sim}{x}_2, \underset{\sim}{\theta}_2)$ yields an improvement in efficiency. Interestingly, the asymptotic theory indicates an improvement for $\hat{\underset{\sim}{\theta}}_{21}$ only whereas the Monte-Carlo study indicates a slight gain in efficiency in the remaining parameters. The imposition of the restriction $\underset{\sim}{\theta} = \underset{\sim}{g}(\underset{\sim}{\rho})$ results in additional improvement.

Table 3 indicates that standard errors obtained from the diagonal elements of $(\hat{\hat{\underset{\sim}{\Omega}}})^{-1}$ or $(\hat{\underset{\sim}{G}}\,\hat{\underset{\sim}{\Omega}}\,\hat{\underset{\sim}{G}}')^{-1}$ as the case may be will lead one astray. In the absence of further evidence, it would not be an unreasonable practice to increase estimates of standard errors by the factor 1.10 before use in setting confidence intervals or testing hypotheses.

ACKNOWLEDGEMENT

The author wishes to express his appreciation to his colleagues Professors Thomas M. Gerig and T. Dudley Wallace for their comments and suggestions.

REFERENCES

[1] Aitken, A. C. (1935) "On least squares and linear combination of observations." Proceedings of the Royal Society of Edinburgh, 55, pp. 42-48.

[2] Gallant, A. R. (1973) "Inference for linear models." Institute of Statistics Mimeograph Series, No. 889. Raleigh: North Carolina State University.

[3] Hartley, H. O. (1961) "The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares." Technometrics, 3, pp. 269-280.

[4] Jennrich, R. I. (1969) "Asymptotic properties of nonlinear least squares estimators." Annals of Mathematical Statistics, 40, pp. 633-643.

[5] Malinvaud, E. (1970) "The consistency of nonlinear regressions." Annals of Mathematical Statistics, 41, pp. 956-969.

[6] Marquardt, D. W. (1963) "An algorithm for least squares estimation of nonlinear parameters." Journal of Industrial Applications Mathematics, 11, pp. 431-441.

[7] Rao, C. R. (1965) Linear Statistical Inference and Its Applications. New York: John Wiley and Sons, Inc.

[8] Theil, H. (1971) Principles of Econometrics. New York: John Wiley and Sons, Inc.

[9] Zellner, A. (1962) "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bais." Journal of the American Statistical Association, 57, pp. 348-368.

APPENDIX: LEMMAS

The appendix contains a sequence of technical lemmas which are used to obtain the asymptotic properties of the estimators $\hat{\hat{\theta}}$ and $\hat{\hat{\rho}}$ in Section 4 and to compare the asymptotic efficiency of $\hat{\theta}$ with $\hat{\hat{\theta}}$ in Section 6. The almost sure convergence of $\hat{\theta}$ to θ° is given in Lemma A.2; the asymptotic distribution of $\hat{\theta}$ is given by Lemma A.7. The almost sure convergence of $\hat{\Sigma}^{-1}$ to Σ^{-1} and the rate of convergence are given by Lemma A.5. The remaining lemmas establish those implications of the weak convergence of the measures $\{\mu_n\}$ which are useful in Sections 4 and 6.

<u>Lemma A.1</u>: Let K be a compact subset of the m-dimensional real numbers for some integer $m \geq 1$ and let $g(x, \tau)$ be continuous on $X \times K$. Under the Assumptions of Section 3, the sequence of integrals $\int g(x, \tau)\, d\mu_n(x)$ converges to the integral $\int g(x, \tau)\, d\mu(x)$ uniformly in τ over K.

