THE UNIVERSITY OF CALGARY

ROBUST ESTIMATION OF THE REGRESSION VECTOR IN THE LINEAR MODEL IN THE PRESENCE OF ASYMMETRY

by

JEROME N. SHEAHAN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

OCTOBER, 1979

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled

Robust Estimation of the Regression Vector in the Linear Model in the Presence of Asymmetry

submitted by Jerome Nicholas Sheahan in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

John R. Collins
J. R. Collins, Chairman
Department of Mathematics & Statistics

M. D. Burke
M. D. Burke
Department of Mathematics & Statistics

P. C. Consul
P. C. Consul
Department of Mathematics & Statistics

E. S. Edgington
E. S. Edgington
Department of Psychology

K. L. Mehra
K. L. Mehra, External Examiner
University of Alberta

October 26, 1979

ABSTRACT

A theory of estimation of the regression parameter vector in the linear model, suitable for asymmetric departures from a symmetric error model distribution, is presented.

Let F be the class of distributions that have the density $\phi(y) = (2\pi)^{-\frac{1}{2}} \exp(\frac{-y^2}{2})$ for $y \in [-d,d]$, where d is a specified number, and are arbitrary outside $[-d,d]$. This F reflects the type of departure from normality that is common in error distributions. Our model is

$$\underset{\sim}{X} = C\underset{\sim}{\theta} + \underset{\sim}{\varepsilon} ,$$

where $\underset{\sim}{X} = (X_1,\ldots,X_n)^T$ is the observation vector, $C = ((c_{ij}))$ is a given matrix of n rows and p columns, $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_p)^T$ is the unknown regression parameter vector to be estimated and $\underset{\sim}{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_p)^T$ where the ε_i are independent identically distributed random variables with distribution $G \in F$. Let Ψ_c be the class of smooth skew-symmetric functions that vanish outside $[-c,c]$ where c depends on d in a realistic fashion.

In the case where σ is known (say $\sigma = 1$), we estimate $\underset{\sim}{\theta}$ by solving the system

$$\sum_{i=1}^{n} c_{ik}\ \psi(X_i - \sum_{j=1}^{p} c_{ij}\ \theta_j) = 0\ ,\ k=1,\ \ldots,\ p \quad \text{where } \psi \in \Psi_c,$$

iteratively, using an appropriate initial value. We show that the resulting estimator $\underset{\sim}{T}_n = \underset{\sim}{T}_n(\psi)$ satisfies $\underset{\sim}{T}_n \xrightarrow{P} \underset{\sim}{\theta}$ and $n^{\frac{1}{2}}\ \underset{\sim}{T}_n$ is asymptotically multivariate normal with mean $\underset{\sim}{\theta}$ and covariance matrix given by

$$C_0^{-1} \frac{\int_{-c}^{c} \psi^2(y)\ \phi(y)dy}{\left[\int_{-c}^{c} \psi(y)\ \phi'(y)dy\right]^2}, \quad \text{where } C_0 = \lim \frac{C^TC}{n} .$$

The problem of identifying robust members of the class $\{\underset{\sim}{T}_n(\psi) : \psi \in \Psi_c\}$ is considered when $G \in F$ and also when G has, in addition, small contamination of its normal centre.

In the case of scale unknown, we proceed in two ways, both of which ensure scale invariance of our estimators of $\underset{\sim}{\theta}$. The first involves getting a prior estimate $\hat{\sigma}_n$ of σ and solving iteratively the system

$$\sum_{i=1}^{n} c_{ik}\ \psi\left(\frac{x_i - \sum_{j=1}^{p} c_{ij}\ \theta_j}{\hat{\sigma}_n}\right), \quad k=1,\ldots,p .$$

The second involves the simultaneous estimation of $\underset{\sim}{\theta}$ and σ by solving iteratively the system of $p + 1$ equations

$$\begin{cases} \sum_{i=1}^{n} c_{ik}\ \psi\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\ \theta_j}{\sigma}\right) = 0, \quad k=1,\ldots,p \\ \sum_{i=1}^{n} \left[\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\ \theta_j}{\sigma}\right) \psi\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\ \theta_j}{\sigma}\right) - \rho\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\ \theta_j}{\sigma}\right)\right] - a_n = 0 , \end{cases}$$

where $\rho(y) = \int_{-\infty}^{y} \psi(x)dx$

$a_n = (n-p)E[U\psi(U) - \rho(U)]$

and U has the standard normal distribution.

In both cases, we arrive at consistent and asymptotically normal estimators and optimal estimators are proposed.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Dr. J.R. Collins, without whose intuition, wealth of ideas, encouragement and careful guidance this research would not have been carried out.

I would also like to thank the members of my Examining Committee, Professor P.C. Consul, Professor K.L. Mehra, Dr. M.D. Burke and Dr. E.S. Edgington.

I am grateful to the University of Calgary for the award of a Ph.D. Dissertation Fellowship.

I would like to thank Mr. Grant Gilmour for the programming required to obtain some numerical results.

Finally my gratitude goes to Mrs. Maria Dourado, Mrs. Betty Teare and especially Mrs. Pat Dalgetty, each of whom cheerfully helped in typing the thesis.

To my parents

Patrick and Eileen Sheahan

TABLE OF CONTENTS

INTRODUCTION

A class of estimators of a location parameter was introduced by Huber (1964). The asymptotic properties of these estimators were studied and members of the class which are robust against symmetric departures from a symmetric model distribution were identified. Collins (1976) adapted Huber's theory to allow for asymmetric departures from the model distribution. Several authors have extended Huber's theory to the estimation of the regression parameter vector in the linear model. A serious inadequacy of the theory in the literature is either an assumption that the distribution of the errors is symmetric or else an assumption is made that is not much weaker than the assumption of symmetry. The purpose of this work is to present a theory of estimation of the regression parameter vector in the linear model that is suitable for asymmetric departures from a symmetric error model distribution. We first give some essential background material.

In the location problem, Huber's M-estimators are defined as solutions of equations of the form

$$(1.1)\quad \sum_{i=1}^{n} \psi(X_i-\theta) = 0$$

where ψ belongs to some suitable class of functions. If the distribution of the independent identically distributed random variables is only approximately known, Huber's minimax criterion is to choose that estimator of the location parameter which minimizes the supremum of the asymptotic variance over all distributions in a neighbourhood of a model distribution.

Collins (1976) considered the location problem when the distribution of X_i is an asymmetric departure from a model distribution. We give some of the elements of Collins' work.

Let F be the class of distributions defined by

(1.2) $G \in F \Rightarrow$

$$G(y) = \begin{cases} \Phi(y), \; y \in [-d,d] \\ \text{arbitrary, otherwise,} \end{cases}$$

where Φ is the standard normal cumulative and

$$d = \Phi^{-1}(1 - \tfrac{\alpha}{2}) \text{ for some "reasonably small" } \alpha.$$

Let Ψ_c be the class of all mappings of $\mathbb{R}^1 \to \mathbb{R}^1$ such that

(1.3) $\psi \in \Psi_c \Rightarrow \psi$ is smooth, skew-symmetric, vanishes outside $[-c,c]$, is non-negative on $[0,c]$ and not identically zero on $[0,c]$.

The choice of c is determined judiciously - for example, the "influence" exerted on (1.1) by tail observations is cut to zero.

When $\psi \in \Psi_c$, (1.1) has multiple roots. Using a good starting value in the iterative solution of (1.1), Collins derived estimators $T_n = T_n(\psi)$ that are consistent and asymptotically normal. The asymptotic variance turned out to be

$$V(\psi,G) = \frac{\int_{-c}^{c} \psi^2(y) \;\; \phi(y)\,dy}{\left(\int_{-c}^{c} \psi(y) \;\; \phi'(y)\,dy\right)^2} . \tag{1.4}$$

where ϕ is the standard normal density function. Note that (1.4) is independent of $G \in F$. Its infimum was shown by Collins to be attained by

$$(1.5) \qquad \psi^*(x) = \begin{cases} x \text{ if } |x| \le c \\ 0 \quad \text{otherwise,} \end{cases}$$

so that ψ^* (or, equivalently, the estimator corresponding to ψ^*) is most robust in the sense of Huber's minimax criterion, but only formally since $\psi^* \notin \Psi_c$.

Then Collins extended the class F in (1.2) to consider, in addition to completely unknown tails, a small amount of symmetric contamination of the normal centre and solved the minimax variance problem. Collins (1977) gave mild necessary and sufficient conditions for the parameter to be identifiable in this model. (By identifiability of a parameter θ in a model $\{G(x-\theta),\ G \in H,\ H$ some class of distributions$\}$, we mean that there do not exist θ_1, θ_2, $\theta_1 \neq \theta_2$ and $G_1, G_2 \in H$ such that

$$G_1(x-\theta_1) = G_2(x-\theta_2) \text{ for all } x.)$$

Collins (1976) extended the results to the case where an unknown scale parameter is present in the model.

Now consider the linear model

$$(1.6) \qquad X_i = \sum_{j=1}^{p} c_{ij}\, \theta_j + \varepsilon_i\ , \quad i=1,\ldots,n$$

where

$\underset{\sim}{X} = (X_1,\ldots,X_n)^T$ is the observation vector (T denotes transpose), $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_p)^T$ is the unknown regression parameter vector to be estimated, $C = ((c_{ij}))$ is the design matrix and

$\underset{\sim}{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)^T$ is the error vector. The ε_i are assumed to be independent identically distributed random variables.

In estimating $\underset{\sim}{\theta}$, there are, of course, several assumptions that can be violated, e.g., the error distribution may have longer tails than supposed (this can be caused by a few grossly erroneous observations, for example), the model may not be quite linear, there may be deviations from the assumption of independence of the errors, systematic inhomogeneity of variance, etc.. In our work, we shall be concerned with distributional robustness, i.e., in deriving estimators that behave well under small changes of the distribution of the errors.

The classical solution to the problem of estimating $\underset{\sim}{\theta}$ is to minimize the sum of squares:

$$\sum_{i=1}^{n} (X_i - \sum_{j=1}^{p} c_{ij}\, \theta_j)^2 = \min! \tag{1.7}$$

or, equivalently, to solve the system

$$\sum_{i=1}^{n} c_{ik}\, (X_i - \sum_{j=1}^{p} c_{ij}\, \theta_j) = 0\ , \quad k=1,\ldots,p\ . \tag{1.8}$$

This classical approach is highly sensitive to heavy tails in the distribution of the errors. The resulting estimator may not be consistent and is not efficient. Note that, by the Gauss-Markov theorem, (robust) alternative estimators must be non-linear in the observations.

One method studied by Relles (1968), Huber (1973) and Yohai (1972) is to replace the square function in (1.7) by some less rapidly increasing function ρ. The resulting family of estimators (Huber M-estimators) are then the solutions of

$$(1.9) \qquad \sum_{i=1}^{n} \rho(X_i - \sum_{j=1}^{p} c_{ij}\, \theta_j) = \min!$$

If ρ is convex and has a derivative ψ, (1.9) is equivalent to solving the system

$$(1.10) \qquad \sum_{i=1}^{n} c_{ik}\, \psi(X_i - \sum_{j=1}^{p} c_{ij}\, \theta_j) = 0 \ , \ k=1,\ldots,p.$$

The assumption that the error distribution be symmetric or that $E\psi(\varepsilon_i) = 0$, where E denotes the expectation operator, has been made. (The consistency condition $E\psi(\varepsilon_i) = 0$ is not easy to satisfy as ψ and the error distribution G range over some classes unless we assume G is symmetric.)

Under a variety of additional regularity conditions, various authors (Relles (1968), Huber (1973), Yohai and Maronna (1979)) have proved the consistency and asymptotic normality of the estimators derived from (1.10). Huber's robustness results carry through due to the form of the asymptotic covariance matrix. Bickel (1975) introduced one-step (M) estimators in the linear model. (these are solutions of a linear approximation to the system (1.10))and showed that their behaviour is much like the actual roots of (1.10). Results corresponding to those above were obtained when scale is also unknown. The case where p, the number of parameters, is allowed to increase with n, the number of observations, has also been treated, but in our work we shall consider only fixed p.

We shall take as our starting point the model (1.6) and extend the current theory of estimation of $\underset{\sim}{\theta}$ to allow for error distributions that are asymmetric departures from a symmetric model distribution. With $\psi \in \Psi_c$, where Ψ_c is given in (1.3), we consider the system of equations (1.10) (note that $\rho^{\underline{s}}$ corresponding to our $\psi^{\underline{s}}$ are not convex). Now, the system (1.10) has multiple roots when $\psi \in \Psi_c$. One procedure open to us is to solve (1.10) iteratively with some good starting value. This is the procedure we adopt, making a restriction on the design matrix C to give us our initial value. We consider first the case where $\mathrm{Var}(\varepsilon_i) = \sigma^2$ is known (so that, without loss of generality, $\sigma = 1$), deferring the case of unknown scale to Chapters 7 and 8. The distribution G of ε_i is first taken to be a member of F, with F given by (1.2). We obtain estimators $\underset{\sim}{T}_n = \underset{\sim}{T}_n(\psi)$ of $\underset{\sim}{\theta}$ that are consistent and we find that

$$(1.11) \qquad \underset{\sim}{T}_n \xrightarrow{\mathcal{D}} MVN\left(\underset{\sim}{\theta},\ C_0^{-1}\ \frac{\int_{-c}^{c} \psi^2(y)\phi(y)dy}{\left(\int_{-c}^{c} \psi(y)\phi'(y)dy\right)^2}\right),$$

where $C_0 = \lim_{n\to\infty} \frac{C^T C}{n}$, the limit being shown to exist and be positive definite.

The efficiencies are independent of the design matrix and (see (1.4)) the optimality results of Collins (1976) apply.

In the case of symmetric contamination of the normal centre, the minimax results of Collins in the location case apply to our linear model.

In the case of scale unknown and $G \in F$ we separate our treatment into two sections. In Section 7, we propose an estimator $\hat{\sigma}_n$ of σ which satisfies

$\hat{\sigma}_n \xrightarrow{p} \beta\sigma$ (following Collins (1976)) where the biasing factor β is unknown, but close to 1. Solutions of

$$\text{(1.12)} \qquad \sum_{i=1}^{n} c_{ik}\ \psi\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\ \theta_j}{\hat{\sigma}_n}\right) = 0\ ,\ k=1,\ldots,p$$

then yield scale invariant estimators of $\underset{\sim}{\theta}$. We solve (1.12) iteratively, with ψ in some appropriate class, using the same starting value as in the scale known case. We again derive a class of consistent and asymptotically normal estimators of $\underset{\sim}{\theta}$. The asymptotic covariance matrix of $n^{\frac{1}{2}}\ \underset{\sim}{T}_n$ is given by

$$\text{(1.13)} \qquad C_0^{-1}\ \frac{\int_{-c}^{c} \psi^2(y)\ \phi(\beta y)dy}{\beta\left[\int_{-c'}^{c'} \psi(y)\ \phi'(\beta y)dy\right]^2}\ .$$

In Section 8, we ensure scale invariance by solving a certain system of equations simultaneously for $\underset{\sim}{\theta}$ and σ. This system is

$$\text{(1.14)} \qquad \begin{cases} \displaystyle\sum_{i=1}^{n} c_{ik}\ \psi\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\ \theta_j}{\sigma}\right) = 0\ ,\quad k=1,\ldots,p \\ \displaystyle\sum_{i=1}^{n}\left[\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\theta_j}{\sigma}\right)\psi\left(\frac{Xi - \sum_{j=1}^{p} c_{ij}\theta_j}{X_i\ \ \sigma}\right) - \rho\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\theta_j}{\sigma}\right)\right] - a_n = 0 \end{cases}$$

where $\rho(y) = \int_{-\infty}^{y} \psi(x)dx$,

$a_n = (n-p)E[U\psi(U)-\rho(U)]$

and where

U has the standard normal distribution, or, what turns out to be equivalent as far as the value of a_n is concerned, U has the standard normal distribution in $[-d,d]$ and arbitrary outside $[-d,d]$. We use the same starting value for $\underset{\sim}{\theta}$ as in Section 7 and the scale known case, and we use the $\hat{\sigma}_n$ of Section 7 as our starting value for σ. We contend that for n finite the estimators of $\underset{\sim}{\theta}$ derived from (1.14) are superior to the corresponding estimators of $\underset{\sim}{\theta}$ from (1.12) because, instead of using a fixed estimator of σ throughout the iteration process, as (1.12) does, we, in (1.14), improve our initial estimator of σ at each step of the iteration process. Finally, optimal estimators in the scale unknown case are proposed.

2

BASIC MATHEMATICAL CONCEPTS AND RESULTS

In this short section, we will review some basic concepts and results from linear algebra and p-dimensional calculus, for we shall have recourse to use them frequently in our work. (See e.g., Simmons (1963) and Ortega and Rheinboldt (1970).)

We first recall that any two norms, $\|\cdot\|$ and $\|\cdot\|'$ on a finite-dimensional linear space N generate the same topology. This is equivalent to saying that there exist constants $c_2 \geq c_1 > 0$ such that

$$(2.1) \quad c_1\|x\| \leq \|x\|' \leq c_2\|x\| \quad \text{for all } x \in N .$$

This usually permits us to use an arbitrary norm in our work and we will do so, except where we specify otherwise.

Let $L(\mathbb{R}^p)$ denote the (linear space) of all linear operators on $\mathbb{R}^p$, where $\mathbb{R}^p$ is the set of all (ordered) p-tuples of real numbers. $L(\mathbb{R}^p)$ becomes a normed linear space under the operator norm:

$$(2.2) \quad \|E\| = \sup_{\|\underset{\sim}{t}\|=1} \|E\underset{\sim}{t}\| , \quad E \in L(\mathbb{R}^p)$$

and this equals

$$\sup_{\|\underset{\sim}{t}\|\leq 1} \|E\underset{\sim}{t}\| = \sup_{\|\underset{\sim}{t}\|<1} \|E\underset{\sim}{t}\|$$

$$= \inf\{K:K \geq 0 \text{ and } \|E\underset{\sim}{t}\| \leq K\|\underset{\sim}{t}\| \text{ for all } \underset{\sim}{t}\} .$$

In our work, context will indicate whether a letter denotes a linear operator or its matrix representation.

For $E,F \in L(\mathbb{R}^p)$, the multiplicative property

(2.3) $\|EF\| \le \|E\|\|F\|$ holds .

If we equip $\mathbb{R}^p$ with the l_1-norm:

$\|\underset{\sim}{t}\|_1 = \sum_{i=1}^{p} |t_i|$, then, if e_{ij} denotes the $(i,j)^{\underline{\text{th}}}$ entry of the matrix E, we let $\|E\|_1$ denote the matrix norm and have the result

(2.4) $\|E\|_1 = \max_{1\le j\le p} \sum_{i=1}^{p} |e_{ij}|$.

Next, we have the perturbation lemma:

(2.5) if $E,F \in L(\mathbb{R}^p)$, E invertible with $\|E^{-1}\| \le \alpha$ and if $\|E-F\| \le \beta$ where $\beta\alpha < 1$, then F is invertible and $\|F^{-1}\| \le \dfrac{\alpha}{1-\alpha\beta}$.

It follows from (2.5) that:

(2.6) if $E : \mathbb{R}^p \to L(\mathbb{R}^p)$ is continuous at $\underset{\sim}{t}_0$ and $E(\underset{\sim}{t}_0)$ is invertible then there exist $\delta,\gamma > 0$ such that $E(\underset{\sim}{t})$ is invertible and $\|E(\underset{\sim}{t})^{-1}\| \le \gamma$ for all $\underset{\sim}{t} \in \overline{S}(\underset{\sim}{t}_0,\gamma)$.

Moreover, $E(\underset{\sim}{t})^{-1}$ is continuous in $\underset{\sim}{t}$ at $\underset{\sim}{t}_0$.

Here, $\overline{S}(\underset{\sim}{t}_0,\delta) = \{\underset{\sim}{t} : \|\underset{\sim}{t}-\underset{\sim}{t}_0\| \le \delta\}$.

For the remainder of this section, F will denote a mapping from $D \subseteq \mathbb{R}^p$ into $\mathbb{R}^p$. We recall that

(2.7) F is Gateaux- (or G-) differentiable at an interior point $\underset{\sim}{t}$ of D if there exists a linear operator $E \in L(\mathbb{R}^p)$ such that, for any $\underset{\sim}{h} \in \mathbb{R}^p$, (2.8) $\lim_{\alpha\to 0} (\frac{1}{\alpha})\|F(\underset{\sim}{t}+\alpha\underset{\sim}{h}) - F(\underset{\sim}{t}) - \alpha E\underset{\sim}{h}\| = 0$.

The linear operator E is denoted by $F'(\underset{\sim}{t})$, is unique and, by (2.1), is independent of the particular norm on $\mathbb{R}^p$.

If F is G-differentiable at each $\underset{\sim}{t} \in D_0 \subseteq \text{int } D$, then for each $\underset{\sim}{t} \in D_0$, $F'(\underset{\sim}{t}_0)$ is a linear operator; that is, F' is a mapping from D_0 into $L(\mathbb{R}^p)$. In particular, F' is continuous at $\underset{\sim}{t} \in D_0$ if

(2.9) $\|F'(\underset{\sim}{t}+\underset{\sim}{h}) - F'(\underset{\sim}{t})\| \to 0$ as $\|\underset{\sim}{h}\| \to 0$.

We recall further that

(2.10) F is Frechet- (or F-) differentiable at $\underset{\sim}{t} \in \text{int}(D)$ if there is an $E \in L(\mathbb{R}^p)$ such that

(2.11) $\lim_{\underset{\sim}{h}\to 0} (\frac{1}{\|\underset{\sim}{h}\|})\|F(\underset{\sim}{t}+\underset{\sim}{h}) - F\underset{\sim}{t} - E\underset{\sim}{h}\| = 0$.

This E is again denoted by $F'(\underset{\sim}{t})$.

If we write $F = (f_1, f_2, \ldots, f_p)^T$ (T denotes transpose), then the matrix representation of $F'(\underset{\sim}{t})$ is given by the Jacobian matrix:

(2.12) $$F'(\underset{\sim}{t}) = \begin{bmatrix} \delta_1 f_1(\underset{\sim}{t}), & \ldots & , \delta_p f_1(\underset{\sim}{t}) \\ \vdots & & \vdots \\ \delta_1 f_p(\underset{\sim}{t}), & \ldots & , \delta_p f_p(\underset{\sim}{t}) \end{bmatrix},$$

where $\delta_j f_i(\underset{\sim}{t})$ is the $j^{\underline{\text{th}}}$ partial derivative of $f_i(\underset{\sim}{t})$.

We note that:

(2.13) if F is F-differentiable at $\underset{\sim}{t}$, then F is G-differentiable at $\underset{\sim}{t}$.

On the other hand,

(2.14) if F has a G-derivative at each point of an open neighbourhood of $\underset{\sim}{t}$ and if F' is continuous at $\underset{\sim}{t}$, then F is F-differentiable at $\underset{\sim}{t}$.

We observe also that:

(2.15) if F is F-differentiable at $\underset{\sim}{t}$, then F is continuous at $\underset{\sim}{t}$.

The last statement is false for G-differentiability.

We also note that:

(2.16) F' is continuous at $\underset{\sim}{t} \Longleftrightarrow$ all the partial derivatives $\delta_j f_i$ are continuous at $\underset{\sim}{t}$.

The mean value theorem for mappings $f:\mathbb{R}^p \to \mathbb{R}'$ does not have a direct analogue for mappings $F:\mathbb{R}^p \to \mathbb{R}^p$ $(p > 1)$. An alternative that we shall use on a few occasions is:

(2.17) $F:D \subseteq \mathbb{R}^p \to \mathbb{R}^p$ is G-differentiable on an open convex set $D_0 \subseteq D$ and $\underset{\sim}{t},\underset{\sim}{s} \in D_0$, then $F\underset{\sim}{t} - F\underset{\sim}{s} = B(\underset{\sim}{s},\underset{\sim}{t})(\underset{\sim}{t}-\underset{\sim}{s})$, where $B(\underset{\sim}{s},\underset{\sim}{t}) \in L(\mathbb{R}^p)$ is given by

$$B(\underset{\sim}{s},\underset{\sim}{t}) = \begin{bmatrix} \delta_1 f_1(\underset{\sim}{s}+\alpha_1(\underset{\sim}{t}-\underset{\sim}{s})), & \dots & ,\delta_p f_1(\underset{\sim}{s}+\alpha_1(\underset{\sim}{t}-\underset{\sim}{s})) \\ \vdots & & \vdots \\ \delta_1 f_p(\underset{\sim}{s}+\alpha_p(\underset{\sim}{t}-\underset{\sim}{s})), & \dots & ,\delta_p f_p(\underset{\sim}{s}+\alpha_p(\underset{\sim}{t}-\underset{\sim}{s})) \end{bmatrix},$$

for some $\alpha_1,\dots,\alpha_p \in (0,1)$.

In general, the α_i will all be distinct and $B(\underset{\sim}{s},\underset{\sim}{t})$ will not be the G-derivative evaluated at an intermediate pcint.

(2.18) (Corollary to Leray-Schauder theorem.)

Let C be an open bounded set in $\mathbb{R}^p$ and assume that $F:\overline{C} \subseteq \mathbb{R}^p \to \mathbb{R}^p$ is continuous and satisfies $(\underset{\sim}{t}-\underset{\sim}{t}_0)^T F(\underset{\sim}{t}) \geq 0$ for some $\underset{\sim}{t}_0 \in C$ and all $\underset{\sim}{t} \in \dot{C}$, where $\dot{C}$ denotes the boundary of C. Then $F(\underset{\sim}{t}) = \underset{\sim}{0}$ has a solution in $\overline{C}$.

(Note that the same result holds if $(\underset{\sim}{t}-\underset{\sim}{t}_0)^T F(\underset{\sim}{t}) \geq 0$ is changed to $(\underset{\sim}{t}-\underset{\sim}{t}_0)^T F(\underset{\sim}{t}) \leq 0$.)

(2.19) Assume that $F:D \subseteq \mathbb{R}^p \to \mathbb{R}^p$ is continuously differentiable on the open convex set D and that for any p points $\underset{\sim}{t}_1,\ldots,\underset{\sim}{t}_p \in D$, the matrix

$$\begin{bmatrix} \delta_1 f_1(\underset{\sim}{t}), & \cdots & ,\delta_p f_1(\underset{\sim}{t}) \\ \vdots & & \vdots \\ \delta_1 f_p(\underset{\sim}{t}), & \cdots & ,\delta_p f_p(\underset{\sim}{t}) \end{bmatrix} \text{ is invertible.}$$

Then F is one-to-one (see Ortega and Rheinboldt (1970), p. 140.)

3

MODEL AND CLASS OF ESTIMATORS

We fix α, $0 < \alpha < .5$ and set

$$(3.1) \qquad d = \Phi^{-1}(1 - \tfrac{\alpha}{2}) ,$$

where $\Phi(y) = \int_{-\infty}^{y} \phi(t)dt$ and $\phi(t) = (2\pi)^{-\frac{1}{2}}\exp(\frac{-t^2}{2})$. A class of distribution functions F is defined as follows:

(3.2) $G \in F$ if and only if there exists $\gamma \in (-\frac{\alpha}{2} , \frac{\alpha}{2})$ such that

$G(y) = \gamma + \Phi(y)$ for all $y \in [-d,d]$.

We set

$$(3.3) \qquad k = \Phi^{-1}(\tfrac{1}{2} + \tfrac{\alpha}{2})$$

and

$$(3.4) \qquad c = d-k .$$

Our model is:

$$(3.5) \qquad \underset{\sim}{X} = C\underset{\sim}{\theta} + \underset{\sim}{\varepsilon} ,$$

where $\underset{\sim}{X} = (X_1,\ldots,X_n)^T$ is a vector of n observations, $C = ((c_{ij}))$ is a given matrix of n rows and p columns (the design matrix), $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_p)^T$ is the unknown regression parameter vector to be estimated, and $\underset{\sim}{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_n)^T$ is the error vector.

We assume that the ε_i are independent, identically distributed random variables with distribution function

$G(\frac{y}{\sigma})$ where G is known to be a member of F.

We shall assume σ to be known here (so that, without loss of generality, $\sigma = 1$,) deferring the case of scale unknown until chapters 7 and 8.

We will now make a restriction on the design matrix. Let $q_1,\ldots,q_p$ be fixed positive weights and for each positive integer n, let $q_1(n),\ldots,q_p(n)$ be weights such that $q_i(n) \to q_i$. Since the theory we present is asymptotic and since we can arrange $q_i(n) \to q_i$ through a sequence of rationals, we may assume $nq_i(n)$ to be an integer.

Now assume that the first $nq_1(n)$ rows of C are the same, that the next $nq_2(n)$ rows of C are the same, and so on. That is

$$C = \begin{bmatrix} c_{11}, & \cdots & ,c_{1p} \\ c_{21}, & \cdots & ,c_{2p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ c_{n1}, & \cdots & ,c_{np} \end{bmatrix} = \begin{array}{l} \left[\begin{array}{ccc} a_{11}, & \cdots & ,a_{1p} \\ a_{11}, & \cdots & ,a_{1p} \\ a_{21}, & \cdots & ,a_{2p} \\ a_{21}, & \cdots & ,a_{2p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{p}, & \cdots & ,a_{pp} \\ a_{p}, & \cdots & ,a_{pp} \end{array}\right] \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}a\\a\end{array}}\right\} \to nq_1(n) \text{ times} \\ \left.\vphantom{\begin{array}{c}a\\a\end{array}}\right\} \to nq_2(n) \text{ times} \\ \vphantom{\begin{array}{c}a\\a\end{array}} \\ \left.\vphantom{\begin{array}{c}a\\a\end{array}}\right\} \to nq_p(n) \text{ times} . \end{array}$$

(Repeating rows of design matrices is not a serious restriction (see Draper and Smith (1966), p. 28). For our purposes it will help us in getting a good initial value for solving a certain system of equations that has multiple roots. The problem could be overcome by minimizing a certain functional instead and there are other possibilities also.) We let A be the matrix whose (i,j)th entry is a_{ij}, that is,

$\underset{p\times p}{A} = ((a_{ij}))$, and we assume that

$B = A^{-1} = ((b_{ij}))$ exists.

Now set

$$(3.6)\qquad \begin{aligned} M_1 &= M_{1,n} = \text{median of } \{X_1,\ldots,X_{nq_1}\} \ , \\ M_2 &= M_{2,n} = \text{median of } \{X_{nq_1+1},\ldots,X_{nq_2}\} \ , \\ &\ \ \vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ M_p &= M_{p,n} = \text{median of } \{X_{nq_{p-1}+1},\ldots,X_{nq_p}\} \end{aligned}$$

and set

$$(3.7)\qquad \underset{\sim}{M} = \underset{\sim}{M}_n = (M_1,\ldots,M_p)^T \ .$$

Next, let $\underset{\sim}{\theta}^* = \underset{\sim}{\theta}^*_n = (\theta^*_1,\ldots,\theta^*_p)^T$ be the solution of the system

$$A\underset{\sim}{\theta} = \underset{\sim}{M} \ , \quad \text{so that}$$

$$(3.8)\qquad \underset{\sim}{\theta}^* = B\underset{\sim}{M} = \begin{pmatrix} \sum_{j=1}^{p} b_{ij}\, M_j \\ \vdots \\ \sum_{j=1}^{p} b_{pj}\, M_j \end{pmatrix} .$$

Strictly, we should write $\underset{\sim}{\theta}^*_n(\underset{\sim}{X})$ in place of $\underset{\sim}{\theta}^*$. Such notational brevity is common in our work.

Now let

(3.9) $m(G)$ be the median of the distribution G and set

$$(3.10)\qquad \underset{\sim}{\theta}^{**} = m(G) \begin{pmatrix} \sum_{j=1}^{p} b_{ij} \\ \vdots \\ \sum_{j=1}^{p} b_{pj} \end{pmatrix}, \text{ which equals } B \begin{pmatrix} m(G) \\ \vdots \\ m(G) \end{pmatrix} .$$

We note that whatever the distribution G may be,

(3.11) $\quad M_j = M_{jn} \xrightarrow{P} m(G)$ for all $j=1,\ldots,p$.

[One way of seeing this is to recall that the variance of a sample quantile is of order n^{-1} for large n (see Cramer (1946) p.369) and then Chebychev's inequality gives the result.]

Then using the l_1 - norm or $\mathbb{R}^p$, we have

$$\|\underset{\sim}{\theta}^* - \underset{\sim}{\theta}^{**}\|_1 = \sum_{i=1}^{p} \left| \sum_{j=1}^{p} b_{ij} \; [M_j - m(G)] \right|$$

$$\leq \sum_{i=1}^{p} \sum_{j=1}^{p} b_{ij} |M_j - m(G)| \xrightarrow{P} 0 \text{ , by (3.11).}$$

Thus, we could have defined $\underset{\sim}{\theta}^{**}$ by:

(3.12) $\quad \underset{\sim}{\theta}^{**} = P\lim \underset{\sim}{\theta}^*$

where $P\lim$ denotes limit in probability.

We define a class Ψ_c of mappings from $\mathbb{R}^1$ to $\mathbb{R}^1$ as follows:

<u>(3.13) DEFINITION</u>:

$\psi \in \Psi_c$ *if and only if* ψ *is smooth (continuously differentiable), skew-symmetric, vanishes outside* $[-c,c]$ *and satisfies* $\psi \geq 0$ *on* $[0,c]$ *but* $\psi \not\equiv 0$ *on* $[0,c]$.

We propose to estimate $\underset{\sim}{\theta}$ by solving the system

(3.14) $$\sum_{i=1}^{p} c_{ik} \, \psi \left(X_i - \sum_{j=1}^{p} c_{ij} \, \theta_j \right) = 0 \text{ , } \quad k = 1,\ldots,p$$

for $\underset{\sim}{\theta}$, where $\psi \in \Psi_c$.

Clearly, since ψ vanishes outside $[-c,c]$, (3.14) has multiple roots, with probability one, even asymptotically. We shall first show that any solution of (3.14) is location invariant, so that later, in analysis of the behaviour of the estimator we shall propose for the true $\underset{\sim}{\theta}$, we will be able to assume without loss of generality that the true value of $\underset{\sim}{\theta}$ is $\underset{\sim}{0}$, for the purpose of simplifying notations and calculations.

We are to show that if $\underset{\sim}{\theta}_1(\underset{\sim}{X})$ is a solution of (3.14) then

$$(3.15) \qquad \underset{\sim}{\theta}_1(\underset{\sim}{X} + C\underset{\sim}{t}) = \theta_1(\underset{\sim}{X}) + \underset{\sim}{t} \quad \text{where } \underset{\sim}{t} \in \mathbb{R}^p .$$

To prove (3.15), we replace $\underset{\sim}{X}$ by $\underset{\sim}{X} + C\underset{\sim}{t}$ in (3.14). Then (3.14) becomes

$$(3.16) \qquad \sum_{i=1}^{n} c_{ik}\, \psi\left(X_i + \sum_{j=1}^{p} c_{ij}\, t_j - \sum_{j=1}^{p} c_{ij}\, \theta_j\right) = 0\,, \quad k = 1,\ldots,p\,,$$

i.e.,

$$(3.17) \qquad \sum_{i=1}^{n} c_{ik}\, \psi\left(X_i - \sum_{j=1}^{p} c_{ij}\, (\theta_j - t_j)\right) = 0\,, \quad k = 1,\ldots,p\,.$$

But (3.17) is of the same form as (3.14) with θ_j becoming $\theta_j - t_j$.

Thus the $j^{\underline{\text{th}}}$ component of the solution of (3.17) minus t_j is the $j^{\underline{\text{th}}}$ component of the solution of (3.14)

i.e., solution of (3.17) minus $\underset{\sim}{t}$ = solution of (3.14);

i.e., solution of (3.16) minus $\underset{\sim}{t}$ = solution of (3.14);

i.e., $\underset{\sim}{\theta}_1(\underset{\sim}{X} + C\underset{\sim}{t}) - \underset{\sim}{t} = \underset{\sim}{\theta}_1(\underset{\sim}{X})$, proving (3.15).

Thus,

(3.18) any solution of (3.14) is location invariant.

In vector form, the system (3.14) reads

$$(3.19) \qquad \begin{pmatrix} \displaystyle\sum_{i=1}^{n} c_{i1}\, \psi\left(X_i - \sum_{j=1}^{p} c_{ij}\, \theta_j\right) \\ \vdots \\ \displaystyle\sum_{i=1}^{n} c_{ip}\, \psi\left(X_i - \sum_{j=1}^{p} c_{ij}\, \theta_j\right) \end{pmatrix} = \underset{\sim}{0}\,.$$

We set

$$(3.20)\qquad \bar{D} = \{\underset{\sim}{t} \in \mathbb{R}^p : \max_{1\le r\le p} |\sum_{j=1}^{p} a_{rj}\, t_j| \le k\}$$

$$= \{\underset{\sim}{t} \in \mathbb{R}^p : A\underset{\sim}{t} \in [-k,k]^p\},$$

and introduce the process $\{F_n(\underset{\sim}{t}) : \underset{\sim}{t} \in \bar{D}\}$ where

$$(3.21)\quad F_n(\underset{\sim}{t}) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} c_{i1}\, \psi\left(X_i - \sum_{j=1}^{p} c_{ij}\, t_j\right) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} c_{ip}\, \psi\left(X_i - \sum_{j=1}^{p} c_{ij}\, t_j\right) \end{pmatrix}.$$

Note that the set of solutions of $F_n(\underset{\sim}{t}) = \underset{\sim}{0}$ coincides with that of (3.19), trivially.

We further introduce the mapping $F : \bar{D} \to \mathbb{R}^p$ defined by

$$(3.22)\quad F(\underset{\sim}{t}) = \begin{pmatrix} \sum_{i=1}^{p} a_{i1}\, q_i\, E_G\, \psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right) \\ \vdots \\ \sum_{i=1}^{p} a_{ip}\, q_i\, E_G\, \psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right) \end{pmatrix},$$

where $X - \sum_{j=1}^{p} a_{ij}\, t_j$ has distribution G and where E_G denotes the expectation operator under the distribution G. We call $F(\underset{\sim}{t})$ the asymptotic deterministic version of $F_n(\underset{\sim}{t})$.

In solving the system $F_n(\underset{\sim}{t}) = \underset{\sim}{0}$, we would like to ensure that the resulting estimator is a consistent estimator of the true $\underset{\sim}{\theta}$.

We require a "good" starting value for some iterative method of solution. Now, we will show in Section 5 that

$$\sup\{\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| : t \in \overline{D}\} \xrightarrow{P} 0 \ . \tag{3.23}$$

For the moment, we interpret this loosely by saying that the process $\{F_n(\underset{\sim}{t}),\ \underset{\sim}{t} \in \overline{D}\}$ resembles the function $\{F(\underset{\sim}{t}),\ \underset{\sim}{t} \in \overline{D}\}$ asymptotically.

Further, in Chapter 4, we will show that:

(3.24) the solution of $F(\underset{\sim}{t}) = \underset{\sim}{0}$

by Newton's method with starting value θ^{**} is the true $\underset{\sim}{\theta}$.

Then (3.12), (3.23) and (3.24) lead one to suspect that if we solve the system $F_n(\underset{\sim}{t}) = \underset{\sim}{0}$ by Newton's method with starting value $\hat{\theta}^*$, we may arrive at a consistent estimator of the true $\underset{\sim}{\theta}$. (This turns out to be true, as we show in Section 5.) Accordingly, for a fixed $\psi \in \Psi_c$, we define the sequence $\{\underset{\sim}{T}_n = \underset{\sim}{T}_n(\psi),\ n = 1,2,\ldots\}$ of estimators of the true $\underset{\sim}{\theta}$ as follows:

<u>(3.25) DEFINITION:</u>

Set $\underset{\sim}{t}^0 = \underset{\sim}{\theta}^*$ *and form the sequence*

$$\underset{\sim}{t}^{k+1} = \underset{\sim}{t}^k - F_n'(\underset{\sim}{t}^k)^{-1} F_n(\underset{\sim}{t}^k) \ , \quad k = 0,1,2,\ldots \ .$$

Then set

$$\underset{\sim}{T}_n = \begin{cases} \lim_{k\to\infty} \underset{\sim}{t}^k \ , & \textit{if this limit exists} \\ \underset{\sim}{\theta}^* \ , & \textit{otherwise} \ . \end{cases}$$

In Section 5 we will be required, of course, to examine if the iteration process is well-defined - in the sense of establishing the invertibility of $F_n'(\underset{\sim}{\theta}^k)$ and the boundedness of $\|F_n'(\underset{\sim}{\theta}^k)^{-1}\|$, $k = 0,1,2,\ldots$. (see (5.6).)

We note that since convergence of the sequence $\{\underset{\sim}{\theta}^k\}$ can never be determined, (3.25) is not actually an algorithm. In the case of estimation of a location parameter, Collins (1976), p. 71, gives an approximate algorithm, which can be applied here. We remark also that the idea of first solving $F(\underset{\sim}{t}) = \underset{\sim}{0}$ (instead of immediately tackling the system $F_n(\underset{\sim}{t}) = 0$) is a mathematically very simplifying technique.

Finally we make the remark that we chose to use Newton's method of solution for the elegance and simplicity of its form, in addition to its fast rate of convergence (e.g., quadratic convergence - $\|\underset{\sim}{t}^{k+1} - \underset{\sim}{t}_0\| \leq \beta\|\underset{\sim}{t}^k - \underset{\sim}{t}_0\|^2$, $\beta < +\infty$ provided the $\underset{\sim}{t}^k$ are sufficiently close to a solution $\underset{\sim}{t}_0$ - which holds under quite natural conditions). In actual practise, difficulties can arise in applications of this method. In specific cases, it should be possible to make suitable modifications.

NEWTON'S METHOD SOLUTION OF THE ASYMPTOTIC DETERMINISTIC EQUATION

As outlined in Section 3, for a first step in showing that $\underset{\sim}{T}_n$, defined in (3.25), is a consistent estimator of the true parameter vector $\underset{\sim}{\theta}$ in the model (3.5), we intend to show here that the Newton iteration method of solving the system $F(\underset{\sim}{t}) = \underset{\sim}{0}$ with starting value $\underset{\sim}{\theta}^{**}$ is the true $\underset{\sim}{\theta}$ (see (3.22), (3.10) and (3.11)). Because of (3.18) we will, without loss of generality, assume from now on that the true value of $\underset{\sim}{\theta}$ is $\underset{\sim}{0}$.

We have

$$(4.1) \qquad F(\underset{\sim}{t}) = \begin{pmatrix} \sum_{i=1}^{p} a_{i1}\, q_i\, E_G\, \psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right) \\ \vdots \\ \sum_{i=1}^{p} a_{ip}\, q_i\, E_G\, \psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right) \end{pmatrix}, \quad \underset{\sim}{t} \in \overline{D},$$

where $\overline{D}$ is given by (3.20). We note first that for each $i = 1,\ldots,p$:

$$(4.2) \qquad E_G\, \psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right) = \int_{-\infty}^{\infty} \psi\left(x - \sum_{j=1}^{p} a_{ij}\, t_j\right) dG(x)$$

$$= \int_{-c + \sum_{j=1}^{p} a_{ij}\, t_j}^{c + \sum_{j=1}^{p} a_{ij}\, t_j} \psi\left(x - \sum_{j=1}^{p} a_{ij}\, t_j\right) dG(x),$$

since ψ vanishes outside $[-c,c]$ (see (3.13)).

From (4.2), (3.20), (3.4) and (3.2) we see that

$$(4.3) \qquad E_G\, \psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right) = \int_{-c}^{c} \psi(x)\,\phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right) dx\ .$$

Noting that (4.3) is independent of $G \in F$ (this was partly the motivation for defining c as we did), we may drop the subscript G in (4.3) from here on.

