INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

UNIVERSITY OF ALBERTA

Approximations to the Distributions of Some Robust Test Statistics

BY

Quanli Wang ©

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

STATISTICS

DEPARTMENT OF MATHEMATICAL SCIENCES

EDMONTON, ALBERTA

FALL 1999

National Library of Canada

Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services

Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre reference

Our file Notre reference

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-47112-8

Canada

University of Alberta

Library Release Form

Name of Author: *Quanli Wang*

Title of Thesis: *Approximations to the Distributions of Some Robust Test Statistics*

Degree: *Master of Science*

Year this Degree Granted: *1999*

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and expect as hereinbefore provided, neither the thesis nor any substantial potion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Quanli Wang
#204, 10305-116street
Edmonton,AB
Canada, T5K 1W5

Date: 99-8-11

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Approximations to the Distributions of Some Robust Test Statistics submitted by Quanli Wang in partial fulfillment of the requirements for the degree of Master of Sciences in Statistics.

Dr. Douglas P. Wiens(Supervisor)
Department of Mathematical Sciences
University of Alberta

Dr. N.G.N. Prasad(Committee Chair & Examiner)
Department of Mathematical Sciences
University of Alberta

Dr. Biao Huang(Examiner)
Department of Chemical & Materials Engineering
University of Alberta

Date: 99-2-9

To

My Wife, Liping Liu

and

My Parents.

Abstract

There are various robust GM–type testing procedures. Some of these procedures are simple but less accurate, while others are more accurate but difficult to use in practice. This thesis then concerns the construction of a testing procedure which has better accuracy and still can be used easily in practice.

First, we review some GM–type testing procedures and some techniques used in this thesis. Then a higher order asymptotic expansion of the GM–estimators is derived by using the Edgeworth expansion. Later, a testing statistic Q_n for the robust M–type linear regression problem is given and its asymptotic distribution is investigated. It turns out that the Q_n statistic is approximately F distributed to the order of $O(n^{-2})$. Finally, the simulation study on some selected testing problems will demonstrate the advantages of using the Q_n statistic.

Acknowledgements

I am very grateful to my supervisor, Dr. Douglas P. Wiens, for leading me to the research field for this thesis, and for his many helpful and thoughtful notes, comments, discussions and suggestions throughout the preparation of this thesis. Without his guidance, this thesis would have been impossible.

Thanks also go to the Department of Mathematical Sciences, University of Alberta, for providing me the opportunity to study here and the financial support. I would like to express my appreciation to the faculty members of the department who provided us with many interesting courses in statistics.

Special thanks go to Dr. N. G. N. Prasad and Dr. Biao Huang for reading my thesis and for their valuable suggestions.

Table of Contents

Chapter 1. Introduction **1**

1.1 The linear model and the robust testing problem 2

1.2 Some robust GM–type testing procedures for linear models 3

1.3 Stochastic and Edgeworth expansions 6

1.3.1 Stochastic asymptotic expansions 6

1.3.2 Edgeworth expansions 7

Chapter 2. Stochastic Asymptotic Expansion of Robust GM-type Estimator **9**

2.1 Kronecker Products and the Calculus of Matrix Differentiation 10

2.1.1 Basic Notation 10

2.1.2 Some Basic Properties 11

2.2 Asymptotic expansion of GM–estimators 12

2.2.1 Preliminaries 12

2.2.2 Derivations 15

2.3 Examples 32

Chapter 3. A Statistic Related to Scores Type Test for Some Ordinary M–estimation Problems **37**

3.1 M–estimation of location parameter with scale known 38

3.1.1 The asymptotic distribution of W_n 38

3.1.2 A test statistic related to W_n 40

3.2 Linear regression M–estimation problem with scale unknown 44

Chapter 4. Simulation Study results 53

4.1 Location case with scale known 53

4.2 Simple linear regression case with scale unknown 58

References 64

Chapter 1. Introduction

Though systematic research on the problem of robust estimation started later than that on testing, still far less attention has been given to robust testing procedures. But the need for robust testing procedures is obvious: we cannot estimate robustly the parameters of a model and then use unmodified procedures to test hypotheses about these parameters. Some robust testing procedures were defined and investigated by Markatou, Stahel and Ronchetti (1991). All these test statistics turned out to be approximately χ_q^2–distributed for some number q. Also some higher order approximations in these cases have been derived by Field and Ronchetti (1990). Their approximations are sometimes amazingly accurate; the problem is that they are computationally very intensive and difficult to use in practice. It motivated us to find an easily explained and implemented modification to the normal theory test statistics, i.e. a 't' or 'F' statistic with the degrees of freedom modified to take into account the estimation method, perhaps with a scaling factor added.

We begin with a short introduction to the linear regression model and the robust GM–type estimation in Section 1. This section serves as a motivation which helps readers understand the robust regression model presented in this dissertation. Section 2 provides an overview of some common testing procedures following GM–estimation. In Section 3, some notations and results about (multiple) asymptotic expansions are reviewed since we will apply them to many of our derivations.

§1.1 The linear model and the robust testing problem

The linear model with which we are working is defined as follows. Suppose that the observed data $\{(\mathbf{x}_i, y_i), i = 1, 2, ..., n\}$, which are independently and identically distributed random variables, can be modeled as

$$Y = X\boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{1.1}$$

where $Y = (y_1, y_2, \cdots, y_n)^T \in \mathbf{R}^n$, the $n \times m$ design matrix X has rank $r \leq m$, $\boldsymbol{\theta} \in \Omega \subseteq \mathbf{R}^m$ is a m-vector of unknown parameters and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^T \in \mathbf{R}^n$ is the error.

The hypotheses of interest are

$$\begin{cases} H_0: & C\boldsymbol{\theta} = 0 \\ H_1: & C\boldsymbol{\theta} \neq 0 \end{cases} \tag{1.2}$$

for some $(m-q) \times m$ matrix C with rank $(m-q) \leq r$.

Under H_0, we have $\boldsymbol{\theta} \in (row(C))^{\perp}$, i.e. $\boldsymbol{\theta} = D_{m\times(r-m+q)}\boldsymbol{\delta}_{(r-m+q)\times 1}$ for some $\boldsymbol{\delta}$, where the columns of D form a basis for $(row(C))^{\perp}$. Thus $X\boldsymbol{\theta} = XD\boldsymbol{\delta}$ under H_0.

Now, let the columns of the $n \times (r-m+q)$ matrix Γ_1 be an orthogonal basis for $col(XD)$. Notice that we can extend it to $(\Gamma_{1_{n\times(r-m+q)}} \vdots \Gamma_{2_{n\times(m-q)}})$, an orthogonal basis for $col(X)$, and to $\Gamma = (\Gamma_{1_{n\times(r-m+q)}} \vdots \Gamma_{2_{n\times(m-q)}} \vdots \Gamma_{3_{n\times(n-r)}})$, an orthogonal basis for $\mathbf{R}^n$. Then (1.1) can be written as

$$\begin{aligned} Y =& \Gamma\Gamma^T X\boldsymbol{\theta} + \boldsymbol{\epsilon} \\ =& \Gamma \begin{pmatrix} \Gamma_1^T X\boldsymbol{\theta} \\ \Gamma_2^T X\boldsymbol{\theta} \\ \Gamma_3^T X\boldsymbol{\theta} \end{pmatrix} + \boldsymbol{\epsilon} \end{aligned}$$

where $\Gamma_3^T X = \mathbf{0}$, and under H_0, $\Gamma_2^T X\boldsymbol{\theta} = \Gamma_2^T XD\boldsymbol{\delta} = \mathbf{0}$.

Put $\boldsymbol{\phi}_1 = \Gamma_1^T X\boldsymbol{\theta}$, $\boldsymbol{\phi}_2 = \Gamma_2^T X\boldsymbol{\theta}$, $\boldsymbol{\phi} = \begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \end{pmatrix}$, $\Gamma_0 = (\Gamma_1 \vdots \Gamma_2)$, then

$$\begin{aligned} Y =& (\Gamma_1 \vdots \Gamma_2) \begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \end{pmatrix} + \boldsymbol{\epsilon} \\ =& \Gamma_0 \boldsymbol{\phi} + \boldsymbol{\epsilon} \end{aligned}$$

with $\Gamma_0^T \Gamma_0 = I_r$, and under H_0, $\boldsymbol{\phi}_2 = \mathbf{0}_{m\times 1}$.

Therefore, without lose of generality, one can always assume that $Y = X\boldsymbol{\theta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix} \begin{matrix} p \\ m-p \end{matrix}$ and $X^T X = I_m$. Furthermore, the hypotheses become

$$\begin{cases} H_0: & \boldsymbol{\theta}_2 = \mathbf{0}_{(m-p)\times 1}, \\ H_1: & \boldsymbol{\theta}_2 \neq \mathbf{0}_{(m-p)\times 1} \end{cases} \tag{1.3}$$

with $\boldsymbol{\theta}_1$ unspecified.

Robust tests usually rely on some GM–estimators of $\boldsymbol{\theta}$ and σ defined by

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n \eta(\mathbf{x}_i, \frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\theta}}_n}{\hat{\sigma}_n})\mathbf{x}_i = 0 \\ \frac{1}{n}\sum_{i=1}^n \chi(\frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\theta}}_n}{\hat{\sigma}_n}) = 0. \end{cases} \tag{1.4}$$

The function η is assumed to be continuous, piecewise differentiable, odd in r and $\eta(x, r) \geq 0$ if $r \geq 0$. The function χ is assumed to be continuous, piecewise differentiable and even. If, however, $\eta(x, r) = r$ and $\chi(r) = r^2 - \beta$ for some suitable β, then $\hat{\boldsymbol{\theta}}_n$ and $\hat{\sigma}_n$ are the least squares estimators.

§1.2 Some robust GM–type testing procedures for linear models

For hypotheses (1.3), three classes of tests have been introduced and investigated by Markatou, Stahel and Ronchetti (1991):

(1) The Wald type test uses a quadratic form of the second part $\hat{\boldsymbol{\theta}}_{n,2}$ of an $GM-$ estimator of $\boldsymbol{\theta}$,

$$W_{C,n}^2 = \hat{\boldsymbol{\theta}}_{n,2}^T C^{-1} \hat{\boldsymbol{\theta}}_{n,2}$$

as its test statistic. Here, C is a suitable positive definite $(m-p) \times (m-p)$ matrix, which will depend on the design. It is most naturally chosen to be an estimate of the covariance matrix V_n of $\hat{\boldsymbol{\theta}}_{n,2}$.

(2) The scores type test is based on the test statistic

$$R_{C,n}^2 = Z_n^T M_{(22.1)} C^{-1} M_{(22.1)} Z_n,$$

where $Z_n = \frac{1}{n}\sum_{i=1}^n \eta(\mathbf{x}_i, \frac{y_i - \mathbf{x}_{i,1}^T \hat{\hat{\theta}}_n}{\hat{\hat{\sigma}}_n})\mathbf{x}_{i,2}$ with $\mathbf{x}_i = \begin{pmatrix} \mathbf{x}_{i,1} \\ \mathbf{x}_{i,2} \end{pmatrix} \begin{matrix} p \\ m-p \end{matrix}$, and $\hat{\hat{\theta}}_{n,1}$, $\hat{\hat{\sigma}}_n$ are the $GM-$estimators obtained assuming H_0 to be true. C is again a suitable matrix and $M_{(22.1)} = M_{22} - M_{21} M_{11}^{-1} M_{12}$ with $M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = E[\eta^2(\mathbf{x}, r)\mathbf{x}\mathbf{x}^T]$.

(3) The drop-in-dispersion type test is given by a test statistic of the form:

$$S_{\tau,n}^2 = \frac{2}{n}\sum_{i=1}^n (\tau(\mathbf{x}_i, \frac{y_i - \mathbf{x}_{i,1}^T \hat{\hat{\boldsymbol{\theta}}}_n}{\hat{\hat{\sigma}}_n}) - \tau(\mathbf{x}_i, \frac{y_i - \mathbf{x}_i^T \hat{\boldsymbol{\theta}}_n}{\hat{\sigma}_n}))$$

where τ is such that $\tau(0) = 0$ and $\eta(x, r) = \frac{\partial}{\partial r}\tau(x, r)$, $\hat{\hat{\boldsymbol{\theta}}}_n$, $\hat{\hat{\sigma}}_n$ are the $GM-$estimators obtained when assuming H_0 is true, $\hat{\boldsymbol{\theta}}_n$, $\hat{\sigma}_n$ are the $GM-$estimators obtained from the full model.

By investigating the influence functions of these test statistics, one can show that:

(1) $nW_{C,n}^2$ is asymptotically χ_{m-p}^2–distributed if C is chosen to be an estimate of the covariance matrix.

(2) The scores type test statistic $R_{C,n}^2$ and the drop-in-dispersion type test statistic $S_{\tau,n}^2$ are asymptotically equivalent to $W_{C,n}^2$, if we choose C, τ properly. Thus they will also be asymptotically χ_{m-p}^2–distributed to the order $O(n^{-1})$.

Example 1.1: Consider the ordinary M–estimation of a location parameter with scale known:

$$\hat{\theta}_n = \text{a solution to } \sum_{i=1}^{n} \psi(y_i - \theta) = 0.$$

The null hypothesis of interest is:

$$H_0: \quad \theta = 0.$$

I) A natural Wald type test statistic is T_n^2 with $T_n = \frac{\hat{\theta}_n}{S(\hat{\theta}_n)}$, where

$$S^2(\hat{\theta}_n) = \frac{\frac{1}{n-1}\sum_{i=1}^{n} \psi^2(y_i - \hat{\theta}_n)}{\frac{1}{n}\sum_{i=1}^{n}(\psi'(y_i - \hat{\theta}_n))^2}$$

is an estimate of $\text{var}(\sqrt{n}\hat{\theta}_n)$. Note that nT_n^2 is approximately χ_1^2–distributed. If, however, $\psi(r) = r$, $\hat{\theta}_n$ then becomes the least squares estimate and $S^2(\hat{\theta}_n) = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{\theta}_n)^2$, therefore nT_n^2 is F_{n-1}^1–distributed since then $\sqrt{n}T_n$ is the ordinary normal theory t–statistic. In another words, we are using a χ_1^2 random variable to approximate an F_{n-1}^1 random variable under least squares.

II) A scores type test statistic is given by W_n^2 with $W_n = \frac{\frac{1}{n}\sum_{i=1}^{n}\psi(y_i)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\psi^2(y_i)}}$. It can be easily shown that under least squares,

$$nW_n^2 = \frac{F_{n-1}^1}{1 - \frac{1}{n} + \frac{1}{n}F_{n-1}^1}.$$

Once again, we are using a χ_1^2 random variable to approximate an F_{n-1}^1–related random variable.

This example motivated us to find some T_n^2–based or W_n^2–based test statistics which are approximately F–distributed in general case and exactly F–distributed under least squares and the normality assumption on the error distribution.

§1.3 Stochastic and Edgeworth expansions

Since the stochastic asymptotic expansions and Edgeworth expansions will be used frequently in this thesis, a brief review of these topics becomes necessary. More details can be found in Field and Ronchetti(1990).

§1.3.1 Stochastic asymptotic expansions

Let $\{Y_n\}$ be a sequence of continuous random variables and

$$Y_n = X_0 + b_{1n}X_1 + b_{2n}X_2 + \cdots + b_{mn}X_m + O_p(b_{(m+1)n}) \ , \tag{1.5}$$

where the distribution of $\{X_1, X_2, \cdots, X_m\}$ does not depend on n: $b_{1n} = a/\sqrt{n}$, $b_{2n} = b/n, \cdots$, or $b_{1n} = a/n$, $b_{2n} = b/n^2, \cdots$ for some constants $a, b, \cdots$. Usually (1.5) is called a stochastic expansion for Y_n.

An important question of interest is the relation between (1.5) and the asymptotic expansion of the corresponding characteristic function.

Example 1.2: Suppose that

$$Y_n = X_0 + \frac{1}{\sqrt{n}}X_1 + \frac{1}{2n}X_2 + O_p(n^{-3/2}),$$

then the characteristic function of Y_n becomes

$$\begin{aligned}
\xi(Y_n, t) &= E[e^{itY_n}] \\
&= E[e^{it(X_0 + \frac{1}{\sqrt{n}}X_1 + \frac{1}{2n}X_2 + O_p(n^{-3/2}))}] \\
&= E[e^{itX_0}(1 + \frac{itX_1}{\sqrt{n}} + \frac{1}{2n}(itX_2 - t^2X_1^2) + O_p(n^{-3/2}))] \\
&= E[e^{itX_0}] + \frac{1}{\sqrt{n}}E[e^{itX_0}itX_1] + \frac{1}{2n}E[e^{itX_0}(itX_2 - t^2X_1^2)] + O(n^{-3/2}),
\end{aligned}$$

provided $E(O_p(n^{-3/2})) = O(n^{-3/2})$. Thus it is possible for us to obtain the asymptotic characteristic function of Y_n from its stochastic expansion under certain conditions. Furthermore, we could also have the corresponding asymptotic density and distribution functions.

§1.3.2 Edgeworth expansions

Let S_n be a random variable with distribution function $F(x)$, characteristic function $\xi(t)$, cumulants κ_r, $r = 1, 2, \cdots$; let Y be a standard normal random variable with distribution function $\Phi(x)$, characteristic function $\eta(t)$, cumulants γ_r, $r = 1, 2, \cdots$. Recall that

$$\begin{cases} \kappa_r = (-i)^r \frac{d^r}{dt^r} \log \xi(t)|_{t=0} \\ \eta(t) = e^{-t^2/2} \\ \gamma_1 = 0, \ \gamma_2 = 1, \ \gamma_r = 0, \ r \geq 3. \end{cases} \tag{1.6}$$

Then by formal Taylor expansion we have

$$\log \frac{\xi(t)}{\eta(t)} = \sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(it)^r}{r!},$$

and

$$\xi(t) = \exp(\sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(it)^r}{r!}). \tag{1.7}$$

Furthermore, by Fourier inversion of (1.7), we obtain

$$H(x) = \exp(\sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(-D)^r}{r!}) \Phi(x), \tag{1.8}$$

where D denotes the differential operator.

If the terms in (1.8) can be collected according to the powers of some index n, then (1.7) and (1.8) form the Edgeworth expansion of S_n.

Example 1.3: Sum of n iid random variables

Let $X_1, \cdots, X_n$ be n iid random variables with distribution $F(x)$ and $E(X_1) = 0$, $\text{var}(X_1) = \sigma^2 > 0$, and cumulants $\beta_r(X_1) = \rho_r\sigma^r$, $r \geq 3$. Let $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i}{\sigma}$, with distribution function $F_n(x)$, characteristic $\xi(t)$, and cumulants κ_r. Then we have

$$\begin{cases} \kappa_1 = 0, \ \kappa_2 = 1, \\ \kappa_r = n^{-r/2}\sigma^{-r}n\beta_r(X_1) = \rho_r n^{-(r/2-1)}, \quad \text{for } r \geq 3 \ . \end{cases} \tag{1.9}$$

By applying (1.6) and (1.7), we have

$$\xi(t) = \exp(\sum_{r=3}^{\infty} \frac{\rho_r}{n^{r/2-1}}\frac{(it)^r}{r!})\exp(-t^2/2),$$

and by expanding $\exp(\sum_{r=3}^{\infty} \frac{\rho_r}{n^{r/2-1}}\frac{(it)^r}{r!})$, we obtain

$$\xi(t) = 1 + \frac{-i\rho_3}{6\sqrt{n}}t^3 - \frac{\rho_3^2}{72n}t^6 + \frac{\rho_4 t^4}{24n} + O(n^{-3/2}).$$

Finally, by applying (1.8), we obtain

$$F_n(x) = \Phi(x) - \phi(x)\{\frac{\rho_3 H_2(x)}{6\sqrt{n}} + \frac{\rho_4 H_3(x)}{24n} + \frac{\rho_3^2 H_5(x)}{72n}\} + O(n^{-3/2}). \tag{1.10}$$

therefore the density function of S_n is

$$f_n(x) = \phi(x)\{1 + \frac{\rho_3 H_3(x)}{6\sqrt{n}} + \frac{\rho_4 H_4(x)}{24n} + \frac{\rho_3^2 H_6(x)}{72n}\} + O(n^{-3/2}), \tag{1.11}$$

where the $H_r(x)$(the Hermite polynomials of order r), $r = 1, 2, \cdots$ are defined by

$$\phi(x)H_r(x) = (-1)^r \frac{d^r}{dx^r}\phi(x).$$

Notice that under certain assumptions, Edgeworth expansions can also be applied to nonlinear functions of sums. An example of this can be found in Barndorff-Nielsen and Cox (1989).

Chapter 2. Stochastic Asymptotic Expansion of Robust GM-type Estimator

We continue our discussion by considering the linear regression model as in §1.1. It is further assumed that

(I) the parameter space Ω is an open and convex set;

(II) the errors ϵ_i, $i = 1, 2, \cdots, n$ are independently identically distributed random variables with the symmetric distribution $G(\frac{\epsilon_1}{\sigma})$, where $\sigma > 0$ is a scale parameter;

(III) the design matrix X satisfies $X^T X = I$;

(IV) ϵ_i and $\mathbf{x}_i$ are independent .

For the hypotheses (1.3), the test statistic under investigation in this thesis is W_n^2 with

$$W_n = \frac{\sum_{i=1}^{n} \eta(\mathbf{x}_i, \frac{y_i - \mathbf{x}_{i,1}^T \hat{\boldsymbol{\theta}}_{n,1}}{\hat{\sigma}_n})\mathbf{x}_{i,2}}{\sqrt{\sum_{i=1}^{n} \eta^2(\mathbf{x}_i, \frac{y_i - \mathbf{x}_{i,1}^T \hat{\boldsymbol{\theta}}_{n,1}}{\hat{\sigma}_n})\mathbf{x}_{i,2}\mathbf{x}_{i,2}^T}}, \tag{2.1}$$

where $\hat{\boldsymbol{\theta}}_{n,1}$, $\hat{\sigma}_n$ are the GM estimators obtained when assuming H_0 in (1.3) is true. We choose W_n^2 as our subject because it has a much simpler form than that of the others.

In this chapter, we first review some notation and results related to Kronecker products and the calculus of matrix differentiation in Section 1. Then we apply them to the derivations of stochastic asymptotic expansions of the robust estimators in Section 2. Section 3 involves some examples of M–estimation problems.

§2.1 Kronecker Products and the Calculus of Matrix Differentiation

§2.1.1 Basic Notation

Definition 2.1 (Kronecker product) Let $A = (a_{ij})_{m\times n}$, $B = (b_{ij})_{p\times q}$, the Kronecker product $A \otimes B$ is defined as the $mp \times nq$ matrix

$$A \otimes B = (a_{ij}B) \ .$$

Definition 2.2 (*vec* operator) Let $A = (a_{ij})_{m\times n} = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)$, the vector operator *vec* of A is defined by

$$vec(A) = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{pmatrix} \ .$$

Definition 2.3 (*vec*–permutation matrix) Let $A = (a_{ij})_{m\times n}$, the matrix $I_{m,n}$ defined by

$$vec(A) = I_{m,n} \cdot vec(A^T)$$

is called a vec-permutation matrix.

Definition 2.4 (Matrix differentiation) Let X be an $m \times n$ matrix, and let Y be a $p \times q$ matrix, whose elements are functions of the elements of X. Let $\frac{\partial}{\partial X}$ be a matrix of derivative operator $\frac{\partial}{\partial x_{ij}}$. Then $\frac{\partial Y}{\partial X}$, the derivative of Y with respect to X is defined symbolically by

$$\frac{\partial Y}{\partial X} = vec(Y) \cdot (vec(\frac{\partial}{\partial X}))^T.$$

§2.1.2 Some Basic Properties

Some basic properties related to Kronecker products and matrix differentiation are listed below.

