ESTIMATION AND HYPOTHESIS TESTING IN STOCHASTIC REGRESSION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

HAKAN SAVAŞ SAZAK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

THE DEPARTMENT OF STATISTICS

DECEMBER 2003

Approval of the Graduate School of Natural and Applied Sciences.

Prof. Dr. Canan Özgen
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. H. Öztaş Ayhan
Head of Department

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Asst. Prof. Dr. Qamarul İslam
Co-Supervisor

Prof. Dr. Moti Lal Tiku
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Ayşen Akkaya

Prof. Dr. Moti Lal Tiku

Prof. Dr. H. Öztaş Ayhan

Prof. Dr. Okay Çelebi

Asst. Prof. Dr. Qamarul İslam

ABSTRACT

ESTIMATION AND HYPOTHESIS TESTING IN STOCHASTIC REGRESSION

SAZAK, Hakan Savaş
Ph. D., Department of Statistics
Supervisor : Prof. Dr. Moti Lal TIKU
Co-Supervisor : Asst. Prof. Dr. Qamarul İSLAM

December 2003, 201 pages

Regression analysis is very popular among researchers in various fields but almost all the researchers use the classical methods which assume that X is nonstochastic and the error is normally distributed. However, in real life problems, X is generally stochastic and error can be nonnormal. Maximum likelihood (ML) estimation technique which is known to have optimal features, is very problematic in situations when the distribution of X (marginal part) or error (conditional part) is nonnormal.

Modified maximum likelihood (MML) technique which is asymptotically giving the estimators equivalent to the ML estimators, gives us the opportunity to conduct the estimation and the hypothesis testing procedures under nonnormal marginal and conditional distributions. In this study we show that MML estimators are highly efficient and robust. Moreover, the test statistics based on the MML estimators are much more powerful and robust compared to the test

statistics based on least squares (LS) estimators which are mostly used in literature. Theoretically, MML estimators are asymptotically minimum variance bound (MVB) estimators but simulation results show that they are highly efficient even for small sample sizes. In this thesis, Weibull and Generalized Logistic distributions are used for illustration and the results given are based on these distributions.

As a future study, MML technique can be utilized for other types of distributions and the procedures based on bivariate data can be extended to multivariate data.

Key Words: Maximum likelihood (ML), modified maximum likelihood (MML), least squares (LS), stochastic regression, nonnormal error distribution, nonnormal marginal distribution, minimum variance bound (MVB), robustness, Weibull, Generalized Logistic.

ÖZ

STOKASTİK REGRESYONDA TAHMİN VE HİPOTEZ TESTİ

SAZAK, Hakan Savaş
Doktora, İstatistik Bölümü
Tez Yöneticisi : Prof. Dr. Moti Lal TIKU
Ortak Tez Yöneticisi : Yrd. Doç. Dr. Qamarul İSLAM

Aralık 2003, 201 sayfa

Regresyon analizi çeşitli alanlardaki araştırmacılar arasında çok popülerdir fakat hemen hemen tüm araştırmacılar, X 'in stokastik olmadığını ve hata dağılımının normal olduğunu varsayan klasik metotlar kullanırlar. Bununla beraber gerçek hayat problemlerinde X , genellikle stokastik değildir ve hatanın dağılımı da normal olmayabilir. En uygun yöntem olarak bilinen en çok olabilirlik (ML) tahmin etme yöntemi, X 'in (marjinal kısım) veya hatanın (koşullu kısım) dağılımının normal olmadığı durumlarda çok problemli bir yöntemdir.

Uyarlanmış en çok olabilirlik (MML) tahmin etme tekniği ki asimptotik olarak ürettiği tahmin ediciler ML tahmin etme yönteminin tahmin edicilerine eşittir, bize normal olmayan marjinal ve koşullu durumlar altında tahmin etme ve hipotez testi yapma imkanı verir. Bu çalışmada MML tahmin edicilerinin sağlam ve yüksek etkinliğe sahip olduklarını gösterdik. Dahası, MML tahmin edicilerine dayalı test istatistikleri de literatürde çokça kullanılan en küçük kareler (LS)

yöntemi ile bulunan tahmin edicilere dayalı test istatistiklerinden çok daha güçlü ve sağlamdır. Teorik olarak MML tahmin edicileri asimptotik olarak en küçük varyans sınırını bulan (MVB) tahmin edicilerine eşittirler fakat simulasyon sonuçları gösteriyor ki küçük örneklem hacimlerinde bile çok etkindirler. Bu tezde normal olmayan dağılımlara örnek olarak Weibull ve Genelleştirilmiş Lojistik dağılımı kullanılmış ve sonuçlar bu dağılımlara göre verilmiştir.

Gelecekteki çalışmalarda MML tekniği daha değişik dağılımlar için kullanılabilir ve bu tezde ikili veri seti için verilen sonuçlar çok değişkenli veri setleri için genellenebilir.

Anahtar Kelimeler: En çok olabilirlik (ML), uyarlanmış en çok olabilirlik (MML), en küçük kareler (LS), stokastik regresyon, normal olmayan hata dağılımı, normal olmayan marjinal dağılım, en küçük varyans sınırı (MVB), sağlamlık, Weibull dağılımı, Genelleştirilmiş Lojistik dağılımı.

To My Father and My Mother

ACKNOWLEDGMENTS

You showed me how a teacher should be,
You showed me how a scientist should be,
You showed me how a human should be.

You never got tired of my mostly irrelevant questions. I am surprised how you adapt immediately to the questions from different fields and answer them in such a precise way. I know that I will never be even thousandth of you, but this does not make upset, on the contrary, I am proud of being your student. It was a big opportunity for me to work such a statistician like you and it was a big pleasure to know such a person like you. You really gave me a vision that I will always use in my academic life. I hope I will have a chance to work with you again. Thanks, Professor Tiku.

I would like to express my deepest gratitude to Asst. Prof. Dr. Qamarul İslam for being my Co-Supervisor and supporting me during my work on the thesis.

I want to thank the Head of the Department, Prof. Dr. Öztaş Ayhan, on behalf of the Department of Statistics because of providing me all the technological equipments and a nice room to work in during my research assistantship as well as his contribution to my thesis.

I would like to thank Assoc. Prof. Dr. Ayşen Akkaya because she always gave support and helped me whenever I needed.

I should also thank Prof. Dr. Okay Çelebi due to his contribution and invaluable critisizm on my thesis.

I would like to thank Asst. Prof. Dr. Birdal Şenoğlu because of the statistical discussions we have done and being a really trusty friend.

I should thank Dr. Barış Sürücü who tidy me up all the time and give suggestions which are invaluable. I always admire his working style which I am afraid I will never have. You do the goodness of fit, I will do the regression.

I should also thank Mahir Sinan Özdemir because of his good friendship and nice conversations.

I want to thank Research Assistants Oya Can, Didem Avcıoğlu, Burçin Akgün and Pelin Özkan for their support throughout my study.

I want to thank my father, Mehmet Emin Sazak; my mother, Fatma Sazak; my sisters, Sevinç Sazak, İldem Sazak and Aydem Sazak Bezcioğlu; and my nephews, Kerem Bora İrten, Hakan Oğuzcan Erkam and Yılmaz Batuhan Bezcioğlu because they always encouraged me in my academic life and cheered me up.

Finally, I want to thank my friends Mustafa Tolga Gündoğmuş, Bülent Şener, Tolga Yangın and Ufuk Özok because of being my friends since High School and continuosly created opportunity to see me to have some conversation.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION
and
HISTORICAL PERSPECTIVE

Summary: In a simple linear model, the design variable X has traditionally been taken to be nonstochastic and the random error e assumed to be normally distributed. It has been recognized, however, that in numerous applications X might be stochastic and e might not be normal. This gives rise to three research areas: (a) X is nonstochastic and e is nonnormal, (b) X is stochastic and e is normal, and (c) X is stochastic and e is nonnormal. In this chapter, we briefly review the work done so far in the areas (a) and (b). There is no previous work in the area (c).

1.1 Linear Model Under Normality

Consider a simple linear regression model

$$y_i = \theta_0 + \theta_1 x_i + e_i, 1 \le i \le n, \tag{1.1.1}$$

where x_i's are nonstochastic design values (assumed to be measurable without error) and e_i's are identically and independently distributed (iid) random errors. The parameter θ_0 is called the intercept and θ_1 is called the slope or the regression coefficient. The parameter θ_1 is of particular importance since $\frac{dy}{dx} = \theta_1$

and represents the rate of change in y as x changes by one unit. Usually, x_i's $(1 \le i \le n)$ are chosen to be equidistant and such an arrangement is called symmetric design. In fact, a symmetric design has certain theoretical and computational advantages which will be pointed out later.

A problem of major importance is to estimate the parameters θ_0 and θ_1 and test the null hypothesis $H_0 : \theta_1 = 0$. Clearly, H_0 implies that the design variable (also called input) has no effect on the response y. Assume in the first place, as the tradition has it, that e_i's $(1 \le i \le n)$ are iid normal $N(0, \sigma^2)$. Since the method of maximum likelihood has all the desirable optimal properties under some general regularity conditions, we employ this method to estimate θ_0, θ_1 and σ. The likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)^2\right\}. \tag{1.1.2}$$

The maximum likelihood (ML) estimators are the solutions of the equations

$$\frac{\partial \ln L}{\partial \theta_0} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i) = 0 \tag{1.1.3}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)x_i = 0 \text{ and} \tag{1.1.4}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)^2 = 0. \tag{1.1.5}$$

The solutions are the ML estimators:

$$\hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x}\,,\ \hat{\theta}_1 = \sum_{i=1}^{n}(x_i - \bar{x})y_i \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2\,, \text{ and} \tag{1.1.6}$$

$$\hat{\sigma} = \left[\sum_{i=1}^{n} \left\{ (y_i - \bar{y}) - \hat{\theta}_1 (x_i - \bar{x}) \right\}^2 / (n-2) \right]^{1/2}. \tag{1.1.7}$$

The divisor n-2 in (1.1.7) makes $\hat{\sigma}^2$ unbiased, i.e., $E(\hat{\sigma}^2) = \sigma^2$.

Lemma 1.1: The estimator $\hat{\theta}_1$ is the minimum variance bound (MVB) estimator of θ_1 with variance $\sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2$ and is normally distributed. This follows from a re-organization of $\partial \ln L / \partial \theta_1$. In view of (1.1.3), we have (Kendall and Stuart, 1979, p.24)

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sigma^2} (\hat{\theta}_1 - \theta_1) \tag{1.1.8}$$

which implies that $\hat{\theta}_1$ is the MVB estimator of θ_1. The normality follows from the fact that $\hat{\theta}_1$ is a linear function of iid normal variates $y_1, y_2, ..., y_n$.

Fisher Information: The Fisher information matrix $I(\theta_0, \theta_1, \sigma)$, consists of the elements $-E(\partial^2 \ln L / \partial \theta_0^2)$, $-E(\partial^2 \ln L / \partial \theta_1^2)$, $-E(\partial^2 \ln L / \partial \theta_0 \partial \theta_1)$, etc., and is given by

$$I(\theta_0, \theta_1, \sigma) = \frac{n}{\sigma^2} \begin{bmatrix} 1 & \bar{x} & 0 \\ \bar{x} & (1/n) \sum_{i=1}^{n} x_i^2 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \tag{1.1.9}$$

The asymptotic variance-covariance matrix of the estimators $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$ is given by I^{-1}. In particular,

$$V(\hat{\theta}_1) = \sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2 \text{ and } V(\hat{\sigma}) \cong \sigma^2 / 2n . \tag{1.1.10}$$

It may be noted, however, that the expression for the variance $V(\hat{\theta}_1)$ given in (1.1.10) is exact for all sample sizes n (Lemma 1.1). We also have the Cochran identity

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 \equiv \hat{\theta}_1 Q + \sum_{i=1}^{n} \left\{ (y_i - \bar{y}) - \hat{\theta}_1 (x_i - \bar{x}) \right\}^2 , \; Q = \sum_{i=1}^{n} (x_i - \bar{x}) y_i ,$$

or
$$(n-1)s^2 \equiv \hat{\theta}_1 Q + (n-2)s_e^2 \tag{1.1.11}$$

with expected values

$$(n-1)\sigma^2 \equiv \sigma^2 + (n-2)\sigma^2 \text{ (if } \theta_1 = 0); \tag{1.1.12}$$

(1.1.12) is, in fact, true even if $\theta_1 \neq 0$ but then the expected values of the SS on the left and the first SS on the right also have a noncentrality parameter. Under the normality assumption, the distribution of $(n-2)s_e^2 / \sigma^2$ is chi-square with $(n-2)$ degrees of freedom, and $\hat{\theta}_1$ and s_e^2 are independently distributed. Note that

$$\hat{\theta}_1 Q = \hat{\theta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 . \tag{1.1.13}$$

Remark: It is well known that if x_i's are equidistant (without loss of generality between –1 and 1) then $\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2$ is maximized, that is, $V(\hat{\theta}_1)$ is minimized. Such an arrangement is called an optimal design.

Hypothesis Testing: As said earlier, testing the null hypothesis $H_0 : \theta_1 = 0$ is of paramount importance. To test H_0 against $H_1 : \theta_1 > 0$, we define the statistic

$$t = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} (\hat{\theta}_1 / s_e). \tag{1.1.14}$$

Large values of t lead to the rejection of H_0. This test is , in fact, uniformly most powerful (UMP). The null distribution of t is Student's t with $n-2$ degrees of freedom. The nonnull distribution of t is noncentral t with $n-2$ degrees of freedom and noncentrality parameter Δ,

$$\Delta^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 (\theta_1 / \sigma)^2. \tag{1.1.15}$$

To test H_0 against $H_1 : \theta_1 \neq 0$, we use the statistic

$$F = \sum_{i=1}^{n} (x_i - \bar{x})^2 (\hat{\theta}_1 / s_e)^2 = \hat{\theta}_1 Q / s_e^2. \tag{1.1.16}$$

The null distribution of F is central-F with $(1, n-2)$ degrees of freedom. The nonnull distribution of F is noncentral-F with $(1, n-2)$ degrees of freedom and noncentrality parameter Δ^2. For details about noncentral distributions, one may refer to Tiku (1985).

Incidentally, the following properties of the likelihood equations (1.1.3)-(1.1.5) may be noted:

(i) $E(\partial^r \ln L / \partial \theta_0^r) = 0$ and $E(\partial^r \ln L / \partial \theta_1^r) = 0$ for $r \geq 3$,

(ii) $E(\partial^{r+s} \ln L / \partial \theta_0^r \partial \sigma^s) = 0$ and $E(\partial^{r+s} \ln L / \partial \theta_1^r \partial \sigma^s) = 0$

for all $r \geq 1$ and $s \geq 1$

and (iii) $E(\partial^2 \ln L/\partial\sigma^2) = -2n/\sigma^2$, $E(\partial^3 \ln L/\partial\sigma^3) = 10n/\sigma^3$, etc; (1.1.17)

(i)-(iii) are called Bartlett (1953) conditions and together with the Cochran identity (1.1.11), imply that $\hat{\theta}_0$ and $\hat{\theta}_1$ are independently distributed of s_e^2, and $\hat{\theta}_1$ is normal and s_e^2 is a multiple of chi-square. We will use such structural relationships from time to time in determining the distributions of statistics like those in (1.1.14) and (1.1.16). We now briefly review the research area (a) mentioned above.

1.2 Nonnormal Error Distribution

Consider the situation where X in the linear model (1.1.1) continues to be nonstochastic but the random error e is nonnormal. We consider location-scale distributions of the type $(1/\sigma) f((y-\mu)/\sigma)$, i.e, the distribution of $z = (y-\mu)/\sigma$ is free of μ and σ. Under nonnormality, using the maximum likelihood methodology is in general problematic, e.g, the likelihood equations have no explicit solutions and solving them by iteration can be problematic for reasons of (i) multiple roots, (ii) nonconvergence of iterations, or (iii) convergence to wrong values; see, for example, Barnett (1966), Lee et al. (1980) and Vaughan (1992). In fact, Puthenpura and Sinha (1986) showed that if the sample contains outliers, iterations with likelihood equations might not converge at all. To alleviate these difficulties , we use the method of modified likelihood due to Tiku (1967; 1968 ; 1980) and Tiku and Suresh (1992). This method linearizes the intractable terms in the likelihood equations and its features are discussed in the literature, e.g., Vaughan and Tiku (2000). Suffice it to say here that this method yields estimators which are explicit functions of sample observations and are, therefore, easy to compute. The estimators are called modified maximum likelihood (MML) estimators and are asymptotically fully efficient , i.e., they are asymptotically minimum variance bound estimators. For

small samples, they have no or negligible bias and are highly efficient; see, for example, Senoglu and Tiku (2001, 2002) and Vaughan (2002).

Islam et al. (2001) and Tiku et al. (2001) develop modified likelihood methodology for treating statistically the situations where X is nonstochastic and e is nonnormal. They consider, for illustration, the following families of distributions: Weibull, Generalized Logistic, Student's t, and short-tailed distributions recently introduced by Tiku and Vaughan (1999), namely,

$$f(e) \propto \frac{1}{\sigma}\left\{1+\frac{\lambda}{2r}\frac{e^2}{\sigma^2}\right\}^{r} \frac{\exp(-e^2/2\sigma^2)}{\sqrt{2\pi}}\ ,\ -\infty < e < \infty\ \ (\lambda > 0)\,.$$

For illustration, we present their results for Student's t family which represents long-tailed symmetric distributions with kurtosis $\mu_4/\mu_2^2 \geq 3$. The distributions in this family are also used to model samples which contain outliers (Tiku et al., 2001).

1.3 Student t Family

Suppose that e has the distribution (p known)

$$f(e) \propto \frac{1}{\sigma}\left\{1+\frac{e^2}{k\sigma^2}\right\}^{-p}\ ,\ -\infty < e < \infty\,; \qquad (1.3.1)$$

$k = 2p-3$ and $p \geq 2$. It may be noted that $E(e) = 0$ and $V(e) = \sigma^2$, and the distribution of $t = \sqrt{(\nu/k)}(e/\sigma)$ is a Student's t distribution with $\nu = 2p-1$ degrees of freedom. Given a random sample $y_1, y_2, ..., y_n$ from (1.3.1), the likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^n \prod_{i=1}^{n} \left\{1+\frac{e_i^2}{k\sigma^2}\right\}^{-p}. \tag{1.3.2}$$

Writing

$$z_i = e_i/\sigma = (y_i - \theta_0 - \theta_1 x_i)/\sigma \text{ and } g(z) = z/\left(1+(1/k)z^2\right), \tag{1.3.3}$$

the likelihood equations are

$$\frac{\partial \ln L}{\partial \theta_0} = \frac{2p}{k\sigma}\sum_{i=1}^{n} g(z_i) = 0 \tag{1.3.4}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{2p}{k\sigma}\sum_{i=1}^{n} x_i g(z_i) = 0 \text{ and} \tag{1.3.5}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{n} z_i g(z_i) = 0. \tag{1.3.6}$$

As explained in Tiku and Suresh (1992), and Vaughan (1992), solving such equations is problematic. In particular, they have multiple roots.

1.4 Modified Likelihood

To formulate the modified likelihood equations, we write (for a given θ_1)

$$w_{(i)} = y_{[i]} - \theta_1 x_{[i]} \text{ and } z_{(i)} = (w_{[i]} - \theta_0)/\sigma,\ 1 \le i \le n; \tag{1.4.1}$$

$(y_{[i]}, x_{[i]})$ may be called concomitants of $z_{(i)}$ and is that pair (y_j, x_j) which determines $z_{(i)}$. Since complete sums are invariant to ordering, the likelihood equations (1.3.4)-(1.3.6) can be written as

$$\frac{\partial \ln L}{\partial \theta_0} = \frac{2p}{k\sigma}\sum_{i=1}^{n} g(z_{(i)}) = 0 \tag{1.4.2}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{2p}{k\sigma}\sum_{i=1}^{n} x_{[i]} g(z_{(i)}) = 0 \text{ and} \tag{1.4.3}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{n} z_{(i)} g(z_{(i)}) = 0\,. \tag{1.4.4}$$

Note that θ_0 and $\sigma\ (>0)$ have no role to play in determining the ordered variates $z_{(i)}$.

Linearization: Since the function $g(z)$ is linear (almost) in a small interval $a < z < b$, and $z_{(i)}$ is located in the vicinity of $t_{(i)} = E\left(z_{(i)}\right),\ (1 \le i \le n)$, we have the first two terms of a Taylor series expansion

$$g(z_{(i)}) \cong g(t_{(i)}) + \left[z_{(i)} - t_{(i)}\right]\left\{\frac{d}{dz} g(z)\right\}_{z=t_{(i)}}.$$

$$= \alpha_i + \beta_i z_{(i)},\ 1 \le i \le n\,. \tag{1.4.5}$$

Thus,

$$\alpha_i = \frac{(2/k)t_{(i)}^3}{\left\{1 + (1/k)t_{(i)}^2\right\}^2} \quad \text{and} \quad \beta_i = \frac{1 - (1/k)t_{(i)}^2}{\left\{1 + (1/k)t_{(i)}^2\right\}^2},\ 1 \le i \le n\,. \tag{1.4.6}$$

The values of $t_{(i)}\ (1 \le i \le n)$ are given in Tiku and Kumra (1981) for $p = 2(.5)10$ and $n \le 20$. For $n \ge 10$, the approximate values of $t_{(i)}$ obtained from the equations

$$\frac{1}{\sqrt{k}\beta(1/2, p - 1/2)}\int_{-\infty}^{t_{(i)}} \left(1 + \frac{z^2}{k}\right)^{-p} dz = \frac{i}{n+1},\ (1 \le i \le n)\,, \tag{1.4.7}$$

are used. Realize that $t = \sqrt{(\nu / k)}z$ has Student's t distribution with $\nu = 2p - 1$ degrees of freedom. An IMSL subroutine is available to compute $t_{(i)}$ from (1.4.7). The use of the approximate values of $t_{(i)}$ for all $n \geq 10$ does not affect the efficiency of the resulting estimators adversely (Tiku and Suresh, 1992; Tiku et al., 2001).

Modified likelihood equations are obtained by incorporating (1.4.5) in (1.4.2)-(1.4.4):

$$\frac{\partial \ln L}{\partial \theta_0} \cong \frac{\partial \ln L^*}{\partial \theta_0} = \frac{2p}{k\sigma}\sum_{i=1}^{n}\{\alpha_i + \beta_i z_{(i)}\} = 0 \qquad (1.4.8)$$

$$\frac{\partial \ln L}{\partial \theta_1} \cong \frac{\partial \ln L^*}{\partial \theta_1} = \frac{2p}{k\sigma}\sum_{i=1}^{n} x_{[i]}\{\alpha_i + \beta_i z_{(i)}\} = 0 \text{ and} \qquad (1.4.9)$$

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{n} z_{(i)}\{\alpha_i + \beta_i z_{(i)}\} = 0\,. \qquad (1.4.10)$$

The equations (1.4.8)-(1.4.10) are asymptotically equivalent to the likelihood equations (1.4.2)-(1.4.4); see Vaughan and Tiku (2000) for a rigorous mathematical proof.