Observe next that for any i,k where i, k=1, 2, ..., p ,

$$(4.4) \quad \frac{\delta}{\delta t_k} E\psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right) = \int_{-c}^{c} \psi(x) \frac{\delta}{\delta t_k}\, \phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right) dx\ .$$

This follows from the continuity of $\psi(x)\phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right)$ and the (existence and) continuity of the partial derivatives of $\psi(x)\phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right)$ with respect to t_k (k=1,...,p).

We now check the continuity of $\frac{\delta}{\delta t_k} E\psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right)$ in $\underset{\sim}{t} \in D$: let $\underset{\sim}{r} \in D$. Then for $\underset{\sim}{s}$ in a neighbourhood of $\underset{\sim}{r}$, we write

$$\left| \left[\frac{\delta\ E\ \psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right)}{\delta t_k}\right]_{\underset{\sim}{t}=\underset{\sim}{s}} - \left[\frac{\delta\ E\ \psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right)}{\delta t_k}\right]_{\underset{\sim}{t}=\underset{\sim}{r}} \right|$$

$$\leq \int_{-c}^{c} |\psi(x)| \left| \left[\frac{\delta\phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right)}{\delta t_k}\right]_{\underset{\sim}{t}=\underset{\sim}{s}} - \left[\frac{\delta\phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right)}{\delta t_k}\right]_{\underset{\sim}{t}=\underset{\sim}{r}} \right| dx$$

and, since ψ is bounded (in fact, it attains its sup and inf, being a continuous function that vanishes outside a compact set), it suffices to show that for $\varepsilon > 0$, there exists $\delta > 0$ such that

$$(4.5) \qquad \left| \left[\frac{\delta\phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right)}{\delta t_k}\right]_{\underset{\sim}{t}=\underset{\sim}{s}} - \left[\frac{\delta\phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right)}{\delta t_k}\right]_{\underset{\sim}{t}=\underset{\sim}{r}} \right| < \varepsilon$$

whenever $\|\underset{\sim}{s}-\underset{\sim}{r}\| < \delta$.

We write

$$\left| \left[\frac{\delta\phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right)}{\delta t_k}\right]_{\underset{\sim}{t}=\underset{\sim}{s}} - \left[\frac{\delta\phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right)}{\delta t_k}\right]_{\underset{\sim}{t}=\underset{\sim}{r}} \right|$$

$$= |a_{ik}| \left| \left(x+\sum_{j=1}^{p} a_{ij} r_j\right)\phi\left(x+\sum_{j=1}^{p} a_{ij} r_j\right) - \left(x+\sum_{j=1}^{p} a_{ij} s_j\right)\phi\left(x+\sum_{j=1}^{p} a_{ij} s_j\right) \right|$$

$$\leq |a_{ik}| \left\{ \left|x+\sum_{j=1}^{p} a_{ij} r_j\right| \left|\phi\left(x+\sum_{j=1}^{p} a_{ij} r_j\right) - \phi\left(x+\sum_{j=1}^{p} a_{ij} s_j\right)\right| \right.$$

$$\left. + \left| \sum_{j=1}^{p} a_{ij} r_j - \sum_{j=1}^{p} a_{ij} s_j \right| \phi\left(x+\sum_{j=1}^{p} a_{ij} s_j\right) \right\}$$

and the rest of the proof is elementary using continuity of the functions $\underset{\sim}{t} \to \phi\left(x + \sum_{j=1}^{p} a_{ij}\, t_j\right)$ and $\underset{\sim}{t} \to \sum_{j=1}^{p} a_{ij}\, t_j$.

Next, we observe that each of the p components of our function F is just a linear combination of the $E\,\psi\left(X - \sum_{j=1}^{p} a_{ij}\, t_j\right)$, $i=1,\ldots,p$.

Accordingly, we have:

(4.6) LEMMA:

The partial derivatives, with respect to $t_1,\ldots,t_p$ *of the* p *components of the function*

$$F(\underset{\sim}{t}) = \begin{pmatrix} \sum_{i=1}^{p} a_{i1}q_i \; E\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right) \\ \vdots \\ \sum_{i=1}^{p} a_{ip}q_i \; E\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right) \end{pmatrix}$$

(exist and) are continuous on D .

We now prove

<u>(4.7) THEOREM:</u>

a) F *is G-differentiable on* D ;

b) F' *is continuous on* D ;

c) F *is F-differentiable on* D ;

d) F *is continuous on* D .

<u>Proof:</u>

Let $\underset{\sim}{t} \in D$ and let $M = M(\underset{\sim}{t})$ be the matrix of partial derivatives of F, i.e.,

(4.8) $$M = \begin{bmatrix} \frac{\delta}{\delta t_1} \sum_i a_{i1}q_i \; E\psi(X-\sum_j a_{ij}t_j), \ldots, \frac{\delta}{\delta t_p} \sum_i a_{i1}q_i \; E\psi(X-\sum_j a_{ij}\; t_j) \\ \vdots \\ \frac{\delta}{\delta t_1} \sum_i a_{ip}q_i \; E\psi(X-\sum_j a_{ij}t_j), \ldots, \frac{\delta}{\delta t_p} \sum_i a_{ip}q_i \; E\psi(X-\sum a_{ij}\; t_j) \end{bmatrix} .$$

To prove a), we will show (see (2.7)) that this M satisfies

(4.9) $$\lim_{\alpha\to 0} \left(\frac{1}{\alpha}\right)\|F(\underset{\sim}{t}+\alpha\underset{\sim}{h}) - F(\underset{\sim}{t}) - \alpha M\underset{\sim}{h}\| = 0 , \quad \underset{\sim}{h} \in \mathbb{R}^p .$$

Using, e.g., the l_1-norm: $\|\underset{\sim}{s}\|_1 = \Sigma|s_i|$, $\underset{\sim}{s} \in \mathbb{R}^p$, we have

$$\frac{1}{\alpha}\ \|F(\underset{\sim}{t}+\alpha\underset{\sim}{h}) - F(\underset{\sim}{t}) - \alpha M\underset{\sim}{h}\|_1$$

$$= \frac{1}{\alpha}\sum_k \Big| \sum_i a_{ik}q_i\ E\psi(X-\sum_j a_{ij}(t_j+\alpha h_j)) - \sum_i a_{ik}q_i\ E\psi(X-\sum_j a_{ij}t_j)$$

$$- \alpha \sum_r h_r \left\{\frac{\delta}{\delta t_r}\sum_i a_{ik}q_i\ E\psi(X-\sum_j a_{ij}t_j)\right\} \Big|$$

$$\leq \frac{1}{\alpha}\sum_k\sum_i a_{ik}q_i\ \Big| E\psi(X-\sum_j a_{ij}(t_j+\alpha h_j)) - E\psi(X-\sum_j a_{ij}t_j)$$

$$- \alpha\sum_r h_r \frac{\delta}{\delta t_r} E\psi(X-\sum_j a_{ij}t_j) \Big| ,$$

so that, to prove (4.9), it suffices to show

$$(4.10)\quad \frac{1}{\alpha}\ \Big| E\psi(X-\sum_j a_{ij}(t_j+\alpha h_j) - E\psi(X-\sum_j a_{ij}t_j)$$

$$- \sum_r h_r \frac{\delta}{\delta t_r} E\ \ (X-\sum_j a_{ij}t_j) \Big| \rightarrow 0 \quad (\alpha \rightarrow 0) ,$$

for each $i = 1,\ldots,p$.

Now, from (4.3) and (4.4), the expression in (4.10) equals

$$\frac{1}{\alpha}\ \Big| \int_{-c}^{c}\psi(x)\,\phi(x+\sum_j a_{ij}(t_j+\alpha h_j)) - \int_{-c}^{c}\psi(x)\,\phi(x+\sum_j a_{ij}t_j)$$

$$- \alpha\sum_r h_r \int_{-c}^{c}\psi(x)\ \frac{\delta}{\delta t_r}\ \phi(x+\sum_j a_{ij}t_j) \Big|$$

$$= \frac{1}{\alpha}\ \Big| \int_{-c}^{c}\psi(x)\ \Big[\phi(x+\sum_j a_{ij}(t_j+\alpha h_j)) - \phi(x+\sum_j a_{ij}t_j)$$

$$- \alpha\ \phi'(x+\sum_j a_{ij}t_j)\ \sum_r a_{ir}h_r \Big] \Big|$$

$$\leq \frac{1}{\alpha}\int_{-c}^{c}|\psi(x)|\ \Big| \phi(x+\sum_j a_{ij}(t_j+\alpha h_j)) - \phi(x+\sum_j a_{ij}t_j)$$

$$- \alpha\ \phi'(x+\sum_j a_{ij}t_j)\ \sum_r a_{ir}h_r \Big| .$$

Thus, to prove (4.10), it suffices to show:

$$(4.11)\qquad \frac{1}{\alpha}\left|\ \phi(x+\sum_j a_{ij}(t_j+\alpha h_j)) - \phi(x+\sum_j a_{ij}t_j) - \alpha\ \phi'(x+\sum_j a_{ij}t_j)\cdot\sum a_{ij}\ h_j\ \right| \longrightarrow 0 \quad (\alpha\to 0)\ ,$$

for each $x \in [-c,c]$.

When $\sum_j a_{ij}h_j = 0$, (4.11) is trivial.

For those values of $\underset{\sim}{h}$ for which $\sum_j a_{ij}h_j \neq 0$, set

$$h = \sum_j a_{ij}h_j\ ,\quad t = x + \sum_j a_{ij}t_j\ ,\quad \beta = \alpha h$$

and then proving (4.11) is equivalent to proving

$$\lim_{\beta\to 0}\frac{1}{\beta}\left|\ \phi(t+\beta) - \phi(t) - \beta\phi'(t)\ \right| = 0\ .$$

But this is the definition of $\phi'(t)$. Thus a) is proved.

Now b) follows from (4.6) and (2.16);

c) follows from a), b) and (2.14)

while d) is proved by c) and (2.15).

This completes the proof of (4.7). □

(4.12) REMARK

We are now in a position to state and prove the main result of this section concerning the Newton's method solution of the system $\underset{\sim}{F}(\underset{\sim}{t}) = \underset{\sim}{0}$. A careful examination of the theorems in Sections 10.1 and 10.2 of Ortega and Rheinboldt (1970) (in particular 10.1.3, p. 300, 10.2.1, p. 311 and 10.2.2, p. 312) show that, under certain conditions, certain iteration processes, in particular Newton's iteration process, will converge. The key point, however, is that while these theorems are instructive they are most non-constructive. Their drawback lies in the fact that while they guarantee convergence if we start in a sufficiently small neighbourhood of the target value, they do not guarantee convergence if our starting value is in a "pre-chosen" neighbourhood (see Ortega and Rheinboldt (1970), p. 302, p. 317 and p. 381).

We aim to show that the Newton's method solution of $F(\underset{\sim}{t}) = \underset{\sim}{0}$ with starting value θ^{**} (see (3.10)) converges to $\underset{\sim}{0}$. Now, with $\overline{D}$ given by (3.20), we shall see that $\theta^{**} \in D = \text{int}\,\overline{D}$ (θ^{**} may be any point of D), but we should have no reason to suppose that D is a small enough neighbourhood of $\underset{\sim}{0}$ to permit convergence of the Newton iterates to $\underset{\sim}{0}$ if we start anywhere in D (although, as we will soon show, D is sufficiently small). Originally, we thought of using the Newton-Kantorovich Theorem (see Ortega and Rheinboldt (1970) and Ortega (1972)) to give us our neighbourhood. However, we decided against this, in view of the fact that the conditions of that theorem are usually very difficult to verify in practise. In any case, the importance of that theorem perhaps lies more in giving us error estimates and in ensuring that a given system does have a root, than in giving us a "starting neighbourhood". Our choice of D was suggested by an examination of the corresponding situation in the case of a location parameter (Collins (1976)) and the starting neighbourhood $(-k,k)$ found there.

For notational convenience in future we shall often write:

$$E\psi(X-\sum_j a_{ij}t_j) = E\psi_i$$

and

$$E\ \frac{d\psi(X-\sum_j a_{ij}t_j)}{d(X-\sum_j a_{ij}t_j)} = E\psi_i'$$

in cases where no confusion can arise.

We now have:

(4.13) THEOREM:

a) $F(\underset{\sim}{0}) = \underset{\sim}{0}$;

b) $\underset{\sim}{\theta}^{**} \in D$;

c) *$F'(\underset{\sim}{t})$ is non-singular for all $\underset{\sim}{t} \in D$* ;

d) *The Newton iterates*

$$\underset{\sim}{t}^{k+1} = \underset{\sim}{t}^{k} - F'(\underset{\sim}{t}^{k})^{-1}F(\underset{\sim}{t}^{k}) \ , \quad k = 0,1,2,\ldots,$$

with starting value $\underset{\sim}{t} = \underset{\sim}{\theta}^{**}$ *are well-defined, remain in* D *and converge to* $\underset{\sim}{0}$.

Proof:

We have

$$F(\underset{\sim}{t}) = \begin{pmatrix} \sum_i a_{i1}q_i \ E\psi(X-\sum_j a_{ij}t_j) \\ \vdots \\ \sum_i a_{ip}q_i \ E\psi(X-\sum_j a_{ij}t_j) \end{pmatrix}$$

so that

$$F(\underset{\sim}{0}) = \begin{pmatrix} \sum_i a_{i1}q_i \ E\psi(X) \\ \vdots \\ \sum_i a_{ip}q_i \ E\psi(X) \end{pmatrix}$$

so, to prove a) we will show that $E\psi(X) = 0$. But,

$$E\psi(X) = \int_{-c}^{c} \psi(x)\phi(x)dx \ , \quad \text{by (4.3)},$$

and the result follows by symmetry of ϕ and skew-symmetry of ψ.

We next prove b): it is easy to see from (3.2) and (3.3) that $m(G) \in (-k,k)$. Then

$$\underset{\sim}{\theta}^{**} = B \begin{pmatrix} m(G) \\ \cdot \\ \cdot \\ \cdot \\ m(G) \end{pmatrix} \in D \quad \text{because}$$

$$A\underset{\sim}{\theta}^{**} = \begin{pmatrix} m(G) \\ \cdot \\ \cdot \\ \cdot \\ m(G) \end{pmatrix} \in (-k,k)^p \ .$$

To prove c), fix $\underset{\sim}{t} \in D$ and write

$$F(\underset{\sim}{t}) = \begin{pmatrix} \sum_i a_{i1} q_i \ E\psi(X - \sum_j a_{ij} t_j) \\ \cdot \\ \cdot \\ \cdot \\ \sum_i a_{ip} q_i \ E\psi(X - \sum_j a_{ij} t_j) \end{pmatrix} = \begin{pmatrix} \sum_i a_{i1} q_i \ E\psi_i \\ \cdot \\ \cdot \\ \cdot \\ \sum_i a_{ip} q_i \ E\psi_i \end{pmatrix} ,$$

in a notation previously introduced.

We have, by (4.7), that $F'(\underset{\sim}{t})$ exists and its matrix representation is given by (4.8):

$$F'(\underset{\sim}{t}) = \begin{pmatrix} \sum_i a_{i1} q_i \frac{\delta}{\delta t_1} E\psi_i \ , & \cdot \ \cdot \ \cdot \ \cdot \ , & \sum_i a_{i1} q_i \frac{\delta}{\delta t_p} E\psi_i \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \sum_i a_{ip} q_i \frac{\delta}{\delta t_1} E\psi_i \ , & \cdot \ \cdot \ \cdot \ \cdot \ , & \sum_i a_{ip} q_i \frac{\delta}{\delta t_p} E\psi_i \end{pmatrix} ,$$

which equals

$$\begin{pmatrix} -\sum_i a_{i1}q_i a_{i1}E\psi_i' , & \ldots , & -\sum_i a_{i1}q_i a_{ip}E\psi_i' \\ \vdots & & \vdots \\ -\sum_i a_{ip}q_i a_{i1}E\psi_i' , & \ldots , & -\sum_i a_{ip}q_i a_{ip}E\psi_i' \end{pmatrix} ,$$

which factors to

$$(-1) \begin{pmatrix} a_{11}q_1 , & \ldots , & a_{p1}q_p \\ \vdots & & \vdots \\ a_{1p}q_1 , & \ldots , & a_{pp}q_p \end{pmatrix} \begin{pmatrix} a_{11}E\psi_1' , & \ldots , & a_{1p}E\psi_1' \\ \vdots & & \vdots \\ a_{p1}E\psi_p' , & \ldots , & a_{pp}E\psi_p' \end{pmatrix}$$

and this further factors to

$$- A^T \operatorname{Diag}((q_i)) \operatorname{Diag}((E\psi_i')) A ,$$

where $\operatorname{Diag}((q_i))$ is the matrix with q_i in the i^{th} row and i^{th} column and zeros elsewhere. Similarly for $\operatorname{Diag}((E\psi_i'))$.

We thus have

$$F'(\underset{\sim}{t}) = - A^T \operatorname{Diag}((q_i)) \operatorname{Diag}((E\psi_i')) A . \tag{4.14}$$

Hence, the determinant of $F'(\underset{\sim}{t})$ is

$$\det F'(\underset{\sim}{t}) = - \prod_{i=1}^{p} (q_i E\psi_i')(\det A^T)(\det A) . \tag{4.15}$$

Now, $q_i \neq 0$, $i=1,\ldots,p$ and

$\det A^T = \det A \neq 0$, by assumption.

Thus, to show $F'(\underset{\sim}{t})$ is non-singular, it is sufficient to show

(4.16) $\qquad E\psi_i' \neq 0 \ , \quad i=1,\ldots,p \ .$

Now, $$-E\psi_i' = -E\,\frac{d\psi(X-\sum_j a_{ij}t_j)}{d(X-\sum_j a_{ij}t_j)} = E\,\frac{\delta\psi(X-\sum_j a_{ij}t_j)}{\delta(\sum_j a_{ij}t_j)} \ .$$

Now also $\underset{\sim}{t} \in D$ and so $t = \sum_j a_{ij}t_j$ satisfies

(4.17) $\qquad t \in (-k,k) \ .$

By (4.17) and Lemma 2.1 (iii) of Collins (1976), we have

$$-E\psi_i' < 0 \ , \quad \text{establishing (4.16).}$$

(Note that Collins (1976) uses the notation $\lambda(t)$ for $E\psi(X-t)$.)

This proves c).

Finally, we prove d). Set

(4.18) $\qquad H(\underset{\sim}{t}) = \underset{\sim}{t} - F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t}) \ , \quad \underset{\sim}{t} \in D \ .$

To show that this is well-defined, we must show that $F'(\underset{\sim}{t})^{-1}$ exists and that $\|F'(\underset{\sim}{t})^{-1}\|$ is bounded (in $\underset{\sim}{t}$) on D.

By c), $F'(\underset{\sim}{t})^{-1}$ exists $\forall\ \underset{\sim}{t} \in D$ and we now sketch two different proofs of the boundedness of $\|F'(\underset{\sim}{t})^{-1}\|$ on D.

Proof (i):

We use the l_1-norm for vectors in $\mathbb{R}^p$, so that, by (2.4), the corresponding norm of a matrix E is

(4.19) $$\|E\|_1 = \max_j \sum_i |e_{ij}| \ .$$

We have $E\psi_i' \neq 0$, $i=1,\ldots,p$, for all $\underset{\sim}{t} \in D$, by (4.16). In fact we have $E\psi_i' \neq 0$ for all $\underset{\sim}{t} \in \overline{D}$. Further, it is easy to check that $E\psi_i'$ is continuous on $\overline{D}$. These two facts imply that $1/E\psi_i'$ is continuous on $\overline{D}$. Since $\overline{D}$ is compact, $1/E\psi_i'$ is bounded, that is, for each i, $i=1,\ldots,p$, there exists M_i such that

$$(4.20) \qquad |1/E\psi_i'| \leq M_i \ .$$

Then, $\qquad \|F'(\underset{\sim}{t})^{-1}\|_1$

$$= \|(-1)\ A^{-1}\ \mathrm{Diag}((\frac{1}{E\psi_i'}))\ \mathrm{Diag}((\frac{1}{q_i}))\ (A^T)^{-1}\|_1\ , \quad \text{by (4.14)}$$

$$\leq \|A^{-1}\|_1 \|\mathrm{Diag}((\frac{1}{E\psi_i'}))\|_1 \|\mathrm{Diag}((\frac{1}{q_i}))\|_1 \|(A^T)^{-1}\|_1\ , \quad \text{by (2.3)}$$

$$= \|A^{-1}\|_1 \|(A^T)^{-1}\|_1 \left(\max_j \left|\frac{1}{E\psi_j'}\right|\right) \left(\max_j \frac{1}{q_j}\right)\ , \quad \text{by (4.19)}$$

$$\leq \|A^{-1}\|_1 \|(A^T)^{-1}\|_1 \left(\max_j M_j\right) \left(\max_j \frac{1}{q_j}\right)\ ,$$

i.e, $\|F'(\underset{\sim}{t})^{-1}\|$ is bounded.

Proof (ii):

By extending the domain of definition of F from $\overline{D}$ to any open set containing $\overline{D}$ (or even $\mathbb{R}^p$) one can easily modify (4.7) to show that the results there hold when D is replaced by $\overline{D}$ (after all, ψ vanishes outside $[-c,c]$). In particular, F is F-differentiable on $\overline{D}$ and F' is continuous on $\overline{D}$. We remark that the statement that F is differentiable on a closed set must be interpreted in the sense of F being differentiable on an open set containing $\overline{D}$, for we do not discuss differentiability on boundary points of the domain of definition of a function if this domain is closed. Now also, $F'(\underset{\sim}{t})^{-1}$ exists on $\overline{D}$.

Then, by (2.6), for any $\underset{\sim}{t}_1 \in \bar{D}$, we can find a ball $S(\underset{\sim}{t}_1, \delta)$ and a number γ_1 such that

$$\|F'(\underset{\sim}{t})^{-1}\| < \gamma_1 \quad \text{for all} \quad \underset{\sim}{t} \in \bar{S}(\underset{\sim}{t}_1, \delta_1).$$

The collection of balls $\{S(\underset{\sim}{t}, \delta_{\underset{\sim}{t}})\}_{\underset{\sim}{t} \in \bar{D}}$ is an open covering of the compact set $\bar{D}$ and hence a finite subcollection

$$\{S(\underset{\sim}{t}_i, \delta_{\underset{\sim}{t}_i})\}_{1 \le i \le m} \quad \text{covers } \bar{D} .$$

If γ_i is the γ corresponding to $\delta_{\underset{\sim}{t}_i}$, we then have

$$\|F'(\underset{\sim}{t})^{-1}\| \le \max_{1 \le i \le m} \gamma_i \text{ for all } \underset{\sim}{t} \in \bar{D} ,$$

i.e., $\|F'(\underset{\sim}{t})^{-1}\|$ is bounded.

To show that the iterates remain in D, we first show

$$\|\underset{\sim}{t} - F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\| < \|\underset{\sim}{t}\| \quad \text{for all } \underset{\sim}{t} \in D , \tag{4.21}$$

for some norm to be specified shortly.

We have, for $\underset{\sim}{t} \in D$,

$$F(\underset{\sim}{t}) = \begin{pmatrix} \sum_i a_{i1} q_i \; E\psi_i \\ \vdots \\ \sum_i a_{ip} q_i \; E\psi_i \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}, & \cdots, & a_{p1} \\ \vdots & & \vdots \\ a_{1p}, & \cdots, & a_{pp} \end{pmatrix} \begin{pmatrix} q_1 \; E\psi_1 \\ \vdots \\ q_p \; E\psi_p \end{pmatrix}$$

or,

$$F(\underset{\sim}{t}) = A^T \operatorname{Diag}((q_i)) \begin{pmatrix} E\psi_1 \\ \cdot \\ \cdot \\ \cdot \\ E\psi_p \end{pmatrix} . \tag{4.22}$$

Now, from (4.14), we have

$$F'(\underset{\sim}{t})^{-1} = -A^{-1} \operatorname{Diag}((\frac{1}{E\psi_i'})) \operatorname{Diag}((\frac{1}{q_i})) (A^T)^{-1} . \tag{4.23}$$

From (4.22) and (4.23) we have

$$F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t}) \tag{4.24}$$

$$= - A^{-1} \operatorname{Diag}((\frac{1}{E\psi_i'})) \operatorname{Diag}((\frac{1}{q_i}))(A^T)^{-1} \operatorname{Diag}((q_i)) \begin{pmatrix} E\psi_1 \\ \cdot \\ \cdot \\ \cdot \\ E\psi_p \end{pmatrix}$$

$$= - A^{-1} \begin{pmatrix} \frac{E\psi_1}{E\psi_1'} \\ \cdot \\ \cdot \\ \cdot \\ \frac{E\psi_p}{E\psi_p'} \end{pmatrix} = A^{-1} \begin{pmatrix} \frac{E\psi_1}{-E\psi_1'} \\ \cdot \\ \cdot \\ \cdot \\ \frac{E\psi_p}{-E\psi_p'} \end{pmatrix} .$$

Now, for each fixed $\underset{\sim}{t} \in D$, we write

$$t = \sum_j a_{ij} t_j \text{ , so that}$$

$$-\frac{E\psi_i}{E\psi_i'} = \frac{E\psi\ (X-t)}{-E\psi'(X-t)} \text{ with } t \in (-k,k) \text{ by definition of } D . \tag{4.25}$$

In the notation of Collins (1976), (4.25) equals $\frac{\lambda(t)}{\lambda'(t)}$ and Collins showed that

$$(4.26) \qquad \left|\frac{\lambda(t)}{\lambda'(t)}\right| < 2|t| \text{ for all } t \in (-k,k) .$$

Further, it was shown in Collins (1976) that

$$(4.27) \qquad \frac{\lambda(t)}{\lambda'(t)} \text{ is } > 0 \text{ or } < 0 \text{ according as } t > 0 \text{ or } < 0, \text{ respectively,}$$

i.e., $\frac{\lambda(t)}{\lambda'(t)}$ retains the same sign as t.

(At $t = 0$, $\frac{\lambda(t)}{\lambda'(t)} = 0$.)

In our notation, (4.26) and (4.27) read

$$(4.28) \qquad \left|\frac{E\psi_i}{E\psi_i'}\right| = \left|\frac{E\psi_i}{-E\psi_i'}\right| < 2|t| \quad \text{and}$$

$$(4.29) \qquad -\frac{E\psi_i}{E\psi_i'} \text{ has the same sign as } t = \sum_j a_{ij}t_j$$

and $\frac{E\psi_i}{E\psi_i'} = 0$ if $t = \sum_j a_{ij}t_j = 0$.

(The function $\frac{E\psi_i}{E\psi_i'}$ and many others are thoroughly examined in Chapter 8.)

In proving (4.21), the appropriate norm to use is

$$(4.30) \qquad \|\underset{\sim}{t}\|_A = \max_{1\le l\le p} \left| \sum_{j=1}^{p} a_{lj}t_j \right| .$$

(The invertibility of A is necessary for (4.30) to be a norm.)

We now have

$$\|\underset{\sim}{t} - F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\|_A$$

$$= \left\| \begin{pmatrix} t_1 \\ \cdot \\ \cdot \\ \cdot \\ t_p \end{pmatrix} - A^{-1} \begin{pmatrix} \dfrac{-E\psi_1}{E\psi_1'} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{-E\psi_p}{E\psi_p'} \end{pmatrix} \right\|_A , \quad \text{from (4.24)}$$

$$= \left\| \begin{pmatrix} t_1 \\ \cdot \\ \cdot \\ \cdot \\ t_p \end{pmatrix} - \begin{pmatrix} \sum_k b_{1k} \left(\dfrac{-E\psi_k}{E\psi_k'} \right) \\ \cdot \\ \cdot \\ \cdot \\ \sum_k b_{pk} \left(\dfrac{-E\psi_k}{E\psi_k'} \right) \end{pmatrix} \right\|_A , \quad \text{where } ((b_{ij})) = A^{-1}$$

$$= \max_l \left| \sum_j a_{lj} \left[t_j - \sum_k b_{jk} \left(\frac{-E\psi_k}{E\psi_k'} \right) \right] \right| , \quad \text{by (4.30)}$$

$$= \max_l \left| \sum_j a_{lj} t_j - \left(- \frac{E\psi_l}{E\psi_l'} \right) \right| , \quad \text{since } ((b_{ij})) = ((a_{ij}))^{-1}$$

$$< \max_l \left| \sum_j a_{lj} t_j \right| , \quad \text{by (4.28) and (4.29)}$$

$$= \|\underset{\sim}{t}\|_A .$$

This proves (4.21) for the norm (4.30).

We are now in a position to prove that the iterates

(4.31) $$\underset{\sim}{t}^{k+1} = \underset{\sim}{t}^{k} - F'(\underset{\sim}{t}^{k})^{-1} F(\underset{\sim}{t}^{k}) , \quad k = 0,1,2,\ldots,$$

remain in D and converge to $\underset{\sim}{0}$.

By (4.21), there exists $\alpha < 1$ such that

(4.32) $\|\underset{\sim}{t} - F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\|_A \leq \alpha\|\underset{\sim}{t}\|_A$ for all $\underset{\sim}{t} \in D$.

Now $\underset{\sim}{t}^0 = \underset{\sim}{\theta}^{**} \in D$, by b) of this theorem.

Hence, the first iterate

$$\underset{\sim}{t}^1 = \underset{\sim}{t}^0 - F'(\underset{\sim}{t}^0)^{-1}F(\underset{\sim}{t}^0) \quad \text{satisfies}$$

$$\|\underset{\sim}{t}^1\|_A = \|\underset{\sim}{t}^0 - F'(\underset{\sim}{t}^0)^{-1}F(\underset{\sim}{t}^0)\|_A \leq \alpha\|\underset{\sim}{t}^0\|_A < \|\underset{\sim}{t}^0\|_A$$

i.e., $\max\limits_{l} \left| \sum\limits_{j} a_{lj}t_j^1 \right| < \max\limits_{l} \left| \sum\limits_{j} a_{lj}t_j^0 \right|$ by (4.30) .

But $\sum\limits_{j} a_{lj}t_j^0 \in (-k,k)$, by the definition of D and the fact that $\underset{\sim}{t}^0 = \underset{\sim}{\theta}^{**} \in D$ and so

$$\max_{l} \left| \sum_{j} a_{lj}t_j^0 \right| < k .$$

Thus ,

$$\underset{\sim}{t}^1 \in D .$$

By induction, all iterates $\underset{\sim}{t}^j$, $j = 0,1,2,\ldots$ lie in D and satisfy

$$\|\underset{\sim}{t}^j\|_A \leq \alpha\ \|\underset{\sim}{t}^0\|_A .$$

Since $\alpha < 1$, we have

$$\lim_{\to\infty} \|\underset{\sim}{t}^j\|_A = 0 \quad \text{which implies}$$

$\lim\limits_{j\to\infty} \underset{\sim}{t}^j = 0$, since any norm is a continuous function.

This completes the proof of (4.13). □

5

CONSISTENCY AND ASYMPTOTIC NORMALITY

In this section, we will show that the system of equations

$$(5.1) \qquad F_n(\underset{\sim}{t}) = 0 ,$$

where

$$(5.2) \quad F_n(\underset{\sim}{t}) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} c_{i1}\,\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} c_{ip}\,\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right) \end{pmatrix} = \begin{pmatrix} f_{n,1}(\underset{\sim}{t}) \\ \vdots \\ f_{n,p}(\underset{\sim}{t}) \end{pmatrix}, \quad \text{say,}$$

solved by Newton's method with initial value $\underset{\sim}{\theta}^*$ (see (3.8) and (3.25)), yields, for each $\psi \in \Psi_c$, a consistent and asymptotically normal estimator of the true $\underset{\sim}{\theta}$, which without loss of generality we have assumed to be $\underset{\sim}{0}$.

We had

$$(5.3) \quad F(\underset{\sim}{t}) = \begin{pmatrix} \sum_{i=1}^{p} a_{i1}q_i\, E\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right) \\ \vdots \\ \sum_{i=1}^{p} a_{ip}q_i\, E\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right) \end{pmatrix} = \begin{pmatrix} f_1(\underset{\sim}{t}) \\ \vdots \\ f_p(\underset{\sim}{t}) \end{pmatrix}, \quad \text{say,}$$

and in Section 4, we showed that the Newton's method solution of $F(\underset{\sim}{t}) = \underset{\sim}{0}$ with initial value $\underset{\sim}{\theta}^{**}$ (see 3.10) is $\underset{\sim}{0}$.

(5.4) REMARK:

We note that for each w in the underlying sample space of the random variables X_i, an identical version of (4.7) holds for the mapping

$F_{n,w}:D$ (or $\overline{D}$ or even $\mathbb{R}^p$) $\rightarrow \mathbb{R}^p$ given by

$$(5.5) \quad F_{n,w}(\underset{\sim}{t}) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} c_{i1}\,\psi\left(X_i(w) - \sum_{j=1}^{p} c_{ij}t_j\right) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} c_{ip}\,\psi\left(X_i(w) - \sum_{j=1}^{p} c_{ij}t_j\right) \end{pmatrix} .$$

Since the proof of this fact is so similar to, only easier than, the proof of (4.7), we omit the details.

(5.6) REMARK:

We cannot discuss the iterates

$$(5.7) \qquad \underset{\sim}{t}^{k+1} = \underset{\sim}{t}^{k} - F_n'(\underset{\sim}{t}^{k})^{-1}F_n(\underset{\sim}{t}^{k}) \ , \quad k = 0,1,2,\ldots, \quad \text{in } (3.25)$$

unless we show that they are well-defined. We thus ask if $F_n'(\underset{\sim}{t})^{-1}$ exists and if $\|F_n'(\underset{\sim}{t})^{-1}\|$ is bounded in some appropriate neighbourhood of $\underset{\sim}{0}$. In the proof of (4.13) we showed that

(5.8) for all $t \in \overline{D}$, $F'(\underset{\sim}{t})^{-1}$ exists and $\|F'(\underset{\sim}{t})^{-1}\|$ is bounded.

Now, in this section, we will soon show that

$$(5.9) \quad \sup\{\|F_n'(\underset{\sim}{t}) - F'(\underset{\sim}{t})\| : t \in \overline{D}\} \xrightarrow{P} 0 \ .$$

Let α satisfy $\|F'(\underset{\sim}{t})^{-1}\| \leq \alpha$ for all $\underset{\sim}{t} \in \overline{D}$. Choose any β such that $\beta < \frac{1}{\alpha}$. Then, by (5.8),

$$\text{(5.10)} \qquad \lim_{n\to\infty} P\{w:\sup\{\|F'_{n,w}(\underset{\sim}{t}) - F'(\underset{\sim}{t})\|:\underset{\sim}{t} \in \overline{D}\} < \beta\} = 1 .$$

By (5.10) and the perturbation lemma (2.5), we see that for any $\delta(0 < \delta < 1)$ we can find $N(\delta)$ such that $n \geq N$ implies $F'_n(\underset{\sim}{t})$ is invertible on $\overline{D}$ and $\|F'_n(\underset{\sim}{t})^{-1}\|$ is bounded (by $\frac{\alpha}{1-\alpha\beta}$), with probability $\geq 1 - \delta$. Thus, in this sense, (5.7) is well-defined. Note that in showing $F'_n(\underset{\sim}{t})^{-1}$ is bounded in probability, we avoided explicitly calculating the matrix $F'_n(\underset{\sim}{t})^{-1}$. In all our future work we shall avoid doing this, simply because $F'_n(\underset{\sim}{t})^{-1}$ does not have the kind of factorization we found for $F'(\underset{\sim}{t})^{-1}$ (see (4.23)).

<u>(5.11) DEFINITION:</u>

Let $C(\overline{D})$ *be the space of continuous functions from* $\overline{D}$ *into* $\mathbb{R}'$ *. We equip* $C(\overline{D})$ *with the uniform topology, i.e., the topology induced by the metric* d *defined by*

$$d(g,h) = \sup\{|g(\underset{\sim}{t}) - h(\underset{\sim}{t})|:\underset{\sim}{t} \in \overline{D}\}, \quad \textit{where} \quad g,h \in C(\overline{D}) .$$

Thus, a set $E \subseteq C(\overline{D})$ is open (i.e., in the topology) if each point (function) $g \in E$ is contained in a ball

$$B_g(\delta) = \{h \in C(\overline{D}):d(h,g) < \delta\} \text{ contained in } E,$$

i.e., $g \in B_g(\delta) \subseteq E$.

It is elementary to show that $(C(\overline{D}),d)$ is complete. It is also separable: one countable dense subset consists of the polygonal functions that are linear on each of the sets $[a_1,b_1] \times [a_2,b_2] \times\ldots\times [a_p,b_p]$ where $a_i,b_i \in \overline{D} \cap \mathbb{Q}$ and each function takes a rational value at each face (hyper-rectangle) of each of the sets $[a_1,b_1] \times [a_2,b_2] \times\ldots\times [a_p,b_p]$. Here X denotes Cartesian product and $\mathbb{Q}$ the set of rational numbers.

We make one final remark before commencing the consistency proof.

(5.12) REMARK: (In this remark, all references to theorems and pages are to Billingsley (1968), as well as the notation we use.)

In the Arzela-Ascoli criterion for tightness in Theorem 8.2, p. 55, there is nothing special about the point 0 in condition (i) of that theorem. To see that we could just as well use tightness at any point $t_0 \in [0,1]$, we go to the proof of the Arzela-Ascoli theorem on p. 221 and note that the uniform boundedness of A could be got from the inequality

$$|x(t)| \leq |x(t_0)| + \sum_{i=1}^{k} \left| x(\tfrac{i}{k}[t-t_0]+t_0) - x(\tfrac{i-1}{k}[t-t_0]+t_0) \right| .$$

Now we note that for each $i = 1,\ldots,k$, $\frac{i}{k}[t-t_0] + t_0$ and $\frac{i-1}{k}[t-t_0] + t_0$ lie in [0,1] when t and t_0 do (e.g., both are convex combinations of t and t_0 and so lie in the convex set [0,1]). Hence, if condition (9) on p. 221 holds and condition (8) on p. 221 holds with 0 replaced by t_0, we get, from the inequality above, the bound (10) of p. 221 in exactly the same way as it was derived there. The rest of the proof of the Arzela-Ascoli theorem is then identical with that given on p. 221.

An obvious generalization of the above comment to our space $C(\overline{D})$ holds - but if our space was $D[0,1]$ (see p. 109 - we are still referring to Billingsley (1968)), or p-dimensional version of $D[0,1]$, then the condition (i) of Theorem 8.2, p. 55 must be replaced by a stronger one.

We now start on a long chain of lemmas leading to our consistency proof.

(5.13) LEMMA:

$$\sup\{\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| : \underset{\sim}{t} \in \overline{D}\} \xrightarrow{P} 0 \ .$$

Proof:

From (5.2) and (5.3) we have

$$(5.14) \qquad F_n(\underset{\sim}{t}) = \begin{pmatrix} f_{n,1}(\underset{\sim}{t}) \\ \vdots \\ f_{n,p}(\underset{\sim}{t}) \end{pmatrix} \quad \text{and} \quad F(\underset{\sim}{t}) = \begin{pmatrix} f_1(\underset{\sim}{t}) \\ \vdots \\ f_p(\underset{\sim}{t}) \end{pmatrix} .$$

We prove the lemma in several steps:

Step I: We show

$$(5.15) \qquad \text{for each } \underset{\sim}{t} \in \overline{D}\ , \quad f_{n,k}(\underset{\sim}{t}) \xrightarrow{P} f_k(\underset{\sim}{t}) \ ,$$
$$\text{for each } k = 1,\ldots,p \ .$$

Note that $f_{n,k}$: sample space $\Omega \to C(\overline{D})$ is measurable, i.e., for each $w \in \Omega$, $f_{n,k,w}$ is a random element of $C(\overline{D})$, because:

a) each projection $\pi_{\underset{\sim}{t}_1,\ldots,\underset{\sim}{t}_r}$ defined by $\pi_{\underset{\sim}{t}_1,\ldots,\underset{\sim}{t}_r}(f_{n,k}) = (f_{n,k}(\underset{\sim}{t}_1),\ldots,f_{n,k}(\underset{\sim}{t}_r))$ is easily seen to be continuous by continuity of ψ. a) says that all sample paths are continuous,

b) for each fixed $\underset{\sim}{t} \in \overline{D}$, $f_{n,k}(\underset{\sim}{t})$ is a random variable

and c) $C(\overline{D})$ is separable (see (5.11) and Billingsley (1968), p. 57).

To avoid undue length in some of the future lemmas, we will often not even state that a given function is random, because it will be clear.

Proof of (5.15):

We have

$$f_{n,k}(\underset{\sim}{t}) = \frac{1}{n}\sum_{i=1}^{n} c_{ik}\ \psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right) \quad \text{and} \tag{5.16}$$

$$f_k(\underset{\sim}{t}) = \sum_{i=1}^{p} a_{ik}q_i\ E\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right) . \tag{5.17}$$

Note that the W.L.L.N$\underline{\text{s}}$ for i.i.d. random variables does not immediately apply to the average in (5.16) since the coefficients c_{ik} of ψ are not all equal. We write

$$f_{n,k}(\underset{\sim}{t}) = \frac{1}{n}\left\{\sum_{s=1}^{nq_1} c_{sk}\ \psi\left(X_s - \sum_{j=1}^{p} c_{sj}t_j\right)\right. \tag{5.18}$$

$$\left. + \sum_{s=nq_1+1}^{nq_1+nq_2} c_{ik}\ \psi\left(X_s - \sum_{j=1}^{p} c_{sj}tj\right) + \ldots + \sum_{s=n\sum_{r=1}^{p-1}q_r+1}^{n} c_{sk}\ \psi\left(X_s - \sum_{j=1}^{p} c_{sj}t_j\right)\right\}$$

$$= \frac{1}{n}\left\{a_{1k}\sum_{s=1}^{nq_1}\psi\left(X_s-\sum_{j=1}^{p}a_{1j}t_j\right) + a_{2k}\sum_{s=nq_1+1}^{nq_1+nq_2}\psi\left(X_s-\sum_{j=1}^{p}a_{2k}t_j\right) + \ldots\right.$$

$$\left. + a_{pk}\sum_{s=n\sum_{r=1}^{p-1}q_r+1}^{n}\psi\left(X_s-\sum_{j=1}^{p}a_{pj}t_j\right)\right\}$$

$$= \sum_{i=1}^{p}a_{ik}q_i\left[\frac{1}{nq_i}\sum_{s=n\sum_{r=1}^{i-1}q_r+1}^{n\sum_{r=1}^{i}q_r}\psi\left(X_s-\sum_{j=1}^{p}a_{ij}t_j\right)\right],$$

where $\sum_{r=1}^{i-1} q_r$ is defined to be 0 when $i = 1$.

Now, the term in square brackets after the last equality of (5.18) is the average of nq_i independent and identically distributed random variables with expectation

$$E\psi(X - \sum_{j=1}^{p} a_{ij}t_j) \ .$$

By the W.L.L.N$\underline{\text{s}}$, this average converges in probability to

$$E\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right) .$$

(Note that partitions like that in (5.18) are needed frequently in our work before applying the W.L.L.N$\underline{\text{s}}$. However, we will not always mention that this partitioning has been done.)

Then, by an elementary result on convergence in probability, we see that

$$f_{n,k}(\underset{\sim}{t}) \xrightarrow{P} \sum_{i=1}^{p} a_{ik}q_i \; E\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right)$$

i.e., $f_{n,k}(\underset{\sim}{t}) \xrightarrow{P} f_k(\underset{\sim}{t})$, completing Step I.

Note that the method above, of showing that $f_{n,k}(\underset{\sim}{t}) \xrightarrow{P} f_k(\underset{\sim}{t})$, also shows that

(5.19) $$E\, f_{n,k}(\underset{\sim}{t}) = f_k(\underset{\sim}{t})$$

and so,

(5.20) $$E\, F_n(\underset{\sim}{t}) = F(\underset{\sim}{t}) .$$

<u>Step II</u>: We show, for each $\underset{\sim}{t} \in \overline{D}$, that

(5.19) $$F_n(\underset{\sim}{t}) \xrightarrow{P} F(\underset{\sim}{t}) .$$

<u>Proof of (5.19)</u>:

This is actually trivial: e.g., using the l_1-norm, we have

$$\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\|_1 = \sum_{k=1}^{p} |f_{n,k}(\underset{\sim}{t}) - f_k(\underset{\sim}{t})| \xrightarrow{P} 0 , \quad \text{by (5.15).}$$

This completes Step II.