P1: $vec(A_{m\times n} \cdot B_{n\times p}) = (I_p \otimes A) \cdot vec(B) = (B^T \otimes I_m) vec(A)$;

P2: $(AC) \otimes (BD) = (A \otimes B)(C \otimes D)$;

P3: $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$;

P4: $(A \otimes B)^T = A^T \otimes B^T$;

P5: $vec(\mathbf{a}\mathbf{b}^T) = \mathbf{b} \otimes \mathbf{a}$;

P6: $B_{p\times q} \otimes A_{m\times n} = I_{m,p}(A \otimes B) I_{q,n}$;

P7: $I_{m,n} \cdot I_{n,m} = I_{mn}$;

P8: $I_{m,1} = I_{1,m} = I_m$;

P9: $(B \otimes A) \cdot vec(X) = vec(AXB^T)$;

P10: $\frac{\partial(Y_{p\times q} \cdot Z_{q\times r})}{\partial X} = (Z^T \otimes I_p)\frac{\partial Y}{\partial X} + (I_r \otimes Y)\frac{\partial Z}{\partial X}$;

P11: If $Z = Z(Y_1, Y_2, ..., Y_m)$ and $Y_i = Y_i(X)$, then $\frac{\partial Z}{\partial X} = \sum_{i=1}^{m} \frac{\partial Z}{\partial Y_i} \cdot \frac{\partial Y_i}{\partial X}$;

P12: $vec(Y_{m\times n} \otimes Z_{p\times s}) = (I_n \otimes I_{m,s} \otimes I_p) \cdot (vec(Y) \otimes vec(Z))$;

P13: $\frac{\partial}{\partial X}(Y_{m\times n} \otimes Z_{p\times s}) = (I_n \otimes I_{m,s} \otimes I_p) \cdot (vec(Y) \otimes \frac{\partial Z}{\partial X} + \frac{\partial Y}{\partial X} \otimes vec(Z))$;

P14: $(\mathbf{b}_p^T \otimes A_{m\times n})\mathbf{a}_{np} = (\mathbf{a}_{np}^T \otimes I_m)(\mathbf{b}_p \otimes vec(A_{m\times n}))$;

P15: $(A_{m\times pt} \otimes \mathbf{a}_s^T)(\mathbf{b}_p \otimes vec(B_{s\times t})) = A_{m\times pt}(\mathbf{b}_p \otimes B_{s\times t}^T)\mathbf{a}_s$;

P16: If $I_q = (\mathbf{e}_1, \mathbf{e}_2, ... \mathbf{e}_q)$, then $vec(I_q) = \sum_{i=1}^{q} \mathbf{e}_i \otimes \mathbf{e}_i$.

More properties of the Kronecker products and the matrix differentiation may be found in Wiens(1985), Graham(1981), Henderson and Searle(1979).

§2.2 Asymptotic expansion of GM–estimators

§2.2.1 Preliminaries

Put $\boldsymbol{\xi}_n = (\hat{\boldsymbol{\theta}}_n^T, \hat{\sigma}_n)$, $\boldsymbol{\xi}_0 = (\boldsymbol{\theta}_0^T, \sigma)^T$, where $\hat{\boldsymbol{\theta}}_n$, $\hat{\sigma}_n$ are the GM–estimators defined by (1.4), and $\boldsymbol{\theta}_0$, σ are the true parameters defined by

$$\begin{cases} E[\eta(\mathbf{x}, \frac{y-\mathbf{x}^T\boldsymbol{\theta}_0}{\sigma})\mathbf{x}] = \mathbf{0} \\ E[\chi(\frac{y-\mathbf{x}^T\boldsymbol{\theta}_0}{\sigma})] = 0 \end{cases} \tag{2.2}$$

Notice that under the null hypothesis in (1.3), this is

$$\begin{cases} E[\eta(\mathbf{x}, \frac{y-\mathbf{x}_1^T\boldsymbol{\theta}_{10}}{\sigma})\mathbf{x}_1] = \mathbf{0} \\ E[\chi(\frac{y-\mathbf{x}_1^T\boldsymbol{\theta}_{10}}{\sigma})] = 0, \end{cases} \tag{2.3}$$

with $\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \begin{matrix} p \\ m-p \end{matrix}$ and $\boldsymbol{\theta}_0 = \begin{pmatrix} \boldsymbol{\theta}_{10} \\ \mathbf{0} \end{pmatrix} \begin{matrix} p \\ m-p \end{matrix}$.

Put

$$\eta_{ij}(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi}_n) = \eta(\mathbf{x}_i, \frac{\epsilon_i - \mathbf{x}_{i,1}^T(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)}{\frac{\hat{\sigma}_n}{\sigma}\sigma})x_{ij}, \quad i = 1, \cdots, n, \ j = 1, \cdots, p,$$

$$\eta_{i,p+1}(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi}_n) = \chi(\frac{\epsilon_i - \mathbf{x}_{i,1}^T(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)}{\frac{\hat{\sigma}_n}{\sigma}\sigma}), \quad i = 1, \cdots, n,$$

$$\boldsymbol{\eta}_i(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi}_n) = (\eta_{i1}(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi}_n), \cdots, \eta_{iq}(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi}_n))^T, \quad \text{with} \quad q = p+1,$$

then under H_0,(1.4),(2.2) can be rewritten as

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n \boldsymbol{\eta}_i(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi}_n) = 0 \\ E_{\boldsymbol{\xi}_0}[\boldsymbol{\eta}_1(\mathbf{x}, \epsilon, \boldsymbol{\xi}_0)] = \mathbf{0}. \end{cases} \tag{2.4}$$

We first list several assumptions made on (2.4)(Bhattacharya and Ghosh, 1978).

Suppose $s \geq 3$, $\boldsymbol{\xi} \in \Theta$.

A1) There is an open subset U of $\mathbf{R}^q$, such that

I) for each $\boldsymbol{\xi} \in \Theta$, one has $K_{\boldsymbol{\xi}}(U) = 1$, where $K_{\boldsymbol{\xi}}$ is the distribution function of y_i given $\boldsymbol{\xi} = (\boldsymbol{\theta}^T, \sigma)^T$;

II) $\boldsymbol{\eta}_i$ has a ν^{th} derivative with respect to $\boldsymbol{\xi}$ for each $|\nu| \le s$.

A2) For each compact $K \subset \Theta$, and $0 \le \nu \le s-1$,

$$\sup_{\xi_0 \in K} E_{\xi_0}[\|vec\frac{\partial^\nu}{\partial \boldsymbol{\xi}^\nu}\boldsymbol{\eta}_1(\mathbf{x}, \epsilon, \boldsymbol{\xi})|_{\xi=\xi_0}\|^s] < \infty,$$

and for each compact $K \subset \Theta$, there exists $\epsilon > 0$, such that

$$\sup_{\xi_0 \in K} E_{\xi_0}[(\max_{\|\xi-\xi_0\| \le \epsilon} \|vec\frac{\partial^s}{\partial \boldsymbol{\xi}^s}\boldsymbol{\eta}_1(\mathbf{x}, \epsilon, \boldsymbol{\xi})\|)^s] < \infty.$$

A3) For each $\boldsymbol{\xi}_0 \in \Theta$, the matrices

$$A_0 = \sigma E_{\xi_0}[\frac{\partial \boldsymbol{\eta}_1(\mathbf{x}, \epsilon, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}}|_{\xi=\xi_0}]$$

and

$$E_{\xi_0}[\boldsymbol{\eta}_1(\mathbf{x}, \epsilon, \boldsymbol{\xi}_0)\boldsymbol{\eta}_1^T(\mathbf{x}, \epsilon, \boldsymbol{\xi}_0)]$$

are nonsingular.

A4) The functions $I(\boldsymbol{\xi}) = -\frac{1}{\sigma}A_0(\boldsymbol{\xi})$ and for $1 \le \nu, \nu' \le s$,

$$E_{\xi}[vec(\frac{\partial^{\nu-1}}{\partial \boldsymbol{\xi}^{\nu-1}}\boldsymbol{\eta}_1(\mathbf{x}, \epsilon, \boldsymbol{\xi}))vec(\frac{\partial^{\nu'-1}}{\partial \boldsymbol{\xi}^{\nu'-1}}\boldsymbol{\eta}_1(\mathbf{x}, \epsilon, \boldsymbol{\xi}))^T]$$

are continuous on Θ.

A5) The map $\boldsymbol{\xi} \to K_{\xi}$ on Θ into the space of all probability measures on $\mathbf{R}^q$ is continuous when the latter space is given the (variation) norm topology.

A6) For each $\boldsymbol{\xi} \in \Theta$, $K_{\boldsymbol{\xi}}$ has a nonzero absolutely continuous component whose density has a version $k(y; \boldsymbol{\xi})$, which is strictly positive on U.

Theorem 2.1 (Bhattacharya and Ghosh, 1978) Assume **A1)–A6)** hold. Then there is a sequence of statistics $\{\boldsymbol{\xi}_n\}$ such that for every compact $K \subset \Theta$ and $\boldsymbol{\theta}_0 \in K$,

$$0 = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\eta}_i(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi}_n)$$
$$= \frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{\eta}_i(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi}_0) + \sum_{\nu=1}^{3}((\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^{[\nu]T} \otimes I_q)\frac{vec(\sum_{i=1}^{n} \frac{\partial^\nu}{\partial \xi^\nu}\boldsymbol{\eta}_i(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi})|_{\xi=\xi_0})}{\nu!}] + o_p(n^{-2+\epsilon}) , \quad (2.5)$$

where $(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^{[1]} = \boldsymbol{\xi}_n - \boldsymbol{\xi}_0$, $(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^{[\nu]} = (\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^{[\nu-1]} \otimes (\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)$ for $\nu = 2, 3$.

Now, for $0 \le \nu \le 3$, $1 \le i \le n$, write

$$\mathbf{U}_{\nu,i} = vec(\frac{\partial^\nu}{\partial \boldsymbol{\xi}^\nu}\boldsymbol{\eta}_i(\mathbf{x}_i, \epsilon_i, \boldsymbol{\xi})|_{\xi=\xi_0}),$$
$$\overline{\mathbf{U}}_\nu = \frac{1}{n}\sum_{i=1}^{n} \mathbf{U}_{\nu,i},$$
$$\overline{\mathbf{U}}^T = (\overline{\mathbf{U}}_0^T, \overline{\mathbf{U}}_1^T, \overline{\mathbf{U}}_2^T, \overline{\mathbf{U}}_3^T),$$
$$\mathbf{a}^T = E[\overline{\mathbf{U}}^T] = (\boldsymbol{\mu}_0^T, \boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, \boldsymbol{\mu}_3^T),$$
$$Q = \sum_{i=1}^{4} q^i,$$

then $\mathbf{U}_{\nu,i}$, $\overline{\mathbf{U}}_\nu$ are $q^{\nu+1} \times 1$ vectors, $\overline{\mathbf{U}}$ and $\mathbf{a}$ are $Q \times 1$ vectors, $\boldsymbol{\mu}_0 = E[\overline{\mathbf{U}}_0] = \mathbf{0}_{q\times 1}$ and (2.5) can be rewritten as

$$\mathbf{0}_{q\times 1} = \overline{\mathbf{U}}_0 + \sum_{\nu=1}^{3}\frac{((\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^{[\nu]T} \otimes I_q)\overline{\mathbf{U}}_\nu}{\nu!} + o_p(n^{-2+\epsilon}). \quad (2.6)$$

Furthermore, define $\mathbf{s}_q : \mathbf{R}^Q \times \mathbf{R}^q \to \mathbf{R}^q$ as

$$\mathbf{s}_q(\mathbf{u}, \mathbf{t}) = \mathbf{u}_0 + \sum_{\nu=1}^{3}\frac{((\mathbf{t} - \boldsymbol{\xi}_0)^{[\nu]T} \otimes I_q)\mathbf{u}_\nu}{\nu!}, \quad (2.7)$$

where $\mathbf{u}^T = (\mathbf{u}_0^T, \mathbf{u}_1^T, \mathbf{u}_2^T, \mathbf{u}_3^T)$ is a $1 \times Q$ vector. Notice that $\mathbf{s}_q(\mathbf{a}, \boldsymbol{\xi}_0) = \mathbf{0}_{q\times 1}$, then by the Implicit Function Theorem, there is a function $H : \mathbf{U} \in \mathbf{R}^Q \to H(\mathbf{U}) \in \mathbf{R}^q$ such that $H(\mathbf{a}) = \boldsymbol{\xi}_0$ and $\mathbf{s}_q(\mathbf{u}, H(\mathbf{u})) = \mathbf{0}$ in a neighbourhood of $\mathbf{a}$.

Following Field and Ronchetti(1990) we have:

Lemma 2.1 $H(\overline{\mathbf{U}}) - \boldsymbol{\xi}_n = o_p(n^{-2+\epsilon})$ for any $\epsilon > 0$.

Proof: cf expression 2.39 and 2.40 of Bhattacharya and Ghosh (1978).

The next step is to expand H in a Taylor series expansion about $\mathbf{a}$. The result is the following expression:

$$\begin{aligned}\boldsymbol{\xi}_n = H(\mathbf{a}) &+ ((\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q) vec(H'(\mathbf{a})) \\ &+ ((\overline{\mathbf{U}} - \mathbf{a})^{[2]^T} \otimes I_q) vec(\frac{H''(\mathbf{a})}{2}) \\ &+ ((\overline{\mathbf{U}} - \mathbf{a})^{[3]^T} \otimes I_q) vec(\frac{H'''(\mathbf{a})}{6}) + o_p(n^{-2+\epsilon}). \end{aligned} \qquad (2.8)$$

Putting this equation together with Lemma 2.1, we have that for every $\boldsymbol{\theta}_0$ in a compact subset of Θ,

$$\begin{aligned}\sqrt{n}(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0) &= \sqrt{n}\sum_{\nu=1}^{3}((\overline{\mathbf{U}} - \mathbf{a})^{[\nu]^T} \otimes I_q) vec(\frac{H^{(\nu)}(\mathbf{a})}{\nu!}) + o_p(n^{(-3/2+\epsilon)}) \\ &= \sqrt{n}\sum_{\nu=1}^{3}((\overline{\mathbf{U}} - \mathbf{a})^{[\nu-1]^T} \otimes I_q) \frac{H^{(\nu)}(\mathbf{a})}{\nu!}(\overline{\mathbf{U}} - \mathbf{a}) + o_p(n^{(-3/2+\epsilon)}). \end{aligned} \qquad (2.9)$$

However, (2.9) cannot be used directly unless we are able to evaluate $H^{(\nu)}(\mathbf{a})$ for $\nu = 1, 2, 3$ in some way. The next section will then focus on the derivations of the first three derivatives of H about $\mathbf{a}$. Later a simplified version of (2.9) will be provided.

§2.2.2 Derivations

We first list some notation which will be used in this section.

Definition 2.5 Define

$$\begin{cases} Z_n = (\overline{\mathbf{U}}_{0_{q\times 1}} \vdots A_{n_{q\times q}} \vdots B_{n_{q\times q^2}} \vdots C_{n_{q\times q^3}}) \,, \\ Z_0 = E[Z_n] = (\mathbf{0}_{q\times 1} \vdots A_{0_{q\times q}} \vdots B_{0_{q\times q^2}} \vdots C_{0_{q\times q^3}}) \end{cases}$$

by

$$\overline{\mathbf{U}} = \begin{pmatrix} \overline{\mathbf{U}}_0 \\ \overline{\mathbf{U}}_1 \\ \overline{\mathbf{U}}_2 \\ \overline{\mathbf{U}}_3 \end{pmatrix} = vec(Z_n) = \begin{pmatrix} vec(\overline{\mathbf{U}}_0) \\ vec(A_n) \\ vec(B_n) \\ vec(C_n) \end{pmatrix},$$

and

$$\mathbf{a} = \begin{pmatrix} \mathbf{0}_{q\times 1} \\ \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_3 \end{pmatrix} = vec(Z_0) = \begin{pmatrix} vec(\mathbf{0}_{q\times 1}) \\ vec(A_0) \\ vec(B_0) \\ vec(C_0) \end{pmatrix},$$

or equivalently by

$$\begin{cases} A_n = \frac{1}{n}\sum_{i=1}^n(\boldsymbol{\eta}_i(\mathbf{x}_i,\epsilon_i,\boldsymbol{\xi})(\frac{\partial}{\partial\boldsymbol{\xi}})^T|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}), & A_0 = E[\boldsymbol{\eta}_1(\mathbf{x},\epsilon,\boldsymbol{\xi})(\frac{\partial}{\partial\boldsymbol{\xi}})^T|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}]\,, \\ B_n = \frac{1}{n}\sum_{i=1}^n(\boldsymbol{\eta}_i(\mathbf{x}_i,\epsilon_i,\boldsymbol{\xi})(\frac{\partial}{\partial\boldsymbol{\xi}})^{[2]^T}|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}), & B_0 = E[\boldsymbol{\eta}_1(\mathbf{x},\epsilon,\boldsymbol{\xi})(\frac{\partial}{\partial\boldsymbol{\xi}})^{[2]^T}|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}]\,; \\ C_n = \frac{1}{n}\sum_{i=1}^n(\boldsymbol{\eta}_i(\mathbf{x}_i,\epsilon_i,\boldsymbol{\xi})(\frac{\partial}{\partial\boldsymbol{\xi}})^{[3]^T}|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}), & C_0 = E[\boldsymbol{\eta}_1(\mathbf{x},\epsilon,\boldsymbol{\xi})(\frac{\partial}{\partial\boldsymbol{\xi}})^{[3]^T}|_{\boldsymbol{\xi}=\boldsymbol{\xi}_0}]\,. \end{cases} \tag{2.10}$$

Next, we continue our discussion by evaluating (2.9) term by term. Define

$$\mathbf{l}(\mathbf{t}) = \begin{pmatrix} 1 \\ \mathbf{t}-\boldsymbol{\xi}_0 \\ \frac{(\mathbf{t}-\boldsymbol{\xi}_0)^{[2]}}{2} \\ \frac{(\mathbf{t}-\boldsymbol{\xi}_0)^{[3]}}{3!} \end{pmatrix}, \quad L_\nu(\mathbf{t}) = \frac{\partial^\nu}{\partial\mathbf{t}^\nu}\mathbf{l}(\mathbf{t})\ , \quad \nu = 1,2,3,$$

then

$$\mathbf{s}_q(\mathbf{u},\mathbf{t}) = (\mathbf{l}^T(\mathbf{t})\otimes I_q)\mathbf{u}, \tag{2.11}$$

and $H(\mathbf{u}) = \mathbf{t}$ is defined by $\mathbf{s}_q(\mathbf{u},\mathbf{t}) = \mathbf{0}$ in a neighbourhood of $\mathbf{u} = \mathbf{a}$. Notice that $H(\mathbf{a}) = \boldsymbol{\xi}_0$.

$\nu = 1$:

Differentiating equation (2.11) with respect to $\mathbf{u}$ once gives

$$\mathbf{0}_{q\times Q} = \mathbf{l}^T(\mathbf{t}) \otimes I_q + (\mathbf{u}^T \otimes I_q)(L_1(\mathbf{t}) \otimes vec(I_q))\frac{\partial \mathbf{t}}{\partial \mathbf{u}}. \tag{2.12}$$

With some algebra, it can be shown that

$$L_1(\mathbf{t}) = \frac{\partial \mathbf{l}(\mathbf{t})}{\partial \mathbf{t}} = Q_3 \begin{pmatrix} \mathbf{0}_{1\times q} \\ I_q \\ (\mathbf{t} - \boldsymbol{\xi}_0) \otimes I_q \\ \frac{1}{2}(\mathbf{t} - \boldsymbol{\xi}_0)^{[2]} \otimes I_q \end{pmatrix},$$

where

$$Q_3 = \begin{pmatrix} 1 & \mathbf{0}_{1\times q} & \mathbf{0}_{1\times q^2} & \mathbf{0}_{1\times q^3} \\ \mathbf{0}_{q\times 1} & I_q & \mathbf{0}_{q\times q^2} & \mathbf{0}_{q\times q^3} \\ \mathbf{0}_{q^2\times 1} & \mathbf{0}_{q^2\times q} & \frac{1}{2}(I_{q^2} + I_{q,q}) & \mathbf{0}_{q^2\times q^3} \\ \mathbf{0}_{q^3\times 1} & \mathbf{0}_{q^2\times q} & \mathbf{0}_{q^2\times q^2} & \frac{1}{3}(I_{q^3} + (I_q \otimes I_{q,q}) + I_{q^2,q}) \end{pmatrix}.$$

Thus from (2.12) we have

$$\frac{\partial}{\partial \mathbf{u}} H(\mathbf{u}) = \frac{\partial \mathbf{t}}{\partial \mathbf{u}} = -B^{-1}(\mathbf{u}, \mathbf{t})(\mathbf{l}^T(\mathbf{t}) \otimes I_q) = -\mathbf{l}^T(\mathbf{t}) \otimes B^{-1}(\mathbf{u}, \mathbf{t}) . \tag{2.13}$$

with

$$B(\mathbf{u}, \mathbf{t}) = (\mathbf{u}^T \otimes I_q)[L_1(\mathbf{t}) \otimes vec(I_q)]. \tag{2.14}$$

When (2.13) is evaluated at $\mathbf{u} = \mathbf{a}$, $\mathbf{t} = \boldsymbol{\xi}_0$, we have

$$\begin{aligned} B(\mathbf{a}, \boldsymbol{\xi}_0) &= (\mathbf{0}_{1\times q} \otimes I_q \vdots \boldsymbol{\mu}_1^T \otimes I_q \vdots \boldsymbol{\mu}_2^T \otimes I_q \vdots \boldsymbol{\mu}_3^T \otimes I_q) \left(\begin{pmatrix} \mathbf{0}_{1\times q} \\ I_q \\ \mathbf{0}_{q^2\times q} \\ \mathbf{0}_{q^3\times q} \end{pmatrix} \otimes vec(I_q) \right) \\ &= (\boldsymbol{\mu}_1^T \otimes I_q)(I_q \otimes vec(I_q)) \\ &= (vec^T(A_0) \otimes I_q)(I_q \otimes vec(I_q)) \\ &= A_0 , \end{aligned}$$

and

$$\mathbf{1}^T(\boldsymbol{\xi}_0) \otimes I_q = (I_q \vdots \mathbf{0}_{q\times q^2} \vdots \mathbf{0}_{q\times q^3} \vdots \mathbf{0}_{q\times q^4}) .$$

Therefore

$$H'(\mathbf{a})_{q\times Q} = (-A_0^{-1} \vdots \mathbf{0}_{q\times q^2} \vdots \mathbf{0}_{q\times q^3} \vdots \mathbf{0}_{q\times q^4}) . \tag{2.15}$$

Now by applying (2.15), (2.9) becomes

$$\begin{aligned}\sqrt{n}(\xi_n - \xi_0) = &- A_0^{-1}(\sqrt{n}\,\overline{\mathbf{U}}_0) \\ &+ \sqrt{n}\sum_{\nu=2}^{3}((\overline{\mathbf{U}} - \mathbf{a})^{[\nu-1]^T} \otimes I_q)\frac{H^{(\nu)}(\mathbf{a})}{\nu!}(\overline{\mathbf{U}} - \mathbf{a}) + o_p(n^{-3/2+\epsilon}),\end{aligned} \tag{2.16}$$

where

$$\overline{\mathbf{U}}_0 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \eta(\mathbf{x}_i, \frac{\epsilon_i}{\sigma})\mathbf{x}_{i,1} \\ \frac{1}{n}\sum_{i=1}^n \chi(\frac{\epsilon_i}{\sigma}) \end{pmatrix},$$

$$A_0 = E[\eta(\mathbf{x}, \epsilon, \boldsymbol{\xi}_0)\mathbf{x}(\frac{\partial}{\partial \boldsymbol{\xi}})^T] = -\frac{1}{\sigma}\begin{pmatrix} E[\eta'(\mathbf{x}, \frac{\epsilon}{\sigma})\mathbf{x}_1\mathbf{x}_1^T] & E[\frac{\epsilon}{\sigma}\eta'(\mathbf{x}, \frac{\epsilon}{\sigma})\mathbf{x}_1] \\ E[\chi'(\frac{\epsilon}{\sigma})\mathbf{x}_1^T] & E[\chi'(\frac{\epsilon}{\sigma})] \end{pmatrix}.$$