<u>Proof</u>: Malinvaud (1970, p. 967). ▯

<u>Lemma A.2</u>: Under the Assumptions of Section 3, the least squares estimators $\hat{\theta}_\alpha$ converge almost surely to θ°_α and are of the form

$$\hat{\theta}_\alpha = \theta^\circ_\alpha + (F'_\alpha F_\alpha)^{-1} F'_\alpha e_\alpha + a_{\alpha n}$$

where $\sqrt{n}\,\underset{\sim}{a}_{\alpha n}$ converges in probability to zero. Moreover, $\sqrt{n}(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}^{o})$ converges in distribution to a p_α-variate normal with mean $\underset{\sim}{0}$ and variance-covariance matrix $\sigma_{\alpha\alpha}\underset{\sim}{V}_{\alpha\alpha}^{-1}$.

<u>Proof</u>: Let $\mathcal{a}_\alpha$ be the Borel subsets of $\underset{\sim}{X}_\alpha$ and A_α a set in $\mathcal{a}_\alpha$. Define $\mu_{\alpha n}(A_\alpha) = n^{-1}\Sigma_{t=1}^{n} I_{A_\alpha}(\underset{\sim}{x}_{t\alpha})$. Set $A = A_\alpha \times X_{\beta\neq\alpha}^{M} \underset{\sim}{X}_\beta$ and set $\mu_\alpha(A_\alpha) = \mu(A)$. The measures $\mu_{\alpha n}$ converge weakly to the measure μ_α so that Assumption 6 of Gallant (1973) is satisfied. If $g(\underset{\sim}{x}) = 0$ a.e. μ_α then $g(\underset{\sim}{x}) = 0$ a.e. μ so that Assumption 7 is satisfied. Since $\underset{\sim}{\Omega}$ is non-singular $\underset{\sim}{V}_{\alpha\alpha}$ must be non-singular and Assumption 11 is satisfied. Assumptions 1, 2, 4, 5, 9, 10, 14, and 15 of Gallant (1973) have been assumed directly and so that the Lemma follows from Theorems 2, 3, 4, and 5 of Gallant (1973). ▯

<u>Lemma A.3</u>: Let the Assumptions of Section 3 hold and let $\underset{\sim}{\theta}_\alpha^{*}$ converge almost surely to $\underset{\sim}{\theta}_\alpha^{o}$ and $\underset{\sim}{\theta}_\beta^{*}$ converge almost surely to $\underset{\sim}{\theta}_\beta^{o}$. Then $n^{-1}\underset{\sim}{F}_\alpha'(\underset{\sim}{\theta}_\alpha^{*})\,\underset{\sim}{F}_\beta(\underset{\sim}{\theta}_\beta^{*})$ converges almost surely to $\underset{\sim}{V}_{\alpha\beta}$ for $\alpha,\beta = 1, 2, \ldots, M$.

<u>Proof</u>: The ijth element of $n^{-1}\underset{\sim}{F}_\alpha'(\underset{\sim}{\theta}_\alpha)\,\underset{\sim}{F}_\beta(\underset{\sim}{\theta}_\beta)$ may be written as

$$v_{ijn}(\underset{\sim}{\theta}_\alpha, \underset{\sim}{\theta}_\beta) = \int \frac{\partial}{\partial\theta_{i\alpha}} f_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha) \frac{\partial}{\partial\theta_{j\beta}} f_\beta(\underset{\sim}{x}_\beta, \underset{\sim}{\theta}_\beta)\, d\mu_n(\underset{\sim}{x}) \;.$$

We may apply Lemma A.1 since $\underset{\sim}{\Theta}_\alpha \times \Theta_\beta$ is compact so that $v_{ijn}(\underset{\sim}{\theta}_\alpha, \underset{\sim}{\theta}_\beta)$ has the uniform limit

$$v_{ij}(\underset{\sim}{\theta}_\alpha, \underset{\sim}{\theta}_\beta) = \int \frac{\partial}{\partial\theta_{i\alpha}} f_\alpha(\underset{\sim}{x}_\alpha, \underset{\sim}{\theta}_\alpha) \frac{\partial}{\partial\theta_{j\beta}} f_\beta(\underset{\sim}{x}_\beta, \underset{\sim}{\theta}_\beta)\, d\mu(\underset{\sim}{x})$$