Of course, component-wise convergence in probability of a sequence of random vectors can be taken as the definition of convergence in probability of the random vector (see Cramer (1946), p. 299).

Step III: (Convergence of the finite-dimensional distributions of F_n to those of F.)

We show, for any $\underset{\sim}{t}_1, \underset{\sim}{t}_2, \ldots, \underset{\sim}{t}_s \in \overline{D}$,

(5.20) $$(F_n(\underset{\sim}{t}_1), \ldots, F_n(\underset{\sim}{t}_s)) \xrightarrow{\mathcal{D}} (F(\underset{\sim}{t}_1), \ldots, F(\underset{\sim}{t}_s)) .$$

Here "$\mathcal{D}$" denotes convergence in distribution.

Proof of (5.20):

By (4.24), for $r = 1,2,\ldots,s$,

$$F_n(\underset{\sim}{t}_r) \xrightarrow{P} F(\underset{\sim}{t}_r) .$$

Equivalently, since $F(\underset{\sim}{t}_r)$ is a degenerate random vector,

(5.21) $$F_n(\underset{\sim}{t}_r) \xrightarrow{\mathcal{D}} F(\underset{\sim}{t}_r) .$$

From (5.21) we get

(5.22) $\sum_{r=1}^{s} \alpha_r F_n(\underset{\sim}{t}_r) \xrightarrow{\mathcal{D}} \sum_{r=1}^{s} \alpha_r F(\underset{\sim}{t}_r)$ for any scalars α_r $(r=1,\ldots,s)$.

By (5.22) and the Cramer-Wold Theorem, (5.20) holds.

This completes Step III.

Step IV: (Relative compactness of the sequence of distributions corresponding to the F_n. Equivalently (-Prohorov), tightness of the sequence of distributions corresponding to the F_n, i.e., tightness of the sequence $\{F_n\}$.)

To establish this tightness, we use the p-dimensional analogue of Theorem 8.2 in Billingsley (1968), As shown in (5.12) we can replace the point 0 in that theorem by an arbitrary point.

We split Step IV into two parts.

Step IV, part a): (Tightness at a single point $\underset{\sim}{t}_0$.)

We show, for each positive η, there exists α such that

(5.23) $$P\{\|F_n(\underset{\sim}{t}_0)\| > \alpha\} \leq \eta \ \forall\ n \geq 1 .$$

Proof of (5.23):

Let $\varepsilon > 0$. By Step II, $F_n(\underset{\sim}{t}_0) \xrightarrow{P} F(\underset{\sim}{t}_0)$ and so, $\exists\ n_0$ such that

(5.24) $$n \geq n_0 \Rightarrow P\{\|F_n(\underset{\sim}{t}_0) - F(\underset{\sim}{t}_0)\| > \varepsilon\} \leq \eta .$$

Now, for any ω,

$$\|F_{n,\omega}(\underset{\sim}{t}_0) - F(\underset{\sim}{t}_0)\| \geq \|F_{n,\omega}(\underset{\sim}{t}_0)\| - \|F(\underset{\sim}{t}_0)\|$$

so that the event

$\|F_n(\underset{\sim}{t}_0)\| - \|F(\underset{\sim}{t}_0)\| > \varepsilon$ implies the event $\|F_n(\underset{\sim}{t}_0) - F(\underset{\sim}{t}_0)\| > \varepsilon$.

Thus,

(5.25) $$P\{\|F_n(\underset{\sim}{t}_0)\| - \|F(\underset{\sim}{t}_0)\| > \varepsilon\} \leq P\{\|F_n(\underset{\sim}{t}_0) - F(\underset{\sim}{t}_0)\| > \varepsilon\} .$$

By (5.24) and (5.25), we have

(5.26) $$P\{\|F_n(\underset{\sim}{t}_0)\| > \varepsilon + \|F(\underset{\sim}{t}_0)\|\} \leq \eta \ \forall\ n \geq n_0 .$$

Next, for $i = 1,\ldots,n_0-1$, we chose ε_i so that

(5.27) $$P\{\|F_i(\underset{\sim}{t}_0)\| > \varepsilon_i\} \leq \eta$$

[(5.27) is clearly possible since a random variable is, by most definitions, finite a.e. - of course here we could, if we prefer, appeal to the fact that F is continuous on $\overline{D}$, so that F is bounded on $\overline{D}$. But $EF_i(\underset{\sim}{t}_0) = F(\underset{\sim}{t}_0)$, by (5.20) with i replacing n. Thus $\|EF_i(\underset{\sim}{t}_0)\| < \infty$.

Thus $|Ef_{ik}(\underset{\sim}{t}_0)| < \infty$, $k = 1,2,\ldots,p$ where the f_{ik} are the components of F_i.

Thus $Ef_{ik}(\underset{\sim}{t}_0) < \infty$ and by e.g., Chung (1974), p. 41, we have $E|f_{ik}(\underset{\sim}{t}_0)| < \infty$. Thus $E\|f_i(\underset{\sim}{t}_0)\| < \infty$ and this implies clearly that $\|F_i(\underset{\sim}{t}_0)\| < \infty$ a.e. so that (5.27) is possible.]

Now choosing $\alpha \geq \max\{\varepsilon + \|F(\underset{\sim}{t}_0)\|$, $\varepsilon_1,\ldots,\varepsilon_{n_0-1}\}$, we have from (5.26) and (5.27),

$$P\{\|F_n(\underset{\sim}{t}_0)\| > \alpha\} \leq \eta \ \forall\ n \geq 1 \text{ , proving (5.23).}$$

This completes Step IV, part a).

<u>Step IV, part b):</u> (With arbitrarily high probability, the random functions are each - for large n - uniformly equicontinuous.)

We show, for each $\varepsilon > 0$,

$$\text{(5.28)} \quad \lim_{\delta\to 0} \limsup_{n\to\infty} P\{\sup\{\|F_n(\underset{\sim}{s}) - F_n(\underset{\sim}{t})\| : \|\underset{\sim}{s}-\underset{\sim}{t}\| < \delta,\ \underset{\sim}{t},\underset{\sim}{s} \in \overline{D}\} \geq \varepsilon\} = 0 \ .$$

<u>Proof of (5.28):</u>

For convenience here, we use the l_∞-norm for vectors in $\mathbb{R}^p$, i.e., if $\underset{\sim}{a} = (a_1,a_2,\ldots,a_p)^T$, then $\|\underset{\sim}{a}\|_\infty = \max\limits_{1\leq i\leq p} |a_i|$.

Let $\varepsilon > 0$, $\eta > 0$.

Since ψ is uniformly continuous as a mapping from $\overline{D}$ to $\mathbb{R}^1$, $\exists\ \delta > 0$ such that

$$(5.29) \qquad \sup\{|\psi(\underset{\sim}{a}) - \psi(\underset{\sim}{b})| : \|\underset{\sim}{a}-\underset{\sim}{b}\|_\infty < \delta\} < \frac{\varepsilon}{\max\limits_{\substack{i,k \\ 1\le i\le p \\ 1\le k\le p}} |c_{ik}|} .$$

Then, for $\|\underset{\sim}{s}-\underset{\sim}{t}\|_\infty < \delta$, $\underset{\sim}{s},\underset{\sim}{t} \in \overline{D}$,

$$\|F_n(\underset{\sim}{s})-F_n(\underset{\sim}{t})\|_\infty = \max_{1\le k\le p}\left|\frac{1}{n}\sum_{i=1}^{n} c_{ik}\ \psi\left(X_i-\sum_{j=1}^{p} c_{ij}s_j\right) - \frac{1}{n}\sum_{i=1}^{n} c_{ik}\ \psi\left(X_i-\sum_{j=1}^{p} c_{ij}t_j\right)\right|$$

$$\le \max_{1\le k\le p} \frac{1}{n}\sum_{i=1}^{n} |c_{ik}|\ \left|\psi\left(X_i-\sum_{j=1}^{p} c_{ij}s_j\right) - \psi\left(X_i-\sum_{j=1}^{p} c_{ij}t_j\right)\right|$$

$$< \frac{\varepsilon}{\left(\max\limits_{i,k} |c_{ik}|\right)} \max_{1\le k\le p} \frac{1}{n}\sum_{i=1}^{n} |c_{ik}| \quad , \quad \text{by (5.29)} \ ;$$

$$\le \frac{\varepsilon}{\max\limits_{i,k} |c_{ik}|} \frac{1}{n}\sum_{i=1}^{n} \max_{k}|c_{ik}| \le \frac{\varepsilon}{\max\limits_{i,k} |c_{ik}|} \frac{1}{n}\sum_{i=1}^{n} \max_{i,k}|c_{ik}| = \varepsilon$$

and so,

$$P\{\sup\{\|F_n(\underset{\sim}{s}) - F_n(\underset{\sim}{t})\|_\infty : \|\underset{\sim}{s}-\underset{\sim}{t}\|_\infty < \delta \ , \ \ \underset{\sim}{s},\underset{\sim}{t} \in \overline{D}\} \ge \varepsilon\} = 0 < \eta \text{ for all } n \ge 1 \ ,$$

so certainly (5.28) holds.

This completes Step IV, part b) and so Step IV is completed.

Step V: The proof of the lemma now follows from Steps III and IV and the p-dimensional generalization of Theorem 8.1 in Billingsley (1968).

This completes the proof of the lemma. □

(5.30) LEMMA:

(5.31) $$\sup\{\|F_n'(\underset{\sim}{t}) - F'(\underset{\sim}{t})\| : t \in \overline{D}\} \xrightarrow{P} 0 .$$

Proof:

Here we will use the maximum column sum norm for matrices, as given in (2.4). For the remainder of this section, we adopt the notation:

(5.32) $$\lambda_i(\underset{\sim}{t}) \text{ in place of } E\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right) .$$

With this notation we have, (see (4.8),

(5.33) $$F'(\underset{\sim}{t}) = \begin{pmatrix} \sum_{i=1}^{p} a_{i1}q_i \frac{\delta\lambda_i(\underset{\sim}{t})}{\delta t_1} & ,\dots, & \sum_{i=1}^{p} a_{i1}q_i \frac{\delta\lambda_i(\underset{\sim}{t})}{\delta t_p} \\ \vdots & & \vdots \\ \sum_{i=1}^{p} a_{ip}q_i \frac{\delta\lambda_i(\underset{\sim}{t})}{\delta t_1} & ,\dots, & \sum_{i=1}^{p} a_{ip}q_i \frac{\delta\lambda_i(\underset{\sim}{t})}{\delta t_p} \end{pmatrix} .$$

We also have

(5.34) $$F_n'(\underset{\sim}{t}) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} c_{i1} \frac{\delta\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)}{\delta t_1} & ,\dots, & \frac{1}{n}\sum_{i=1}^{n} c_{i1} \frac{\delta\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)}{\delta t_p} \\ \vdots & & \vdots \\ \frac{1}{n}\sum_{i=1}^{n} c_{ip} \frac{\delta\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)}{\delta t_1} & ,\dots, & \frac{1}{n}\sum_{i=1}^{n} c_{ip} \frac{\delta\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)}{\delta t_p} \end{pmatrix} .$$

We now proceed with the proof of (5.30) in a manner similar to the proof of (5.13).

Step I: We show, for each $\underset{\sim}{t} \in \overline{D}$,

$$\frac{1}{n}\sum_{i=1}^{n} c_{il} \frac{\delta\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)}{\delta t_k} \xrightarrow{P} \sum_{i=1}^{p} a_{il}q_i \frac{\delta\lambda_i(\underset{\sim}{t})}{\delta t_k} \tag{5.35}$$

for all $l,k = 1,\dots,p$.

Proof of (5.35):

We have

$$\frac{1}{n}\sum_{i=1}^{n} c_{il} \frac{\delta\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)}{\delta t_k}$$

$$= \sum_{i=1}^{p} a_{il}q_i \left[\frac{1}{nq_i} \sum_{s=n\sum_{r=1}^{i-1} q_r+1}^{n\sum_{r=1}^{i} q_r} \frac{\delta\psi\left(X_s - \sum_{j=1}^{p} a_{ij}t_j\right)}{\delta t_k}\right] \quad \text{(see (5.18))}$$

$$\xrightarrow{P} \sum_{i=1}^{p} a_{il}q_i\, E\, \frac{\delta\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right)}{\delta t_k}$$, by W.L.L.N$^{\underline{s}}$ applied to each of the terms in squared brackets above

$$= \sum_{i=1}^{p} a_{il}q_i \frac{\delta\lambda_i(\underset{\sim}{t})}{\delta t_k}$$, proving (5.35).

This completes Step I.

Step II: We show

$$\text{for each } \underset{\sim}{t} \in \overline{D}\,, \quad F_n'(\underset{\sim}{t}) \xrightarrow{P} F'(\underset{\sim}{t}) . \tag{5.36}$$

Proof of (5.36):

We have

$$\|F_n'(\underset{\sim}{t}) - F'(\underset{\sim}{t})\|_1 = \max_{1\le k\le p} \sum_{l=1}^{p} \left| \frac{1}{n} \sum_{i=1}^{n} c_{il} \frac{\delta\psi\left(X_i - \sum_{j=1}^{p} c_{ij} t_j\right)}{\delta t_k} - \sum_{i=1}^{p} a_{il} q_i \frac{\delta\lambda_i(\underset{\sim}{t})}{\delta t_k} \right|$$

from (2.4), (5.33) and (5.34) ;

$$\xrightarrow{P} \max_{1\le k\le p} \sum_{l=1}^{p} 0 \;, \quad \text{by (5.35)} ;$$

$$= 0 \;, \quad \text{proving (5.36).}$$

Again, Step II is a bit superfluous, since we could have defined convergence of random matrices by element convergence. This completes Step II.

Step III: We show, for any $\underset{\sim}{t}_1, \underset{\sim}{t}_2, \ldots, \underset{\sim}{t}_s \in \overline{D}$,

(5.37) $$(F_n'(\underset{\sim}{t}_1), \ldots, F_n'(\underset{\sim}{t}_s)) \xrightarrow{\mathcal{D}} (F(\underset{\sim}{t}_1), \ldots, F(\underset{\sim}{t}_s)) \;.$$

Proof of (5.37):

By (5.36), for $r=1,\ldots,s$, $F_n'(\underset{\sim}{t}_r) \xrightarrow{P} F'(\underset{\sim}{t}_r)$. Equivalently, since $F'(\underset{\sim}{t}_r)$ is a degenerate random matrix,

(5.38) $$F_n'(\underset{\sim}{t}_r) \xrightarrow{\mathcal{D}} F'(\underset{\sim}{t}_r) \;.$$

From this we get

(5.39) $$\sum_{r=1}^{s} \alpha_r F_n'(\underset{\sim}{t}_r) \xrightarrow{\mathcal{D}} \sum_{r=1}^{s} \alpha_r F'(\underset{\sim}{t}_r) \text{ for any scalars } \alpha_r \; (r=1,\ldots,s).$$

By (5.39) and the Cramer-Wold theorem, (5.37) holds. This completes Step III.

Step (IV): Relative compactness of the sequence of distributions corresponding to the F_n' . As with Step IV of Lemma 4.1, we split Step IV here into two parts:

Step IV, part a): We show, for each positive η , $\exists$ α such that

$$(5.40) \qquad P\{\|F_n'(\underset{\sim}{t}_0)\| > \alpha\} \le \eta \quad \text{for all} \quad n \ge 1 \ .$$

Here $\underset{\sim}{t}_0$ is any point of our choosing for which we can get (5.40) to hold, but since (5.40) actually holds for each $\underset{\sim}{t} \in \overline{D}$, the $\underset{\sim}{t}_0$ in (5.40) above is arbitrary (but fixed).

Proof of (5.40):

Let $\varepsilon > 0$. By Step II, $F_n'(\underset{\sim}{t}_0) \xrightarrow{P} F'(\underset{\sim}{t}_0)$ and so, there exists n_0 such that

$$(5.41) \qquad n \ge n_0 \Rightarrow P\{\|F_n'(\underset{\sim}{t}_0) - F'(\underset{\sim}{t}_0)\| > \varepsilon\} \le \eta \ .$$

Exactly as in the calculations leading from (5.24) to (5.26), we have from (5.41),

$$(5.42) \qquad P\{\|F_n'(\underset{\sim}{t}_0)\| > \varepsilon + \|F'(\underset{\sim}{t}_0)\|\} \le \eta \quad \text{for all } n \ge n_0 \ .$$

Next, for $i = 1,\ldots,n_0-1$, we choose ε_i so that

$$(5.43) \qquad P\{\|F_i'(\underset{\sim}{t}_0)\| > \varepsilon_i\} \le \eta \ .$$

(5.43) is justified in exactly the same way as (5.27) (see the bracketed comments following (5.27)).

Now choose any α such that

$$\alpha \geq \max\{\varepsilon + \|F'(\underset{\sim}{t}_0)\| \;,\; \varepsilon_1,\ldots,\varepsilon_{n_0-1}\}$$

and then we have, from (5.42) and (5.43),

$$P\{\|F_n'(\underset{\sim}{t}_0)\| > \alpha\} \leq \eta \quad \text{for all } n \geq 1 \quad \text{proving (5.40).}$$

This completes Step IV, part a).

<u>Step IV, part b):</u> We show, for each $\varepsilon > 0$,

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \{\sup\{\|F_n'(\underset{\sim}{s}) - F'(\underset{\sim}{t})\| : \|\underset{\sim}{s}-\underset{\sim}{t}\| < \delta,\ \underset{\sim}{t},\underset{\sim}{s} \in \overline{D}\} \geq \varepsilon\} = 0 \;. \tag{5.44}$$

<u>Proof of (5.44):</u>

Let $\varepsilon > 0$, $\eta > 0$. Since ψ' is uniformly continuous on $\overline{D}$,

so is $\displaystyle \delta_k\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right) = \frac{\delta\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)}{\delta t_k}$,

since $\displaystyle \delta_k\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right) = -\, c_{ik}\ \psi'\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)$.

Thus $\exists\ \delta > 0$ such that $\forall\ i,k = 1,2,\ldots,p$:

$$\sup\left\{\left|\delta_k\psi\left(X_i - \sum_{j=1}^{p} c_{ij}a_j\right) - \delta_k\psi\left(X_i - \sum_{j=1}^{p} c_{ij}b_j\right)\right| : \|\underset{\sim}{a}-\underset{\sim}{b}\| < \delta\right\} < \frac{\varepsilon}{p\,\max\limits_{i,r}|c_{ir}|} \;. \tag{5.45}$$

Then for $\|\underset{\sim}{s}-\underset{\sim}{t}\| < \delta$, we have

$$\|F_n'(\underset{\sim}{s}) - F_n'(\underset{\sim}{t})\|_1$$

$$= \max_{1\le k\le p} \sum_{r=1}^{p} \left| \frac{1}{n} \sum_{i=1}^{n} c_{ir}\delta_k\psi\left(X_i - \sum_{j=1}^{p} c_{ij}s_j\right) - \frac{1}{n} \sum_{i=1}^{n} c_{ir}\delta_k\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right) \right| ,$$

by (2.4) and (5.34) ;

$$\le \max_k \sum_{r=1}^{p} \frac{1}{n} \sum_{i=1}^{n} |c_{ir}| \left|\delta_k\psi\left(X_i - \sum_{j=1}^{p} c_{ij}s_j\right) - \delta_k\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)\right|$$

$$< \left(\max_k \sum_{r=1}^{p} \frac{1}{n} \sum_{i=1}^{n} |c_{ir}|\right) \left(\frac{\varepsilon}{p \max_{i,r}|c_{ir}|}\right) , \text{ by (5.45) ;}$$

$$= \left(\sum_{r=1}^{p} \frac{1}{n} \sum_{i=1}^{n} |c_{ir}|\right) \left(\frac{\varepsilon}{p \max_{i,r}|c_{ir}|}\right)$$

$$\le \left(\sum_{r=1}^{p} \frac{1}{n} \sum_{i=1}^{n} \max_i |c_{ir}|\right) \left(\frac{\varepsilon}{p \max_{i,r}|c_{ir}|}\right)$$

$$= \left(\sum_{r=1}^{p} \max_i |c_{ir}|\right) \left(\frac{\varepsilon}{p \max_{i,r}|c_{ir}|}\right)$$

$$\le \left(p \max_{i,r} |c_{ir}|\right) \left(\frac{\varepsilon}{p \max_{i,r}|c_{ir}|}\right)$$

$$= \varepsilon .$$

Thus,

$$P\{\sup\{\|F_n'(\underset{\sim}{s}) - F_n'(\underset{\sim}{t})\|_1 : \|\underset{\sim}{s}-\underset{\sim}{t}\| < \delta,\ \underset{\sim}{s},\underset{\sim}{t} \in \overline{D}\} \ge \varepsilon\} = 0 < \eta,\ \forall\, n \ge 1,$$

and so certainly (5.44) holds.

This completes Step IV, part b) and so Step IV is completed.

Step V: The proof of the lemma now follows from Steps III and IV and a generalization of Theorem 8.1 in Billingsley (1968).

This completes the proof of the lemma. □

(5.46) LEMMA:

F is continuous on $\overline{D}$ and satisfies $\underset{\sim}{t}^T F(\underset{\sim}{t}) \le 0$ for all $\underset{\sim}{t} \in \dot{D}$ where $\dot{D}$ denotes the boundary of D.

Proof:

Continuity of F was shown earlier (see (4.7) and the second proof of the boundedness of $\|F'(\underset{\sim}{t})^{-1}\|$ in (4.13d)). Now, for the second assertion, we have

$$F(\underset{\sim}{t}) = \begin{pmatrix} \sum_{i=1}^{p} a_{i1} q_i \; E\psi\left(X - \sum_{j=1}^{p} a_{ij} t_j\right) \\ \vdots \\ \sum_{i=1}^{p} a_{ip} q_i \; E\psi\left(X - \sum_{j=1}^{p} a_{ij} t_j\right) \end{pmatrix},$$

so that

$$\underset{\sim}{t}^T F(\underset{\sim}{t}) = \sum_{k=1}^{p} t_k \sum_{i=1}^{p} a_{ik} q_i \; E\psi\left(X - \sum_{j=1}^{p} a_{ij} t_j\right)$$

$$= \sum_{i=1}^{p} q_i \left\{ \left[E\psi\left(X - \sum_{j=1}^{p} a_{ij} t_j\right) \right] \left[\sum_{k=1}^{p} a_{ik} t_k \right] \right\} .$$

Now

$$\sum_{j=1}^{p} a_{ij} t_j > 0 \implies E\psi\left(X - \sum_{j=1}^{p} a_{ij} t_j\right) < 0$$

while $$\sum_{j=1}^{p} a_{ij} t_j < 0 \implies E\psi\left(X - \sum_{j=1}^{p} a_{ij} t_j\right) > 0 .$$

This is clear from (8.29) further out with $\sigma = 1$. (Note that in (8.29) Δ denotes any of the $X - \sum_{j=1}^{p} a_{ij} t_j$ and t is written for any of the $\sum_{j=1}^{p} a_{ij} t_j$,

$i=1, \ldots, p$.) It also follows from the work of Collins (1973). Thus each term in curly brackets in the expression above for $\underset{\sim}{t}^T F(\underset{\sim}{t})$ is negative and the lemma follows. Note that we actually have

$$\underset{\sim}{t}^T F(\underset{\sim}{t}) \le 0 \quad \text{for all} \quad \underset{\sim}{t} \in \overline{D} \text{ , not just } \underset{\sim}{t} \in \dot{D}$$

and we have strict inequality except when $\sum_{j=1}^{p} a_{ij} t_j = 0$ for all $i = 1, \ldots, p$, i.e. except when $\underset{\sim}{t} = \underset{\sim}{0}$ (since A is invertible). □

<u>(5.47) LEMMA</u>:

Let the event $E_{1,n}$ be defined by

(5.48) $$E_{1,n} = \{w : \underset{\sim}{t}^T F_n(\underset{\sim}{t}, w) \le 0 \quad \text{for all} \quad \underset{\sim}{t} \in \dot{D}$$

$$\text{and } F_n(\underset{\sim}{t}_1, w) = F_n(\underset{\sim}{t}_2, w) \quad \text{implies}$$

$$\underset{\sim}{t}_1 = \underset{\sim}{t}_2, \quad \underset{\sim}{t}_1, \underset{\sim}{t}_2 \in \overline{D}\} \text{ .}$$

Then

(5.49) $$P(E_{1,n}) \to 1 \quad \text{as} \quad n \to \infty \text{ .}$$

<u>(5.50) REMARK</u>:

We observe that the event

$$\{\underset{\sim}{t}^T F_n(\underset{\sim}{t}) \le 0 \quad \text{for all} \quad \underset{\sim}{t} \in \dot{D}\}$$

implies, by (2.18) with $\underset{\sim}{t}_0 = \underset{\sim}{0}$, that the equation

(5.51) $$F_n(\underset{\sim}{t}) = \underset{\sim}{0} \quad \text{has a solution in } \overline{D} \text{ .}$$

Further, the event $\{F_n(\underset{\sim}{t}_1) = F_n(\underset{\sim}{t}_2) \Rightarrow \underset{\sim}{t}_1 = \underset{\sim}{t}_2,\ \ \underset{\sim}{t}_1, \underset{\sim}{t}_2 \in \overline{D}\}$ (one-oneness) implies

(5.52) $F_n(\underset{\sim}{t}) = \underset{\sim}{0}$ has at most one root in $\overline{D}$.

Thus the event $E_{1,n}$ can be stated as saying that

(5.53) the system $F_n(\underset{\sim}{t}) = \underset{\sim}{0}$ has exactly one root in $\overline{D}$.

When $E_{1,n}$ obtains ,

(5.54) we denote the unique zero of $F_n(\underset{\sim}{t}) = \underset{\sim}{0}$ by $\underset{\sim}{Z}_n$.

We define $\underset{\sim}{Z}_n$ to be $\underset{\sim}{0}$ otherwise.

We remark also that there are many other events that imply the existence and uniqueness of a root of $F_n(\underset{\sim}{t}) = \underset{\sim}{0}$, but since our consistency proof is already laboriously long, we shall omit a discussion of them.

Proof of (5.47): By (5.13),

(5.55) $$\sup\{\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| : \underset{\sim}{t} \in \overline{D}\} \xrightarrow{P} 0 .$$

Since

$$|t^T F_n(\underset{\sim}{t}) - t^T F(\underset{\sim}{t})| = |t^T (F_n(\underset{\sim}{t}) - F(\underset{\sim}{t}))|$$

$$\leq \|\underset{\sim}{t}^T\|\ \|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| \quad \text{(Schwarz)}$$

$$\leq \sup_{\underset{\sim}{t} \in D} \|t^T\| \ \sup_{\underset{\sim}{t} \in D} \|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\|$$

$$= M \sup_{\underset{\sim}{t} \in D} \| F_n(\underset{\sim}{t}) - F(\underset{\sim}{t}) \| \ , \quad \text{say,}$$

we have, for a given $\varepsilon > 0$,

$$P\{\sup_{\underset{\sim}{t} \in D} |t^T F_n(\underset{\sim}{t}) - t^T F(\underset{\sim}{t})| > \varepsilon\}$$
$$\leq P\{\sup_{\underset{\sim}{t} \in D} \| F_n(\underset{\sim}{t}) - F(\underset{\sim}{t}) \| > \frac{\varepsilon}{M}\}$$
$$\longrightarrow 0 \ (n \to \infty) \quad \text{by (5.55).}$$

That is, for all $\varepsilon > 0$

(5.56) $$P\{\sup_{\underset{\sim}{t} \in D} |\underset{\sim}{t}^T F_n(\underset{\sim}{t}) - \underset{\sim}{t}^T F(\underset{\sim}{t})| > \varepsilon\} \longrightarrow 0 \ (n \to \infty) \ .$$

From (5.56) and (5.46) it is elementary to show that

(5.57) $$P\{\sup\{\underset{\sim}{t}^T F_n(\underset{\sim}{t}) : \underset{\sim}{t} \in \overline{D}\} \leq 0\} \longrightarrow 1 \ (n \to \infty) \ .$$

In particular,

(5.58) $$P\{\sup\{\underset{\sim}{t}^T F_n(\underset{\sim}{t}) : \underset{\sim}{t} \in \dot{D}\} \leq 0\} \longrightarrow 1 \ (n \to \infty) \ .$$

Next we show that the event

(5.59) $$\{F_n(\underset{\sim}{t}_1) = F_n(\underset{\sim}{t}_2) \Rightarrow \underset{\sim}{t}_1 = \underset{\sim}{t}_2 \ , \quad \underset{\sim}{t}_1, \underset{\sim}{t}_2 \in \overline{D}\}$$

has probability tending to 1 as $n \to \infty$.

Here is just one method of proceeding:

let $\underset{\sim}{t}^1, \ldots, \underset{\sim}{t}^p \in D$ and let $\underset{\sim}{f}_1, \ldots, \underset{\sim}{f}_p$, as usual, denote the components of F. The matrix (where as usual δ_i denotes $i^{\underline{\text{th}}}$ partial derivative),

$$N = \begin{bmatrix} \delta_1 f_1(\underset{\sim}{t}^1), & \delta_2 f_1(\underset{\sim}{t}^1), & \ldots, & \delta_p f_1(\underset{\sim}{t}^1) \\ \delta_1 f_2(\underset{\sim}{t}^2), & \delta_2 f_2(\underset{\sim}{t}^2), & \ldots, & \delta_p f_2(\underset{\sim}{t}^2) \\ \vdots & & & \vdots \\ \delta_1 f_p(\underset{\sim}{t}^p), & \delta_2 f_p(\underset{\sim}{t}^p), & \ldots, & \delta_p f_p(\underset{\sim}{t}^p) \end{bmatrix},$$

can be written as

$$N = -A^T \operatorname{Diag}((q_i)) \operatorname{Diag}\left(\left(E\psi_i'\left(X - \sum_{j=1}^{p} a_{ij} t_j^i\right)\right)\right) A$$

(see the analogous expression for $F'(\underset{\sim}{t})$ in (4.14))

and the invertibility of N follows from (4.16) and the invertibility of A.

From this and (2.19) we have that F is one-one on D (note that D is convex). Now replace D by an open convex set $C \supset \overline{D}$ such that (4.16) holds for $\underset{\sim}{t} \in C$. (An examination of (3.3) and e.g. (8.72) shows that such a set C exists.) Then by the argument just used in showing the one-oneness of F on D, we see that F is one-one on C and hence on $\overline{D}$. From this and (5.55), we get (5.59). Combining (5.58) and (5.59), we arrive at (5.48). This completes the proof of the lemma. □

(5.60) LEMMA:

(5.61) $$\sup\{\|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\| : \underset{\sim}{t} \in \overline{D}\} \xrightarrow{P} 0$$

and

(5.62) $$\sup\left\{\left|\|F_n'(\underset{\sim}{t})^{-1}\| - \|F'(\underset{\sim}{t})^{-1}\|\right| : \underset{\sim}{t} \in \overline{D}\right\} \xrightarrow{P} 0 .$$

Proof:

By (5.30),

(5.63) $$\sup\{\|F_n'(\underset{\sim}{t}) - F'(\underset{\sim}{t})\| : \underset{\sim}{t} \in \overline{D}\} \xrightarrow{P} 0 .$$

$$(5.64)\qquad \left|\ \|F_n'(\underset{\sim}{t})^{-1}\| - \|F'(\underset{\sim}{t})^{-1}\|\ \right|$$

$$\leq \|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\| \ ,$$

(by, e.g., Simmons (1963), p. 212) ;

$$= \|F_n'(\underset{\sim}{t})^{-1}\left[F'(\underset{\sim}{t}) - F_n'(\underset{\sim}{t})\right]F'(\underset{\sim}{t})^{-1}\| \ , \quad \text{clearly} \ ;$$

$$\leq \|F_n'(\underset{\sim}{t})^{-1}\| \ \|F'(\underset{\sim}{t}) - F_n'(\underset{\sim}{t})\| \ \|F'(\underset{\sim}{t})^{-1}\| \ .$$

Let $M_1 < \infty$ be such that

$$(5.65)\qquad \|F'(\underset{\sim}{t})^{-1}\| \leq M_1 \quad \text{for all} \quad \underset{\sim}{t} \in \overline{D} \ .$$

[M_1 exists by our work following (4.23).]

Next, by our work in (5.6), there exists $M_2 < \infty$ such that

$$(5.66)\qquad P\{\sup\{\|F_n'(\underset{\sim}{t})^{-1}\| : \underset{\sim}{t} \in \overline{D}\} \leq M_2\} \longrightarrow 1 \ (n \to \infty) \ .$$

Now, let $\varepsilon > 0$. Then

$$P\{\sup_{\underset{\sim}{t}\in D} \left|\ \|F_n'(\underset{\sim}{t})^{-1}\| - \|F'(\underset{\sim}{t})^{-1}\|\ \right| > \varepsilon\}$$

$$\leq P\{\sup_{\underset{\sim}{t}\in D} \|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\| > \varepsilon\} \ , \quad \text{by (5.64)} \ ;$$

$$\leq P\{\sup_{\underset{\sim}{t}\in D} \|F_n'(\underset{\sim}{t})^{-1}\| \ \sup_{\underset{\sim}{t}\in D} \|F'(\underset{\sim}{t}) - F_n'(\underset{\sim}{t})\| \ \sup_{\underset{\sim}{t}\in D} \|F'(\underset{\sim}{t})^{-1}\| > \varepsilon\} \ ,$$

by (5.64) again ;

$$\leq P\{\sup_{\underset{\sim}{t}\in D} \|F_n'(\underset{\sim}{t})^{-1}\| \ \sup_{\underset{\sim}{t}\in D} \|F'(\underset{\sim}{t}) - F_n'(\underset{\sim}{t})\| M_1 > \varepsilon\} \ , \quad \text{by (5.65)}$$

$$= P\{\sup_{\underset{\sim}{t}\in\bar{D}}\|F_n'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t}\in\bar{D}}\|F'(\underset{\sim}{t})-F_n'(\underset{\sim}{t})\| > \frac{\varepsilon}{M_1} \cap \sup_{\underset{\sim}{t}\in\bar{D}}\|F_n'(\underset{\sim}{t})^{-1}\| \le M_2\}$$

$$+ P\{\sup_{\underset{\sim}{t}\in\bar{D}}\|F_n'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t}\in\bar{D}}\|F'(\underset{\sim}{t})-F_n'(\underset{\sim}{t})\| > \frac{\varepsilon}{M_1} \cap \sup_{\underset{\sim}{t}\in\bar{D}}\|F_n'(\underset{\sim}{t})^{-1}\| > M_2\}$$

$$\le P\{M_2 \sup_{\underset{\sim}{t}\in\bar{D}}\|F'(\underset{\sim}{t})-F_n'(\underset{\sim}{t})\| > \frac{\varepsilon}{M_1}\} + P\{\sup_{\underset{\sim}{t}\in\bar{D}}\|F_n'(\underset{\sim}{t})^{-1}\| > M_2\} \ , \quad \text{clearly} \ ;$$

$$\xrightarrow{n\to\infty} 0 + 0 \ , \quad \text{by (5.63) and (5.66)} \ .$$

We have shown that, for each $\varepsilon > 0$,

$$P\{\sup_{\underset{\sim}{t}\in\bar{D}} \left| \ \|F_n'(\underset{\sim}{t})^{-1}\| - \|F'(\underset{\sim}{t})^{-1}\| \ \right| > \varepsilon\}$$

$$\le P\{\sup_{\underset{\sim}{t}\in\bar{D}}\|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\| > \varepsilon\}$$

$\longrightarrow 0 \ (n \to \infty)$. Thus (5.16) and (5.62) are proved.

This completes the proof of the lemma. □

[Note that in the above proof we again (see (5.6)) avoided explicitly computing the matrix $F_n'(\underset{\sim}{t})^{-1}$.]

<u>(5.67) LEMMA</u>:

(5.68) $$\sup\{\|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t}) - F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\| : \underset{\sim}{t} \in \bar{D}\} \xrightarrow{P} 0 \ .$$

<u>Proof</u>:

We have

$$\|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t}) - F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\|$$

(5.69)
$$\begin{aligned}
&\le \|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t}) - F_n'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\| + \|F_n'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t}) - F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\| \\
&= \|F_n'(\underset{\sim}{t})^{-1}[Fn(\underset{\sim}{t}) - F(\underset{\sim}{t})]\| + \|[F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}]F(\underset{\sim}{t})\| \\
&\le \|F_n'(\underset{\sim}{t})^{-1}\| \; \|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| + \|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\| \; \|F(\underset{\sim}{t})\| .
\end{aligned}$$

The last inequality follows from the fact that if $E:H \to G$ is a linear mapping between normed linear spaces H and G, then

(5.70) for all $\underset{\sim}{t} \in H: \|E\underset{\sim}{t}\| \le \|E\| \; \|\underset{\sim}{t}\|$.

(5.70) is easily seen to be valid upon looking at the last equality in (2.2).

[Note that when E is continuous, $\|E\|$ in (5.70) is a finite number - this follows from the fact that continuity and boundedness are equivalent for linear mappings between normed vector spaces - e.g., Simmons (1963), p. 220, Theorem A.]

Now, let ε be given, say $0 < \varepsilon < 1$. Let $M_1 \in \mathbb{R}^1$ be such that

(5.71) $\|F(\underset{\sim}{t})\| \le M_1 \quad \forall \; \underset{\sim}{t} \in \overline{D}$.

[M_1 exists by continuity of F on the compact set $\overline{D}$.]

Next, find $M_2 \in \mathbb{R}^1$ and $N_1 \in \mathbb{N}$ so that

(5.72) for all $n \ge N_1$: $P\{\sup_{\underset{\sim}{t} \in D}\|F_n'(t)^{-1}\| > M_2\} < \frac{\varepsilon}{3}$.

[This is possible by, e.g., (5.66).]

Next, find $N_2 \in \mathbb{N}$ so that

$$(5.73) \quad \forall \quad n \geq N_2 : P\{\sup_{\underset{\sim}{t} \in D} \|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\| > \frac{\varepsilon}{2M_1}\} < \frac{\varepsilon}{3} .$$

[This is possible by (5.61).]

Finally, choose $N_3 \in \mathbb{N}$ so that,

$$(5.74) \quad \forall \; n \geq N_3 : P\{\sup_{\underset{\sim}{t} \in D} \|F_n \underset{\sim}{t} - F(\underset{\sim}{t})\| > \frac{\varepsilon}{2M_2}\} < \frac{\varepsilon}{3} .$$

[This is possible by (5.13).]

Now let $N \in \mathbb{N}$ be such that $N \geq \max\{N_1,N_2,N_3\}$. Then, for all $n \geq N$:

$$P\{\sup_{\underset{\sim}{t} \in D} \|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t}) - F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\| > \varepsilon\}$$

$$\leq P\{\sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t} \in D}\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| + \sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t} \in D}\|F(\underset{\sim}{t})\| > \varepsilon\},$$

by (5.69) ;

$$\leq P\{\sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t} \in D}\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| > \frac{\varepsilon}{2} \cup \sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t} \in D}\|F(\underset{\sim}{t})\| > \frac{\varepsilon}{2} \}$$

$$\leq P\{\sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t} \in D}\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| > \frac{\varepsilon}{2}\} + P\{\sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t} \in D}\|F(\underset{\sim}{t})\| > \frac{\varepsilon}{2}\}$$

$$< P\{\sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t} \in D}\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| > \frac{\varepsilon}{2}\} + \frac{\varepsilon}{3}, \quad \text{by (5.73) ;}$$

$$= P\{\sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t} \in D}\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| > \frac{\varepsilon}{2} \cap \sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1}\| \leq M_2\}$$

$$+ P\{\sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1}\| \sup_{\underset{\sim}{t} \in D}\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| > \frac{\varepsilon}{2} \cap \sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1}\| > M_2\} + \frac{\varepsilon}{3}$$

$$\leq P\{M_2 \sup_{\underset{\sim}{t} \in D}\|F_n(\underset{\sim}{t}) - F(\underset{\sim}{t})\| > \frac{\varepsilon}{2} + P\{\sup_{\underset{\sim}{t} \in D}\|F_n'(\underset{\sim}{t})^{-1}\| > M_2\} + \frac{\varepsilon}{3}$$

$$< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}, \quad \text{by (5.74) and (5.72)}$$

$$= \varepsilon .$$

We have shown that

$$P\{\sup_{\underset{\sim}{t}\in D}\|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})-F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\| > \varepsilon\} < \varepsilon \quad \forall\, n \geq N \ .$$

This proves (5.68) and the lemma. □

(5.75) LEMMA:

Let E_{2n} be the event

$$\underset{\sim}{\theta}_n^* < \|\underset{\sim}{\theta}^{**}\| + (k - \|\underset{\sim}{\theta}^{**}\|)/4 \ .$$

Then

(5.76) $$P\{E_{2n}\} \longrightarrow 1 \quad (n \to \infty) \ .$$

Proof:

This is immediate from (3.12).

(5.76) LEMMA:

Let ε be fixed, $0 < \varepsilon \leq k$. Then

(5.77) $$P\{\|\underset{\sim}{Z}_n\| \leq \varepsilon\} \xrightarrow{n\to\infty} 1 \ .$$

Proof:

Let $\overline{C} = \{\underset{\sim}{t} \in R^p : \max_{1\leq k\leq p} |\sum_{j=1}^{p} a_{kj}t_j| \leq \varepsilon\}$. Then $\overline{C} \subset \overline{D}$, since $\varepsilon \leq k$. See diagram below.

Let $E_{1,n,\overline{C}}$ be the event

$$\underset{\sim}{t}^T F_n(\underset{\sim}{t}) \leq 0 \quad \forall\, \underset{\sim}{t} \in \dot{C}$$

and F_n is one-one on $\overline{C}$.

The event $E_{1,n,\overline{C}}$ implies the existence and uniqueness of a root $\underset{\sim}{Z}_n$ in $\overline{C}$ (see 5.50). Further

$$(5.78) \qquad P(E_{1,n,\overline{C}}) \longrightarrow 1 \quad (n \to \infty) .$$

The proof of (5.78) is exactly the same as the proof of (5.49) - all we have done here is replaced $\overline{D}$ by $\overline{C}$. The key point is that $\overline{C}$ contains the point $\underset{\sim}{0}$ and $\overline{C} \subseteq \overline{D}$ [the former because, in fact, if $\overline{C}$ does not contain $\underset{\sim}{0}$ then there does not exist $t_0 \in C$ such that $(t-t_0)^T Ft \le 0 \;\; \forall\, t \in \dot{C}$, so that we could not apply (2.18) to ensure the existence of a zero in $\overline{C}$; and the latter, i.e., $\overline{C} \subset \overline{E}$ so that (5.49) applies with $\overline{D}$ replaced by $\overline{C}$].