$\nu = 2$:

Since

$$\begin{cases} \frac{\partial}{\partial \mathbf{u}} vec(B^{-1}(\mathbf{u}, \mathbf{t})|_{\mathbf{t}\text{ fixed}}) = -[B^{-T}(\mathbf{u}, \mathbf{t}) \otimes B^{-1}(\mathbf{u}, \mathbf{t})]\ \frac{\partial}{\partial \mathbf{u}} vec(B(\mathbf{u}, \mathbf{t})|_{\mathbf{t}\text{ fixed}}), \\ \frac{\partial}{\partial \mathbf{t}} vec(B^{-1}(\mathbf{u}, \mathbf{t})|_{\mathbf{u}\text{ fixed}}) = -[B^{-T}(\mathbf{u}, \mathbf{t}) \otimes B^{-1}(\mathbf{u}, \mathbf{t})]\ \frac{\partial}{\partial \mathbf{t}} vec(B(\mathbf{u}, \mathbf{t})|_{\mathbf{u}\text{ fixed}}), \end{cases} \tag{2.17}$$

differentiating (2.13) with respect to $\mathbf{u}$ gives

$$\begin{aligned}
\frac{\partial^2}{\partial \mathbf{u}^2}H(\mathbf{u}) &= \frac{\partial^2 \mathbf{t}}{\partial \mathbf{u}^2} = -\frac{\partial}{\partial \mathbf{u}}[B^{-1}(\mathbf{u},\mathbf{t})(\mathbf{l}^T(\mathbf{t}) \otimes I_q)] \\
&= -(\mathbf{l}(\mathbf{t}) \otimes I_q \otimes I_q)[\frac{\partial}{\partial \mathbf{u}} vec(B^{-1}(\mathbf{u},\mathbf{t})|_{\mathbf{t}} \text{ fixed}) \\
&\quad + \frac{\partial}{\partial \mathbf{t}} vec(B^{-1}(\mathbf{u},\mathbf{t})|_{\mathbf{u}} \text{ fixed})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}] - (I_Q \otimes B^{-1}(\mathbf{u},\mathbf{t}))(\frac{\partial \mathbf{l}(\mathbf{t})}{\partial \mathbf{u}} \otimes vec(I_q)) \\
&= (\mathbf{l}(\mathbf{t}) \otimes I_q \otimes I_q)(B^{-T}(\mathbf{u},\mathbf{t}) \otimes B^{-1}(\mathbf{u},\mathbf{t}))(\frac{\partial}{\partial \mathbf{u}} vec(B(\mathbf{u},\mathbf{t})|_{\mathbf{t}} \text{ fixed}) \\
&\quad + \frac{\partial}{\partial \mathbf{t}} vec(B(\mathbf{u},\mathbf{t})|_{\mathbf{u}} \text{ fixed})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) - (I_Q \otimes B^{-1}(\mathbf{u},\mathbf{t}))[(L_1(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) \otimes vec(I_q)] \\
&= (\mathbf{l}(\mathbf{t}) \otimes B^{-T}(\mathbf{u},\mathbf{t}) \otimes B^{-1}(\mathbf{u},\mathbf{t}))(\frac{\partial}{\partial \mathbf{u}} vec(B(\mathbf{u},\mathbf{t})|_{\mathbf{t}} \text{ fixed}) \\
&\quad + \frac{\partial}{\partial \mathbf{t}} vec(B(\mathbf{u},\mathbf{t})|_{\mathbf{u}} \text{ fixed})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) - (L_1(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t})). \qquad (2.18)
\end{aligned}$$

Now, from (2.14) we have

$$\begin{aligned}
I) \quad \frac{\partial}{\partial \mathbf{u}} vec(B(\mathbf{u},\mathbf{t})|_{\mathbf{t}} \text{ fixed}) &= \frac{\partial}{\partial \mathbf{u}}\{[L_1^T(\mathbf{t}) \otimes vec(I_q)^T \otimes I_q](\mathbf{u}^T \otimes vec(I_q))|_{\mathbf{t}} \text{ fixed}\} \\
&= (L_1^T(\mathbf{t}) \otimes vec(I_q)^T \otimes I_q)(I_Q \otimes vec(I_q)) \\
&= (L_1^T(\mathbf{t}) I_{\frac{Q}{q}}) \otimes [(vec(I_q)^T \otimes I_q)(I_q \otimes vec(I_q))] \\
&= L_1^T(\mathbf{t}) \otimes [(\sum_i \mathbf{e}_i^T \otimes \mathbf{e}_i^T \otimes I_q)(\sum_j I_q \otimes \mathbf{e}_j \otimes \mathbf{e}_j)] \\
&= L_1^T(\mathbf{t}) \otimes \{\sum_i \sum_j (\mathbf{e}_i^T \cdot I_q) \otimes [(\mathbf{e}_i^T \otimes I_q)(\mathbf{e}_i \otimes \mathbf{e}_j)]\} \\
&= L_1^T(\mathbf{t}) \otimes (\sum_i \mathbf{e}_i^T \otimes \mathbf{e}_i) \\
&= L_1^T(\mathbf{t}) \otimes I_q,
\end{aligned}$$

and

$$\begin{aligned}
&[\mathbf{l}(\mathbf{t}) \otimes B^{-T}(\mathbf{u},\mathbf{t}) \otimes B^{-1}(\mathbf{u},\mathbf{t})]\frac{\partial}{\partial \mathbf{u}} vec(B(\mathbf{u},\mathbf{t})|_{\mathbf{t}} \text{ fixed}) \\
&= \{[\mathbf{l}(\mathbf{t}) \otimes B^{-T}(\mathbf{u},\mathbf{t})]L_1^T(\mathbf{t})\} \otimes B^{-1}(\mathbf{u},\mathbf{t}) \\
&= \mathbf{l}(\mathbf{t}) \otimes [B^{-T}(\mathbf{u},\mathbf{t})L_1^T(\mathbf{t})] \otimes B^{-1}(\mathbf{u},\mathbf{t}) \;;
\end{aligned}$$

$$II) \quad \frac{\partial}{\partial \mathbf{t}} vec(B(\mathbf{u},\mathbf{t})|_{\mathbf{u}\text{ fixed}}) = \frac{\partial}{\partial \mathbf{t}}[(I_q \otimes \mathbf{u}^T \otimes I_q)(vecL_1(\mathbf{t}) \otimes vec(I_q))|_{\mathbf{u}\text{ fixed}}]$$
$$= (I_q \otimes \mathbf{u}^T \otimes I_q)[\frac{\partial}{\partial \mathbf{t}} vec(L_1(\mathbf{t})) \otimes vec(I_q)] \ ,$$

and

$$[\mathbf{l}(\mathbf{t}) \otimes B^{-T}(\mathbf{u},\mathbf{t}) \otimes B^{-1}(\mathbf{u},\mathbf{t})]\frac{\partial}{\partial \mathbf{t}} vec(B(\mathbf{u},\mathbf{t})|_{\mathbf{u}\text{ fixed}})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}$$
$$= \{[(\mathbf{l}(\mathbf{t}) \otimes B^{-T}(\mathbf{u},\mathbf{t}))(I_q \otimes \mathbf{u}^T)] \otimes B^{-1}(\mathbf{u},\mathbf{t})\}[\frac{\partial}{\partial \mathbf{t}} vec(L_1(\mathbf{t})) \otimes vec(I_q)]\frac{\partial \mathbf{t}}{\partial \mathbf{u}}$$
$$= [\mathbf{l}(\mathbf{t}) \otimes B^{-T}(\mathbf{u},\mathbf{t}) \otimes \mathbf{u}^T \otimes B^{-1}(\mathbf{u},\mathbf{t})][\frac{\partial}{\partial \mathbf{t}} vec(L_1(\mathbf{t})) \otimes vec(I_q)]\frac{\partial \mathbf{t}}{\partial \mathbf{u}} \ .$$

Thus (2.18) can be rewritten as

$$\frac{\partial^2 \mathbf{t}}{\partial \mathbf{u}^2} = \mathbf{l}(\mathbf{t}) \otimes [B^{-T}(\mathbf{u},\mathbf{t})L_1^T(\mathbf{t})] \otimes B^{-1}(\mathbf{u},\mathbf{t}) - (L_1(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t}))$$
$$+ [\mathbf{l}(\mathbf{t}) \otimes B^{-T}(\mathbf{u},\mathbf{t}) \otimes \mathbf{u}^T \otimes B^{-1}(\mathbf{u},\mathbf{t})][\frac{\partial}{\partial \mathbf{t}}(vec(L_1(\mathbf{t})) \otimes vec(I_q))]\frac{\partial \mathbf{t}}{\partial \mathbf{u}} \ . \quad (2.19)$$

However, since

$$\mathbf{l}(\mathbf{t}) \otimes [B^{-T}(\mathbf{u},\mathbf{t})L_1^T(\mathbf{t})[= (\mathbf{l}(\mathbf{t}) \otimes I_q)B^{-T}(\mathbf{u},\mathbf{t})L_1^T(\mathbf{t})$$
$$= -(L_1(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T,$$
$$\mathbf{l}(\mathbf{t}) \otimes B^{-T}(\mathbf{u},\mathbf{t}) = (\mathbf{l}(\mathbf{t}) \otimes I_q)B^{-T}(\mathbf{u},\mathbf{t}) = -(\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T,$$

(2.19) becomes

$$\frac{\partial^2 \mathbf{t}}{\partial \mathbf{u}^2} = -(L_1(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T \otimes B^{-1}(\mathbf{u},\mathbf{t}) - (L_1(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t}))$$
$$- [(\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T \otimes \mathbf{u}^T \otimes B^{-1}(\mathbf{u},\mathbf{t})][(\frac{\partial L_1(\mathbf{t})}{\partial \mathbf{t}}\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) \otimes vec(I_q)]$$
$$= -(L_1(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T \otimes B^{-1}(\mathbf{u},\mathbf{t}) - (L_1(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t}))$$
$$- [(\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T \otimes (B^{-1}(\mathbf{u},\mathbf{t})Z_n)](L_2(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) \ , \quad (2.20)$$

where

$$L_2(\mathbf{t}) = \frac{\partial}{\partial \mathbf{t}} L_1(\mathbf{t}) = (I_q \otimes Q_3) \frac{\partial}{\partial \mathbf{t}} vec(\begin{pmatrix} \mathbf{0}_{1\times q} \\ I_q \\ (\mathbf{t}-\boldsymbol{\xi}_0)\otimes I_q \\ \frac{(\mathbf{t}-\boldsymbol{\xi}_0)^{[2]}\otimes I_q}{2} \end{pmatrix})$$

$$= (I_q \otimes Q_3) I_{\frac{Q}{q},q} \frac{\partial}{\partial \mathbf{t}} vec(\mathbf{0}_{q\times 1}, I_q, (\mathbf{t}-\boldsymbol{\xi}_0)^T \otimes I_q, \frac{(\mathbf{t}-\boldsymbol{\xi}_0)^{[2]^T} \otimes I_q}{2})$$

$$= (I_q \otimes Q_3) I_{\frac{Q}{q},q} \frac{\partial}{\partial \mathbf{t}} \begin{pmatrix} \mathbf{0}_{1\times q} \\ vec(I_q) \\ (\mathbf{t}-\boldsymbol{\xi}_0)\otimes vec(I_q) \\ \frac{(\mathbf{t}-\boldsymbol{\xi}_0)^{[2]}\otimes vec(I_q)}{2} \end{pmatrix}$$

$$= (I_q \otimes Q_3) I_{\frac{Q}{q},q} \begin{pmatrix} \mathbf{0}_{q\times q} \\ \mathbf{0}_{q^2\times q} \\ I_q \otimes vec(I_q) \\ \frac{I_q\otimes(\mathbf{t}-\boldsymbol{\xi}_0)+(\mathbf{t}-\boldsymbol{\xi}_0)\otimes I_q}{2} \otimes vec(I_q) \end{pmatrix} .$$

When (2.20) is evaluated at $\mathbf{u} = \mathbf{a}$, $\mathbf{t} = \boldsymbol{\xi}_0$, we obtain

$$\begin{aligned} H''(\mathbf{a})_{qQ\times Q} = &- (\mathbf{0}_{Q\times 1} \vdots H'(\mathbf{a})^T_{Q\times q} \vdots \mathbf{0}_{Q\times q^2} \vdots \mathbf{0}_{Q\times q^3}) \otimes A_0^{-1} \\ &- \begin{pmatrix} \mathbf{0}_{1\times Q} \\ H'(\mathbf{a})_{q\times Q} \\ \mathbf{0}_{q^2\times Q} \\ \mathbf{0}_{q^3\times Q} \end{pmatrix} \otimes vec(A_0^{-1}) \\ &- (H'(\mathbf{a})^T \otimes A_0^{-1} Z_0) L_2(\boldsymbol{\xi}_0) H'(\mathbf{a}) . \end{aligned} \tag{2.21}$$

Now by applying (2.21), (2.16) becomes

$$\begin{aligned} \sqrt{n}(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0) = &- A_0^{-1}(\sqrt{n}\overline{\mathbf{u}}_0) - \frac{1}{2}\sqrt{n} I_1 - \frac{1}{2}\sqrt{n} I_2 - \frac{1}{2}\sqrt{n} I_3 \\ &+ \sqrt{n}((\overline{\mathbf{U}} - \mathbf{a})^{[2]^T} \otimes I_q) \frac{H^{(3)}(\mathbf{a})}{3!} (\overline{\mathbf{U}} - \mathbf{a}) + o_p(n^{-3/2+\epsilon}) , \end{aligned} \tag{2.22}$$

where

$$\begin{cases} I_1 &= [(\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q][(\mathbf{0}_{Q\times 1} \vdots H'(\mathbf{a})^T \vdots \mathbf{0}_{Q\times q^2} \vdots \mathbf{0}_{Q\times q^3}) \otimes A_0^{-1}](\overline{\mathbf{U}} - \mathbf{a}), \\ I_2 &= [(\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q][\begin{pmatrix} \mathbf{0}_{1\times Q} \\ H'(\mathbf{a}) \\ \mathbf{0}_{q^2\times Q} \\ \mathbf{0}_{q^3\times Q} \end{pmatrix} \otimes vec(A_0^{-1})](\overline{\mathbf{U}} - \mathbf{a}), \\ I_3 &= [(\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q][H'(\mathbf{a})^T \otimes (A_0^{-1}Z_0)]L_2(\boldsymbol{\xi}_0)H'(\mathbf{a})(\overline{\mathbf{U}} - \mathbf{a}) \ . \end{cases}$$

The next step is devoted to the simplification of I_1, I_2 and I_3.

Define $\mathbf{v}_0$ by $\mathbf{v}_0 = A_0^{-1}\overline{\mathbf{U}}_0$, then

$$\begin{aligned} I)I_1 =& \{[(\overline{\mathbf{U}} - \mathbf{a})^T(\mathbf{0}_{Q\times 1} \vdots H'(\mathbf{a})^T \vdots \mathbf{0}_{Q\times q^2} \vdots \mathbf{0}_{Q\times q^3})] \otimes A_0^{-1}\}(\overline{\mathbf{U}} - \mathbf{a}) \\ =& - (\mathbf{0}_{q\times q} \vdots \mathbf{v}_0^T \otimes A_0^{-1} \vdots \mathbf{0}_{q\times q^3} \vdots \mathbf{0}_{q\times q^4}) \begin{pmatrix} \overline{\mathbf{U}}_0 - \boldsymbol{\mu}_0 \\ \overline{\mathbf{U}}_1 - \boldsymbol{\mu}_1 \\ \overline{\mathbf{U}}_2 - \boldsymbol{\mu}_2 \\ \overline{\mathbf{U}}_3 - \boldsymbol{\mu}_3 \end{pmatrix} \\ =& - (\mathbf{v}_0^T \otimes A_0^{-1})(\overline{\mathbf{U}}_1 - \boldsymbol{\mu}_1) \\ =& - A_0^{-1}(A_n - A_0)\mathbf{v}_0 \ ; \end{aligned} \tag{2.23}$$

$$\begin{aligned} II)I_2 =& [(\overline{\mathbf{U}}_1 - \boldsymbol{\mu}_1)^T \otimes I_q][H'(\mathbf{a}) \otimes vec(A_0^{-1})](\overline{\mathbf{U}} - \mathbf{a}) \\ =& - [(\overline{\mathbf{U}}_1 - \boldsymbol{\mu}_1)^T \otimes I_q][\mathbf{v}_0 \otimes vec(A_0^{-1})] \\ =& - A_0^{-1}(A_n - A_0)\mathbf{v}_0 \ , \end{aligned} \tag{2.24}$$

where the last line is obtained by noticing that $vec(A_0^{-1})$ can be written as $vec(A_0^{-1}) =$

$\sum_i \mathbf{e}_i \otimes \mathbf{d}_i$, where $I_q = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_q)$ and $A_0^{-1} = (\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_q)$;

$$
\begin{aligned}
III) I_3 =& [(H'(\mathbf{a})(\overline{\mathbf{U}} - \mathbf{a}))^T \otimes (A_0^{-1} Z_0)] L_2(\boldsymbol{\xi}_0) H'(\mathbf{a})(\overline{\mathbf{U}} - \mathbf{a}) \\
=& [\mathbf{v}_0^T \otimes (A_0^{-1} Z_0)] L_2(\boldsymbol{\xi}_0) \mathbf{v}_0 \\
=& [\mathbf{v}_0^T \otimes (A_0^{-1} Z_0)] (I_q \otimes Q_3) I_{\frac{Q}{q}, q} vec((\mathbf{0}_{q\times 1} \vdots \mathbf{0}_{q\times q} \vdots \mathbf{v}_0^T \otimes I_q \vdots \mathbf{0}_{q\times q^3})) \\
=& [\mathbf{v}_0^T \otimes (A_0^{-1} Z_0)] vec(E) \\
=& vec(A_0^{-1} Z_0 E \mathbf{v}_0) \\
=& vec(A_0^{-1} B_0 \mathbf{v}_0^{[2]}) \\
=& A_0^{-1} B_0 \mathbf{v}_0^{[2]},
\end{aligned}
\tag{2.25}
$$

where

$$
E = L_2(\boldsymbol{\xi}_0)\mathbf{v}_0 = \begin{pmatrix} \mathbf{0}_{1\times q} \\ \mathbf{0}_{q\times q} \\ \frac{1}{2}\mathbf{v}_0 \otimes I_q + \frac{1}{2} I_q \otimes \mathbf{v}_0 \\ \mathbf{0}_{q^3 \times q} \end{pmatrix} .
$$

From (2.22)–(2.25) we now have

$$
\begin{aligned}
\sqrt{n}(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0) =& - \sqrt{n}\mathbf{v}_0 + \sqrt{n} A_0^{-1}(A_n - A_0)\mathbf{v}_0 - \frac{1}{2}\sqrt{n} A_0^{-1} B_0 \mathbf{v}_0^{[2]} \\
&+ \sqrt{n}(\overline{\mathbf{U}} - \mathbf{a})^{[2]T} \frac{H^{(3)}(\mathbf{a})}{3!}(\overline{\mathbf{U}} - \mathbf{a}) + o_p(n^{-3/2+\epsilon}) .
\end{aligned}
\tag{2.26}
$$

$\nu = 3$:

Define

$$
\begin{cases}
R_1(\mathbf{u}, \mathbf{t}) = L_1(\mathbf{t}) \frac{\partial \mathbf{t}}{\partial \mathbf{u}}, \\
R_2(\mathbf{u}, \mathbf{t}) = L_2(\mathbf{t}) \frac{\partial \mathbf{t}}{\partial \mathbf{u}}, \\
R_3(\mathbf{u}, \mathbf{t}) = (\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T \otimes (B^{-1}(\mathbf{u}, \mathbf{t}) Z_n) ,
\end{cases}
$$

then (2.18) can be rewritten as

$$\frac{\partial^2 \mathbf{t}}{\partial \mathbf{u}^2} = R_1^T(\mathbf{u},\mathbf{t}) \otimes B^{-1}(\mathbf{u},\mathbf{t}) + R_1(\mathbf{u},\mathbf{t}) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t})) + R_3(\mathbf{u},\mathbf{t})R_2(\mathbf{u},\mathbf{t}). \quad (2.27)$$

Now notice that

I) $vec(R_1(\mathbf{u},\mathbf{t}) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t})) = vec(R_1(\mathbf{u},\mathbf{t})) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t}))$;

II) since for any $\mathbf{c}_{Q\times 1}, \mathbf{d}_{\frac{Q}{q}\times 1}, \mathbf{e}_{q\times 1}$ *and* $\mathbf{f}_{q\times 1}$

$$\begin{aligned}(I_{\frac{Q}{q}} \otimes I_{Q,q} \otimes I_q)(I_{Q,\frac{Q}{q}} \otimes I_{q^2})(\mathbf{c}\otimes\mathbf{d}\otimes\mathbf{e}\otimes\mathbf{f}) \\ =&(I_{\frac{Q}{q}} \otimes I_{Q,q} \otimes I_q)(\mathbf{d}\otimes\mathbf{c}\otimes\mathbf{e}\otimes\mathbf{f}) \\ =&(\mathbf{d}\otimes\mathbf{e}\otimes\mathbf{c}\otimes\mathbf{f}) \\ =&(I_{Q,Q}\otimes I_q)(\mathbf{c}\otimes\mathbf{d}\otimes\mathbf{e}\otimes\mathbf{f}),\end{aligned}$$

we have $(I_{\frac{Q}{q}} \otimes I_{Q,q} \otimes I_q)(I_{Q,\frac{Q}{q}} \otimes I_{q^2}) = I_{Q,Q} \otimes I_q$, and therefore

$$\begin{aligned}vec(R_1^T(\mathbf{u},\mathbf{t}) \otimes B^{-1}(\mathbf{u},\mathbf{t})) =&(I_{\frac{Q}{q}} \otimes I_{Q,q} \otimes I_q)(I_{Q,\frac{Q}{q}} \otimes I_{q^2})[vec(R_1(\mathbf{u},\mathbf{t})) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t}))] \\ =&(I_{Q,Q}\otimes I_q)[vec(R_1(\mathbf{u},\mathbf{t})) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t}))]\ . \qquad (2.28)\end{aligned}$$

Then from (2.27)–(2.28) we have

$$\begin{aligned}vec(\frac{\partial^2 \mathbf{t}}{\partial \mathbf{u}^2}) =& -(I_{qQ^2} + I_{Q,Q}\otimes I_q)[vec(R_1(\mathbf{u},\mathbf{t})) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t}))] \\ &+ vec(R_3(\mathbf{u},\mathbf{t})R_2(\mathbf{u},\mathbf{t}))\ ,\end{aligned}$$

and so

$$\begin{aligned}\frac{\partial^3 \mathbf{t}}{\partial \mathbf{u}^3} =& -Q_4[\frac{\partial}{\partial \mathbf{u}}R_1(\mathbf{u},\mathbf{t}) \otimes vec(B^{-1}(\mathbf{u},\mathbf{t})) + vec(R_1(\mathbf{u},\mathbf{t})) \otimes \frac{\partial}{\partial \mathbf{u}}B^{-1}(\mathbf{u},\mathbf{t})] \\ &- [R_2^T(\mathbf{u},\mathbf{t}) \otimes I_{qQ}]\frac{\partial R_3(\mathbf{u},\mathbf{t})}{\partial \mathbf{u}} - [I_Q \otimes R_3(\mathbf{u},\mathbf{t})]\frac{\partial R_2(\mathbf{u},\mathbf{t})}{\partial \mathbf{u}}\ , \qquad (2.29)\end{aligned}$$

where $Q_4 = I_{qQ^2} + I_{Q,Q} \otimes I_q$.