which is continuous on $\underset{\sim}{\Theta}_\alpha \times \Theta_\beta$ being the uniform limit of continuous functions. The difference $v_{ijn}(\underset{\sim}{\theta}_\alpha^*, \underset{\sim}{\theta}_\beta^*) - v_{ij}(\underset{\sim}{\theta}_\alpha^*, \underset{\sim}{\theta}_\beta^*)$ converges almost surely to zero by uniform convergence and the difference $v_{ij}(\underset{\sim}{\theta}_\alpha^*, \underset{\sim}{\theta}_\beta^*) - v_{ij}(\underset{\sim}{\theta}_\alpha^\circ, \underset{\sim}{\theta}_\beta^\circ)$ converges almost surely to zero by continuity. The sum of these two differences is $v_{ijn}(\underset{\sim}{\theta}_\alpha^*, \underset{\sim}{\theta}_\beta^*) - v_{ij}(\underset{\sim}{\theta}_\alpha^\circ, \underset{\sim}{\theta}_\beta^\circ)$ which, therefore, converges almost surely to zero. ▯

<u>Lemma A.4</u>: Let $\underset{\sim}{K}$ be a compact subset of the m-dimensional real numbers for some integer $m \geq 1$. Let the Assumptions of Section 3 hold and let $g(\underset{\sim}{x}, \underset{\sim}{\tau})$ be continuous on $\underset{\sim}{X} \times \underset{\sim}{K}$. Then for almost all realizations of $\{e_{t\alpha}\}_{t=1}^{\infty}$ the series $n^{-1}\Sigma_{t=1}^{n} g(\underset{\sim}{x}_t, \underset{\sim}{\tau})\, e_{t\alpha}$ converges to zero uniformly in $\underset{\sim}{\tau}$ over $\underset{\sim}{K}$.

<u>Proof</u>: By Lemma A.1, $n^{-1}\,\Sigma_{t=1}^{n}\, g^2(\underset{\sim}{x}_t, \underset{\sim}{\tau})$ converges to $\int g^2(\underset{\sim}{x}, \underset{\sim}{\tau})\, d\mu(\underset{\sim}{x})$ uniformly in $\underset{\sim}{\tau}$ over $\underset{\sim}{K}$. The Lemma follows from Theorem 4 of Jennrich (1969). ▯

<u>Lemma A.5</u>: Let the Assumptions of Section 3 hold. Then the estimators $\hat{\sigma}_{\alpha\beta} = n^{-1}\, \hat{e}_\alpha'\, \hat{e}_\beta$ converge almost surely to the corresponding elements $\sigma_{\alpha\beta}$ of $\underset{\sim}{\Sigma}$. For n sufficiently large, the matrix $\hat{\underset{\sim}{\Sigma}}$ with elements $\hat{\sigma}_{\alpha\beta}$ is non-singular except on an event occurring with probability zero. The elements $\hat{\sigma}^{\alpha\beta}$ of $\hat{\underset{\sim}{\Sigma}}^{-1}$ converge almost surely to the corresponding elements $\sigma^{\alpha\beta}$ of $\underset{\sim}{\Sigma}^{-1}$. Moreover, $\sqrt{n}(\hat{\sigma}^{\alpha\beta} - \sigma^{\alpha\beta})$ is bounded in probability; that is, given $\delta > 0$ there is a bound M such that for all n sufficiently large we have $P[\sqrt{n}|\hat{\sigma}^{\alpha\beta} - \sigma^{\alpha\beta}| < M] > 1-\delta$.