Since $\underset{\sim}{Z}_n = (Z_{1n}, Z_{2n}, \ldots, Z_{pn})^T \in \overline{C}$, we have

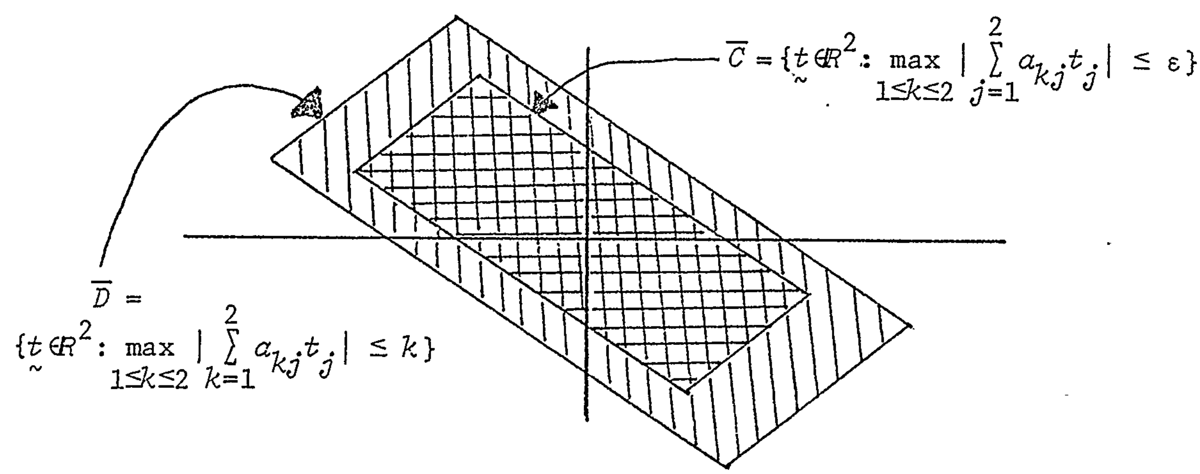
$$\|\underset{\sim}{Z}_n\|_A = \max_{1\le k\le p} \left| \sum_{j=1}^{p} a_{kj} Z_{jn} \right| \le \varepsilon$$

by definition of $\overline{C}$.

From this and (5.78), we obtain (5.77).

This completes the proof of the lemma. □

<u>Diagram showing $\overline{D}$ and $\overline{C}$ (defined in (5.76)) for $p = 2$</u>.



(5.79) LEMMA:

Let $0 < \varepsilon_1 \leq \dfrac{k - \|\underset{\sim}{\theta}^{**}\|}{4}$, $\varepsilon_2 > 0$. Then,

$$(5.80)\quad P\left\{\sup\{\left|\frac{1}{\|\underset{\sim}{t}\|}\|F_n'(\underset{\sim}{t}+\underset{\sim}{Z}_n)^{-1}F_n(\underset{\sim}{t}+\underset{\sim}{Z}_n)\| - \frac{1}{\|\underset{\sim}{t}\|}\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\|\right| : 0 < \|\underset{\sim}{t}\| \leq \frac{\|\underset{\sim}{\theta}^{**}\| + k}{2}\} < \varepsilon_2 \cap \|\underset{\sim}{Z}_n\| < \varepsilon_1\right\} \xrightarrow{(n\to\infty)} 1 .$$

Proof:

Set

$$h(\underset{\sim}{t}) = \begin{cases} \dfrac{1}{\|\underset{\sim}{t}\|}\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\| , & 0 < \|\underset{\sim}{t}\| \leq \dfrac{\|\underset{\sim}{\theta}^{**}\| + k}{2} \\ 1 , & \text{otherwise} \end{cases}$$

and

$$h_n(t) = \begin{cases} \dfrac{1}{\|\underset{\sim}{t}\|}\|F_n'(\underset{\sim}{t}+\underset{\sim}{Z}_n)^{-1}F_n(\underset{\sim}{t}+\underset{\sim}{Z}_n)\| , & 0 < \|\underset{\sim}{t}\| \leq \dfrac{\|\underset{\sim}{\theta}^{**}\| + k}{2} \\ & \underline{\text{and}}\ E_{1,n} \text{ obtains, where } E_{1,n} \\ & \text{is defined in (5.48)} \\ 1 , & \text{otherwise.} \end{cases}$$

By (5.47) and (5.76), we will have proved (5.80) if we prove that, for each $\varepsilon > 0$,

$$(5.81)\quad P\left\{\sup\{|h_n(\underset{\sim}{t}) - h(\underset{\sim}{t})| : \|\underset{\sim}{t}\| \leq \frac{\|\underset{\sim}{\theta}^{**}\| + k}{2}\} < \varepsilon\right\} \xrightarrow{n\to\infty} 1 .$$

We now follow the proof of Lemma 2.6 of Collins (1973) as closely as possible. We use the same letters ε, ε', η, δ_1, etc., as used there, with the same use.

Curly brackets around the number of an equation or statement will mean that we are referring to or calling to mind the equation or etc. with that number in Lemma 2.6 of Collins (1973).

The main complication that arises in our proof is that the expressions {2.57} and {2.59} for the $h(t)$ and $h_n(t)$ of Collins (1973) do not have quite as simple an analogue for our $h(t)$ and $h_n(t)$. This is because the mean value theorem for (differentiable) mappings $g:\mathbb{R}^1 \to \mathbb{R}^1$ does not hold for (G-differentiable) mappings $G:\mathbb{R}^p \to \mathbb{R}^p$ $(p > 1)$, as stated in Chapter 1. However, we may write (see (2.17)), for our F and F_n,

$$F(\underset{\sim}{t}) - F(\underset{\sim}{s}) = B(\underset{\sim}{s},\underset{\sim}{t})(\underset{\sim}{t}-\underset{\sim}{s}) \tag{5.82}$$

$$F_n(\underset{\sim}{t}) - F_n(\underset{\sim}{s}) = B_n(\underset{\sim}{s},\underset{\sim}{t})(\underset{\sim}{t}-\underset{\sim}{s}) \tag{5.83}$$

for all $\underset{\sim}{s},\underset{\sim}{t} \in \mathbb{R}^p$, where

$$B(\underset{\sim}{s},\underset{\sim}{t}) = \begin{bmatrix} \delta_1 f_1(\underset{\sim}{s}+\alpha_1(\underset{\sim}{t}-\underset{\sim}{s})), & \cdots & ,\delta_p f_1(\underset{\sim}{s}+\alpha_1(\underset{\sim}{t}-\underset{\sim}{s})) \\ \vdots & & \vdots \\ \delta_1 f_p(\underset{\sim}{s}+\alpha_p(\underset{\sim}{t}-\underset{\sim}{s})), & \cdots & ,\delta_p f_p(\underset{\sim}{s}+\alpha_p(\underset{\sim}{t}-\underset{\sim}{s})) \end{bmatrix},$$

$$B_n(\underset{\sim}{s},\underset{\sim}{t}) = \begin{bmatrix} \delta_1 f_{n1}(\underset{\sim}{s}+\beta_1(\underset{\sim}{t}-\underset{\sim}{s})), & \cdots & ,\delta_p f_{n1}(\underset{\sim}{s}+\beta_1(\underset{\sim}{t}-\underset{\sim}{s})) \\ \vdots & & \vdots \\ \delta_1 f_{np}(\underset{\sim}{s}+\beta_p(\underset{\sim}{t}-\underset{\sim}{s})), & \cdots & ,\delta_p f_{np}(\underset{\sim}{s}+\beta_p(\underset{\sim}{t}-\underset{\sim}{s})) \end{bmatrix}.$$

and where

the f_i $(i=1,\ldots,p)$ are given in (5.2),

the f_{ni} $(i=1,\ldots,p)$ are given in (5.3), the α_i and β_i $(i=1,\ldots,p)$ satisfy $0 < \alpha_i, \beta_i < 1$ and δ_i $(i=1,\ldots,p)$ denotes $i^{\underline{\text{th}}}$ partial derivative.

(We are now writing f_{ni} instead of $f_{n,i}$.)

In particular and using $F(\underset{\sim}{0}) = \underset{\sim}{0}$ (see (4.13a)) and $F_n(\underset{\sim}{Z}_n) = \underset{\sim}{0}$ (see (5.54)), we may write, from (5.82) and (5.83), expressions such as

$$(5.84) \qquad F(\underset{\sim}{t}) = B(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}$$

and, e.g.,

$$(5.85) \qquad F_n(\underset{\sim}{t}+\underset{\sim}{Z}_n) = B_n(\underset{\sim}{Z}_n,\underset{\sim}{t}+\underset{\sim}{Z}_n)\underset{\sim}{t}$$

where (5.84) is from (5.82) with $\underset{\sim}{s} = \underset{\sim}{0}$ and (5.85) from (5.83) with $\underset{\sim}{t}$ replaced by $\underset{\sim}{t}+\underset{\sim}{Z}_n$ and $\underset{\sim}{s}$ replaced by $\underset{\sim}{Z}_n$.

Using (5.84) and (5.85), we may now write

$$(5.86) \quad h(\underset{\sim}{t}) = \left\{ \begin{array}{ll} \dfrac{1}{\|\underset{\sim}{t}\|}\|F'(\underset{\sim}{t})^{-1}B(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}\| \; , & 0 < \|\underset{\sim}{t}\| \le \dfrac{\|\underset{\sim}{\theta}^{**}\| + k}{2} \\ \\ 1 & , \text{ otherwise} \end{array} \right\}$$

and

$$(5.87) \quad h_n(\underset{\sim}{t}) = \left\{ \begin{array}{ll} \dfrac{1}{\|\underset{\sim}{t}\|}\|F_n'(\underset{\sim}{t}+\underset{\sim}{Z}_n)^{-1}B_n(\underset{\sim}{Z}_n,\underset{\sim}{t}+\underset{\sim}{Z}_n)\underset{\sim}{t}\| \; , & 0 < \|\underset{\sim}{t}\| \le \dfrac{\|\theta^{**}\| + k}{2} \\ \\ 1 & , \text{ otherwise .} \end{array} \right\}$$

To apply weak convergence theory, we show that

$$\{h(\underset{\sim}{t}):\underset{\sim}{t} \in \overline{D}\} \in C(\overline{D}) \quad \text{and} \tag{5.88}$$

$$\{h_n(\underset{\sim}{t}):\underset{\sim}{t} \in \overline{D}_1\} \in C(\overline{D}_1) \tag{5.89}$$

where $C(\overline{D})$ is defined in (5.11) and

$$\overline{D}_1 = \left\{\underset{\sim}{t} \in \mathbb{R}^p : \max_{1\le k\le p} \sum_{j=1}^{p} a_{kj}t_j \in \left[-\frac{\|\underset{\sim}{\theta}^{**}\| + k}{2}, \frac{\|\underset{\sim}{\theta}^{**}\| + k}{2} \right]\right\}. \tag{5.90}$$

<u>Proof of (5.88):</u>

(5.88) is easy for $\underset{\sim}{t} \in \overline{D}|\{\underset{\sim}{0}\}$, by continuity of F and F' and the fact that $\|F'(\underset{\sim}{t})^{-1}\|$ is bounded on $\overline{D}$ (proved earlier).
Note that "$\|F'(\underset{\sim}{t})^{-1}\|$ bounded" is the analogue of the statement "λ' is non-zero" on p. 36 of Collins (1973).

We must now check continuity of h at $\underset{\sim}{0}$. We give three proofs of this fact.

<u>1st proof of continuity of h at $\underset{\sim}{0}$:</u>

In a neighbourhood of $\underset{\sim}{t} = \underset{\sim}{0}$, write :

$$\left| \frac{\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\|}{\|\underset{\sim}{t}\|} - 1 \right| \quad \text{(note that } h(\underset{\sim}{0}) = 1\text{, by definition)}$$

$$= \left| \frac{\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\|}{\|\underset{\sim}{t}\|} - \frac{\|F'(\underset{\sim}{t})^{-1}F'(\underset{\sim}{t})\underset{\sim}{t}\|}{\|\underset{\sim}{t}\|} \right| \quad \text{clearly ;}$$

$$\le \frac{\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t}) - F'(\underset{\sim}{t})^{-1}F'(\underset{\sim}{t})\underset{\sim}{t}\|}{\|\underset{\sim}{t}\|}, \quad \text{since } \left|\|\underset{\sim}{a}\| - \|\underset{\sim}{b}\|\right| \le \|\underset{\sim}{a}-\underset{\sim}{b}\| \text{ for } \underset{\sim}{a},\underset{\sim}{b} \in \mathbb{R}^p \ ;$$

$$\le \sup\|F'(\underset{\sim}{t})^{-1}\| \frac{\|F(\underset{\sim}{t}) - F'(\underset{\sim}{t})\underset{\sim}{t}\|}{\|\underset{\sim}{t}\|},$$

$\leq M \frac{1}{\|t\|} \|F(0+t) - F(0) - F'(t)t\|$, since $F(0) = 0$;

here M is a bound for $\|F'(t)^{-1}\|$

$\xrightarrow[t\to 0]{} 0$, on, say, $\overline{D}$

by definition of $F'(0)$ (see (2.11)) .

2nd proof of continuity of h at 0 :

$$\left| \frac{\|F'(t)^{-1}F(t)\|}{\|t\|} - 1 \right\|$$

$$= \left| \frac{\|F'(t)^{-1}B(0,t)t\|}{\|t\|} - 1 \right\| , \quad \text{by (5.82)}$$

$$= \left| \frac{\|F'(t)^{-1}B(0,t)t\|}{\|t\|} - \frac{\|It\|}{\|t\|} \right| , \quad \text{where } I \text{ is the } p\times p \text{ identity matrix}$$

$$\leq \frac{\|F'(t)^{-1}B(0,t)t - It\|}{\|t\|}$$

$$\leq \|F'(t)^{-1} B(0,t) - I\| , \quad \text{by (2.2) with } E = F'(t)^{-1}B(0,t) - I$$

$\xrightarrow[t\to 0]{} \|F'(0)^{-1}F'(0) - I\|$, since from the definition of B with $s = 0$ it is trivial that $B(0,t) \xrightarrow[t\to 0]{} B(0,0) = F'(0)$; of course continuity of all partial derivatives is used.

$= 0$.

3rd proof of continuity of h at 0 :

Set

(5.91) $\quad G(t) = F'(t)^{-1}F(t)$.

We claim that

(5.92) G is differentiable at $\underset{\sim}{0}$ and $G'(\underset{\sim}{0}) = I$.

The proof of (5.92) is simplified by the fact that $F(\underset{\sim}{0}) = \underset{\sim}{0}$ (see (5.13a)) and does not require that the second derivative of F exist at $\underset{\sim}{0}$. We omit the proof of (5.92) since it is entirely similar to the proof of (2) in (10.2.1) p. 311 of Ortega and Rheinboldt (1970).

From (5.92), our third proof of continuity of h at $\underset{\sim}{0}$ follows easily as follows:

$$\left| \frac{\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\|}{\|\underset{\sim}{t}\|} - 1 \right|$$

$$= \left| \frac{\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t}) - F'(\underset{\sim}{0})^{-1}F(\underset{\sim}{0})\|}{\|\underset{\sim}{t}\|} - 1 \right|$$

since $F(\underset{\sim}{0}) = \underset{\sim}{0}$;

$$= \left| \frac{\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t}) - F'(\underset{\sim}{0})^{-1}F(\underset{\sim}{0})\| - \|I\underset{\sim}{t}\|}{\|\underset{\sim}{t}\|} \right| \text{, clearly ;}$$

$$\leq \frac{\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t}) - F'(\underset{\sim}{0})^{-1}F(\underset{\sim}{0}) - I\underset{\sim}{t}\|}{\|\underset{\sim}{t}\|}$$

$$\xrightarrow[\underset{\sim}{t}\to\underset{\sim}{0}]{} 0 \text{ , by (5.92) .}$$

This completes the proof of (5.88). □

To establish (5.89), we merely check continuity at $\underset{\sim}{0}$ (and give just one proof of this): if $E_{1,n}$ obtains, then by (5.54) there exists $\underset{\sim}{Z}_n$ such that $F_n(\underset{\sim}{Z}_n) = \underset{\sim}{0}$.

Then for $\underset{\sim}{t}$ in a neighbourhood of $\underset{\sim}{0}$, we get from (5.87)

$$(5.93)\quad |h_n(\underset{\sim}{t}) - 1|$$

$$= \left| \frac{1}{\|\underset{\sim}{t}\|} \|F_n'(\underset{\sim}{t}+\underset{\sim}{Z}_n)^{-1} B_n(\underset{\sim}{Z}_n,\underset{\sim}{t}+\underset{\sim}{Z}_n)\underset{\sim}{t}\| - 1 \right|$$

$$= \left| \frac{1}{\|\underset{\sim}{t}\|} \|F_n'(\underset{\sim}{t}+\underset{\sim}{Z}_n)^{-1} B_n(\underset{\sim}{Z}_n,\underset{\sim}{t}+\underset{\sim}{Z}_n)\underset{\sim}{t}\| - \frac{1}{\|\underset{\sim}{t}\|} \|I\underset{\sim}{t}\| \right|$$

$$\leq \frac{1}{\|\underset{\sim}{t}\|} \|F_n'(\underset{\sim}{t}+\underset{\sim}{Z}_n)^{-1} B_n(\underset{\sim}{Z}_n,\underset{\sim}{t}+\underset{\sim}{Z}_n)\underset{\sim}{t} - I\underset{\sim}{t}\|$$

$$\leq \|F_n'(\underset{\sim}{t}+\underset{\sim}{Z}_n)^{-1} B_n(\underset{\sim}{Z}_n,\underset{\sim}{t}+\underset{\sim}{Z}_n) - I\| .$$

Now it is almost trivial from the expression for B_n, that

$$B_n(\underset{\sim}{Z}_n,\underset{\sim}{t}+\underset{\sim}{Z}_n) = \begin{pmatrix} \delta_1 f_{n1}(\underset{\sim}{Z}_n+\beta_1\underset{\sim}{t}), & \cdots & ,\delta_p f_{n1}(\underset{\sim}{Z}_n+\beta_1\underset{\sim}{t}) \\ \vdots & & \vdots \\ \delta_1 f_{np}(\underset{\sim}{Z}_n+\beta_p\underset{\sim}{t}), & \cdots & ,\delta_p f_{np}(\underset{\sim}{Z}_n+\beta_p\underset{\sim}{t}) \end{pmatrix}$$

$$\xrightarrow[\underset{\sim}{t}\to\underset{\sim}{0}]{} \begin{pmatrix} \delta_1 f_{n1}(\underset{\sim}{Z}_n), & \cdots & ,\delta_p f_{n1}(\underset{\sim}{Z}_n) \\ \vdots & & \vdots \\ \delta_1 f_{np}(\underset{\sim}{Z}_n), & \cdots & ,\delta_p f_{np}(\underset{\sim}{Z}_n) \end{pmatrix} = F_n'(\underset{\sim}{Z}_n).$$

From this and (5.93) we see that

$$|h_n(\underset{\sim}{t}) - 1| \leq \lim_{\underset{\sim}{t}\to\underset{\sim}{0}} \|F_n'(\underset{\sim}{Z}_n)^{-1} F_n'(\underset{\sim}{Z}_n) - I\| = 0 ,$$

proving continuity at $\underset{\sim}{0}$ of h_n .

We remark that the type of proof used as our first proof of continuity of F at $\underset{\sim}{0}$ could have been used here and would have been easier.

We now give the analogues of some of the equations etc. in Lemma 2.6 of Collins (1973) that we need in proving tightness of the sequence

$\{h_n(\underset{\sim}{t}):\underset{\sim}{t} \in \overline{D}_1 (\overline{D}_1$ is defined in 5.90)$\}$.

(Note that convergence of the finite dimensional distributions of the h_n to those of h is proved in exactly the same way as [a], p. 38 of Collins (1973), using (5.13), (5.30) and (5.76) of the present work to arrive at the analogues of the expressions on pages 38 and 39 of Collins (1973).)

(Note also that some of the analogues below are stronger statements than the corresponding ones in Collins (1973) . This is because we need stronger statements to combat the complications arising from the non-existence of a direct analogue of the mean value theorem for mappings $f:\mathbb{R}' \to \mathbb{R}'$ to mappings $F:\mathbb{R}^p \to \mathbb{R}^p$ $(p > 1)$.)

(5.94) $\{2.74\} \leftrightarrow P\{\sup\{|h_n(\underset{\sim}{t})-h_n(\underset{\sim}{s})|:\|\underset{\sim}{t}-\underset{\sim}{s}\| < \delta\} < 2\varepsilon\} > 1 - 4\eta$;

(5.95) $\{2.75\} \leftrightarrow \sup\left|\|F_n'(\underset{\sim}{t}+\underset{\sim}{Z}_n)^{-1}F_n(\underset{\sim}{t}+\underset{\sim}{Z}_n)\|-\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\|\right| \xrightarrow{P} 0$;

(5.96) $\{2.85\} \leftrightarrow \sup_{\underset{\sim}{t},\underset{\sim}{s}}\|F_n'(\underset{\sim}{s})^{-1}B_n^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})-F'(\underset{\sim}{s})^{-1}B^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\| \xrightarrow{P} 0$,

for all $\underset{\sim}{\alpha} = (\alpha_1,\ldots,\alpha_p)$ such that

$0 < \alpha_i < 1 \quad (i=1,\ldots,p)$

where B_n and B are defined below (5.83). We put the superscript $\underset{\sim}{\alpha}$ on the B_n and B in (5.96) to indicate that we insist the same vector of constants appears in the expression for B_n and B in (5.96).
Note that (5.96) is a stronger statement than

(5.97) $\sup_{\underset{\sim}{t},\underset{\sim}{s}}\|F_n'(\underset{\sim}{s})^{-1}F_n'(\underset{\sim}{t})-F'(\underset{\sim}{s})^{-1}F'(\underset{\sim}{t})\| \xrightarrow{P} 0$

because (5.97) follows from (5.96) by continuity - let $\underset{\sim}{\alpha} \to \underset{\sim}{1}$

(where $\underset{p\times 1}{\underset{\sim}{1}} = (1,\ldots,1)^T$).

Of course, (5.97) can be established independently of (5.96). We will indicate the proof of (5.96) in a moment.

The key inequality we are getting at is the analogue of {2.89}:

$$\text{(5.98)} \quad \{2.89\} \leftrightarrow \sup\{|h_n(\underset{\sim}{t})-h_n(\underset{\sim}{s})| : \|\underset{\sim}{t}\|,\|\underset{\sim}{s}\| \le \delta_1\}$$

$$= \sup \left|\frac{1}{\|\underset{\sim}{t}\|}\|F_n'(\underset{\sim}{t})^{-1}B_n^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}\| - \frac{1}{\|\underset{\sim}{s}\|}\|F_n'(\underset{\sim}{s})^{-1}B_n^{\underset{\sim}{\beta}}(\underset{\sim}{0},\underset{\sim}{s})\underset{\sim}{s}\|\right|$$

$$\le 2 \sup\left|\frac{1}{\|\underset{\sim}{t}\|}\|F'(\underset{\sim}{t})^{-1}B_n^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}\| - \frac{1}{\|\underset{\sim}{t}\|}\|F'(\underset{\sim}{t})^{-1}B^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}\|\right|$$

$$+ \sup\left|\frac{1}{\|\underset{\sim}{t}\|}\|F'(\underset{\sim}{t})^{-1}B^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}\| - \frac{1}{\|\underset{\sim}{s}\|}\|F'(\underset{\sim}{s})^{-1}B^{\underset{\sim}{\beta}}(\underset{\sim}{0},\underset{\sim}{s})\underset{\sim}{s}\|\right| .$$

We extend this inequality further to

$$\text{(5.99)} \quad \le 2 \sup \left\{\frac{1}{\|\underset{\sim}{t}\|}\|F_n'(\underset{\sim}{t})^{-1}B_n^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}-F'(\underset{\sim}{t})^{-1}B^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}\|\right\}$$

$$+ \sup\left|\frac{1}{\|\underset{\sim}{t}\|}\|F'(\underset{\sim}{t})^{-1}B^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}\| - \frac{1}{\|\underset{\sim}{s}\|}\|F'(\underset{\sim}{s})^{-1}B^{\underset{\sim}{\beta}}(\underset{\sim}{0},\underset{\sim}{s})\underset{\sim}{s}\|\right|$$

and still further to

$$\text{(5.100)} \quad \le 2 \sup\|F_n'(\underset{\sim}{t})^{-1}B_n^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})-F'(\underset{\sim}{t})^{-1}B^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\|$$

$$+ \sup\left|\frac{1}{\|\underset{\sim}{t}\|}\|F'(\underset{\sim}{t})^{-1}B^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}\| - \frac{1}{\|\underset{\sim}{s}\|}\|F'(\underset{\sim}{s})^{-1}B^{\underset{\sim}{\beta}}(\underset{\sim}{0},\underset{\sim}{s})\underset{\sim}{s}\|\right| .$$

The first term after the inequality sign can be made less than, say, $2\,\frac{\varepsilon}{3}$ by (5.96) and the second less than $\frac{\varepsilon}{3}$ by uniform continuity of the function

$$\frac{1}{\|\underset{\sim}{t}\|}\|F'(\underset{\sim}{t})^{-1}B^{\underset{\sim}{\gamma}}(\underset{\sim}{0},\underset{\sim}{t})\underset{\sim}{t}\| .$$

This establishes the analogue of {2.90}:

(5.101) $\quad \{2.90\} \leftrightarrow \{\sup\{|h_n(\underset{\sim}{t})-h_n(\underset{\sim}{s})|:\|\underset{\sim}{t}\|,\|\underset{\sim}{s}\| \le \delta_1\} < \varepsilon\} > 1 - 2\eta$.

The rest of the proof of the lemma is identical with that of Lemma 2.6 in Collins (1973). For example, the analogue of {2.96} is

$$\{2.96\} \leftrightarrow |h_n(\underset{\sim}{t})-h_n(\underset{\sim}{s})|$$

$$= \left|\frac{1}{\|\underset{\sim}{t}\|}\|F_n'(\underset{\sim}{t}+\underset{\sim n}{Z})^{-1}F_n(\underset{\sim}{t}+\underset{\sim n}{Z})\| - \frac{1}{\|\underset{\sim}{s}\|}\|F_n'(\underset{\sim}{s}+\underset{\sim n}{Z})^{-1}F_n(\underset{\sim}{s}+\underset{\sim n}{Z})\|\right|$$

$$\le \ldots\ldots\ldots$$

$$< \varepsilon \ .$$

This inequality and the analogues of {2.94} and {2.95} lead to the analogue of {2.97}:

(5.102) $\{2.97\} \leftrightarrow P\{\sup\{|h_n(\underset{\sim}{t})-h_n(\underset{\sim}{s})|:\|\underset{\sim}{t}\|,\|\underset{\sim}{s}\| \ge \delta_1,\|\underset{\sim}{t}-\underset{\sim}{s}\| < \delta\} < \varepsilon\} > 1 - 2\eta$

and, exactly as on p. 46 of Collins (1973), we get from (5.101) and (5.102)

(5.103) $\{2.101\} \leftrightarrow P\{\sup\{|h_n(\underset{\sim}{t})-h_n(\underset{\sim}{s})|:\|\underset{\sim}{t}-\underset{\sim}{s}\| < \delta\} < 2\varepsilon\} > 1 - 4\eta$,

establishing (5.94) .

To complete our discussion of this lemma we sketch a proof of (5.96). We have, from (5.61),

(5.104) $\quad \sup\{\|F_n'(\underset{\sim}{t})^{-1} - F'(\underset{\sim}{t})^{-1}\|:\underset{\sim}{t} \in \overline{D}\} \xrightarrow{P} 0$.

We also have, for any $\underset{\sim}{\alpha} = (\alpha_1,\ldots,\alpha_p)^T$, $0 < \alpha_i < 1$ $(i=1,\ldots,p)$

(5.105) $\quad \sup\{\|B_n^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t}) - B^{\underset{\sim}{\alpha}}(\underset{\sim}{0},\underset{\sim}{t})\| \xrightarrow{P} 0$.

The proof of (5.105) is exactly the same as that of (5.30). For example, the analogue of (5.45) is

$$\sup\{|\delta_k\psi(X_i-\alpha_k\sum_{j=1}^{p}c_{ij}a_j) - \delta_k\psi(X_i-\alpha_k\sum_{j=1}^{p}c_{ij}b_j)| : \|\underset{\sim}{a}-\underset{\sim}{b}\| < \delta\}$$

$$< \frac{\varepsilon}{p \max_{i,r}|c_{ir}|} .$$

From (5.104) and (5.105), we arrive at (5.96) by an argument similar to that used on pages 41 and 42 of Collins (1973).

This completes the proof of the lemma. □

(5.106) Given the vector $\underset{\sim}{x} = (x_1,\ldots,x_p)^T$, we shall denote the l^{th} component of the vector $A\underset{\sim}{x}$ by $[A\underset{\sim}{x}]_l$. Thus

$$[A\underset{\sim}{x}]_l = \sum_{j=1}^{p} a_{lj}x_j, \quad l=1,\ldots,p .$$

(5.107) LEMMA:

Let

(5.108) $$E^*_{1n} = \{w: [A(t-Z_n(w))]_l\,[AF'_n(t,w)^{-1}F_n(t,w)]_l \geq 0 \text{ for all } \underset{\sim}{t} \in \overline{D} , \; l=1,\ldots,p\}.$$

Then

(5.109) $$P(E^*_{1n}) \longrightarrow 1 \quad (n \to \infty) .$$

We note that

$$[A(\underset{\sim}{t}-\underset{\sim}{Z}_n)]_l[AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})]_l \geq 0 \ , \quad l=1,\ldots,p$$

says that each component of the vector $AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})$ lies on the same side of the origin as the corresponding component of $A(\underset{\sim}{t}-\underset{\sim}{Z}_n)$.

Proof of (5.107):

We must show that

$$P\{\underline{\inf}_{\underset{\sim}{t}\in D}[A(\underset{\sim}{t}-\underset{\sim}{Z}_n)]_l[AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})]_l \geq 0 \ , \quad l=1,\ldots,p\} \xrightarrow[(n\to\infty)]{} 1$$

or, equivalently, we must show that

$$(5.110) \quad P\{\underline{\sup}_{\underset{\sim}{t}\in D}[A(\underset{\sim}{Z}_n-\underset{\sim}{t})]_l[AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})]_l \leq 0 \ , \quad l=1,\ldots,p\} \xrightarrow{(n\to\infty)} 1 \ ,$$

since $\inf x_n = -\sup(-x_n)$

and, by linearity of A, $A(-\underset{\sim}{x}) = -A(\underset{\sim}{x})$.

Now, for all l, $1 \leq l \leq p$,

$$\sup[A(\underset{\sim}{Z}_n-\underset{\sim}{t})]_l[AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})]_l$$

$$= \sup\Big\{[A(-\underset{\sim}{t})]_l[AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})-AF'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})]_l$$

$$+ [A(-\underset{\sim}{t})]_l[AF'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})]_l + [A(\underset{\sim}{Z}_n)]_l[AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})]_l\Big\}$$

$$\leq \sup\Big\{\Big|[A(-\underset{\sim}{t})]_l\Big| \ \Big|[AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})-AF_n'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})]_l\Big|$$

$$+ [A(-\underset{\sim}{t})]_l[AF'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})]_l + \Big|[A(\underset{\sim}{Z}_n)]_l\Big| \ \Big|[AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})]_l\Big|\Big\} \ ,$$

and now, taking the sup inside and recalling from the definition of $\|\cdot\|_A$, $\|\underset{\sim}{x}\|_A = \max_{1\leq l\leq p} |[A(\underset{\sim}{x})]_l|$, we get that the last expression above is

$$(5.111)\quad \le \sup\|\underset{\sim}{t}\|_A \ \sup\|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})-F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\|_A$$

$$+ \sup[A(-\underset{\sim}{t})]_l[AF'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})]_l$$

$$+ \|\underset{\sim}{Z}_n\|_A \ \sup\|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})\|_A \ .$$

The first term in (5.111) approaches zero in probability, since $\sup\|\underset{\sim}{t}\|_A < \infty$ on $\overline{D}$ and $\sup\|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})-F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\| \xrightarrow{P} 0$ (with respect to any norm, of course) by (5.68).

Now also the third term in (5.111) converges to zero in probability because

$$\|\underset{\sim}{Z}_n\| \xrightarrow{P} 0 \ , \quad \text{by (5.76)}$$

and $\sup\|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})\| \le \sup\|F_n'(\underset{\sim}{t})\|^{-1} \ \sup\|F_n(\underset{\sim}{t})\|$

and the last two terms are bounded in probability by previous work.

Finally, the middle term in (5.111) is

$$\sup_{\underset{\sim}{t}\in D}[A(-\underset{\sim}{t})]_l[AF'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})]_l$$

$$= -\inf_{\underset{\sim}{t}}[A(\underset{\sim}{t})]_l[AF'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})]_l$$

$$= -\inf_{\underset{\sim}{t}}\left(\sum_{j=1}^{p} a_{lj}t_j\right)\left|AA^{-1}\left(-\frac{E\psi_l}{E\psi_l'}\right)\right| \qquad \text{(see the note preceding (5.13) and see also (4.24))}$$

$$= -\inf_{\underset{\sim}{t}}\left(\sum_{j=1}^{p} a_{lj}t_j\left(\frac{-E\psi_l}{E\psi_l'}\right)\right)$$

$$= -\inf_{\underset{\sim}{t}}\left|\sum_{j=1}^{p} a_{lj}t_j\right|\left|\frac{E\psi_l}{E\psi_l'}\right| \ , \quad \text{(by (4.29))}$$

$$\le 0 \ .$$

Thus we have that

$$\sup[A(\underset{\sim}{Z}_n-\underset{\sim}{t})]\ [AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})]_l \xrightarrow{P} \text{ non-positive quantity.}$$

This proves (5.110) and the lemma. □

(5.112) REMARK:

We will be omitting $\underset{\sim}{t} = \underset{\sim}{0}$ from our domain in the following final lemma and so we note from our proof above that we have strict inequality (> 0) appearing in the event E_{1n}^{*} (see (5.108)) if $\overline{D}$ is replaced by $\overline{D} | \{\underset{\sim}{0}\}$, because $\left(\sum_{j=1}^{p} a_{lj}t_j\right)\left(\frac{E\psi_l}{E\psi_l'}\right) = 0$ if and only if $\sum_{j=1}^{p} a_{lj}t_j = 0$ (see (4.29)) so that $[A(\underset{\sim}{t}-\underset{\sim}{Z}_n)]_l[AF_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})]_l = 0$ for all l if and only if $\sum_{j=1}^{p} a_{lj}t_j = 0$ for all l, i.e., if and only if $A\underset{\sim}{t} = \underset{\sim}{0}$, i.e., if and only if $\underset{\sim}{t} = \underset{\sim}{0}$, since A is invertible.

In the following lemma, the norm (4.30), without the subscript A, is understood.

(5.113) LEMMA:

The events

E_{1n} (defined in (5.47)) ,

E_{1n}^{*} (defined in (5.107)) ,

E_{2n} (defined in (5.75)) ,

(5.114) $$E_{3n}: \|\underset{\sim}{Z}_n\| < \min\left[\frac{k - \|\underset{\sim}{\theta}^{**}\|}{4},\ \frac{2-\gamma}{\gamma}\|\underset{\sim}{\theta}^{**}\|\right]$$

where $1 < \gamma < 2$

and

(5.115) $$E_{4n}: \sup\left\{\|F_n'(\underset{\sim}{t}+\underset{\sim}{Z}_n)^{-1}F_n(\underset{\sim}{t}+\underset{\sim}{Z}_n)\|\frac{1}{\|\underset{\sim}{t}\|} : \|\underset{\sim}{t}\| \in \left[0, \frac{k+\|\underset{\sim}{\theta}^{**}\|}{2}\right]\right\} < \gamma$$

jointly imply

(5.116) $\underset{\sim}{T}_n = \underset{\sim}{Z}_n$, where $\underset{\sim}{T}_n$ is given in (3.25).

Note that it makes sense to discuss $\underset{\sim}{Z}_n$ above, by the comments in (5.53). Note also that

$P(E_{1n}) \longrightarrow 1$ by (5.47)

$P(E_{1n}^*) \longrightarrow 1$ by (5.107)

$P(E_{2n}) \longrightarrow 1$ by (5.75)

$P(E_{3n}) \longrightarrow 1$ by (5.76)

and $P(E_{4n}) \longrightarrow 1$ as a consequence of (5.79) and the fact that

$$\|F'(\underset{\sim}{t})^{-1}F(\underset{\sim}{t})\| < 2\|\underset{\sim}{t}\| \ , \quad \underset{\sim}{t} \in \overline{D} \ .$$

(Note that this last inequality follows trivially (triangle inequality) from (4.21).)

Thus (5.113) may be re-stated as

(5.117) $$P\{\underset{\sim}{T}_n = \underset{\sim}{Z}_n\} \longrightarrow 1 \ (n\to\infty).$$

<u>Proof of (5.113):</u>

If $\|\underset{\sim}{t}\| < \|\underset{\sim}{\theta}^{**}\| + \dfrac{k-\|\underset{\sim}{\theta}^{**}\|}{4}$ then, since

$$\|\underset{\sim}{Z}_n\| < \frac{k-\|\underset{\sim}{\theta}^{**}\|}{4} \ , \quad \text{by (5.114),}$$

we have

$$\|t-Z_n\| \le \|t\| + \|Z_n\| < \|\theta^{**}\| + \frac{k-\|\theta^{**}\|}{2} = \frac{k+\|\theta^{**}\|}{2} \quad .$$

Thus (5.115) implies

$$(5.118) \qquad \|F_n'(t)^{-1}F_n(t)\| < \gamma\|t-Z_n\|, \ \|t\| < \|\theta^{**}\| + \frac{k-\|\theta^{**}\|}{4}$$

$$\text{and} \quad \|t-Z_n\| \ne 0.$$

We now claim that

$$(5.119) \qquad \|t-Z_n - F_n'(t)^{-1}F_n(t)\| < \|t-Z_n\| \ ,$$

$$\|t-Z_n\| \in \left(0 \ , \ \frac{k + \|\theta^{**}\|}{2}\right] \ .$$

Using the notation introduced in (5.106) and recalling that our norm is the norm (5.30), we have that (5.118) reads

$$\max_{1\le l\le p} [AF_n'(t)^{-1}F_n(t)]_l < \gamma \max_{1\le l\le p} [A(t-Z_n)]_l$$

which, since $\gamma < 2$, implies

$$(5.120) \qquad \max_{1\le l\le p} [AF_n'(t)^{-1}F_n(t)]_l < 2 \max_{1\le l\le p} [A(t-Z_n)]_l \ .$$

Then (note the analogy between the following lines and the few lines following (4.30))

$$\|t-Z_n - F_n'(t)^{-1}F_n(t)\|$$

$$= \max_{1\le l\le p} \left|[A(t-Z_n)]_l - [AF_n'(t)^{-1}F_n(t)]_l\right| \ ,$$ (by definition of our norm in (5.30) and linearity of A) ;

$$< \max_{1 \le l \le p} [A(\underset{\sim}{t} - \underset{\sim}{Z}_n)]_l \ , \quad (\text{by (5.120) and } E^*_{1n})$$

$$= \|\underset{\sim}{t} - \underset{\sim}{Z}_n\| \ , \quad \text{proving (5.119)} \ .$$

We write (5.119) as

$$\text{(5.121)} \qquad \|\underset{\sim}{t} - \underset{\sim}{Z}_n - F'_n(\underset{\sim}{t})^{-1} F_n(\underset{\sim}{t})\| \le \alpha \|\underset{\sim}{t} - \underset{\sim}{Z}_n\|$$

$$\text{for} \quad \|\underset{\sim}{t}\| < \|\underset{\sim}{\theta}^{**}\| + \frac{k - \|\underset{\sim}{\theta}^{**}\|}{4} \ , \quad \text{some } \alpha < 1 \ ,$$

so that the first iterate

$$\underset{\sim}{t}^1 = \underset{\sim}{\theta}^*_n - F'_n(\underset{\sim}{\theta}^*_n)^{-1} F_n(\underset{\sim}{\theta}^*_n) \quad \text{(see (3.25)) satisfies}$$

$$\|\underset{\sim}{t}^1 - \underset{\sim}{Z}_n\| \le \alpha \|\underset{\sim}{\theta}^*_n - \underset{\sim}{Z}_n\| \ ,$$

since

$$\underset{\sim}{t}^0 = \underset{\sim}{\theta}^*_n \quad \text{satisfies} \quad \|\underset{\sim}{\theta}^*_n\| < \|\underset{\sim}{\theta}^{**}\| + \frac{k - \|\underset{\sim}{\theta}^{**}\|}{4}$$

by definition of $E_{2,n}$.

To show that this and the remaining iterates remain in

$$\{\underset{\sim}{t} : \|\underset{\sim}{t}\| < \|\underset{\sim}{\theta}^{**}\| + \frac{k - \|\underset{\sim}{\theta}^{**}\|}{4}\} \ , \quad \text{observe that, by (5.118),}$$

$$\|F'_n(\underset{\sim}{t})^{-1} F_n(\underset{\sim}{t})\| < \gamma \|\underset{\sim}{t} - \underset{\sim}{Z}_n\|$$

so that $\|F'_n(\underset{\sim}{t})^{-1} F_n(\underset{\sim}{t})\| < \gamma \left(\|\underset{\sim}{\theta}^{**}\| + \frac{k - \|\underset{\sim}{\theta}^{**}\|}{4} + \frac{2 - \gamma}{\gamma} \|\underset{\sim}{\theta}^{**}\| \right)$

(by definition of $E_{2,n}$ and (5.114))

$$= \gamma \|\underset{\sim}{\theta}^{**}\| + \frac{\gamma k - \gamma \|\underset{\sim}{\theta}^{**}\|}{4} + 2\|\underset{\sim}{\theta}^{**}\| - \gamma \|\underset{\sim}{\theta}^{**}\|$$

$$= 2\left[\|\underset{\sim}{\theta}^{**}\| + \frac{\gamma(k - \|\underset{\sim}{\theta}^{**}\|)}{8}\right]$$

$$< 2\left[\|\underset{\sim}{\theta}^{**}\| + \frac{k - \|\underset{\sim}{\theta}^{**}\|}{4}\right] \quad \text{(since } \gamma < 2\text{)} \ , \quad \text{i.e.,}$$

(5.122) $$\|F_n'(\underset{\sim}{t})^{-1}F_n(\underset{\sim}{t})\| < 2\left[\|\underset{\sim}{\theta}^{**}\| + \frac{k - \|\underset{\sim}{\theta}^{**}\|}{4}\right] .$$

From (5.122) and the fact that

$$\underset{\sim}{t}^0 = \underset{\sim}{\theta}^* \quad \text{satisfies} \quad \|\underset{\sim}{\theta}^*\| < \|\underset{\sim}{\theta}^{**}\| + \frac{k - \|\underset{\sim}{\theta}^{**}\|}{4}$$

one sees by the same argument as used in deriving (5.119) from (5.120) that

$$\|\underset{\sim}{t}^1\| = \|\underset{\sim}{t}^0 - F_n'(\underset{\sim}{t}^0)^{-1}F_n(\underset{\sim}{t}^0)\| < \|\underset{\sim}{\theta}^{**}\| + \frac{k - \|\underset{\sim}{\theta}^{**}\|}{4}$$

so that the first iterate remains in the set

$$\left\{\underset{\sim}{t} : \|\underset{\sim}{t}\| < \|\underset{\sim}{\theta}^{**}\| + \frac{k - \|\underset{\sim}{\theta}^{**}\|}{4}\right\} .$$

By induction, all iterates remain in this set and satisfy

$$\|\underset{\sim}{t}^{j+1} - \underset{\sim}{Z}_n\| \le \alpha^{j+1}\|\underset{\sim}{\theta}^0 - \underset{\sim}{Z}_n\| \ , \quad j = 0,1,2,\ldots .$$

Then,

$\underset{\sim}{Z}_n = \lim_{j\to\infty} \underset{\sim}{t}^j = \underset{\sim}{T}_n$ and this completes the proof of the lemma. □

<u>(5.123) THEOREM:</u> (Consistency of $\underset{\sim}{T}_n$) .

Consider the model (3.5) and let $\{\underset{\sim}{T}_n\}$ be given by (3.25). Then

(5.124) $$\underset{\sim}{T}_n \xrightarrow{P} \underset{\sim}{0} .$$

Proof:

Let $\varepsilon > 0$ and write

$$P\{\|\underset{\sim}{T}_n\| \le \varepsilon\}$$

$$= P\{[\underset{\sim}{T}_n = \underset{\sim}{\theta}^*] \cap [\|\underset{\sim}{T}_n\| \le \varepsilon]\} + P\{[\underset{\sim}{T}_n = \underset{\sim}{Z}_n] \cap [\|\underset{\sim}{T}_n\| \le \varepsilon]\}$$

$$\ge P\{[\underset{\sim}{T}_n = \underset{\sim}{Z}_n] \cap [\|\underset{\sim}{T}_n\| \le \varepsilon]\}$$

$$= P\{[\underset{\sim}{T}_n = \underset{\sim}{Z}_n] \cap [\|\underset{\sim}{Z}_n\| \le \varepsilon]\}$$

$$\longrightarrow 1 \quad (n \to \infty)$$

by (5.113) (see (5.117)) and (5.76).