1.5 The MML Estimators

The solutions of the equations (1.4.8)-(1.4.10) are the following MML estimators:

$$\hat{\theta}_0 = \bar{y}_{[.]} - \hat{\theta}_1 \bar{x}_{[.]}, \quad \hat{\theta}_1 = K + L\hat{\sigma}\text{, and} \qquad (1.5.1)$$

$$\hat{\sigma} = \left\{B + \sqrt{(B^2 + 4nC)}\right\} \Big/ 2\sqrt{\{n(n-2)\}} \qquad (1.5.2)$$

where

$$m = \sum_{i=1}^{n} \beta_i \ , \ \bar{y}_{[.]} = (1/m)\sum_{i=1}^{n} \beta_i y_{[i]} \ , \ \bar{x}_{[.]} = (1/m)\sum_{i=1}^{n} \beta_i x_{[i]} \ ,$$

$$K = \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]} \Big/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2 \ ,$$

$$L = \sum_{i=1}^{n} \alpha_i (x_{[i]} - \bar{x}_{[.]}) \Big/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2 \ ,$$

$$B = \frac{2p}{k} \sum_{i=1}^{n} \alpha_i \{ y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]}) \}, \text{ and}$$

$$C = \frac{2p}{k} \sum_{i=1}^{n} \beta_i \{ y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]}) \}^2$$

$$= \frac{2p}{k} \left\{ \sum_{i=1}^{n} \beta_i (y_{[i]} - \bar{y}_{[.]})^2 - K \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[i]}) y_{[i]} \right\}. \qquad (1.5.3)$$

Note that $B/\sqrt{(nC)} \cong 0$ for large n and, consequently, $\hat{\sigma}^2 \cong C/(n-2)$.

Since for symmetric distributions $t_{(n-i+1)} = -t_{(i)}$, it immediately follows that

$$\sum_{i=1}^{n} \alpha_i = 0 \quad \text{and} \quad \beta_{n-i+1} = \beta_i \, , \, 1 \le i \le n \, . \qquad (1.5.4)$$

Computations: The MML estimators are computed in two iterations. In the first place, they are computed from the order statistics $w_{(i)}$ of $w_i = y_i - \tilde{\theta}_1 x_i$ $(1 \le i \le n)$, $\tilde{\theta}_1 = \sum_{i=1}^{n} (x_i - \bar{x}) y_i \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2$ being the least squares (LS) estimator of θ_1. In the second iteration, the estimate $\tilde{\theta}_1$ is replaced by $\hat{\theta}_1$ and the computations repeated. Tiku et al. (2001) and Islam et al. (2001) show that no more than two iterations are needed for the estimates to stabilize sufficiency enough.

Comment: It is clear from (1.5.2)-(1.5.3) that $\hat{\sigma}$ is real and positive if $\beta_i \geq 0$ for all $i = 1,2,...,n$. Now, β_i is an increasing sequence until the middle value and then decreases again in a symmetric fashion. Thus, if $\beta_1 \geq 0$ then all β_i coefficients are nonnegative. For small p and large n, however, β_1 (and possibly few other coefficients) can be negative (Tiku and Suresh, 1992; Vaughan, 1992). Consequently, $\hat{\sigma}$ can cease to be real. In such situations, we recast the linear approximation (1.4.5). This can be done since $g(z_{(i)})$ is bounded and so are α_i and β_i $(1 \leq i \leq n)$. Now (Tiku et al., 2001)

$$g(z_{(i)}) \cong \alpha_i^* + \beta_i^* z_{(i)},\ 1 \leq i \leq n, \tag{1.5.5}$$

where

$$\alpha_i^* = 0 \quad \text{and} \quad \beta_i^* = 1/\left\{1 + (1/k)t_{(i)}^2\right\}. \tag{1.5.6}$$

Asymptotically,

$$\alpha_i + \beta_i z_{(i)} \cong \alpha_i^* + \beta_i^* z_{(i)} \quad \text{since} \quad z_{(i)} - t_{(i)} \cong 0. \tag{1.5.7}$$

Remark: Since β_i is negative only for small p and large n, the linear functional (1.5.5) is not used that often.

Asymptotic Variances-Covariances: Since the MML estimators (1.5.1)-(1.5.2) are asymptotically equivalent to the ML estimators, their asymptotic variance-covariance matrix is given by $I^{-1}(\theta_0, \theta_1, \sigma)$, where I is the Fisher information matrix consisting of the elements $-E\left(\partial^2 \ln L/\partial\theta_0^2\right)$, $-E\left(\partial^2 \ln L/\partial\theta_0\partial\theta_1\right)$, etc.

The Fisher information matrix is

$$I(\theta_0,\theta_1,\sigma)=\frac{n}{\sigma^2}\frac{p(p-1/2)}{(p+1)(p-3/2)}\begin{pmatrix}1 & \bar{x} & 0\\ \bar{x} & (1/n)\sum_{i=1}^{n}x_i^2 & 0\\ 0 & 0 & \dfrac{2(p-3/2)}{p}\end{pmatrix}\ (p\geq 2).\quad (1.5.8)$$

This gives, in particular, the following MVB:

$$MVB(\theta_0)=\frac{\sigma^2}{n}\frac{(p+1)(p-3/2)}{p(p-1/2)}\left\{1+\frac{\bar{x}^2}{s_x^2}\right\},\ s_x^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

$$MVB(\theta_1)=\frac{\sigma^2}{n}\frac{(p+1)(p-3/2)}{p(p-1/2)}\frac{1}{s_x^2},\ \text{and}\ MVB(\sigma)=\frac{\sigma^2}{2n}\frac{(p+1)}{(p-1/2)}.\qquad (1.5.9)$$

Tiku et al. (2001) simulated the means and variances of $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$ for $p=2,3,4\ \text{and}\ 5$ and $n=20,50\ \text{and}\ 100$. They showed that these estimators have negligible bias and their variances are very close to the minimum variance bounds above. In other words, the MML estimators are highly efficient as expected.

Remark: It may be noted that $V(\hat{\theta}_1)$ is inversely proportional to $\sum_{i=1}^{n}(x_i-\bar{x})^2$, as in the situation when the error e is normally distributed. Therefore, a design which is optimal for normal is also optimal for Student's t family, at any rate for large n. This is a very useful result since there is no necessity of re-inventing optimal designs.

Least Squares: No distributional assumptions as such are made in deriving the LS estimators. The only requirement is the existence of the mean and the variance of e. The estimators of θ_0 and θ_1 are obtained by minimizing

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2 \qquad (1.5.10)$$

which gives

$$\tilde{\theta}_0 = \bar{y} - \tilde{\theta}_1 \bar{x}\,, \quad \tilde{\theta}_1 = \sum_{i=1}^{n} (x_i - \bar{x}) y_i \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 \text{ , and}$$

$$\tilde{\sigma}^2 = \min \sum_{i=1}^{n} e_i^2 \Big/ (n-2) = \sum_{i=1}^{n} \left\{ y_i - \bar{y} - \tilde{\theta}_1 (x_i - \bar{x}) \right\}^2 \Big/ (n-2)\,. \qquad (1.5.11)$$

The LS estimators are used very widely. For the family (1.3.1) and other families, however, Tiku et al. (2001) and Islam et al. (2001) showed that the MML estimators are enormously more efficient than the LS estimators.

Another important issue is that of robustness. An estimator is said to be robust if it is fully efficient (or nearly so) for an assumed model but maintains high efficiency for plausible alternatives (Tiku et al., 1986, Preface). The MML estimators are known to be remarkably robust (Tiku, 1980; Tiku et al., 2001; Islam et al., 2001; Akkaya and Tiku, 2001). We will address the issue of robustness in later chapters.

1.6. Hypothesis Testing for Symmetric Family

To test $H_0 : \theta_1 = 0$, we have the following result regarding the distribution of the MML estimator $\hat{\theta}_1$ given in (1.5.1).

Lemma 1.2: Conditionally (σ known) the asymptotic distribution of $\hat{\theta}_1(\sigma) = K + L\sigma$ is normal with mean θ_1 and variance $\sigma^2 \Big/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2$. This follows from a re-organization of (1.4.9). In the light of (1.4.8), it can be expressed in the form

$$\frac{\partial \ln L^*}{\partial \theta_1} = \frac{\sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2}{\sigma^2} \{(K + L\sigma) - \theta_1\}. \tag{1.6.1}$$

Also

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2 = \frac{p(p-1/2)}{(p+1)(p-3/2)} s_x^2. \tag{1.6.2}$$

To test $H_0 : \theta_1 = 0$ against $H_1 : \theta_1 > 0$, we use the statistic $(p \geq 2)$

$$T = \sqrt{\left\{ \frac{np(p-1/2)}{(p+1)(p-3/2)} s_x^2 \right\}} \left(\frac{\hat{\theta}_1}{\hat{\sigma}} \right). \tag{1.6.3}$$

The null distribution of T is asymptotically normal $N(0,1)$. This follows from (1.6.1) and the fact that $\hat{\sigma}$ converges to σ as n becomes large. For small n, the null distribution of T is referred to Student's t with $n-2$ degrees of freedom.

The statistic based on the LS estimators is

$$G = \sqrt{(ns_x^2)} \left(\tilde{\theta}_1 / \tilde{\sigma} \right), \tag{1.6.4}$$

$\tilde{\theta}_1$ and $\tilde{\sigma}_1$ are given in (1.5.11). The null distribution of G is asymptotically normal $N(0,1)$. This follows from the fact that $\tilde{\sigma}$ converges to σ and $\tilde{\theta}_1$ is asymptotically normal by Central Limit Theorem. The asymptotic nonnull distributions of T and G are also normal with noncentrality parameters Δ and Δ_0, respectively:

$$\Delta^2 = \left[\{np(p-1/2)/(p+1)(p-3/2)\}s_x^2\right](\theta_1/\sigma)^2 \text{ and } \Delta_0 = (ns_x^2)(\theta_1/\sigma)^2 . \quad (1.6.5)$$

The ratio of the two noncentrality parameters is

$$\Delta^2/\Delta_0^2 = p(p-1/2)/(p+1)(p-3/2) \quad (1.6.6)$$

which is greater than 1 for all $p < \infty$. For $p = \infty$, $\Delta^2/\Delta_0^2 = 1$ which was to be expected since the distribution (1.3.1) reduces to normal, and the estimators (1.5.1)-(1.5.2) reduce to the LS estimators. The T test is, therefore, asymptotically more powerful than the classical t test for all $p < \infty$. For $p = \infty$, T reduces to t.

Tiku et al. (2001) investigate the power properties and robustness of the T and G tests. Realize that the latter is very commonly used in practice. They show that the T test is robust and more powerful than the G test. They have in fact a very interesting result, namely, if the sample contains outliers or the sample is contaminated, the T test has not only smaller Type I error but has generally higher power. For illustration, we reproduce their results in Table 1.1. The assumed model is (1.3.1) with $p = 3$, i.e. $f(3,\sigma)$, but the sample comes from ($\sigma = 1$ without loss of generality)

(a) the outlier model: $(n-1)$ observations from $f(3,1)$ and 1 (we do not know which) comes from $f(3,8)$;

(b) contamination model: $0.90f(3,1) + 0.1N(0,8^2)$.

Their simulated values of the Type I error and power are given in Table 1.1 and are based on $[100,000/n]$ (integer value) Monte Carlo runs.

Table 1.1 Values of the power of the T and G tests; $n = 30$.

	Model (a)		Model (b)	
θ_1	T	G	T	G
0.0	0.045	0.080	0.025	0.044
0.4	0.31	0.34	0.28	0.30
0.8	0.69	0.64	0.66	0.58
1.2	0.88	0.80	0.86	0.77
1.6	0.95	0.89	0.97	0.89
2.0	0.97	0.93	0.99	0.95

Not only has the G test lower power but it has also substantially higher Type I error than the presumed level 0.050 for model (a). The G test, therefore, has neither criterion robustness nor efficiency robustness.

1.7 Stochastic Design Variable

In numerous applications the design variable X in the model (1.1.1) is also stochastic. This is indeed the research area (b) mentioned earlier. Consider, for example, the following well-known data given in Table 1.2 where X represents 100 times the white blood counts and Y represents the survival times (in weeks) of patients who died of acute myelogenous leukemia (Gross and Clark, 1975).

Table 1.2 Gross and Clark data.

i	1	2	3	4	5	6	7	8
X_i:	23	7.5	43	26	60	105	100	170
Y_i:	65	156	100	134	16	108	121	4
i	9	10	11	12	13	14	15	16
X_i:	54	70	94	320	350	1000	1000	520
Y_i:	39	143	56	26	22	1	1	5

Source: Gross and Clark (1975)

Clearly, X and Y are both random variables. In fact, Vaughan and Tiku (2000) showed that it is reasonable to regard X as a Weibull random variable and the conditional distribution of Y given $X = x$ as normal. To initiate a proper analysis of the Gross and Clark data given in Table 1.2, we proceed step by step as follows.

Marginal and Conditional both normal: In the first place assume that in the model (1.1.1), X is normal $N(\mu_1, \sigma_1^2)$ and Y given $X = x$ is normal $N(\mu_{2.1}, \sigma_{2.1}^2)$, where $\mu_{2.1} = \mu_2 + \rho(\sigma_2 / \sigma_1)(x - \mu_1) = \theta_0 + \theta_1 x$ and $\sigma_{2.1}^2 = \sigma_2^2(1 - \rho^2)$; $\theta_0 = \mu_2 - \theta_1 \mu_1$ and $\theta_1 = \rho(\sigma_2 / \sigma_1)$. Realize that the distribution of e is, in fact, the conditional distribution of Y given $X = x$ other than the mean $E(e) = 0$. The parameter ρ is the correlation coefficient between X and Y. Let (x_i, y_i), $1 \le i \le n$, be a random sample of size n. The likelihood function is

$$L \propto \left(\frac{1}{\sigma_1}\right)^n e^{-\sum_{i=1}^{n}(x_i-\mu_1)^2/2\sigma_1^2} \left(\frac{1}{\sigma_{2.1}}\right)^n e^{-\sum_{i=1}^{n}\{y_i-\mu_2-\rho(\sigma_2/\sigma_1)(x_i-\mu_1)\}^2/2\sigma_{2.1}^2}$$

$$= L_x L_{y|x}, \qquad (1.7.1)$$

which is essentially of the form

$$\prod_{i=1}^{n} g(x_i) h(y_i \mid x_i) \qquad (1.7.2)$$

as it should be.

The ML estimators are solutions of the following equations:

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{1}{\sigma_1^2}\sum_{i=1}^{n}(x_i - \mu_1)^2 + \frac{\rho}{\sigma_{2.1}^2}\left(\frac{\sigma_2}{\sigma_1}\right)\sum_{i=1}^{n}\left\{y_i - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_i - \mu_1)\right\} = 0 \qquad (1.7.3)$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1^3}\sum_{i=1}^{n}(x_i - \mu_1)^2$$

$$-\frac{\rho}{\sigma_1\sigma_{2.1}^2}\left(\frac{\sigma_2}{\sigma_1}\right)\sum_{i=1}^{n}(x_i - \mu_1)\left\{y_i - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_i - \mu_1)\right\} = 0 \qquad (1.7.4)$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{1}{\sigma_{2.1}^2}\sum_{i=1}^{n}\left\{y_i - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_i - \mu_1)\right\} = 0 \qquad (1.7.5)$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2\sigma_{2.1}^2}\sum_{i=1}^{n}\left\{y_i - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_i - \mu_1)\right\}^2$$

$$+\frac{\rho}{\sigma_1\sigma_{2.1}^2}\sum_{i=1}^{n}(x_i - \mu_1)\left\{y_i - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_i - \mu_1)\right\} = 0 \text{ and} \qquad (1.7.6)$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_{2.1}^2(1-\rho^2)}\sum_{i=1}^{n}\left\{y_i - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_i - \mu_1)\right\}^2$$

$$+\frac{1}{\sigma_{2.1}^2}\left(\frac{\sigma_2}{\sigma_1}\right)\sum_{i=1}^{n}(x_i - \mu_1)\left\{y_i - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_i - \mu_1)\right\} = 0\,. \qquad (1.7.7)$$

Also $\left(\theta_1 = \rho\frac{\sigma_2}{\sigma_1}\right)$,

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{1}{\sigma_{2.1}^2}\sum_{i=1}^{n}(x_i - \mu_1)\{(y_i - \mu_2) - \theta_1(x_i - \mu_1)\} = 0 \qquad (1.7.8)$$

which gives

$$\theta_1 = \sum_{i=1}^{n}(x_i - \mu_1)(y_i - \mu_2) \Big/ \sum_{i=1}^{n}(x_i - \mu_1)^2\,. \qquad (1.7.9)$$

The solutions of the equations above are the well-known ML estimators of μ_1, μ_2, σ_1^2, σ_2^2, ρ and θ_1, respectively:

$$\bar{x} = (1/n)\sum_{i=1}^{n} x_i\ ,\ \bar{y} = (1/n)\sum_{i=1}^{n} y_i\ ,\ s_1^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \Big/ (n-1)\,,$$

$$s_2^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 \Big/ (n-1)\,,\ \ \hat{\rho} = \sum_{i=1}^{n}(x_i - \bar{x})y_i \Big/ \sqrt{\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2\right\}}\,,\ \text{and}$$

$$\hat{\theta}_1 = \sum_{i=1}^{n}(x_i - \bar{x})y_i \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2\,. \tag{1.7.10}$$

Of course, s_1^2 and s_2^2 are bias-corrected estimators. The bias-corrected ML estimator of $\sigma_{2.1}^2$ is $s_{2.1}^2 = s_e^2 = s_2^2(1-\hat{\rho}^2) = \sum_{i=1}^{n}\left\{y_i - \bar{y} - \hat{\theta}_1(x_i - \bar{x})\right\}^2 \Big/ (n-2)$.

Remark: It is very important to realize that the ML estimators of μ_2, σ_2^2 and θ_1 obtained from $L_{y|x}$ (with μ_1 and σ_1^2 replaced by $\bar{x}$ and s_1^2, respectively) are exactly the same as those obtained from the entire likelihood function L. This has unfortunately created the impression that using the conditional likelihood function $L_{y|x}$ (with μ_1 and σ_1^2 replaced by $\bar{x}$ and s_1^2, respectively) will suffice and the marginal likelihood L_x has no role to play in the estimation of μ_2, σ_2^2 and θ_1. This is not true if X has a nonnormal distribution. This is illustrated in the following section.

1.8 Nonnormal Marginal Distribution

Suppose that the marginal distribution of X is the extreme-value distribution (Vaughan and Tiku, 2000)

$$g(x) = \frac{1}{\sigma_1}\exp\left[-\left(\frac{x-\mu_1}{\sigma_1}\right) - \exp\left(-\frac{x-\mu_1}{\sigma_1}\right)\right],\ -\infty < x < \infty\,. \tag{1.8.1}$$

This distribution is important not only on its own but also for the fact that if U has the Weibull distribution

$$f(u)=\frac{\alpha}{\beta}\left(\frac{u}{\beta}\right)^{\alpha-1}\exp\left(-\left(\frac{u}{\beta}\right)^{\alpha}\right),\ 0<u<\infty, \tag{1.8.2}$$

then $X=\ln U$ has the extreme-value distribution (1.8.1) with $\mu_1=\ln\beta$ and $\sigma_1=1/\alpha$. Since α determines the failure rate of U, the parameter σ_1 in (1.8.1) takes on an added importance.

Suppose that the distribution of e is normal as in the previous section. Given a random sample (x_i, y_i), the likelihood function is

$$L\propto\sigma_1^{-n}\sigma_2^{-n}(1-\rho^2)^{-n/2}\exp\left(-\sum_{i=1}^{n}(z_i-e^{-z_i})-\frac{1}{2\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n}e_i^2\right) \tag{1.8.3}$$

where $z_i=(x_i-\mu_1)/\sigma_1$ and $e_i=y_i-\mu_2-\rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_i-\mu_1)$ $(1\le i\le n)$.

The likelihood equations for estimating μ_1, σ_1, μ_2, σ_2 and ρ are

$$\frac{\partial\ln L}{\partial\mu_1}=\frac{n}{\sigma_1}-\frac{1}{\sigma_1}\sum_{i=1}^{n}\exp(-z_i)-\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\sum_{i=1}^{n}e_i=0 \tag{1.8.4}$$

$$\frac{\partial\ln L}{\partial\sigma_1}=-\frac{n}{\sigma_1}-\frac{1}{\sigma_1}\sum_{i=1}^{n}z_i\exp(-z_i)+\frac{1}{\sigma_1}\sum_{i=1}^{n}z_i-\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\sum_{i=1}^{n}z_ie_i=0 \tag{1.8.5}$$

$$\frac{\partial\ln L}{\partial\mu_2}=\frac{1}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n}e_i=0 \tag{1.8.6}$$

$$\frac{\partial\ln L}{\partial\sigma_2}=-\frac{n}{\sigma_2}+\frac{1}{\sigma_2^3(1-\rho^2)}\sum_{i=1}^{n}e_i^2+\frac{\rho}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n}z_ie_i=0 \quad \text{and} \tag{1.8.7}$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2^2(1-\rho^2)^2}\sum_{i=1}^{n} e_i^2 + \frac{1}{\sigma_2(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0\,. \qquad (1.8.8)$$

Writing $\theta_1 = \rho\left(\frac{\sigma_2}{\sigma_1}\right)$, we also have

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0\,. \qquad (1.8.9)$$

Due to the intractable nature of the first two equations, (1.8.4)-(1.8.8) have no explicit solutions. Solving them by iteration is indeed problematic as is, in general, true with likelihood equations (Tiku et. al., 1986).

1.9 Modified Likelihood Equations

Since the ML estimators are intractable, we derive the MML estimators. We reiterate that under some very general regularity conditions, the MML estimators have the following properties (Tiku and Suresh, 1992; Vaughan 1992; Vaughan and Tiku, 2000):

(a) asymptotically, the MML estimators are fully efficient, i.e., they are unbiased and their variances are equal to the MVB (minimum variance bounds);

(b) for small samples, the MML estimators are almost fully efficient, that is, they have no or negligible bias and their variances are only marginally bigger than the MVB;

(c) the MML estimators are explicit functions of sample observations and are, therefore, easy to compute.

To derive the MML estimators in the present situation, let

$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)} \tag{1.9.1}$$

be the order statistics of the random sample $x_1, x_2, ..., x_n$. Let $y_{[i]}$ be the y_j observation which corresponds to $x_{(i)}$; $y_{[i]}$ may be called concomitant of $x_{(i)}$. The sample observations are now denoted by $(x_{(i)}, y_{[i]})$, $1 \le i \le n$. Write

$$z_{(i)} = (x_{(i)} - \mu_1)/\sigma_1 \text{ and } e_{[i]} = y_{[i]} - \mu_2 - \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_{(i)} - \mu_1) \quad (1 \le i \le n). \tag{1.9.2}$$

The fact that complete sums are invariant to ordering implies that

$$\sum_{i=1}^{n} e_{[i]} = 0 \text{ and } \sum_{i=1}^{n} z_{(i)} e_{[i]} = 0 \tag{1.9.3}$$

from (1.8.6) and (1.8.9). Thus, the equations (1.8.4)-(1.8.8) reduce to

$$\begin{aligned}
&\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{1}{\sigma_1}\sum_{i=1}^{n} \exp(-z_{(i)}) = 0 \\
&\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} - \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{(i)} \exp(-z_{(i)}) = 0 \\
&\frac{\partial \ln L}{\partial \mu_2} = \frac{1}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} e_{[i]} = 0 \\
&\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^3(1-\rho^2)}\sum_{i=1}^{n} e_{[i]}^2 = 0 \text{ and} \\
&\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2^2(1-\rho^2)^2}\sum_{i=1}^{n} e_{[i]}^2 = 0 .
\end{aligned} \tag{1.9.4}$$

Solving these equations is problematic because of the function $\exp(-z_{(i)})$.