<u>Proof</u>: The estimator may be rewritten as

$$\hat{\sigma}_{\alpha\beta} = \frac{1}{n}\, \underset{\sim}{e}_\alpha' \underset{\sim}{e}_\beta + \frac{1}{n}\, \underset{\sim}{\delta}_\alpha'(\hat{\underset{\sim}{\theta}}_\alpha) \underset{\sim}{e}_\beta + \frac{1}{n}\, \underset{\sim}{\delta}_\beta'(\hat{\underset{\sim}{\theta}}_\beta) \underset{\sim}{e}_\alpha + \frac{1}{n}\, \underset{\sim}{\delta}_\alpha'(\hat{\underset{\sim}{\theta}}_\alpha)\, \underset{\sim}{\delta}_\beta(\hat{\underset{\sim}{\theta}}_\beta) .$$

where $\underset{\sim}{\delta}_\alpha(\underset{\sim}{\theta}_\alpha) = \underset{\sim}{f}_\alpha(\underset{\sim}{\theta}_\alpha^\circ) - \underset{\sim}{f}_\alpha(\underset{\sim}{\theta}_\alpha)$.

The first term $\frac{1}{n} \underset{\sim}{e}'_{\alpha} \underset{\sim}{e}_{\beta}$ converges almost surely to $\sigma_{\alpha\beta}$ by the Strong Law of Large Numbers. The second term $\frac{1}{n} \underset{\sim}{\delta}_{\alpha}(\hat{\underset{\sim}{\theta}}_{\alpha})\underset{\sim}{e}_{\beta}$ converges almost surely to zero by Lemma A.4. Similarly for the third term. Letting $\delta_{\alpha}(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}_{\alpha}) = f(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}^{o}_{\alpha}) - f(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}_{\alpha})$ we may write

$$w_{\alpha\beta_n}(\underset{\sim}{\theta}_{\alpha}, \underset{\sim}{\theta}_{\beta}) = \frac{1}{n} \underset{\sim}{\delta}_{\alpha}(\underset{\sim}{\theta}_{\alpha}) \underset{\sim}{\delta}_{\beta}(\underset{\sim}{\theta}_{\beta}) = \int \delta_{\alpha}(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}_{\alpha}) \delta_{\beta}(\underset{\sim}{x}_{\beta}, \underset{\sim}{\theta}_{\beta}) \, d\mu_n(\underset{\sim}{x})$$

which, by Lemma A.1, converges uniformly in $(\underset{\sim}{\theta}'_{\alpha}, \underset{\sim}{\theta}'_{\beta})'$ over the compact set $\underset{\sim}{\Theta}_{\alpha} \times \underset{\sim}{\Theta}_{\beta}$ to

$$w_{\alpha\beta}(\underset{\sim}{\theta}_{\alpha}, \underset{\sim}{\theta}_{\beta}) = \int \delta_{\alpha}(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}_{\alpha}) \delta_{\beta}(\underset{\sim}{x}_{\beta}, \underset{\sim}{\theta}_{\beta}) \, d\mu(\underset{\sim}{x}) .$$

By the same arguments employed in Lemma A.3 we have that $w_{\alpha\beta n}(\hat{\underset{\sim}{\theta}}_{\alpha}, \hat{\underset{\sim}{\theta}}_{\beta})$ converges almost surely to $w_{\alpha\beta}(\underset{\sim}{\theta}^{o}_{\alpha}, \underset{\sim}{\theta}^{o}_{\beta}) = 0$.

The eventual non-singularity of $\hat{\underset{\sim}{\Sigma}}$ follows from the continuity of $\det(\underset{\sim}{\Sigma})$ and the assumption that $\underset{\sim}{\Sigma}$ is positive definite. The almost sure convergence of $\hat{\underset{\sim}{\Sigma}}^{-1}$ to $\underset{\sim}{\Sigma}^{-1}$ follows from the standard continuity argument.