In order to get $H'''(\mathbf{a})$, (2.29) has to be evaluated term by term as follows.

$$
\begin{aligned}
I)\frac{\partial R_1(\mathbf{u},\mathbf{t})}{\partial \mathbf{u}} =&[(\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T \otimes I_{\frac{Q}{q}}]\frac{L_1(\mathbf{t})}{\partial \mathbf{u}}\frac{\partial \mathbf{t}}{\partial \mathbf{u}} + [I_Q \otimes L_1(\mathbf{t})]\frac{\partial^2 \mathbf{t}}{\partial \mathbf{u}^2} \\
=&(H'(\mathbf{t})^T \otimes I_{\frac{Q}{q}})L_2(\mathbf{t})H'(\mathbf{t}) + [I_Q \otimes L_1(\mathbf{t})]H''(\mathbf{t});
\end{aligned}
\tag{2.30}
$$

II) From (2.17), we have

$$
\begin{aligned}
\frac{\partial B^{-1}(\mathbf{u},\mathbf{t})}{\partial \mathbf{u}} =& - [B^{-T}(\mathbf{u},\mathbf{t}) \otimes B^{-1}(\mathbf{u},\mathbf{t})](\frac{\partial B(\mathbf{u},\mathbf{t})}{\partial \mathbf{u}}|_{\mathbf{t}\text{ fixed}} + \frac{\partial B(\mathbf{u},\mathbf{t})}{\partial \mathbf{t}}|_{\mathbf{u}\text{ fixed}}\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) \\
=& - [B^{-T}(\mathbf{u},\mathbf{t}) \otimes B^{-1}(\mathbf{u},\mathbf{t})] \cdot \\
&\{L_1^T(\mathbf{t}) \otimes I_q + (I_q \otimes \mathbf{u}^T \otimes I_q)[(L_2(\mathbf{t})\frac{\partial \mathbf{t}}{\partial \mathbf{u}}) \otimes vec(I_q)]\} \\
=& - [B^{-T}(\mathbf{u},\mathbf{t})L_1^T(\mathbf{t})] \otimes B^{-1}(\mathbf{u},\mathbf{t}) - [B^{-T}(\mathbf{u},\mathbf{t}) \otimes B^{-1}(\mathbf{u},\mathbf{t})] \cdot \\
&[(L_2(\mathbf{t})H'(\mathbf{t})) \otimes vec(I_q)] \\
=& - [B^{-T}(\mathbf{u},\mathbf{t})L_1^T(\mathbf{t})] \otimes B^{-1}(\mathbf{u},\mathbf{t}) - [B^{-T}(\mathbf{u},\mathbf{t}) \otimes (B^{-1}(\mathbf{u},\mathbf{t})Z_n)]L_2(\mathbf{t})H'(\mathbf{u})\ ;
\end{aligned}
\tag{2.31}
$$

$$
\begin{aligned}
III)\ \frac{\partial R_3(\mathbf{u},\mathbf{t})}{\partial \mathbf{u}} =& \frac{\partial}{\partial \mathbf{u}}[(I_Q \otimes B^{-1}(\mathbf{u},\mathbf{t}))((\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T \otimes Z_n)] \\
=&(\frac{\partial \mathbf{t}}{\partial \mathbf{u}} \otimes Z_n \otimes I_{qQ})(I_Q \otimes I_{Q,q} \otimes I_q)[vec(I_Q) \otimes \frac{\partial B^{-1}(\mathbf{u},\mathbf{t})}{\partial \mathbf{u}}] + [I_{Q^2} \otimes B^{-1}(\mathbf{u},\mathbf{t})] \cdot \\
&(I_q \otimes I_{Q,\frac{Q}{q}} \otimes I_q)[vec((\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T) \otimes \frac{\partial Z_n}{\partial \mathbf{u}} + \frac{\partial}{\partial \mathbf{u}}((\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T) \otimes vec(Z_n)] \\
=&[\frac{\partial \mathbf{t}}{\partial \mathbf{u}} \otimes ((Z_n^T \otimes I_Q)I_{Q,q}) \otimes I_q][vec(I_Q) \otimes \frac{\partial B^{-1}(\mathbf{u},\mathbf{t})}{\partial \mathbf{u}}] \\
&+ (I_q \otimes I_{Q,\frac{Q}{q}} \otimes B^{-1}(\mathbf{u},\mathbf{t}))[vec((\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T) \otimes I_Q + (I_{Q,q}\frac{\partial^2 \mathbf{t}}{\partial \mathbf{u}^2}) \otimes \mathbf{u}];
\end{aligned}
\tag{2.32}
$$

$$IV)\frac{\partial R_2(\mathbf{u},\mathbf{t})}{\partial \mathbf{u}} = [(\frac{\partial \mathbf{t}}{\partial \mathbf{u}})^T \otimes I_Q]\frac{\partial L_2(\mathbf{t})}{\partial \mathbf{t}}\frac{\partial \mathbf{t}}{\partial \mathbf{u}} + [I_Q \otimes L_2(\mathbf{t})]\frac{\partial^2 \mathbf{t}}{\partial \mathbf{u}^2} , \tag{2.33}$$

and since

$$\begin{aligned}
vec(L_2(\mathbf{t})) = &\{I_q \otimes [(I_q \otimes Q_3)I_{\frac{Q}{q},q}]\}I_{Q,q} \cdot \\
&vec((\mathbf{0}_{q\times q} \vdots \mathbf{0}_{q\times q^2} \vdots I_q \otimes vec^T(I_q) \vdots \frac{I_q \otimes (\mathbf{t}-\boldsymbol{\xi}_0)^T + (\mathbf{t}-\boldsymbol{\xi}_0)^T \otimes I_q}{2} \otimes vec^T(I_q))) \\
=&(I_q \otimes I_q \otimes Q_3)(I_q \otimes I_{\frac{Q}{q},q})I_{Q,q} \cdot \\
&\begin{pmatrix} \mathbf{0}_{q^2\times 1} \\ \mathbf{0}_{q^3\times 1} \\ vec(I_q \otimes vec(I_q)) \\ \frac{1}{2}(I_{q^3,q} \otimes I_q + I_q \otimes I_{q^2,q} \otimes I_q)((\mathbf{t}-\boldsymbol{\xi}_0) \otimes vec(I_q) \otimes vec(I_q)) \end{pmatrix} ,
\end{aligned}$$

we have

$$\begin{aligned}
L_3(\mathbf{t}) =&\frac{\partial L_2(\mathbf{t})}{\partial \mathbf{t}} \\
=&(I_q \otimes I_q \otimes Q_3)(I_q \otimes I_{\frac{Q}{q},q})I_{Q,q} \\
&\begin{pmatrix} \mathbf{0}_{q^2\times q} \\ \mathbf{0}_{q^3\times q} \\ \mathbf{0}_{q^4\times q} \\ \frac{1}{2}(I_{q^3,q} \otimes I_q + I_q \otimes I_{q^2,q} \otimes I_q)(I_q \otimes vec(I_q) \otimes vec(I_q)) \end{pmatrix} .
\end{aligned} \tag{2.34}$$

Now when (2.29)–(2.34) are evaluated at $(\mathbf{u},\mathbf{t}) = (\mathbf{a},\boldsymbol{\xi}_0)$, we obtain

$$H'''(\mathbf{a}) = T_{11} + T_{12} + T_{13} + T_{21} + T_{22} + T_{31} + T_{32} , \tag{2.35}$$

where

$$\begin{cases} T_{11} = -Q_4\{[(H'^T(\mathbf{a}) \otimes I_{\frac{Q}{q}})L_2(\boldsymbol{\xi}_0)H'(\mathbf{a})] \otimes vec(A_0^{-1})\} , \\ T_{12} = -Q_4\{[(I_Q \otimes L_1(\boldsymbol{\xi}_0))H''(\mathbf{a})] \otimes vec(A_0^{-1})\} , \\ T_{13} = Q_4\{vec(L_1(\boldsymbol{\xi}_0)H'(\mathbf{a})) \otimes [(A_0^{-T}L_1^T(\boldsymbol{\xi}_0)) \otimes A_0^{-1} + (A_0^{-T} \otimes (A_0^{-T}Z_0))L_2(\boldsymbol{\xi}_0)H'(\mathbf{a})]\} , \\ T_{21} = [(L_2(\boldsymbol{\xi}_0)H'(\mathbf{a}))^T \otimes I_{qQ}][H'(\mathbf{a}) \otimes ((Z_0^T \otimes I_Q)I_{Q,q}) \otimes I_q] \cdot \\ \qquad \{vec(I_Q) \otimes [(A_0^{-T}L_1^T(\boldsymbol{\xi}_0)) \otimes A_0^{-1} + (A_0^{-T} \otimes (A_0^{-1}Z_0))L_2(\boldsymbol{\xi}_0)H'(\mathbf{a})]\} , \\ T_{22} = -[(L_2(\boldsymbol{\xi}_0)H'(\mathbf{a}))^T \otimes I_{qQ}](I_q \otimes I_{Q,\frac{Q}{q}} \otimes A_0^{-1})[vec(H'^T(\mathbf{a})) \otimes I_Q + (I_{Q,q}H''(\mathbf{a})) \otimes \mathbf{a}] , \\ T_{31} = -[I_Q \otimes H'^T(\mathbf{a}) \otimes (A_0^{-1}Z_0)](H'^T(\mathbf{a}) \otimes I_Q)L_3(\boldsymbol{\xi}_0)H'(\mathbf{a}), \\ T_{32} = -[I_Q \otimes H'^T(\mathbf{a}) \otimes (A_0^{-1}Z_0)](I_Q \otimes L_2(\boldsymbol{\xi}_0))H''(\mathbf{a}) . \end{cases}$$

Again, $[(\overline{\mathbf{U}} - \mathbf{a})^{[2]^T} \otimes I_q]H'''(\mathbf{a})(\overline{\mathbf{U}} - \mathbf{a})$ has to be evaluated term by term as follows:

$$\begin{aligned} I) \ & [(\overline{\mathbf{U}}-\mathbf{a})^{[2]^T} \otimes I_q]T_{11}(\overline{\mathbf{U}} - \mathbf{a}) \\ &= -2[(\overline{\mathbf{U}} - \mathbf{a})^{[2]^T} \otimes I_q]\{[(H'^T(\mathbf{a}) \otimes I_{\frac{Q}{q}})(-L_2(\boldsymbol{\xi}_0)\mathbf{v}_0)] \otimes vec(A_0^{-1})\} \\ &=2[(\overline{\mathbf{U}} - \mathbf{a})^{[2]^T} \otimes I_q](H'^T(\mathbf{a}) \otimes I_{\frac{Q}{q}} \otimes I_{q^2})[vec(E) \otimes vec(A_0^{-1})] \\ &= -2(\mathbf{v}_0^T \otimes (\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q)(vec(E) \otimes vec(A_0^{-1})) \\ &= -2(vec^T(E) \otimes A_0^{-1})(\mathbf{v}_0 \otimes (\overline{\mathbf{U}} - \mathbf{a})) \\ &= -2(vec^T(E) \otimes I_q)(I_Q \otimes A_0^{-1})(\mathbf{v}_0 \otimes I_Q)(\overline{\mathbf{U}} - \mathbf{a}) \\ &= -2[(vec^T(E)(\mathbf{v}_0 \otimes I_{\frac{Q}{q}})) \otimes A_0^{-1}]vec(Z_n - Z_0) \\ &= -2((E\mathbf{v}_0)^T \otimes A_0^{-1})vec(Z_n - Z_0) \\ &= -2A_0^{-1}(Z_n - Z_0)E\mathbf{v}_0 \\ &= -2A_0^{-1}(B_n - B_0)(\frac{1}{2}\mathbf{v}_0 \otimes I_q + \frac{1}{2}I_q \otimes \mathbf{v}_0)\mathbf{v}_0 \\ &= -2A_0^{-1}(B_n - B_0)\mathbf{v}_0^{[2]} ; \end{aligned} \tag{2.36}$$

$II)$ $[(\overline{\mathbf{U}}-\mathbf{a})^{[2]^T} \otimes I_q]T_{12}(\overline{\mathbf{U}}-\mathbf{a})$

$$
\begin{aligned}
&= -2[(\overline{\mathbf{U}}-\mathbf{a})^{[2]^T} \otimes I_q]\{[(I_Q \otimes L_1(\boldsymbol{\xi}_0))H''(\mathbf{a})(\overline{\mathbf{U}}-\mathbf{a})] \otimes vec(A_0^{-1})\} \\
&= -2\{(\overline{\mathbf{U}}-\mathbf{a})^T \otimes [(\overline{\mathbf{U}}-\mathbf{a})^T(L_1(\boldsymbol{\xi}_0) \otimes I_q)] \otimes I_q\}[(H''(\mathbf{a})(\overline{\mathbf{U}}-\mathbf{a})) \otimes vec(A_0^{-1})] \\
&= -2[(\overline{\mathbf{U}}-\mathbf{a})^T \otimes vec^T(A_n - A_0) \otimes I_q]((H''(\mathbf{a})(\overline{\mathbf{U}}-\mathbf{a})) \otimes vec(A_0^{-1})) \\
&= -2\{[H''(\mathbf{a})(\overline{\mathbf{U}}-\mathbf{a})]^T \otimes A_0^{-1}\}[(\overline{\mathbf{U}}-\mathbf{a}) \otimes vec(A_n - A_0)] \\
&= -2\{(\overline{\mathbf{U}}-\mathbf{a})^T \otimes [A_0^{-1}(A_n - A_0)]\}H''(\mathbf{a})(\overline{\mathbf{U}}-\mathbf{a}) \\
&= -2A_0^{-1}(A_n - A_0)[(\overline{\mathbf{U}}-\mathbf{a})^T \otimes I_q]H''(\mathbf{a})(\overline{\mathbf{U}}-\mathbf{a}) \\
&= -2A_0^{-1}(A_n - A_0)[2A_0^{-1}(A_n - A_0)\mathbf{v}_0 - A_0^{-1}B_0\mathbf{v}_0^{[2]}] \\
&= -4[A_0^{-1}(A_n - A_0)]^2\mathbf{v}_0 + 2A_0^{-1}(A_n - A_0)A_0^{-1}B_0\mathbf{v}_0^{[2]} \; ; \qquad (2.37)
\end{aligned}
$$

$III)$ Since

$$
\begin{aligned}
&[(A_0^{-T}L_1^T(\boldsymbol{\xi}_0)) \otimes A_0^{-1}](\overline{\mathbf{U}}-\mathbf{a}) + [A_0^{-T} \otimes (A_0^{-1}Z_0)]L_2(\boldsymbol{\xi}_0)H'(\mathbf{a})(\overline{\mathbf{U}}-\mathbf{a}) \\
=&vec(A_0^{-1}(Z_n - Z_0)L_1(\boldsymbol{\xi}_0)A_0^{-1} - A_0^{-1}Z_0EA_0^{-1}) \\
=&vec(A_0^{-1}(A_n - A_0)A_0^{-1} - A_0^{-1}B_0(\mathbf{v}_0 \otimes I_q)A_0^{-1}) \\
:=&vec(P_{q\times q}) \; ,
\end{aligned}
$$

we have

$$
\begin{aligned}
&[(\overline{\mathbf{U}}-\mathbf{a})^{[2]^T} \otimes I_q]T_{13}(\overline{\mathbf{U}}-\mathbf{a}) \\
=&2[(\overline{\mathbf{U}}-\mathbf{a})^{[2]^T} \otimes I_q][vec(L_1(\boldsymbol{\xi}_0)H'(\mathbf{a})) \otimes vec(P)] \\
=&2[vec^T(L_1(\boldsymbol{\xi}_0)H'(\mathbf{a})) \otimes P][(\overline{\mathbf{U}}-\mathbf{a}) \otimes vec(Z_n - Z_0)] \\
=&2[(\overline{\mathbf{U}}-\mathbf{a})^T \otimes (P(Z_n - Z_0))]vec(L_1(\boldsymbol{\xi}_0)H'(\mathbf{a}))
\end{aligned}
$$

$$=2vec(P(Z_n - Z_0)L_1(\boldsymbol{\xi}_0)H'(\mathbf{a})(\overline{\mathbf{U}} - \mathbf{a}))$$

$$= -2P(A_n - A_0)\mathbf{v}_0$$

$$= -2[A_0^{-1}(A_n - A_0)]^2\mathbf{v}_0 + 2A_0^{-1}B_0\{\mathbf{v}_0 \otimes [A_0^{-1}(A_n - A_0)\mathbf{v}_0]\}$$

$$= -2[A_0^{-1}(A_n - A_0)]^2\mathbf{v}_0 + 2A_0^{-1}B_0\{I_q \otimes [A_0^{-1}(A_n - A_0)]\}\mathbf{v}_0^{[2]}\ ; \tag{2.38}$$

$IV)$ $[(\overline{\mathbf{U}}-\mathbf{a})^{[2]^T} \otimes I_q]T_{21}(\overline{\mathbf{U}} - \mathbf{a})$

$$=[vec(L_2(\boldsymbol{\xi}_0)H'(\mathbf{a})(\overline{\mathbf{U}} - \mathbf{a})) \otimes (\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q](H'(\mathbf{a}) \otimes Z_0^T \otimes I_{qQ}) \cdot$$

$$(I_Q \otimes I_{Q,q} \otimes I_q)[vec(I_Q) \otimes vec(P)]$$

$$= -[vec(E) \otimes (\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q](H'(\mathbf{a}) \otimes Z_0^T \otimes I_{qQ})vec(I_Q \otimes P)$$

$$= -[vec(Z_0EH'(\mathbf{a})) \otimes (\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q]vec(I_Q \otimes P)$$

$$= -[(\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q](I_Q \otimes P)vec(Z_0EH'(\mathbf{a}))$$

$$= -PZ_0EH'(\mathbf{a})(\overline{\mathbf{U}} - \mathbf{a})$$

$$=PZ_0E\mathbf{v}_0$$

$$=A_0^{-1}(A_n - A_0)A_0^{-1}B_0\mathbf{v}_0^{[2]} - A_0^{-1}B_0[\mathbf{v}_0 \otimes (A_0^{-1}B_0)]\mathbf{v}_0^{[2]}$$

$$=A_0^{-1}(A_n - A_0)A_0^{-1}B_0\mathbf{v}_0^{[2]} - A_0^{-1}B_0[I_q \otimes (A_0^{-1}B_0)]\mathbf{v}_0^{[3]}\ ; \tag{2.39}$$

$V)$ define F_1 by $vec(F_{1_{q\times Q}}) = H''(\mathbf{a})(\overline{\mathbf{U}} - \mathbf{a})$, then by applying the similar procedure as in $IV)$, we have

$$[(\overline{\mathbf{U}}-\mathbf{a})^{[2]^T} \otimes I_q]T_{22}(\overline{\mathbf{U}} - \mathbf{a})$$

$$=[vec^T(E) \otimes (\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q](I_q \otimes I_{Q,\frac{Q}{q}} \otimes A_0^{-1}) \cdot$$

$$[vec(H'^T(\mathbf{a})) \otimes vec(Z_n - Z_0) + vec(F_1^T) \otimes \mathbf{a}]$$

$$
\begin{aligned}
=&[vec^T(E) \otimes (\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q] vec(H'^T(\mathbf{a}) \otimes (A_0^{-1}(Z_n - Z_0)) + F_1^T \otimes (A_0^{-1} Z_0)) \\
=&[(\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q][H'^T(\mathbf{a}) \otimes (A_0^{-1}(Z_n - Z_0)) + F_1^T \otimes (A_0^{-1} Z_0)] vec(E) \\
=& - A_0^{-1}(Z_n - Z_0) E \mathbf{v}_0 + A_0^{-1} Z_0 E F_1 (\overline{\mathbf{U}} - \mathbf{a}) \\
=& - A_0^{-1}(B_n - B_0)\mathbf{v}_0^{[2]} + A_0^{-1} B_0 (\mathbf{v}_0 \otimes I_q)[(\overline{\mathbf{U}} - \mathbf{a})^T \otimes I_q] H''(\mathbf{a})(\overline{\mathbf{U}} - \mathbf{a}) \\
=& - A_0^{-1}(B_n - B_0)\mathbf{v}_0^{[2]} + 2A_0^{-1} B_0 \{\mathbf{v}_0 \otimes [A_0^{-1}(A_n - A_0)\mathbf{v}_0]\} - A_0^{-1} B_0 [\mathbf{v}_0 \otimes (A_0^{-1} B_0 \mathbf{v}_0^{[2]})] \\
=& - A_0^{-1}(B_n - B_0)\mathbf{v}_0^{[2]} + 2A_0^{-1} B_0 \{I_q \otimes [A_0^{-1}(A_n - A_0)]\}\mathbf{v}_0^{[2]} - A_0^{-1} B_0 [I_q \otimes (A_0^{-1} B_0)]\mathbf{v}_0^{[2]} \ ;
\end{aligned}
\tag{2.40}
$$

$VI)$ define F_2 by

$$
vec(F_{2_{q\times q^4}}) = \frac{1}{2}(I_{q^3,q} \otimes I_q + I_q \otimes I_{q^2,q} \otimes I_q)[\mathbf{v}_0 \otimes vec(I_q) \otimes vec(I_q)] \ ,
$$

then

$$
\begin{aligned}
F_2^T \mathbf{v}_0 =& vec(\mathbf{v}_0^T F_2) \\
=& \frac{1}{2}\{[I_{q^3,q} + (I_q \otimes I_{q^2,q})] \otimes \mathbf{v}_0^T\}[\mathbf{v}_0 \otimes vec(I_q) \otimes vec(I_q)] \\
=& \mathbf{v}_0^{[2]} \otimes vec(I_q) \ ,
\end{aligned}
$$

and therefore

$$
\begin{aligned}
&[(\overline{\mathbf{U}} - \mathbf{a})^{[2]^T} \otimes I_q] T_{31} (\overline{\mathbf{U}} - \mathbf{a}) \\
&= \{[(\overline{\mathbf{U}} - \mathbf{a})^{[2]^T}(I_Q \otimes H'^T(\mathbf{a}))] \otimes (A_0^{-1} Z_0)\}(H'^T(\mathbf{a}) \otimes I_Q) L_3(\boldsymbol{\xi}_0)\mathbf{v}_0 \\
&= [\mathbf{v}_0^{[2]^T} \otimes (A_0^{-1} Z_0)] L_3(\boldsymbol{\xi}_0)\mathbf{v}_0 \\
&= [(A_0^{-1} Z_0 Q_3) \otimes \mathbf{v}_0^{[2]^T}] vec((\mathbf{0}_{q\times q} \vdots \mathbf{0}_{q\times q^2} \vdots \mathbf{0}_{q\times q^3} \vdots F_{2_{q\times q^4}})) \\
&= vec(\mathbf{v}_0^T (\mathbf{0}_{q\times q} \vdots \mathbf{0}_{q\times q^2} \vdots \mathbf{0}_{q\times q^3} \vdots F_{2_{q\times q^4}})[(A_0^{-1} Z_0)^T \otimes \mathbf{v}_0])
\end{aligned}
$$

$$=((A_0^{-1}Z_0)\otimes \mathbf{v}_0^T)\begin{pmatrix} \mathbf{0}_{q\times 1} \\ \mathbf{0}_{q^2\times 1} \\ \mathbf{0}_{q^3\times 1} \\ F_2^T\mathbf{v}_0 \end{pmatrix}$$

$$=((A_0^{-1}C_0)\otimes \mathbf{v}_0^T)F_2^T\mathbf{v}_0 = A_0^{-1}C_0\mathbf{v}_0^{[3]} \;; \qquad (2.41)$$