This completes the proof. □

(5.124) THEOREM: (Asymptotic Normality of $\underset{\sim}{T}_n$)

For the model (3.5) we have

$$(5.125) \qquad n^{\frac{1}{2}} \underset{\sim}{T}_n \xrightarrow{\mathcal{D}} MVN \left(0,\ C_0^{-1} \frac{\int \psi^2 dG}{(\int \psi' dG)^2}\right)$$

where the sequence $\{\underset{\sim}{T}_n\}$ is given by (3.25) and where $C_0 = \lim\limits_{n\to\infty} \frac{C^T C}{n}$.

(Note that, by this result, the efficiencies (of the components of $\underset{\sim}{T}_n$) are clearly independent of the design matrix and, further, the limiting covariance matrix is independent of $G \in F$, since, clearly,

$$(5.126) \qquad \frac{\int \psi^2 dG}{(\int \psi' dG)^2} = \frac{\int_{-c}^{c} \psi^2(y)\phi(y)dy}{(\int_{-c}^{c} \psi(x)\phi'(x)dx)^2} \Bigg) .$$

Before proving (5.124), we give the following lemma:

(5.127) LEMMA:

$$C_0 = \lim_{n\to\infty} \frac{C^T C}{n} \quad \text{exists and is positive definite}$$

and, further,

$$\lim_{n\to\infty} \max_{i,j} \frac{|c_{ij}|}{n^{\frac{1}{2}}} = 0 .$$

Proof of (5.127):

We have (see the conditions given after (3.5))

$$C = \begin{bmatrix} c_{11}\ c_{12}, & \cdots & ,c_{1p} \\ c_{21}\ c_{22}, & \cdots & ,c_{2p} \\ \vdots & & \vdots \\ c_{n1}\ c_{n2}, & \cdots & ,c_{np} \end{bmatrix} = \begin{array}{l} \left.\begin{bmatrix} a_{11}, & \cdots & ,a_{1p} \\ a_{11}, & \cdots & ,a_{p} \\ a_{21}, & \cdots & ,a_{2p} \\ a_{21}, & \cdots & ,a_{2p} \\ \vdots & & \vdots \\ a_{p1}, & \cdots & ,a_{pp} \\ a_{p1}, & \cdots & ,a_{pp} \end{bmatrix}\right. \end{array} \begin{array}{l} \Big\} \longrightarrow n_{q_1}(n) \text{ times} \\ \Big\} \longrightarrow n_{q_2}(n) \text{ times} \\ \\ \Big\} \longrightarrow n_{q_p}(n) \text{ times.} \end{array}$$

Then

$$C^T C = \begin{bmatrix} n\sum_{k=1}^{p} q_k a_{k1}^2, & \cdots & ,n\sum_{k=1}^{p} q_k a_{k1} a_{kp} \\ \vdots & & \vdots \\ n\sum_{k=1}^{p} a_{k1} a_{kp}, & \cdots & ,n\sum_{k=1}^{p} q_k a_{kp}^2 \end{bmatrix}$$

$$= \left(\left(n\sum_{k=1}^{p} q_k a_{kj} a_{ki}\right)\right) = n\ A^T\ \mathrm{Diag}((q_i))A .$$

Put

$$(5.128) \qquad C_0 = A^T \operatorname{Diag}((q_i))A$$

so that

$$\lim_{n\to\infty} \frac{C^T C}{n} = C_0$$

and note that C_0 is positive definite since

$$C_0 = (\operatorname{Diag}((q_i^{\frac{1}{2}}))A)^T(\operatorname{Diag}((q_i^{\frac{1}{2}}))A)$$

and the matrix $\operatorname{Diag}((q_i^{\frac{1}{2}}))A$ is non-singular, by assumption. (Recall that if E is any matrix, then $E^T E$ is always positive semi-definite and is positive definite if and only if E is non-singular.)

The last assertion of (5.127) is trivial since

$$\lim_{n\to\infty} \max_{i,j} \frac{|c_{ij}|}{n^{\frac{1}{2}}} = \max_{k,l} |a_{kl}| \lim_{n\to\infty} n^{-\frac{1}{2}} = 0 .$$

This completes the proof of (5.127). □

<u>Proof of (5.124)</u>:

Write

$$(5.129) \quad H_n(\underset{\sim}{t}) = nF_n(\underset{\sim}{t}) = \begin{bmatrix} \sum_{i=1}^{n} c_{i1}\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right) \\ \vdots \\ \sum_{i=1}^{n} c_{ip}\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right) \end{bmatrix} = \begin{bmatrix} h_{n1}(\underset{\sim}{t}) \\ \vdots \\ h_{np}(\underset{\sim}{t}) \end{bmatrix} , \text{ say },$$

and note that $H_n(\underset{\sim}{t})$ is the sum of the n independent but not identically distributed random vectors $\left(c_{i1}\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right), \ldots, c_{ip}\psi\left(X_i - \sum_{j=1}^{p} c_{ij}t_j\right)\right)^T$.

We may write (see (2.17))

$$H_n(\underset{\sim}{T}_n) = H_n(\underset{\sim}{0}) + B_n(\underset{\sim}{0},\underset{\sim}{T}_n)\underset{\sim}{T}_n \tag{5.130}$$

where $B_n(\underset{\sim}{0},\underset{\sim}{T}_n)$ denotes the matrix

$$B_n(\underset{\sim}{0},\underset{\sim}{T}_n) = \begin{bmatrix} \delta_1 h_{n1}(\beta_1\underset{\sim}{T}_n), & \cdots & ,\delta_p h_{n1}(\beta_1\underset{\sim}{T}_n) \\ \vdots & & \vdots \\ \delta_1 h_{np}(\beta_p\underset{\sim}{T}_n), & \cdots & ,\delta_p h_{np}(\beta_p\underset{\sim}{T}_n) \end{bmatrix}, \tag{5.131}$$

for some β_i , $0 < \beta_i < 1$ $(i=1,\ldots,p)$.

We write (5.129) in the form

$$n^{\frac{1}{2}}\underset{\sim}{T}_n = -\left(\frac{1}{n} B_n(\underset{\sim}{0},\underset{\sim}{T}_n)\right)^{-1}[n^{-\frac{1}{2}}H_n(\underset{\sim}{0}) - n^{-\frac{1}{2}}H_n(\underset{\sim}{T}_n)] . \tag{5.132}$$

Note that the proof of the invertibility of $\frac{1}{n} B_n(\underset{\sim}{0},\underset{\sim}{T}_n)$ (in the sense that it converges weakly to an invertible matrix, namely $F'(\underset{\sim}{0})$ is entirely similar to the proof of the invertibility of $F_n'(\underset{\sim}{t})$, so we omit it (see the work following (5.6)).

Now

$$n^{-\frac{1}{2}}H_n(\underset{\sim}{T}_n) \xrightarrow{P} \underset{\sim}{0} \tag{5.133}$$

since, for any $\varepsilon > 0$,

$$P\{n^{-\frac{1}{2}}\|H_n(\underset{\sim}{T}_n)\| < \varepsilon\} \geq P\{\underset{\sim}{T}_n = \underset{\sim}{Z}_n\}$$

[because $\underset{\sim}{T}_n = \underset{\sim}{Z}_n \Rightarrow n^{-\frac{1}{2}}\|H_n(\underset{\sim}{T}_n)\| < \varepsilon$

owing to $H_n(\underset{\sim}{Z}_n) = F_n(\underset{\sim}{Z}_n) = 0$ (see (5.54))

and $P\{\underset{\sim}{T}_n = \underset{\sim}{Z}_n\} \longrightarrow 1$ $(n \to \infty)$ by (5.113) (see (5.117)] .

Next, we consider the term $n^{-\frac{1}{2}}H_n(\underset{\sim}{0})$ in (5.132).

For each r,k where $r,k = 1,\ldots,p$, the covariance of

$$\sum_{i=1}^{n} c_{ir}\ \psi(X_i) \text{ and } \sum_{i=1}^{n} c_{ik}\ \psi(X_i) \quad \text{is}$$

$$E\left[\left(\sum_{i=1}^{n} c_{ir}\ \psi(X_i)\right)\left(\sum_{i=1}^{n} c_{ik}\ \psi(X_i)\right)\right] - \left[E\sum_{i=1}^{n} c_{ir}\ \psi(X_i)\right]\left[E\sum_{i=1}^{n} c_{ik}\ \psi(X_i)\right]$$

$$= E\left[\left(\sum_{i=1}^{n} c_{ir}\ \psi(X_i)\right)\left(\sum_{i=1}^{n} c_{ik}\ \psi(X_i)\right)\right] - 0\ , \quad \text{since } E\psi(X) = 0 \text{ (see the proof of (5.13a))}$$

$$= \sum_{i=1}^{n} c_{ir}c_{ik}\ E\psi^2 + \sum_{i\neq j}\sum c_{ir}c_{jk}(0)\ , \quad \text{again using } E\psi(X) = 0\ .$$

(5.134) Thus the covariance matrix of

$H_n(\underset{\sim}{0})$ is $C^T CE\psi^2$.

We now claim:

(5.135) Lindeberg's condition applies to each of the p components

$$n^{-\frac{1}{2}}\sum_{i=1}^{n} c_{ik}\ \psi(X_i) \quad \text{of} \quad n^{-\frac{1}{2}}H_n(\underset{\sim}{0})\ .$$

Proof:

Let $\eta > 0$ and let α be a bound for ψ. Then, by (5.126), as soon as n is large enough we will have

$$n^{-\frac{1}{2}}|c_{ik}\ \psi(X_i)| \leq n^{-\frac{1}{2}}|c_{ik}|\alpha \leq \eta$$

so that, with $\xi_{ikn} = n^{-\frac{1}{2}}c_{ik}\ \psi(X_i)$, we have $\displaystyle\int_{|\xi_{ik}| > \eta} y^2\ d\,G_{\xi_{ikn}}(y) = 0,$

where $G_{\xi_{ikn}}$ denotes the distribution of ξ_{ikn}, from which Lindeberg's condition is immediate (see Chung (1974), p. 205).

From (5.134), (5.135) and (5.125) we conclude

$$(5.136)\qquad n^{-\frac{1}{2}} \underset{\sim}{T}_n \xrightarrow{\mathcal{D}} MVN(\underset{\sim}{0}, C_0\, E\psi^2)\ .$$

Finally (see (5.129) and (5.131),

$$-\left[\frac{1}{n} B_n(\underset{\sim}{0}, \underset{\sim}{T}_n)\right]^{-1} =$$

$$-\left(\frac{1}{n}\begin{bmatrix} \delta_1 h_{n1}(\beta_1 \underset{\sim}{T}_n), & \cdots & , \delta_p h_{n1}(\beta_1 \underset{\sim}{T}_n) \\ \vdots & & \vdots \\ \delta_1 h_{np}(\beta_p \underset{\sim}{T}_n), & \cdots & , \delta_p h_{np}(\beta_p \underset{\sim}{T}_n) \end{bmatrix}\right)^{-1}$$

$$= \begin{bmatrix} \frac{-1}{n}\sum_{i=1}^{n} c_{i1}\delta_1\psi\left(X_i - \beta_1 \sum_{j=1}^{p} c_{ij}T_{nj}\right), \ldots, \frac{-1}{n}\sum_{i=1}^{n} c_{i1}\delta_p\psi\left(X_i - \beta_1 \sum_{j=1}^{p} c_{ij}T_{nj}\right) \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ \frac{-1}{n}\sum_{i=1}^{n} c_{ip}\delta_1\psi\left(X_i - \beta_p \sum_{j=1}^{p} c_{ij}T_{nj}\right), \ldots, \frac{-1}{n}\sum_{i=1}^{n} c_{ip}\delta_p\psi\left(X_i - \beta_p \sum_{j=1}^{p} c_{ij}T_{nj}\right) \end{bmatrix}^{-1}$$

(where we have written T_{nj} for the j^{th} component of $\underset{\sim}{T}_n'$) ;

$$= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} c_{i1}c_{i1}\psi'\left(X_i - \beta_1 \sum_{j=1}^{p} c_{ij}T_{nj}\right), \ldots, \frac{1}{n}\sum_{i=1}^{n} c_{i1}c_{ip}\psi'\left(X_i - \beta_1 \sum_{j=1}^{p} c_{ij}T_{nj}\right) \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ \frac{1}{n}\sum_{i=1}^{n} c_{ip}c_{i1}\psi'\left(X_i - \beta_p \sum_{j=1}^{p} c_{ij}T_{nj}\right), \ldots, \frac{1}{n}\sum_{i=1}^{n} c_{ip}c_{ip}\psi'\left(X_i - \beta_p \sum_{j=1}^{p} c_{ij}T_{nj}\right) \end{bmatrix}^{-1}$$

$$\xrightarrow{P} \left(\left(\lim_{n\to\infty} \frac{C^T C}{n}\right) E\psi'\right)^{-1}$$ using uniform continuity of ψ' on compact sets, (5.123) and a partitioning similar to that in (5.18);

that is,

$$(5.137) \qquad -\left(\frac{1}{n} B_n(\underset{\sim}{0},\underset{\sim}{T}_n)\right)^{-1} \xrightarrow{P} \frac{1}{E\psi'} C_0^{-1} .$$

From (5.137), (5.136), (5.133), (5.132) and Slutsky's theorem, we obtain

$$n^{\frac{1}{2}} \underset{\sim}{T}_n \xrightarrow{\mathcal{D}} MVN\left(\underset{\sim}{0}, \left[\frac{1}{E\psi'} C_0^{-1}\right]^2 C_0 E\psi^2\right) ,$$

i.e.,

$$n^{\frac{1}{2}} \underset{\sim}{T}_n \xrightarrow{\mathcal{D}} MVN\left(\underset{\sim}{0}, C_0^{-1} \frac{E\psi^2}{(E\psi')^2}\right) ,$$

completing the proof of the theorem. □

6

ROBUST ESTIMATORS AND MODEL EXTENSIONS

For the model (3.5):

$$X_i = \sum_{j=1}^{p} c_{ij}\theta_j + \varepsilon_i , \quad i=1,\ldots,n \tag{6.1}$$

with $\varepsilon_1,\ldots,\varepsilon_n$ being i.i.d. random variables with distribution function $G \in F$ (where F is the class of all distribution having normal centres and arbitrary tails) and $C = ((c_{ij}))$ having the form given in (3.5), we

a) defined a class of estimators $\{\underset{\sim}{T}_n = \underset{\sim}{T}_n(\psi) : \psi \in \Psi_c\}$ of $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_p)^T$ (see (3.13) and (3.25))

and b) investigated their asymptotic properties.

We found that for each $\psi \in \Psi_c$:

$$\underset{\sim}{T}_n = \underset{\sim}{T}_n(\psi) \xrightarrow{P} \underset{\sim}{0} \quad \text{(see (5.123))} \tag{6.2}$$

and

$$\underset{\sim}{T}_n = \underset{\sim}{T}_n(\psi) \xrightarrow{\mathcal{D}} MVN \left(\underset{\sim}{0},\ C_0^{-1} \frac{\int_{-c}^{c} \psi^2(y)\phi(y)\,dy}{\left[\int_{-c}^{c} \psi(y)\phi'(y)\,dy\right]^2} \right) \quad \text{(see (5.124))} \tag{6.3}$$

where $C_0 = \lim_{n\to\infty} \dfrac{C^T C}{n}$, and without loss of generality, the true value of $\underset{\sim}{\theta}$ is $\underset{\sim}{0}$ (see (3.18)).

In this section, we consider the problem of finding the optimal estimator $\underset{\sim}{T}_n = \underset{\sim}{T}_n(\psi)$ (optimal according to some criterion we shall specify.) We will then extend our class of distributions F to consider, in addition, a small amount of symmetric contamination of the normal centre and will give the optimal estimator in this case. Again, the case of scale unknown is deferred to sections 7 and 8.

Consider the case of estimation of a location parameter first. Huber (1964) proposed judging the "goodness" of an estimator (or "robustness", meaning, roughly, good performance when reasonably small deviations, from the assumptions made in the model, occur) by it's asymptotic variance. Huber (1964) says:

"Since ill effects from contamination are mainly felt for large sample sizes, it seems that one should primarily optimize large sample robustness properties. Therefore a convenient measure of robustness for asymptotically normal estimators seems to be the supremum of the asymptotic variance ($n \to \infty$) when G ranges over some suitable set of underlying distributions ..." and goes on to give reasons why, even for moderate sample sizes, the asymptotic variance is a better measure of performance than the actual variance. We shall adopt Huber's criterion in our work (of course our "asymptotic variance" is a covariance matrix.)

The model (6.1) with p=1 and c_{i1}=1, i=1,...,n was studied by Collins (1976). It is clear that any optimality results of Collins carry over to our model because of the form of our asymptotic covariance matrix. For example, order matrices by positive definiteness, so that $M < N$ means

$$\tilde{t}^T(N-M)\tilde{t} \geq 0 \text{ for all } \tilde{t} \in R^p \text{ and}$$

(6.4)

$$\tilde{t}^T(N-M)\tilde{t} = 0 \text{ if and only if } \tilde{t} = 0 \text{ .}$$

Then, noting that we may use the phrases "optimal estimator" synonomously with "optimal ψ-function" we suppose that ψ_o is optimal for the model (6.3) with p=1, c_{i1}=1, i=1,...,n, meaning that, according to our criterion above,

(6.5) $\quad \psi_o$ minimizes $\sup\limits_{G \in F} V(\psi,G)$ where

$$(6.6) \qquad V(\psi,G) = \frac{\int_{-c}^{c} \psi^2(y)\phi(y)dy}{\left(\int_{-c}^{c} \psi(y)\phi\ (y)dy\right)^2} \text{ .}$$

Note that, of course, (6.6) is independent of $G \in F$ so in this case, the supremum in (6.5) is redundant.

It follows immediately that

$C_0^{-1}\ V(\psi_o,G) \leq C_0^{-1}\ V(\psi,G)$ in the sense of (6.4) so that the optimal ψ in the model (6.1) coincides with the optimal ψ in the same model with p=1 and c_{i1}=1, i=1,...,n.

It thus suffices to give the results of Collins (1976) and we will do this briefly:

The infimum of $V(\psi,\phi)$ for $\psi \in \Psi_c$ is

$$\frac{1}{\int_{-c}^{c} x^2\phi(x)dx} \text{ .}$$

This infimum is attained by ψ^* where

$$\psi^*(x) = \begin{cases} x, & x \in (-c,c) \\ 0, & \text{otherwise} \end{cases} .$$

$\psi^* \notin \Psi_c$, clearly, but any dominated sequence $\{\psi_j\}$ in Ψ_c satisfying $\psi_j(x) \longrightarrow \psi^*(x)$ a.e. as $j\to\infty$ will satisfy $V(\psi_j,\phi) \longrightarrow V(\psi^*,\phi)$.

Next, define a class of distributions P_ε as follows: fix ε, $0 < \varepsilon < 1$ and say that

(6.7) $G \in P_\varepsilon$ if the density g of G satisfies

$$g(x) = (1-\varepsilon)\ \phi(x) + \varepsilon h(x) \text{ for all } x \in [-d,d]$$

where h is symmetric and smooth.

Collins showed that, according to our criterion above, any discontinuous ψ (in particular ψ^*) cannot be robust, for we have $\sup\{V(\psi,g)\colon G \in P_\varepsilon\} = \infty$ if ψ is discontinuous.

Collins next defined a class

(6.8) Ψ_c' by replacing, in the definition of Ψ_c, the condition that ψ be smooth by the condition that ψ has a piecewise continuous derivative and showed that if

$\frac{\varepsilon}{1-\varepsilon} < 2c\ \phi(0) - 2\Phi(c) + 1$, then

(6.9) $\psi \in \Psi_c'$ minimizes $\sup\{V(\psi,g)\colon G \in P_\varepsilon\}$ if and only if ψ is a nonzero multiple of

$$\psi_o(x) = \begin{cases} x, & |x| \le x_o \\ x_1 \tanh\left[\frac{1}{2} x_1(c-|x|)\right] \text{sign}(x), & x_o \le |x| \le c \\ 0 & |x| \ge c \end{cases}$$

where x_o and x_1 are uniquely determined by

$$x_o = x_1 \tanh\left[\frac{1}{2} x_1(c-x_o)\right]$$

and $\int_c^c [g_o(x) - (1-\varepsilon)\phi(x)]dx = \varepsilon$ and

where

$$g_o(x) = \begin{cases} (1-\varepsilon)\phi(x), & |x| \le x_o \\ \dfrac{(1-\varepsilon)\phi(x_o)}{\cosh^2\left[\frac{1}{2} x_1(c-x_o)\right]} \cosh^2\left[\frac{1}{2} x_1(c-|x|)\right], & x_o \le |x| \le c. \end{cases}$$

Collins (1976') validated this minimax result in the sense that under certain conditions, $V(\psi_o,g)$ is the asymptotic variance of a consistent estimator for all G.

The problem of estimating $\underset{\sim}{\theta}$ in the linear model with $\psi \in \Psi_c'$ and $G \in P_\varepsilon$ has not been undertaken but we have no reason to doubt that with our conditions on C and the conditions of Theorem 2 of Collins (1976') that consistent and asymptotically normal estimators of $\underset{\sim}{\theta}$ can be found.

In any case, a comparison of the matrices

$$C_0^{-1}\left[\sup_{G\in P_\varepsilon} \frac{\int_{-c}^{c} \psi_1^2(y)g(y)dy}{\left(\int_{-c}^{c} \psi_1(y)g'(y)dy\right)^2}\right] \text{ and } C^{-1}\left[\sup_{G\in P_\varepsilon} \frac{\int_{-c}^{c} \psi_2^2(y)g(y)dy}{\left(\int_{-c}^{c} \psi_2(y)g'(y)dy\right)^2}\right]$$

is a natural way of comparing two estimators $\underset{\sim}{T}_n(\psi_1)$ and $\underset{\sim}{T}_n(\psi_2)$ so that the minimax results of Collins (1976) apply.

7

EXTENSION TO THE CASE OF SCALE UNKNOWN

As in section 3, we fix α, $0 < \alpha < .5$, we define $d = \Phi^{-1}(1 - \frac{\alpha}{2})$ and $G \in F$ if and only if there exists $\gamma \in (-\frac{\alpha}{2}, \frac{\alpha}{2})$ such that $G(y) = \Phi(y) + \gamma$ for all $y \in [-d,d]$. Our model is

$$X_i = \sum_{j=1}^{p} c_{ij}\theta_j + \varepsilon_i, \quad i=1,\ldots,n \tag{7.1}$$

where the ε_i are i.i.d. random variables with distribution function $G_\sigma(y)$, where

$$G_\sigma(y) = G(\tfrac{y}{\sigma}) \;, \quad G \in F, \tag{7.2}$$

and $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)^T$ and σ are unknown.

The design matrix C has the same restriction we imposed on it in section 3. Our problem is to estimate $\underset{\sim}{\theta}$.

In the case where σ was known ($\sigma = 1$ without loss of generality), we proposed estimate $\underset{\sim}{\theta}$ by solving the system of equations

$$\sum_{i=1}^{n} c_{ik}\, \psi(x_i - \sum_{j=1}^{p} c_{ij}\theta_j) \;, \quad k=1,\ldots,p, \tag{7.3}$$

for $\psi \in \Psi_c$ where Ψ_c was given by (3.13).

Unfortunately, the resulting estimators are not scale invariant. (Recall that an estimator $\underset{\sim}{\theta}_1$ of $\underset{\sim}{\theta}$ is location invariant or equivariant if $\underset{\sim}{\theta}_1(\underset{\sim}{X} + C\underset{\sim}{t}) = \underset{\sim}{\theta}_1(\underset{\sim}{X}) + \underset{\sim}{t}$, $\underset{\sim}{t} \in \mathbb{R}^p$ and scale invariant if $\underset{\sim}{\theta}_1(\lambda\underset{\sim}{X}) = \lambda\theta_1(\underset{\sim}{X})$

where λ is any scalar. An estimator $\hat{\sigma}_n$ of σ is location invariant if $\hat{\sigma}(X + C\underset{\sim}{t}) = \hat{\sigma}(\underset{\sim}{X})$ and scale invariant if $\hat{\sigma}(\lambda \underset{\sim}{X}) = |\lambda|\hat{\sigma}(\underset{\sim}{X})$.)

In fact, unless the function ψ in (7.3) is of the form

$\psi(x) = |x|^{\delta}$ sign(x), solutions of (7.3) are not location invariant and, clearly, such $\psi^{\underline{s}}$ do not vanish outside a compact set.

To obtain acceptable procedures, we must estimate $\underset{\sim}{\theta}$ so that the estimator is scale as well as location invariant. There are two common procedures for doing this. One is to estimate σ simultaneously with $\underset{\sim}{\theta}$. This is the method we shall use in section 8 and shall comment further on it there. The other, which is the method we employ here, is to first estimate σ from the data and then to solve the system

$$\sum_{i=1}^{n} c_{ik}\, \psi\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\theta_j}{\hat{\sigma}_n}\right) = 0, \quad k=1,\ldots,p \quad \text{for } \underset{\sim}{\theta}.$$

Accordingly, following Collins (1976), we let G_n be the empirical distribution function of the sample and set

$$\hat{\sigma}_n = \frac{G_n^{-1}(1-\alpha) - G_n^{-1}(\alpha)}{\Phi^{-1}(1-\alpha) - \Phi^{-1}(\alpha)} \tag{7.4}$$

where $G_n^{-1}(t) = \inf\{y:\ G_n(y) \geq t\}$, $0 < t < 1$.

Also put

$$b = \frac{\Phi^{-1}(1-\frac{\alpha}{2}) - \Phi^{-1}(\frac{3\alpha}{2})}{\Phi^{-1}(1-\alpha) - \Phi^{-1}(\alpha)} \quad . \tag{7.5}$$

Then it was shown in Collins (1976) that

(7.6) there is a number $\beta_\gamma \in [1,b)$ such that

(7.7) $\hat{\sigma}_n \xrightarrow{P} \beta_\gamma \sigma$.

β_γ is given by

(7.8) $$\frac{\Phi^{-1}(1-\alpha+\gamma) - \Phi^{-1}(\alpha+\gamma)}{\Phi^{-1}(1-\alpha) - \Phi^{-1}(\alpha)} .$$

The upper limit on the biasing factor is small for reasonably small α.

Now set

(7.9) $c' = \frac{d-k}{b}$ and define

(7.10) $\Psi_{c'}$ to be the class Ψ_c of (3.13) with c replaced by c'.

We propose to estimate the true $\underset{\sim}{\theta}$ by solving

(7.11) $$\sum_{i=1}^{n} c_{ik} \, \psi\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\theta_j}{\hat{\sigma}_n} \right) = 0, \quad k = 1,\ldots,p \text{ where } \psi \in \Psi_{c'}.$$

The proof of location invariance of solutions of (7.11) follows in exactly the same way as our proof of (3.15). We will check that any solution of (7.11) is also scale invariant.

(Note that $\hat{\sigma}_n$ is location and scale invariant.)

We must show that for any scalar λ

(7.12) $\underset{\sim}{\theta}_1(\lambda \underset{\sim}{X}) = \lambda\theta_1(\underset{\sim}{X})$ where $\theta_1(\underset{\sim}{X})$ is any solution of (7.11).

To prove (7.12), replace $\underset{\sim}{X}$ by $\lambda\underset{\sim}{X}$ in (7.11). (We may assume $\lambda \neq 0$, since the result is trivial otherwise.)

$\hat{\sigma}_n = \hat{\sigma}_n(\underset{\sim}{X})$ becomes $\hat{\sigma}_n(\lambda\underset{\sim}{X})$ which, by scale invariance of $\hat{\sigma}_n$, equals $|\lambda|\ \hat{\sigma}_n(\underset{\sim}{X}) = |\lambda|\hat{\sigma}_n$.

Then (7.11) becomes

$$\sum_{i=1}^{n} c_{ik}\ \psi\left(\frac{\lambda X_i - \sum_{j=1}^{p} c_{ij}\theta_j}{|\lambda|\hat{\sigma}_n}\right) = 0, \quad k = 1,\ldots,p \tag{7.13}$$

i.e.

$$\sum_{i=1}^{n} c_{ik}\ \psi\left(\frac{\lambda}{|\lambda|}\left[\frac{X_i - \sum_{j=1}^{p} c_{ij}\left(\frac{\theta_j}{\lambda}\right)}{\hat{\sigma}_n}\right]\right) = 0, \quad k = 1,\ldots,p \tag{7.14}$$

i.e.

$$\sum_{i=1}^{n} c_{ik}\ \psi\left(\frac{X_i - \sum_{j=1}^{p} c_{ij}\left(\frac{\theta_j}{\lambda}\right)}{\hat{\sigma}_n}\right) = 0, \quad k = 1,\ldots,p, \tag{7.15}$$

by skew-symmetry of ψ,

from which we see that the j^{th} component of the solution of (7.13) upon division by λ is the j^{th} component of the solution of (7.11)

i.e.

$$\frac{\theta_1(\lambda\underset{\sim}{X})}{\lambda} = \theta_1(\underset{\sim}{X}), \quad \text{proving (7.12).}$$

We may thus assume throughout this section that the true values of $\underset{\sim}{\theta}$ and σ are

$$\underset{\sim}{\theta} = \underset{\sim}{0}\ , \quad \sigma = 1. \tag{7.16}$$

For $\eta > 0$, set

$$(7.17)\quad F_\eta(\underset{\sim}{t}) = \begin{pmatrix} \sum_{i=1}^{p} a_{i1} q_i E\psi\left(\frac{X - \sum_{j=1}^{p} a_{ij} t_j}{\eta}\right) \\ \vdots \\ \sum_{i=1}^{p} a_{ip} q_i E\psi\left(\frac{X - \sum_{j=1}^{p} a_{ij} t_j}{\eta}\right) \end{pmatrix} = \begin{pmatrix} f_{\eta 1}(\underset{\sim}{t}) \\ \vdots \\ f_{\eta p}(\underset{\sim}{t}) \end{pmatrix}$$

$$(7.18)\quad F_{n\eta}(t) = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} c_{i1} \psi\left(\frac{X_i - \sum_{j=1}^{p} c_{ij} t_j}{\eta}\right) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^{n} c_{ip} \psi\left(\frac{X_i - \sum_{j=1}^{p} c_{ij} t_j}{\eta}\right) \end{pmatrix} = \begin{pmatrix} f_{n\eta 1}(\underset{\sim}{t}) \\ \vdots \\ f_{n\eta p}(\underset{\sim}{t}) \end{pmatrix}$$

<u>(7.19) DEFINITION</u>:

For $\psi \in \Psi_{c'}$, the sequence of estimators $\{\underset{\sim}{T}_{n\hat{\sigma}_n}\} = \{\underset{\sim}{T}_{n\hat{\sigma}_n}(\psi)\}$ of the true $\underset{\sim}{\theta}$ is defined as follows:

set $\underset{\sim}{t}^{o} = \underset{\sim}{\theta}^{*}$ (see (3.8)) and then form the sequence

$$\underset{\sim}{t}^{k+1} = \underset{\sim}{t}^{k} - F_{n\hat{\sigma}_n}(\underset{\sim}{t}^{k})^{-1} F_{n\hat{\sigma}_n}(\underset{\sim}{t}^{k}) \qquad k = 0,1,2,\ldots,$$

in analogy with (3.25).

Then set

$$\underset{\sim}{T}_{n\hat{\sigma}_n} = \begin{cases} \lim_{k\to\infty} \underset{\sim}{t}^{k}, & \text{if this limit exists} \\ \underset{\sim}{\theta}^{*}, & \text{otherwise.} \end{cases}$$

(7.20) THEOREM: (*Consistency of* $\underset{\sim}{T}_{n\hat{\sigma}_n}$) .

(7.21) $\underset{\sim}{T}_{n\hat{\sigma}_n} \xrightarrow{P} \underset{\sim}{0}$.

Proof: By (7.16) and (7.7), there exists

$\beta = \beta_\gamma \in [1,b)$ such that

(7.22) $\hat{\sigma}_n \xrightarrow{P} \beta$.

Elementary properties of

$$E\,\psi\left(\frac{X - \sum_{j=1}^{p} a_{ij}t_j}{\beta}\right) \text{ and } E\,\psi'\left(\frac{X - \sum_{j=1}^{p} a_{ij}t_j}{\beta}\right) \text{ can be derived,}$$

as previously for $E\,\psi\left(X - \sum_{j=1}^{p} a_{ij}t_j\right)$ and $E\,\psi'\left(X - \sum_{j=1}^{p} a_{ij}t_j\right)$

in section 3 (see Collins (1976) section 4 and see also chapter 8 of this work.)

In particular

(7.23) $$\left|\frac{E\,\psi\left(\frac{X - \sum_{j=1}^{p} a_{ij}t_j}{\beta}\right)}{E\psi'\left(\frac{X - \sum_{j=1}^{p} a_{ij}t_j}{\beta}\right)}\right| < 2\left|\sum_{j=1}^{p} a_{ij}t_j\right| \text{ for all } \underset{\sim}{t} \in \overline{D} \text{ (see (3.20).)}$$

The analogue of (5.15) in this case is

(7.24) $$\sup\{\ \| F_{n\hat{\sigma}_n}(\underset{\sim}{t}) - F_\beta(\underset{\sim}{t}) \| : \underset{\sim}{t} \in \overline{D}\ \} \xrightarrow{P} 0.$$

To check this, write

$$(7.25)\qquad \sup\{\ \|F_{n\hat{\sigma}_n}(\underset{\sim}{t}) - F_\beta(\underset{\sim}{t})\| : \underset{\sim}{t} \in \overline{D}\ \}$$

$$\leq \sup_{\underset{\sim}{t}\in\overline{D}} \|F_{n\hat{\sigma}_n}(\underset{\sim}{t}) - F_{n\beta}(\underset{\sim}{t})\| + \sup_{\underset{\sim}{t}\in\overline{D}} \|F_{n\beta}(\underset{\sim}{t}) - F_\beta(\underset{\sim}{t})\| .$$

The first term on the right hand side of the last inequality tends to zero in probability by uniform continuity of F_n (or just ψ, say) and the fact that $\hat{\sigma}_n \xrightarrow{P} \beta$ (we omit details since several proofs of this sort have been done previously in this work.)

The last term in (7.25) tends to zero in probability by exactly the same argument as used in (5.15), and so we arrive at (7.24).

The rest of the proof of consistency is the same as our previous consistency proof so we omit further details.

This completes the proof of (7.20) ||

(7.26) THEOREM: *(Asymptotic normality of $\underset{\sim}{T}_{n\hat{\sigma}_n}$)*

We have for $\psi \in \Psi_c$, and $\underset{\sim}{T}_{n\hat{\sigma}_n}$ as given in (7.19):

$$(7.27)\qquad n^{\frac{1}{2}}\,\underset{\sim}{T}_{n\hat{\sigma}_n} \xrightarrow{\mathcal{D}} MVN\left(\underset{\sim}{0}\ ,\quad C_0^{-1}\,\frac{\int_c^{c'} \psi^2(y)\,\phi(\beta y)\,dy}{\beta\left[\int_c^{c'} \psi(y)\,\phi'(\beta y)\,dy\right]^2}\right) .$$

In the proof below

(7.28) the first p components of $\underset{\sim}{T}_{n\hat{\sigma}_n}$ will be denoted by T_{ni} $(i=1,\ldots,p)$.

Proof: We set, for any $\eta > 0$,

$$(7.29) \qquad H_n \begin{pmatrix} \underset{\sim}{t} \\ \\ \eta \end{pmatrix} = H_n \begin{pmatrix} t_1 \\ \cdot \\ \cdot \\ \cdot \\ t_p \\ \eta \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} c_{i1} \, \psi \left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij} t_j}{\eta} \right) \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} c_{ip} \, \psi \left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij} t_j}{\eta} \right) \end{pmatrix} ,$$

a mapping from

$$\underbrace{\overline{D} \times \mathbb{R}^+}_{p+1 \text{ dimensions}} \longrightarrow \mathbb{R}^p \qquad (\text{here } \mathbb{R}^+ = \{\eta \in \mathbb{R}^1 : \eta > 0\}).$$

Following notation similar to that in (5.130),

write

$$(7.30) \quad H_n \begin{pmatrix} T_{n1} \\ \cdot \\ \cdot \\ \cdot \\ T_{np} \\ \hat{\sigma}_n \end{pmatrix} = H_n \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \beta \end{pmatrix} + B_n \left(\begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \beta \end{pmatrix} , \begin{pmatrix} T_{n1} \\ \cdot \\ \cdot \\ \cdot \\ T_{np} \\ \hat{\sigma}_n \end{pmatrix} \right) \begin{pmatrix} T_{n1}-0 \\ \cdot \\ \cdot \\ \cdot \\ T_{np}-0 \\ \hat{\sigma}_n-\beta \end{pmatrix} ,$$

Here,

$$(7.31)\quad B_n\left(\begin{pmatrix}0\\0\\ \cdot\\ \cdot\\ \cdot\\ \beta\end{pmatrix},\ \begin{pmatrix}T_{n1}\\ \cdot\\ \cdot\\ \cdot\\ T_{np}\\ \hat{\sigma}_n\end{pmatrix}\right)$$

equals

$$\begin{bmatrix}\sum_{i=1}^{n} c_{i1}\delta_1\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right), & \cdots & \sum c_{i1}\,\delta_{p+1}\,\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right)\\ \vdots & & \vdots\\ \sum_{i=1}^{n} c_{ip}\delta_1\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right), & \cdots & \sum c_{ip}\,\delta_{p+1}\,\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right)\end{bmatrix}$$

some α_i, $0 < \alpha_i < 1$, $i=1,\ldots,p$.

This may be written as

$$(7.32)\quad B_n\left(\begin{pmatrix}0\\ \cdot\\ \cdot\\ \cdot\\ 0\\ \beta\end{pmatrix},\ \begin{pmatrix}T_{n1}\\ \cdot\\ \cdot\\ \cdot\\ T_{np}\\ \hat{\sigma}_n\end{pmatrix}\right)\begin{pmatrix}T_{n1}\\ \cdot\\ \cdot\\ \cdot\\ T_{np}\\ \hat{\sigma}_n-\beta\end{pmatrix} =$$

$$L_n \left(\begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \beta \end{pmatrix}, \begin{pmatrix} T_{n1} \\ \cdot \\ \cdot \\ \cdot \\ T_{np} \\ \hat{\sigma}_n \end{pmatrix} \right) \underset{\sim}{T}_{n\hat{\sigma}_n} + M_n \left(\begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ \beta \end{pmatrix}, \begin{pmatrix} T_{n1} \\ \cdot \\ \cdot \\ \cdot \\ T_{np} \\ \hat{\sigma}_n \end{pmatrix} \right) (\hat{\sigma}_n - \beta)$$

$$= L_n\ (\underset{\sim}{x},\underset{\sim}{y})\ \underset{\sim}{T}_{n\hat{\sigma}_n} + M_n\ (\underset{\sim}{x},\underset{\sim}{y})\ (\hat{\sigma}_n-\beta)\ \text{, say,}$$

where we have written

(7.33) $\qquad (0,\ldots,0,\beta)^T$ as $\underset{\sim}{x}$,

(7.34) $\qquad (T_{n1},\ldots,T_{np},\hat{\sigma}_n)^T = \underset{\sim}{T}_{n\hat{\sigma}_n}$ as $\underset{\sim}{y}$,

(7.35)
$$L_n(\underset{\sim}{x},\underset{\sim}{y}) = \begin{bmatrix} \sum_{i=1}^{n} c_{i1}\delta_1\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right), & \ldots\ldots & \sum_{i=1}^{n} c_{i1}\delta_p\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \sum c_{ip}\delta_1\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right), & \ldots\ldots & \sum c_{ip}\delta_p\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right) \end{bmatrix}_{p\times p}$$

and

$$(7.36) \qquad M_n(\underset{\sim}{x},\underset{\sim}{y}) = \begin{bmatrix} \sum_{i=1}^{n} c_{i1}\ \delta_{p+1}\ \psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right) \\ \vdots \\ \sum_{i=1}^{n} c_{ip}\ \delta_{p+1}\ \psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right) \end{bmatrix}_{p \times 1} .$$

On re-arranging terms in (7.30) and multiplying through by $n^{\frac{1}{2}}$ we get, using the last equality in (7.32),

$$(7.37) \qquad n^{\frac{1}{2}}\underset{\sim}{T}_{n\hat{\sigma}_n} = -\left(\frac{1}{n} L_n(\underset{\sim}{x},\underset{\sim}{y})\right)^{-1}\left[n^{-\frac{1}{2}} H_n(\underset{\sim}{x}) + n^{-\frac{1}{2}} M_n(\underset{\sim}{x},\underset{\sim}{y})(\hat{\sigma}_n-\beta) - n^{-\frac{1}{2}} H_n(\underset{\sim}{y})\right].$$

Now,

$$(7.38) \qquad n^{-\frac{1}{2}} H_n(\underset{\sim}{x}) = n^{-\frac{1}{2}} H_n\left[(0,\ldots,0,\beta)^T\right] \quad \text{(from (7.33))};$$

$$= n^{-\frac{1}{2}} \begin{bmatrix} \sum_{i=1}^{n} c_{i1}\ \psi\left(\dfrac{X_i}{\beta}\right) \\ \vdots \\ \sum c_{ip}\ \psi\left(\dfrac{X_i}{\beta}\right) \end{bmatrix} \quad \text{(from (7.29))};$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} \begin{bmatrix} c_{i1}\ \psi\left(\dfrac{X_i}{\beta}\right) \\ \vdots \\ c_{ip}\ \psi\left(\dfrac{X_i}{\beta}\right) \end{bmatrix}$$

$$\xrightarrow[\text{by C.L.T.}]{\mathcal{D}} MVN\left(\underset{\sim}{0},\ E\,\psi^2\left(\frac{X}{\beta}\right) \lim_{n\to\infty} \frac{C^T C}{n}\right)$$

$$= MVN\left(\underset{\sim}{0},\ C_o\, E\,\psi^2\left(\frac{X}{\beta}\right)\right).$$

Note that in arriving at this limiting distribution, we used the fact that for $r,k = 1,\ldots,p$,

$$\operatorname{cov}\left(n^{-\frac{1}{2}} \sum_{i=1}^{n} c_{ir}\, \psi\left(\frac{X_i}{\beta}\right),\ n^{-\frac{1}{2}} \sum_{i=1}^{n} c_{ik}\, \psi\left(\frac{X_i}{\beta}\right) \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} c_{ir}\, c_{ik}\, E\,\psi^2\left(\frac{X}{\beta}\right) + \frac{1}{n} \sum_{i\neq j}\sum c_{ir}\, c_{jk}(0)$$

to give us the expression for the variance - covariance matrix. The condition (5.126) is used to ensure that Lindeberg's condition holds (see (5.135).)

Next,

$$(7.39) \qquad n^{-\frac{1}{2}} H_n(\underset{\sim}{y}) = n^{-\frac{1}{2}} H_n \begin{pmatrix} T_{n1} \\ \cdot \\ \cdot \\ \cdot \\ T_{np} \\ \hat{\sigma}_n \end{pmatrix}, \qquad \text{from (7.34)}$$

$$= \quad n^{-\frac{1}{2}} \begin{pmatrix} \sum_{i=1}^{n} c_{i1} \, \psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij} T_{nj}}{\hat{\sigma}_n}\right) \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} c_{ip} \, \psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij} T_{nj}}{\hat{\sigma}_n}\right) \end{pmatrix}$$

$$\xrightarrow{P} \underset{\sim}{0} \, ,$$

since $P\left(\underset{\sim}{T}_{n\hat{\sigma}_n} = \text{zero of } H_n\right) \xrightarrow{n\to\infty} 1$, by (7.20).