By Taylor's theorem, for given $\underset{\sim}{\theta}_{\alpha}$ in $\underset{\sim}{S}_{\alpha}$ and $\underset{\sim}{x}$ in $\underset{\sim}{X}$ there is a $\bar{\underset{\sim}{\theta}}_{\alpha}$ on the line segment joining $\underset{\sim}{\theta}_{\alpha}$ and $\underset{\sim}{\theta}^{o}_{\alpha}$ such that

$$f_{\alpha}(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}_{\alpha}) = f_{\alpha}(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}^{o}_{\alpha}) + \underset{\sim}{\nabla}'_{\alpha} f_{\alpha}(\underset{\sim}{x}_{\alpha}, \bar{\underset{\sim}{\theta}}_{\alpha})(\underset{\sim}{\theta}_{\alpha} - \underset{\sim}{\theta}^{o}_{\alpha}) .$$

Let $\underset{\sim}{A}_{\alpha}$ be the n by p_{α} matrix with t^{th} row $\underset{\sim}{\nabla}'_{\alpha} f_{\alpha}(\underset{\sim}{x}_{t\alpha}, \bar{\underset{\sim}{\theta}}_{\alpha})$ so that

$$\begin{aligned} \sqrt{n}|\hat{\sigma}_{\alpha\beta} - \sigma_{\alpha\beta}| \le \sqrt{n}|\tfrac{1}{n} e'_{\alpha} e_{\beta} - \sigma_{\alpha\beta}| &+ |\tfrac{1}{n} \underset{\sim}{e}'_{\beta} \underset{\sim}{A}_{\alpha}(\hat{\underset{\sim}{\theta}}_{\alpha} - \underset{\sim}{\theta}^{o}_{\alpha}) \sqrt{n}| \\ &+ |\tfrac{1}{n} \underset{\sim}{e}'_{\alpha} \underset{\sim}{A}_{\beta}(\hat{\underset{\sim}{\theta}}_{\beta} - \underset{\sim}{\theta}^{o}_{\beta}) \sqrt{n}| \\ &+ |\sqrt{n} (\hat{\underset{\sim}{\theta}}_{\alpha} - \underset{\sim}{\theta}^{o}_{\alpha})(\tfrac{1}{n} \underset{\sim}{A}'_{\alpha} \underset{\sim}{A}_{\beta})(\hat{\underset{\sim}{\theta}}_{\beta} - \underset{\sim}{\theta}^{o}_{\beta})| . \end{aligned}$$

The first term on the right is bounded in probability because $\sqrt{n}(\frac{1}{n} \underset{\sim}{e}_\alpha' \underset{\sim}{e}_\beta - \sigma_{\alpha\beta})$ is asymptotically normally distributed by the Central Limit Theorem. The second term on the right converges in probability to zero because $\sqrt{n}(\hat{\underset{\sim}{\theta}}_\alpha - \underset{\sim}{\theta}_\alpha^\circ)$ is asymptotically normally distributed by Lemma A.2 and $\frac{1}{n} \underset{\sim}{e}_\beta' \underset{\sim}{A}_\alpha$ converges almost surely to zero by Lemma A.4. Similarly for the third term on the right. The fourth term converges in probability to zero because $\sqrt{n}\ (\hat{\underset{\sim}{\theta}}_\alpha - \underset{\sim}{\theta}_\alpha^\circ)$ is asymptotically normally distributed, $\hat{\underset{\sim}{\theta}}_\beta - \underset{\sim}{\theta}_\beta^\circ$ converges almost surely to zero, and the elements of $\frac{1}{n} \underset{\sim}{A}_\alpha' \underset{\sim}{A}_\beta$ are uniformly bounded by

$$\sup\{\frac{\partial}{\partial\theta_{i\alpha}} f_\alpha(\underset{\sim}{x}_\alpha,\underset{\sim}{\theta}_\alpha) \frac{\partial}{\partial\theta_{j\beta}} f_\beta(\underset{\sim}{x}_\beta,\underset{\sim}{\theta}_\beta): \ (\underset{\sim}{x}_\alpha',\underset{\sim}{\theta}_\alpha',\underset{\sim}{x}_\beta',\underset{\sim}{\theta}_\beta')' \in \underset{\sim}{X}_\alpha \times \underset{\sim}{\Theta}_\alpha \times \underset{\sim}{X}_\beta \times \underset{\sim}{\Theta}_\beta\}$$