Linearization: To linearize $\exp(-z_{(i)})$, we need the values of $t_{(i)} = E\left(z_{(i)}\right)$, $1 \le i \le n$. They are given by (Lieblien, 1953; White, 1969)

$$t_{(i)} = c + i\binom{n}{i}\sum_{j=0}^{n-i}\binom{n-i}{j}(-1)^j \frac{\ln(i+j)}{i+j} \tag{1.9.5}$$

where $c = \int_0^\infty \ln(u)\exp(-u)du \cong 0.57722$ is the Euler constant. For $n \ge 10$, however, $t_{(i)}$ obtained from the following equation are used (David, 1981),

$$\int_{-\infty}^{t_{(i)}} g(z)dz = \frac{i}{n+1}, \qquad g(z) = \exp(-z - e^{-z}) \tag{1.9.6}$$

which gives

$$t_{(i)} \cong -\ln\left[-\ln(i/(n+1))\right],\ 1 \le i \le n. \tag{1.9.7}$$

Now we have the linear functional, obtained from the first two terms of a Taylor series expansion,

$$e^{-z_{(i)}} \cong \alpha_i - \beta_i z_{(i)},\ 1 \le i \le n, \tag{1.9.8}$$

where

$$\beta_i = -\left\{\frac{d}{dz}e^{-z}\right\}_{z=t_{(i)}} = e^{-t_{(i)}} \quad \text{and} \quad \alpha_i = e^{-t_{(i)}}(1 + t_{(i)}). \tag{1.9.9}$$

It may be noted that β_i is positive for all $i = 1,2,...,n$.

The modified likelihood equations are obtained by incorporating (1.9.8) in the first two equations in (1.9.4). Their solutions are the following MML estimators:

$$\hat{\mu}_1 = K + D\hat{\sigma}_1 \tag{1.9.10}$$

$$\hat{\sigma}_1 = \frac{\left(B + \sqrt{B^2 + 4nC}\right)}{2n} \tag{1.9.11}$$

$$\hat{\mu}_2 = \bar{y} - \hat{\rho}\left(\frac{\hat{\sigma}_2}{\hat{\sigma}_1}\right)(\bar{x} - \hat{\mu}_1) \tag{1.9.12}$$

$$\hat{\sigma}_2 = \left(s_y^2 + \frac{s_{xy}^2}{s_x^2}\left(\frac{\hat{\sigma}_1^2}{s_x^2} - 1\right)\right)^{1/2} \tag{1.9.13}$$

and

$$\hat{\rho} = \frac{s_{xy}}{s_x^2}\frac{\hat{\sigma}_1}{\hat{\sigma}_2};\ \ \hat{\theta}_1 = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)(y_i - \hat{\mu}_2)}{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)^2}. \tag{1.9.14}$$

Here, $n\bar{x} = \sum_{i=1}^{n} x_i$, $n\bar{y} = \sum_{i=1}^{n} y_i$, $(n-1)s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$, $(n-1)s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2$,

$(n-1)s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$, and

$K = \frac{1}{m}\sum_{i=1}^{n}\beta_i x_{(i)}$, $D = \frac{1}{m}\sum_{i=1}^{n}(1-\alpha_i)$, $m = \sum_{i=1}^{n}\beta_i$

$B = \sum_{i=1}^{n}(1-\alpha_i)(x_{(i)} - K)$ and $C = \sum_{i=1}^{n}\beta_i(x_{(i)} - K)^2 = \sum_{i=1}^{n}\beta_i x_{(i)}^2 - mK^2$. (1.9.15)

Note how different the estimators of μ_2, σ_2 and θ_1 are than those based on the conditional likelihood $L_{y|x}$ (with $\hat{\mu}_1$ and $\hat{\sigma}_1^2$ replaced by $\bar{x}$ and s_x^2, respectively).

Inevitably, the estimators $\hat{\sigma}_2^2$ and $\hat{\rho}^2$ have to have the property that $\hat{\sigma}_2^2 > 0$ and $\hat{\rho}^2 \leq 1$. We now establish these results as follows.

Remark: The estimator $\hat{\sigma}_2^2$ is always positive. This follows from the fact that $s_{xy}^2 \leq s_x^2 s_y^2$ so that $s_y^2 - (s_{xy}^2 / s_x^2) \geq s_y^2 - s_y^2 = 0$.

Remark: The estimator $\hat{\rho}^2$ is bounded above by 1. This follows from the fact that

$$\hat{\rho}^2 = \frac{1}{\left\{1 + \frac{s_x^2 s_y^2}{s_{xy}^2 \hat{\sigma}_1^2}\left(1 - \frac{s_{xy}^2}{s_x^2 s_y^2}\right)\right\}} \tag{1.9.16}$$

and $0 \leq s_{xy}^2 \leq s_x^2 s_y^2$; see also Tiku and Kambo (1992).

Lemma 1.5: The asymptotic distribution of $\hat{\mu}_1$ is conditionally (for known σ_1) normal with mean μ_1 and variance σ_1^2 / m.

For known σ_1, $\hat{\mu}_1 = \hat{\mu}_1(\sigma_1) = K + D\sigma_1$, and

$$\frac{\partial \ln L^*}{\partial \mu_1} = \frac{m}{\sigma_1^2}\{\hat{\mu}_1(\sigma_1) - \mu_1\}. \tag{1.9.17}$$

The result follows from the fact that $\partial \ln L^* / \partial \mu_1$ is asymptotically equivalent to $\partial \ln L / \partial \mu_1$ and $E(\partial^r \ln L^* / \partial \mu_1^r) = 0$ for all $r \geq 3$. It also follows from (1.9.17) that $\hat{\mu}_1$ is the MVB estimator (asymptotically).

Lemma 1.6: Asymptotically, the estimator $\hat{\sigma}_1$ is conditionally (for known μ_1) the MVB estimator of σ_1.

For known μ_1, $\hat{\sigma}_1 = \hat{\sigma}_1(\mu_1)$ is given by

$$\hat{\sigma}_1(\mu_1) = \frac{\left(B_0 + \sqrt{B_0^2 + 4nC_0}\right)}{2n} \tag{1.9.18}$$

where

$$B_0 = \sum_{i=1}^{n}(1-\alpha_i)(x_{(i)} - \mu_1) \text{ and } C_0 = \sum_{i=1}^{n}\beta_i (x_{(i)} - \mu_1)^2 .$$

Further,

$$\frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1^3}\left(\frac{B_0 + \sqrt{B_0^2 + 4nC_0}}{2n} - \sigma_1\right)\left(\frac{B_0 - \sqrt{B_0^2 + 4nC_0}}{2n} - \sigma_1\right). \tag{1.9.19}$$

The only admissible solution of (1.9.19) is $\sigma_1 = \hat{\sigma}_1(\mu_1)$ and the result follows; see Kendall and Stuart (1979).

Corollary: Since for large n, B_0 is very small as compared to $\sqrt{(nC_0)}$, $B_0 \big/ \sqrt{nC_0}$ is negligibly small. Thus,

$$\frac{\partial \ln L^*}{\partial \sigma_1} \cong \frac{n}{\sigma_1^3}\left(\frac{C_0}{n} - \sigma_1^2\right). \tag{1.9.20}$$

Therefore, $\hat{\sigma}_1^2(\mu_1) \cong C_0/n$ is asymptotically the MVB of σ_1^2 and C_0/σ_1^2 is for large n distributed as chi-square with n degrees of freedom; see Kendall and Stuart (1979).

Remark: For small $n\ (\leq 15)$, $\hat{\mu}_1$ and $\hat{\sigma}_1$ have some bias. The bias corrected estimators are given by

$$\hat{\mu}_1 = K - \frac{1}{m}\left(\sum_{i=1}^{n} \beta_i t_{(i)}\right)\hat{\sigma}_1 \quad \text{with} \quad \hat{\sigma}_1 = \frac{\left(B + \sqrt{B^2 + 4nC}\right)}{2m} \tag{1.9.21}$$

Lemma 1.7: For known μ_1 and μ_2, $\hat{\theta}_1(\mu_1, \mu_2)$ is the MVB estimator of θ_1 and is normally distributed with mean θ_1 and variance $\sigma_2^2(1-\rho^2)\Big/\sum_{i=1}^{n}(x_i - \mu_1)^2$.

This follows from the fact that

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\partial \ln L^*}{\partial \theta_1} = \frac{\sum_{i=1}^{n}(x_i - \mu_1)^2}{\sigma_2^2(1-\rho^2)}\left(\hat{\theta}_1(\mu_1, \mu_2) - \theta_1\right). \tag{1.9.22}$$

1.10 Asymptotic Covariance Matrix

The asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1$, $\hat{\mu}_2$, $\hat{\sigma}_2$ and $\hat{\rho}$ is given by $I^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$, where I is the Fisher information matrix consisting of the elements $-E(\partial^2 \ln L/\partial \mu_1^2)$, $-E(\partial^2 \ln L/\partial \sigma_1^2)$, etc. Writing $\kappa = (\pi^2 + 6c^2 + 6 - 12c)/\pi^2 \cong 1.10866$ ($c \cong 0.57722$ being the Euler constant), the elements of this matrix are (Vaughan and Tiku, 2000):

$$-E(\partial^2 \ln L/\partial \mu_1^2) = \kappa\sigma_1^2/n,\ -E(\partial^2 \ln L/\partial \mu_1 \partial \sigma_1) = 6(1-c)\sigma_1^2\big/n\pi^2,$$

and so on.

Testing the null hypothesis $H_0 : \rho = 0$ is of primary importance. That can be done as follows.

Since the MML estimators are asymptotically equivalent to the ML estimators, the likelihood function L is maximized (asymptotically) by the MML estimators (Vaughan and Tiku, 2000). Thus, the likelihood ratio is (asymptotically)

$$\hat{\lambda} = \frac{\max(L \mid H_0)}{\max(L \mid H_1)}$$

$$= \left(\frac{\hat{\sigma}_2^2}{s_y^2} \right)^{n/2} (1 - \hat{\rho}^2)^{n/2} \exp\left[\frac{(n-1)s_y^2}{2(1-\hat{\rho}^2)\hat{\sigma}_2^2}(1 - \hat{\rho}_0^2) - \frac{(n-1)}{2} \right] \qquad (1.10.1)$$

where $\hat{\rho}_0 = \dfrac{s_{xy}}{s_x s_y}$ is the usual Pearson sample correlation coefficient. Since s_y^2 and $\hat{\sigma}_2^2$ both converge to σ_2^2, and $\hat{\rho}_0$ and $\hat{\rho}$ both converge to ρ as n tends to infinity, the exponent is essentially zero for large n, so that the likelihood ratio is a monotonic function of $\hat{\rho}^2$. Thus, to test H_0 against $H_1 : \rho < 0$ (or $\rho > 0$), the test based on $\hat{\rho}$ will be uniformly most powerful (asymptotically). Vaughan and Tiku (2000), therefore, propose the statistic

$$W = \hat{\rho} \Big/ \sqrt{6/(n\pi^2)} \qquad (1.10.2)$$

$$= \hat{\rho}\pi\sqrt{\frac{n}{6}};$$

$6/(n\pi^2)$ is the asymptotic variance of $\hat{\rho}$ under H_0, obtained from

$$\left\{ I^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho) \right\}_{\rho=0}. \qquad (1.10.3)$$

The null distribution of W is referred to normal $N(0,1)$. Large values of W lead to the rejection of H_0 against $H_1 : \rho > 0$ and small values of W lead to the rejection of H_0 against $H_1 : \rho < 0$.

For the Gross-Clark data (see Table 1.2) reported in Section 1.7,

$$\hat{\rho} = -0.7121 \quad \text{and} \quad W = -3.654 . \tag{1.10.4}$$

The computed value of W being less than -3.09, the null hypothesis H_0 is rejected even at 0.1% level, let alone the 1% and 5% significance levels used so often. The conclusion is, therefore, that a higher white blood count tends to shorten the survival time of a patient. This is in conformity with the medical opinion.

Remark: The estimate $\hat{\rho} = -0.7121$ above is estimating the correlation coefficient between ln of white blood count and the survival time of a patient. To estimate the correlation coefficient between white blood count and survival time, however, we need to take the distribution of X as Weibull and the conditional distribution of Y given $X = x$ as normal. That is the model proposed originally by Vaughan and Tiku (2000) but they did not give the mathematical analyses. We consider this situation in Chapter 2, and develop the required mathematics.

Another interesting situation is that X has Student's t distribution and the conditional distribution of Y given $X = x$ is normal. Tiku and Kambo (1992) mention that this model has genetic applications. We now briefly review the estimation of parameters under this model.

1.11 Student t Marginal and Normal Conditional

Let the marginal distribution of X be

$$g(x) \propto \frac{1}{\sigma_1}\left\{1+\frac{(x-\mu_1)^2}{k\sigma_1^2}\right\}^{-p}, \; -\infty < x < \infty; \tag{1.11.1}$$

$k = 2p-3$ and $p \geq 2$. It is easy to show that $E(X) = \mu_1$ and $V(X) = \sigma_1^2$, and the distribution of $t = \sqrt{(\nu/k)}z$, $z = (x-\mu_1)/\sigma_1$, has a Student's t distribution with $\nu = 2p-1$ degrees of freedom.

Here, the likelihood function is

$$L \propto \left(\frac{1}{\sigma_1}\right)^n \prod_{i=1}^{n}\left\{1+\frac{(x_i-\mu_1)^2}{k\sigma_1^2}\right\}^{-p}\left\{\frac{1}{\sigma_2^2(1-\rho^2)}\right\}^{n/2} e^{-\sum_{i=1}^{n}\varepsilon_i^2 \Big/ 2\sigma_2^2(1-\rho^2)}, \tag{1.11.2}$$

$\varepsilon_i = y_i - \mu_2 - \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_i-\mu_1)$. The likelihood equations work in terms of the intractable functions

$$g(z_i) = z_i \Big/ \left(1+(1/k)z_i^2\right), \; 1 \leq i \leq n. \tag{1.11.3}$$

Consequently, the likelihood equations have no explicit solutions. In fact, they have multiple roots. Solving them by iteration and locating the ML estimates is very problematic (Tiku and Kambo, 1992). To obtain the MML estimators, we first express the likelihood equations in terms of $z_{(i)} = \{x_{(i)} - \mu_1\}/\sigma_1 \;\; (1 \leq i \leq n)$. Writing $t_{(i)} = E\left(z_{(i)}\right) \;\; (1 \leq i \leq n)$, modified likelihood equations are obtained by replacing $g\left(z_{(i)}\right)$ by

$$g\left(z_{(i)}\right) \cong \alpha_i + \beta_i z_{(i)}, \; 1 \leq i \leq n. \tag{1.11.4}$$

The coefficients α_i and β_i are determined by the first two terms of a Taylor series expansion of $g(z_{(i)})$ around $t_{(i)}$. In fact,

$$\alpha_i = \frac{(2/k)t_{(i)}^3}{\left\{1+(1/k)t_{(i)}^2\right\}^2} \quad \text{and} \quad \beta_i = \frac{1-(1/k)t_{(i)}^2}{\left\{1+(1/k)t_{(i)}^2\right\}^2}. \tag{1.11.5}$$

The values of $t_{(i)}$ are available in Tiku and Kumra (1985) and Vaughan (1992b); $t_{(n-i+1)} = -t_{(i)}$ $(1 \le i \le n)$. The modified likelihood equations have explicit solutions, giving the following MML estimators:

$$\hat{\mu}_1 = \left\{\sum_{i=1}^{n} \beta_i x_{(i)}\right\} \Big/ m \quad (m = \sum_{i=1}^{n} \beta_i), \ \hat{\sigma}_1 = \left\{B + \sqrt{(B^2 + 4nC)}\right\} \Big/ 2\sqrt{\{n(n-1)\}}, \tag{1.11.6}$$

$$\hat{\mu}_2 = \bar{y} - \hat{\rho}\left(\frac{\hat{\sigma}_2}{\hat{\sigma}_1}\right)(\bar{x} - \hat{\mu}_1), \tag{1.11.7}$$

$$\hat{\sigma}_2 = \left(s_y^2 + \frac{s_{xy}^2}{s_x^2}\left(\frac{\hat{\sigma}_1^{\,2}}{s_x^2} - 1\right)\right)^{1/2} \quad \text{and} \quad \hat{\rho} = \frac{s_{xy}}{s_x^2}\frac{\hat{\eta}}{\hat{\sigma}}; \tag{1.11.8}$$

$$B = (2p/k)\sum_{i=1}^{n} \alpha_i x_{(i)} \quad \text{and} \quad C = (2p/k)\sum_{i=1}^{n} \beta_i (x_{(i)} - \hat{\mu}_1)^2. \tag{1.11.9}$$

Notice the remarkable property of the MML estimators, namely, in spite of the fact that the extreme-value distribution (1.8.1) is very different from the long-tailed symmetric distribution (1.11.1), the formulae (1.9.10)-(1.9.14) and (1.11.6)-(1.11.8) are exactly similar to one another.

Comment: The estimator $\hat{\sigma}_1$ is real and positive if β_i is positive. For small p and large n, however, β_1 (and possibly few other coefficients) can be negative (Tiku and Suresh, 1992; Vaughan, 1992) as said earlier. In such situations, we replace α_i and β_i by α_i^* and β_i^*, respectively:

$$\alpha_i^* = 0 \quad \text{and} \quad \beta_i^* = 1/\left\{1 + (1/k)t_{(i)}^2\right\}. \tag{1.11.10}$$

This does not alter the asymptotic properties of the estimators.

1.12 Asymptotic Variances and Covariances

The asymptotic variance-covariance matrix V of the MML estimators (1.11.6)-(1.11.8) are obtained by inverting the Fisher information matrix. Because of the symmetry of (1.11.1), V assumes an interesting form:

$$V = \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}_{5\times 5} \tag{1.12.1}$$

where

$$V_1 = \frac{k}{2p}\begin{bmatrix} \dfrac{\rho^2}{\sigma_2^2} + \dfrac{2p}{k}\dfrac{(1-\rho^2)}{\sigma_2^2}\dfrac{m}{n} & \dfrac{\rho}{\sigma_1\sigma_2} \\ \dfrac{\rho}{\sigma_2\sigma_1} & \dfrac{1}{\sigma_1^2} \end{bmatrix}\frac{\sigma_1^2\sigma_2^2}{m} \tag{1.12.2}$$

and, similarly, V_2. Realize that V_1 is different from the variance-covariance matrix of the sample means $\bar{x}$ and $\bar{y}$, namely,

$$V_0 = Cov(\bar{x}, \bar{y}) = \begin{bmatrix} \dfrac{1}{\sigma_2^2} & \dfrac{\rho}{\sigma_1\sigma_2} \\ \dfrac{\rho}{\sigma_2\sigma_1} & \dfrac{1}{\sigma_1^2} \end{bmatrix}\frac{\sigma_1^2\sigma_2^2}{n}. \tag{1.12.3}$$

Since (Tiku and Suresh, 1992)

$$\lim_{n\to\infty}\frac{m}{n}=\frac{(p-1/2)}{p+1}, \tag{1.12.4}$$

V_1-V_0 is always a positive semi-definite matrix, even asymptotically; $V_1-V_0=0$ only if $p=\infty$ in which case (1.11.1) reduces to normal $N(\mu_1,\sigma_1^2)$. It is, therefore, very important to pay attention to nonnormality of the marginal distribution.

Tiku and Kambo (1992) develop a Hotelling type T^2 statistic for testing the mean vector

$$H_0:\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}. \tag{1.12.5}$$

Their statistic is

$$T_p^2=\hat{\mu}'V_1^{-1}\hat{\mu}, \quad \hat{\mu}'=(\hat{\mu}_1,\hat{\mu}_2). \tag{1.12.6}$$

They show that T_p^2 is more powerful than the bivariate normal-theory Hotelling T^2 statistic which they denote by T_∞^2.

Tiku and Kambo (1992) give the likelihood ratio statistic, similar to (1.10.1), for testing $H_0:\rho=0$. They also extend the methodology to censored samples. Censored samples occur due to experimental constraints.

CHAPTER 2

NORMAL CONDITIONAL AND NONNORMAL MARGINAL DISTRIBUTIONS

Summary: We first consider the situation when X has a three parameter Weibull distribution and the conditional distribution of Y given $X = x$ is normal. Then we develop estimation and hypothesis testing procedure for the case when the marginal distribution is Generalized Logistic and the conditional is normal.

2.1 Marginal Weibull and Conditional Normal

In this section we consider the situation when the marginal distribution of X is three parameter Weibull and the conditional distribution of Y given $X = x$ is normal. We first estimate the parameters and then develop the hypothesis testing procedures based on the MML estimators.

2.1.1 Estimation of Parameters

Suppose that the marginal distribution of X is the Weibull distribution with density

$$g(x) = \frac{p}{\sigma_1}\left(\frac{x-\mu_1}{\sigma_1}\right)^{p-1} \exp\left(-\left(\frac{x-\mu_1}{\sigma_1}\right)^{p}\right), \qquad (2.1.1.1)$$

and the conditional distribution of Y given $X = x$ is the normal distribution with density

$$h(y \mid x) = \frac{1}{\sqrt{2\pi}\sigma_2(1-\rho^2)^{1/2}} \times$$

$$\exp\left[-\frac{1}{2\sigma_2^2(1-\rho^2)} \left\{ y - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) \right\}^2 \right] \qquad (2.1.1.2)$$

where $\mu_1 < x < \infty$, $-\infty \langle y \langle \infty$, $\mu_1, \mu_2 \in \Re$, $\sigma_1, \sigma_2 \rangle 0$, $-1 \langle \rho \langle 1$, $p > 0$.

Given a random sample (x_i, y_i) $(1 \le i \le n)$, the likelihood function is

$$L \propto \sigma_1^{-n}\sigma_2^{-n}(1-\rho^2)^{-n/2}\prod_{i=1}^{n} z_i^{p-1} \exp\left(-\sum_{i=1}^{n} z_i^{p} - \frac{1}{2\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} e_i^2 \right) \qquad (2.1.1.3)$$

where $z_i = (x_i - \mu_1)/\sigma_1$ and $e_i = y_i - \mu_2 - \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_i - \mu_1)$, $1 \le i \le n$; $\rho^2 < 1$.

The likelihood equations for estimating μ_1, σ_1, μ_2, σ_2 and ρ are

$$\frac{\partial \ln L}{\partial \mu_1} = -\frac{(p-1)}{\sigma_1}\sum_{i=1}^{n} z_i^{-1} + \frac{p}{\sigma_1}\sum_{i=1}^{n} z_i^{p-1} - \frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\sum_{i=1}^{n} e_i = 0 \qquad (2.1.1.4)$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} - \frac{(p-1)}{\sigma_1}\sum_{i=1}^{n} z_i z_i^{-1} + \frac{p}{\sigma_1}\sum_{i=1}^{n} z_i z_i^{p-1} - \frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0 \qquad (2.1.1.5)$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{1}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} e_i = 0 \qquad (2.1.1.6)$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^3(1-\rho^2)}\sum_{i=1}^{n} e_i^2 + \frac{\rho}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0 \text{ and} \qquad (2.1.1.7)$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2^2(1-\rho^2)^2}\sum_{i=1}^{n} e_i^2 + \frac{1}{\sigma_2(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0. \qquad (2.1.1.8)$$

Writing $\theta_1 = \rho\left(\frac{\sigma_2}{\sigma_1}\right)$, we also have the additional equation

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0. \qquad (2.1.1.9)$$

Due to the intractable nature of the first two equations, (2.1.1.4)-(2.1.1.8) have no explicit solutions. Solving them by iteration is indeed problematic as is, in general, true with likelihood equations (Tiku et. al., 1986; Tiku and Akkaya, 2003).