(Note that we have been writing

T_{nk} for the k^{th} component of $\underset{\sim}{T}_{n\hat{\sigma}_n}$

instead of the more accurate

$T_{n\hat{\sigma}_n k}$, but we dropped the $\hat{\sigma}_n$ for reasons of

notational brevity.)

Next, we examine the term

(7.40) $\quad n^{-\frac{1}{2}} M_n(x,y) \; (\hat{\sigma}_n - \beta)$

$$= n^{-\frac{1}{2}} \begin{Bmatrix} \sum_{i=1}^{n} c_{i1}\ \delta_{p+1}\ \psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij}\alpha_1 T_{nj}}{\beta + \alpha_p(\hat{\sigma}_n - \beta)}\right) \\ \vdots \\ \sum c_{ip}\ \delta_{p+1}\ \psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij}\alpha_p T_{nj}}{\beta + \alpha_p(\hat{\sigma}_n - \beta)}\right) \end{Bmatrix} \left(\hat{\sigma}_n - \beta\right).$$

Splitting the $n^{-\frac{1}{2}}$ as $n^{\frac{1}{2}}$ times n^{-1} and recalling that

(7.41) δ_{p+1} means partial derivative with respect to $\hat{\sigma}_n$,

we get from (7.40) ,

(7.42) $$n^{-\frac{1}{2}} M_n(\underset{\sim}{x},\underset{\sim}{y})(\hat{\sigma}_n-\beta) =$$

$$n^{\frac{1}{2}}(\hat{\sigma}_n-\beta) \begin{pmatrix} -\dfrac{1}{n}\sum_{i=1}^{n}\left\{\left[c_{i1}\dfrac{X_i-\sum_{j=1}^{p}c_{ij}\alpha_1 T_{nj}}{\{\beta+\alpha_1(\hat{\sigma}_n-\beta)\}^2}\right]\left[\psi'\ \dfrac{X_i-\sum_{j=1}^{p}c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right]\right\} \\ \\ -\dfrac{1}{n}\sum_{i=1}^{n}\left\{\left[c_{i1}\dfrac{X_i-\sum_{j=1}^{p}c_{ij}\alpha_p T_{nj}}{[\beta+\alpha_p(\hat{\sigma}_n-\beta)]^2}\right]\left[\psi'\ \dfrac{X_i-\sum_{j=1}^{p}c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right]\right\} \end{pmatrix}.$$

We see that each component of the vector in (7.42) approaches zero in probability, using (7.21), (7.22), uniform continuity of ψ' on $[-c', c']$ and the fact that $E[X\psi'(\frac{X}{\beta})] = \int_{-c'}^{c'} x\psi'(\frac{x}{\beta})\phi(x)dx = 0$ (see

Collins (1973), p.101). (Of course, a partitioning such as that done in (5.18) is done here before applying the W.L.L.N$\underline{\text{S}}$).

Also, $n^{\frac{1}{2}}(\hat{\sigma}_n-\beta) \xrightarrow{P} 0$ since β is the normalized difference of two quantiles and $\hat{\sigma}_n$ the normalized difference of the corresponding sample quantiles. Thus

(7.43) $\quad n^{-\frac{1}{2}} M_n(\underset{\sim}{x},\underset{\sim}{y})\ (\hat{\sigma}_n-\beta) \xrightarrow{P} \underset{\sim}{0}$. $\quad$ Finally,

(7.44) $\quad -\left(\frac{1}{n} L_n(\underset{\sim}{x},\underset{\sim}{y})\right)^{-1}$

$$= -\left(\frac{1}{n}\begin{bmatrix} \sum_{i=1}^{n} c_{i1}\,\delta_1\,\psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right), & \cdots & \sum_{i=1}^{n} c_{i1}\,\delta_p\,\psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right) \\ \vdots & & \vdots \\ \sum_{i=1}^{n} c_{ip}\,\delta_1\,\psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right), & \cdots & \sum_{i=1}^{n} c_{ip}\,\delta_p\,\psi\left(\dfrac{X_i - \sum_{j=1}^{p} c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right) \end{bmatrix}\right)^{-1}$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{c_{i1}c_{i1}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\,\psi'\left(\frac{X_i-\sum_{j=1}^{p}c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right)\right\}, & \ldots, & \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{c_{i1}c_{ip}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\,\psi'\left(\frac{X_i-\sum_{j=1}^{p}c_{ij}\alpha_1 T_{nj}}{\beta+\alpha_1(\hat{\sigma}_n-\beta)}\right)\right\} \\ \vdots & & \vdots \\ \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{c_{ip}c_{i1}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\,\psi'\left(\frac{X_i-\sum_{j=1}^{p}c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right)\right\}, & \ldots, & \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{c_{ip}c_{ip}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\,\psi'\left(\frac{X_i-\sum_{j=1}^{p}c_{ij}\alpha_p T_{nj}}{\beta+\alpha_p(\hat{\sigma}_n-\beta)}\right)\right\} \end{bmatrix}^{-1}$$

$$\xrightarrow{P} \left(\lim_{n\to\infty}\frac{C^T C}{n}\right)^{-1}\left(\lim_{n\to\infty}[\beta+\alpha_p(\hat{\sigma}_n-\beta)]\right)\frac{1}{E\psi'\left(\frac{X}{\lim_{n\to\infty}[\beta+\alpha_p(\hat{\sigma}_n-\beta)]}\right)}$$

$$= \beta\, C_o^{-1}\frac{1}{E\psi'\left(\frac{X}{\beta}\right)} .$$

From (7.38), (7.39), (7.43), (7.44) and (7.37) we have, using Slutsky's theorem,

$$n^{\frac{1}{2}}\,\underset{\sim}{T}_n \xrightarrow{\mathcal{D}} MVN\left(\underset{\sim}{0},\; C_o\; E\,\psi^2\left(\frac{X}{\beta}\right)\left(\beta C_o^{-1}\frac{1}{E\,\psi'\left(\frac{X}{\beta}\right)}\right)^2\right)$$

and this is easily seen to be the same as

$$n^{\frac{1}{2}}\,\underset{\sim}{T}_n \xrightarrow{\mathcal{D}} MVN\left(0,\; C_o^{-1}\frac{\int_{c'}^{c'}\psi^2(y)\,\phi(\beta y)\,dy}{\beta\left[\int_{c'}^{c'}\psi(y)\,\phi'(\beta y)\,dy\right]^2}\right),$$

and this completes the proof of the theorem. □

8

CONTINUATION OF THE CASE OF SCALE UNKNOWN – SIMULTANEOUS ESTIMATION OF THE REGRESSION VECTOR AND SCALE PARAMETER

In this section, we retain the model of section 7 and consider the problem of estimating $\underset{\sim}{\theta}$ by estimating simultaneously with it the scale parameter σ. We shall defer comparison of this method and the method of the previous section until after we have done the analysis. A discussion of the ranges allowed by certain parameters will be given in (8.120).

(8.1) Fix α, $\quad 0 < \alpha \leq .05$

and

let d, k, $\underset{\sim}{\theta}^*$, $\underset{\sim}{\theta}^{**}$, $\hat{\sigma}$, $\beta = \beta_\gamma$ and b have the same definitions as before (see section 3 and section 7). We let $\underset{\sim}{\theta}_0$ denote the true value of $\underset{\sim}{\theta}$ and σ_0 denote the true value of σ.

Let c be any number satisfying

(8.2) $$c \leq .90 \; .$$

We set

(8.3) $$c' = \frac{c}{2b-1}$$

and define the class

(8.4) $\Psi_{c'} = \{\psi : \mathbb{R}^1 \to \mathbb{R}^1$ such that ψ is smooth, ψ vanishes outside $[-c', c']$, is non-negative on $[0,c]$ and is not identically zero on $[0,c]$.$\}$

Now define

$$\rho(x) = \int_{-\infty}^{x} \psi(y)\,dy \ , \qquad \psi \in \Psi_{c'} \ , \tag{8.5}$$

$$a_n = (n-p)E\left[U\psi(U) - \rho(U)\right] \quad \text{where} \tag{8.6}$$

U has the standard normal distribution,

and put

$$\Delta_i = X - \sum_{j=1}^{p} a_{ij}t_j, \quad i=1,\ldots,p \ , \ \underset{\sim}{t} \in R^1 \ , \tag{8.7}$$

where each Δ_i has distribution $G \in F$ (see (7.1) and (7.2)).

Sometimes, when the index i is unimportant in $X - \sum_{j=1}^{p} a_{ij}t_j$, we will write

$$\Delta = X - t = X - \sum_{j=1}^{p} a_{ij}t_j . \tag{8.8}$$

Finally, set

$$a = \lim_{n\to\infty} \frac{a_n}{n} . \tag{8.9}$$

We propose to estimate $\underset{\sim}{\theta}$ and σ jointly in the model (7.1) by solving the system

$$H_n\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} = \underset{(p+1)\times 1}{\tilde{0}} \tag{8.10}$$

where

$$(8.11)\quad H_n\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} c_{i1}\,\psi\left(\frac{X_i-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{p} c_{ip}\,\psi\left(\frac{X_i-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right) \\ \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{X_i-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right)\psi\left(\frac{X_i-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right) - \rho\left(\frac{X_i-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right)\right] - \frac{a_n}{n} \end{pmatrix}.$$

It is clear that (8.10) has multiple roots so we make the following definition:

(8.12) DEFINITION: Set $\begin{pmatrix}\underset{\sim}{t}^0\\ \sigma^0\end{pmatrix} = \begin{pmatrix}\underset{\sim}{\theta}^*\\ \hat{\sigma}_n\end{pmatrix}$ (note that we dropped the subscript n from $\underset{\sim}{\theta}^*$ — see (3.8))

and form the sequence

$$(8.13)\qquad \begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{k+1} = \begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{k} - \left(H_n'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{k}\right)^{-1} H_n\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{k}, \quad k = 0,1,2,\ldots .$$

Then set

$$\underset{(p+1)\times 1}{\underset{\sim}{T}_n} = T_n(\psi) = \begin{cases} \lim_{k\to\infty}\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{k}, & \text{if this limit exists} \\ \begin{pmatrix}\underset{\sim}{\theta}^*\\ \hat{\sigma}_n\end{pmatrix}, & \text{otherwise .}\end{cases}$$

(8.14) REMARK: The existence of $H_n\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1}$ and boundedness of $\left\|H_n'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1}\right\|$, in the probability limit sense in some range will follow from the existence of $H'\binom{\underset{\sim}{t}}{\sigma}^{-1}$ and boundedness of $\|H'\binom{\underset{\sim}{t}}{\sigma}^{-1}\|$ where $H\binom{\underset{\sim}{t}}{\sigma}$ is defined below (see the argument in (5.6).)

Set

(8.15)
$$H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{p} a_{i1}q_i E\psi\left(\frac{\Delta_i}{\sigma}\right) \\ \vdots \\ \sum_{i=1}^{p} a_{ip}q_i E\psi\left(\frac{\Delta_i}{\sigma}\right) \\ \sum_{i=1}^{p} q_i E\left[\frac{\Delta_i}{\sigma}\,\psi\left(\frac{\Delta_i}{\sigma}\right) - \rho\left(\frac{\Delta_i}{\sigma}\right)\right] - a \end{pmatrix}.$$

It is easily checked that

(8.16)
$$EH_n\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} = H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}$$

and the fact that

(8.17) $\sup \|H_n\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} - H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}\| \xrightarrow{P} 0$, where the supremum is taken over all $\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}$ in some appropriate closed neighbourhood of $\begin{pmatrix}\underset{\sim}{\theta}_0\\ \sigma_0\end{pmatrix}$, is proved by the same argument as in (5.13).

We shall call (8.15) the asymptotic deterministic version of (8.11).

(8.18) REMARK: We remark that any solution of (8.10) is location and scale invariant. This can be checked as follows:

Replace $\underset{\sim}{X}$ by $\underset{\sim}{X} + C\underset{\sim}{s}$ in (8.10), $\underset{\sim}{s} \in \mathbb{R}^p$. We get

$$(8.19)\qquad \begin{pmatrix} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} c_{i1}\,\psi\left(\dfrac{X_i+\sum_{j=1}^{p} c_{ij}s_j-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right) \\ \vdots \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} c_{ip}\,\psi\left(\dfrac{X_i+\sum_{j=1}^{p} c_{ij}s_j-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right) \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left[\left(\dfrac{X_i+\sum_{j=1}^{p} c_{ij}s_j-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right)\psi\left(\dfrac{X_i+\sum_{j=1}^{p} c_{ij}s_j-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right)\right. \\ \left.\qquad -\rho\left(\dfrac{X_i+\sum_{j=1}^{p} c_{ij}s_j-\sum_{j=1}^{p} c_{ij}t_j}{\sigma}\right)\right]-\dfrac{a_n}{n} \end{pmatrix} = \underset{\sim}{0}$$

or

$$\begin{pmatrix} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} c_{i1}\,\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}(t_j-s_j)}{\sigma}\right) \\ \vdots \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} c_{ip}\,\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}(t_j-s_j)}{\sigma}\right) \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left[\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}(t_j-s_j)}{\sigma}\right)\psi\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}(t_j-s_j)}{\sigma}\right)\right. \\ \left.\qquad -\rho\left(\dfrac{X_i-\sum_{j=1}^{p} c_{ij}(t_j-s_j)}{\sigma}\right)\right]-\dfrac{a_n}{n} \end{pmatrix} = \underset{\sim}{0}$$

so that any solution of (8.19) minus $(s_1,\ldots s_p,0)^T$ equals the corresponding solution of (8.10), proving location invariance of solutions of (8.10).

Similarly, if in (8.10) we replace $\underset{\sim}{X}$ by $\lambda\underset{\sim}{X}$ we get the system

$$(8.20)\quad \begin{pmatrix} \frac{1}{n}\sum\limits_{i=1}^{n} c_{i1}\,\psi\left(\dfrac{\lambda X_i-\sum\limits_{j=1}^{p} c_{ij}t_j}{\sigma}\right) \\ \vdots \\ \frac{1}{n}\sum\limits_{i=1}^{n} c_{ip}\,\psi\left(\dfrac{\lambda X_i-\sum\limits_{j=1}^{p} c_{ij}t_j}{\sigma}\right) \\ \frac{1}{n}\sum\limits_{i=1}^{n}\left[\left(\dfrac{\lambda X_i-\sum\limits_{j=1}^{p} c_{ij}t_j}{\sigma}\right)\psi\left(\dfrac{\lambda X_i-\sum\limits_{j=1}^{p} c_{ij}t_j}{\sigma}\right)-\rho\left(\dfrac{\lambda X_i-\sum\limits_{j=1}^{p} c_{ij}t_j}{\sigma}\right)\right]-\dfrac{a_n}{n} \end{pmatrix} = \underset{\sim}{0}$$

or

$$(8.21)\quad \begin{pmatrix} \frac{1}{n}\sum\limits_{i=1}^{n} c_{i1}\dfrac{\lambda}{|\lambda|}\,\psi\left(\dfrac{X_i-\sum\limits_{j=1}^{p} c_{ij}\frac{t_j}{\lambda}}{\sigma/|\lambda|}\right) \\ \vdots \\ \frac{1}{n}\sum\limits_{i=1}^{n} c_{ip}\dfrac{\lambda}{|\lambda|}\,\psi\left(\dfrac{X_i-\sum\limits_{j=1}^{p} c_{ij}\frac{t_j}{\lambda}}{\sigma/|\lambda|}\right) \\ \frac{1}{n}\sum\limits_{i=1}^{n}\left[c_{ip}\dfrac{\lambda^2}{|\lambda|^2}\left(\dfrac{X_i-\sum\limits_{j=1}^{p} c_{ij}\frac{t_j}{\lambda}}{\sigma/|\lambda|}\right)\psi\left(\dfrac{X_i-\sum\limits_{j=1}^{p} c_{ij}\frac{t_j}{\lambda}}{\sigma/|\lambda|}\right)-\rho\left(\dfrac{X_i-\sum\limits_{j=1}^{p} c_{ij}\frac{t_j}{\lambda}}{\sigma/|\lambda|}\right)\right]-\dfrac{a_n}{n} \end{pmatrix} = \underset{\sim}{0}\,.$$

so that any solution of (8.20) with t_j replaced by t_j/λ and σ replaced by $\sigma/|\lambda|$ equals the corresponding solution of (8.10) proving scale invariance of the solutions of (8.10).

We may thus assume in what follows that the true value of

$$\begin{pmatrix}\underset{\sim}{\theta}\\ \sigma\end{pmatrix} \quad \text{is} \quad \begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix}$$

i.e.

$$\begin{pmatrix}\underset{\sim}{\theta}_0\\ \sigma_0\end{pmatrix} = \begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix} \quad . \tag{8.22}$$

We shall show that the estimator of $\underset{\sim}{\theta}$ resulting from (8.12) is consistent and asymptotically normal.

As a first step in doing this, we will show (following the technique we used in the scale known case) that the Newton's method solution of $H\binom{\underset{\sim}{t}}{\sigma} = \underset{\sim}{0}$, with starting value $\binom{\underset{\sim}{\theta}^{**}}{\beta}$, is $\binom{\underset{\sim}{0}}{1}$.

Note that

$$\begin{pmatrix}\underset{\sim}{\theta}^{**}\\ \cdot\\ \beta\end{pmatrix} \in D \; X \; [1,b), \text{ where } X \text{ here denotes Cartesian product,} \tag{8.23}$$

(see (4.13b)) and (7.6)).

Now $D \; X \; [1,b)$ is a closed set. We shall have to perform our analysis in an open neighbourhood of $\binom{\underset{\sim}{\theta}^{**}}{\beta}$, for, otherwise, the attraction theorems we wish to use fail.

It is easy to check that H is continuously F-differentiable on any neighbourhood of $\binom{\underset{\sim}{0}}{1}$ that we wish to work in (see (4.7) for the proof of smoothness on D of the function F there.)

We will see later that $H'\binom{t}{\sigma}$ is a symmetric matrix. Consequently, by Theorem 4.1.6, p. 95 of Ortega and Rheinboldt (1970), H is a gradient mapping. This fact is of considerable interest in the solution of non-linear equations in general. It turns the problem of solving systems of equations into minimization of non-linear functionals (although, in fact, there is a way of doing this in general.) In all our work to date, we have solved systems of equations rather than minimized functionals and will continue to do so in the present problem. For our purposes, symmetry of $H'\binom{t}{\sigma}$ will serve to simplify our analysis.

Now we shall need expressions for expectations of various functions. We summarize these in a lemma.

(8.24) LEMMA: (In this lemma, U denotes a standard normal random variable, so that

$$P(U \leq x) = \Phi(x).)$$

We have:

(8.25) $$E\rho(U) = m - \int_0^{c'} \psi(x)\left[\Phi(x)-\Phi(-x)\right]dx$$

where $\rho(x) = m$ for all x such that $|x| > c'$

(8.26) $$E\left[U\psi(U)\right] = 2\int_0^{c'} \psi(x)x\phi(x)\,dx$$

(8.27) $$E\rho\left(\frac{\Delta}{\sigma}\right) = m - \int_0^{c'} \psi(x)\left[\Phi(\sigma x+t) - \Phi(-\sigma x+t)\right]dx$$

(8.28) $$E\left[\frac{\Delta}{\sigma}\,\psi\left(\frac{\Delta}{\sigma}\right)\right] = 2\sigma\exp(-t^2/2)\int_0^{c'} \psi(x)x\phi(\sigma x)\sum_{j=0}^{\infty}\frac{(\sigma x t)^{2j}}{(2j)!}\,dx$$

(8.29) $$E\psi\left(\frac{\Delta}{\sigma}\right) = -2\sigma\exp(-t^2/2)\int_0^{c'}\psi(x)\phi(\sigma x)\sigma xt\sum_{j=0}^{\infty}\frac{(\sigma xt)^{2j}}{(2j+1)!}\,dx$$

(8.30) $$E\psi'\left(\frac{\Delta}{\sigma}\right) = 2\sigma^2\exp(-t^2/2)\int_0^{c'}\psi(x)\phi(\sigma x)\sigma x\sum_{j=0}^{\infty}(\sigma xt)^{2j}\left(\frac{1}{(2j)!}-\frac{t^2}{(2j+1)!}\right)dx$$

(8.31) $$E\left[\Delta\psi'\left(\frac{\Delta}{\sigma}\right)\right] = 2\sigma^2\exp(-t^2/2)\text{ times}$$
$$\int_0^{c'}\psi(x)\phi(\sigma x)\sigma xt\sum_{j=0}^{\infty}(\sigma xt)^{2j}\left(\frac{1}{(2j)!}+\frac{1-(\sigma x)^2}{(2j+1)!}\right)dx$$

(8.32) $$E\left[\Delta^2\psi'\left(\frac{\Delta}{\sigma}\right)\right] = 2\sigma^3\exp(-t^2/2)\text{ times}$$
$$\int_0^{c'}\left[x\psi(x)-\rho(x)\right]\phi(\sigma x)\sum_{j=0}^{\infty}(\sigma xt)^{2j}\left\{\frac{(\sigma x)^2-1}{(2j)!}-\frac{(\sigma x)^2t^2}{(2j+1)!}\right\}dx$$

(8.33) $$= 2\sigma^2\exp(-t^2/2)\text{ times}$$
$$\int_0^{c'}\psi(x)\phi(\sigma x)\sigma x\sum_{j=0}^{\infty}(\sigma xt)^{2j}\left(\frac{\sigma^2x^2-2}{(2j)!}-\frac{(\sigma x)^2t^2}{(2j+1)!}\right)dx$$

(8.34) $$E\left[\frac{\Delta}{\sigma}\psi\left(\frac{\Delta}{\sigma}\right)-\rho\left(\frac{\Delta}{\sigma}\right)\right] = 2\sigma\exp(-t^2/2)\text{ times}$$
$$\int_0^{c'}\left[x\psi(x)-\rho(x)\right]\phi(\sigma x)\cosh(\sigma xt)\,dx - m\left[\Phi(-\sigma c'+t)+1-\Phi(\sigma c'+t)\right]$$

(8.35) $$= (8.28) - (8.27)$$

(8.36) $$E\left[\rho(U)-\rho\left(\frac{\Delta}{\sigma}\right)\right] = \int_0^{c'}\psi(x)\left[\Phi(\sigma x+t)+\Phi(\sigma x-t)-2\Phi(x)\right]\,dx\ .$$

(8.37) REMARK: Perhaps the easiest way to prove some of these results is to use integration by parts. For example, write

$$E\left[\Delta^2\psi'\right] = \sigma^3 \int_{-c'}^{c'} x^2\psi'(x)\,\phi(\sigma x+t)\ dx.$$

Then set $du = \psi'(x)dx$, $v = x^2\phi(\sigma x+t)$ to arrive rapidly at the result (8.33).

Proof:

$$(8.38)\quad E\rho(U) = \int_{-\infty}^{\infty} \rho(U)\phi(U)\ du$$

$$= \int_{|U|\le c'} \rho(U)\phi(U)\,du + m \int_{|U|>c'} \phi(U)\,du, \text{ since } \rho(U) = m \text{ for } |U| > c'$$

$$= \int_{|U|\le c'} \rho(U)\phi(U)\,du + m\left[\Phi(-c')+1-\Phi(c')\right] .$$

Applying integration by parts to the last written integral in (8.38), we get that (8.38) equals

$$\left[\rho(U)\,\Phi(U)\right]_{-c'}^{c'} - \int_{|U|\le c'} \psi(U)\Phi(U)\,du + m\left[\Phi(-c') + 1 - \Phi(c')\right]$$

$$= m\Phi(c') - m\Phi(-c') - \int_{|U|\le c'} \psi(U)\Phi(U)\,du + m\left[\Phi(-c') + 1 - \Phi(c')\right]$$

$$= m - \left[\int_0^{c'} \psi(U)\Phi(U)\,du + \int_{-c'}^{0} \psi(U)\,\Phi(U)\ du\right] .$$

$$= m - \int_0^{c'} \psi(U)\left[\Phi(u)-\Phi(-u)\right]\ du ,$$

proving (8.25) .

Next,

$$E\left[U\psi(U)\right] = \int_{-\infty}^{\infty} u\psi(u)\,\phi(u)\,du$$

$$= \int_{|U|\le c'} u\psi(U)\,\phi(U)\,du + 0, \text{ since } \psi(x) = 0 \text{ for } |x| > c'$$

$$= 2\int_{0}^{c'} x\psi(x)\,\phi(x)\,dx, \text{ since } x\psi(x) \text{ is even,}$$

proving (8.26) .

REMARK: It is easy to see from the calculations just done that $E\rho(U)$ and $E\left[U\psi(U)\right]$ have the same values as given above even if U has distribution function $G \in F$, where F is given at the start of section 7, so that in defining a_n in (8.6), we could have put $a_n = (n-p)E\left[U\psi(U)-\rho(U)\right]$ with U having distribution function G for any $G \in F$.

Next,

$$E\rho\left(\frac{\Delta}{\sigma}\right) = \int_{-\infty}^{\infty} \rho\left(\frac{\Delta}{\sigma}\right) dG(x) \qquad \text{(see (8.7))}$$

$$= \int_{\{x:\left|\frac{\Delta}{\sigma}\right|\le c'\}} \rho\left(\frac{\Delta}{\sigma}\right) dF(x) + \int_{\{x:\left|\frac{\Delta}{\sigma}\right|> c'\}} \rho\left(\frac{\Delta}{\sigma}\right) dG(x)$$

$$= \int_{\{x:\left|\frac{\Delta}{\sigma}\right|\le c'\}} \rho\left(\frac{\Delta}{\sigma}\right) \phi(x)\,dx + \int_{\{x:\left|\frac{\Delta}{\sigma}\right|> c'\}} dG(x) \quad ,$$

$$\text{since } \left|\frac{\Delta}{\sigma}\right| \le c' \Rightarrow |x| \le d \quad ;$$

$$= \sigma \int_{-c'}^{c'} \rho(x)\phi(\sigma x+t)\,dx + m\left[\Phi(-\sigma c'+t)+1-\Phi(c'\sigma+t)\right]$$

(Put $u = \rho(x)$, $dv = \phi(\sigma x+t)dx \Rightarrow du = \psi(x)dx$, $v = \frac{1}{\sigma}\,\Phi(\sigma x+t)$.)

$$= \left[\rho(x)\Phi(\sigma x+t)\right]_{-c'}^{c'} - \int_{-c'}^{c'} \psi(x)\Phi(\sigma x+t)\,dx + m\left[\Phi(-\sigma c'+t)+1-\Phi(\sigma c'+t)\right]$$

$$= m - \int_{-c'}^{c'} \psi(x)\Phi(\sigma x+t)\,dx$$

$$= m - \int_{0}^{c'} \psi(x)\left[\Phi(\sigma x+t) - \Phi(-\sigma x+t)\right]dx \ ,$$

proving (8.27) .

Next we prove (8.36):

$E\left[\rho(U)-\rho\left(\frac{\Delta}{\sigma}\right)\right]$

$$= m - \int_{0}^{c'} \psi(x)\left[\Phi(x)-\Phi(-x)\right]dx - \left[m-\int_{0}^{c'} \psi(x)\left[\Phi(\sigma x+t)-\Phi(-\sigma x+t)\right]dx\right] \ ,$$

(from (8.25) and (8.27))

$$= \int_{0}^{c'} \psi(x)\left[\Phi(\sigma x+t)-\Phi(-\sigma x+t)-\Phi(x)+\Phi(-x)\right]dx$$

$$= \int_{0}^{c'} \psi(x)\left[\Phi(\sigma x+t)-\{1-\Phi(\sigma x-t)\}-\Phi(x)+\{1-\Phi(x)\}\right]dx$$

$$= \int_{0}^{c'} \psi(x)\left[\Phi(\sigma x+t)+\Phi(\sigma x-t)-2\Phi(x)\right]dx \ ,$$

proving (8.36) .

We now prove (8.29) and then (8.30). Note that these two expressions have been derived in Collins (1973) and (essentially in) Collins (1976) but we enter the proofs here for completeness and also because our notation differs a little bit.

We have

$$E\psi\left(\frac{\Delta}{\sigma}\right) = \int_{-\infty}^{\infty} \psi\left(\frac{\Delta}{\sigma}\right) d(x)$$

$$= \int_{\{x:\left|\frac{\Delta}{\sigma}\right| \le c'\}} \psi\left(\frac{\Delta}{\sigma}\right) \phi(x)dx + 0$$

$$= \sigma \int_{-c'}^{c'} \psi(x)\,\phi(\sigma x+t)\,dx$$

$$= \sigma \int_{0}^{c'} \psi(x)\left[\phi(\sigma x+t)-\phi(\sigma x-t)\right] dx$$

$$= \sigma \exp(-t^2/2) \int_{0}^{c'} \psi(x)\,\phi(\sigma x)\left[e^{-\sigma x t}-e^{\sigma x t}\right] dx$$

$$= -2\sigma \exp(-t^2/2) \int_{0}^{c'} \psi(x)\,\phi(\sigma x) \sum_{j=0}^{\infty} \frac{(\sigma x t)^{2j+1}}{(2j+1)!} \quad ,$$

which is the same as (8.29).

Next,

$$E\psi'\left(\frac{\Delta}{\sigma}\right) = -\sigma E \frac{\delta}{\delta t}\, \psi\left(\frac{\Delta}{\sigma}\right)$$

$$= -\sigma \frac{\delta}{\delta t} E\psi\left(\frac{\Delta}{\sigma}\right)$$

$$= -\sigma \frac{\delta}{\delta t} \int_{\{x:\left|\frac{\Delta}{\sigma}\right| \le c'\}} \psi\left(\frac{\Delta}{\sigma}\right) \phi(x)\,dx$$

$$= -\sigma^2 \frac{\delta}{\delta t} \int_{-c'}^{c'} \psi(x)\,\phi(\sigma x+t)\,dx$$

$$= -\sigma^2 \frac{\delta}{\partial t} \int_0^{c'} \psi(x)\left[\phi(\sigma x+t)-\phi(\sigma x-t)\right] dx$$

$$= -\sigma^2 \int_0^{c'} \psi(x)\left[-(\sigma x+t)\phi(\sigma x+t)-(\sigma x-t)\phi(\sigma x-t)\right] dx$$

$$= \sigma^2 \exp(-t^2/2) \int_0^{c'} \psi(x)\phi(\sigma x)\left[\sigma x\{e^{-\sigma xt}+e^{\sigma xt}\} + t\{e^{-\sigma xt}-e^{\sigma xt}\}\right] dx$$

$$= 2\sigma^2 \exp(-t^2/2) \int_0^{c'} \psi(x)\phi(\sigma x)\left[\sigma x \sum_{j=0}^{\infty} \frac{(\sigma xt)^{2j}}{(2j)!} - t \sum_{j=0}^{\infty} \frac{(\sigma xt)^{2j+1}}{(2j+1)!}\right] dx$$

$$= 2\sigma^2 \exp(-t^2/2) \int_0^{c'} \psi(x)\phi(\sigma x)\sigma x \sum_{j=0}^{\infty} (\sigma xt)^{2j}\left[\frac{1}{(2j)!} - \frac{t^2}{(2j+1)!}\right] dx \quad ,$$

which is (8.30).

(8.39) REMARK: Since we are dealing with a ψ-function that vanishes outside an interval, the interchange of integral and partial derivatives above and in future calculations is most easily justified by the existence and continuity of the integrands involved.

Next we prove (8.28):

$$E\left[\frac{\Delta}{\sigma}\psi\left(\frac{\Delta}{\sigma}\right)\right] = \int_{-\infty}^{\infty} \frac{\Delta}{\sigma}\, \psi\left(\frac{\Delta}{\sigma}\right) dG(x)$$

$$= \int_{\{x: \left|\frac{\Delta}{\sigma}\right| \le c'\}} \frac{\Delta}{\sigma}\, \psi\left(\frac{\Delta}{\sigma}\right) \phi(x)\,dx$$

$$= \sigma \int_{-c'}^{c'} x\psi(x)\,\phi(\sigma x+t)\,dx$$

$$= \sigma \int_{0}^{c'} x\psi(x)\left[\phi(\sigma x+t) + \phi(\sigma x-t)\right] dx$$

$$= \sigma\exp(-t^2/2) \int_{0}^{c'} x\psi(x)\,\phi(\sigma x)\left[e^{-\sigma xt}+e^{\sigma xt}\right] dx$$

$$= 2\sigma\exp(-t^2/2) \int_{0}^{c'} x\psi(x)\,\phi(\sigma x)\cosh(\sigma xt)\,dx \ ,$$

agreeing with (8.28).

Next,

$$E\left[\frac{\Delta}{\sigma}\psi\left(\frac{\Delta}{\sigma}\right) - \rho\left(\frac{\Delta}{\sigma}\right)\right]$$

$$= 2\sigma\exp(-t^2/2) \int^{c'} \left[x\psi(x)-\rho(x)\right]\phi(\sigma x)\cosh(\sigma xt)\,dx$$

$$+ \int_{\{x:\left|\frac{\Delta}{\sigma}\right|>c'\}} \left[\frac{\Delta}{\sigma}\,\psi\left(\frac{\Delta}{\sigma}\right) - \rho\left(\frac{\Delta}{\sigma}\right)\right] dG(x) \ ,$$

(where the first term on the right of the equality sign is derived from calculations exactly the same as those in the derivation of (8.28))

$$= 2\sigma\exp(-t^2/2) \int_{0}^{c'} \left[x\psi(x)-\rho(x)\right]\phi(\sigma x)\cosh(\sigma xt)\,dx - \int_{\{x:\left|\frac{\Delta}{\sigma}\right|>c'\}} \rho\left(\frac{\Delta}{\sigma}\right) dG(x) \ ,$$

since $\left|\frac{\Delta}{\sigma}\right| > c' \Rightarrow \psi\left(\frac{\Delta}{\sigma}\right) = 0;$

$$= 2\sigma \exp(-t^2/2) \int_0^{c'} [x\psi(x) - \rho(x)]\, \phi(\sigma x) \cosh(\sigma x t)\, dx - m[\Phi(-\sigma c'+t)+1$$

$$- \Phi(\sigma c'+t)] \quad , \quad \text{proving (8.34)} \; .$$

Of course, (8.35) is trivial from (8.28) and (8.27).

Next, we have

$$E[\Delta\psi'(\tfrac{\Delta}{\sigma})] = -\sigma^2 \frac{\delta}{\delta\sigma} E\psi(\tfrac{\Delta}{\sigma}) \qquad \text{(clearly)}$$

$$= -\sigma^2 \frac{\delta}{\delta\sigma} [\sigma \int_{-c'}^{c'} \psi(x)\, \phi(\sigma x+t)\, dx]$$

$$= -\sigma^2 \frac{\delta}{\delta\sigma} \int_0^{c'} \sigma\psi(x) [\phi(\sigma x+t) - \phi(\sigma x-t)]\, dx$$

$$= -\sigma^2 \int_0^{c'} \psi(x) \{\sigma \frac{\delta}{\delta\sigma} [\phi(\sigma x+t) - \phi(\sigma x-t)] + [\phi(\sigma x+t) - \phi(\sigma x-t)] \frac{\delta\sigma}{\delta\sigma}\}\, dx$$

$$= -\sigma^2 \int_0^{c'} \psi(x) \{\sigma[-x(\sigma x+t)\phi(\sigma x+t) + x(\sigma x-t)\phi(\ x-t)] + \phi(\sigma x+t)$$

$$- \phi(\sigma x-t)\}\, dx$$

$$= \sigma^2 \int_0^{c'} \psi(x) \{\sigma[x(\sigma x+t)\phi(\sigma x+t) - x(\sigma x-t)\phi(\sigma x-t)] + \phi(\sigma x-t)$$

$$- \phi(\sigma x+t)\}\, dx$$

$$= \sigma^2 \int_0^{c'} \psi(x)\, \sigma x \exp(-t^2/2)\, \phi(\sigma x) [(\sigma x+t) e^{-\sigma x t} - (\sigma x-t) e^{\sigma x t}]\, dx$$

$$+ \sigma^2 \int_0^{c'} \psi(x) \exp(-t^2/2)\, \phi(\sigma x) [e^{\sigma x t} - e^{-\sigma x t}]\, dx$$

$$= \sigma^2 \int_0^{c'} \psi(x)\exp(-t^2/2)\phi(\sigma x)\sigma x\left[\sigma x(e^{-\sigma xt}-e^{\sigma xt}) + t(e^{-\sigma xt}+e^{\sigma xt})\right] dx$$

$$+ \sigma^2 \int_0^{c'} \psi(x)\exp(-t^2/2)\phi(\sigma x)(2)\sinh(\sigma xt)dx$$

$$= 2\sigma^2 \int_0^{c'} \psi(x)\exp(-t^2/2)\phi(\sigma x)\sigma x\left[t\sum_{j=0}^{\infty} \frac{(\sigma xt)^{2j}}{(2j)!} - \sigma x \sum_{j=0}^{\infty} \frac{(\sigma xt)^{2j+1}}{(2j+1)!}\right] dx$$

$$+ 2\sigma^2 \int_0^{c'} \psi(x)\exp(-t^2/2)\phi(\sigma x) \sum_{j=0}^{\infty} \frac{(\sigma xt)^{2j+1}}{(2j+1)!}\, dx$$

$$= 2\sigma^2\exp(-t^2/2) \int_0^{c'} \psi(x)\phi(\sigma x)\sigma xt \sum_{j=0}^{\infty} (\sigma xt)^{2j}\left[\frac{1}{(2j)!} + \frac{1-(\sigma x)^2}{(2j+1)!}\right] dx,$$

proving (8.31).

(8.40) REMARK: We could have expressed $E\left[\Delta\psi'\left(\frac{\Delta}{\sigma}\right)\right]$ in terms of ρ by writing $E\left[\Delta\psi'\left(\frac{\Delta}{\sigma}\right)\right] = \sigma^2 \frac{\delta}{\partial\sigma}\{\sigma \frac{\delta}{\delta t} E\rho\left(\frac{\Delta}{\sigma}\right)\}$ and showing that this equals

$$\sigma^2 \int_0^{c'} \rho(x)\left[\phi(\sigma x+t)\{\sigma^2(\sigma x+t)^2 x-2\sigma(\sigma x+t)-\sigma^2 x\}\right.$$

$$\left.+ \phi(\sigma x-t)\{-\sigma^2(\sigma x-t)^2 x + 2\sigma(\sigma x-t) + \sigma^2 x\}\right.\,.$$

Of course, this last expression could be derived from (8.31) also, by parts or otherwise, but the fact that it is so messy leads one to suspect that we are better off expressing everything in terms of ψ.

Note that as a check on the calculations that led to (8.31), we can also write $E\Delta\psi'\left(\frac{\Delta}{\sigma}\right) = -\sigma \frac{\delta}{\partial t} E\left[\Delta\psi\left(\frac{\Delta}{\sigma}\right)\right] - E\psi\left(\frac{\Delta}{\sigma}\right)$ and, proceeding as above, this leads us to (8.31) again.

Finally, we derive the expressions (8.32) and (8.33) for $E\left[\Delta^2\psi'\left(\frac{\Delta}{\sigma}\right)\right]$:

$$E\left[\Delta^2\psi'\left(\frac{\Delta}{\sigma}\right)\right] = -\sigma^3 E \frac{\delta}{\partial\sigma}\left[\frac{\Delta}{\sigma}\psi\left(\frac{\Delta}{\sigma}\right) - \rho\left(\frac{\Delta}{\sigma}\right)\right]$$

$$= -\sigma^3 \int_{-c'}^{c'} \left[x\psi(x)-\rho(x)\right] \frac{\delta}{\delta\sigma}\left[\sigma\phi(\sigma x+t)\right] dx$$

$$- \sigma^3 \int_{\{x:\left|\frac{\Delta}{\sigma}\right|>c'\}} \frac{\delta}{\delta\sigma}\left[\frac{\Delta}{\sigma}\psi\left(\frac{\Delta}{\sigma}\right) - \rho\left(\frac{\Delta}{\sigma}\right)\right] dF(x)$$

$$= -\sigma^3 \int_{-c'}^{c'} \left[x\psi(x)-\rho(x)\right] \frac{\delta}{\delta\sigma}\left[\sigma\phi(\sigma x+t)\right] dx + 0,$$

(since ψ vanishes outside $[-c',c']$ and ρ, being constant outside $[-c',c']$, has zero partial derivatives)

$$= -\sigma^3 \int_0^{c'} \left[x\psi(x)-\rho(x)\right] \frac{\delta}{\delta\sigma}\left[\sigma\{\phi(\sigma x+t) + \phi(\sigma x-t)\}\right] dx$$

$$= -\sigma^3 \int_0^{c'} \left[x\psi(x)-\rho(x)\right]\{\sigma\left[-x(\sigma x+t)\phi(\sigma x+t)-x(\sigma x-t)\phi(\sigma x-t)\right]$$

$$+ \phi(\sigma x+t) + \phi(\sigma x-t)\}dx$$

$$= \sigma^3 \int_0^{c'} \left[x\psi(x)-\rho(x)\right]\{\sigma x(\sigma x+t)\phi(\sigma x+t)+\sigma x(\sigma x-t)\phi(\sigma x-t)$$

$$- \left[\phi(\sigma x+t) + \phi(\sigma x-t)\right]\}dx$$

$$= \sigma^3 \int_0^{c'} [x\psi(x)-\rho(x)]\{[(\sigma x)^2-1][\phi(\sigma x+t)+\phi(\sigma x-t)]$$

$$- \sigma xt[\phi(\sigma x-t)-\phi(\sigma x+t)]\}dx$$

$$= \sigma^3\exp(-t^2/2) \int_0^{c'} [x\psi(x)-\rho(x)]\phi(\sigma x)\{[(\sigma x)^2-1][e^{-\sigma xt}+e^{\sigma xt}]$$

$$- \sigma xt[e^{\sigma xt}-e^{-\sigma xt}]\}dx$$

$$= 2\sigma^3\exp(-t^2/2) \int_0^{c'} [x\psi(x)-\rho(x)]\phi(\sigma x)\{[(\sigma x)^2-1] \sum_{j=0}^{\infty} \frac{(\sigma xt)^{2j}}{(2j)!}$$

$$- \sigma xt \sum_{j=0}^{\infty} \frac{(\sigma xt)^{2j+1}}{(2j+1)!}\} dx$$

$$= 2\sigma^3\exp(-t^2/2) \int_0^{c'} [x\psi(x)-\rho(x)]\phi(\sigma x) \sum_{j=0}^{\infty} (\sigma xt)^{2j}\{\frac{(\sigma x)^2-1}{(2j)!} - \frac{(\sigma x)^2 t^2}{(2j+1)!}\}dx,$$

proving (8.32) .

<u>(8.41) REMARK:</u> The expression (8.32) just derived for $E[\Delta^2\psi'(\frac{\Delta}{\sigma})]$ does not make clear at all the range of integration (or portion thereof) for which the expression is positive (or negative.) The expression (8.33) will be more useful for this purpose. Later, we will be interested in comparing $E[\Delta^2\psi'(\frac{\Delta}{\sigma})]$ and $E[\frac{\Delta}{\sigma}\psi(\frac{\Delta}{\sigma})-\rho(\frac{\Delta}{\sigma})]$. We can do this by comparing (8.32) and (8.34) or else (8.33) and (8.35). The latter comparison seems to be

the more fruitful.