$VII)\quad [(\overline{\mathbf{U}}-\mathbf{a})^{[2]^T}\otimes I_q]T_{32}(\overline{\mathbf{U}}-\mathbf{a})$

$$=-\{[(\overline{\mathbf{U}}-\mathbf{a})^{[2]^T}(I_Q\otimes H'^T(\mathbf{a}))]\otimes (A_0^{-1}Z_0)\}[I_Q\otimes L_2(\boldsymbol{\xi}_0)]vec(F_1)$$

$$=\{(\overline{\mathbf{U}}-\mathbf{a})^T\otimes[(\mathbf{v}_0^T\otimes(A_0^{-1}Z_0))L_2(\boldsymbol{\xi}_0)]\}vec(F_1)$$

$$=vec([\mathbf{v}_0^T\otimes(A_0^{-1}Z_0)]L_2(\boldsymbol{\xi}_0)F_1(\overline{\mathbf{U}}-\mathbf{a}))$$

$$=A_0^{-1}Z_0Q_3\begin{pmatrix} \mathbf{0}_{q\times q} \\ \mathbf{0}_{q^2\times q} \\ I_q\otimes vec(I_q) \\ \mathbf{0}_{q^4\times q} \end{pmatrix}F_1(\overline{\mathbf{U}}-\mathbf{a})$$

$$=\frac{1}{2}\{[A_0^{-1}B_0(I_{q^2}+I_{q,q})]\otimes \mathbf{v}_0^T\}\{[F_1(\overline{\mathbf{U}}-\mathbf{a})]\otimes vec(I_q)\}$$

$$=\frac{1}{2}[A_0^{-1}B_0(I_{q^2}+I_{q,q})]\{[F_1(\overline{\mathbf{U}}-\mathbf{a})]\otimes \mathbf{v}_0\}$$

$$=A_0^{-1}B_0\{[F_1(\overline{\mathbf{U}}-\mathbf{a})]\otimes \mathbf{v}_0+\mathbf{v}_0\otimes[F_1(\overline{\mathbf{U}}-\mathbf{a})]\}$$

$$=2A_0^{-1}B_0\{[A_0^{-1}(A_n-A_0)]\otimes I_q\}\mathbf{v}_0^{[2]}-A_0^{-1}B_0[(A_0^{-1}B_0)\otimes I_q]\mathbf{v}_0^{[3]}\;. \qquad (2.42)$$

Now from (2.26) and (2.35)–(2.42), we have

$$\begin{aligned}\sqrt{n}(\boldsymbol{\xi}_n-\boldsymbol{\xi}_0)=&-\sqrt{n}\mathbf{v}_0+\sqrt{n}A_0^{-1}(A_n-A_0)\mathbf{v}_0-\frac{\sqrt{n}}{2}A_0^{-1}B_0\mathbf{v}_0^{[2]}\\ &-\frac{\sqrt{n}}{6}\{6[A_0^{-1}(A_n-A_0)]^2\mathbf{v}_0+3A_0^{-1}(B_n-B_0)\mathbf{v}_0^{[2]}\\ &-3A_0^{-1}(A_n-A_0)A_0^{-1}B_0\mathbf{v}_0^{[2]}-4A_0^{-1}B_0[I_q\otimes(A_0^{-1}(A_n-A_0))]\mathbf{v}_0^{[2]}\\ &-2A_0^{-1}B_0[(A_0^{-1}(A_n-A_0))\otimes I_q]\mathbf{v}_0^{[2]}+2A_0^{-1}B_0[I_q\otimes(A_0^{-1}B_0)]\mathbf{v}_0^{[3]}\\ &+A_0^{-1}B_0[(A_0^{-1}B_0)\otimes I_q]\mathbf{v}_0^{[3]}-A_0^{-1}C_0B_0\mathbf{v}_0^{[3]}\}+o_p(n^{-3/2+\epsilon})\;. \qquad (2.43)\end{aligned}$$

Finally we summarize this result with theorem 2.2.

Theorem 2.2 Assume that **A1)**–**A6)** hold and $\tilde{\mathbf{v}}_0 = \frac{\mathbf{v}_0}{\sigma}$, $R_n = A_0^{-1}A_n$, $S_n = \sigma A_0^{-1}B_n$, $S_0 = \sigma A_0^{-1}B_0$, $T_0 = \sigma^2 A_0^{-1}C_0$, where $\mathbf{v}_0 = A_0^{-1}\overline{\mathbf{U}}_0$, $\overline{\mathbf{U}}_0 = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n \eta(\mathbf{x}_i, \frac{\epsilon_i}{\sigma})\mathbf{x}_{i,1} \\ \frac{1}{n}\sum_{i=1}^n \chi(\frac{\epsilon_i}{\sigma}) \end{pmatrix}$ and A_0, B_0, C_0, A_n, B_n, C_n are defined as in (2.10). Then the following is valid uniformly on compact subsets of the parameter space for any $\epsilon > 0$.

$$\begin{aligned}\frac{\sqrt{n}(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)}{\sigma} =& -\sqrt{n}\tilde{\mathbf{v}}_0 + \sqrt{n}[(R_n - I_q)\tilde{\mathbf{v}}_0 - \frac{1}{2}S_0\tilde{\mathbf{v}}_0^{[2]}] \\ &- \frac{\sqrt{n}}{6}\{6(R_n - I_q)^2\tilde{\mathbf{v}}_0 + 3(S_n - S_0)\tilde{\mathbf{v}}_0^{[2]} \\ &- 3(R_n - I_q)S_0\tilde{\mathbf{v}}_0^{[2]} - 4S_0[I_q \otimes (R_n - I_q)]\tilde{\mathbf{v}}_0^{[2]} \\ &- 2S_0[(R_n - I_q) \otimes I_q]\tilde{\mathbf{v}}_0^{[2]} + 2S_0(I_q \otimes S_0)\tilde{\mathbf{v}}_0^{[3]} \\ &+ S_0(S_0 \otimes I_q)\tilde{\mathbf{v}}_0^{[3]} - T_0\tilde{\mathbf{v}}_0^{[3]}\} + o_p(n^{-3/2+\epsilon}) \ . \end{aligned} \tag{2.44}$$

From (2.44), the approximate cumulants of $\frac{\sqrt{n}(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)}{\sigma}$ can then be derived and by applying an Edgeworth expansion, the approximate distribution could also be obtained if necessary.

§2.3 Examples

In this section, we will present two examples of M–type estimation problems. As we will see, (2.44) can be further simplified in these cases.

Example 2.1: Ordinary M–estimation of a location parameter with scale known.

The model of interest in this example is

$$y_i = \theta + \epsilon_i, \qquad i = 1, 2, \cdots, n \ ,$$

where ϵ_i's are independently and symmetrically distributed with known scale parameter σ (without loss of generality we assume $\sigma = 1$). Then the ordinary M-estimator $\hat{\theta}$ of the location parameter θ solves the equation $\sum_{i=1}^{n} \psi(y_i - \theta) = 0$. By applying the notation in Theorem 2.2 we have $q = 1$, $\xi_n = \hat{\theta}$, $\xi_0 = \theta_0$, $\sigma = 1$ and

$$\begin{cases} R_n = \frac{\frac{1}{n}\sum \psi'(\epsilon_i)}{E[\psi'(\epsilon)]}, \\ S_n = \frac{\frac{1}{n}\sum \psi''(\epsilon_i)}{E[\psi'(\epsilon)]}, \quad S_0 = \frac{E[\psi''(\epsilon)]}{E[\psi'(\epsilon)]}, \\ T_0 = \frac{E[\psi'''(\epsilon)]}{E[\psi'(\epsilon)]}, \\ \overline{\mathbf{U}}_0 = \frac{1}{n}\sum \psi(\epsilon_i), \quad \tilde{\mathbf{v}}_0 = \frac{\frac{1}{n}\sum \psi(\epsilon_i)}{E[\psi'(\epsilon)]}. \end{cases}$$

If we further define

$$\begin{cases} \beta_i = E[\psi^{(i)}(\epsilon)], & i = 0, 1, 2, 3 , \\ \overline{X}_i = \frac{1}{n}\sum_{j=1}^{n} \psi^{(i)}(\epsilon_j) - \beta_i; & i = 0, 1, 2, 3 , \end{cases} \tag{2.45}$$

then we have $\beta_0 = \beta_2 = 0$ and (2.44) can be written as

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) =& \sqrt{n}\left(-\frac{\overline{X}_0}{\beta_1} + \frac{\overline{X}_1\overline{X}_0}{\beta_1^2} + \frac{1}{6}\frac{\beta_3\overline{X}_0^3}{\beta_1^4} - \frac{1}{2}\frac{\overline{X}_2\overline{X}_0^2}{\beta_1^3} - \frac{\overline{X}_1^2\overline{X}_0}{\beta_1^3}\right) \\ &+ o_p(n^{-3/2+\epsilon}) \end{aligned} \tag{2.46}$$

for any $\epsilon > 0$.

Once we have (2.46), we may then take the Edgeworth expansion to get the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$. In order to do this, we first need to evaluate the cumulants of $\sqrt{n}(\hat{\theta} - \theta_0)$, which with some algebra are given by

$$\begin{cases} \kappa_1 = O(n^{-2}), \\ \kappa_2 = \frac{\nu_0}{\beta_1^2} + \frac{\kappa_{22}}{n} + O(n^{-2}), \\ \kappa_3 = O(n^{-2}), \\ \kappa_4 = \frac{\kappa_{42}}{n} + O(n^{-2}), \end{cases}$$

where

$$\begin{cases} \kappa_{22} = -\frac{2}{\beta_1^3}E[X_0^2X_1] - \frac{\nu_0}{\beta_1^2} + 3\frac{\nu_0}{\beta_1^4}E[X_0X_2] - \frac{\beta_3\nu_0^2}{\beta_1^5} + \frac{3\nu_0\nu_1}{\beta_1^4}, \\ \kappa_{42} = -\frac{12\nu_0}{\beta_1^5}E[X_0^2X_1] + 12\frac{\nu_0^2}{\beta_1^6}E[X_0X_2] + 12\frac{\nu_0^2\nu_1}{\beta_1^6} - \frac{4\beta_3\nu_0^3}{\beta_1^7}, \end{cases}$$

with $\beta_i = E[\psi^{(i)}(\epsilon)]$, $\nu_i = E[(\psi^{(i)}(\epsilon))^2]$ and $X_i = \psi^{(i)}(\epsilon_1)$.

Next, the expression

$$\exp\{(it)\kappa_1 + \frac{(it)^2}{2}(\kappa_2 - \frac{\nu_0}{\beta_1^2}) + \frac{(it)^3}{3!}\kappa_3 + \frac{(it)^4}{4!}\kappa_4\}\exp(-\frac{t^2}{2}\frac{\nu_0}{\beta_1^2}) \tag{2.47}$$

gives us an approximation of the characteristic function of $\sqrt{n}(\hat{\theta} - \theta_0)$. Expanding the first exponential factor, one may reduce (2.47) to

$$\exp(-\frac{\nu_0 t^2}{2\beta_1^2})(1 + \frac{1}{n}(\frac{(it)^2\kappa_{22}}{2} + \frac{(it)^4\kappa_{42}}{24})) + O(n^{-2}) \ . \tag{2.48}$$

It follows that the density and distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$ can be approximated by

$$\begin{cases} f(x) = (1 + \frac{1}{n}(\frac{\kappa_{22}}{2}H_2(x) + \frac{\kappa_{42}}{24}H_4(x)))\phi_{\sigma_0^2}(x) \ + O(n^{-2}), \\ F(x) = \Phi_{\sigma_0^2}(x) - \frac{1}{n}(\frac{\kappa_{22}}{2}H_1(x) + \frac{\kappa_{42}}{24}H_3(x))\phi_{\sigma_0^2}(x) + O(n^{-2}) \ , \end{cases} \tag{2.49}$$

where $\sigma_0^2 = \frac{\nu_0}{\beta_1^2}$, $\phi_{\sigma^2}(x)$, $\Phi_{\sigma^2}(x)$ are the normal $N(0, \sigma^2)$ density and distribution functions respectively, and $H_k(x)$ is defined by

$$H_k(x)\phi_{\sigma^2}(x) = (-\frac{d}{dx})^k\phi_{\sigma^2}(x) \ .$$

If the assumptions **A1)–A6)** hold, Bhattacharya and Ghosh have proven that for every compact $K \subset \Theta$, one has

$$\sup_{\theta_0 \in K} |P_{\theta_0}(\sqrt{n}(\hat{\theta} - \theta_0) \in B) - \int_B f(x)dx| = o(n^{-1})$$

uniformly over every class $\mathfrak{B}$ of Borel sets satisfying

$$\sup_{\theta_0 \in K}\sup_{B \in \mathfrak{B}} \int_{(\partial B)^\epsilon} \phi_{\sigma_0^2}(x)dx = O(\epsilon) \qquad \text{as} \qquad \epsilon \downarrow 0$$

and therefore our asymptotic expansion is valid under **A1)–A6)**. However, (2.49) gives a somewhat poor approximation in the tails when n is small.

Example 2.2: Ordinary linear regression M–estimation problem with scale unknown

The model in this example is the same as in §1.1. Then the ordinary M–estimator $\xi_n = (\hat{\theta}, \hat{\sigma})$ of regression/scale parameter $\xi_0 = (\theta_0, \sigma)$ solves equations

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} \psi(\frac{y_i - \mathbf{x}_i^T\theta}{\sigma})\mathbf{x}_i = 0 \\ \frac{1}{n}\sum_{i=1}^{n} \chi(\frac{y_i - \theta}{\sigma}) = 0 \ , \end{cases} \tag{2.50}$$

for some odd function ψ and even function χ, where ψ and χ are both assumed to be continuous and piecewise differentiable at least three times.

Proceeding as in example 2.1, we can derive the asymptotic expansions of $\frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sigma}$ and $\frac{\sqrt{n}(\hat{\sigma}-\sigma)}{\sigma}$, and then calculate their cumulants and joint/marginal distributions by using Edgeworth expansions. However, the expressions become lengthy quickly. For example, if we define

$$\begin{cases} \alpha_{1,0} = \sigma E[(\psi(\frac{y-\mathbf{x}^T\theta_0}{\sigma})\mathbf{x})(\frac{\partial}{\partial\theta})^T], \\ \alpha_{i,j} = \sigma^{i+j}\alpha_{1,0}^{-1}E[(\frac{\partial^j}{\partial\sigma^j}\psi(\frac{y-\mathbf{x}^T\theta_0}{\sigma})\mathbf{x})(\frac{\partial}{\partial\theta})^{[i]^T}] \text{ if } (i,j) \neq (1,0), \\ \gamma_{0,1} = \sigma E[\frac{\partial}{\partial\sigma}\chi(\frac{y-\mathbf{x}^T\theta_0}{\sigma})], \\ \gamma_{i,j} = \sigma^{i+j}\gamma_{0,1}^{-1}E[(\frac{\partial^j}{\partial\sigma^j}\chi(\frac{y-\mathbf{x}^T\theta_0}{\sigma}))(\frac{\partial}{\partial\theta})^{[i]^T}] \text{ , if } (i,j) \neq (0,1), \\ \overline{X}_{i,j} = \frac{1}{n}\sigma^{i+j}\alpha_{1,0}^{-1}\sum_{k=1}^{n}[(\frac{\partial^j}{\partial\sigma^j}\psi(\frac{y_k-\mathbf{x}_k^T\theta_0}{\sigma})\mathbf{x}_k)(\frac{\partial}{\partial\theta})^{[i]^T}] - \alpha_{i,j} \ , \\ \overline{Y}_{i,j} = \frac{1}{n}\sigma^{i+j}\sum_{k=1}^{n}[(\frac{\partial^j}{\partial\sigma^j}\chi(\frac{y_k-\mathbf{x}_k^T\theta_0}{\sigma}))(\frac{\partial}{\partial\theta})^{[i]^T}] - \gamma_{i,j} \ , \end{cases} \tag{2.51}$$

then one can show that

$$\begin{cases} \frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sigma} = \sqrt{n}\Theta = \sqrt{n}(\Theta_1 + \Theta_2 + \Theta_3) + o_p(n^{-3/2+\epsilon}), \\ \frac{\sqrt{n}(\hat{\sigma}-\sigma)}{\sigma} = \sqrt{n}\Sigma = \sqrt{n}(\Sigma_1 + \Sigma_2 + \Sigma_3) + o_p(n^{-3/2+\epsilon}) \end{cases} \tag{2.52}$$

for any $\epsilon > 0$, where

$$\begin{cases} \Theta_1 = -\overline{X}_{0,0}, \\ \Theta_2 = \overline{X}_{1,0}\overline{X}_{0,0} + \overline{X}_{0,1}\overline{Y}_{0,0} - \alpha_{1,1}\overline{X}_{0,0}\overline{Y}_{0,0}, \\ \Theta_3 = \alpha_{1,1}\overline{X}_{0,0}\overline{Y}_{0,0}\overline{Y}_{0,1} + (\frac{\gamma_{0,2}}{2} + \alpha_{1,1})\overline{X}_{0,1}\overline{Y}^2_{0,0} - \overline{X}_{1,1}\overline{X}_{0,0}\overline{Y}_{0,0} \\ \qquad - \overline{X}^2_{1,0}\overline{X}_{0,0} - \overline{X}_{1,0}\overline{X}_{0,1}\overline{Y}_{0,0} + (-\alpha^2_{1,1} - \frac{1}{2}\alpha_{1,1}\gamma_{0,2} + \frac{1}{2}\alpha_{1,2})\overline{X}_{0,0}\overline{Y}^2_{0,0} \\ \qquad + \frac{1}{6}(\alpha_{3,0}\overline{X}^{[3]}_{0,0} - 3\gamma_{2,0}\overline{X}^{[2]}_{0,0}\alpha_{1,1}\overline{X}_{0,0}) - \overline{X}_{0,1}\overline{Y}_{1,0}\overline{X}_{0,0} + \alpha_{1,1}\overline{X}_{0,0}\overline{Y}_{1,0}\overline{X}_{0,0} \\ \qquad + \frac{1}{2}\gamma_{2,0}\overline{X}^{[2]}_{0,0}\overline{X}_{0,1} + (\overline{X}_{1,0}\alpha_{1,1}\overline{X}_{0,0}\overline{Y}_{0,0} + \alpha_{1,1}\overline{X}_{1,0}\overline{X}_{0,0}\overline{Y}_{0,0}) \\ \qquad - \frac{1}{2}\overline{X}_{2,0}\overline{X}^{[2]}_{0,0} - \frac{1}{2}\overline{X}_{0,2}\overline{Y}^2_{0,0} - \overline{X}_{0,1}\overline{Y}_{0,1}\overline{Y}_{0,0}, \end{cases} \tag{2.53}$$

and

$$\begin{cases} \Sigma_1 = -\overline{Y}_{0,0}, \\ \Sigma_2 = \overline{Y}_{1,0}\overline{X}_{0,0} + \overline{Y}_{0,1}\overline{Y}_{0,0} - \frac{1}{2}\gamma_{2,0}\overline{X}^{[2]}_{0,0} - \frac{1}{2}\gamma_{0,2}\overline{Y}^2_{0,0}, \\ \Sigma_3 = \frac{1}{6}(-3\gamma^2_{0,2} + \gamma_{0,3})\overline{Y}^3_{0,0} - \overline{Y}_{0,0}\overline{Y}_{1,0}\overline{X}_{0,1} - \overline{Y}_{0,0}\overline{Y}_{1,1}\overline{X}_{0,0} - \frac{1}{2}\overline{Y}_{2,0}\overline{X}^{[2]}_{0,0} \\ \qquad - \frac{1}{2}\overline{Y}^2_{0,0}\overline{Y}_{0,2} - \overline{Y}_{1,0}\overline{X}_{1,0}\overline{X}_{0,0} - \overline{Y}_{0,0}\overline{Y}^2_{0,1} - \overline{Y}_{0,0}\overline{Y}_{1,0}\overline{X}_{0,1} \\ \qquad + \frac{1}{2}\gamma_{2,0}\overline{X}^{[2]}_{0,0}\overline{Y}_{0,1} + (\overline{Y}_{1,0}\alpha_{1,1}\overline{X}_{0,0}\overline{Y}_{0,0} + \gamma_{0,2}\overline{Y}_{1,0}\overline{X}_{0,0}\overline{Y}_{0,0}) \\ \qquad + \frac{3}{2}\gamma_{0,2}\overline{Y}^2_{0,0}\overline{Y}_{0,1} + \gamma_{2,0}((\overline{X}_{1,0}\overline{X}_{0,0}) \otimes \overline{X}_{0,0}) + \gamma_{2,0}(\overline{X}_{0,0} \otimes \overline{X}_{0,1})\overline{Y}_{0,0} \\ \qquad - \gamma_{2,0}((\alpha_{1,1}\overline{X}_{0,0}) \otimes \overline{X}_{0,0})\overline{Y}_{0,0} - \frac{1}{2}\gamma_{0,2}\gamma_{2,0}\overline{X}^{[2]}_{0,0}\overline{Y}_{0,0} + \frac{1}{2}\gamma_{2,1}\overline{X}^{[2]}_{0,0}\overline{Y}_{0,0}. \end{cases} \tag{2.54}$$

Chapter 3. A Statistic Related to Scores Type Test for Some Ordinary M–estimation Problems

In Chapter 2, we discussed the asymptotic expansion of an ordinary GM–estimator and ended up with an explicit asymptotic expression (2.44). Thus any test statistic based on GM–estimators can be further investigated by using this result.

To make it clear, let $g(\epsilon, \boldsymbol{\xi}_n)$ be any function of a GM–estimator $\boldsymbol{\xi}_n$ that has a Taylor expansion with respect to $\boldsymbol{\xi}_n$ in a neighbourhood of $\boldsymbol{\xi}_0$:

$$g(\epsilon, \boldsymbol{\xi}_n) = g(\epsilon, \boldsymbol{\xi}_0) + \sum_{\nu=1}^{s}((\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^{[\nu]^T} \otimes I_q)\frac{vec(\frac{\partial \nu}{\partial \boldsymbol{\xi}^\nu} g(\epsilon, \boldsymbol{\xi}))|_{\boldsymbol{\xi}_0}}{\nu!} + R_{s+1}(\epsilon, \boldsymbol{\xi}_n) , \quad (3.1)$$

then by plugging in (2.44), the terms in (3.1) can be collected according to the powers of n to get an asymptotic expansion of $g(\epsilon, \boldsymbol{\xi}_n)$. Next, by applying Edgeworth expansion again, one can find the asymptotic distribution of $g(\epsilon, \boldsymbol{\xi}_n)$. The validity of this procedure is assured by the result from Bhattacharya and Ghosh(1978) under the assumptions **A1)**–**A6)**. Furthermore, by choosing $g(\epsilon, \boldsymbol{\xi}_n)$ appropriately, it is possible to get some statistic $g_0(\epsilon, \boldsymbol{\xi}_n)$ which is an easily explained and implemented modification to the normal theory test statistics such as 't' or 'F'.

In this chapter, we will apply this procedure to some function of the scores type test statistic W_n^2 described in Chapter 2 for location and regression M–estimation problems. We choose W_n^2 under investigation because it relies only on the estimation of scale and $g(\epsilon, \boldsymbol{\xi}_n)$ will have a much simpler form than that of the others. In Section 1, the simplest case—M–estimation of location with scale known—will be investigated. As we will see that, $g(\epsilon, \xi_n)$ in this case is a constant with respect to ξ_n. In Section 2, the linear regression M–estimation problem, where $g(\epsilon, \xi_n)$ is a multivariate function of estimators, will be discussed.

§3.1 M–estimation of location parameter with scale known

The model of interest is the same as in example 2.1. Then the scores type test statistic for testing the null hypothesis

$$H_0 : \theta = 0$$

is given by W_n^2, with

$$W_n = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}\psi(\epsilon_i)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\psi^2(\epsilon_i)}} . \tag{3.2}$$

Notice that this form is quite simple since no estimator of parameter has been involved, i.e. we do not even need to know the distribution of the M–estimator, hence the result from Chapter 2 will not be used here. However, since this case shares many techniques with some other common cases, we will investigate it first.