To derive the MML estimators for this situation, let

$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)} \qquad (2.1.1.10)$$

be the order statistics of the random sample $x_1, x_2, ..., x_n$. Let $y_{[i]}$ be the y_j observation which corresponds to $x_{(i)}$; $y_{[i]}$ may be called concomitant of $x_{(i)}$. The sample observations are now denoted by $(x_{(i)}, y_{[i]})$, $1 \le i \le n$. Write

$$z_{(i)} = (x_{(i)} - \mu_1)/\sigma_1 \text{ and } e_{[i]} = y_{[i]} - \mu_2 - \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_{(i)} - \mu_1),\ 1 \le i \le n. \qquad (2.1.1.11)$$

The fact that complete sums are invariant to ordering implies that

$$\sum_{i=1}^{n} e_{[i]} = 0 \text{ and } \sum_{i=1}^{n} z_{(i)} e_{[i]} = 0 \qquad (2.1.1.12)$$

from (2.1.1.6) and (2.1.1.9). Thus, the equations (2.1.1.4)-(2.1.1.8) reduce to $(p > 1)$

$$\frac{\partial \ln L}{\partial \mu_1} = -\frac{(p-1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)}^{-1} + \frac{p}{\sigma_1}\sum_{i=1}^{n} z_{(i)}^{p-1} = 0$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} - \frac{(p-1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)} z_{(i)}^{-1} + \frac{p}{\sigma_1}\sum_{i=1}^{n} z_{(i)} z_{(i)}^{p-1} = 0$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{1}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} e_{[i]} = 0$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^3(1-\rho^2)}\sum_{i=1}^{n} e_{[i]}^2 = 0 \text{ and}$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2^2(1-\rho^2)^2}\sum_{i=1}^{n} e_{[i]}^2 = 0. \qquad (2.1.1.13)$$

Solving these equations is problematic because of the functions $z_{(i)}^{-1}$ and $z_{(i)}^{p-1}$. To alleviate this difficulty, we linearize these functions by using the first two terms of Taylor series expansions around $E(z_{(i)}) = t_{(i)}$ as follows:

$$z_{(i)}^{-1} \cong t_{(i)}^{-1} + (z_{(i)} - t_{(i)})\left(\frac{dz_{(i)}^{-1}}{dz}\right)_{z_{(i)}=t_{(i)}}$$

$$= \alpha_{i0} - \beta_{i0} z_{(i)}, \quad 1 \le i \le n. \qquad (2.1.1.14)$$

where $\alpha_{i0} = 2t_{(i)}^{-1}$ and $\beta_{i0} = t_{(i)}^{-2}$.

And

$$z_{(i)}^{p-1} \cong t_{(i)}^{p-1} + (z_{(i)} - t_{(i)})\left(\frac{dz_{(i)}^{p-1}}{dz}\right)_{z_{(i)}=t_{(i)}}$$

$$= \alpha_i + \beta_i z_{(i)}, \quad 1 \le i \le n. \qquad (2.1.1.15)$$

where $\alpha_i = (2-p)t_{(i)}^{p-1}$ and $\beta_i = (p-1)t_{(i)}^{p-2}$.

Remark: We assume that $p > 1$ which is true in most engineering and biomedical applications, since $p > 1$ represents increasing failure rate $r(t) = f(t)/\big(1 - F(t)\big)$.

Substituting (2.1.1.14) and (2.1.1.15) in (2.1.1.13), we obtain the following modified likelihood equations:

$$\frac{\partial \ln L}{\partial \mu_1} \cong \frac{\partial \ln L^*}{\partial \mu_1} = -\frac{(p-1)}{\sigma_1}\sum_{i=1}^{n}\big(\alpha_{i0} - \beta_{i0} z_{(i)}\big) + \frac{p}{\sigma_1}\sum_{i=1}^{n}\big(\alpha_i + \beta_i z_{(i)}\big) = 0 \qquad (2.1.1.16)$$

$$\frac{\partial \ln L}{\partial \sigma_1} \cong \frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1} - \frac{(p-1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)}\big(\alpha_{i0} - \beta_{i0} z_{(i)}\big)$$

$$+ \frac{p}{\sigma_1}\sum_{i=1}^{n} z_{(i)}\big(\alpha_i + \beta_i z_{(i)}\big) = 0 \text{ and} \qquad (2.1.1.17)$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{\partial \ln L^*}{\partial \mu_2} = 0\,,\ \frac{\partial \ln L}{\partial \sigma_2} = \frac{\partial \ln L^*}{\partial \sigma_2} = 0\,,\ \frac{\partial \ln L}{\partial \rho} = \frac{\partial \ln L^*}{\partial \rho} = 0\,. \qquad (2.1.1.18)$$

The MML estimators are the solutions of the equations (2.1.1.16)-(2.1.1.18):

$$\hat{\mu}_1 = \hat{\mu}_n - \frac{\Delta}{m}\hat{\sigma}_1 \qquad (2.1.1.19)$$

$$\hat{\sigma}_1 = \frac{-B + \sqrt{B^2 + 4nC}}{2n} \qquad (2.1.1.20)$$

$$\hat{\mu}_2 = \bar{y} - \hat{\rho}\left(\frac{\hat{\sigma}_2}{\hat{\sigma}_1}\right)(\bar{x} - \hat{\mu}_1) \qquad (2.1.1.21)$$

$$\hat{\sigma}_2 = \left(s_y^2 + \frac{s_{xy}^2}{s_x^2}\left(\frac{\hat{\sigma}_1^{\,2}}{s_x^2} - 1\right)\right)^{1/2} \text{ and} \qquad (2.1.1.22)$$

$$\hat{\rho} = \frac{s_{xy}}{s_x^2}\frac{\hat{\sigma}_1}{\hat{\sigma}_2}\,. \qquad (2.1.1.23)$$

Additionally,

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)(y_i - \hat{\mu}_2)}{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)^2}. \qquad (2.1.1.24)$$

As usual,

$$n\bar{x} = \sum_{i=1}^{n} x_i \ , \ n\bar{y} = \sum_{i=1}^{n} y_i \ , \ (n-1)s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \ ,$$

$$(n-1)s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 \ , \ (n-1)s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \text{, and}$$

$$\hat{\mu}_n = \frac{1}{m}\sum_{i=1}^{n}\delta_i x_{(i)} \ , \ \ \delta_i = (p-1)\beta_{i0} + p\beta_i \ , \ \ m = \sum_{i=1}^{n}\delta_i \ ,$$

$$\Delta_i = (p-1)\alpha_{i0} - p\alpha_i \ , \ \ \Delta = \sum_{i=1}^{n}\Delta_i \ ,$$

$$B = \sum_{i=1}^{n}\Delta_i (x_{(i)} - \hat{\mu}_n) \text{ and } C = \sum_{i=1}^{n}\delta_i (x_{(i)} - \hat{\mu}_n)^2 = \sum_{i=1}^{n}\delta_i x_{(i)}^2 - m\hat{\mu}_n^2 . \qquad (2.1.1.25)$$

Lemma 1: The estimator $\hat{\sigma}_2$ is always positive. This follows from the fact that $s_{xy}^2 \le s_x^2 s_y^2$ so that $s_y^2 - (s_{xy}^2 / s_x^2) \ge s_y^2 - s_y^2 = 0$. Since $s_{xy}^2 \hat{\sigma}_1^2 \big/ s_x^2$ is always positive, the result follows.

Lemma 2: The estimator $\hat{\rho}^2$ always assumes values between 0 and 1. This follows from the fact that

$$\hat{\rho}^2 = \frac{1}{\left[1 + \left(s_x^4 s_y^2 \big/ s_{xy}^2 \hat{\sigma}_1^2\right)\left(1 - s_{xy}^2 \big/ s_x^2 s_y^2\right)\right]} \text{ and } 0 \le s_{xy}^2 \le s_x^2 s_y^2 .$$

2.1.2 Conditional and Marginal Likelihood Functions

We now separate out the likelihood function into two parts as conditional and marginal density functions and consider a reparametrization in the conditional part. If we let $w_i = y_i - \theta_1 x_i$, $\mu_{2.1} = \mu_2 - \theta_1 \mu_1$ and $\sigma_{2.1}^2 = \sigma_2^2(1-\rho^2)$, we have

$$L_{y|x} \propto \sigma_{2.1}^{-n} \exp\left(-\frac{1}{2\sigma_{2.1}^2} \sum_{i=1}^{n} (w_i - \mu_{2.1})^2 \right) . \tag{2.1.2.1}$$

It is important to realize that e_i is distributed as normal $N(0, \sigma_{2.1}^2)$ and w_i is distributed as normal $N(\mu_{2.1}, \sigma_{2.1}^2)$.

Realize that $e_i = (w_i - \mu_{2.1}) = y_i - \mu_2 - \theta_1(x_i - \mu_1)$, $1 \le i \le n$. Since $L = L_x L_{y|x}$,

$$\ln L = n \ln p - n \ln \sigma_1 + (p-1) \sum_{i=1}^{n} \ln z_i - \sum_{i=1}^{n} z_i^p$$
$$- \frac{n}{2} \ln 2\pi - n \ln \sigma_{2.1} - \frac{1}{2\sigma_{2.1}^2} \sum_{i=1}^{n} e_i^2 . \tag{2.1.2.2}$$

Here, the first part represents the loglikelihood of the marginal distribution and the second part represents the log likelihood of the conditional distribution.

The likelihood equations for estimating μ_1, σ_1, $\mu_{2.1}$, $\sigma_{2.1}$ and θ_1 are

$$\frac{\partial \ln L}{\partial \mu_1} = -\frac{(p-1)}{\sigma_1} \sum_{i=1}^{n} z_i^{-1} + \frac{p}{\sigma_1} \sum_{i=1}^{n} z_i^{p-1} = 0 \tag{2.1.2.3}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} - \frac{(p-1)}{\sigma_1} \sum_{i=1}^{n} z_i z_i^{-1} + \frac{p}{\sigma_1} \sum_{i=1}^{n} z_i z_i^{p-1} = 0 \tag{2.1.2.4}$$

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{1}{\sigma_{2.1}} \sum_{i=1}^{n} e_i = 0 \tag{2.1.2.5}$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^3} \sum_{i=1}^{n} e_i^2 = 0 \quad \text{and} \tag{2.1.2.6}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}^2} \sum_{i=1}^{n} z_i e_i = 0 . \tag{2.1.2.7}$$

Since (2.1.2.3)-(2.1.2.7) have no explicit solutions, we again derive MML estimators. We first order x_i's in the increasing order

$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)} \tag{2.1.2.8}$$

Let $y_{[i]}$ be the y_j observation which corresponds to $x_{(i)}$. The sample observations are now denoted by $(x_{(i)}, y_{[i]})$, $1 \le i \le n$. We now define

$$z_{(i)} = (x_{(i)} - \mu_1)/\sigma_1, \; w_{[i]} = y_{[i]} - \theta_1 x_{(i)}, \text{ and}$$
$$e_{[i]} = y_{[i]} - \mu_2 - \theta_1 (x_{(i)} - \mu_1), \; 1 \le i \le n . \tag{2.1.2.9}$$

Realize that ordering of $z_{(i)}$ is invariant to μ_1 and $\sigma_1 (> 0)$. That is why $z_{(i)}$ corresponds to $x_{(i)}$ $(1 \le i \le n)$.

Since complete sums are invariant to ordering, we have

$$\frac{\partial \ln L}{\partial \mu_1} = -\frac{(p-1)}{\sigma_1} \sum_{i=1}^{n} z_{(i)}^{-1} + \frac{p}{\sigma_1} \sum_{i=1}^{n} z_{(i)}^{p-1} = 0$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} - \frac{(p-1)}{\sigma_1} \sum_{i=1}^{n} z_{(i)} z_{(i)}^{-1} + \frac{p}{\sigma_1} \sum_{i=1}^{n} z_{(i)} z_{(i)}^{p-1} = 0$$

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{1}{\sigma_{2.1}} \sum_{i=1}^{n} e_{[i]} = 0$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^3}\sum_{i=1}^{n} e_{[i]}^2 = 0 \quad \text{and}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}^2}\sum_{i=1}^{n} z_{(i)} e_{[i]} = 0\,. \tag{2.1.2.10}$$

Replacing $z_{(i)}^{-1}$ by $\alpha_{i0} - \beta_{i0} z_{(i)}$ and $z_{(i)}^{p-1}$ by $\alpha_i + \beta_i z_{(i)}$ gives the modified likelihood equations below (the coefficients α_{i0}, β_{i0}, α_i and β_i are the same as in (2.1.1.14) and (2.1.1.15)):

$$\frac{\partial \ln L}{\partial \mu_1} \cong \frac{\partial \ln L^*}{\partial \mu_1} = -\frac{(p-1)}{\sigma_1}\sum_{i=1}^{n}\left(\alpha_{i0} - \beta_{i0} z_{(i)}\right) + \frac{p}{\sigma_1}\sum_{i=1}^{n}\left(\alpha_i + \beta_i z_{(i)}\right) = 0 \tag{2.1.2.11}$$

$$\frac{\partial \ln L}{\partial \sigma_1} \cong \frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1} - \frac{(p-1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)}\left(\alpha_{i0} - \beta_{i0} z_{(i)}\right)$$

$$+ \frac{p}{\sigma_1}\sum_{i=1}^{n} z_{(i)}\left(\alpha_i + \beta_i z_{(i)}\right) = 0 \quad \text{and} \tag{2.1.2.12}$$

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{\partial \ln L^*}{\partial \mu_{2.1}} = 0\,,\ \frac{\partial \ln L}{\partial \sigma_{2.1}} = \frac{\partial \ln L^*}{\partial \sigma_{2.1}} = 0\,,\ \frac{\partial \ln L}{\partial \theta_1} = \frac{\partial \ln L^*}{\partial \theta_1} = 0\,. \tag{2.1.2.13}$$

The MML estimators are the solutions of the equations (2.1.2.11)-(2.1.2.13):

$$\hat{\mu}_1 = \hat{\mu}_n - \frac{\Delta}{m}\hat{\sigma}_1 \tag{2.1.2.14}$$

$$\hat{\sigma}_1 = \frac{-B + \sqrt{B^2 + 4nC}}{2n} \tag{2.1.2.15}$$

$$\hat{\mu}_{2.1} = \frac{1}{n}\sum_{i=1}^{n} w_{[i]} = \bar{y} - \hat{\theta}_1 \bar{x} \tag{2.1.2.16}$$

$$\hat{\sigma}_{2.1} = \frac{1}{(n-2)}\sum_{i=1}^{n}(w_{[i]} - \hat{\mu}_{2.1})^2 = \frac{1}{(n-2)}\sum_{i=1}^{n}(y_{[i]} - \bar{y} - \hat{\theta}_1(x_{(i)} - \bar{x}))^2 \quad \text{and} \tag{2.1.2.17}$$

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)(y_i - \hat{\mu}_2)}{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)^2}; \qquad (2.1.2.18)$$

$\hat{\mu}_n$, Δ, m, B, C and related terms are given in (2.1.1.25). It is interesting to note that $\hat{\theta}_1$ in (2.1.2.18) reduces to s_{xy} / s_x^2.

Remark: The MML estimators (2.1.1.19)-(2.1.1.23) are very different than those based on bivariate normality. However, the conditional estimators $\hat{\mu}_{2.1}$, $\hat{\sigma}_{2.1}$ and $\hat{\theta}_1$ are the same as the LSE. This is due to the fact that e_i's in the linear model $y_i = \mu_{2.1} + \theta_1 x_i + e_i$ $(1 \le i \le n)$, are assumed to be iid normal $N(0, \sigma^2)$.

2.1.3 Properties of the MML Estimators

Lemma 2.1: The asymptotic distribution of $\hat{\mu}_1$ is conditionally (for known σ_1) normal with mean μ_1 and variance σ_1^2 / m.

The result follows for the fact that for known σ_1, $\hat{\mu}_1 = \hat{\mu}_1(\sigma_1) = K + D\sigma_1$ and

$$\frac{\partial \ln L^*}{\partial \mu_1} = \frac{m}{\sigma_1^2}\{\hat{\mu}_1(\sigma_1) - \mu_1\}. \qquad (2.1.3.1)$$

Lemma 2.2: Asymptotically, the estimator $\hat{\sigma}_1$ is conditionally (for known μ_1) the MVB estimator of σ_1.

For known μ_1, $\hat{\sigma}_1 = \hat{\sigma}_1(\mu_1)$ is given by

$$\hat{\sigma}_1(\mu_1) = \frac{\left(-B_0 + \sqrt{B_0^2 + 4nC_0}\right)}{2n} \tag{2.1.3.2}$$

where

$$B_0 = \sum_{i=1}^{n} \Delta_i (x_{(i)} - \mu_1) \text{ and } C_0 = \sum_{i=1}^{n} \delta_i (x_{(i)} - \mu_1)^2 . \tag{2.1.3.3}$$

Further,

$$\frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1^3}\left(\frac{-B_0 + \sqrt{B_0^2 + 4nC_0}}{2n} - \sigma_1\right)\left(\frac{-B_0 - \sqrt{B_0^2 + 4nC_0}}{2n} - \sigma_1\right). \tag{2.1.3.4}$$

The only admissible solution of (2.1.3.4) is $\sigma_1 = \hat{\sigma}_1(\mu_1)$ and the result follows as before.

Corollary: Since for large n, B_0 is very small as compared to $\sqrt{(nC_0)}$, $B_0 \big/ \sqrt{nC_0}$ is negligibly small. Consequently, $\frac{\partial \ln L^*}{\partial \sigma_1}$ assumes the form

$$\frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1^3}\left[\left(\sqrt{\frac{C_0}{n}}\left(-\frac{1}{2}\left(\frac{B_0}{\sqrt{nC_0}}\right) + \sqrt{1 + \frac{1}{4}\left(\frac{B_0}{\sqrt{nC_0}}\right)^2}\right) - \sigma_1\right)\times\right.$$
$$\left.\left(\sqrt{\frac{C_0}{n}}\left(-\frac{1}{2}\left(\frac{B_0}{\sqrt{nC_0}}\right) - \sqrt{1 + \frac{1}{4}\left(\frac{B_0}{\sqrt{nC_0}}\right)^2}\right) - \sigma_1\right)\right], \tag{2.1.3.5}$$

$$\frac{\partial \ln L^*}{\partial \sigma_1} \cong -\frac{n}{\sigma_1^3}\left(\left(\sqrt{\frac{C_0}{n}}-\sigma_1\right)\left(-\sqrt{\frac{C_0}{n}}-\sigma_1\right)\right), \tag{2.1.3.6}$$

$$\frac{\partial \ln L^*}{\partial \sigma_1} \cong \frac{n}{\sigma_1^3}\left(\frac{C_0}{n}-\sigma_1^2\right). \tag{2.1.3.7}$$

Since $\frac{\partial \ln L^*}{\partial \sigma_1} = \frac{\partial \ln L^*}{\partial \sigma_1^2}\frac{\partial \sigma_1^2}{\partial \sigma_1}$ and $\frac{\partial \sigma_1^2}{\partial \sigma_1} = 2\sigma_1$, $\hat{\sigma}_1^2(\mu_1) \cong C_0/n$ is asymptotically the MVB of σ_1^2 with variance $\frac{2\sigma_1^4}{n}$, C_0/σ_1^2 is for large n distributed as chi-square with n degrees of freedom; see Kendall and Stuart (1969).

Bias Correction: For small $n\,(\leq 15)$, $\hat{\mu}_1$ and $\hat{\sigma}_1$ have some bias. The bias corrected estimators are given by (see Appendix B for details)

$$\hat{\mu}_1 = \hat{\mu}_n - \frac{1}{m}\left(\sum_{i=1}^{n}\delta_i t_{(i)}\right)\hat{\sigma}_1 \quad \text{with} \quad \hat{\sigma}_1 = \frac{\left\{B+\sqrt{B^2+4nC}\right\}}{2\sqrt{n(n-1)}}. \tag{2.1.3.8}$$

Lemma 2.3: For known μ_1 and μ_2, $\hat{\theta}_1(\mu_1,\mu_2)$ is the MVB estimator of θ_1 and is normally distributed with mean θ_1 and variance $\sigma_2^2(1-\rho^2)\Big/\sum_{i=1}^{n}(x_i-\mu_1)^2$.

This follows from the fact that

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\partial \ln L^*}{\partial \theta_1} = \frac{\sum_{i=1}^{n}(x_i-\mu_1)^2}{\sigma_2^2(1-\rho^2)}\left(\hat{\theta}_1(\mu_1,\mu_2)-\theta_1\right). \tag{2.1.3.9}$$

2.1.4 Asymptotic Covariance Matrix

The asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1$, $\hat{\mu}_2$, $\hat{\sigma}_2$ and $\hat{\rho}$ is given by $I^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$, where I is the Fisher information matrix. The Fisher information matrix is given by $(i, j = 1,2,3,4,5)$

$$\boldsymbol{I} = \left[I_{ij}\right] = \left[-E\left(\frac{\partial^2 \ln L}{\partial \tau_i \partial \tau_j} \right) \right],\ \tau_1 = \mu_1,\ \tau_2 = \sigma_1,\ \tau_3 = \mu_2,\ \tau_4 = \sigma_2,\ \tau_5 = \rho. \quad (2.1.4.1)$$

If we let $\boldsymbol{I} = \frac{n}{(1-\rho^2)}\ \boldsymbol{A}$, the elements of the matrix $\boldsymbol{A}$ are

$$A_{11} = \frac{1}{\sigma_1^2}\left[(p-1)^2(1-\rho^2)\Gamma(1-2/p) + \rho^2\right],$$

$$A_{12} = \frac{1}{\sigma_1^2}\left[p^2(1-\rho^2)\Gamma(2-1/p) + \rho^2\Gamma(1+1/p)\right],$$

$$A_{13} = -\frac{\rho}{\sigma_1\sigma_2},\ A_{14} = -\frac{\rho^2}{\sigma_1\sigma_2}\Gamma(1+1/p),\ A_{15} = -\frac{\rho}{\sigma_1}\Gamma(1+1/p),$$

$$A_{22} = \frac{1}{\sigma_1^2}\left[p^2(1-\rho^2) + \rho^2\Gamma(1+2/p)\right],\ A_{23} = -\frac{\rho}{\sigma_1\sigma_2}\Gamma(1+1/p),$$

$$A_{24} = -\frac{\rho^2}{\sigma_1\sigma_2}\Gamma(1+2/p),\ A_{25} = -\frac{\rho}{\sigma_1}\Gamma(1+2/p),\ A_{33} = \frac{1}{\sigma_2^2},$$

$$A_{34} = \frac{\rho}{\sigma_2^2}\Gamma(1+1/p),\ A_{35} = \frac{1}{\sigma_2}\Gamma(1+1/p),$$

$$A_{44} = \frac{1}{\sigma_2^2}\left[2(1-\rho^2) + \rho^2\Gamma(1+2/p)\right],\ A_{45} = \frac{\rho}{\sigma_2}\left[-2 + \Gamma(1+2/p)\right]$$

and

$$A_{55} = \frac{2\rho^2}{(1-\rho^2)} + \Gamma(1+2/p). \quad (2.1.4.2)$$

Here, the asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1, \hat{\mu}_2, \hat{\sigma}_2$ and $\hat{\rho}$ is given by $\boldsymbol{\Sigma} \equiv I^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$. See Appendix D for details.