To derive (8.33), we write

$$E[\Delta^2\psi'] = -2\sigma^2 E\left[\frac{\Delta}{\sigma}\psi\left(\frac{\Delta}{\sigma}\right)\right] - \sigma^3 \frac{\delta}{\delta t} E\left[\left(\frac{\Delta}{\sigma}\right)^2 \psi\left(\frac{\Delta}{\sigma}\right)\right]$$

(which is easily seen to be true by working out the partial derivative)

$$= -2\sigma^3 \int_{-c'}^{c'} x\psi(x)\phi(\sigma x+t)\,dx - \sigma^4 \frac{\sigma}{\delta t}\int_{-c'}^{c'} x^2\psi(x)\phi(\sigma x+t)\,dx$$

$$= -2\sigma^3 \int_0^{c'} x\psi(x)\left[\phi(\sigma x+t)+\phi(\sigma x-t)\right]dx$$

$$- \sigma^4 \frac{\delta}{\delta t}\int_0^{c'} x^2\psi(x)\left[\phi(\sigma x+t)-\phi(\sigma x-t)\right]dx$$

$$= -2\sigma^3 \int^{c'} x\psi(x)\left[\phi(\sigma x+t)+\phi(\sigma x-t)\,dx\right.$$

$$- \sigma^4 \int_0^{c'} x^2\psi(x)\left[-(\sigma x+t)\phi(\sigma x+t)-(\sigma x-t)\phi(\sigma x-t)\right]dx$$

$$= \sigma^4 \int_0^{c'} x^2\psi(x)\left[(\sigma x+t)\phi(\sigma x+t)+(\sigma x-t)\phi(\sigma x-t)\right]dx$$

$$- 2\sigma^3 \int_0^{c'} x\psi(x)\left[\phi(\sigma x+t)+\phi(\sigma x-t)\right]dx$$

$$= \sigma^4 \int_0^{c'} x^2\psi(x)\left[\sigma x\{\phi(\sigma x+t)+\phi(\sigma x-t)\} - t\{\phi(\sigma x-t)-\phi(\sigma x+t)\}\right]dx$$

$$- 2\sigma^3 \int_0^{c'} x\psi(x)\left[\phi(\sigma x+t)+\phi(\sigma x-t)\right]dx$$

$$= \sigma^2 \int_0^{c'} \sigma x\psi(x)\left[(\sigma^2x^2-2)(\phi(\sigma x+t)+\phi(\sigma x-t))-\sigma xt(\phi(\sigma x-t)-\phi(\sigma x+t))\right]dx$$

$$= \sigma^2\exp(-t^2/2)\int_0^{c'} \sigma x\psi(x)\phi(\sigma x)\{\left[\sigma^2x^2-2\right]\left[e^{-\sigma xt}+e^{\sigma xt}\right]-\sigma xt\left[e^{\sigma xt}-e^{-\sigma xt}\right]\}dx$$

$$= 2\sigma^2\exp(-t^2/2)\int_0^{c'} \sigma x\psi(x)\phi(\sigma x)\{\left[\sigma^2x^2-2\right]\sum_{j=0}^{\infty}\frac{(\sigma xt)^{2j}}{(2j)!} - \sigma xt\sum\frac{(\sigma xt)^{2j+1}}{(2j+1)!}\}dx$$

$$= 2\sigma^2\exp(-t^2/2)\int_0^{c'} \sigma x\psi(x)\phi(\sigma x)\sum_{j=0}^{\infty}(\sigma xt)^{2j}\left[\frac{\sigma^2x^2-2}{(2j)!} - \frac{\sigma^2x^2t^2}{(2j+1)!}\right]dx\ ,$$

proving (8.33) .

As a further check (see also (8.37)) on (8.33), we derive the same expression for $E[\Delta^2\psi']$ as follows:

$$E\left[\Delta^2\psi'\left(\frac{\Delta}{\sigma}\right)\right] = -\sigma^2\frac{\delta}{\delta\sigma}E\left[\Delta\psi\left(\frac{\Delta}{\sigma}\right)\right] = -\sigma^2\frac{\delta}{\delta\sigma}E\left[\sigma\left(\frac{\Delta}{\sigma}\right)\psi\left(\frac{\Delta}{\sigma}\right)\right]$$

$$= -\sigma^2\frac{\partial}{\partial\sigma}\int_{-c'}^{c'}\sigma^2x\psi(x)\phi(\sigma x+t)dx$$

$$= -\sigma^2\int_{-c'}^{c'}x\psi(x)\{-\sigma^2x(\sigma x+t)\phi(\sigma x+t) + 2\sigma\phi(\sigma x+t)\}dx$$

$$= -\sigma^2\int_{-c'}^{c'}x\psi(x)\{-\sigma^3x^2\phi(\sigma x+t)-\sigma^2xt\phi(\sigma x+t)+2\sigma\phi(\sigma x+t)\}dx$$

$$= -\sigma^2\int_0^{c'}x\psi(x)\{-\sigma^3x^2\left[\phi(\sigma x+t)+\phi(\sigma x-t)\right]-\sigma^2xt\left[\phi(\sigma x+t)-\phi(\sigma x-t)\right]$$

$$+ 2\sigma\left[\phi(\sigma x+t)+\phi(\sigma x-t)\right]\}dx$$

$$= -\sigma^2 \int_0^{c'} x\psi(x)\{[-\sigma^3x^2+2\sigma][\phi(\sigma x+t)+\phi(\sigma x-t)]-\sigma^2xt[\phi(\sigma x+t)$$

$$-\phi(\sigma x-t)]\}dx$$

$$= -\sigma^2\exp(-t^2/2) \int_0^{c'} x\psi(x)\{[-\sigma^3x^2+2\sigma][e^{-\sigma xt}+e^{\sigma xt}]-\sigma^2xt[e^{-\sigma xt}-e^{\sigma xt}]\}dx$$

$$= -2\sigma^3\exp(-t^2/2) \int_0^{c'} x\psi(x)\{[-\sigma^2x^2+2] \sum_{j=0}^{\infty}[\frac{(\sigma xt)^{2j}}{(2j)!} + \sigma xt \sum_{j=0}^{\infty} \frac{(\sigma xt)^{2j+1}}{(2j+1)!}]dx$$

$$= 2\sigma^3\exp(-t^2/2) \int_0^{c'} x\psi(x) \sum_{j=0}^{\infty} (\sigma xt)^{2j}[\frac{(\sigma x)^2-2}{(2j)!} - \frac{(\sigma x)^2t^2}{(2j+1)!}]dx ,$$

returning us to (8.33) again.

This completes the proof of (8.24). □

Before proceeding, we equip ourselves with a table of parameter values and some functions of interest.

All entries are rounded to four decimal places.

(8.42) TABLE:

α	.001	.01	.05
$d=\Phi^{-1}(1-\frac{\alpha}{2})$	3.2905	2.5758	1.9600
$k=\Phi^{-1}(\frac{1}{2}+\frac{\alpha}{2})$	.0012	.0125	.0627
$b=[\Phi^{-1}(1-\frac{\alpha}{2})-\Phi^{-1}(\frac{3\alpha}{2})]/[\Phi^{-1}(1-\alpha)-\Phi^{-1}(\alpha)]$	1.0126	1.0200	1.0334
$[b-1]/k$	10.5000	1.6000	.5327
$2b-1$	1.0252	1.0400	1.0668
$-2b+3$	.9748	.9600	.9332
$\cosh(.9k)/[1-k^2]$	1.0000	1.0003	1.0055

<u>(8.43) DEFINITION:</u> Let $\mathcal{D}$ be the region in R^{p+1} defined by (see (8.119) for comments)

$$\mathcal{D} = \left\{ \begin{pmatrix} \tilde{t} \\ s \end{pmatrix} \in R^{p+1} : |s-1| + \frac{2(b-1)}{k} \max_{1\le i\le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| < 2(b-1) \right\} . \tag{8.44}$$

Note that this region is "well-defined" even when $k = 0$. Clearly, $\begin{pmatrix} \tilde{t} \\ s \end{pmatrix} \in \mathcal{D}$ implies (but not conversely)

$$\left| \sum_{j=1}^{p} a_{ij} t_j \right| < k , \qquad i=1,\ldots,p \tag{8.45}$$

and

$$|s-1| < 2(b-1) \tag{8.46}$$

or, equivalently,

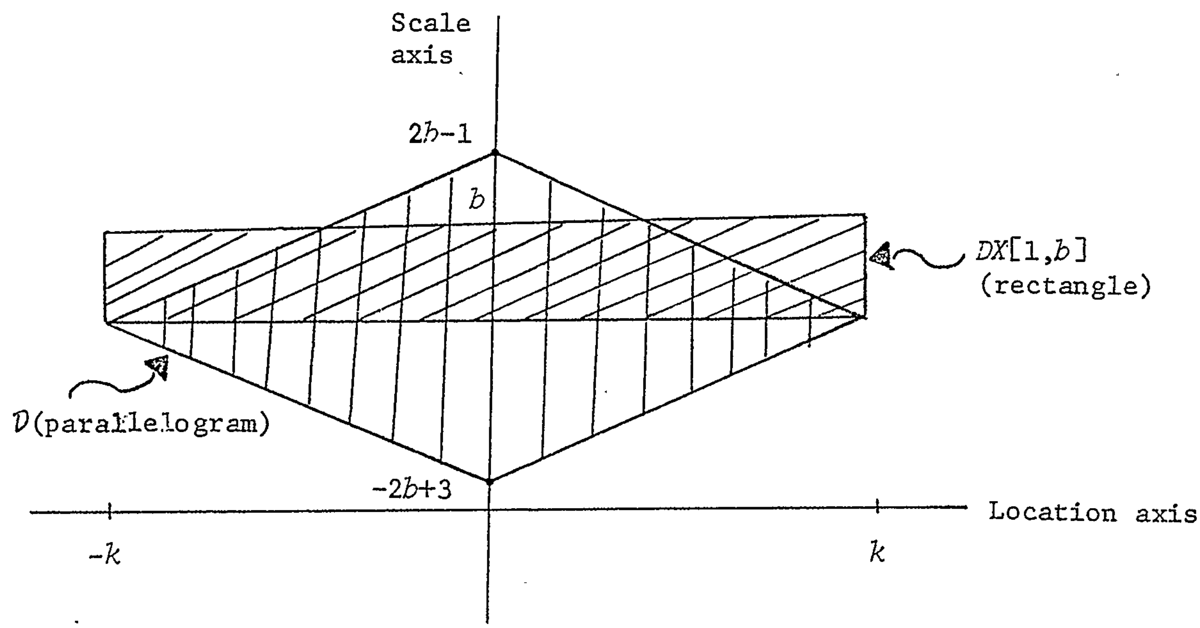
$$-2b+3 < s < 2b-1 . \tag{8.47}$$

Now define a norm on R^{p+1} as follows:

$$\left\| \begin{pmatrix} \tilde{t} \\ s \end{pmatrix} \right\| = |s| + \frac{2(b-1)}{k} \max_{1\le i\le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| . \tag{8.48}$$

Throughout the remainder of this section, this norm is the one understood to be in use.

(8.48) Diagram showing $\mathcal{D}$ (defined in (8.43)) and $DX[1,b)$ (see (8.23) and recall that $\begin{pmatrix}\theta^{**}\\ \beta\end{pmatrix} \in D \times [1,b))$ for the case $p = 1$ and, for convenience



<u>(8.49) THEOREM:</u> We have $H\begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix} = \underset{\sim}{0}$. Also $H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}$ is non-singular for all $\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} \in \mathcal{D}$. Finally, the Newton iterates

(8.50) $$\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{k+1} = \begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{k} - \left(H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{k}\right)^{-1} H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{k}, \quad k = 0,1,2,\ldots$$

with starting value $\begin{pmatrix}\underset{\sim}{\theta}^{**}\\ \beta\end{pmatrix}$ all lie in $\mathcal{D}$ and converge to $\begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix}$. Note that the starting value $\begin{pmatrix}\underset{\sim}{\theta}^{**}\\ \beta\end{pmatrix}$ may not lie in $\mathcal{D}$! (see the diagram in (8.49).)

Proof of (8.49):

We have, from (8.15),

$$H\binom{\underset{\sim}{t}}{\sigma} = \begin{bmatrix} \sum_{i=1}^{p} a_{i1}q_i E\psi\left(\frac{\Delta_i}{\sigma}\right) \\ \vdots \\ \sum_{i=1}^{p} a_{ip}q_i E\psi\left(\frac{\Delta_i}{\sigma}\right) \\ \sum_{i=1}^{p} q_i E\left[\frac{\Delta_i}{\sigma}\,\psi\left(\frac{\Delta_i}{\sigma}\right) - \rho\left(\frac{\Delta_i}{\sigma}\right)\right] - a \end{bmatrix}.$$

We first check the non-singularity of $H'\binom{\underset{\sim}{t}}{\sigma}$.

(8.52) $H'\binom{\underset{\sim}{t}}{\sigma} =$

$$-\begin{bmatrix} \sum_{i=1}^{p} a_{i1}a_{i1}\frac{q_i}{\sigma}E\psi'\left(\frac{\Delta_i}{\sigma}\right), \ldots\ldots\ldots \sum_{i=1}^{p} a_{i1}a_{ip}\frac{q_i}{\sigma}E\psi'\left(\frac{\Delta_i}{\sigma}\right), \sum_{i=1}^{p} a_i\frac{q_i}{\sigma^2}E\left[\Delta_i\psi'\left(\frac{\Delta_i}{\sigma}\right)\right] \\ \vdots \qquad\qquad \vdots \\ \sum_{i=1}^{p} a_{ip}a_{i1}\frac{q_i}{\sigma}E\psi'\left(\frac{\Delta_i}{\sigma}\right), \ldots\ldots\ldots \sum_{i=1}^{p} a_{ip}a_{ip}\frac{q_i}{\sigma}E\psi'\left(\frac{\Delta_i}{\sigma}\right), \sum_{i=1}^{p} a_{ip}\frac{q_i}{\sigma^2}E\left[\Delta_i\psi'\left(\frac{\Delta_i}{\sigma}\right)\right] \\ \sum_{i=1}^{p} q_iE\left[\frac{\Delta_i}{\sigma}\frac{a_{i1}}{\sigma}\psi'\left(\frac{\Delta_i}{\sigma}\right) + \frac{a_{i1}}{\sigma}\psi\left(\frac{\Delta_i}{\sigma}\right) - \frac{a_{i1}}{\sigma}\psi\left(\frac{\Delta_i}{\sigma}\right)\right], \ldots \sum_{i=1}^{p} q_iE\left[\frac{\Delta_i}{\sigma}\frac{a_{ip}}{\sigma}\psi'\left(\frac{\Delta_i}{\sigma}\right) \right. \\ \left. \qquad + \frac{a_{ip}}{\sigma}\psi\left(\frac{\Delta_i}{\sigma}\right) - \frac{a_{ip}}{\sigma}\psi\left(\frac{\Delta_i}{\sigma}\right)\right], \\ \sum_{i=1}^{p} q_iE\left[\frac{\Delta_i^2}{\sigma^2}\psi'\left(\frac{\Delta_i}{\sigma}\right) + \frac{\Delta_i}{\sigma^2}\psi\left(\frac{\Delta_i}{\sigma}\right) - \frac{\Delta_i}{\sigma^2}\psi\left(\frac{\Delta_i}{\sigma}\right)\right] \end{bmatrix}$$

or

(8.53) $H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} =$

$$
-\begin{bmatrix}
\sum_{i=1}^{p} a_{i1}a_{i1}\frac{q_i}{\sigma}E\psi_i', \ldots\ldots\ldots & \sum_{i=1}^{p} a_{i1}a_{ip}\frac{q_i}{\sigma}E\psi_i', & \sum_{i=1}^{p} a_{i1}\frac{q_i}{\sigma^2}E[\Delta_i\psi_i'] \\
\vdots & \vdots & \\
\sum_{i=1}^{p} a_{ip}a_{i1}\frac{q_i}{\sigma}E\psi_i', \ldots\ldots\ldots & \sum_{i=1}^{p} a_{ip}a_{ip}\frac{q_i}{\sigma}E\psi_i', & \sum_{i=1}^{p} a_{ip}\frac{q_i}{\sigma^2}E[\Delta_i\psi_i'] \\
\sum_{i=1}^{p} a_{i1}\frac{q_i}{\sigma^2}E[\Delta_i\psi_i'], \ldots\ldots\ldots & \sum_{i=1}^{p} a_{ip}\frac{q_i}{\sigma^2}E[\Delta_i\psi_i'], & \sum_{i=1}^{p}\frac{q_i}{\sigma^3}E[\Delta_i^2\psi_i']
\end{bmatrix},
$$

where, for brevity, we have written ψ_i' in place of $\psi'\left(\frac{\Delta_i}{\sigma}\right)$. We also shall use the notation ψ_i for $\psi\left(\frac{\Delta_i}{\sigma}\right)$ and ρ_i for $\rho\left(\frac{\Delta_i}{\sigma}\right)$, $i=1,\ldots,p$. Notice that $H'\binom{t}{\sigma}$ is symmetric. It has the following factorization as the product of three matrices:

(8.54) $$H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} = -A^T BA$$

where

(8.55) $$A = \begin{bmatrix}
a_{11}, & a_{12}, & \cdots & a_{1p}, & \frac{E\Delta_1\psi_1'}{\sigma E\psi_1'} \\
 & & \vdots & & \\
a_{p1}, & a_{p2}, & \cdots & a_{pp}, & \frac{E[\Delta_p\psi_p']}{\sigma E\psi_p'} \\
0, & 0, & \ldots & 0, & 1
\end{bmatrix}$$

and

$$(8.56)\quad B = \begin{bmatrix} \frac{q_1}{\sigma}E\psi_1' & & & & & \\ & \frac{q_2}{\sigma}E\psi_2' & & & \mathit{0} & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & \frac{q_p}{\sigma}E\psi_p' & \\ & \mathit{0} & & & & \sum_{i=1}^{p} \frac{q_i}{\sigma^3}\left\{E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right\} \end{bmatrix}$$

and where $\mathit{0}$ denotes a triangular block of zeros (so that B is diagonal.)

Consider now the invertibility of $H'\left(\frac{\tilde{t}}{\sigma}\right)$. We have

$$\det A^T = \det A = \det A,$$

clearly, and since $\det A \neq 0$ by assumption in our model, we are left with checking the invertibility of B. Now also $q_i \neq 0$ by assumption and we have, from (8.30),

$$E\psi_i' = 2\sigma^2\exp(-t^2/2)\int_0^{c'}\psi(x)\phi(\sigma x)\sigma x \sum_{j=0}^{\infty}(\sigma x t)^{2j}\left(\frac{1}{(2j)!} - \frac{t^2}{(2j+1)!}\right)dx\ ,$$

where we have written t for any of the $\sum_{j=1}^{p} a_{ij}t_j$.

Hence,

$$E\psi_i' \geq \left[2\sigma^2\exp(-t^2/2)\int_0^{c'}\psi(x)\phi(\sigma x)\sigma x dx\right]\left[1-t^2\right]\ .$$

Thus $E\psi_i' > 0$ because $\psi \not\equiv 0$ on $[0,c']$ (see (8.4)) and

$$\begin{aligned} 1-t^2 &= 1-\left(\sum_{j=1}^{p} a_{ij}t_j\right)^2 \\ &\geq 1-k^2 \qquad \text{(by (8.45))} \\ &> 0 \qquad \text{(by (8.42))} \end{aligned}$$

and furthermore $\sigma > 0$ by (8.47) and (8.42). This leaves us with checking if we can invert the last entry of $\mathcal{B}$. Now $E\psi_i' > 0$ by work just done and since $E^2[\Delta_i\psi_i'] \geq 0$, we have

$$-\sum_{i=1}^{p} q_i \left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'} \right) \geq -\sum_{i=1}^{p} q_i E[\Delta_i^2\psi_i']$$

and it suffices to show $-E[\Delta_i^2\psi_i'] > 0$, $i=1,\ldots,p$. Again writing t for $\sum_{j=1}^{p} a_{ij}t_j$, we have from (8.33),

$$E[\Delta_i^2\psi_i'] = 2\sigma^2\exp(-t^2/2)\int_0^{c'} \psi(x)\phi(\sigma x)\sigma x \sum_{j=0}^{\infty} (\sigma x t)^{2j}\left(\frac{\sigma^2x^2-2}{(2j)!} - \frac{\sigma^2x^2t^2}{(2j+1)!}\right)dx$$

and so

$$-E[\Delta_i^2\psi_i'] > 2\sigma^2\exp(-t^2/2)\int_0^{c'} \psi(x)\phi(\sigma x)\sigma x \sum_{j=0}^{\infty} (\sigma x t)^{2j}\left(\frac{2-\sigma^2x^2}{(2j)!}\right) dx$$

and it will suffice to show that

(8.57) $\quad 2-\sigma^2x^2 > 0 \quad$ for all $x \in [0,c']$ and $\sigma \in (-2b+3,2b-1)$ (see 8.47).

(8.57) is equivalent to $\sigma^2x^2 < 2$ and this is implied by $\sigma^2(c')^2 < 2$ which in turn is true if $((2b-1)c')^2 < 2$ (since $(-2b+3)^3 < (2b-1)^2$ —see (8.42)) or $c^2 < 2$ (see (8.3)) and this is true by (8.2). (Of course, $c^2 < 2$ is true for $c < 2^{\frac{1}{2}}$ but the restriction on c in (8.2) was chosen much sharper than needed here, for bounds computed later on.)

We may now proceed and invert our matrix $H'\left(\underset{\sim}{t}\atop\sigma\right)$.

We have

$$A = \begin{bmatrix} a_{11} & \cdots\cdots & a_{1p}, & E[\Delta_1\psi_1']/(\sigma E\psi_1') \\ \cdot & \cdots\cdots & \cdot & \cdot \\ \cdot & \cdots\cdots & \cdot & \cdot \\ \cdot & \cdots\cdots & \cdot & \cdot \\ a_{p1} & \cdots\cdots & a_{pp}, & E[\Delta p\psi_p']/(\sigma E\psi_p') \\ 0, & \cdots & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} p\times p & p\times 1 \\ A & \underset{\sim}{b} \\ \underset{\sim}{0}^T & 1 \\ 1\times p & 1\times 1 \end{bmatrix} \quad \text{, say}$$

where

(8.59) $$\underset{\sim}{b} = \left(\frac{E[\Delta_1\psi_1']}{\sigma E\psi_1'}, \ldots \frac{E[\Delta_p\psi_p']}{\sigma E\psi_p'}\right)^T$$

so that

(8.60) $$A^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}\underset{\sim}{b} \\ \\ \underset{\sim}{0}^T & 1 \end{bmatrix} .$$

Then,

(8.61) $$(A^T)^{-1} = (A^{-1})^T = \begin{bmatrix} (A^T)^{-1} & \underset{\sim}{0} \\ \\ -\underset{\sim}{b}^T(A^T)^{-1} & 1 \end{bmatrix}$$

and so $H'\binom{\underset{\sim}{t}}{\sigma}^{-1} = (-A^T BA)^{-1} = -A^{-1}B^{-1}(A^T)^{-1}$ becomes

(8.62)

$$H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1} = -\begin{pmatrix} A^{-1} & -A^{-1}\underset{\sim}{b} \\ \underset{\sim}{0}^T & 1 \end{pmatrix} \begin{pmatrix} \frac{\sigma}{q_1 E\psi_1'} & & & 0 \\ & \ddots & & \\ & & \frac{\sigma}{q_p E\psi_p'} & \\ 0 & & & \frac{1}{\sum_{i=1}^{p} \frac{q_i}{\sigma^2}\left[E\Delta_i^2\psi_i' - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right]} \end{pmatrix} \begin{pmatrix} (A^T)^{-1} & \underset{\sim}{0} \\ -\underset{\sim}{b}^T(A^T)^{-1} & 1 \end{pmatrix}$$

Now also

$$H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} = \begin{bmatrix} \sum_{i=1}^{p} a_{i1}q_i E\psi_i \\ \vdots \\ \sum_{i=1}^{p} a_{ip}q_i E\psi_i \\ \sum_{i=1}^{p} q_i\left(E\left[\frac{\Delta_i}{\sigma}\psi_i\right] - E\rho_i\right) - a \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{p1} & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ a_{1p} & a_{2p} & \cdots & a_{pp} & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} q_1 E\psi_1 \\ \vdots \\ q_p E\psi_p \\ \sum_{i=1}^{p} q_i\left(E\left[\frac{\Delta_i}{\sigma}\psi_i\right] - E\rho_i\right) - a \end{bmatrix}$$

or

$$(8.63)\qquad H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} = \begin{pmatrix} A^T & \underset{\sim}{0} \\ \underset{\sim}{0}^T & 1 \end{pmatrix}\begin{pmatrix} q_1 E\psi_1 \\ \vdots \\ q_p E\psi_p \\ \sum_{i=1}^{p} q_i \left(E\left[\frac{\Delta_i}{\sigma}\,\psi_i\right] - E\rho_i\right) - a \end{pmatrix}.$$

From (8.63) and (8.62) we have

$$H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1} H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} =$$

$$-\begin{pmatrix} A^{-1} & -A^{-1}\underset{\sim}{b} \\ \underset{\sim}{0} & 1 \end{pmatrix}\begin{pmatrix} \frac{\sigma}{q_1 E\psi_1'} & & & 0 \\ & \ddots & & \\ & & \frac{\sigma}{q_p E\psi_p'} & \\ 0 & & & \frac{1}{\sum_{i=1}^{p} \frac{q_i}{\sigma^3}\left\{E\left[\Delta_i^2\psi_i'\right] - \frac{E^2\left[\Delta_i\psi_i'\right]}{E\psi_i'}\right\}} \end{pmatrix}\begin{pmatrix} (A^T)^{-1} & \underset{\sim}{0} \\ -\underset{\sim}{b}^T (A^T)^{-1} & 1 \end{pmatrix}\begin{pmatrix} A^T & \underset{\sim}{0} \\ \underset{\sim}{0}^T & 1 \end{pmatrix} \text{ times}$$

$$\begin{pmatrix} q_1 E\psi_1 \\ \vdots \\ q_p E\psi_p \\ \sum_{i=1}^{p} q_i \left(E\left[\frac{\Delta_i}{\sigma}\,\psi_i\right] - E\rho_i\right) - a \end{pmatrix}$$

$$= -A^{-1}\begin{pmatrix} \frac{\sigma}{q_1 E\psi_1'} & & & 0 \\ & \ddots & & \\ & & \frac{\sigma}{q_p E\psi_p'} & \\ 0 & & & \frac{1}{\sum_{i=1}^{p}\frac{q_i}{\sigma^3}\left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i^2\psi_i']}{E\psi_i'}\right)} \end{pmatrix}\begin{pmatrix} I & \underset{\sim}{0} \\ -\underset{\sim}{b}^T & 1 \end{pmatrix}\begin{pmatrix} q_1 E\psi_1 \\ \vdots \\ q_p E\psi_p \\ \sum_i^p q_i\left(E[\frac{\Delta_i}{\sigma}\psi_i] - E\rho_i\right) - a \end{pmatrix}$$

(where I is the $p\times p$ unit matrix), that is,

$$(8.64)\quad H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1} H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} =$$

$$-A^{-1}\begin{pmatrix} \frac{\sigma}{q_1 E\psi_1'} & & & 0 \\ & \ddots & & \\ & & \frac{\sigma}{q_p E\psi_p'} & \\ 0 & & & \frac{1}{\sum_{i=1}^{p}\frac{q_i}{\sigma^3}\left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right)} \end{pmatrix}\begin{pmatrix} q_1 E\psi_1 \\ \vdots \\ q_p E\psi_p \\ -\sum_i^p q_i E\psi_i b_i + \sum_{i=1}^{p} q_i E[\frac{i}{\sigma}\psi_i - \rho_i] - a \end{pmatrix}$$

where b_i is the ith component of $\underset{\sim}{b}$, i.e.

$$\underset{\sim}{b} = \left(\frac{E[\Delta_1\psi_1']}{\sigma\ \psi_1'}, \ldots, \frac{E[\Delta_p\psi_p']}{\sigma\ \psi_p'}\right)^T \qquad \text{(see (8.59))}$$

$$= (b_1, \ldots, b_p)^T .$$

(8.64) equals

$$-A^{-1}\begin{pmatrix} \sigma\,\dfrac{E\psi_1}{E\psi_1'} \\ \cdot \\ \cdot \\ \cdot \\ \sigma\,\dfrac{E\psi_p}{E\psi_p'} \\ \dfrac{\sum\limits_{i=1}^{p} q_i E\left[\frac{\Delta_i}{\sigma}\psi_i - \rho_i - \psi_i b_i\right] - a}{\sum\limits_{i=1}^{p} \frac{q_i}{\sigma^3}\left[E(\Delta_i^2\psi_i') - \frac{E^2(\Delta_i\psi_i')}{E\psi_i'}\right]} \end{pmatrix}$$

i.e.

(8.65)

$$H'\begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}^{-1} H\begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix} = A^{-1}\begin{pmatrix} -\dfrac{E\psi_1}{E\psi_1'} \\ \cdot \\ \cdot \\ \cdot \\ -\dfrac{E\psi_p}{E\psi_p'} \\ \dfrac{\sum\limits_{i=1}^{p} q_i E\left[\rho_i + \psi_i\,\frac{E(\Delta_i\psi_i')}{\sigma E\psi_i'} - \frac{\Delta_i}{\sigma}\,\psi_i\right] + a}{\sum\limits_{i=1}^{p} \frac{q_i}{\sigma^3}\left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right)} \end{pmatrix}$$

(compare with (4.24).)

We now check that

(8.66) $$H\begin{pmatrix} \underset{\sim}{0} \\ 1 \end{pmatrix} = \underset{\sim}{0} \,.$$

Since

$$H\begin{pmatrix}\underset{\sim}{t} \\ \sigma\end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{p} a_{i1} q_i E\psi\left(\frac{\Delta_i}{\sigma}\right) \\ \vdots \\ \sum_{i=1}^{p} a_{ip} q_i E\psi\left(\frac{\Delta_i}{\sigma}\right) \\ \sum_{i=1}^{p} q_i E\left[\frac{\Delta_i}{\sigma}\,\psi\left(\frac{\Delta_i}{\sigma}\right) - \rho\left(\frac{\Delta_i}{\sigma}\right)\right] - a \end{pmatrix},$$

we have

$$H\begin{pmatrix}\underset{\sim}{0} \\ 1\end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{p} a_{i1} q_i E\psi(X) \\ \vdots \\ \sum_{i=1}^{p} a_{ip} q_i E\psi(X) \\ \sum_{i=1}^{p} q_i E[X\psi(X) - \rho(X)] - a \end{pmatrix} .$$

The first p co-ordinates of $H\binom{\underset{\sim}{0}}{1}$ are thus zero by (4.13a)) and it remains to show that

(8.67) $$\sum_{i=1}^{p} q_i E\left[X\psi(X)-\rho(X)\right] - a = 0 .$$

We have

$$\sum_{i=1}^{p} q_i E\left[X\psi(X)-\rho(X)\right] - a$$

$$= E\left[X\psi(X)-\rho(X)\right] - a$$

$$= E[X\psi(X)-\rho(X)] - E[U\psi(U)-\rho(U)]$$

(see (8.6) and (8.9))

$$= E[X\psi(X)] - E[U\psi(U)] + E\rho(U) - E\rho(X)$$

$$= 2\int_0^{c'} x\psi(x)\phi(x)dx - 2\int_0^{c'} x\psi(x)\phi(x)dx$$

$$+ \int_0^{c'} \psi(x)[\Phi(x)+\Phi(x)-2\Phi(x)]\,dx,$$

(from (8.28), with $\sigma = 1$ and $t = 0$, (8.26) and (8.36))

$$= 0 ,$$ proving (8.67).

This proves (8.66).

Before proving the last assertion of (8.49), namely that the iterates lie in $\mathcal{D}$ and converge to $\binom{0}{1}$, we will need several lemmas. For reasons of brevity a great deal of analysis of the functions involved below will be omitted. It is unfortunate that some of the lemmas rely on numerical techniques for their "proofs" — where this occurs it is almost an understatement to say that the functions involved are extremely difficult to treat mathematically.

In these lemmas, t will denote any of the $\sum_{j=1}^{p} a_{ij}t_j$, so that, from (8.45) we have $|t| < k$. Also, the points $t = 0$ and $\sigma = 1$ will be omitted since if $t = 0$ and $\sigma = 1$, the inequalities appearing will become trivial equalities. Finally, frequent appeal is made to (8.24) for expressions needed.

(8.68) LEMMA:

$$\left| t - \left(\frac{-E\psi\sigma}{E\psi'} \right) \right| < .0055|t| \quad .$$

Proof: We have

$$-E\psi = 2\sigma \exp(-t^2/2) \int_0^{c'} \psi(x)\phi(\sigma x)\sigma x t \sum_{j=0}^{\infty} \frac{(\sigma x t)^{2j}}{(2j+1)!}\, dx \quad .$$

Then, for $t > 0$,

$$(8.69) \qquad -E\psi > 2\sigma^2 t \exp(-t^2/2)(1) \int_0^{c'} x\psi(x)\phi(\sigma x)\,dx$$

and

$$(8.70) \qquad -E\psi < 2\sigma^2 t \exp(-t^2/2) \cosh[2(b-1)c't] \int_0^{c'} x\psi(x)\phi(\sigma x)\,dx$$

using the facts that

$$\sum_{j=0}^{\infty} \frac{(\sigma x t)^{2j}}{(2j+1)!} \le \cosh[\sigma x t]$$

and

$$\sigma < 2b-1 \qquad \text{(see (8.47).)}$$

(8.70) is the same as

$$(8.71) \qquad -E\psi < 2\sigma^2 t \exp(-t^2/2) \cosh[ct] \int_0^{c'} x\psi(x)\phi(\sigma x)\,dx \qquad \text{(see (8.3).)}$$

Also,

$$E\psi' = 2\sigma^2 \exp(-t^2/2) \int_0^{c'} \psi(x)\phi(\sigma x)\sigma x \sum_{j=0}^{\infty} (\sigma x t)^{2j} \left(\frac{1}{(2j)!} - \frac{t^2}{(2j+1)!} \right) dx$$

and so,

(8.72) $$E\psi' > (1-t^2)2\sigma^3\exp(-t^2/2)\int_0^{c'} x\psi(x)\phi(\sigma x)dx$$

and

(8.73) $$E\psi' < (\cosh t-t^2)2\sigma^3\exp(-t^2/2)\int_0^{c'} x\psi(x)\phi(\sigma x)dx \quad .$$

From (8.69), (8.71), (8.72) and (8.73) we get:

$t > 0$ implies

(8.74) $$\frac{-E\psi\sigma}{E\psi'} > \frac{t}{\cosh t-t^2}$$

and

(8.75) $$\frac{-E\psi\sigma}{E\psi'} < t\frac{\cosh(ct)}{1-t^2} \quad .$$

Now $\cosh t-t^2$ is decreasing in t for t less than (about) 2.1 because its derivative $\sinh t-2t$ is negative for $t <$ (about) 2.1. (Note that we have $t < k$ so that $t < .0627$, from (8.42).) Thus $\cosh t-t^2 < \cosh 0 - (0)^2 = 1$ and (8.74) then gives

(8.76) $$\frac{-E\psi\sigma}{E\psi'} > t \quad .$$

Also

$$\frac{\cosh(ct)}{1-t^2} < \frac{\cosh(ck)}{1-k^2} \leq \frac{\cosh(.9k)}{1-k^2} \qquad \text{(see (8.2))}$$

$$< 1.0055 \qquad \text{(see (8.42))}$$

and so from (8.75) we have

(8.77) $$\frac{-E\psi\sigma}{E\psi'} < (1.0055)t \ .$$

From (8.76), (8.77) and corresponding inequalities when $t < 0$, we get

(8.78) $$|t| < \left|\frac{-E\psi\sigma}{E\psi'}\right| < 1.0055|t| \ .$$

Since, clearly, $E\psi' > 0$ and $-E\psi$ retains the same sign as t, we get from (8.78)

$$\left|t - \left(\frac{-E\psi\sigma}{E\psi'}\right)\right| < .0055|t| \ ,$$

proving (8.68). □

(8.79) LEMMA:

$$1.19\,\frac{|t|}{\sigma} < \left|\frac{E|\Delta\psi'|}{\sigma E\psi'}\right| < 2(1.0055)\,\frac{|t|}{\sigma} \ .$$

Proof: We have

$$E[\Delta\psi'] = 2\sigma^2\exp(-t^2/2)\int_0^{c'}\psi(x)\phi(\sigma x)\sigma xt\sum_{j=0}^{\infty}(\sigma xt)^{2j}\left(\frac{1}{(2j)!}+\frac{1-(\sigma x)^2}{(2j+1)!}\right)dx$$

and, taking $t > 0$ first, we get

(8.80) $$E[\Delta\psi'] > \left(\int_0^{c'}\psi(x)\phi(\sigma x)x\,dx\right)t(2)\sigma^3\exp(-t^2/2)(2-c^2)$$

because

$$\sum_{j=0}^{\infty} (\sigma x t)^{2j} \left[\frac{1}{(2j)!} + \frac{1-(\sigma x)^2}{(2j+1)!}\right]$$

$$= \cosh(\sigma x t) + \frac{1-(\sigma x)^2}{\sigma x t}\sinh(\sigma x t)$$

$$> 1 + 1 - [2(b-1)c']^2 \frac{\sinh(\sigma x t)}{\sigma x t}$$

$$> 1 + 1 - [2(b-1)c']^2$$

$$= 2 - c^2 \qquad \text{(from (8.3).)}$$

Further,

$$(8.81) \qquad E[\Delta\psi'] <$$

$$\left(\int_0^{c'} \psi(x)\phi(\sigma x)x dx\right) t(2)\sigma^3 \exp(-t^2/2)\left[\cosh(ct) + \cosh(ct)\right] ,$$

since

$$[1-(\sigma x)^2] \frac{\sinh(\sigma x t)}{\sigma x t}$$

$$< [1-(\sigma x)^2] \cosh(\sigma x t)$$

$$< [1-c^2] \cosh(\sigma x t)$$

$$< \cosh(\sigma x t) \qquad \text{since } c \leq .9 \qquad \text{(see (8.2).)}$$

Then, for $t > 0$, we get from (8.80) and (8.81),

$$\frac{t[2-c^2]}{\sigma(\cosh t - t^2)} < \frac{E[\Delta\psi']}{\sigma E\psi'} < \frac{2t\cosh(ct)}{\sigma(1-t^2)}$$

which implies

(8.82) $$(2-c^2)\ \frac{t}{\sigma} < \frac{E[\Delta\psi']}{\sigma E\psi'} < 2(1.0055)\ \frac{t}{\sigma}\ ,$$

using again $\cosh t - t^2 < 1$ (in our range of t), and using (8.42) for the upper bound.

From (8.82) and corresponding inequalities when $t < 0$, we have

$$(2-c^2)\ \frac{|t|}{\sigma} < \left|\frac{E[\Delta\psi']}{E\psi'}\right| < 2(1.0055)\ \frac{|t|}{\sigma}$$

which implies, since $c \leq .9$,

$$1.19\ \frac{|t|}{\sigma} < \left|\frac{E[\Delta\psi']}{E\psi'}\right| < 2(1.0055)\ \frac{|t|}{\sigma}\ ,$$

proving (8.79).

(8.83) LEMMA:

$$0 < \frac{2\sigma t^2 \exp(-t^2/2)}{\cosh t - t^2} \left(\int_0^{c'} x\psi(x)\,\phi(\sigma x)\,dx\right) < \frac{-E\psi E[\Delta\psi']}{\sigma E\psi'}$$

$$< 4(1.0055)\sigma t^2 \exp(-t^2/2)\cosh(ct)\left(\int_0^{c'} x\psi(x)\,\phi(\sigma x)\,dx\right)\ .$$

Proof: Note that the first inequality above is trivial. Since, from the expressions for $E\psi$, $E[\Delta\psi']$ and $E\psi'$ in (8.24), it is clear that $\frac{-E\psi E[\Delta\psi']}{\sigma E\psi'} > 0$ for all $t \neq 0$, and is symmetric in t, we may restrict ourselves to $t > 0$.

We have

(8.84) $$\frac{-E\psi E[\Delta\psi']}{\sigma E\psi'} = \left(\frac{-E\psi\sigma}{E\psi'}\right)\left(\frac{E[\Delta\psi']}{\sigma^2}\right)$$

$$< 1.0055t\left(\frac{E[\Delta\psi']}{\sigma^2}\right) \qquad \text{(from (8.77))}$$

and

(8.85) $$\frac{-E\psi E[\Delta\psi']}{\sigma E\psi'} > \frac{t}{\cosh t - t^2}\left(\frac{E[\Delta\psi']}{\sigma^2}\right) \qquad \text{(from 8.74).)}$$

From (8.84), (8.85), (8.80) and (8.81) we have

$$\left(\frac{t}{\cosh - t^2}\right) t(2)\sigma\exp(-t^2/2)(2-c^2)\left(\int_0^{c'} x\psi(x)\phi(\sigma x)\,dx\right) < \frac{-E\psi E[\Delta\psi']}{\sigma E\psi'}$$

$$< 1.0055t(t[2]\sigma\exp(-t^2/2)(2)\cosh(ct))\left(\int_0^{c'} x\psi(x)\phi(\sigma x)dx\right)$$

and, using $2-c^2 > 1$, this can be weakened to the result stated in the lemma.

This completes the proof of (8.83). □

(8.86) LEMMA:

$$2\sigma(\exp(-t^2/2)-1) < 2\sigma[\exp(-t^2/2)\cosh(\sigma xt)-1] < 2\sigma[\exp(-t^2/2)\cosh t-1] < 0 .$$

Proof: The first two inequalities follow from the fact that cosh is an increasing function, so that $1 < \cosh(\sigma xt) < \cosh(ct) < \cosh t$ (see again (8.2).) The last inequality, negativity of $\exp(-t^2/2)\cosh t - 1$, is easily checked by elementary calculus.

This completes the proof of (8.86). □

(8.87) LEMMA:

$$\frac{\Phi(c+t)+\Phi(c-t)-2\Phi(c)}{c'\phi(c)}$$

$$< \frac{\Phi(\sigma x+t)+\Phi(\sigma x-t)-2\Phi(\sigma x)}{x\phi(\sigma x)}$$

$$< 2\sigma\left(\exp(-t^2/2)-1\right) > 0 \ .$$

Proof: The last inequality is trivial. The others have been verified by computer. □

(8.88) REMARK: We note that

$$2[\exp(-\frac{t^2}{2})-1] = \lim_{x\to 0}\left[\frac{\Phi(\sigma x+t)+\Phi(\sigma x-t)-2\Phi(\sigma x)}{\sigma x\ \phi(\sigma x)}\right].$$

This is most easily seen by using L'Hôpital's rule:

set $f(x) = \Phi(\sigma x+t)+\Phi(\sigma x-t)-2\Phi(\sigma x)$

and $g(x) = \sigma x\ \phi(\sigma x)$.

Then $\lim_{x\to 0}\frac{f(x)}{g(x)} = \lim_{x\to 0}\frac{f'(x)}{g'(x)}$

$$= \lim_{x\to 0}\frac{\sigma\phi(\sigma x+t)+\sigma\phi(\sigma x-t)-2\sigma\phi(\sigma x)}{\sigma\phi(\sigma x)-\sigma^3x^2\phi(\sigma x)}$$

$$= \frac{\phi(t)+\phi(-t)-2\phi(0)}{\phi(0)}$$

$$= 2[\exp(-\frac{t^2}{2})-1] \ .$$

We note that $\frac{\Phi(\sigma x+t)+\Phi(\sigma x-t)-2\Phi(\sigma x)}{\sigma x\phi(\sigma x)} = h(\sigma x)$, say, is not monotone in σx for fixed t. Further, for each fixed t, h has at least one zero which depends on that t.

We remark finally that h has an inflexion point at $\sigma x = 0$ (this can be verified by computing the second derivative of h and applying l'Hospital's Rule) but, as (8.87) says, h takes its maximum value at $x = 0$ for the range under consideration.