§3.1.1 The asymptotic distribution of W_n

Define

$$\overline{Z} = \begin{pmatrix} \overline{Z}_1 \\ \overline{Z}_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} Z_{1i} \\ \frac{1}{n}\sum_{i=1}^{n} Z_{2i} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\psi(\epsilon_i) \\ \frac{1}{n}\sum_{i=1}^{n}\psi^2(\epsilon_i) \end{pmatrix}, \tag{3.3}$$

then $E(\overline{Z}) = \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ E[\psi^2(\epsilon)] \end{pmatrix}$ and W_n can be rewritten as $W_n = \sqrt{n}(H(\overline{Z}) - H(\mu))$ where $H(a, b) = a/\sqrt{b}$ (so $H(\mu) = 0$). Next, we follow the common practice to calculate the "approximate moments" of W_n by expanding $H(\overline{Z})$

around μ, keeping a certain number of terms, raising to an appropriate power and taking expectations term by term (the so-called delta method). These "approximate moments" will then be used to obtain a formal Edgeworth expansion of the distribution function of W_n. The validity of this procedure has been proven by Bhattacharya and Ghosh(1978).

Now, a Taylor expansion of $H(\overline{Z})$ around $\mu = \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}$ yields the statistic

$$W_n' = \sqrt{n}(\frac{\overline{Z}_1}{\sqrt{\mu_2}}\sum_{k=0}^{s-2}(-\frac{1}{2})^k(\frac{(2k-1)!!}{\mu_2})(\frac{Z_2-\mu_2}{\mu_2})^k)$$

with $W_n = W_n' + o_p(n^{-(s-2)/2})$ for $s \geq 2$. Take $s = 4$, then we have

$$W_n' = \sqrt{n}\frac{\overline{Z}_1}{\sqrt{\mu_2}}(1 - \frac{1}{2}\frac{(\overline{Z}_2-\mu_2)}{\mu_2} + \frac{3}{8}(\frac{\overline{Z}_2-\mu_2}{\mu_2})^2) \tag{3.4}$$

and $W_n = W_n' + o_p(n^{-1})$. It can be expected that an asymptotic expansion of the distribution function of W_n' will coincide with that of W_n.

In order to get an asymptotic expansion of the distribution function of W_n', the first four cumulants have to be calculated (approximately) first. For example, if we put $U_i = \frac{\psi(\epsilon_i)}{\sqrt{\mu_2}}$, $V_i = \frac{\psi^2(\epsilon_i)}{\mu_2} - 1 = U_i^2 - 1$, then $E(U_i) = 0$, $var(U_i) = 1$, $E(V_i) = 0$ and

$$\begin{aligned}
E[W_n'] =&\sqrt{n}E[\overline{u}(-1 - \frac{1}{2}\overline{v} + \frac{3}{8}\overline{v}^2)] \\
=&\sqrt{n}(-\frac{1}{2}\sum_{i,j=1}^{n}\frac{E[U_iV_j]}{n^2} + \frac{3}{8}\sum_{i,j,k=1}^{n}\frac{E[U_iV_jV_k]}{n^3}) \\
=&-\frac{1}{2\sqrt{n}}E[U_1^3] \\
=&O(n^{-2}) \qquad \text{(Assuming } \epsilon \text{ has a symmetric distribution).}
\end{aligned}$$

By raising W_n' to an appropriate power and taking expectations term by term, it is easy to check that

$$\begin{cases} E[W'_n] = O(n^{-2}), \\ E[W'^2_n] = 1 + O(n^{-2}), \\ E[W'^3_n] = O(n^{-2}), \\ E[W'^4_n] = 3 - \frac{2}{n}E[U_1^4] + O(n^{-2}). \end{cases} \tag{3.5}$$

Now if we write $\kappa = E[U_1^4] - 3$, then from (3.5) the first four cumulants of W'_n are given by

$$\begin{cases} \kappa_1 = O(n^{-2}), \\ \kappa_2 = 1 + O(n^{-2}), \\ \kappa_3 = O(n^{-2}), \\ \kappa_4 = -\frac{2k+6}{n} + O(n^{-2}). \end{cases} \tag{3.6}$$

Proceeding exactly as in example 2.1 yields the approximate density and distribution functions of W_n:

$$\begin{cases} f_{W_n}(x) = (1 - \frac{\kappa+3}{12n}H_4(x))\phi(x) + O(n^{-2}) \\ F_{W_n}(x) = \Phi(x) + \frac{\kappa+3}{12n}H_3(x)\phi(x) + O(n^{-2}) \end{cases} \tag{3.7}$$

where $H_3(x) = x^3 - x$, $H_4(x) = x^4 - 6x + 3$ are the Hermite polynomials of order 3 and 4 respectively.

§3.1.2 A test statistic related to W_n

Once the probability density function of W_n is obtained as in (3.7), we can then proceed to get the moment generating function of W_n^2:

$$E[e^{tW_n^2}] = \int e^{tx^2}(1 - \frac{\kappa+3}{12n}H_4(x))\phi(x)dx + O(n^{-2}) \ .$$

With some algebra this is

$$E[e^{tW_n^2}] = (1-\frac{A}{n})(1-2t)^{-1/2} + \frac{2A}{n}(1-2t)^{-3/2} - \frac{A}{n}(1-2t)^{-5/2} + O(n^{-2}) \tag{3.8}$$

with $A = \frac{\kappa+3}{4}$.

It can be further verified that the mean and variance of W_n^2 are

$$\begin{cases} \mu(W_n^2) = 1 + O(n^{-2}), \\ \sigma^2(W_n^2) = 2 - \frac{2\kappa+6}{n} + O(n^{-2}) \ . \end{cases} \tag{3.9}$$

Remark 1. From (3.8), it is obvious that to the order of $O(n^{-1})$, $E[e^{tW_n^2}] = (1-2t)^{-1/2}$, which is the moment generating function of the χ_1^2 distribution. i.e. the χ_1^2 approximation of W_n^2 is of order $O_p(n^{-1})$.

Remark 2. It is known that if the random variable X_n has a distribution tending to the χ_q^2 as $n \longrightarrow \infty$, then the monent generating function of X_n is of the form $(1-2t)^{-d/2}(1+\frac{2at}{n(1-2t)}) + O(n^{-2})$ for some constant a(Barndorff-Nielsen and Cox, 1989). Thus to the order of $O_p(n^{-2})$, W_n^2 can not be approximated by a single χ^2 with a scaling constant and an adjustment to the degree of freedom.

Remark 3. When the higher order moments of W_n^2 have been calculated, it has been found that W_n^2 behaves like a random variable with a F distribution. In fact, if we let $\psi(x) = x$ (i.e. the least squares estimation) and assume that ϵ has a $N(0,1)$ distribution, then the statistic $Q_n = (1-\frac{1}{n})\frac{W_n^2}{1-\frac{W_n^2}{n}}$ has a F_{n-1}^1 distribution. However, what happens to Q_n if we do not assume normality and least squares criteria? It seems natural to expect that Q_n would also be approximately F distributed, possibly with a scaling constant and an adjustment to the degrees of freedom.

From Remark 3, it then becomes necessary to investigate the statistic Q_n first. Now, since Q_n is a function of W_n, from (3.7), the moment generating function of Q_n is given by

$$
\begin{aligned}
E[e^{tQ_n}] =& E[e^{\frac{t(1-\frac{1}{n})W_n^2}{1-\frac{W_n^2}{n}}}] \\
=& \int e^{\frac{t(1-\frac{1}{n})x^2}{1-\frac{x^2}{n}}} \phi(x)(1 - \frac{\kappa+3}{12n}H_4(x))dx + O(n^{-2}) \ . \qquad (3.10)
\end{aligned}
$$

By noticing that

$$
\begin{aligned}
e^{\frac{tx^2(1-\frac{1}{n})}{1-\frac{x^2}{n}}} =& e^{tx^2(1+\frac{x^2-1}{n}+O(n^{-2}))} \\
=& e^{tx^2} + \frac{1}{n}e^{tx^2}tx^2(x^2-1) + O(n^{-2}) \ .
\end{aligned}
$$

(3.10) becomes

$$
E[e^{tQ_n}] = E[e^{tW_n^2}] + \frac{t}{n}I + O(n^{-2}) \ , \qquad (3.11)
$$

where

$$
\begin{aligned}
I =& \int e^{tx^2}x^2(x^2-1)\phi(x)dx \\
=& 12t^2(1-2t)^{-5/2} + 10t(1-2t)^{-3/2} + 2(1-2t)^{-1/2} \\
=& \frac{1}{\sqrt{1-2t}}(\frac{3}{(1-2t)^2} - \frac{1}{1-2t}) \ . \qquad (3.12)
\end{aligned}
$$

Substituting (3.8), (3.12) into (3.11), one has

$$
\begin{aligned}
E[e^{tQ_n}] =& (1-\frac{A}{n})(1-2t)^{-1/2} + (\frac{2A-t}{n})(1-2t)^{-3/2} \\
&+ \frac{(3t-A)}{n}(1-2t)^{-5/2} + O(n^{-2}) \ . \qquad (3.13)
\end{aligned}
$$

From (3.13) one can further have

$$
E[Q_n^r] = (2r-1)!!(1 + \frac{2r^2}{n} + \frac{r(1-r)(\kappa+3)}{3n}) + O(n^{-2}) \ , \quad r = 1, 2, \cdots \qquad (3.14)
$$

which gives us the approximate uncentered moments of statistic Q_n. A deeper investigation of (3.14) shows that all the leading terms of these moments agree with that of an F–distributed random variable. Therefore, it is natural to use a scaled and degree of freedoms modified F to approximate Q_n. Our proposal here is to assume that Q_n has a $(1 - \frac{\alpha}{n-1})F^1_{(n-1)(1-\beta)}$ distribution. Then by matching the first two moments of Q_n with $(1 - \frac{\alpha}{n-1})F^1_{(n-1)(1-\beta)}$, we can fix the values for α and β and then check if they also agree to the high orders.

To make it clear, let μ'_1, μ'_2 be the first two uncentered moments of $(1-\frac{\alpha}{n-1})F^1_{(n-1)(1-\beta)}$, then we have

$$\begin{cases} \mu'_1 = 1 + \frac{2-\alpha+\alpha\beta}{n(1-\beta)} + O(n^{-2}) , \\ \mu'_2 = 3 + \frac{3(6-2\alpha+2\alpha\beta)}{n(1-\beta)} + O(n^{-2}) . \end{cases} \tag{3.15}$$

On the other hand, from (3.14), it follows that

$$\begin{cases} E[Q_n] = 1 + \frac{2}{n} + O(n^{-2}). \\ E[Q_n^2] = 3 + \frac{18-2\kappa}{n} + O(n^{-2}). \end{cases} \tag{3.16}$$

Now, by matching $E[Q_n]$ with μ'_1 and $E[Q_n^2]$ with μ'_2, we have $\alpha = -\frac{2}{3}\kappa$ and $\beta = \frac{\kappa}{\kappa-3}$. Next, in order to compare the higher order moments, the moments of $(1 - \frac{\alpha}{n-1})F^1_{(n-1)(1-\beta)}$ have to be evaluated first as follows. It is well known that the r^{th} uncentered moment of a $F^{\nu_1}_{\nu_2}$ random variable X is given by

$$E[X^r] = \frac{\Gamma(\frac{\nu_1+2r}{2})\Gamma(\frac{\nu_2-2r}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})}(\frac{\nu_2}{\nu_1})^r, \quad r < \frac{\nu_2}{2} . \tag{3.17}$$

When (3.17) is applied to our problem, it becomes

$$E[((1+\frac{2\kappa}{3(n-1)})F^1_{\frac{3(n-1)}{3-\kappa}})^r] = \frac{(1+\frac{2\kappa}{3(n-1)})^r\Gamma(\frac{2r+1}{2})\Gamma(\frac{\frac{3(n-1)}{3-\kappa}-2r}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{3(n-1)}{2(3-\kappa)})}(\frac{3(n-1)}{3-\kappa})^r ,$$

which can be further rewritten as

$$E[((1+\frac{2\kappa}{3(n-1)})F^1_{\frac{3(n-1)}{3-\kappa}})^r] = (2r-1)!!(1+\frac{2\kappa}{3(n-1)})^r\Pi^r_{j=1}\frac{\frac{3(n-1)}{3-\kappa}}{\frac{3(n-1)}{3-\kappa}-2j} \ . \tag{3.18}$$

When (3.18) is expanded in powers of n^{-1}, we obtain

$$E[((1+\frac{2\kappa}{3(n-1)})F^1_{\frac{3(n-1)}{3-\kappa}})^r] = (2r-1)!!(1+\frac{2r^2}{n}+\frac{r(1-r)(\kappa+3)}{3n})+O(n^{-2}) \ . \tag{3.19}$$

Now, by comparing (3.19) with (3.14), we see that all the moments of statistic Q_n agree with that of a $(1-\frac{2\kappa}{3(n-1)})F^1_{\frac{3(n-1)}{3-\kappa}}$ random variable to the order $O(n^{-2})$ and therefore the characteristic functions also agree to that order. This can be summarized by Theorem 3.1.

Theorem 3.1 Assume that **A1)–A6)** hold. Then to the order $O_p(n^{-2})$, the test statistic $Q_n = (1-\frac{1}{n})\frac{W_n^2}{1-\frac{W_n^2}{n}}$ is approximately $(1-\frac{\alpha}{n-1})F^1_{(n-1)(1-\beta)}$-distributed, with $W_n = \frac{\sum\psi(\epsilon_i)}{\sqrt{\sum\psi^2(\epsilon_i)}}$, $\alpha = -\frac{2}{3}\kappa$, $\beta = \frac{\kappa}{\kappa-3}$ and $\kappa = \frac{E\psi^4(\epsilon)}{(E\psi^2(\epsilon))^2} - 3$.

Remark 1 Under normality and least squares, we have $\psi(x) = x$, $\kappa = 0$, therefore $\alpha = \beta = 0$ and Q_n will have exactly a F^1_{n-1} distribution which agrees with the ordinary theorem.

Remark 2 This approximation is valid for $\kappa < 3$. As we know,since $\kappa = 3$ corresponds to the case where $\psi(x) = x$ and ϵ has a double exponential distribution, which has a very thick tail, it is virtually not a restraint at all.

§3.2 Linear regression M–estimation problem with scale unknown

In this section, the linear regression model under investigation is as given in Example 2.2. The hypotheses of interest are the same as in (1.3) except we will only focus on the special case $m-p=1$. i.e.,we only test if some specified scalar parameter θ_2 is equal to zero. Then the scores type test statistic for testing the null hypothesis in (1.3) is also given by W_n^2, but with

$$W_n = \frac{\sum_{i=1}^n \psi(\frac{y_i - \mathbf{x}_{i,1}^T \hat{\theta}_n}{\hat{\sigma}_n}) x_{i,2}}{\sqrt{\sum_{i=1}^n \psi^2(\frac{y_i - \mathbf{x}_{i,1}^T \hat{\theta}_n}{\hat{\sigma}_n}) x_{i,2}^2}} \tag{3.20}$$

where $\hat{\boldsymbol{\theta}}_n$ $\hat{\sigma}$ are the M-estimators under null hypothesis defined by

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n \psi(\frac{y_i - \mathbf{x}_{i,1}^T \hat{\theta}_n}{\hat{\sigma}_n}) x_{i,1} = 0 \\ \frac{1}{n}\sum_{i=1}^n \chi(\frac{y_i - \mathbf{x}_{i,1}^T \hat{\theta}}{\hat{\sigma}}) = 0 \ , \end{cases} \tag{3.21}$$

for some odd function ψ and even function χ, where ψ and χ are both assumed to be continuous and piecewise differentiable at least three times.

Unlike the situation in Section 3.1 where no estimator has been involved in the test statistic, we are unable to write W_n as a function $H(\overline{Z})$, where $\overline{Z}$ is a vector of averages of independent random variables. So the delta method used in section 3.1 can not be applied here directly to W_n. In fact, W_n has to be approximated first by some statistic W_n^* to which the delta method applies.

W_n^* can be obtained by using Taylor expansion twice as follows. First, define

$$\begin{cases} \overline{Z}(\boldsymbol{\xi}_n) = \frac{1}{n}\sum_{i=1}^n \psi(\frac{y_i - \mathbf{x}_{i,1}^T \hat{\theta}_n}{\hat{\sigma}_n}) x_{i,2}, \\ \overline{H}(\boldsymbol{\xi}_n) = \frac{1}{n}\sum_{i=1}^n \psi^2(\frac{y_i - \mathbf{x}_{i,1}^T \hat{\theta}_n}{\hat{\sigma}_n}) x_{i,2}^2, \end{cases} \tag{3.22}$$

then $W_n = \sqrt{n}\frac{\overline{Z}(\boldsymbol{\xi}_n)}{\sqrt{\overline{H}(\boldsymbol{\xi}_n)}}$ and a Taylor expansion of W_n with respect to $\boldsymbol{\xi}_n = (\hat{\boldsymbol{\theta}}_n, \hat{\sigma})$

around the true parameter $\boldsymbol{\xi}_0 = (\theta_0, \sigma)$ gives

$$\begin{aligned} W_n &= W_n(\boldsymbol{\xi}_n) \\ &= Wn(\boldsymbol{\xi}_0) + \sum_{\nu=1}^{3} (\frac{\boldsymbol{\xi}_n - \boldsymbol{\xi}_0}{\sigma})^{[\nu-1]T} \frac{W_n^{(\nu)}(\boldsymbol{\xi}_0)}{\nu!} (\frac{\boldsymbol{\xi}_n - \boldsymbol{\xi}_0}{\sigma}) + O_p(n^{-2}) \ , \end{aligned} \quad (3.23)$$

where $W_n^{(\nu)}(\boldsymbol{\xi}_0) = \frac{\partial^\nu}{\partial \xi^\nu} W_n(\boldsymbol{\xi}_n)|_{\xi=\xi_0}$ for $\nu = 1, 2, 3$ are given as follows.

$$\begin{cases} W_n^{(1)}(\boldsymbol{\xi}_0) = \sqrt{n}[\overline{H}^{-\frac{1}{2}} \frac{\partial \overline{Z}}{\partial \xi} - \frac{1}{2} \overline{Z}\ \overline{H}^{-\frac{3}{2}} \frac{\partial \overline{H}}{\partial \xi}]|_{\xi_n=\xi_0}, \\ W_n^{(2)}(\boldsymbol{\xi}_0) = \sqrt{n}[\overline{H}^{-\frac{1}{2}} \frac{\partial^2 \overline{Z}}{\partial \xi^2} - \frac{1}{2} \overline{Z}\ \overline{H}^{-\frac{3}{2}} \frac{\partial^2 \overline{H}}{\partial \xi^2} - \frac{1}{2} \overline{H}^{-\frac{3}{2}} (\frac{\partial \overline{Z}}{\partial \xi})^T \frac{\partial \overline{H}}{\partial \xi} \\ \qquad - \frac{1}{2} \overline{H}^{-\frac{3}{2}} (\frac{\partial \overline{H}}{\partial \xi})^T \frac{\partial \overline{Z}}{\partial \xi} + \frac{3}{4} \overline{Z}\ \overline{H}^{-\frac{5}{2}} (\frac{\partial \overline{H}}{\partial \xi})^T \frac{\partial \overline{H}}{\partial \xi}]|_{\xi_n=\xi_0}, \\ W_n^{(3)}(\boldsymbol{\xi}_0) = \sqrt{n}[\overline{H}^{-\frac{1}{2}} \frac{\partial^3 \overline{Z}}{\partial \xi^3} - \frac{1}{2} \overline{H}^{-\frac{3}{2}} \frac{\partial \overline{H}}{\partial \xi} \otimes vec(\frac{\partial^2 \overline{Z}}{\partial \xi^2}) + \overline{Z}\ \overline{H}^{-\frac{3}{2}} \frac{\partial^3 \overline{H}}{\partial \xi^3} \\ \qquad + \overline{H}^{-\frac{3}{2}} \frac{\partial \overline{Z}}{\partial \xi} \otimes vec(\frac{\partial^2 \overline{H}}{\partial \xi^2}) - \overline{Z}\ \overline{H}^{-\frac{5}{2}} \frac{\partial \overline{H}}{\partial \xi} \otimes vec(\frac{\partial^2 \overline{H}}{\partial \xi^2}) \\ \qquad + \overline{H}^{-\frac{3}{2}} \frac{\partial}{\partial \xi}((\frac{\partial \overline{Z}}{\partial \xi})^T \frac{\partial \overline{H}}{\partial \xi}) - \frac{3}{2} \overline{H}^{-\frac{5}{2}} \frac{\partial \overline{H}}{\partial \xi} \otimes vec((\frac{\partial \overline{X}}{\partial \xi})^T \frac{\partial \overline{H}}{\partial \xi}) \\ \qquad + \overline{H}^{-\frac{3}{2}} \frac{\partial}{\partial \xi}((\frac{\partial \overline{H}}{\partial \xi})^T \frac{\partial \overline{Z}}{\partial \xi}) - \frac{3}{2} \overline{H}^{-\frac{5}{2}} \frac{\partial \overline{H}}{\partial \xi} \otimes vec((\frac{\partial \overline{H}}{\partial \xi})^T \frac{\partial \overline{Z}}{\partial \xi}) \\ \qquad + \overline{Z}\ \overline{H}^{-\frac{5}{2}} \frac{\partial}{\partial \xi}((\frac{\partial \overline{H}}{\partial \xi})^T \frac{\partial \overline{H}}{\partial \xi}) + \overline{H}^{-\frac{5}{2}} \frac{\partial \overline{Z}}{\partial \xi} \otimes vec((\frac{\partial \overline{H}}{\partial \xi})^T \frac{\partial \overline{H}}{\partial \xi}) \\ \qquad - \frac{5}{2} \overline{Z}\ \overline{H}^{-\frac{7}{2}} \frac{\partial \overline{H}}{\partial \xi} \otimes vec((\frac{\partial \overline{H}}{\partial \xi})^T \frac{\partial \overline{H}}{\partial \xi})]|_{\xi_n=\xi_0}. \end{cases} \quad (3.24)$$

Now, if we define

$$\begin{cases} L_Z = \frac{\partial \overline{Z}}{\partial \xi}(\frac{\xi_n - \xi_0}{\sigma}), L_H = \frac{\partial \overline{H}}{\partial \xi}(\frac{\xi_n - \xi_0}{\sigma}), \\ Q_Z = (\frac{\xi_n - \xi_0}{\sigma})^T \frac{\partial^2 \overline{Z}}{\partial \xi^2}(\frac{\xi_n - \xi_0}{\sigma}), Q_H = (\frac{\xi_n - \xi_0}{\sigma})^T \frac{\partial^2 \overline{H}}{\partial \xi^2}(\frac{\xi_n - \xi_0}{\sigma}), \\ C_Z = (\frac{\xi_n - \xi_0}{\sigma})^{[2]^T} B_Z(\frac{\xi_n - \xi_0}{\sigma}), C_H = (\frac{\xi_n - \xi_0}{\sigma})^{[2]^T} B_H(\frac{\xi_n - \xi_0}{\sigma}), \end{cases} \quad (3.25)$$

with $B_Z = \frac{\partial \overline{Z}}{\partial \xi}(\frac{\partial}{\partial \xi})^{[2]}, B_H = \frac{\partial \overline{H}}{\partial \xi}(\frac{\partial}{\partial \xi})^{[2]}$, then with some algebra,we have

$$\begin{cases} Wn(\boldsymbol{\xi}_0) = \sqrt{n}\overline{Z}\ \overline{H}^{-\frac{1}{2}}, \\ W_n^{(1)}(\boldsymbol{\xi}_0)(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0) = \sqrt{n}(\overline{H}^{-\frac{1}{2}}L_Z - \frac{1}{2}\overline{Z}\ \overline{H}^{-\frac{3}{2}}L_H), \\ (\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^T W_n^{(2)}(\boldsymbol{\xi}_0)(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0) = \sqrt{n}(\overline{H}^{-\frac{1}{2}}Q_Z - \frac{1}{2}\overline{Z}\ \overline{H}^{-\frac{3}{2}}Q_H - \overline{H}^{-\frac{3}{2}}L_Z L_H + \frac{3}{4}\overline{Z}\ \overline{H}^{-\frac{5}{2}}L_H^2), \\ (\boldsymbol{\xi}_n - \boldsymbol{\xi}_0)^{[2]T} W_n^{(3)}(\boldsymbol{\xi}_0)(\boldsymbol{\xi}_n - \boldsymbol{\xi}_0) = \sqrt{n}(\overline{H}^{-\frac{1}{2}}C_Z - \frac{3}{2}\overline{H}^{-\frac{3}{2}}L_H Q_Z - \frac{1}{2}\overline{Z}\ \overline{H}^{-\frac{3}{2}}C_H - \frac{3}{2}\overline{H}^{-\frac{3}{2}}L_Z Q_H \\ \qquad + \frac{9}{4}\overline{Z}\ \overline{H}^{-\frac{5}{2}}L_H Q_H + \frac{9}{4}\overline{H}^{-\frac{5}{2}}L_H^2 L_Z - \frac{15}{8}\overline{Z}\ \overline{H}^{-\frac{7}{2}}L_H^3), \end{cases} \tag{3.26}$$

and therefore W_n can be represented in terms of $\overline{Z}, \overline{H}, L_Z, L_H, Q_Z, Q_H, C_Z$ and C_H.