Hypothesis Testing: Our major interest is in testing the null hypothesis $H_0 : \rho = 0$. Since the MML estimators are asymptotically equivalent to the ML estimators, the likelihood function L is maximized (asymptotically) by the MML estimators. Thus, the likelihood ratio statistic is (asymptotically)

$$\hat{\lambda} = \frac{\max(L \mid H_0)}{\max(L \mid H_1)}$$

$$= \left(\frac{\hat{\sigma}_2^2}{s_y^2} \right)^{n/2} (1 - \hat{\rho}^2)^{n/2} \exp\left[\frac{(n-1)s_y^2}{2(1-\hat{\rho}^2)\hat{\sigma}_2^2}(1 - \hat{\rho}_0^2) - \frac{(n-1)}{2} \right] \qquad (2.1.4.3)$$

where $\hat{\rho}_0 = \frac{s_{xy}}{s_x s_y}$ is the usual Pearson sample correlation coefficient. For the same reasons as before, the likelihood ratio is a monotonic function of $\hat{\rho}^2$. Thus, to test H_0 against $H_1 : \rho < 0$ (or $\rho > 0$), the test based on $\hat{\rho}$ is uniformly most powerful (asymptotically). We propose the statistic

$$W = \hat{\rho} \Bigg/ \sqrt{\frac{1}{n\left[\Gamma\left(1 + \frac{2}{p}\right) - \Gamma^2\left(1 + \frac{1}{p}\right) \right]}}$$

$$= \hat{\rho} \sqrt{n\left[\Gamma\left(1 + \frac{2}{p}\right) - \Gamma^2\left(1 + \frac{1}{p}\right) \right]}; \qquad (2.1.4.4)$$

$\frac{1}{n\left[\Gamma\left(1 + \frac{2}{p}\right) - \Gamma^2\left(1 + \frac{1}{p}\right) \right]}$ is the asymptotic variance of $\hat{\rho}$ under H_0, obtained from

$$\left\{I^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)\right\}_{\rho=0} \ . \tag{2.1.4.5}$$

The null distribution of W is referred to normal $N(0,1)$. Small values of W lead to rejection of H_0 against $H_1 : \rho < 0$.

Also, testing the mean vector

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.1.4.6}$$

is of enormous practical interest. Since $\hat{\mu}_1$ and $\hat{\mu}_2$ are asymptotically equivalent to the ML estimators, the distribution of the random vector $\sqrt{n}(\hat{\mu}_1,\hat{\mu}_2)$ is bivariate normal with zero mean vector and (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n\begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{13} \\ \hat{\sigma}_{13} & \hat{\sigma}_{33} \end{bmatrix} \tag{2.1.4.7}$$

where, $\hat{\sigma}_{11}$, $\hat{\sigma}_{13}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij} = \mathbf{\Sigma_{ij}} = I_{ij}^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)$. Since in these elements $\hat{\sigma}_1$ and $\hat{\sigma}_2$ converge to σ_1 and σ_2, respectively, the asymptotic null distribution of

$$\hat{\mathrm{T}}^2 = n(\hat{\mu}_1,\hat{\mu}_2)\hat{\mathbf{\Omega}}^{-1}\begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \tag{2.1.4.8}$$

is chi-square with 2 df. We reject H_0 at the 5% significance level if the value of $\hat{\mathrm{T}}^2$ is greater than $\chi^2_{0.05}(2)$. The nonnull distribution (asymptotic) of $\hat{\mathrm{T}}^2$ is noncentral chi-square with 2 df and noncentrality parameter

$$\lambda^2 = n(\mu_1, \mu_2)\boldsymbol{\Omega}^{-1}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}. \qquad (2.1.4.9)$$

For small n, the null distribution of $\frac{(n-2)}{2(n-1)}\hat{T}^2$ is approximately central-F with $(2, n-2)$ degrees of freedom and the nonnull distribution is approximately noncentral-F with $(2, n-2)$ degrees of freedom and noncentrality parameter λ^2. In Section 2.1.2 we expressed the likelihood function as the product of the marginal and the conditional likelihood functions and used a reparametrization. We now define the Fisher information matrix, $I(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$ for estimating $\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}$ and θ_1 instead of $\mu_1, \sigma_1, \mu_2, \sigma_2$ and ρ. If we let $\boldsymbol{I} = n\ \boldsymbol{A}$, the elements of the matrix $\boldsymbol{A}$ are

$$A_{11} = \frac{(p-1)^2}{\sigma_1^2}\Gamma(1-2/p)\ , A_{12} = \frac{p^2}{\sigma_1^2}\Gamma(2-1/p), A_{13} = 0,\ \ A_{14} = 0,\ A_{15} = 0,$$

$$A_{22} = \frac{p^2}{\sigma_1^2},\ A_{23} = 0,\ A_{24} = 0, A_{25} = 0,\ A_{33} = \frac{1}{\sigma_{2.1}^2},\ A_{34} = 0,$$

$$A_{35} = -\frac{\sigma_1}{\sigma_{2.1}^2}\Gamma(1+1/p), A_{44} = \frac{2}{\sigma_{2.1}^2},$$

$$A_{45} = 0 \text{ and } A_{55} = -\frac{\sigma_1^2}{\sigma_{2.1}^2}\Gamma(1+2/p). \qquad (2.1.4.10)$$

Here the asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1, \hat{\mu}_{2.1}, \hat{\sigma}_{2.1}$ and $\hat{\theta}_1$ is given by $\boldsymbol{\Sigma} \equiv I^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$. See Appendix D for the details of the derivation of the Fisher information matrix.

Hypothesis Testing: Now, we want to test the mean vector

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_{2.1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \qquad (2.1.4.11)$$

Since $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$ are asymptotically equivalent to the ML estimators, the distribution of the random vector $\sqrt{n}(\hat{\mu}_1, \hat{\mu}_{2.1})$ is bivariate normal with zero mean vector and (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n \begin{bmatrix} \hat{\sigma}_{11} & 0 \\ 0 & \hat{\sigma}_{33} \end{bmatrix} \qquad (2.1.4.12)$$

where $\hat{\sigma}_{11}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij} = \mathbf{\Sigma_{ij}} = I_{ij}^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$. The covariance between $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$ is zero since they are orthogonal components, so there is no need to estimate it. Since in these elements $\hat{\sigma}_1$ and $\hat{\sigma}_{2.1}$ converge to σ_1 and $\sigma_{2.1}$, respectively, the asymptotic null distribution of

$$\hat{\mathrm{T}}_1^2 = n(\hat{\mu}_1, \hat{\mu}_{2.1}) \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_{2.1} \end{pmatrix} \qquad (2.1.4.13)$$

is chi-square with 2 df. Since the inverse of the estimated covariance matrix is simply

$$\hat{\mathbf{\Omega}}^{-1} = \frac{1}{n} \begin{bmatrix} \hat{\sigma}_{11}^{-1} & 0 \\ 0 & \hat{\sigma}_{33}^{-1} \end{bmatrix}, \qquad (2.1.4.14)$$

our test statistics $\hat{T}_1^2$ is equal to

$$\hat{T}_1^2 = \hat{\mu}_1{}^2 / \hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2 / \hat{\sigma}_{33} \,. \tag{2.1.4.15}$$

We reject H_0 at the 5% significance level if the value of $\hat{T}_1^2$ is greater than $\chi^2_{0.05}(2)$. The nonnull distribution (asymptotic) of $\hat{T}_1^2$ is noncentral chi-square with 2 df and noncentrality parameter

$$\begin{aligned} \lambda^2 &= n(\mu_1, \mu_{2.1}) \boldsymbol{\Omega}^{-1} \begin{pmatrix} \mu_1 \\ \mu_{2.1} \end{pmatrix} \\ &= \mu_1{}^2 / \sigma_{11} + \mu_{2.1}{}^2 / \sigma_{33} . \end{aligned} \tag{2.1.4.16}$$

For small n, the null distribution of $\frac{(n-2)}{2(n-1)} \hat{T}_1^2$ is approximately central-F with $(2, n-2)$ degrees of freedom and the nonnull distribution is approximately noncentral-F with $(2, n-2)$ degrees of freedom and noncentrality parameter λ^2.

2.2 Marginal Generalized Logistic and Conditional Normal

We again take the conditional distribution as normal but the marginal distribution as Generalized Logistic. We derive the estimators and the hypothesis testing procedure for this situation.

2.2.1 Estimation of Parameters

Generalized Logistic distribution is flexible in nature. It is negatively skewed if the shape parameter b is less than 1 and positively skewed if b is greater than 1. It is symmetric if b is equal to 1 in which case it is the well known logistic

distribution. Generalized Logistic distribution is also preferred since its range is between minus infinity and plus infinity.

Suppose that the marginal distribution of X is the Generalized Logistic distribution with density

$$g(x) = \frac{b}{\sigma_1} \frac{e^{-\left(\frac{x-\mu_1}{\sigma_1}\right)}}{\left(1+e^{-\left(\frac{x-\mu_1}{\sigma_1}\right)}\right)^{b+1}}, \qquad (2.2.1.1)$$

and the conditional distribution of Y given $X = x$ is the normal distribution with density

$$h(y \mid x) = \frac{1}{\sqrt{2\pi}\sigma_2(1-\rho^2)^{1/2}} \times \exp\left[-\frac{1}{2\sigma_2^2(1-\rho^2)} \left\{ y - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) \right\}^2 \right] \qquad (2.2.1.2)$$

where $-\infty < x < \infty$, $-\infty \langle y \langle \infty$, $\mu_1, \mu_2 \in \Re$, $\sigma_1, \sigma_2 \rangle 0$, $-1 \langle \rho \langle 1$, $b > 0$.

Given a random sample (x_i, y_i) $(1 \le i \le n)$, the likelihood function is

$$L \propto \sigma_1^{-n}\sigma_2^{-n}(1-\rho^2)^{-n/2} \frac{1}{\prod_{i=1}^{n}\left(1+e^{-z_i}\right)^{b+1}} \exp\left(-\sum_{i=1}^{n} z_i - \frac{1}{2\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} e_i^2 \right) \qquad (2.2.1.3)$$

where $z_i = (x_i - \mu_1)/\sigma_1$ and $e_i = y_i - \mu_2 - \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_i - \mu_1)$, $1 \le i \le n$; $\rho^2 < 1$.

The loglikelihood function is

$$\ln L = n\ln b - n\ln\sigma_1 - \sum_{i=1}^{n} z_i - (b+1)\sum_{i=1}^{n}\ln\left(1+e^{-z_i}\right)$$

$$-\frac{n}{2}\ln(2\pi) - n\ln\sigma_2 - \frac{n}{2}\ln(1-\rho^2) - \frac{1}{2\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} e_i^2 \,. \qquad (2.2.1.4)$$

The likelihood equations expressed in terms of $x_{(i)}$ and the corresponding concomitants $y_{[i]}$ $(1 \le i \le n)$ are

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b+1)}{\sigma_1}\sum_{i=1}^{n}\frac{e^{-z_{(i)}}}{\left(1+e^{-z_{(i)}}\right)} = 0$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)}\frac{e^{-z_{(i)}}}{\left(1+e^{-z_{(i)}}\right)} = 0$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{1}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} e_{[i]} = 0$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^3(1-\rho^2)}\sum_{i=1}^{n} e_{[i]}^2 = 0 \quad \text{and}$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2^2(1-\rho^2)^2}\sum_{i=1}^{n} e_{[i]}^2 = 0 \,, \qquad (2.2.1.5)$$

where $z_{(i)} = (x_{(i)} - \mu_1)/\sigma_1$ and $e_{[i]} = y_{[i]} - \mu_2 - \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_{(i)} - \mu_1)$, $1 \le i \le n$.

Writing $\theta_1 = \rho\left(\frac{\sigma_2}{\sigma_1}\right)$, we also have

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0 \,. \qquad (2.2.1.6)$$

Solving these equations is problematic because of the function $g(z_{(i)}) = \frac{e^{-z_{(i)}}}{\left(1+e^{-z_{(i)}}\right)}$. We linearize this function by using the first two terms of a Taylor series expansion around $E(z_{(i)}) = t_{(i)}$ as follows.

$$\begin{aligned} g(z_{(i)}) &\cong g(t_{(i)}) + (z_{(i)} - t_{(i)})\left(\frac{dg(z_{(i)})}{dz}\right)_{z_{(i)}=t_{(i)}} \\ &= \frac{e^{-t_{(i)}}}{\left(1+e^{-t_{(i)}}\right)} + (z_{(i)} - t_{(i)})\left(-\frac{e^{-t_{(i)}}}{\left(1+e^{-t_{(i)}}\right)^2}\right) \\ &= \alpha_i - \beta_i z_{(i)}, \quad 1 \le i \le n. \end{aligned} \tag{2.2.1.7}$$

where $\alpha_i = \frac{e^{-t_{(i)}}}{\left(1+e^{-t_{(i)}}\right)^2}\left(1 + t_{(i)} + e^{-t_{(i)}}\right)$ and $\beta_i = \frac{e^{-t_{(i)}}}{\left(1+e^{-t_{(i)}}\right)^2}$.

Realize that β_i are positive for all $i = 1,2,...,n$.

Substituting (2.2.1.7) in (2.2.1.5), we obtain the following modified likelihood equations:

$$\frac{\partial \ln L}{\partial \mu_1} \cong \frac{\partial \ln L^*}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b+1)}{\sigma_1}\sum_{i=1}^{n}\left(\alpha_i - \beta_i z_{(i)}\right) = 0 \tag{2.2.1.8}$$

$$\frac{\partial \ln L}{\partial \sigma_1} \cong \frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)}\left(\alpha_i - \beta_i z_{(i)}\right) = 0 \quad \text{and} \tag{2.2.1.9}$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{\partial \ln L^*}{\partial \mu_2} = 0, \ \frac{\partial \ln L}{\partial \sigma_2} = \frac{\partial \ln L^*}{\partial \sigma_2} = 0, \ \frac{\partial \ln L}{\partial \rho} = \frac{\partial \ln L^*}{\partial \rho} = 0. \tag{2.2.1.10}$$

The MML estimators are the solutions of the equations (2.2.1.8)-(2.2.1.10):

$$\hat{\mu}_1 = K + D\hat{\sigma}_1 \tag{2.2.1.11}$$

$$\hat{\sigma}_1 = \frac{B + \sqrt{B^2 + 4nC}}{2n} \tag{2.2.1.12}$$

$$\hat{\mu}_2 = \bar{y} - \hat{\rho}\left(\frac{\hat{\sigma}_2}{\hat{\sigma}_1}\right)(\bar{x} - \hat{\mu}_1) \tag{2.2.1.13}$$

$$\hat{\sigma}_2 = \left(s_y^2 + \frac{s_{xy}^2}{s_x^2}\left(\frac{\hat{\sigma}_1^{\,2}}{s_x^2} - 1\right)\right)^{1/2} \quad \text{and} \tag{2.2.1.14}$$

$$\hat{\rho} = \frac{s_{xy}}{s_x^2}\frac{\hat{\sigma}_1}{\hat{\sigma}_2}\,. \tag{2.2.1.15}$$

Additionally,

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)(y_i - \hat{\mu}_2)}{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)^2} \tag{2.2.1.16}$$

Here, $n\bar{x} = \sum_{i=1}^{n} x_i$, $n\bar{y} = \sum_{i=1}^{n} y_i$, $(n-1)s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$, $(n-1)s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2$,

$(n-1)s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ and

$K = \frac{1}{m}\sum_{i=1}^{n}\beta_i x_{(i)}$, $D = \frac{1}{m}\sum_{i=1}^{n}\left(\frac{1}{b+1} - \alpha_i\right)$,

$m = \sum_{i=1}^{n}\beta_i$,

$$B = (b+1)\sum_{i=1}^{n}(x_{(i)} - K)\left(\frac{1}{b+1} - \alpha_i\right), \quad C = (b+1)\sum_{i=1}^{n}\beta_i(x_{(i)} - K)^2\,. \tag{2.2.1.17}$$

2.2.2 Conditional and Marginal Likelihood Functions

By separating the likelihood function as before, we obtain

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b+1)}{\sigma_1}\sum_{i=1}^{n}\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)} = 0 \qquad (2.2.2.1)$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_i - \frac{(b+1)}{\sigma_1}\sum_{i=1}^{n} z_i \frac{e^{-z_i}}{\left(1+e^{-z_i}\right)} = 0 \qquad (2.2.2.2)$$

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} e_i = 0 \qquad (2.2.2.3)$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^3}\sum_{i=1}^{n} e_i^2 = 0 \quad \text{and} \qquad (2.2.2.4)$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}^2}\sum_{i=1}^{n} z_i e_i = 0\,. \qquad (2.2.2.5)$$

The modified likelihood equations are

$$\frac{\partial \ln L}{\partial \mu_1} \cong \frac{\partial \ln L^*}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b+1)}{\sigma_1}\sum_{i=1}^{n}\left(\alpha_i - \beta_i z_{(i)}\right) = 0 \qquad (2.2.2.6)$$

$$\frac{\partial \ln L}{\partial \sigma_1} \cong \frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)}\left(\alpha_i - \beta_i z_{(i)}\right) = 0 \quad \text{and} \qquad (2.2.2.7)$$

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{\partial \ln L^*}{\partial \mu_{2.1}} = 0\,,\ \frac{\partial \ln L}{\partial \sigma_{2.1}} = \frac{\partial \ln L^*}{\partial \sigma_{2.1}} = 0\,,\ \frac{\partial \ln L}{\partial \theta_1} = \frac{\partial \ln L^*}{\partial \theta_1} = 0\,. \qquad (2.2.2.8)$$

The MML estimators are the solutions of the equations (2.2.2.6)-(2.2.2.8)

$$\hat{\mu}_1 = K + D\hat{\sigma}_1 \qquad (2.2.2.9)$$

$$\hat{\sigma}_1 = \frac{B + \sqrt{B^2 + 4nC}}{2n} \qquad (2.2.2.10)$$

$$\hat{\mu}_{2.1} = \frac{1}{n}\sum_{i=1}^{n} w_{[i]} = \bar{y} - \hat{\theta}_1 \bar{x} \qquad (2.2.2.11)$$

$$\hat{\sigma}_{2.1} = \frac{1}{(n-2)}\sum_{i=1}^{n}(w_{[i]} - \hat{\mu}_{2.1})^2 = \frac{1}{(n-2)}\sum_{i=1}^{n}(y_{[i]} - \bar{y} - \hat{\theta}_1(x_{(i)} - \bar{x}))^2 \qquad (2.2.2.12)$$

and

$$\hat{\theta}_1 = \frac{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)(y_i - \hat{\mu}_2)}{\sum_{i=1}^{n}(x_i - \hat{\mu}_1)^2} = \frac{s_{xy}}{s_x^2}; \qquad (2.2.2.13)$$

K, D, m, B, C and related terms are given in (2.2.1.17).

2.2.3 Properties of the MML Estimators

Lemma 2.4: The asymptotic distribution of $\hat{\mu}_1$ is conditionally (for known σ_1) normal with mean μ_1 and variance $\frac{\sigma_1^2}{m(b+1)}$.

For known σ_1, $\hat{\mu}_1 = \hat{\mu}_1(\sigma_1) = K + D\sigma_1$. Since

$$\frac{\partial \ln L^*}{\partial \mu_1} = \frac{m(b+1)}{\sigma_1^2}\{\hat{\mu}_1(\sigma_1) - \mu_1\}, \qquad (2.2.3.1)$$

the result follows.

Lemma 2.5: Asymptotically, the estimator $\hat{\sigma}_1$ is conditionally (for known μ_1) the MVB estimator of σ_1.

For known μ_1, $\hat{\sigma}_1 = \hat{\sigma}_1(\mu_1)$ is given by

$$\hat{\sigma}_1(\mu_1) = \frac{\left(B_0 + \sqrt{B_0^2 + 4nC_0}\right)}{2n} \qquad (2.2.3.2)$$

where

$$B_0 = (b+1)\sum_{i=1}^{n}(x_{(i)} - \mu_1)\left(\frac{1}{(b+1)} - \alpha_i\right) \text{ and } C_0 = (b+1)\sum_{i=1}^{n}\beta_i (x_{(i)} - \mu_1)^2 \quad (2.2.3.3)$$

Further,

$$\frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1^3}\left(\frac{B_0 + \sqrt{B_0^2 + 4nC_0}}{2n} - \sigma_1\right)\left(\frac{B_0 - \sqrt{B_0^2 + 4nC_0}}{2n} - \sigma_1\right). \quad (2.2.3.4)$$

The only admissible solution of (2.2.3.4) is $\sigma_1 = \hat{\sigma}_1(\mu_1)$ and the result follows as before.

Bias Correction: For small $n\ (\leq 15)$, $\hat{\mu}_1$ and $\hat{\sigma}_1$ have some bias. The bias corrected estimators are given by (see Appendix B for details)

$$\hat{\mu}_1 = K - \frac{\hat{\sigma}_1}{m}\left(\sum_{i=1}^{n}\beta_i t_{(i)}\right) \quad \text{with} \quad \hat{\sigma}_1 = \frac{\left\{B + \sqrt{B^2 + 4nC}\right\}}{2\sqrt{n(n-1)}}. \quad (2.2.3.5)$$

Lemma 2.6: For known μ_1 and μ_2, $\hat{\theta}_1(\mu_1, \mu_2)$ is the MVB estimator of θ_1 and is normally distributed with mean θ_1 and variance $\sigma_2^2(1-\rho^2)\Big/\sum_{i=1}^{n}(x_i - \mu_1)^2$.

This follows from the fact that

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\partial \ln L^*}{\partial \theta_1} = \frac{\sum_{i=1}^{n}(x_i - \mu_1)^2}{\sigma_2^2(1-\rho^2)}\left(\hat{\theta}_1(\mu_1, \mu_2) - \theta_1\right). \quad (2.2.3.6)$$

2.2.4 Asymptotic Covariance Matrix

The asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1$, $\hat{\mu}_2$, $\hat{\sigma}_2$ and $\hat{\rho}$ is given by $I^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$, where I is the Fisher information matrix. The Fisher information matrix is given by $(i, j = 1,2,3,4,5)$

$$\boldsymbol{I} = \left[I_{ij}\right] = \left[-E\left(\frac{\partial^2 \ln L}{\partial \tau_i \partial \tau_j} \right) \right],\ \tau_1 = \mu_1,\ \tau_2 = \sigma_1,\ \tau_3 = \mu_2,\ \tau_4 = \sigma_2,\ \tau_5 = \rho. \quad (2.2.4.1)$$

If we let $\boldsymbol{I} = \frac{n}{(1-\rho^2)}\ \boldsymbol{A}$, the elements of the matrix $\boldsymbol{A}$ are

$$A_{11} = \frac{1}{\sigma_1^2}\left[\frac{b(1-\rho^2)}{(b+2)} + \rho^2 \right],$$

$$A_{12} = \frac{1}{\sigma_1^2}\left[\frac{b(1-\rho^2)}{(b+2)}\left(\psi(b+1) - \psi(2)\right) + \rho^2\left(\psi(b) - \psi(1)\right) \right],$$

$$A_{13} = -\frac{\rho}{\sigma_1 \sigma_2},\ A_{14} = -\frac{\rho^2}{\sigma_1 \sigma_2}\left(\psi(b) - \psi(1)\right), A_{15} = -\frac{\rho}{\sigma_1}\left(\psi(b) - \psi(1)\right),$$

$$A_{22} = \frac{1}{\sigma_1^2}\left[1 + \frac{b}{(b+2)}\left\{ \psi'(b+1) + \psi'(2) + \left(\psi(b+1) - \psi(2)\right)^2 \right\} \right.$$

$$\left. + \rho^2\left(\psi'(b) + \psi'(1) + \left(\psi(b) - \psi(1)\right)^2 \right) \right],$$

$$A_{23} = -\frac{\rho}{\sigma_1 \sigma_2}\left(\psi(b) - \psi(1)\right), A_{24} = -\frac{\rho^2}{\sigma_1 \sigma_2}\left(\psi'(b) + \psi'(1) + \left(\psi(b) - \psi(1)\right)^2 \right),$$

$$A_{25} = -\frac{\rho}{\sigma_1}\left(\psi'(b) + \psi'(1) + \left(\psi(b) - \psi(1)\right)^2 \right),\ A_{33} = \frac{1}{\sigma_2^2},$$

$$A_{34} = \frac{\rho}{\sigma_2^2}\left(\psi(b) - \psi(1)\right),\ A_{35} = \frac{1}{\sigma_2}\left(\psi(b) - \psi(1)\right),$$

$$A_{44} = \frac{1}{\sigma_2^2}\left[2(1-\rho^2) + \rho^2\left(\psi'(b) + \psi'(1) + (\psi(b) - \psi(1))^2\right)\right],$$

$$A_{45} = \frac{\rho}{\sigma_2}\left[-2 + \left(\psi'(b) + \psi'(1) + (\psi(b) - \psi(1))^2\right)\right] \text{ and}$$

$$A_{55} = \frac{2\rho^2}{(1-\rho^2)} + \left(\psi'(b) + \psi'(1) + (\psi(b) - \psi(1))^2\right). \qquad (2.2.4.2)$$

Here, the asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1, \hat{\mu}_2, \hat{\sigma}_2$ and $\hat{\rho}$ is given by $\boldsymbol{\Sigma} \equiv I^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$. See Appendix D for details about the derivation of Fisher Information matrix and Appendix A for details about ψ (psi) functions.