(8.89) LEMMA:

$$\left|\frac{\Phi(\sigma x)-\Phi(x)}{x\phi(\sigma x)} + \sigma - \frac{\phi(x)}{\phi(\sigma x)}\right| < 2|\sigma-1| \ .$$

Proof: This again is computer verification. □

(8.90) REMARK: It is easy to get bounds on $\sigma - \frac{\phi(x)}{\phi(\sigma x)}$ using elementary calculus, but the function $\frac{\Phi(\sigma x)-\Phi(x)}{x\phi(\sigma x)}$ is not easily tackled.

$\sigma - \frac{\phi(x)}{\phi(\sigma x)}$ increases in x while $\frac{\Phi(\sigma x)-\Phi(x)}{x\phi(\sigma x)}$ decreases in x. The sum of the two functions changes sign at a point close to 1.4 (the critical point varies with σ, of course.) This was one reason why we ruled out large values of x from our analysis. From numerical work, we state:

(8.91) LEMMA:

$$\frac{G(\sigma x)-G(x)}{x\phi(\sigma x)} + \sigma - \frac{\phi(x)}{\phi(\sigma x)}$$

retains the same sign as $\sigma-1$. □

<u>(8.92) LEMMA</u>:

$$2\sigma(\exp(-t^2/2)-1) + \frac{\Phi(c+t)+\Phi(c-t)-2\Phi(c)}{c'\phi(c)}$$

$$+ \frac{2\sigma t^2 \exp(-t^2/2)}{\cosh t - t^2} > 0 \quad .$$

<u>Proof</u>: Again, this had to be verified by computer. □

((8.91) is not absolutely necessary, but does simplify our work later.)

<u>(8.93) LEMMA</u>:

$$\frac{1}{\sigma^3}\left(-E[\Delta^2\psi'] + \frac{E^2[\Delta\psi']}{E\psi'}\right)$$

$$> \frac{-E[\Delta^2\psi']}{\sigma^3} \qquad (t \neq 0\text{, of course})$$

$$> 2(1.19)\exp(-t^2/2)\int_0^{c'} x\psi(x)\phi(\sigma x)\,dx$$

$$> 0 \quad .$$

<u>Proof</u>: We need only check the second inequality.

We have

$$-\frac{1}{\sigma^3}E[\Delta^2\psi'] = 2\exp(-t^2/2)\int_0^{c'} x\psi(x)\phi(\sigma x)\sum_{j=0}^{\infty}(\sigma xt)^{2j}\left[\frac{2-\sigma^2x^2}{(2j)!} + \frac{\sigma^2x^2t^2}{(2j+1)!}\right]dx$$

$$> 2\exp(-t^2/2)\int_0^{c'} x\psi(x)\phi(\sigma x)\sum_{j=0}^{\infty}(\sigma xt)^{2j}\frac{2-\sigma^2x^2}{(2j)!} \quad \text{(there is little loss of sharpness here)}$$

$$> 2\exp(-t^2/2)(2-c^2)\int_0^{c'} x\psi(x)\phi(\sigma x)\cosh(\sigma xt)\,dx$$

$$2\exp(-t^2/2)(2-c^2)\int_0^{c'} x\psi(x)\phi(\sigma x)\,dx$$

$$2(1.19)\exp(-t^2/2)\int_0^{c'} x\psi(x)\phi(\sigma x)\,dx\,, \qquad \text{by (8.2)}$$

and this completes the proof of (8.92). □

We now have all the equipment we need to finish our proof of (8.49).

We write

$$-E\left[\rho_i+\psi_i\frac{E\left[\Delta_i\psi_i'\right]}{\sigma E\psi_i'}-\frac{\Delta_i\psi_i}{\sigma}\right]-a$$

$$=E\left[\frac{\Delta_i\psi_i}{\sigma}-U\psi(U)\right]+E\left[\rho(U)-\rho_i\right]-\frac{E\psi_i E\left[\Delta_i\psi_i'\right]}{\sigma E\psi_i'}$$

(from (8.6) and (8.9)) and, using (8.24), this equals (dropping the subscript i and writing, as usual, $t=\sum_{j=1}^{p}a_{ij}t_j$)

$$2\sigma\exp(-t^2/2)\int_0^{c'} x\psi(x)\phi(\sigma x)\cosh(\sigma xt)\;-\;2\int_0^{c'} x\psi(x)\phi(x)\,dx$$

$$+\int_0^{c'}\psi(x)\left[\Phi(\sigma x+t)+\Phi(\sigma x-t)-2\Phi(x)\right]dx-\frac{E\psi_i E\left[\Delta_i\psi_i'\right]}{\sigma E\psi_i'}$$

which in turn equals, upon adding and subtracting certain quantities,

$$\int_0^{c'} x\psi(x)\phi(\sigma x)\left\{2\sigma\left[\exp(-t^2/2)\cosh(\sigma xt)-1\right]\right.$$

$$+\frac{\Phi(\sigma x+t)+\Phi(\sigma x-t)-2\Phi(\sigma x)}{x\phi(\sigma x)}$$

$$\left.+2\left[\frac{\Phi(\sigma x)-\Phi(x)}{x\phi(\sigma x)}+\sigma-\frac{\phi(x)}{\phi(\sigma x)}\right]\right\}dx$$

$$- \frac{E\psi E[\Delta\psi']}{\sigma E\psi'}$$

$= \zeta_1 + \zeta_2$, say where

$$\zeta_1 = \int_0^{c'} x\psi(x)\phi(\sigma x)\left\{2\sigma\left[\exp(-t^2/2)\cosh(\sigma xt)-1\right] + \frac{\Phi(\sigma x+t)+\Phi(\sigma x-t)-2\Phi(\sigma x)}{x\phi(\sigma x)}\right\} dx - \frac{E\psi E[\Delta\psi']}{\sigma E\psi'} \tag{8.94}$$

and

$$\zeta_2 = 2\int_0^{c'} x\psi(x)\phi(\sigma x)\left[\frac{\Phi(\sigma x)-\Phi(x)}{x\phi(\sigma x)} + \sigma - \frac{\phi(x)}{\phi(\sigma x)}\right] dx \quad . \tag{8.95}$$

By (8.89), we have

$$|\zeta_2| < 4|\sigma-1| \int_0^{c'} x\psi(x)\phi(\sigma x)dx \quad . \tag{8.96}$$

By (8.86), (8.87) and (8.83),

$$\zeta_1 > \left[2\sigma(\exp(-t^2/2)-1) + \frac{\Phi(c+t)+\Phi(c-t)-2\Phi(c)}{c'\phi(c)}\right] \int_0^{c'} x\psi(x)\phi(\sigma x)\,dx + \frac{2\sigma t^2\exp(-t^2/2)}{\cosh t-t^2} \int_0^{c'} x\psi(x)\phi(\sigma x)\,dx$$

and hence

$$\zeta_1 > 0, \quad \text{by (8.92)} \ . \tag{8.97}$$

Then, since in (8.94) the term in curly brackets is negative (by (8.86) and (8.87) again), we have

(8.98) $$|\zeta_1| = \zeta_1 < \frac{-E\psi E[\Delta\psi']}{\sigma E\psi'}$$

and so, by (8.83),

(8.99) $$|\zeta_1| = \zeta_1 < 4(1.0055)\sigma t^2 \exp(-t^2/2)\cosh(ct) \int_0^{c'} x\psi(x)\phi(\sigma x)dx \quad .$$

From (8.93) and (8.96), we have that

$$\frac{|\zeta_2|}{\frac{1}{\sigma^3}\left|E[\Delta^2\psi'] - \frac{E^2[\Delta\psi']}{E\psi'}\right|} = \frac{|\zeta_2|}{\frac{1}{\sigma^3}\left[-E[\Delta^2\psi'] + \frac{E^2[\Delta\psi']}{E\psi'}\right]}$$

$$< \frac{4|\sigma-1|}{2(1.19)\exp(t^2/2)} < 2|\sigma-1| \frac{\exp(k^2/2)}{1.19}$$

$$\leq 2|\sigma-1| \frac{\exp((.0627)^2/2)}{1.19} \qquad \text{(see (8.42))}$$

$$< 2|\sigma-1| \frac{1.0020}{1.19}$$

$$< 2|\sigma-1|$$

i.e.

(8.100) $$|\zeta_2| < 2|\sigma-1| \left|\frac{E[\Delta^2\psi'] - \frac{E^2[\Delta\psi']}{E\psi'}}{\sigma^3}\right| \quad .$$

Also, from (8.99) and (8.93) we obtain

$$\frac{|\zeta_1|}{\frac{1}{\sigma^3}\left|E[\Delta^2\psi'] - \frac{E^2[\Delta\psi']}{E\psi'}\right|} < \frac{4(1.0055)\sigma t^2 \cosh(ct)\exp(-t^2/2)}{2(1.19)\exp(-t^2/2)}$$

$$< \frac{2(1.0055)[2b-1]k\cosh(.9k)}{1.19}|t| \qquad (\text{since } \sigma < 2b-1 \text{ and } t < k)$$

$$\le \frac{2(1.0055)[2(1.0334)-1](.0627)\cosh[.9(.0627)]}{1.19}|t| \quad (\text{see } 8.42))$$

$$< .1133\ |t|\ .$$

Thus

$$(8.101) \qquad |\zeta_1| < .1133\ |t| \left| \frac{E[\Delta^2\psi'] - \frac{E^2[\Delta\psi']}{E\psi'}}{\sigma^3} \right| \ .$$

From (8.100), (8.101) and the triangle inequality, we have that, for each i, $i=1,\ldots,p$, (recall that we have been writing t for any of the $\sum_{j=1}^{p} a_{ij}t_j$)

$$\left| E\left[\rho_i + \psi_i \frac{E[\Delta_i\psi_i']}{\sigma E\Delta_i'} - \frac{\Delta_i\psi_i}{\sigma}\right] + a \right|$$

$$< \left(2|\sigma-1| + .1133 \left|\sum_{j=1}^{p} a_{ij}t_j\right|\right) \left| \frac{E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}}{\sigma^3} \right|$$

from which it follows that

$$(8.102) \quad \eta = \frac{\sum_{i=1}^{p} q_i E\left[\rho_i + \psi_i \frac{E[\Delta_i\psi_i']}{\sigma E\psi_i'} - \frac{\Delta_i\psi_i}{\sigma}\right] + a}{\sum_{i=1}^{p} \frac{q_i}{\sigma^3}\left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right)} < 2|\sigma-1| + .1133 \max_{1\le i\le p} \left|\sum_{j=1}^{p} a_{ij}t_j\right| \ .$$

Returning now to (8.65) we write

$$H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1} H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} = A^{-1}\begin{pmatrix}\frac{-E\psi_1}{E\psi_1'}\,\sigma\\ \cdot\\ \cdot\\ \cdot\\ \frac{-E\psi_p}{E\psi_p'}\,\sigma\\ \eta\end{pmatrix} \tag{8.103}$$

and we now recall our norm (8.48):

$$\left\|\begin{pmatrix}\underset{\sim}{t}\\ s\end{pmatrix}\right\| = |s| + \frac{2(b-1)}{k}\max_{1\le i\le p}\left|\sum_{j=1}^{p} a_{ij}t_j\right| \quad . \tag{8.104}$$

If we write $[\underset{\sim}{x}]_l$ to denote the absolute value of the lth co-ordinate of a vector $\underset{\sim}{x}$, then it is easy to see that (8.104) is equivalent to:

$$\left\|\begin{pmatrix}\underset{\sim}{t}\\ s\end{pmatrix}\right\| = \left[\begin{bmatrix}\frac{2(b-1)}{k} & & & \\ & \ddots & & \\ & & \frac{2(b-1)}{k} & \\ & & & 1\end{bmatrix}\begin{bmatrix}A & \underset{\sim}{0}\\ \underset{\sim}{0}^T & 1\end{bmatrix}\begin{pmatrix}t_1\\ \cdot\\ \cdot\\ \cdot\\ t_p\\ s\end{pmatrix}\right]_{p+1}$$
$$+ \max_{1\le i\le p}\left[\begin{bmatrix}\frac{2(b-1)}{k} & & & \\ & \ddots & & \\ & & \frac{2(b-1)}{k} & \\ & & & 1\end{bmatrix}\begin{bmatrix}A & \underset{\sim}{0}\\ \underset{\sim}{0}^T & 1\end{bmatrix}\begin{pmatrix}t_1\\ \cdot\\ \cdot\\ \cdot\\ t_p\\ s\end{pmatrix}\right]_{i} \quad . \tag{8.105}$$

Note now that

(8.106)
$$\begin{bmatrix} A & \underset{\sim}{0} \\ \underset{\sim}{0}^T & 1 \end{bmatrix} A^{-1} \begin{pmatrix} \dfrac{-E\psi_1}{E\psi_1'}\,\sigma \\ \vdots \\ \dfrac{-E\psi_p}{E\psi_p'}\,\sigma \\ \eta \end{pmatrix} = \begin{pmatrix} \dfrac{-E\psi_1}{E\psi_1'}\,\sigma - \dfrac{E[\Delta_1\psi_1']}{E\psi_p'}\,\eta \\ \vdots \\ \dfrac{-E\psi_p}{E\psi_p'}\,\sigma - \dfrac{E[\Delta_p\psi_p']}{E\psi_p'}\,\eta \\ \eta \end{pmatrix}.$$

(8.106) follows from

$$\begin{bmatrix} A & \underset{\sim}{0} \\ \underset{\sim}{0}^T & 1 \end{bmatrix} A^{-1} \begin{pmatrix} \dfrac{-E\psi_1}{E\psi_1'}\,\sigma \\ \vdots \\ \dfrac{-E\psi_p}{E\psi_p'}\,\sigma \\ \eta \end{pmatrix}$$

$$= \begin{bmatrix} A & \underset{\sim}{0} \\ \underset{\sim}{0}^T & 1 \end{bmatrix} \begin{bmatrix} A^{-1} & -A^{-1}\underset{\sim}{b} \\ \underset{\sim}{0}^T & 1 \end{bmatrix} \begin{pmatrix} \dfrac{-E\psi_1}{E\psi_1'}\,\sigma \\ \vdots \\ \dfrac{-E\psi_p}{E\psi_p'}\,\sigma \\ \eta \end{pmatrix} \qquad \text{(from (8.60))}$$

$$= \begin{bmatrix} I & -\underset{\sim}{b} \\ \underset{\sim}{0}^T & 1 \end{bmatrix} \begin{pmatrix} \dfrac{-E\psi_1}{E\psi_p'}\,\sigma \\ \vdots \\ \dfrac{-E\psi_p}{E\psi_p'}\,\sigma \\ \eta \end{pmatrix}$$

$$= \begin{pmatrix} \frac{-E\psi_1}{E\psi_1'}\sigma & - & \eta\,\frac{E[\Delta_1\psi_1']}{E\psi_1'} \\ & \vdots & \\ \frac{-E\psi_p}{E\psi_p'}\sigma & - & \eta\,\frac{E[\Delta_p\psi_p']}{\sigma E\psi_p'} \\ & \eta & \end{pmatrix} \qquad \text{(from 6.42).)}$$

From (8.103), (8.105) and (8.106) we have

$$(8.107)\quad \left\| \begin{pmatrix} \tilde{t} \\ \sigma \end{pmatrix} - \begin{pmatrix} \tilde{0} \\ 1 \end{pmatrix} - H'\begin{pmatrix} \tilde{t} \\ \sigma \end{pmatrix}^{-1} H\begin{pmatrix} \tilde{t} \\ \sigma \end{pmatrix} \right\|$$

$$= |\sigma-1-\eta| + \frac{2(b-1)}{k} \max_{1\le i\le p} \left| \sum_{i=1}^{p} a_{ij}t_j - \left[\frac{-E\psi_i}{E\psi_i'}\sigma - \eta\,\frac{E[\Delta_i\psi_i']}{\sigma E\psi_i'}\right] \right| .$$

Now, by (8.91), ζ_2, given in (8.95) has the same sign as $\sigma-1$. Also, $\zeta_1 > 0$ (see (8.97)) and

$$-\frac{1}{\sigma^3}\left(E[\Delta_i{}^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right) > 0 .$$

Now write

$$\eta = \frac{-\sum_{i=1}^{p} q_i E\left|\rho_i+\psi_i\,\frac{E[\Delta_i\psi_i']}{\sigma E\psi_i'} - \frac{\Delta_i\psi_i}{\sigma}\right| - a}{-\sum_{i=1}^{p}\frac{q_i}{\sigma^3}\left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i]}{E\psi_i'}\right)}$$

(8.108)

$$= \frac{\sum_{i=1}^{p} q_i\,[\xi_{1i}+\xi_2]}{-\sum_{i=1}^{p}\frac{q_i}{\sigma^3}\left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right)}$$

(here, in an obvious notation, for each i, $i=1,\ldots,p$, ξ_{1i} is the ξ_1 of

(8.94) corresponding to $t = \sum_{j=1}^{p} a_{ij}t_j$)

and we have that

$$\frac{\sum_{i=1}^{p} q_i \xi_{1i}}{-\sum_{i=1}^{p} \frac{q_i}{\sigma^3}\left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right)} \quad \text{has the same sign as } \sigma-1.$$

From this, (8.102) and (8.108), we have

$$(8.109)\quad |\sigma-1-\eta| \le \left|\sigma-1 - \frac{\xi_2}{-\sum_{i=1}^{p} \frac{q_i}{\sigma^3}\left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right)}\right| + \frac{\sum_{i=1}^{p} q_i \xi_{1i}}{-\sum_{i=1}^{p} \frac{q_i}{\sigma^3}\left(E[\Delta_i^2\psi_i'] - \frac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right)}$$

(no absolute value sign is required on the last term)

$$< |\sigma-1| + .1133 \max_{1\le i\le p} \left|\sum_{j=1}^{p} a_{ij}t_j\right| .$$

From (8.107) and (8.109) we have (using the triangle inequality)

$$\left\| \begin{pmatrix} \dot{\underset{\sim}{t}} \\ \sigma \end{pmatrix} - \begin{pmatrix} \underset{\sim}{0} \\ 1 \end{pmatrix} - H' \begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}^{-1} H \begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix} \right\|$$

$$< |\sigma-1| + .1133 \max_{1\le i\le p} \left|\sum_{j=1}^{p} a_{ij}t_j\right|$$

$$+ \frac{2(b-1)}{k}\left\{\max_{1\le i\le p}\left|\sum_{j=1}^{p} a_{ij}t_j - \left(\frac{-E\psi_i}{E\psi_i'}\right)\sigma\right| + |\eta| \max_{1\le i\le p}\left|\frac{E[\Delta_i\psi_i']}{\sigma E\psi_i'}\right|\right\}$$

$$< |\sigma-1| + .1133 \max_{1\le i\le p} \left|\sum_{j=1}^{p} a_{ij}t_j\right|$$

$$+ \frac{2(b-1)}{k} \left\{ .0055 \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| + \left[2|\sigma-1| + .1133 \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| \right] \max_{1 \le i \le p} \left| \frac{E[\Delta_i \psi_i']}{\sigma E \Delta_i'} \right| \right\}$$

(using (8.68) and (8.102))

$$< |\sigma-1| + .1133 \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| + \frac{2(b-1)}{k} \left\{ .0055 \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| \right.$$

$$\left. + \left[2|\sigma-1| + .1133 \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| \right] (2)(1.0055) \max_{1 \le i \le p} \left| \sum_{j=1} a_{ij} t_j \right| \left(\frac{1}{\sigma}\right) \right\}$$

(using (8.79))

$$= |\sigma-1| + \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| \left\{ .1133 + \frac{2(b-1)}{k} \left[.0055 \right.\right.$$

$$\left.\left. + \left\{ 2|\sigma-1| + .1133 \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| \right\} \frac{2(1.0055)}{\sigma} \right] \right\}$$

$$< |\sigma-1| + \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| \left\{ .1133 + \frac{2(b-1)}{k} \left[.0055 \right.\right.$$

$$\left.\left. + \left\{ 2[2(b-1)] + .1133\, k \right\} \frac{2(1.0055)}{(-2b+3)} \right] \right\}$$

(using (8.45), (8.46), (8.47) and (8.42))

$$< |\sigma-1| + \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| \left\{ .1133 + \frac{2(b-1)}{k} \left[.0055 \right.\right.$$

$$\left.\left. + \left\{ 4[.0334] + .1133\, [.0627] \right\} \frac{2(1.0055)}{.9332} \right] \right\}$$

(using (8.42))

$$< |\sigma-1| + \left[.1133 + \frac{2(b-1)}{k} (.3033) \right] \max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right| ,$$

that is,

$$(8.110)\quad \left\|\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}-\begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix}-H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1}H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}\right\| < |\sigma-1|+\left[.1133+\frac{2(b-1)}{k}(.3033)\right]\max_{1\le i\le p}\left|\sum_{j=1}^{p}a_{ij}t_j\right| .$$

We note now first that

$$.1133+\frac{2(b-1)}{k}(.3033) < \frac{2(b-1)}{k}$$

if $\quad \dfrac{b-1}{k} > \dfrac{.1133}{2(1-.3033)}$

and this is true if

$$\frac{b-1}{k} > .0814$$

and this is easily true (see (8.42)).

Thus, (8.110) implies

$$(8.111)\quad \left\|\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}-\begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix}-H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1}H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}\right\| < |\sigma-1| + \frac{2(b-1)}{k}\max_{1\le i\le p}\left|\sum_{j=1}^{p}a_{ij}t_j\right|$$

or equivalently, (by (8.48)),

$$\left\|\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}-\begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix}-H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1}H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}\right\| < \left\|\begin{pmatrix}\underset{\approx}{t}\\ \sigma-1\end{pmatrix}\right\| ,$$

that is,

$$(8.112)\quad \left\|\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}-\begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix}-H'\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{-1}H\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}\right\| < \left\|\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}-\begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix}\right\| \quad \text{for all} \quad \begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix} \in \mathcal{D} .$$

From (8.112), we get

$$(8.113)\quad \left\|\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{1}-\begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix}\right\| < \left\|\begin{pmatrix}\underset{\sim}{t}\\ \sigma\end{pmatrix}^{0}-\begin{pmatrix}\underset{\sim}{0}\\ 1\end{pmatrix}\right\| .$$

Now (8.113) does not imply that the first iterate $\begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}^1$ belongs to $\mathcal{D}$ because the starting value $\begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}^0 = \begin{pmatrix} \underset{\sim}{\theta}^{**} \\ \beta \end{pmatrix}$ only satisfies

$$\left\| \begin{pmatrix} \underset{\sim}{\theta}^{**} \\ \beta \end{pmatrix} - \begin{pmatrix} \underset{\sim}{0} \\ 1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} \underset{\sim}{\theta}^{**} \\ \beta-1 \end{pmatrix} \right\| = |\beta-1| + \frac{2(b-1)}{k} \max_{1\le i\le p} \left| \sum_{j=1}^{p} a_{ij}\theta_i^{**} \right| \quad \text{(see (8.48))}$$

$$< b - 1 + \frac{2(b-1)}{k} k \quad \text{(see (8.23))}$$

$$= 3(b-1) ,$$

that is

$\left\| \begin{pmatrix} \underset{\sim}{\theta}^{**} \\ \beta \end{pmatrix} - \begin{pmatrix} \underset{\sim}{0} \\ 1 \end{pmatrix} \right\| < 3(b-1)$ and we cannot deduce from this and (8.113) that $\begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}^1 \in \mathcal{D}$ (see (8.44) and (8.49)).

However, (8.110) does imply that the first and all remaining iterates lie in $\mathcal{D}$, for we have from (8.110):

$$\text{(8.114)} \quad \left\| \begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}^1 - \begin{pmatrix} \underset{\sim}{0} \\ 1 \end{pmatrix} \right\| < |\beta-1| + \left[.1133 + \frac{2(b-1)}{k} (3.033) \right] \max_{1\le j\le p} \left| \sum_{j=1}^{p} a_{ij}\theta_j^{**} \right|$$

$$< b - 1 + \left[.1133 + \frac{2(b-1)}{k} (3.033) \right] k \quad \text{(from (8.23))}$$

$$< \quad 2(b-1) \quad \text{so that} \quad \begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}^1 \in \mathcal{D} \quad \text{(see (8.44))}.$$

The last inequality here follows from

$$\frac{b-1}{k} \ge .5327 \quad \text{(see (8.42))}$$

so that, certainly, $\dfrac{b-1}{k} > \dfrac{.1133}{1-2(.3033)}$

or $\quad .1133 + \frac{2(b-1)}{k}\,(.3033) < \frac{b-1}{k}$

or $\quad \left[.1133 + \frac{2(b-1)}{k}\,(.3033)\right]k < b - 1\ .$

From (8.112) and (8.114) we have

$$\left\| \binom{\underset{\sim}{t}}{\sigma}^{2} - \binom{\underset{\sim}{0}}{1} \right\| < \left\| \binom{\underset{\sim}{t}}{\sigma}^{1} - \binom{\underset{\sim}{0}}{1} \right\| < 2(b-1)\ .$$

Thus the second iterate lies in $\mathcal{D}$ and a simple induction as used before shows that all iterates lie in $\mathcal{D}$ and converge to $\binom{\underset{\sim}{0}}{1}$.

This, finally, completes the proof of (8.49). □

We have shown that the Newton's method solution of $H\binom{\underset{\sim}{t}}{\sigma} = \underset{\sim}{0}$ with starting value $\binom{\underset{\sim}{\theta}^{**}}{\beta}$ is $\binom{\underset{\sim}{0}}{1}$. Using this we can show, among other things, that $\left[\text{writing } \underset{\sim}{T}_n \text{ (see (8.12)) as } (T_{n1},\dots,T_{np},T_{n\,p+1})^T,\right]$ $(T_{n1},\dots,T_{np})^T$ is a consistent and asymptotically normal estimator of the true $\underset{\sim}{\theta}$, which we have assumed to be $\underset{\sim}{0}$. We shall not write out the proof, for it is entirely similar to our proofs of (5.123) and (5.124). Indeed consistency and asymptotic normality of $(T_{n1},\dots,T_{np})^T$ should be obvious from (7.20) and (7.26). Accordingly, we shall merely state the results (note that (8.49) plays the same role in their proof as (4.13) played in the proofs of (5.123) and (5.124)).

(8.115) THEOREM:

Let $\psi \in \Psi_{c'}$ (defined in (8.4)). Then

$$(T_{n1},\dots,T_{np})^T(\psi) \xrightarrow{P} \underset{\sim}{0} \tag{8.116}$$

where T_{ni} is the i^{th} co-ordinate of $\underset{\sim}{T}_n(\psi)$ (see (8.12)).

Furthermore $T_{n\ p+1}(\psi) \xrightarrow{P} 1$.

(8.117) THEOREM:

We have

$$n^{\frac{1}{2}}(T_{n1},\dots,T_{np})^T(\psi) \xrightarrow{D} MVN\left(\underset{\sim}{0},\ C_0^{-1}\frac{\int_{-c'}^{c'}\psi^2(y)\phi(\beta y)dy}{\beta\left[\int_{-c'}^{c'}\psi(y)\phi'(\beta y)dy\right]^2}\right). \tag{8.118}$$

Furthermore $n^{\frac{1}{2}}\ T_{n\ p+1}(\psi) \xrightarrow{D} N\ (1,\ E[X^2\psi']\mathrm{Var}[X\psi(X)-\rho(X)])$ and the estimators $(T_{n1},\dots,T_{np})^T(\psi)$ and $T_{n,p+1}(\psi)$ are asymptotically stochastically independent.

(8.119) REMARK:

A few points concerning our choice of norm (8.48) and neighbourhood (8.44) should be made. Recall first that in Section 4 our neighbourhood D (see (3.20)) was chosen there because it was the natural p-dimensional analogue of the interval $(-k,k)$ used by Collins (1976) in the location case. In (4.13) we were able to show that the starting value and all iterates belonged to D. In our work in Section 4, we used the norm (4.30). This was definitely the most appropriate norm since the p-dimensional ball of radius k for the norm (4.30) coincided with the region D. If, in Section 4, we had used, for example, the elliptic norm $\|\underset{\sim}{t}\|_{A^TA} = (\underset{\sim}{t}^T A^T A \underset{\sim}{t})^{\frac{1}{2}}$ we would then have defined D to be a hyper-ellipse (otherwise one can run into difficulty in showing that the iterates belong to D). But this choice of D would not be as good as the hyper-parallelogram actually used, for we would have no reason to allow the range of $\sum_{j=1}^{p} a_{ij} t_j$ to be different for any two values of i. Now in Section 8, the situation is much more complicated. It is impossible to show that the iterates lie in $D \times (2-b,b)$. $\left\{(D \times (2-b,b)\right.$ is an open neighbourhood containing the starting value $\left.\begin{pmatrix} \underset{\sim}{\theta}^{**} \\ \beta \end{pmatrix}.\right\}$ This is because of a flaw in the nature of things - it just happens that the last co-ordinate of $\begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix} - \begin{pmatrix} \sim \end{pmatrix} - H' \begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}^{-1} H \begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}$ has a component which does not go to zero as the last component σ of $\begin{pmatrix} \underset{\sim}{t} \\ \sigma \end{pmatrix}$ goes to zero (see (8.65) and notice that the bound in (8.102) could not be chosen to involve σ only - unless, of course, we put a numerical upper bound on $\max_{1 \le i \le p} \left| \sum_{j=1}^{p} a_{ij} t_j \right|$). Our choice of $\mathcal{D}$ in (8.44) was to accommodate this fact. Note that in (8.49) there is no need to extend down as far as $-2b+3$. We could go to $1-\varepsilon$ for any positive ε just as well. The key point, however, is that we must go above b. Finally, the choice of norm (8.48) is a natural one for our choice of $\mathcal{D}$.

(8.120) COMPARISON OF THE METHODS OF SECTIONS 7 AND 8 IN ESTIMATING θ

In Section 7 the range of α was taken to be $0 < \alpha < .5$ while in Section 8, α was not allowed to exceed .05. By refining our inequalities in Section 8, we can get our results to hold for larger values of α and by another method that we shall outline below we can get our results to hold for much larger values of α again although this method may involve additional assumptions on the class of ψ-functions used. Now also in Section 8 we made the restriction $c \leq .9$ (see (8.2)). This is a serious restriction because when α is small, d is large, so that the errors are normal except in small tails. Thus it is unreasonable to truncate the ψ-functions in Ψ_c, (see (8.4)) as severely as the restriction $c \leq .9$ forces. An examination of the analysis in Section 8 shows why we made this restriction on c. It was necessary to put a lower bound on $-E[\Delta^2\psi']$ (see (8.93)), for it occurs in the denominator of η (see (8.103)). Now,

$$(8.121)\quad -E[\Delta^2\psi'] = 2\sigma^3 \exp\left(-t^2/2\right) \int_0^{c'} x\psi(x)\phi(\sigma x) \sum_{j=0}^{\infty} (\sigma x t)^{2j} \left[\frac{2-\sigma^2x^2}{(2j)!} + \frac{\sigma^2x^2t^2}{(2j+1)!}\right] dx$$

and it is clearly impossible to tell even the sign of this function of t and σ when σx is "large". When σ^2x^2 exceeds 2 we can, by ruling out certain ψ-functions, ensure that (8.121) is positive. We chose, in our analysis, not to make any additional assumptions about the ψ-functions than we made in (8.4). On the other hand, it is clear that we can get an adequately large positive lower bound for (8.121) for large values of c' by making (not very restrictive) assumptions on the $\psi^{\underline{S}}$. This allows us, in addition, to increase the values allowed by the parameters k and b. For example, we could have defined our class of ψ-functions so that each ψ in our class satisfies, in addition to our usual conditions,

$-E[\Delta^2\psi'] > \lambda\left(2\sigma^3\exp(-t^2/2)\right)\int_0^{c'} x\ \psi(x)\phi(\sigma x)\ dx$ in some range of values of t and σ and some range of the parameter c' (all depending on α), for some appropriate positive λ. We chose not to do this since this is not a condition that can be easily verified without knowledge of the functional form of ψ. However, the following is a most interesting idea. First recall the analytic result:

(8.122) let f,g,h be real functions defined for x in some set S.

Suppose f,g are integrable and h measurable and bounded on S. Put

$$A(y) = \{x{:}h(x) \geq y\}\ ,\quad B(y) = A - A(y) = \{x{:}h(x) < y\}\ .$$

If $\displaystyle\int_{A(y)} f(x)dx \geq \int_{A(y)} g(x)dx$ for all $y \in [0,\infty)$

and if $\displaystyle\int_{B(y)} f(x)dx \leq \int_{B(y)} g(x)dx$ for all $y \in (-\infty,0)$,

then $\displaystyle\int_S g(x)h(x)dx \leq \int_S f(x)h(x)dx$ (see Mitrinovič (1970), p. 307) .

Consider the problem of applying this result with a view to getting a suitable bound for η (see (8.102)) without making serious restrictions on c',k and b. We note that it is sufficient to get a suitable bound for

$$(8.123)\quad \frac{-E\left[\rho_i+\psi_i\,\dfrac{E[\Delta_i\psi_i']}{\sigma E\psi_i'}-\dfrac{\Delta_i\psi_i}{\sigma}\right]-a}{-\dfrac{1}{\sigma^3}\left[E[\Delta_i^2\psi_i']+\dfrac{E^2[\Delta_i\psi_i']}{E\psi_i'}\right]}\quad \text{for each } i=1,\ldots,p\ ,$$

and, as usual, we shall drop the subscript i and write t for $\sum_{j=1}^{p} a_{ij}t_j$.

Now,

$$-E\left[\Delta^2\psi'\left(\frac{\Delta}{\sigma}\right)\right]+\frac{E^2\left[\Delta\psi'\left(\frac{\Delta}{\sigma}\right)\right]}{E\psi'\left(\frac{\Delta}{\sigma}\right)}>0 \quad \text{if}$$

$$-E\left[\Delta^2\psi'\left(\frac{\Delta}{\sigma}\right)\right]>0 \quad (\text{recall that } E\psi'\left(\frac{\Delta}{\sigma}\right)>0 \text{ if } t^2<1)$$

and this is true if

$$(8.124)\quad \int_0^{c'} x\psi(x)\phi(\sigma x)(2-\sigma^2x^2)\cosh(\sigma xt)>0 \quad (\text{see } (8.33))\ .$$

Note that even for "large" values of c', this condition will be satisfied for a large class of ψ-functions of the type $\Psi_{c'}$, in view of the rapid rate of decrease of ϕ.

Now an examination of the proof of (8.24) shows that we may write

$$-E\left[\rho\left(\frac{\Delta}{\sigma}\right)-\frac{\Delta\psi\left(\frac{\Delta}{\sigma}\right)}{\sigma}\right]-a=E\left[\frac{\Delta\psi\left(\frac{\Delta}{\sigma}\right)}{\sigma}\right]-E[U\psi(U)]-E\rho\left(\frac{\Delta}{\sigma}\right)+E\rho(U)$$

in the form

$$\int_{-c'}^{c'} x\psi(x)\phi(\sigma x+t)dx-\int_{-c'}^{c'} x\psi(x)\phi(x)dx+\int_{-c'}^{c'}\psi(x)\Phi(\sigma x+t)dx-\int_{-c'}^{c'}\psi(x)\Phi(x)dx\ ,$$

that is,

$$(8.125)\quad \mu_1=-E\left[\rho\left(\frac{\Delta}{\sigma}\right)-\frac{\Delta\psi\left(\frac{\Delta}{\sigma}\right)}{\sigma}\right]-a=\int_{-c'}^{c'}\psi(x)\{x[\sigma\phi(\sigma x+t)-\phi(x)]+\Phi(\sigma x+t)-\Phi(x)\}dx\ .$$

Now also it is easily seen that

$$\mu_2 = -\frac{1}{\sigma^3} E[\Delta^2 \psi'(\tfrac{\Delta}{\sigma})] \quad .$$

can be written as (using integration by parts or otherwise)

(8.126) $$\mu_2 = \int_{-c'}^{c'} \psi(x)\ x[-\sigma^2 x^2+2-\sigma x t]\phi(\sigma x+t)\ dx \ .$$

Now $\left|\frac{\mu_1}{\mu_2}\right| \le \lambda$ for some "suitably small" λ which depends on t and σ

if (assuming (8.124))

(8.127) $$-\lambda\mu_2 \le \mu_1 \le \lambda\mu_2 \ .$$

We shall give a condition under which the right hand inequality in (8.127) is satisfied. Applying (8.122) with

$$f(x) = x[-\sigma^2x^2+2-\sigma xt]\phi(\sigma x+t)$$

$$g(x) = x[\sigma\phi(\sigma x+t)-\phi(x)]+\Phi(\sigma x+t)-\Phi(x)$$

and $h(x) = \psi(x)$,

we see from (8.125) and (8.126) that

$$\mu_1 < \lambda\mu_2 \text{ if}$$

(8.128)
$$\int_{A(y)} \{x[\sigma\phi(\sigma x+t)-\phi(x)]+\Phi(\sigma x+t)-\Phi(x)\}dx \le \lambda \int_{A(y)} x[-\sigma^2x^2+2-\sigma xt]\phi(\sigma x+t)dx$$

and

$$\int_{B(y)} \{x[\sigma\phi(\sigma x+t)-\phi(x)]+\Phi(\sigma x+t)-\Phi(x)\}dx$$

(8.129)

$$> \lambda \int_{B(y)} x[-\sigma^2x^2+2-\sigma xt]\phi(\sigma x+t)dx \ .$$

We shall deal with the simplification of (8.128) only, since the simplification of (8.129) is similar. Write $A(y) = [a(y),b(y)]$. Then the condition (8.128) reads

$$(8.130) \qquad \int_{a(y)}^{b(y)} x[\sigma\phi(\sigma x+t)-\phi(x)]dx + \int_{a(y)}^{b(y)} [\Phi(\sigma x+t)-\Phi(x)]dx$$

$$\leq \lambda \int_{a(y)}^{b(y)} x[-\sigma^2x^2+2-\sigma dt]\phi(\sigma x+t)dx \ .$$

Noting that $\displaystyle\int_{a(y)}^{b(y)} [\Phi(\sigma x+t)-\Phi(x)]dx$

$$= x[G(\sigma x+t)-G(x)]\Big|_{a(y)}^{b(y)} - \int_{a(y)}^{b(y)} x[\sigma\phi(\sigma x+t)-\phi(x)]dx$$

and that

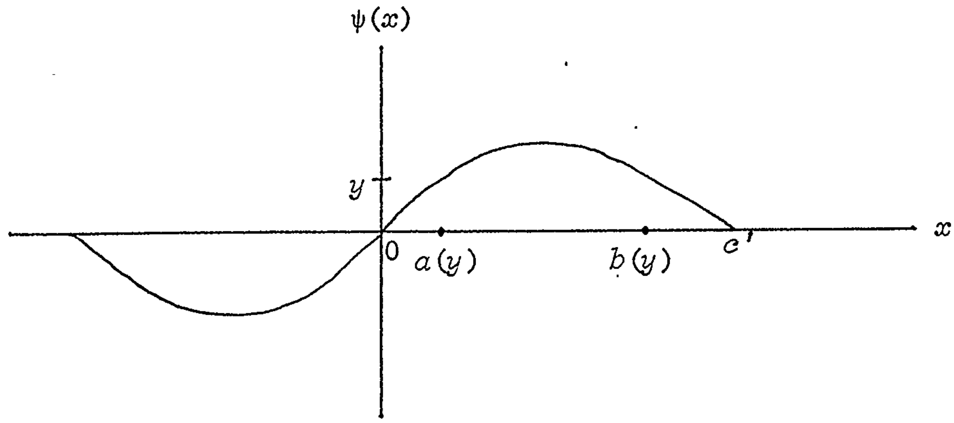
$$\int_{a(y)}^{b(y)} x[-\sigma^2x^2+2-\sigma xt]\phi(\sigma x+t)dx = x^2\phi(\sigma x+t)\Big|_{a(y)}^{b(y)} \ ,$$

(8.130) reads

$$(8.131) \qquad x[G(\sigma x+t)-G(x)]\Big|_{a(y)}^{b(y)} \leq \lambda \ x^2\phi(\sigma x+t)\Big|_{a(y)}^{b(y)} \ .$$

In a suitable range of parameter values, this condition seems to be not difficult to check. Note that (8.131) in no way says that

$\lambda x^2\phi(\sigma x+t) - x[G(\sigma x+t)-G(x)]$ is an increasing function.



Conditions similar to (8.131) can be given to ensure the boundedness of

$$\frac{E\psi\left(\frac{\Delta}{\sigma}\right) \dfrac{E[\Delta\psi'\left(\frac{\Delta}{\sigma}\right)]}{\sigma E\psi'\left(\frac{\Delta}{\sigma}\right)}}{\frac{1}{\sigma^3} E[\Delta^2\psi'\left(\frac{\Delta}{\sigma}\right)]} .$$

Consequently, relatively easy conditions can be given under which (8.123) (and hence η) is bounded by an appropriate function of t and σ (to aid in ensuring that the iterates (8.50) converge) without assuming c' is small and without knowing the functional form of ψ.

As a final remark in our comparison of the methods of Sections 7 and 8, we make the point that we feel the estimator of $\underset{\sim}{\theta}$ found in Section 8 is superior to that in Section 7. We make our contention on the basis that the method of Section 8 improved the initial estimator $\hat{\sigma}_n$ (which was fixed in Section 7) of σ at each step of the iteration process. Note also that the method of Section 8 supplies us with a consistent and asymptotically normal estimator of σ.

In conclusion, let $\Psi'_{c'}$ be the class Ψ'_c of (6.9) with c replaced by c'. Then when $G \in P_\varepsilon$ (see (6.8)) and the errors have distribution $G(\frac{y}{\sigma})$, we recommend statistics of the form $(T_{n1},\ldots,T_{np})^T(\psi)$, for ψ-functions given by (6.10), as estimators of $\underset{\sim}{\theta}$.

BIBLIOGRAPHY

BICKEL, P.J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.*, **70**, 428-433.

BILLINGSLEY, P. (1968). Convergence of Probability Measures. *John Wiley, New York.*

COLLINS, J.R. (1973). Robust estimation of a location parameter in the presence of asymmetry. Ph.D. dissertation, University of California, Berkeley.

COLLINS, J.R. (1976). Robust estimation of a location parameter in the presence of asymmetry. *Ann. Statist.*, **4**, 68-85.

COLLINS, J.R. (1976'). On the consistency of M-estimators. *Purdue Univ., Dept. of Statistics Mineograph* Series No.425.

COLLINS, J.R. (1977). Identifiability of a centre of symmetry. *The Univ. of Calgary, Dept. of Mathematics and Statistics* Research Paper No.340.

CHUNG, K.L. (1974). A Course in Probability Theory. *Academic Press, New York.*

CRAMER, H. (1946). Mathematical Methods of Statistics. *Princeton Univ. Press.*

DRAPER, N.R., and SMITH H. (1966). Applied Regression Analysis. *John Wiley, New York.*

HUBER, P.J. (1964). Robust Estimation of a location parameter. *Ann. Math. Statist.* **35**, 73-101.

HUBER, P.J. (1973). Robust Regression. *Ann. Statist.* **1**, 799-821.

HUBER, P.J. (1977). Robust Statistical Procedures. *SIAM.*

MITRINOVIČ, D.S. (1970). Analytic Inequalities. *Springer-Verlag, New York.*

ORTEGA, J.M. (1972). Numerical Analysis. *Academic Press, New York.*

ORTEGA, J.M. and RHEINBOLDT, W.C. (1970). Iterative Solution of Non-Linear Equations in Several Variables. *Academic Press, New York.*

RELLES, D. (1968). Robust Regression by Modified Least Squares. Ph.D. Thesis, Yale Univ.

SIMMONS, G.F. (1963). Introduction to Topology and Modern Analysis. *McGraw-Hill, New York.*

YOHAI, V.J. (1972). Robust (M) Estimates for the general Linear Model. Notas de Matemática, No.23, Univ. Nacional de La Plata, Facultad de Ciencias Exactas *y* Naturales.

YOHAI, V.J. and MARONNA, R.A. (1979). Asymptotic Behaviour of M-Estimators for the Linear Model. *Ann. Statist.* 7, 258-268.