Secondly, if we define

$$\begin{cases} \nu_{0,0} = E[\psi^2(\frac{y-\mathbf{x}_1^T\theta_0}{\sigma})x_2], \\ \nu_{i,j} = \sigma^{i+j}\nu_{0,0}^{-1}E[(\frac{\partial^j}{\partial\sigma^j}\psi^2(\frac{y-\mathbf{x}_1^T\theta_0}{\sigma})x_2)(\frac{\partial}{\partial\theta})^{[i]T}] \text{ if } (i,j) \neq (0,0), \\ \beta_{i,j} = \sigma^{i+j}\nu_{0,0}^{-\frac{1}{2}}E[(\frac{\partial^j}{\partial\sigma^j}\psi(\frac{y-\mathbf{x}_1^T\theta_0}{\sigma})x_2)(\frac{\partial}{\partial\theta})^{[i]T}], \\ \overline{Z}_{i,j} = \frac{1}{n}\sigma^{i+j}\nu_{0,0}^{-\frac{1}{2}}\sum_{k=1}^n[(\frac{\partial^j}{\partial\sigma^j}\psi(\frac{y_k-\mathbf{x}_{k,1}^T\theta_0}{\sigma})x_{k,2})(\frac{\partial}{\partial\theta})^{[i]T}], \qquad \overline{Z}^*_{i,j} = \overline{Z}_{i,j} - \beta_{i,j}, \\ \overline{H}_{i,j} = \frac{1}{n}\sigma^{i+j}\nu_{0,0}^{-1}\sum_{k=1}^n[(\frac{\partial^j}{\partial\sigma^j}\psi^2(\frac{y-\mathbf{x}_{k,1}^T\theta_0}{\sigma})x_2)(\frac{\partial}{\partial\theta})^{[i]T}], \qquad \overline{H}^*_{i,j} = \overline{H}_{i,j} - \nu_{i,j}, \end{cases} \tag{3.27}$$

then by noticing that $X^TX = I_m$ and ψ is an odd function, we have

$$\begin{cases} \beta_{i,j} = \mathbf{0}, \quad \text{if } i,j < 3 \text{ and } i+j \leq 3, \\ \nu_{1,0} = \nu_{1,1} = \nu_{1,2} = \mathbf{0}_{1\times p}, \nu_{3,0} = \mathbf{0}_{1\times p^3}, \end{cases} \tag{3.28}$$

and thus if we let

$$\begin{cases} L_Z^* = (\frac{\partial\overline{Z}}{\partial\xi} - E[\frac{\partial\overline{Z}}{\partial\xi}])(\frac{\xi_n-\xi_0}{\sigma}), L_H^* = (\frac{\partial\overline{H}}{\partial\xi} - E[\frac{\partial\overline{H}}{\partial\xi}])(\frac{\xi_n-\xi_0}{\sigma}), \\ Q_Z^* = (\frac{\xi_n-\xi_0}{\sigma})^T(\frac{\partial^2\overline{Z}}{\partial\xi^2} - E[\frac{\partial^2\overline{Z}}{\partial\xi^2}])(\frac{\xi_n-\xi_0}{\sigma}), Q_H^* = (\frac{\xi_n-\xi_0}{\sigma})^T(\frac{\partial^2\overline{H}}{\partial\xi^2} - E[\frac{\partial^2\overline{H}}{\partial\xi^2}])(\frac{\xi_n-\xi_0}{\sigma}), \\ C_Z^* = (\frac{\xi_n-\xi_0}{\sigma})^T(B_Z - E[B_Z])(\frac{\xi_n-\xi_0}{\sigma})^{[2]}, C_H^* = (\frac{\xi_n-\xi_0}{\sigma})^T(B_H - E[B_H])(\frac{\xi_n-\xi_0}{\sigma})^{[2]}, \end{cases} \tag{3.29}$$

then one can show that

$$\begin{cases} L_Z = L_Z^*, Q_Z = Q_Z^*, C_Z = C_Z^* + \beta_{3,0}\Theta^{[3]}, \\ L_H = L_H^* + \nu_{0,1}\Sigma, \\ Q_H = Q_H^* + \nu_{2,0}\Theta^{[2]} + \nu_{0,2}\Sigma^2, \\ C_H = C_H^* + 3\nu_{2,1}\Theta^{[2]}\Sigma + \nu_{0,3}\Sigma^3, \end{cases} \tag{3.30}$$

with $\Theta = (\frac{\hat{\theta}-\theta_0}{\sigma})$, $\Sigma = (\frac{\hat{\sigma}-\sigma}{\sigma})$ and W_n can be rewritten as

$$W_n = \widetilde{\widetilde{\widetilde{W_n^*}}} + O_p(n^{-2}) = \widetilde{\widetilde{\widetilde{W_1^*}}} + \widetilde{\widetilde{\widetilde{W_2^*}}} + \widetilde{\widetilde{\widetilde{W_3^*}}} + \widetilde{\widetilde{\widetilde{W_4^*}}} + O_p(n^{-2}) \tag{3.31}$$

where

$$\begin{cases} \widetilde{\widetilde{\widetilde{W_1^*}}} = \sqrt{n}\overline{Z}\ \overline{H}^{-\frac{1}{2}}, \\ \widetilde{\widetilde{\widetilde{W_2^*}}} = \sqrt{n}(L_Z^*\overline{H}^{-\frac{1}{2}} - \frac{1}{2}\nu_{0,1}\overline{Z}\Sigma\overline{H}^{-\frac{3}{2}}), \\ \widetilde{\widetilde{\widetilde{W_3^*}}} = \sqrt{n}(-\frac{1}{2}\overline{Z}L_H^*\overline{H}^{-\frac{3}{2}} + \frac{1}{8}\overline{H}^{-\frac{5}{2}}(-2\nu_{0,2}\overline{H} + 3\nu_{0,1}^2)\overline{Z}\Sigma^2 - \frac{1}{4}\nu_{2,0}\Theta^{[2]}\overline{Z}\ \overline{H}^{-\frac{3}{2}} \\ \qquad - \frac{1}{2}\nu_{0,1}\overline{H}^{-\frac{3}{2}}L_Z^*\Sigma + \frac{1}{2}\overline{H}^{-\frac{1}{2}}Q_Z^* + \frac{1}{6}\overline{H}^{-\frac{1}{2}}\beta_{3,0}\Theta^{[3]}), \\ \widetilde{\widetilde{\widetilde{W_4^*}}} = \sqrt{n}(\frac{3}{4}\nu_{0,1}\overline{H}^{-\frac{5}{2}}\overline{Z}L_H^*\Sigma - \frac{1}{4}\overline{Z}\ \overline{H}^{-\frac{3}{2}}Q_H^* - \frac{1}{48}\overline{H}^{-\frac{7}{2}}(4\nu_{0,3}\overline{H}^2 + 15\nu_{0,1}^3 - 18\nu_{0,1}\nu_{0,2})\overline{Z}\Sigma^3 \\ \qquad - \frac{1}{8}\overline{H}^{-\frac{7}{2}}(-3\overline{H}\nu_{0,1}\nu_{2,0} + 2\overline{H}^2\nu_{2,1})\Theta^{[2]}\overline{Z}\Sigma - \frac{1}{8}\overline{H}^{-\frac{7}{2}}(-3\overline{H}\nu_{0,1}^2 + 2\overline{H}^2\nu_{0,2})L_Z^*\Sigma^2 \\ \qquad - \frac{1}{2}\overline{H}^{-\frac{3}{2}}L_Z^*L_H^* - \frac{1}{4}\nu_{2,0}\Theta^{[2]}\overline{H}^{-\frac{3}{2}}L_Z^* - \frac{1}{4}\nu_{0,1}\overline{H}^{-\frac{3}{2}}\Sigma Q_Z^* + \frac{1}{6}\overline{H}^{-\frac{1}{2}}C_Z^*). \end{cases} \tag{3.32}$$

Now a Taylor expansion of $\widetilde{\widetilde{\widetilde{W_n}}}$ with respect to $\overline{H}$ around its expectation gives

$$\widetilde{\widetilde{\widetilde{W_n^*}}} = \widetilde{W_n^*} + O_p(n^{-2}) = \widetilde{W_1^*} + \widetilde{W_2^*} + \widetilde{W_3^*} + \widetilde{W_4^*} + O_p(n^{-2}) \tag{3.33}$$

with

$$
\begin{cases}
\widetilde{W_1^*} = \sqrt{n}\overline{Z}^*_{0,0}, \\
\widetilde{W_2^*} = \sqrt{n}(L^*_Z - \frac{1}{2}\nu_{0,1}\overline{Z}^*_{0,0}\Sigma - \frac{1}{2}\overline{Z}^*_{0,0}\overline{H}^*_{0,0}), \\
\widetilde{W_3^*} = \sqrt{n}(-\frac{1}{2}\overline{Z}^*_{0,0}L^*_H + \frac{1}{8}(-2\nu_{0,2} + 3\nu_{0,1}^2)\overline{Z}^*_{0,0}\Sigma^2 - \frac{1}{4}\nu_{2,0}\Theta^{[2]}\overline{Z}^*_{0,0} - \frac{1}{2}L^*_Z\overline{H}^*_{0,0} \\
\qquad - \frac{1}{2}\nu_{0,1}L^*_Z\Sigma + \frac{1}{2}Q^*_Z + \frac{1}{6}\beta_{3,0}\Theta^{[3]} + \frac{3}{4}\nu_{0,1}\overline{Z}^*_{0,0}\overline{H}^*_{0,0}\Sigma + \frac{3}{8}\overline{Z}^*_{0,0}\overline{H}^{*\,2}_{0,0}), \\
\widetilde{W_4^*} = \sqrt{n}(\frac{3}{4}\nu_{0,1}\overline{Z}^*_{0,0}L^*_H\Sigma - \frac{1}{4}\overline{Z}^*_{0,0}Q^*_H - \frac{1}{48}(4\nu_{0,3} + 15\nu_{0,1}^3 - 18\nu_{0,1}\nu_{0,2})\overline{Z}^*_{0,0}\Sigma^3 \\
\qquad - \frac{1}{8}(-3\nu_{0,1}\nu_{2,0} + 2\nu_{2,1})\Theta^{[2]}\overline{Z}^*_{0,0}\Sigma - \frac{1}{8}(-3\nu_{0,1}^2 + 2\nu_{0,2})L^*_Z\Sigma^2 + \frac{3}{4}\nu_{0,1}L^*_Z\Sigma\overline{H}^*_{0,0} \\
\qquad - \frac{1}{2}L^*_ZL^*_H - \frac{1}{4}\nu_{2,0}\Theta^{[2]}L^*_Z - \frac{1}{4}\nu_{0,1}\Sigma Q^*_Z + \frac{1}{6}C^*_Z - \frac{1}{12}\beta_{3,0}\Theta^{[3]}\overline{H}^*_{0,0} - \frac{5}{16}\overline{Z}^*_{0,0}\overline{H}^{*\,3}_{0,0} \\
\qquad - \frac{15}{16}\nu_{0,1}\overline{Z}^*_{0,0}\Sigma\overline{H}^{*\,2}_{0,0} + \frac{3}{8}\nu_{2,0}\Theta^{[2]}\overline{Z}^*_{0,0}\overline{H}^*_{0,0} + (\frac{3}{8}\nu_{0,2} - \frac{15}{16}\nu_{0,1}^2)\overline{Z}^*_{0,0}\Sigma^2\overline{H}^*_{0,0} \\
\qquad + \frac{3}{4}\overline{Z}^*_{0,0}L^*_H\overline{H}^*_{0,0} - \frac{1}{4}Q^*_Z\overline{H}^*_{0,0} + \frac{3}{8}L^*_Z\overline{H}^{*\,2}_{0,0}).
\end{cases}
\tag{3.34}
$$

Finally, by noticing that (3.29) can be rewritten as

$$
\begin{cases}
L^*_Z = \overline{Z}^*_{1,0}\Theta + \overline{Z}^*_{0,1}\Sigma, \\
L^*_H = \overline{H}^*_{1,0}\Theta + \overline{H}^*_{0,1}\Sigma, \\
Q^*_Z = \overline{Z}^*_{2,0}\Theta^{[2]} + 2\overline{Z}^*_{1,1}\Theta\Sigma + \overline{Z}^*_{0,2}\Sigma^2, \\
Q^*_H = \overline{H}^*_{2,0}\Theta^{[2]} + 2\overline{H}^*_{1,1}\Theta\Sigma + \overline{H}^*_{0,2}\Sigma^2, \\
C^*_Z = \overline{Z}^*_{3,0}\Theta^{[3]} + 3\overline{Z}^*_{2,1}\Theta^{[2]}\Sigma + 3\overline{Z}^*_{1,2}\Theta\Sigma^2 + \overline{Z}^*_{0,3}\Sigma^3, \\
C^*_H = \overline{H}^*_{3,0}\Theta^{[3]} + 3\overline{H}^*_{2,1}\Theta^{[2]}\Sigma + 3\overline{H}^*_{1,2}\Theta\Sigma^2 + \overline{H}^*_{0,3}\Sigma^3,
\end{cases}
\tag{3.35}
$$

substituting (3.35) and (2.52) into (3.34) yields

$$
\widetilde{W_n^*} = W_n^* + O_p(n^{-2}) = W_1^* + W_2^* + W_3^* + W_4^* + O_p(n^{-2}) \tag{3.36}
$$

with

$$\begin{cases} W_1^* = \sqrt{n}\overline{Z}^*_{0,0}, \\ W_2^* = \sqrt{n}(\frac{1}{2}\nu_{0,1}\overline{Z}^*_{0,0}\overline{Y}^*_{0,0} - \frac{1}{2}\overline{Z}^*_{0,0}\overline{H}^*_{0,0} - \overline{Z}^*_{0,1}\overline{Y}^*_{0,0} - \overline{Z}^*_{1,0}\overline{X}^*_{0,0}), \\ W_3^* = \sqrt{n}(-\frac{1}{2}\overline{Z}^*_{0,1}\gamma_{2,0}\overline{X}^{*\,[2]}_{0,0} + (\frac{1}{2}\gamma_{0,2} - \frac{1}{2}\nu_{0,1})\overline{Y}^{*\,2}_{0,0} + \overline{Z}^*_{0,1}\overline{Y}^*_{0,1}\overline{Y}^*_{0,0} + \overline{Z}^*_{0,1}\overline{Y}^*_{1,0}\overline{X}^*_{0,0} \\ \quad - \alpha_{1,1}\overline{X}^*_{0,0}\overline{Z}^*_{1,0}\overline{Y}^*_{0,0} - \frac{1}{2}\nu_{0,1}\overline{X}^*_{0,0}\overline{Z}^*_{1,0}\overline{Y}^*_{0,0} - \frac{3}{4}\nu_{0,1}\overline{H}^*_{0,0}\overline{Z}^*_{0,0}\overline{Y}^*_{0,0} - \frac{1}{2}\nu_{0,1}\overline{Y}^*_{1,0}\overline{X}^*_{0,0}\overline{Z}^*_{0,0} \\ \quad - \frac{1}{2}\nu_{0,1}\overline{Y}^*_{0,1}\overline{Y}^*_{0,0}\overline{Z}^*_{0,0} + (\frac{1}{4}\nu_{0,1}\gamma_{2,0} - \frac{1}{4}\nu_{2,0})\overline{X}^{*\,[2]}_{0,0}\overline{Z}^*_{0,0} + \overline{Z}^*_{1,1}\overline{X}^*_{0,0}\overline{Y}^*_{0,0} \\ \quad + \frac{1}{2}\overline{Z}^*_{0,2}\overline{Y}^{*\,2}_{0,0} + \overline{Z}^*_{1,0}\overline{X}^*_{0,1}\overline{Y}^*_{0,0} + \overline{Z}^*_{1,0}\overline{X}^*_{1,0}\overline{X}^*_{0,0} + (\frac{1}{4}\nu_{0,1}\gamma_{0,2} + \frac{3}{8}\nu_{0,1}^2 - \frac{1}{4}\nu_{0,2})\overline{Y}^{*\,2}_{0,0}\overline{Z}^*_{0,0} \\ \quad + \frac{1}{2}\overline{H}^*_{0,0}\overline{Z}^*_{0,1}\overline{Y}^*_{0,0} + \frac{1}{2}\overline{H}^*_{0,0}\overline{Z}^*_{1,0}\overline{X}^*_{0,0} + \frac{1}{2}\overline{H}^*_{0,1}\overline{Z}^*_{0,0}\overline{Y}^*_{0,0} + \frac{1}{2}\overline{H}^*_{1,0}\overline{X}^*_{0,0}\overline{Z}^*_{0,0} + \frac{1}{2}\overline{Z}^*_{2,0}\overline{X}^{*\,2}_{0,0} \\ \quad + \frac{3}{8}\overline{Z}^*_{0,0}\overline{H}^{*\,2}_{0,0} - \frac{1}{6}\beta_{3,0}\overline{X}^{*\,[3]}_{0,0} - \frac{1}{4}\nu_{0,2}\overline{Y}^{*\,2}_{0,0}\overline{Z}^*_{0,0}), \\ W_4^* = g(\frac{\epsilon_0}{\sigma}) \text{ for some odd function } g, \text{ see below for details.} \end{cases} \tag{3.37}$$

Notice that we did not give the explicit expression for W_4^* in (3.37). There are two reasons for this. First, the explicit expression for W_4^* is so lengthy that it will take at least one whole page to hold it. Secondly, knowing that W_4^* is an odd function is enough for our later derivations.

Now, by putting (3.31), (3.33) and (3.36) together, we have

$$W_n = W_n^* + O_p(n^{-2}) = W_1^* + W_2^* + W_3^* + W_4^* + O_p(n^{-2}) \tag{3.38}$$

where $W_i^{*\prime}s$ are shown as in (3.37).

In order to get an asymptotic expansion of the distribution function of W_n^*, again the first four cumulants have to be calculated approximately first. The method used here to obtain those approximate cumulants is similar to what has been used in location case (Section 3.1), but the calculations become more complicated. However, with the help of some mathematical software packages such as Maple, we have obtained the first four uncentered moments of W_n^* as follows.

(I). By checking the expressions in (3.37) term by term, it is noticed that $W_i^{*\prime}s$ are in fact some odd functions of $\frac{\epsilon_0}{\sigma}$, and by realizing that $\frac{\epsilon_0}{\sigma}$ is assumed to be symmetrically distributed, it follows that $E[W_n^*] = 0$ and therefore

$$E[W_n] = O(n^{-2}). \tag{3.39}$$

(II). By noticing that W_n^{*3} can be written as

$$\begin{aligned} W_n^{*3} &= W_1^{*3} + 3W_1^{*2}W_2^* + 3W_1^*W_2^{*2} + 3W_1^{*2}W_3^* + 3W_1^{*2}W_4^* \\ &\quad +6W_1^*W_2^*W_3^* + W_2^{*3} + O_p(n^{-2}), \end{aligned} \tag{3.40}$$

which is also an odd function of $\frac{\epsilon_0}{\sigma}$, we have

$$E[W_n^3] = O(n^{-2}). \tag{3.41}$$

(III). By noticing that W_n^{*2} can be written as

$$W_n^{*2} = W_1^{*2} + 2W_1^*W_2^* + W_2^{*2} + 2W_1^*W_3^* + O_p(n^{-2}), \tag{3.42}$$

and by taking expectations term by term(with the help of Maple), we have

$$E[W_n^2] = 1 + O(n^{-2}). \tag{3.43}$$

(IV). By noticing that W_n^{*4} can be written as

$$W_n^{*4} = W_1^{*4} + 4W_1^{*3}W_2^* + 4W_1^{*3}W_3^* + 6W_1^{*2}W_2^{*2} + O_p(n^{-2}), \tag{3.44}$$

and by taking expectations term by term(with the help of Maple), we have

$$E[W_n^4] = 3 - 2\frac{s}{n} + O(n^{-2}) \text{ with } s = \frac{E[\psi^4(\frac{y-\mathbf{x}_1^T\theta_0}{\sigma})x_2^4]}{(E[\psi^2(\frac{y-\mathbf{x}_1^T\theta_0}{\sigma})x_2^2])^2}. \tag{3.45}$$

Now if we write $\kappa = s - 3$, then from (3.39)–(3.45), the first four cumulants of W_n are given by

$$\begin{cases} \kappa_1 = O(n^{-2}), \\ \kappa_2 = 1 + O(n^{-2}), \\ \kappa_3 = O(n^{-2}), \\ \kappa_4 = -\frac{2\kappa+6}{n} + O(n^{-2}) \,. \end{cases} \tag{3.46}$$

Comparing (3.46) with (3.6) , we find that we have reached exactly the same result as we had in the location case. Thus proceeding exactly as in Section 3.1 yields the same conclusion as in the location case. Finally, we summarize this result with theorem 3.2.

Theorem 3.2 Assume that **A1)–A6)** hold. Then to the order $O_p(n^{-2})$, the test statistic $Q_n = (1 - \frac{1}{n})\frac{W_n^2}{1-\frac{W_n^2}{n}}$ is approximately $(1 - \frac{\alpha}{n-1})\ F^1_{(n-1)(1-\beta)}$-distributed, with

$$W_n = \frac{\sum_{i=1}^n \psi(\frac{y_i - \mathbf{x}_{i,1}^T\hat{\theta}_{n,1}}{\hat{\sigma}_n})x_{i,2}}{\sqrt{\sum_{i=1}^n \psi^2(\frac{y_i - \mathbf{x}_{i,1}^T\hat{\theta}_{n,1}}{\hat{\sigma}_n})x_{i,2}^2}}, \quad \alpha = -\frac{2}{3}\kappa, \quad \beta = \frac{\kappa}{\kappa - 3}, \quad \kappa = s - 3 \,.$$

and s is defined as in (3.45).

Chapter 4. Simulation Study results

Up to this point, we have developed approximations for the cumulative distribution of the Q_n test statistic for both the location and the simultaneous linear regression/scale case. It turns out that the order of approximating Q_n by F(short for $(1-\frac{\alpha}{n-1})F^{1}_{(n-1)(1-\beta)}$) is higher than that of approximating W_n^2 by χ_1^2. It is expected that the F approximation should be better than the χ_1^2 approximation especially in the tail area of the distribution which is of interest. This chapter is then devoted to the comparison of the accuracy of these two approximations.