Hypothesis Testing: Our major interest is testing the null hypothesis $H_0 : \rho = 0$. The likelihood ratio statistic obtained exactly along the same lines as before is

$$\hat{\lambda} = \frac{\max(L \mid H_0)}{\max(L \mid H_1)}$$

$$= \left(\frac{\hat{\sigma}_2^2}{s_y^2}\right)^{n/2} (1-\hat{\rho}^2)^{n/2} \exp\left[\frac{(n-1)s_y^2}{2(1-\hat{\rho}^2)\hat{\sigma}_2^2}(1-\hat{\rho}_0^2) - \frac{(n-1)}{2}\right] \qquad (2.2.4.3)$$

which gives the statistic

$$W = \hat{\rho} \Big/ \sqrt{\frac{1}{n\left(\psi'(b) + \psi'(1)\right)}}$$

$$= \hat{\rho}\sqrt{n\left(\psi'(b) + \psi'(1)\right)}; \qquad (2.2.4.4)$$

$\frac{1}{n\left(\psi'(b) + \psi'(1)\right)}$ is the asymptotic variance of $\hat{\rho}$ under H_0, obtained from

$$\left\{I^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)\right\}_{\rho=0}\ . \tag{2.2.4.5}$$

The null distribution of W is referred to normal $N(0,1)$. Small values of W lead to rejection of H_0 against $H_1:\rho<0$.

It is of great practical interest to test the null hypothesis

$$H_0:\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}. \tag{2.2.4.6}$$

Since $\hat{\mu}_1$ and $\hat{\mu}_2$ are asymptotically equivalent to the ML estimators, the distribution of the random vector $\sqrt{n}(\hat{\mu}_1,\hat{\mu}_2)$ is bivariate normal with zero mean vector and (estimated) covariance matrix

$$\hat{\mathbf{\Omega}}=n\begin{bmatrix}\hat{\sigma}_{11} & \hat{\sigma}_{13}\\ \hat{\sigma}_{13} & \hat{\sigma}_{33}\end{bmatrix} \tag{2.2.4.7}$$

where, $\hat{\sigma}_{11}$, $\hat{\sigma}_{13}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij}=\mathbf{\Sigma_{ij}}=I_{ij}^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)$. Since in these elements $\hat{\sigma}_1$ and $\hat{\sigma}_2$ converge to σ_1 and σ_2, respectively, the asymptotic null distribution of

$$\hat{\mathrm{T}}^2=n(\hat{\mu}_1,\hat{\mu}_2)\hat{\mathbf{\Omega}}^{-1}\begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{pmatrix} \tag{2.2.4.8}$$

is chi-square with 2 df. We reject H_0 at the 5% significance level if the value of $\hat{\mathrm{T}}^2$ is greater than $\chi^2_{0.05}(2)$. The nonnull distribution (asymptotic) of $\hat{\mathrm{T}}^2$ is noncentral chi-square with 2 df and noncentrality parameter

$$\lambda^2 = n(\mu_1, \mu_2)\,\mathbf{\Omega}^{-1}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}. \qquad (2.2.4.9)$$

For small n, the null distribution of $\frac{(n-2)}{2(n-1)}\hat{T}^2$ is approximately central-F with $(2, n-2)$ degrees of freedom and the nonnull distribution is approximately noncentral-F with $(2, n-2)$ degrees of freedom and noncentrality parameter λ^2. In Section 2.2.2 we expressed the likelihood as the product of the marginal and conditional likelihood functions and made a reparametrization.

Now we define the Fisher information matrix, $I(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$, consisting of μ_1, σ_1, $\mu_{2.1}$, $\sigma_{2.1}$ and θ_1 instead of μ_1, σ_1, μ_2, σ_2 and ρ. If we let $\boldsymbol{I} = n\ \boldsymbol{A}$, the elements of the matrix $\boldsymbol{A}$ are

$$A_{11} = \frac{b}{\sigma_1^2(b+2)}\ , A_{12} = \frac{b}{\sigma_1^2(b+2)}\big(\psi(b+1) - \psi(2)\big),$$

$$A_{13} = 0,\ \ A_{14} = 0,\ A_{15} = 0,$$

$$A_{22} = \frac{1}{\sigma_1^2}\left[1 + \frac{b}{(b+2)}\left\{\psi'(b+1) + \psi'(2) + \big(\psi(b+1) - \psi(2)\big)^2\right\}\right],$$

$$A_{23} = 0,\ A_{24} = 0, A_{25} = 0, A_{33} = \frac{1}{\sigma_{2.1}^2}, A_{34} = 0,$$

$$A_{35} = -\frac{\sigma_1}{\sigma_{2.1}^2}\big(\psi(b) - \psi(1)\big),\ A_{44} = \frac{2}{\sigma_{2.1}^2},\ A_{45} = 0 \text{ and}$$

$$A_{55} = -\frac{\sigma_1^2}{\sigma_{2.1}^2}\Big(\psi'(b) + \psi'(1) + \big(\psi(b) - \psi(1)\big)^2\Big). \qquad (2.2.4.10)$$

Here the asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1$, $\hat{\mu}_{2.1}$, $\hat{\sigma}_{2.1}$ and $\hat{\theta}_1$ is given by $\mathbf{\Sigma} \equiv I^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$. See Appendix D for details.

Hypothesis Testing: Now, we want to test the mean vector

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_{2.1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.2.4.11}$$

Since $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$ are asymptotically equivalent to the ML estimators, the distribution of the random vector $\sqrt{n}(\hat{\mu}_1, \hat{\mu}_{2.1})$ is bivariate normal with zero mean vector and (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n \begin{bmatrix} \hat{\sigma}_{11} & 0 \\ 0 & \hat{\sigma}_{33} \end{bmatrix} \tag{2.2.4.12}$$

where $\hat{\sigma}_{11}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij} = \mathbf{\Sigma_{ij}} = I_{ij}^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$. The covariance between $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$ is zero since they are orthogonal components, so there is no need to estimate it. Since in these elements $\hat{\sigma}_1$ and $\hat{\sigma}_{2.1}$ converge to σ_1 and $\sigma_{2.1}$, respectively, the asymptotic null distribution of

$$\begin{aligned} \hat{\mathrm{T}}_1^2 &= n(\hat{\mu}_1, \hat{\mu}_{2.1}) \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_{2.1} \end{pmatrix} \\ &= \hat{\mu}_1{}^2 / \hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2 / \hat{\sigma}_{33} \end{aligned} \tag{2.2.4.13}$$

is chi-square with 2 df. We reject H_0 at the 5% significance level if the value of $\hat{\mathrm{T}}_1^2$ is greater than $\chi^2_{0.05}(2)$. The nonnull distribution (asymptotic) of $\hat{\mathrm{T}}_1^2$ is noncentral chi-square with 2 df and noncentrality parameter

$$\lambda^2 = n(\mu_1, \mu_{2.1})\,\mathbf{\Omega}^{-1}\begin{pmatrix}\mu_1\\ \mu_{2.1}\end{pmatrix}$$

$$= {\mu_1}^2/\sigma_{11} + {\mu_{2.1}}^2/\sigma_{33}. \qquad (2.2.4.14)$$

For small n, the null distribution of $\frac{(n-2)}{2(n-1)}\hat{T}_1^2$ is approximately central-F with $(2, n-2)$ degrees of freedom and the nonnull distribution is approximately noncentral-F with $(2, n-2)$ degrees of freedom and noncentrality parameter λ^2.

CHAPTER 3

NONNORMAL CONDITIONAL AND NORMAL MARGINAL DISTRIBUTIONS

Summary: In this chapter we develop estimation and hypothesis testing procedure for the situation when the marginal distribution is normal (stochastic) and the conditional distribution is nonnormal. We will, for illustration, take the conditional distribution as Generalized Logistic. The methodology, however, applies to any other location-scale distribution.

3.1 Estimation of Parameters

Suppose that the marginal distribution of X is normal with density

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(x-\mu_1)^2 \right), \tag{3.1.1}$$

and the conditional distribution of Y given $X = x$ is Generalized Logistic given as

$$h(y \mid x) = \frac{b}{\sigma_2 (1-\rho^2)^{1/2}} \frac{\exp\left(-\dfrac{y-\mu_2-\rho\dfrac{\sigma_2}{\sigma_1}(x-\mu_1)}{\sigma_2\sqrt{(1-\rho^2)}} \right)}{\left[1+\exp\left(-\dfrac{y-\mu_2-\rho\dfrac{\sigma_2}{\sigma_1}(x-\mu_1)}{\sigma_2\sqrt{(1-\rho^2)}} \right) \right]^{b+1}} \tag{3.1.2}$$

where $-\infty < x < \infty$, $-\infty \langle y \langle \infty$, $\mu_1, \mu_2 \in \Re$, $\sigma_1, \sigma_2 \rangle 0$, $-1 \langle \rho \langle 1$, $b > 0$.

Given a random sample (x_i, y_i) $(1 \le i \le n)$, the marginal and conditional likelihood functions are

$$L_X = \frac{1}{(2\pi)^n \sigma_1^n} e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(x_i-\mu_1)^2} \quad \text{and} \tag{3.1.3}$$

$$L_{Y|X=x} = \frac{b^n}{\sigma_2^n (1-\rho^2)^{n/2}} \frac{e^{-\frac{1}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\left(y_i-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_i-\mu_1)\right)}}{\prod_{i=1}^{n}\left(1+e^{-\frac{y_i-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_i-\mu_1)}{\sigma_2\sqrt{(1-\rho^2)}}}\right)^{b+1}} . \tag{3.1.4}$$

If we write $z_i = (x_i - \mu_1)/\sigma_1$, $e_i = y_i - \theta_1 x_i - \mu_{2.1}$, $\mu_{2.1} = \mu_2 - \theta_1 \mu_1$, $\sigma_{2.1}^2 = \sigma_2^2(1-\rho^2)$ and $\theta_1 = \rho\frac{\sigma_2}{\sigma_1}$, the loglikelihood functions based on the marginal and the conditional distributions are, respectively,

$$\ln L_X = -\frac{n}{2}\ln(2\pi) - n\ln\sigma_1 - \frac{1}{2}\sum_{i=1}^{n} z_i^2 , \quad -\infty < z < \infty \tag{3.1.5}$$

and

$$\ln L_{Y|X=x} = n\ln b - n\ln\sigma_{2.1} - \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} e_i$$

$$-(b+1)\sum_{i=1}^{n}\ln\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right), \quad -\infty < e < \infty . \tag{3.1.6}$$

The loglikelihood function is

$$\ln L = -\frac{n}{2}\ln(2\pi) - n\ln\sigma_1 - \frac{1}{2}\sum_{i=1}^{n} z_i^2$$
$$+ n\ln b - n\ln\sigma_{2.1} - \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} e_i - (b+1)\sum_{i=1}^{n}\ln\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right) \qquad (3.1.7)$$

where $-\infty < z < \infty$, $-\infty < e < \infty$, $\mu_1, \mu_{2.1} \in \Re$, $\sigma_1, \sigma_{2.1} \rangle 0$, $b > 0$.

The likelihood equations for estimating μ_1, σ_1, $\mu_{2.1}$, $\sigma_{2.1}$ and θ_1 are

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{1}{\sigma_1}\sum_{i=1}^{n} z_i = 0 \qquad (3.1.8)$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_i^2 = 0 \qquad (3.1.9)$$

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{n}{\sigma_{2.1}} - \frac{(b+1)}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_{2.1}}}}{\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right)} = 0 \qquad (3.1.10)$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^2}\sum_{i=1}^{n} e_i - \frac{(b+1)}{\sigma_{2.1}^2}\sum_{i=1}^{n} e_i \frac{e^{-\frac{e_i}{\sigma_{2.1}}}}{\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right)} = 0 \quad \text{and} \qquad (3.1.11)$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_i - \frac{(b+1)\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_i \frac{e^{-\frac{e_i}{\sigma_{2.1}}}}{\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right)} = 0\,. \qquad (3.1.12)$$

The ML estimators of μ_1 and σ_1 found from (3.1.8)-(3.1.9) are

$$\hat{\mu}_1 = \bar{x} \text{ and } \hat{\sigma}_1 = s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}}\,. \tag{3.1.13}$$

Because of the intractable function $\dfrac{e^{-\frac{e_i}{\sigma_{2.1}}}}{\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right)}$, the equations (3.1.10)-(3.1.12) can not be solved. Hence, we use the modified likelihood methodology to solve them.

To find the MML estimators, we define $w_i = y_i - \theta_1 x_i$. We order w_i $(1 \le i \le n)$, in ascending order of magnitude. Let

$$w_{(1)} \le w_{(2)} \le \ldots \le w_{(n)} \tag{3.1.14}$$

be the ordered variates (for a given θ_1). Then, $e_{(i)} = w_{(i)} - \mu_{2.1}$ has the same order with $w_{(i)}$ since $\mu_{2.1}$ is a constant. We also define $a_i = e_i / \sigma_{2.1}$ and write $a_{(i)} = e_{(i)} / \sigma_{2.1} = (w_{(i)} - \mu_{2.1}) / \sigma_{2.1}$. Also, $w_{(i)} = y_{[i]} - \theta_1 x_{[i]}$, $(x_{[i]}, y_{[i]})$ and $z_{[i]} = (x_{[i]} - \mu_1) / \sigma_1$. If we write $g(a) = \dfrac{e^{-a}}{(1+e^{-a})}$, the likelihood equations expressed in terms of $a_{(i)}$, $e_{(i)}$ and $z_{(i)}$ are

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{n}{\sigma_{2.1}} - \frac{(b+1)}{\sigma_{2.1}} \sum_{i=1}^{n} g(a_{(i)}) = 0 \tag{3.1.15}$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^2} \sum_{i=1}^{n} e_{(i)} - \frac{(b+1)}{\sigma_{2.1}^2} \sum_{i=1}^{n} e_{(i)} g(a_{(i)}) = 0 \text{ and} \tag{3.1.16}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}} \sum_{i=1}^{n} z_{[i]} - \frac{(b+1)\sigma_1}{\sigma_{2.1}} \sum_{i=1}^{n} z_{[i]} g(a_{(i)}) = 0 . \tag{3.1.17}$$

In order to derive the MML estimators, we linearize the function $g(a_{(i)})$ by using the first two terms of a Taylor expansion around $E(a_{(i)}) = t_{(i)}$ as follows:

$$\begin{aligned} g(a_{(i)}) &\cong g(t_{(i)}) + (a_{(i)} - t_{(i)}) \left(\frac{dg(a_{(i)})}{a} \right)_{a_{(i)}=t_{(i)}} \\ &= \frac{e^{-t_{(i)}}}{\left(1+e^{-t_{(i)}}\right)} + (a_{(i)} - t_{(i)}) \left(- \frac{e^{-t_{(i)}}}{\left(1+e^{-t_{(i)}}\right)^2} \right) \\ &= \alpha_i - \beta_i a_{(i)}, \ 1 \le i \le n . \end{aligned} \tag{3.1.18}$$

where $\alpha_i = \frac{e^{-t_{(i)}}}{\left(1+e^{-t_{(i)}}\right)^2} \left(1 + t_{(i)} + e^{-t_{(i)}}\right)$ and $\beta_i = \frac{e^{-t_{(i)}}}{\left(1+e^{-t_{(i)}}\right)^2}$.

Substituting (3.1.18) in (3.1.15)-(3.1.17), we obtain the following modified maximum likelihood equations:

$$\frac{\partial \ln L^*}{\partial \mu_{2.1}} = \frac{n}{\sigma_{2.1}} - \frac{(b+1)}{\sigma_{2.1}} \sum_{i=1}^{n} \left(\alpha_i - \beta_i a_{(i)} \right) = 0 \tag{3.1.19}$$

$$\frac{\partial \ln L^*}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^2} \sum_{i=1}^{n} e_{(i)} - \frac{(b+1)}{\sigma_{2.1}^2} \sum_{i=1}^{n} e_{(i)} \left(\alpha_i - \beta_i a_{(i)} \right) = 0 \quad \text{and} \tag{3.1.20}$$

$$\frac{\partial \ln L^*}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}} \sum_{i=1}^{n} z_{[i]} - \frac{(b+1)\sigma_1}{\sigma_{2.1}} \sum_{i=1}^{n} z_{[i]} \left(\alpha_i - \beta_i a_{(i)} \right) = 0 . \tag{3.1.21}$$

The MML estimators are the solutions of the equations (3.1.19)-(3.1.21):

$$\hat{\mu}_{2.1} = \bar{y}_{[.]} - \hat{\theta}_1 \bar{x}_{[.]} - \frac{\Delta}{m}\hat{\sigma}_{2.1} \tag{3.1.22}$$

$$\hat{\sigma}_{2.1} = \frac{-B + \sqrt{B^2 + 4nC}}{2\sqrt{n(n-2)}} \quad \text{and} \tag{3.1.23}$$

$$\hat{\theta}_1 = K - D\hat{\sigma}_{2.1} \ . \tag{3.1.24}$$

Here, $B = (b+1)\sum_{i=1}^{n} \Delta_i \left\{ y_{[i]} - \bar{y}_{[.]} - K\left(x_{[i]} - \bar{x}_{[.]}\right)\right\}$

$$C = (b+1)\left\{ \sum_{i=1}^{n} \beta_i \left(y_{[i]} - \bar{y}_{[.]}\right)^2 - K\sum_{i=1}^{n} \beta_i \left(x_{[i]} - \bar{x}_{[.]}\right) y_{[i]} \right\}$$

$$K = \sum_{i=1}^{n} \beta_i \left(x_{[i]} - \bar{x}_{[.]}\right) y_{[i]} \Big/ \sum_{i=1}^{n} \beta_i \left(x_{[i]} - \bar{x}_{[.]}\right)^2$$

$$D = \sum_{i=1}^{n} \Delta_i \left(x_{[i]} - \bar{x}_{[.]}\right) \Big/ \sum_{i=1}^{n} \beta_i \left(x_{[i]} - \bar{x}_{[.]}\right)^2$$

$$\bar{x}_{[.]} = \frac{1}{m}\sum_{i=1}^{n} \beta_i x_{[i]} \ , \ \bar{y}_{[.]} = \frac{1}{m}\sum_{i=1}^{n} \beta_i y_{[i]} \quad \text{and}$$

$$\Delta_i = \alpha_i - \frac{1}{(b+1)} \ , \ \Delta = \sum_{i=1}^{n} \Delta_i \ , \ m = \sum_{i=1}^{n} \beta_i \ . \tag{3.1.25}$$

If we write $z_i = (x_i - \mu_1)/\sigma_1$ and $e_i = y_i - \mu_2 - \rho\left(\frac{\sigma_2}{\sigma_1}\right)(x_i - \mu_1)$, $1 \le i \le n$, the loglikelihood function is

$$\ln L = -\frac{n}{2}\ln(2\pi) - n\ln\sigma_1 - \frac{1}{2}\sum_{i=1}^{n} z_i^2$$

$$+ n\ln b - n\ln\sigma_2 - \frac{n}{2}\ln(1-\rho^2) - \frac{1}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_i$$

$$-(b+1)\sum_{i=1}^{n}\ln\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right),\quad -\infty<z<\infty,\ -\infty<e<\infty. \tag{3.1.26}$$

The likelihood equations for estimating μ_1, σ_1, μ_2, σ_2 and ρ are

$$\frac{\partial \ln L}{\partial \mu_1}=\frac{1}{\sigma_1}\sum_{i=1}^{n}z_i-\frac{n\rho}{\sigma_1\sqrt{(1-\rho^2)}}+\frac{(b+1)\rho}{\sigma_1\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0 \tag{3.1.27}$$

$$\frac{\partial \ln L}{\partial \sigma_1}=-\frac{n}{\sigma_1}+\frac{1}{\sigma_1}\sum_{i=1}^{n}z_i^2-\frac{\rho}{\sigma_1\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i$$

$$+\frac{(b+1)\rho}{\sigma_1\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0 \tag{3.1.28}$$

$$\frac{\partial \ln L}{\partial \mu_2}=\frac{n}{\sigma_2\sqrt{(1-\rho^2)}}-\frac{(b+1)}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0 \tag{3.1.29}$$

$$\frac{\partial \ln L}{\partial \sigma_2}=-\frac{n}{\sigma_2}+\frac{1}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}e_i+\frac{\rho}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i$$

$$-\frac{(b+1)\rho}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}$$

$$-\frac{(b+1)}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}e_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0\quad\text{and} \tag{3.1.30}$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n} e_i + \frac{1}{\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} z_i$$

$$-\frac{(b+1)}{\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} z_i \frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}$$

$$+\frac{(b+1)\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n} e_i \frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0. \qquad (3.1.31)$$

Also, writing $\theta_1 = \rho\frac{\sigma_2}{\sigma_1}$,

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} z_i - \frac{(b+1)\sigma_1}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} z_i \frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0. \qquad (3.1.32)$$

From (3.1.29) and (3.1.32),

$$n-(b+1)\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0$$

and

$$\sum_{i=1}^{n} z_i - (b+1)\sum_{i=1}^{n} z_i \frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0. \qquad (3.1.33)$$

Using (3.1.33), the likelihood equations reduce to

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{1}{\sigma_1}\sum_{i=1}^{n} z_i = 0 \tag{3.1.34}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_i^2 = 0 \tag{3.1.35}$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{n}{\sigma_2\sqrt{(1-\rho^2)}} - \frac{(b+1)}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0 \tag{3.1.36}$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_i$$

$$-\frac{(b+1)}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_i \frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0 \text{ and} \tag{3.1.37}$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n} e_i$$

$$+\frac{(b+1)\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n} e_i \frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0. \tag{3.1.38}$$

It is easy to find the ML estimators of μ_1 and σ_1 from (3.1.34)-(3.1.35) as

$$\hat{\mu}_1 = \bar{x} \text{ and } \hat{\sigma}_1 = s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}}. \tag{3.1.39}$$

Because of the function, $g(e_i) = \dfrac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}$, (3.1.36)-(3.1.38) can not be solved and solving them by iteration is problematic.