which is finite by the compactness of $\underset{\sim}{X}_\alpha \times \underset{\sim}{\Theta}_\alpha \times \underset{\sim}{X}_\beta \times \underset{\sim}{\Theta}_\beta$ and the continuity of the partial derivatives. Using a Taylor series expansion one can show that if $g(\underset{\sim}{\tau})$ is a function with continuous first derivatives on a bounded open sphere $\underset{\sim}{S}$, $P(\hat{\underset{\sim}{\tau}} \in \underset{\sim}{S})$ tends to one as n tends to infinity, and $\sqrt{n}\ (\hat{\underset{\sim}{\tau}} - \underset{\sim}{\tau}^\circ)$ is bounded in probability then $\sqrt{n}\ (g(\hat{\underset{\sim}{\tau}}) - g(\underset{\sim}{\tau}^\circ))$ is bounded in probability. The matrix $\underset{\sim}{\Sigma}$ is positive-definite so there is an $\epsilon > 0$ such that $\det(\underset{\sim}{\Sigma}) - \epsilon > 0$. Since $P(\det(\underset{\sim}{\Sigma}) - \epsilon < \det(\hat{\underset{\sim}{\Sigma}}) < \det(\underset{\sim}{\Sigma}) + \epsilon)$ tends to one as n tends to infinity we have that the elements of $\sqrt{n}\ (\hat{\underset{\sim}{\Sigma}}^{-1} - \underset{\sim}{\Sigma}^{-1})$ are bounded in probability. ∎

Lemma A.6: Under the Assumptions listed in Section 3, $\frac{1}{\sqrt{n}} \underset{\sim}{F}'(\underset{\sim}{\Sigma}^{-1} \otimes \underset{\sim}{I})\, \underset{\sim}{e}$ converges in distribution to a p-variate normal with mean $\underset{\sim}{0}$ and variance-covariance matrix $\underset{\sim}{\Omega}^{-1}$.

Proof: We will apply the Central Limit Theorem stated in Problem 4.7 of Rao (1965, p. 118). We may write

$$\frac{1}{\sqrt{n}} \underset{\sim}{F}'(\underset{\sim}{\Sigma}^{-1} \otimes \underset{\sim}{I})\, \underset{\sim}{e} = \frac{1}{\sqrt{n}} \Sigma_{t=1}^{n} \underset{\sim}{K}_t\, \underset{\sim}{\Sigma}^{-1}\, \underset{\sim}{u}_t = \frac{1}{\sqrt{n}} \Sigma_{t=1}^{n} \underset{\sim}{v}_t$$

where $\underset{\sim}{K}_t$ is the p by M matrix

$$\underset{\sim}{K}_t = \begin{pmatrix} \underset{\sim}{\nabla}_1 f_1(\underset{\sim}{x}_{t1}, \underset{\sim}{\theta}_1^{\circ}) & \underset{\sim}{0} & & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\nabla}_2 f_2(\underset{\sim}{x}_{t2}, \underset{\sim}{\theta}_2^{\circ}) & \cdots & \underset{\sim}{0} \\ \vdots & \vdots & & \vdots \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{\nabla}_M f_M(\underset{\sim}{x}_{tM}, \underset{\sim}{\theta}_M^{\circ}) \end{pmatrix}.$$

The $\underset{\sim}{v}_t$ are independent by the independence of the $\underset{\sim}{u}_t$ and $\frac{1}{n} \Sigma_{t=1}^{n} \mathcal{E}(\underset{\sim}{v}_t\, \underset{\sim}{v}_t') = \frac{1}{n} \Sigma_{t=1}^{n} \underset{\sim}{K}_t\, \underset{\sim}{\Sigma}^{-1} \underset{\sim}{K}_t'$ which converges to $\underset{\sim}{\Omega}$ by the weak convergence of the measures μ_n. Let $G_t(\underset{\sim}{v})$ be the distribution function of $\underset{\sim}{v}_t$ and let $H(\underset{\sim}{w})$ be the distribution function of $\underset{\sim}{w}_t = \underset{\sim}{\Sigma}^{-1} \underset{\sim}{u}_t$ (which does not depend on t). For given $\epsilon > 0$