§4.1 Location case with scale known

In order to give an indication of the accuracy of the approximations in this case, we consider the following situation.

I) Suppose that ϵ has a contaminated normal distribution

$$\epsilon \sim (1-\delta)N(0,1) + \delta N(0,\tau^2) \tag{4.1}$$

for some small $0 \le \delta \le 1$ and specified τ (τ is fixed to be 3 in our study).

II) ψ is chosen to be the Huber function, i.e.

$$\psi(x) = \begin{cases} x & \text{if} \quad |x| < k \\ k \cdot sign(x) & \text{if} \quad |x| \ge k \end{cases} \tag{4.2}$$

for some specified k. In our study, k is chosen to be 1.345.

Under I), II) and for each combination of $\delta = 0, 0.05, 0.10$ and $n = 5, 10, 20$, we

a) obtain a random sample $\{y_i\}_{i=1}^{n}$ with distribution (4.1) by using Monte Carlo method;

b) calculate the W_n^2 and Q_n statistics from $\{y_i\}_{i=1}^n$;

c) repeat a) and b) $N = 30,000$ times to get the random samples $\{w_{n_j}^2\}_{j=1}^N$ and $\{q_{n_j}\}_{j=1}^N$;

d) calculate the sample means, variances and selected percentiles from $\{w_{n_j}^2\}_{j=1}^N$ and $\{q_{n_j}\}_{j=1}^N$ respectively;

e) calculate the corresponding quantities from their approximations.

The calculated means, variances and their F and χ_1^2 approximations are as given in Exhibit 4.1, while the percentiles (tail area) and their F and χ_1^2 approximations are as given in Exhibit 4.2.

n		5			10			20		
δ		0	.05	.10	0	.05	.10	0	.05	.10
mean	MC	1.000	1.003	1.004	0.991	1.001	1.005	0.994	1.000	1.005
	F		1.000			1.000			1.000	
	χ_1^2		1.000			1.000			1.000	
var	MC	1.239	1.251	1.242	1.591	1.604	1.607	1.784	1.823	1.820
	F	1.239	1.234	1.219	1.619	1.617	1.609	1.810	1.808	1.805
	χ_1^2		2.000			2.000			2.000	

Exhibit 4.1

Means and variances for W_n^2 from Monte Carlo method (using contaminated normal error distribution and Huber function with k=1.345) and their corresponding F and χ_1^2 approximations.

Notice that in both Exhibits, MC represents the quantities calculated from the Monte Carlo results for W_n^2, χ_1^2 represents the corresponding values by using χ_1^2 to approximate W_n^2, and F represents the corresponding values by using F to approximate Q_n, which is, however, re-represented in terms of W_n^2 so that the comparison between F and χ_1^2 approximations can be performed. The blank cells appeared in the Exhibits mean the values in these cells are the same as in the nearest cell in the same row. Also notice that, although α and β are related to δ in the F approximation, we will always choose $\delta = 0.10$ to get them, since for any real data set we never have a chance to know exactly how many observations have been contaminated and

Tail Area	n / δ	5: 0	5: .05	5: .10	10: 0	10: .05	10: .10	20: 0	20: .05	20: .10
0.2500	MC	1.537	1.538	1.547	1.387	1.396	1.427	1.356	1.357	1.362
	F	1.474	1.475	1.479	1.397	1.398	1.399	1.359	1.360	1.360
	χ_1^2		1.323			1.323			1.323	
0.1000	MC	2.749	2.790	2.765	2.699	2.751	2.742	2.679	2.740	2.760
	F	2.679	2.679	2.679	2.719	2.719	2.719	2.716	2.716	2.716
	χ_1^2		2.706			2.706			2.706	
0.0500	MC	3.440	3.454	3.448	3.705	3.723	3.754	3.750	3.806	3.810
	F	3.401	3.400	3.396	3.693	3.692	3.690	3.780	3.780	3.779
	χ_1^2		3.842			3.842			3.842	
0.0250	MC	3.940	3.926	3.929	4.619	4.634	4.644	4.801	4.867	4.844
	F	3.935	3.933	3.928	4.598	4.597	4.591	4.839	4.838	4.835
	χ_1^2		5.024			5.024			5.024	
0.0100	MC	4.375	4.349	4.351	5.748	5.687	5.691	6.134	6.288	6.175
	F	4.399	4.397	4.391	5.661	5.658	5.649	6.200	6.198	6.192
	χ_1^2		6.635			6.635			6.635	
0.0050	MC	4.561	4.557	4.558	6.418	6.417	6.336	7.179	7.279	7.202
	F	4.616	4.615	4.609	6.355	6.352	6.340	7.188	7.185	7.175
	χ_1^2		7.879			7.879			7.879	
0.0010	MC	4.835	4.825	4.826	7.659	7.608	7.508	9.576	9.296	9.437
	F	4.889	4.868	4.864	7.623	7.618	7.603	9.306	9.301	9.285
	χ_1^2		10.83			10.83			10.83	
0.0001	MC	4.947	4.949	4.972	8.646	8.494	8.393	12.13	11.83	11.71
	F	4.972	4.972	4.971	8.757	8.752	8.737	11.88	11.87	11.84
	χ_1^2		15.14			15.14			15.14	

Exhibit 4.2

Percentiles for W_n^2 from Monte Carlo method (using contaminated normal error distribution and Huber function with k=1.345) and their corresponding F and χ_1^2 approximations.

$\delta = 0.10$ is a typical value for the proportion of contamination. This kind of approximation could reduce the accuracy of our F approximation. However, as we will see, even if α, β are treated in this way, using F to approximate Q_n is still better than using χ_1^2 to approximate W_n^2 in almost all the situations, especially in the tail area which is often of interest.

A glance at the Exhibit 4.1 shows that while both F and χ_1^2 approximate the means obtained from the Monte Carlo simulations very well, F gives a much better approximation than χ_1^2 does regarding the variance. In fact, the relative errors $(|\frac{\text{approximate-simulated}}{\text{simulated}}| \times 100\%)$ from F approximation are never greater than 1.8% for all the cases. As a comparison, for $n = 5$ and with no contaminated observations, the relative error from the χ_1^2 approximation is as large as 61.3%. However, it seems that better accuracy does not always come with larger sample size although we know that the apppoximation is of order $O(n^{-2})$.

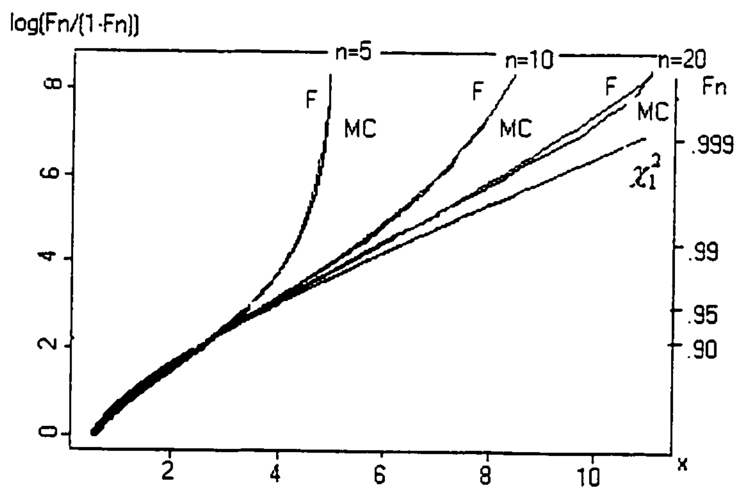

Exhibit 4.3a

$\log(F_n/(1 - F_n))$ *versus* x *for normal error distribution and Huber function (k=1.345).*

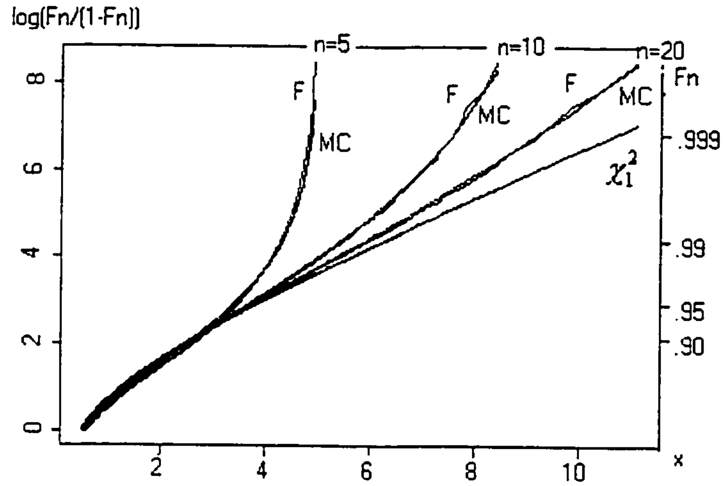

Exhibit 4.3b

$\log(F_n/(1-F_n))$ *versus* x *for contaminated normal error distribution* ($\delta = 0.05$) *and Huber function(k=1.345).*

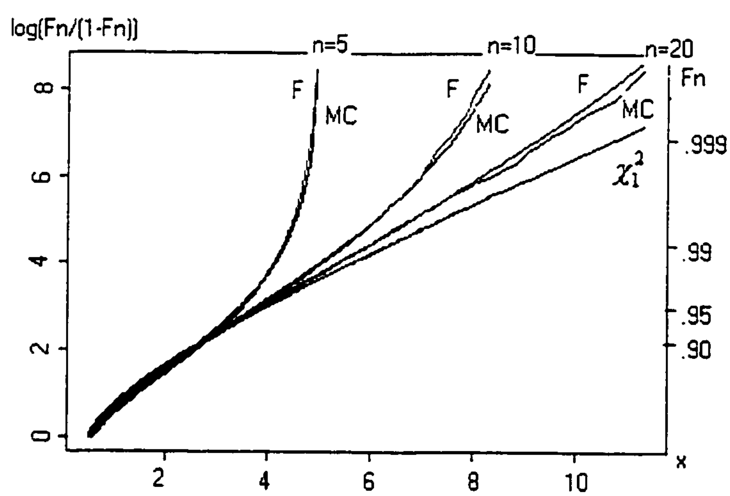

Exhibit 4.3c

$\log(F_n/(1-F_n))$ *versus* x *for contaminated normal error distribution* ($\delta = 0.10$) *and Huber function (k=1.345).*

When we look into Exhibit 4.2, we can find that the F approximation also gives us more accurate percentiles. The relative errors remain well under control even well out into the very tails for n as small as 5, while the χ_1^2 approximation gives us relatively poor results. For example, for the samples with 5 observations, the relative errors from the F approximation when tail area= 0.0001 are about 0.6%, and this number from the χ_1^2 approximation is 206%. A graphical display of Exhibit 4.2 is given in Exhibit 4.3, where the value of $\log(F_n/(1-F_n))$ is plotted against percentile x. Here F_n represents the cumulative probability. These graphs again strongly suggest the advantage of using the F approximation instead of using the χ_1^2 approximation, especially when sample size is small.

§4.2 Simple linear regression case with scale unknown

We now turn to the simple linear regression case(with scale unknown) where the robust M−type estimates of the regression/scale parameters have to be obtained first to get the W_n^2 and Q_n statistics. In this case, suppose that the simple linear regression model is given by

$$y_i = \theta_0 + \theta_1 x_i + \epsilon_i, \quad i = 1, ..n \tag{4.3}$$

where x_i has a $N(0,1)$ distribution and is independent of ϵ_i, which again has the contaminated normal distribution (4.1). The null hypothesis of interest is

$$H_0 : \theta_1 = 0. \tag{4.4}$$

Now, given (4.2), we choose χ to be $\chi(\frac{\epsilon}{\sigma}) = \frac{1}{2}(\psi^2(\frac{\epsilon}{\sigma}) - C)$ with $C = E_{(\theta_0,\sigma)}[\psi^2(\frac{\epsilon}{\sigma})]$. This corresponds to "Proposal 2" of Huber (1964) and gives translation and scale equivariant estimates.

Similarly, for each combination of $\delta = 0, 0.05, 0.10$ and $n = 10, 20, 40$, we calculate the same quantities as in Section 4.1 except that we apply (3.20) to obtain the W_n statistic. The following iterative algorithm gives an outline of how to calculate the regression/scale estimates $\widehat{\theta}_0, \hat{\sigma}$ from a random sample $\{x_i, y_i\}_{i=1}^n$(Huber, 1981).

a) Choose

$$\begin{cases} \theta^{(0)} = \frac{1}{n}\sum_{i=1}^n \psi(y_i) , \\ \sigma^{(0)} = \sqrt{\frac{1}{n-1}\sum_{i=1}^n \psi^2(y_i)} \end{cases} \tag{4.5}$$

as the initial values of $\widehat{\theta}, \widehat{\sigma}$.

b) Given $\sigma^{(m)}, \theta^{(m)}, m \geq 0$, put $r_i = y_i - \theta^{(m)}$ then obtain $\sigma^{(m-1)}$ by using the formula

$$\sigma^{(m+1)} = \sqrt{\frac{1}{(n-1)C}\sum_{i=1}^n \psi^2(\frac{r_i}{\sigma^{(m)}})(\sigma^{(m)})^2} \text{ with } C = E_{(\theta_0,\sigma)}[\psi^2(\frac{\epsilon}{\sigma})].$$

c) Put

$$\begin{cases} r_i = y_i - \theta^{(m)} , \\ r_i^* = \psi(\frac{r_i}{\sigma^{(m+1)}})\sigma^{(m+1)} , \end{cases} \tag{4.6}$$

n		10			20			40		
δ		0	.05	.10	0	.05	.10	0	.05	.10
mean	*MC*	0.997	1.001	0.998	1.006	0.996	0.990	1.003	1.008	1.006
	F		1.000			1.000			1.000	
	χ_1^2		1.000			1.000			1.000	
var	*MC*	1.252	1.245	1.248	1.576	1.537	1.493	1.761	1.823	1.795
	F	0.858	0.851	0.828	1.429	1.425	1.414	1.717	1.713	1.707
	χ_1^2		2.000			2.000			2.000	

Exhibit 4.4

Means and variances for W_n^2 from Monte Carlo method (using contaminated normal error distribution and Huber function with k=1.345) and their corresponding F and χ_1^2 approximations.

Tail Area	n / δ	10: 0	10: .05	10: .10	20: 0	20: .05	20: .10	40: 0	40: .05	40: .10
0.2500	MC	1.503	1.513	1.523	1.436	1.411	1.414	1.379	1.355	1.375
	F	1.517	1.517	1.520	1.425	1.426	1.427	1.375	1.376	1.377
	χ_1^2		1.323			1.323			1.323	
0.1000	MC	2.623	2.613	2.618	2.720	2.682	2.688	2.730	2.772	2.735
	F	2.691	2.691	2.689	2.720	2.720	2.720	2.719	2.719	2.719
	χ_1^2		2.706			2.706			2.706	
0.0500	MC	3.339	3.349	3.356	3.676	3.643	3.548	3.761	3.939	3.829
	F	3.447	3.445	3.437	3.662	3.661	3.657	3.758	3.757	3.755
	χ_1^2		3.842			3.842			3.842	
0.0250	MC	3.998	4.001	3.983	4.537	4.539	4.408	4.756	4.877	4.875
	F	4.091	4.087	4.073	4.545	4.542	4.533	4.782	4.781	4.776
	χ_1^2		5.024			5.024			5.024	
0.0100	MC	4.695	4.745	4.712	5.638	5.469	5.412	6.050	6.157	6.181
	F	4.799	4.792	4.768	5.611	5.606	5.589	6.093	6.090	6.080
	χ_1^2		6.635			6.635			6.635	
0.0050	MC	5.203	5.185	5.149	6.319	6.310	6.120	7.043	7.095	7.172
	F	5.245	5.236	5.206	6.344	6.337	6.314	7.044	7.040	7.025
	χ_1^2		7.879			7.879			7.879	
0.0010	MC	6.333	6.030	6.038	7.748	7.760	7.902	9.364	9.268	9.138
	F	6.064	6.050	6.006	7.829	7.817	7.777	9.107	9.098	9.071
	χ_1^2		10.83			10.83			10.83	
0.0001	MC	7.331	6.808	7.139	9.533	9.280	9.137	10.96	9.991	10.88
	F	6.880	6.862	6.801	9.525	9.505	9.441	11.72	11.70	11.65
	χ_1^2		15.14			15.14			15.14	

Exhibit 4.5

Percentiles for W_n^2 from Monte Carlo method (using contaminated normal error distribution and Huber function with k=1.345) and their corresponding F and χ_1^2 approximations.

then obtain $\theta^{(m+1)}$ by using the formula

$$\theta^{(m+1)} = \theta^{(m)} + l\frac{1}{n}\sum_{i=1}^{n} r_i^* ,$$

where $0 < l < 2$ is an arbitrary relaxation factor.

d) Repeat b) and c) until $\theta^{(m+1)}$ and $\sigma^{(m+1)}$ converge.

The numerical results are as given in Exhibit 4.4. 4.5 and 4.6. These exhibits are similar to what we have gotten in Section 4.1. Although the results are not as perfect as those in the location case, we are still convinced that the F approximation is better than the χ_1^2 approximation.

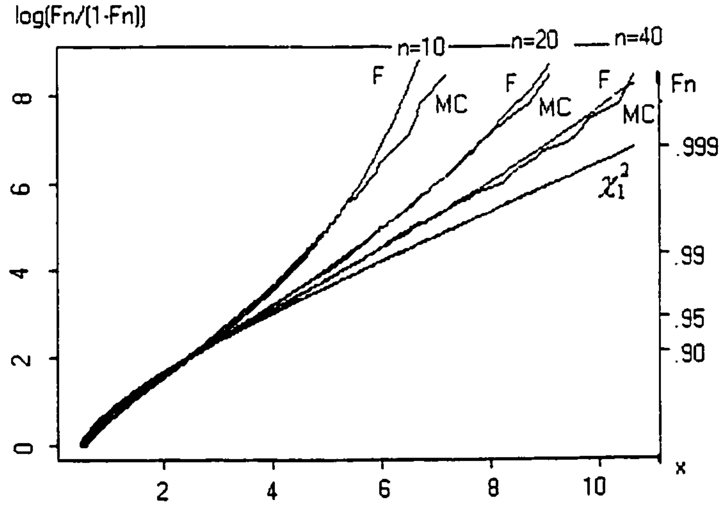

Exhibit 4.6a

$\log(F_n/(1-F_n))$ *versus x for normal error distribution and Huber function (k=1.345).*

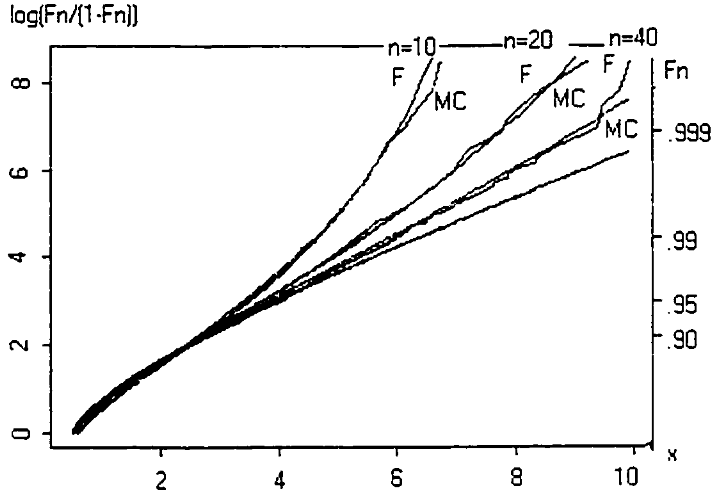

Exhibit 4.6b

$\log(F_n/(1-F_n))$ versus x for contaminated normal error distribution ($\delta = 0.05$) and Huber function (k=1.345).

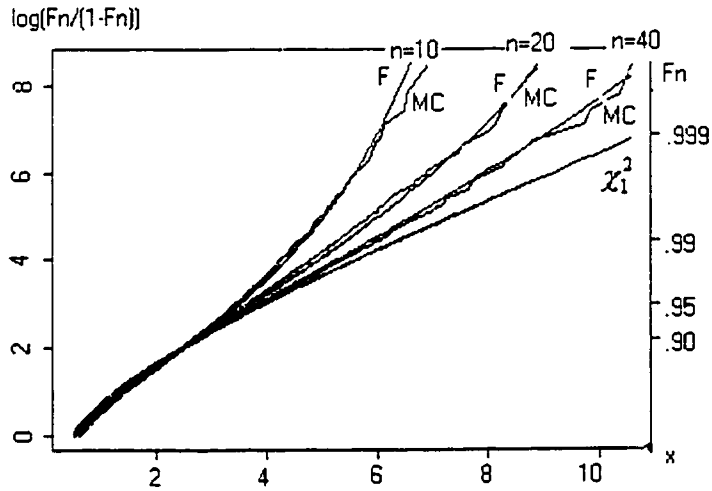

Exhibit 4.6c

$\log(F_n/(1-F_n))$ versus x for contaminated normal error distribution ($\delta = 0.10$) and Huber function (k=1.345).

To give a feeling of the difference between F approximation and χ^2 approximation in tail area., Exhibit 4.7 shows some p-values for normal, contaminated normal($\delta = 0.10, \tau = 10$) and Cauchy distributions when using two kind of approximations with sample size $n = 16$. It is noticed that our F approximation usually gives a smaller p-value.

N(0,1)		Contam		Cauchy	
χ^2	F	χ^2	F	χ^2	F
1.60	0.65	1.60	0.79	1.60	0.62
3.60	2.57	3.60	2.53	3.60	2.76
5.30	4.49	5.30	4.46	5.30	4.65
7.70	7.33	7.70	7.31	7.70	7.42

Exhibit 4.7

Selected p-values for normal, contaminated normal ($\delta = 0.10, \tau = 10$) and Cauchy distributions from F approximation and χ^2 approximation

References

Barndorff-Nielsen, O. E. and Cox, D. R.(1989). Asymptotic Techniques for Use in Statistics. Chapman and Hull, London.

Bhattacharya, R. N. and Ghosh, J. K.(1978). On the Validity of the Formal Edgeworth Expansion. Annals of Statistics, 6, 434–451.

Field, C. A. and Ronchetti, E.(1990). Small Sample Asymptotics. IMS Lecture Notes–Monograph Series.

Field, C. A.(1982). Small Sample Asymptotic Expansions for Multivariate M–estimators. Annals of Statistics, 10, 672–689.

Graham, A.(1981). Kronecker Products and Matrix Calculus with Application. Halsted(Wiley), New York.

Hall, P.(1992). The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York.

Hamilton, D. and Wiens, D. P.(1987). Correction Factors for F Ratios in Nonlinear Regression. Biometrika, 74, 423–425.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A(1986). Robust Statistics: The Approach Based on Influence Functions. Wiley, New York.

Henderson, H. V. and Searle, S. R.(1979). Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics. Canad. J. Statist., 7, 65–81.

Huber, P. J.(1964). Robust Estimation of a Location Parameter. Annals of Mathematical Statistics, 35, 73–101.

Huber, P. J.(1981). Robust Statistics. Wiley, New York.

Jensen, J. L.(1995). Saddlepoint Approximations. Clarendon Press, Oxford.

Johansen, S.(1983). Some Topics in Regression. Scandinavian J. of Statist., 10, 161–194.

Markatou, M., Stahel, W. A. and Ronchetti, E. M.(1991). Robust M-type Testing Procedures for Linear Models. In: Directions in Robust Statistics and Diagnostics I.(Publisher and location). pp. 201–220.

Wiens, D. P.(1985). On Some Pattern-Reduction Matrices Which Appear in Statistics. Linear Algebra and its Application, 67, 233–258.

Wiens, D. P.(1998). Robust Testing, manuscript.

Wiens, D. P.(1996). Asymptotics of Generalized M–Estimation of Regression and Scale with Fixed Carriers in an Approximately Linear Model. Statist. Probab. Lett., 30, 271–285.