To find the MML estimators, we again define $w_i = y_i - \theta_1 x_i$ and order w_i $(1 \le i \le n)$, in ascending order of magnitude. Let

$$w_{(1)} \le w_{(2)} \le \ldots \le w_{(n)} \tag{3.1.40}$$

be the ordered variates. Since $a_{(i)} = \dfrac{e_{(i)}}{\sigma_2\sqrt{(1-\rho^2)}}$, the problematic function $g(e_i) = \dfrac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}$ can be expressed in terms of a_i, which is equal to $g(a_i) = \dfrac{e^{-a_i}}{(1+e^{-a_i})}$. When we order the values of w_i's, we have $a_{(i)} = \dfrac{e_{(i)}}{\sigma_2\sqrt{(1-\rho^2)}} = \dfrac{(w_{(i)} - (\mu_2 - \theta_1\mu_1))}{\sigma_2\sqrt{(1-\rho^2)}}$. Now, $w_{(i)} = y_{[i]} - \theta_1 x_{[i]}$ where $(x_{[i]}, y_{[i]})$ are the concomitants of $w_{(i)}$ and also $g(a_{(i)}) = \dfrac{e^{-a_{(i)}}}{(1+e^{-a_{(i)}})}$. The likelihood equations which are the same as (3.1.36)-(3.1.38) are

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{n}{\sigma_2\sqrt{(1-\rho^2)}} - \frac{(b+1)}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} g(a_{(i)}) = 0 \tag{3.1.41}$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_{(i)} - \frac{(b+1)}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_{(i)} g(a_{(i)}) = 0 \text{ and} \tag{3.1.42}$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2(1-\rho^2)^{3/2}} \sum_{i=1}^{n} e_{(i)} + \frac{(b+1)\rho}{\sigma_2(1-\rho^2)^{3/2}} \sum_{i=1}^{n} e_{(i)} g(a_{(i)}) = 0 . \quad (3.1.43)$$

We again expand $g(a_{(i)})$ around its expected value by using a Taylor expansion as in (3.1.18). Substituting $g(a_{(i)}) \cong \alpha_i - \beta_i a_{(i)}$ in the equations (3.1.41)-(3.1.43) gives the following modified maximum likelihood equations:

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{n}{\sigma_2\sqrt{(1-\rho^2)}} - \frac{(b+1)}{\sigma_2\sqrt{(1-\rho^2)}} \sum_{i=1}^{n} (\alpha_i - \beta_i a_{(i)}) = 0 \quad (3.1.44)$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^2\sqrt{(1-\rho^2)}} \sum_{i=1}^{n} e_{(i)}$$

$$- \frac{(b+1)}{\sigma_2^2\sqrt{(1-\rho^2)}} \sum_{i=1}^{n} e_{(i)}(\alpha_i - \beta_i a_{(i)}) = 0 \quad \text{and} \quad (3.1.45)$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2(1-\rho^2)^{3/2}} \sum_{i=1}^{n} e_{(i)}$$

$$+ \frac{(b+1)\rho}{\sigma_2(1-\rho^2)^{3/2}} \sum_{i=1}^{n} e_{(i)}(\alpha_i - \beta_i a_{(i)}) = 0 . \quad (3.1.46)$$

We find the MML estimators by solving (3.1.44)-(3.1.46) and the functional relations $\rho = \theta_1 \frac{\sigma_1}{\sigma_2}$ and $\sigma_{2.1} = \sigma_2\sqrt{(1-\rho^2)}$ as

$$\hat{\mu}_2 = \bar{y}_{[.]} - \hat{\theta}_1(\bar{x}_{[.]} - \hat{\mu}_1) - \frac{\Delta}{m}\hat{\sigma}_{2.1} \quad (3.1.47)$$

$$\hat{\sigma}_2 = \sqrt{\hat{\sigma}_{2.1}^2 + \hat{\theta}_1^2\hat{\sigma}_1^2} \quad \text{and} \quad (3.1.48)$$

$$\hat{\rho} = \hat{\theta}_1 \frac{\hat{\sigma}_1}{\hat{\sigma}_2} . \quad (3.1.49)$$

Related terms are given in (3.1.25). Realizing that $\hat{\mu}_1 = \bar{x}$ and $\hat{\sigma}_1 = s_x$, we have

$$\hat{\mu}_2 = \bar{y}_{[.]} - \hat{\theta}_1(\bar{x}_{[.]} - \bar{x}) - \frac{\Delta}{m}\hat{\sigma}_{2.1} \tag{3.1.50}$$

$$\hat{\sigma}_2 = \sqrt{\hat{\sigma}_{2.1}^2 + \hat{\theta}_1^2 s_x^2} \text{ and} \tag{3.1.51}$$

$$\hat{\rho} = \hat{\theta}_1 \frac{s_x}{\hat{\sigma}_2}. \tag{3.1.52}$$

Computations: The computation of the estimators is carried out in two iterations. In the first iteration, $w_i = y_i - \theta_1 x_i \ (1 \le i \le n)$ is calculated by replacing θ_1 by the LSE $\tilde{\theta}_1 = \sum_{i=1}^{n}(x_i - \bar{x})y_i \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2$. The MMLE $\hat{\theta}_1$ is then calculated from (3.1.23)-(3.1.24). In the second iteration, we obtain $w_{(i)}$'s by ordering $w_i = y_i - \hat{\theta}_1 x_i \ (1 \le i \le n)$ and the estimators are calculated by using these $w_{(i)}$'s and the concomitants $(x_{[i]}, y_{[i]})$.

3.2 Properties of the MML Estimators

Lemma 3.1: For known μ_1 and μ_2, $\hat{\theta}_1(\mu_1, \mu_2)$ is asymptotically the MVB estimator of θ_1 and is asymptotically normally distributed with mean θ_1 and variance $\dfrac{\sigma_{2.1}^2}{(b+1)\sum_{i=1}^{n}\beta_i(x_{(i)} - \mu_1)^2}$. This follows from the fact that

$$\frac{\partial \ln L}{\partial \theta_1} \cong \frac{\partial \ln L^*}{\partial \theta_1} = \frac{(b+1)\sum_{i=1}^{n}\beta_i(x_{(i)} - \mu_1)^2}{\sigma_{2.1}^2}\left(\hat{\theta}_1(\mu_1, \mu_2) - \theta_1\right) \tag{3.2.1}$$

and the difference $(1/n)\{(\partial \ln L/\partial\theta_1)-(\partial \ln L^*/\partial\theta_1)\}$ converges to zero as n tends to infinity.

Lemma 3.2: For μ_1 and μ_2 unknown, $\hat{\theta}_1$ is asymptotically the MVB estimator of θ_1 and is asymptotically normally distributed with mean θ_1 and variance $\dfrac{\sigma_{2.1}^2}{(b+1)\sum_{i=1}^{n}\beta_i(x_{(i)}-\bar{x}_{[.]})^2}$. This follows from the fact that

$$\frac{\partial \ln L}{\partial\theta_1}\cong\frac{\partial \ln L^*}{\partial\theta_1}=\frac{(b+1)\sum_{i=1}^{n}\beta_i(x_{(i)}-\bar{x}_{[.]})^2}{\sigma_{2.1}^2}\left(\hat{\theta}_1-\theta_1\right) \qquad (3.2.2)$$

and $\hat{\mu}_1$ and $\hat{\mu}_2$ converge to μ_1 and μ_2, respectively, as n tends to infinity.

3.3 Asymptotic Covariance Matrix

The asymptotic covariance matrix of the estimators $\hat{\mu}_1$, $\hat{\sigma}_1$, $\hat{\mu}_2$, $\hat{\sigma}_2$ and $\hat{\rho}$ is given by $I^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)$, where I is the Fisher information matrix. The Fisher information matrix is given by $(i, j = 1,2,3,4,5)$

$$\boldsymbol{I}=\left[I_{ij}\right]=\left[-E\left(\frac{\partial^2 \ln L}{\partial\tau_i\partial\tau_j}\right)\right],\ \tau_1=\mu_1,\ \tau_2=\sigma_1,\ \tau_3=\mu_2,\ \tau_4=\sigma_2,\ \tau_5=\rho.$$

If we let $\boldsymbol{I}=n\ \boldsymbol{A}$, the elements of the matrix $\boldsymbol{A}$ are

$$A_{11}=\frac{1}{\sigma_1^2}\left[1+\frac{b}{(b+2)}\frac{\rho^2}{(1-\rho^2)}\right],\ A_{12}=0,\ A_{13}=-\frac{b}{(b+2)}\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)},$$

$$A_{14} = -\frac{b}{(b+2)}\frac{\rho}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\big(\psi(b+1)-\psi(2)\big),$$

$$A_{15} = \frac{b}{(b+2)}\frac{\rho^2}{\sigma_1(1-\rho^2)^{3/2}}\big(\psi(b+1)-\psi(2)\big),$$

$$A_{22} = \frac{1}{\sigma_1^2}\left(2+\frac{b}{(b+2)}\frac{\rho^2}{(1-\rho^2)}\right), A_{23} = 0\,, A_{24} = -\frac{b}{(b+2)}\frac{\rho^2}{\sigma_1\sigma_2(1-\rho^2)},$$

$$A_{25} = -\frac{b}{(b+2)}\frac{\rho}{\sigma_1(1-\rho^2)},\ A_{33} = \frac{b}{(b+2)}\frac{1}{\sigma_2^2(1-\rho^2)},$$

$$A_{34} = \frac{b}{(b+2)}\frac{\big(\psi(b+1)-\psi(2)\big)}{\sigma_2^2\sqrt{1-\rho^2}},\ A_{35} = -\frac{b}{(b+2)}\frac{\rho}{\sigma_2(1-\rho^2)^{3/2}}\big(\psi(b+1)-\psi(2)\big),$$

$$A_{44} = \frac{1}{\sigma_2^2}\left[1+\frac{b}{(b+2)}\left(\frac{\rho^2}{(1-\rho^2)}+\big(\psi'(b+1)+\psi'(2)+\big(\psi(b+1)-\psi(2)\big)^2\big)\right)\right],$$

$$A_{45} = -\frac{\rho}{\sigma_2(1-\rho^2)}\left[1+\frac{b}{(b+2)}\left(-1+\big(\psi'(b+1)+\psi'(2)+\big(\psi(b+1)-\psi(2)\big)^2\big)\right)\right] \text{ and}$$

$$A_{55} = \frac{\rho^2}{(1-\rho^2)^2}+\frac{b}{(b+2)}\left(\frac{1}{(1-\rho^2)}\right.$$

$$\left.+\frac{\rho^2}{(1-\rho^2)^2}\big(\psi'(b+1)+\psi'(2)+\big(\psi(b+1)-\psi(2)\big)^2\big)\right). \qquad (3.3.1)$$

Here, the asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1,\ \hat{\mu}_2,\ \hat{\sigma}_2$ and $\hat{\rho}$ is given by $\mathbf{\Sigma} \equiv I^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)$. See Appendix D for details.

Hypothesis Testing: Our major interest is testing the null hypothesis $H_0 : \rho = 0$. Since the MML estimators are asymptotically equivalent to the ML estimators, the likelihood function L is maximized (asymptotically) by the MML estimators. Thus, the likelihood ratio statistic is (asymptotically)

$$\hat{\lambda} = \frac{\max(L \mid H_0)}{\max(L \mid H_1)}$$

$$= \frac{\dfrac{1}{\hat{\sigma}_{20}^{n}} \dfrac{e^{-\frac{1}{\hat{\sigma}_{20}}\sum_{i=1}^{n}(y_i-\hat{\mu}_{20})}}{\prod_{i=1}^{n}\left(1+e^{-\frac{y_i-\hat{\mu}_{20}}{\hat{\sigma}_{20}}}\right)^{b+1}}}{\dfrac{1}{\hat{\sigma}_{2}^{n}(1-\hat{\rho}^2)^{n/2}} \dfrac{e^{-\frac{1}{\hat{\sigma}_{2.1}}\sum_{i=1}^{n}(y_i-\hat{\mu}_2-\hat{\theta}_1(x_i-\hat{\mu}_1))}}{\prod_{i=1}^{n}\left(1+e^{-\frac{y_i-\hat{\mu}_2-\hat{\theta}_1(x_i-\hat{\mu}_1)}{\hat{\sigma}_{2.1}}}\right)^{b+1}}}. \qquad (3.3.2)$$

Here, $\hat{\mu}_{20}$ and $\hat{\sigma}_{20}$ are the MML estimators of μ_2 and σ_2 under $H_0 : \rho = 0$. Under $H_0 : \rho = 0$, since the conditional likelihood function (the marginal part cancels with the denominator) becomes a univariate Generalized Logistic likelihood function for Y, they are found exactly in the same way as in Section 2.1 but here y_i's take the place of x_i's. Since $\hat{\mu}_{20}$ and $\hat{\mu}_2$ both converge to μ_2, $\hat{\sigma}_{20}$ and $\hat{\sigma}_2$ both converge to σ_2, $\hat{\sigma}_{2.1}$ converges to $\sigma_{2.1}$, $\hat{\theta}_1$ converges to θ_1, and $\hat{\mu}_1$ converges to μ_1 as n tends to infinity, the likelihood ratio is a monotonic function of $\hat{\rho}^2$. Thus, to test H_0 against $H_1 : \rho < 0$ (or $\rho > 0$), the test based on $\hat{\rho}$ will be uniformly most powerful (asymptotically). We propose the statistic

$$W = \hat{\rho} \Big/ \sqrt{\frac{1}{n}\frac{(b+2)}{b}}$$

$$= \hat{\rho}\sqrt{n\frac{b}{(b+2)}}; \qquad (3.3.3)$$

$\frac{1}{n}\frac{(b+2)}{b}$ is the asymptotic variance of $\hat{\rho}$ under H_0, obtained from

$$\left\{I^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)\right\}_{\rho=0} . \tag{3.3.4}$$

The null distribution of W is referred to normal $N(0,1)$. Small values of W lead to rejection of H_0 against $H_1 : \rho < 0$.

Also we want to test the mean vector

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3.3.6}$$

Since $\hat{\mu}_1$ and $\hat{\mu}_2$ are asymptotically equivalent to the ML estimators, the distribution of the random vector $\sqrt{n}(\hat{\mu}_1, \hat{\mu}_2)$ is bivariate normal with zero mean vector and (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n\begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{13} \\ \hat{\sigma}_{13} & \hat{\sigma}_{33} \end{bmatrix} \tag{3.3.7}$$

where, $\hat{\sigma}_{11}$, $\hat{\sigma}_{13}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij} = \mathbf{\Sigma_{ij}} = I_{ij}^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)$. Since in these elements $\hat{\sigma}_1$ and $\hat{\sigma}_2$ converge to σ_1 and σ_2, respectively, the asymptotic null distribution of

$$\hat{\mathrm{T}}^2 = n(\hat{\mu}_1, \hat{\mu}_2)\,\hat{\mathbf{\Omega}}^{-1}\begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \tag{3.3.8}$$

is chi-square with 2 df. We reject H_0 at the 5% significance level if the value of $\hat{\mathrm{T}}^2$ is greater than $\chi^2_{0.05}(2)$. The nonnull distribution (asymptotic) of $\hat{\mathrm{T}}^2$ is noncentral chi-square with 2 df and noncentrality parameter

$$\lambda^2 = n(\mu_1, \mu_2)\,\mathbf{\Omega}^{-1}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}. \tag{3.3.9}$$

For small n, the null distribution of $\frac{(n-2)}{2(n-1)}\hat{T}^2$ is approximately central-F with $(2, n-2)$ degrees of freedom and the nonnull distribution is approximately noncentral-F with $(2, n-2)$ degrees of freedom and noncentrality parameter λ^2.

Now, we define the Fisher information matrix $I(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$ for estimating $\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}$ and θ_1. If we write $\boldsymbol{I} = n\,\boldsymbol{A}$, the elements of the matrix $\boldsymbol{A}$ are

$$A_{11} = \frac{1}{\sigma_1^2}, A_{12} = 0,\ A_{13} = 0, A_{14} = 0,\ A_{15} = 0, A_{22} = \frac{2}{\sigma_1^2},$$

$$A_{23} = 0, A_{24} = 0, A_{25} = 0, A_{33} = \frac{b}{(b+2)}\frac{1}{\sigma_{2.1}^2}, A_{34} = \frac{b}{(b+2)}\frac{\left(\psi(b+1) - \psi(2)\right)}{\sigma_{2.1}^2},$$

$$A_{35} = 0,\ A_{44} = \frac{1}{\sigma_{2.1}^2}\left[1 + \frac{b}{(b+2)}\left\{\psi'(b+1) + \psi'(2) + \left(\psi(b+1) - \psi(2)\right)^2\right\}\right]$$

and

$$A_{45} = 0,\ A_{55} = \frac{b}{(b+2)}\frac{\sigma_1^2}{\sigma_{2.1}^2}. \tag{3.3.10}$$

As usual, the asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1, \hat{\mu}_{2.1}, \hat{\sigma}_{2.1}$ and $\hat{\theta}_1$ is given by $\mathbf{\Sigma} \equiv I^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$. See Appendix D for details.

Hypothesis Testing: We are interested in the null hypothesis $(\mu_1, \mu_2) = (0,0)$ which is the same as

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_{2.1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3.3.11}$$

Since $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$ are asymptotically equivalent to the ML estimators, the distribution of the random vector $\sqrt{n}(\hat{\mu}_1, \hat{\mu}_{2.1})$ is bivariate normal with zero mean vector and (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n \begin{bmatrix} \hat{\sigma}_{11} & 0 \\ 0 & \hat{\sigma}_{33} \end{bmatrix} \tag{3.3.12}$$

where $\hat{\sigma}_{11}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij} = \mathbf{\Sigma_{ij}} = I_{ij}^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$. The covariance between $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$ is zero since they are orthogonal components, so there is no need to estimate it. Since in these elements $\hat{\sigma}_1$ and $\hat{\sigma}_{2.1}$ converge to σ_1 and $\sigma_{2.1}$, respectively, the asymptotic null distribution of

$$\begin{aligned} \hat{\mathrm{T}}_1^2 &= n(\hat{\mu}_1, \hat{\mu}_{2.1}) \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_{2.1} \end{pmatrix} \\ &= \hat{\mu}_1{}^2 / \hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2 / \hat{\sigma}_{33} \end{aligned} \tag{3.3.13}$$

is chi-square with 2 df. We reject H_0 at the 5% significance level if the value of $\hat{\mathrm{T}}_1^2$ is greater than $\chi^2_{0.05}(2)$. The nonnull distribution (asymptotic) of $\hat{\mathrm{T}}_1^2$ is noncentral chi-square with 2 df and noncentrality parameter

$$\begin{aligned} \lambda^2 &= n(\mu_1, \mu_{2.1}) \mathbf{\Omega}^{-1} \begin{pmatrix} \mu_1 \\ \mu_{2.1} \end{pmatrix} \\ &= \mu_1{}^2 / \sigma_{11} + \mu_{2.1}{}^2 / \sigma_{33}. \end{aligned} \tag{3.3.14}$$

For small n, the null distribution of $\frac{(n-2)}{2(n-1)}\hat{T}_1^2$ is approximately central-F with $(2, n-2)$ degrees of freedom and the nonnull distribution is approximately noncentral-F with $(2, n-2)$ degrees of freedom and noncentrality parameter λ^2.

CHAPTER 4

NONNORMAL CONDITIONAL AND MARGINAL DISTRIBUTIONS

Summary: Finally, we develop estimation and hypothesis testing procedures for the most difficult situation when both the marginal as well as the conditional distributions are nonnormal. For illustration, we take both of them as Generalized Logistic with shape parameter b_1 for the marginal and b_2 for the conditional distribution. The results extend to any other location-scale non-normal distribution.

4.1 Estimation of Parameters

Suppose that the marginal distribution of X is Generalized Logistic with density

$$g(x) = \frac{b_1}{\sigma_1} \frac{\exp\left(-\frac{(x-\mu_1)}{\sigma_1}\right)}{\left[1+\exp\left(-\frac{(x-\mu_1)}{\sigma_1}\right)\right]^{b_1+1}}, \tag{4.1.1}$$

and the conditional distribution of Y given $X = x$ is again Generalized Logistic with density

$$h(y \mid x) = \frac{b_2}{\sigma_2 (1-\rho^2)^{1/2}} \frac{\exp\left(-\dfrac{y - \mu_2 - \rho \dfrac{\sigma_2}{\sigma_1}(x-\mu_1)}{\sigma_2 \sqrt{(1-\rho^2)}} \right)}{\left[1 + \exp\left(-\dfrac{y - \mu_2 - \rho \dfrac{\sigma_2}{\sigma_1}(x-\mu_1)}{\sigma_2 \sqrt{(1-\rho^2)}} \right) \right]^{b_2+1}} \tag{4.1.2}$$

where $-\infty < x < \infty$, $-\infty \langle y \langle \infty$, $\mu_1, \mu_2 \in \Re$, $\sigma_1, \sigma_2 \rangle 0$, $-1 \langle \rho \langle 1$, $b_1, b_2 > 0$.

Given a random sample (x_i, y_i) $(1 \le i \le n)$, the marginal and conditional likelihood functions are

$$L_X = \frac{b_1^n}{\sigma_1^n} \frac{e^{-\sum_{i=1}^{n}\left(\frac{(x_i-\mu_1)}{\sigma_1}\right)}}{\prod_{i=1}^{n}\left(1 + e^{-\left(\frac{(x_i-\mu_1)}{\sigma_1}\right)}\right)^{b_1+1}} \quad \text{and} \tag{4.1.3}$$

$$L_{Y|X=x} = \frac{b_2^n}{\sigma_2^n (1-\rho^2)^{n/2}} \frac{e^{-\frac{1}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\left(y_i-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_i-\mu_1)\right)}}{\prod_{i=1}^{n}\left(1 + e^{-\frac{y_i-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_i-\mu_1)}{\sigma_2\sqrt{(1-\rho^2)}}}\right)^{b_2+1}} . \tag{4.1.4}$$

If we write $z_i = (x_i - \mu_1)/\sigma_1$, $e_i = y_i - \theta_1 x_i - \mu_{2.1}$, $\mu_{2.1} = \mu_2 - \theta_1 \mu_1$, $\sigma_{2.1}^2 = \sigma_2^2 (1-\rho^2)$ and $\theta_1 = \rho \dfrac{\sigma_2}{\sigma_1}$, the loglikelihoods based on the marginal and the conditional distributions are

$$\ln L_X = n \ln b_1 - n \ln \sigma_1 - \sum_{i=1}^{n} z_i - (b_1 + 1) \sum_{i=1}^{n} \ln\left(1 + e^{-z_i}\right), \quad -\infty < z < \infty , \tag{4.1.5}$$

and

$$\ln L_{Y|X=x} = n\ln b_2 - n\ln\sigma_{2.1} - \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} e_i$$

$$-(b_2+1)\sum_{i=1}^{n}\ln\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right), \quad -\infty < e < \infty. \tag{4.1.6}$$

The full loglikelihood function is

$$\ln L = n\ln b_1 - n\ln\sigma_1 - \sum_{i=1}^{n} z_i - (b_1+1)\sum_{i=1}^{n}\ln\left(1+e^{-z_i}\right)$$

$$+ n\ln b_2 - n\ln\sigma_{2.1} - \frac{1}{\sigma_{2.1}}\sum_{i=1}^{n} e_i - (b_2+1)\sum_{i=1}^{n}\ln\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right) \tag{4.1.7}$$

where $-\infty < z < \infty$, $-\infty < e < \infty$, $\mu_1, \mu_{2.1} \in \Re$, $\sigma_1, \sigma_{2.1} \rangle 0$, $b_1, b_2 > 0$.