$$0 \le \frac{1}{n} \Sigma_{t=1}^{n} \int_{A_{tn}} \|\underset{\sim}{v}\|^2 \, dG_t(\underset{\sim}{v})$$

where $A_{tn} = \{\underset{\sim}{v}: \ ||\underset{\sim}{v}|| > \sqrt{n}\,\epsilon\}$

$$= \frac{1}{n}\,\Sigma_{t=1}^{n} \int_{B_{tn}} \Sigma_{\alpha=1}^{M} \,||\underset{\sim}{\nabla}_{\alpha} f_{\alpha}(\underset{\sim}{x}_{t\alpha}, \underset{\sim}{\theta}_{\alpha}^{o})||^2 \, w_{\alpha}^2 \, dH(\underset{\sim}{w})$$

where $B_{tn} = \{\underset{\sim}{w}: \ \Sigma_{\alpha=1}^{M} \,||\underset{\sim}{\nabla}_{\alpha} f_{\alpha}(\underset{\sim}{x}_{t\alpha}, \underset{\sim}{\theta}^{o})||^2 w_{\alpha}^2 > n\,\epsilon^2\}$

$$\leq \frac{1}{n}\,\Sigma_{t=1}^{n} \int_{B_{tn}} K\, \underset{\sim}{w}'\underset{\sim}{w} \, dH(\underset{\sim}{w})$$

where $K = \sup_{\alpha} \sup_{\underset{\sim}{X}_{\alpha}} ||\underset{\sim}{\nabla}_{\alpha} f_{\alpha}(\underset{\sim}{x}_{\alpha}, \underset{\sim}{\theta}_{\alpha}^{o})||^2$ and is finite because $\underset{\sim}{X}_{\alpha}$ is compact

$$\leq K \int_{C_n} \underset{\sim}{w}'\underset{\sim}{w} \, dH(\underset{\sim}{w})$$

where $C_n = \{\underset{\sim}{w}: \ K\,\underset{\sim}{w}'\underset{\sim}{w} > n\,\epsilon^2\}$ because $B_{tn} \subset C_n$ for $t = 1, 2, \ldots, n$; this last term converges to zero as n tends to infinity because $\int \underset{\sim}{w}'\underset{\sim}{w} \, dH(\underset{\sim}{w})$ is finite. ▯

<u>Lemma A.7</u>: Under the Assumptions listed in Section 3, $\sqrt{n}\,(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}^{o})$ converges in distribution to a p-variate normal with mean $\underset{\sim}{0}$ and variance-covariance matrix $\underset{\sim}{V}^{-1}\underset{\sim}{T}\,\underset{\sim}{V}^{-1}$.

<u>Proof</u>: By Lemma A.2

$$\sqrt{n}\,(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}^{o}) = (\tfrac{1}{n}\,\underset{\sim}{F}'\underset{\sim}{F})^{-1} \frac{1}{\sqrt{n}}\,\underset{\sim}{F}'\underset{\sim}{e} + \sqrt{n}\,\underset{\sim}{a}$$

where $\sqrt{n}\,\underset{\sim}{a}$ converges in probability to $\underset{\sim}{0}$. The matrix $(\frac{1}{n}\,\underset{\sim}{F}'\underset{\sim}{F})^{-1}$ converges to $\underset{\sim}{V}^{-1}$ by Lemma A.3. Using arguments similar to those used in the proof of Lemma A.6' we have that $\frac{1}{\sqrt{n}} \underset{\sim}{F}' \underset{\sim}{e} = \frac{1}{\sqrt{n}} \Sigma_{t=1}^{n} \underset{\sim}{K}_t \underset{\sim}{u}_t$ converges in distribution to a p-variate normal with mean $\underset{\sim}{0}$ and variance-covariance matrix $\underset{\sim}{T}$. ‖