The likelihood equations for estimating μ_1, σ_1, $\mu_{2.1}$, $\sigma_{2.1}$ and θ_1 are

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n}\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)} = 0 \tag{4.1.8}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_i - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n} z_i\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)} = 0 \tag{4.1.9}$$

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{n}{\sigma_{2.1}} - \frac{(b_2+1)}{\sigma_{2.1}}\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_{2.1}}}}{\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right)} = 0 \tag{4.1.10}$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^2}\sum_{i=1}^{n} e_i - \frac{(b_2+1)}{\sigma_{2.1}^2}\sum_{i=1}^{n} e_i\frac{e^{-\frac{e_i}{\sigma_{2.1}}}}{\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right)} = 0 \tag{4.1.11}$$

and

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}} \sum_{i=1}^{n} z_i - \frac{(b_2+1)\sigma_1}{\sigma_{2.1}} \sum_{i=1}^{n} z_i \frac{e^{-\frac{e_i}{\sigma_{2.1}}}}{\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right)} = 0 . \tag{4.1.12}$$

Because of the functions $\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}$ and $\frac{e^{-\frac{e_i}{\sigma_{2.1}}}}{\left(1+e^{-\frac{e_i}{\sigma_{2.1}}}\right)}$, the equations (4.1.8)-(4.1.12) and (4.1.14)-(4.1.15), respectively, are almost impossible to solve. Hence, we use the MML method to solve the equations (4.1.8)-(4.1.9) and (4.1.10)-(4.1.12). To find the MML estimators, we define $w_i = y_i - \theta_1 x_i$. We order the values x_i and w_i (for a given θ_1), $1 \le i \le n$, in ascending order of magnitude as

$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)} \tag{4.1.13}$$

$$w_{(1)} \le w_{(2)} \le ... \le w_{(n)} . \tag{4.1.14}$$

Then $e_{(i)} = w_{(i)} - \mu_{2.1}$ has the same order with $w_{(i)}$ since $\mu_{2.1}$ is a constant and $z_{(i)} = (x_{(i)} - \mu_1)/\sigma_1$ has the same order with $x_{(i)}$ since μ_1 is a constant and σ_1 is positive. We also define $a_i = e_i / \sigma_{2.1}$ so that $a_{(i)} = e_{(i)} / \sigma_{2.1} = (w_{(i)} - \mu_{2.1}) / \sigma_{2.1}$. Now $w_{(i)} = y_{[i]} - \theta_1 x_{[i]}$; $(x_{[i]}, y_{[i]})$ and $z_{[i]} = (x_{[i]} - \mu_1)/\sigma_1$ are the concomitants of $w_{(i)}$. Since complete sums are invariant to ordering, the likelihood equations can be written as

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b_1+1)}{\sigma_1} \sum_{i=1}^{n} \frac{e^{-z_{(i)}}}{\left(1+e^{-z_{(i)}}\right)} = 0 \tag{4.1.15}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{(i)} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)}\left(\frac{e^{-z_{(i)}}}{1+e^{-z_{(i)}}}\right) = 0 \tag{4.1.16}$$

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{n}{\sigma_{2.1}} - \frac{(b_2+1)}{\sigma_{2.1}}\sum_{i=1}^{n}\left(\frac{e^{-a_{(i)}}}{1+e^{-a_{(i)}}}\right) = 0 \tag{4.1.17}$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^2}\sum_{i=1}^{n} e_{(i)} - \frac{(b_2+1)}{\sigma_{2.1}^2}\sum_{i=1}^{n} e_{(i)}\left(\frac{e^{-a_{(i)}}}{1+e^{-a_{(i)}}}\right) = 0 \text{ and} \tag{4.1.18}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{[i]} - \frac{(b_2+1)\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{[i]}\left(\frac{e^{-a_{(i)}}}{1+e^{-a_{(i)}}}\right) = 0\,. \tag{4.1.19}$$

If we write $g_1(z) = \dfrac{e^{-z}}{(1+e^{-z})}$ and $g_2(a) = \dfrac{e^{-a}}{(1+e^{-a})}$, the likelihood equations expressed in terms of the ordered variates are

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n} g_1(z_{(i)}) = 0 \tag{4.1.20}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{(i)} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)} g_1(z_{(i)}) = 0 \tag{4.1.21}$$

$$\frac{\partial \ln L}{\partial \mu_{2.1}} = \frac{n}{\sigma_{2.1}} - \frac{(b_2+1)}{\sigma_{2.1}}\sum_{i=1}^{n} g_2(a_{(i)}) = 0 \tag{4.1.22}$$

$$\frac{\partial \ln L}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^2}\sum_{i=1}^{n} e_{(i)} - \frac{(b_2+1)}{\sigma_{2.1}^2}\sum_{i=1}^{n} e_{(i)} g_2(a_{(i)}) = 0 \text{ and} \tag{4.1.23}$$

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{[i]} - \frac{(b_2+1)\sigma_1}{\sigma_{2.1}}\sum_{i=1}^{n} z_{[i]} g_2(a_{(i)}) = 0\,. \tag{4.1.24}$$

To derive the MML estimators, we linearize the functions $g_1(z_{(i)})$ and $g_2(a_{(i)})$ by using the first two terms of Taylor series expansions around $E(z_{(i)}) = t_{1(i)}$ and $E(a_{(i)}) = t_{2(i)}$, respectively, as follows

$$g_1(z_{(i)}) \cong g_1(t_{1(i)}) + (z_{(i)} - t_{1(i)})\left(\frac{dg_1(z_{(i)})}{z}\right)_{z_{(i)}=t_{1(i)}}$$

$$= \frac{e^{-t_{1(i)}}}{\left(1+e^{-t_{1(i)}}\right)} + (z_{(i)} - t_{1(i)})\left(-\frac{e^{-t_{1(i)}}}{\left(1+e^{-t_{1(i)}}\right)^2}\right)$$

$$= \alpha_{1i} - \beta_{1i} z_{(i)}, \quad 1 \le i \le n, \tag{4.1.25}$$

where $\alpha_{1i} = \frac{e^{-t_{1(i)}}}{\left(1+e^{-t_{1(i)}}\right)^2}\left(1 + t_{1(i)} + e^{-t_{1(i)}}\right)$ and $\beta_{1i} = \frac{e^{-t_{1(i)}}}{\left(1+e^{-t_{1(i)}}\right)^2}$

and

$$g_2(a_{(i)}) \cong g_2(t_{2(i)}) + (a_{(i)} - t_{2(i)})\left(\frac{dg_2(a_{(i)})}{a}\right)_{a_{(i)}=t_{2(i)}}$$

$$= \frac{e^{-t_{2(i)}}}{\left(1+e^{-t_{2(i)}}\right)} + (a_{(i)} - t_{2(i)})\left(-\frac{e^{-t_{2(i)}}}{\left(1+e^{-t_{2(i)}}\right)^2}\right)$$

$$= \alpha_{2i} - \beta_{2i} a_{(i)}, \quad 1 \le i \le n. \tag{4.1.26}$$

where $\alpha_{2i} = \frac{e^{-t_{2(i)}}}{\left(1+e^{-t_{2(i)}}\right)^2}\left(1 + t_{2(i)} + e^{-t_{2(i)}}\right)$ and $\beta_{2i} = \frac{e^{-t_{2(i)}}}{\left(1+e^{-t_{2(i)}}\right)^2}$.

Substituting (4.1.25) and (4.1.26) in (4.1.20)-(4.1.24), we obtain the following modified maximum likelihood equations:

$$\frac{\partial \ln L^*}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n}(\alpha_{1i} - \beta_{1i} z_{(i)}) = 0 \tag{4.1.27}$$

$$\frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{(i)} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)}(\alpha_{1i} - \beta_{1i} z_{(i)}) = 0 \tag{4.1.28}$$

$$\frac{\partial \ln L^*}{\partial \mu_{2.1}} = \frac{n}{\sigma_{2.1}} - \frac{(b_2+1)}{\sigma_{2.1}} \sum_{i=1}^{n} (\alpha_{2i} - \beta_{2i} a_{(i)}) = 0 \tag{4.1.29}$$

$$\frac{\partial \ln L^*}{\partial \sigma_{2.1}} = -\frac{n}{\sigma_{2.1}} + \frac{1}{\sigma_{2.1}^2} \sum_{i=1}^{n} e_{(i)} - \frac{(b_2+1)}{\sigma_{2.1}^2} \sum_{i=1}^{n} e_{(i)} (\alpha_{2i} - \beta_{2i} a_{(i)}) = 0 \text{ and} \tag{4.1.30}$$

$$\frac{\partial \ln L^*}{\partial \theta_1} = \frac{\sigma_1}{\sigma_{2.1}} \sum_{i=1}^{n} z_{[i]} - \frac{(b_2+1)\sigma_1}{\sigma_{2.1}} \sum_{i=1}^{n} z_{[i]} (\alpha_{2i} - \beta_{2i} a_{(i)}) = 0 \,. \tag{4.1.31}$$

The equations have explicit solutions which are the following MML estimators:

$$\hat{\mu}_1 = K_1 + D_1 \hat{\sigma}_1 \tag{4.1.32}$$

$$\hat{\sigma}_1 = \frac{B_1 + \sqrt{B_1^2 + 4nC_1}}{2n} \tag{4.1.33}$$

$$\hat{\mu}_{2.1} = \bar{y}_{[.]} - \hat{\theta}_1 \bar{x}_{[.]} - \frac{\Delta}{m_2} \hat{\sigma}_{2.1} \tag{4.1.34}$$

$$\hat{\sigma}_{2.1} = \frac{-B_2 + \sqrt{B_2^2 + 4nC_2}}{2n} \text{ and} \tag{4.1.35}$$

$$\hat{\theta}_1 = K_2 - D_2 \hat{\sigma}_{2.1} \,. \tag{4.1.36}$$

Here,

$$B_1 = (b_1+1) \sum_{i=1}^{n} (x_{(i)} - K_1) \left(\frac{1}{(b_1+1)} - \alpha_{1i} \right)$$

$$C_1 = (b_1+1) \sum_{i=1}^{n} \beta_{1i} (x_{(i)} - K_1)^2$$

$$K_1 = \frac{1}{m_1} \sum_{i=1}^{n} \beta_{1i} x_{(i)} \,,\; D_1 = \frac{1}{m_1} \sum_{i=1}^{n} \left(\frac{1}{(b_1+1)} - \alpha_{1i} \right), m_1 = \sum_{i=1}^{n} \beta_{1i}$$

$$B_2 = (b_2+1) \sum_{i=1}^{n} \Delta_i \left\{ y_{[i]} - \bar{y}_{[.]} - K_2 \left(x_{[i]} - \bar{x}_{[.]} \right) \right\}$$

$$C_2 = (b_2+1) \left\{ \sum_{i=1}^{n} \beta_{2i} \left(y_{[i]} - \bar{y}_{[.]} \right)^2 - K_2 \sum_{i=1}^{n} \beta_{2i} \left(x_{[i]} - \bar{x}_{[.]} \right) y_{[i]} \right\}$$

$$K_2 = \sum_{i=1}^{n} \beta_{2i}\left(x_{[i]} - \bar{x}_{[.]}\right) y_{[i]} \Big/ \sum_{i=1}^{n} \beta_{2i}\left(x_{[i]} - \bar{x}_{[.]}\right)^2$$

$$D = \sum_{i=1}^{n} \Delta_i \left(x_{[i]} - \bar{x}_{[.]}\right) \Big/ \sum_{i=1}^{n} \beta_i \left(x_{[i]} - \bar{x}_{[.]}\right)^2$$

$$\bar{x}_{[.]} = \frac{1}{m_2} \sum_{i=1}^{n} \beta_{2i} x_{[i]}\,,\ \bar{y}_{[.]} = \frac{1}{m_2} \sum_{i=1}^{n} \beta_{2i} y_{[i]} \text{ and}$$

$$\Delta_i = \alpha_{2i} - \frac{1}{(b_2+1)},\ \Delta = \sum_{i=1}^{n} \Delta_i\ ,\ m_2 = \sum_{i=1}^{n} \beta_{2i}\ . \tag{4.1.37}$$

All the estimators are very different than those based on a bivariate normal distribution. It is, therefore, very important to recognize the true underlying distribution.

In order to develop the estimation procedure for the parameters μ_1, σ_1, μ_2, σ_2 and ρ directly, we consider the loglikelihood function

$$\begin{aligned} \ln L = {} & n \ln b_1 - n \ln \sigma_1 - \sum_{i=1}^{n} z_i - (b_1+1) \sum_{i=1}^{n} \ln\left(1 + e^{-z_i}\right) \\ & + n \ln b_2 - n \ln \sigma_2 - \frac{n}{2} \ln(1-\rho^2) - \frac{1}{\sigma_2 \sqrt{(1-\rho^2)}} \sum_{i=1}^{n} e_i \\ & - (b_2+1) \sum_{i=1}^{n} \ln\left(1 + e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right). \end{aligned} \tag{4.1.38}$$

The likelihood equations for estimating μ_1, σ_1, μ_2, σ_2 and ρ are

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b_1+1)}{\sigma_1} \sum_{i=1}^{n} \frac{e^{-z_i}}{\left(1 + e^{-z_i}\right)}$$

$$-\frac{n\rho}{\sigma_1\sqrt{(1-\rho^2)}}+\frac{(b_2+1)\rho}{\sigma_1\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0 \qquad (4.1.39)$$

$$\frac{\partial \ln L}{\partial \sigma_1}=-\frac{n}{\sigma_1}+\frac{1}{\sigma_1}\sum_{i=1}^{n}z_i-\frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n}z_i\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}$$

$$-\frac{\rho}{\sigma_1\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i+\frac{(b_2+1)\rho}{\sigma_1\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0 \qquad (4.1.40)$$

$$\frac{\partial \ln L}{\partial \mu_2}=\frac{n}{\sigma_2\sqrt{(1-\rho^2)}}-\frac{(b_2+1)}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0 \qquad (4.1.41)$$

$$\frac{\partial \ln L}{\partial \sigma_2}=-\frac{n}{\sigma_2}+\frac{1}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}e_i+\frac{\rho}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i$$

$$-\frac{(b_2+1)\rho}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}$$

$$-\frac{(b_2+1)}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}e_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0 \quad \text{and} \qquad (4.1.42)$$

$$\frac{\partial \ln L}{\partial \rho}=\frac{n\rho}{(1-\rho^2)}-\frac{\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n}e_i+\frac{1}{\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i$$

$$-\frac{(b_2+1)}{\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}$$

$$+\frac{(b_2+1)\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n}e_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0. \tag{4.1.43}$$

Also, writing $\theta_1=\rho\frac{\sigma_2}{\sigma_1}$,

$$\frac{\partial \ln L}{\partial \theta_1}=\frac{\sigma_1}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i-\frac{(b_2+1)\sigma_1}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}z_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0. \tag{4.1.44}$$

From (4.1.41) and (4.1.44),

$$n-(b_2+1)\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0$$

and

$$\sum_{i=1}^{n}z_i-(b_2+1)\sum_{i=1}^{n}z_i\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}=0. \tag{4.1.45}$$

Thus, the likelihood equations reduce to

$$\frac{\partial \ln L}{\partial \mu_1}=\frac{n}{\sigma_1}-\frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n}\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}=0 \tag{4.1.46}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_i - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n} z_i \frac{e^{-z_i}}{\left(1+e^{-z_i}\right)} = 0 \tag{4.1.47}$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{n}{\sigma_2\sqrt{(1-\rho^2)}} - \frac{(b_2+1)}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} \frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0 \tag{4.1.48}$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_i$$

$$-\frac{(b_2+1)}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_i \frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0 \quad \text{and} \tag{4.1.49}$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n} e_i$$

$$+\frac{(b_2+1)\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n} e_i \frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)} = 0\,. \tag{4.1.50}$$

Because of the intractable functions $\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}$ and $\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}$, the equations (4.1.46)-(4.1.47) and (4.1.48)-(4.1.50), respectively, can not be solved. Hence, we use MML method to solve the equations (4.1.46)-(4.1.47) and (4.1.48)-(4.1.50).

To find the MML estimators, we define $w_i = y_i - \rho\frac{\sigma_2}{\sigma_1}x_i$. We order the values x_i and w_i (for a given θ_1), $1 \le i \le n$, in ascending order of magnitude as

$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)} \tag{4.1.51}$$

$$w_{(1)} \le w_{(2)} \le ... \le w_{(n)} \tag{4.1.52}$$

Thus, $e_{(i)} = w_{(i)} - (\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1)$ has the same order as $w_{(i)}$ since $(\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1)$ is a constant and $z_{(i)} = (x_{(i)} - \mu_1)/\sigma_1$ has the same order as $x_{(i)}$ since μ_1 is a constant and σ_1 is positive. We also write $a_i = e_i \Big/ \left(\sigma_2\sqrt{1-\rho^2}\right)$. Thus, we have $a_{(i)} = e_{(i)} \Big/ \left(\sigma_2\sqrt{1-\rho^2}\right) = \left(w_{(i)} - (\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1)\right) \Big/ \left(\sigma_2\sqrt{1-\rho^2}\right)$. Now, $w_{(i)} = y_{[i]} - \rho\frac{\sigma_2}{\sigma_1}x_{[i]}$, $(x_{[i]}, y_{[i]})$ and $z_{[i]} = (x_{[i]} - \mu_1)/\sigma_1$ are the concomitants of $w_{(i)}$. Since complete sums are invariant to ordering, the likelihood equations can be written as

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n}\frac{e^{-z_{(i)}}}{\left(1+e^{-z_{(i)}}\right)} = 0 \tag{4.1.53}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n}z_{(i)} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n}z_{(i)}\frac{e^{-z_{(i)}}}{\left(1+e^{-z_{(i)}}\right)} = 0 \tag{4.1.54}$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{n}{\sigma_2\sqrt{(1-\rho^2)}} - \frac{(b_2+1)}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{e^{-a_{(i)}}}{\left(1+e^{-a_{(i)}}\right)} = 0 \tag{4.1.55}$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}e_{(i)} - \frac{(b_2+1)}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}e_{(i)}\frac{e^{-a_{(i)}}}{\left(1+e^{-a_{(i)}}\right)} = 0 \tag{4.1.56}$$

and

$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n}e_{(i)} + \frac{(b_2+1)\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n}e_{(i)}\frac{e^{-a_{(i)}}}{\left(1+e^{-a_{(i)}}\right)} = 0\,. \tag{4.1.57}$$

If we write $g_1(z) = \dfrac{e^{-z}}{(1+e^{-z})}$ and $g_2(a) = \dfrac{e^{-a}}{(1+e^{-a})}$ and linearize them as before, the modified likelihood equations are

$$\frac{\partial \ln L^*}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n}(\alpha_{1i} - \beta_{1i} z_{(i)}) = 0 \tag{4.1.58}$$

$$\frac{\partial \ln L^*}{\partial \sigma_1} = -\frac{n}{\sigma_1} + \frac{1}{\sigma_1}\sum_{i=1}^{n} z_{(i)} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n} z_{(i)}(\alpha_{1i} - \beta_{1i} z_{(i)}) = 0 \tag{4.1.59}$$

$$\frac{\partial \ln L^*}{\partial \mu_2} = \frac{n}{\sigma_2\sqrt{(1-\rho^2)}} - \frac{(b_2+1)}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}(\alpha_{2i} - \beta_{2i} a_{(i)}) = 0 \tag{4.1.60}$$

$$\frac{\partial \ln L^*}{\partial \sigma_2} = -\frac{n}{\sigma_2} + \frac{1}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_{(i)}$$

$$-\frac{(b_2+1)}{\sigma_2^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_{(i)}(\alpha_{2i} - \beta_{2i} a_{(i)}) = 0 \quad \text{and} \tag{4.1.61}$$

$$\frac{\partial \ln L^*}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n} e_{(i)}$$

$$+\frac{(b_2+1)\rho}{\sigma_2(1-\rho^2)^{3/2}}\sum_{i=1}^{n} e_{(i)}(\alpha_{2i} - \beta_{2i} a_{(i)}) = 0\,. \tag{4.1.62}$$

To solve (4.1.58)-(4.1.62), we use the functional relations $\rho = \theta_1 \dfrac{\sigma_1}{\sigma_2}$ and $\sigma_{2.1} = \sigma_2\sqrt{(1-\rho^2)}$, and obtain the following MMLE:

$$\hat{\mu}_1 = K_1 + D_1\hat{\sigma}_1 \tag{4.1.63}$$

$$\hat{\sigma}_1 = \frac{B_1 + \sqrt{B_1^2 + 4nC_1}}{2n} \tag{4.1.64}$$

$$\hat{\mu}_2 = \bar{y}_{[.]} - \hat{\theta}_1(\bar{x}_{[.]} - \hat{\mu}_1) - \frac{\Delta}{m_2}\hat{\sigma}_2\sqrt{(1-\hat{\rho}^2)} \tag{4.1.65}$$

$$\hat{\sigma}_2 = \sqrt{\hat{\sigma}_{2.1}^2 + \hat{\theta}_1^2\hat{\sigma}_1^2} \quad \text{and} \tag{4.1.66}$$

$$\hat{\rho} = \hat{\theta}_1 \frac{\hat{\sigma}_1}{\hat{\sigma}_2}. \tag{4.1.67}$$

Since the estimators (4.1.63)-(4.1.67) can be obtained from (4.1.32)-(4.1.36), by simple substitution, it follows that, like the ML estimators, the MML estimators have the invariance property even in this complex situation. This is a very interesting result indeed.

Computations: The computations are carried out exactly along the same lines as those in Section 3.1. See specifically page 77.

4.2 Properties of the MML Estimators

Lemma 4.1: For known μ_1 and μ_2, $\hat{\theta}_1(\mu_1, \mu_2)$ is asymptotically the MVB estimator of θ_1 and is asymptotically normally distributed with mean θ_1 and variance $\dfrac{\sigma_{2.1}^2}{(b_2+1)\sum_{i=1}^{n} \beta_{2i}(x_{(i)} - \mu_1)^2}$. This follows from the representation

$$\frac{\partial \ln L}{\partial \theta_1} \cong \frac{\partial \ln L^*}{\partial \theta_1} = \frac{(b_2+1)\sum_{i=1}^{n} \beta_{2i}(x_{(i)} - \mu_1)^2}{\sigma_{2.1}^2} \left(\hat{\theta}_1(\mu_1, \mu_2) - \theta_1\right). \tag{4.2.1}$$

Lemma 4.2: For μ_1 and μ_2 unknown, $\hat{\theta}_1$ is asymptotically the MVB estimator of θ_1 and is asymptotically normally distributed with mean θ_1 and variance $\dfrac{\sigma_{2.1}^2}{(b_2+1)\sum_{i=1}^{n} \beta_{2i}(x_{(i)} - \bar{x}_{[.]})^2}$. This follows from the fact that $\hat{\mu}_1$ and $\hat{\mu}_2$ converge to μ_1 and μ_2, respectively, as n tends to infinity and

$$\frac{\partial \ln L}{\partial \theta_1} \cong \frac{\partial \ln L^*}{\partial \theta_1} = \frac{(b_2+1)\sum_{i=1}^{n} \beta_{2i}(x_{(i)} - \bar{x}_{[.]})^2}{\sigma_{2.1}^2}\left(\hat{\theta}_1 - \theta_1\right). \qquad (4.2.2)$$

Bias Correction: For small $n\ (\le 15)$, $\hat{\mu}_1$, $\hat{\sigma}_1$ and $\hat{\sigma}_{2.1}$ have some bias. The bias corrected estimators are given by (see Appendix B for details)

$$\hat{\mu}_1 = K_1 - \frac{\hat{\sigma}_1}{m_1}\left(\sum_{i=1}^{n} \beta_{1i} t_{1(i)}\right) \quad \text{with} \quad \hat{\sigma}_1 = \frac{\left\{B_1 + \sqrt{B_1^2 + 4nC_1}\right\}}{2\sqrt{n(n-1)}}$$

$$\text{and} \quad \hat{\sigma}_{2.1} = \frac{-B_2 + \sqrt{B_2^2 + 4nC_2}}{2\sqrt{n(n-2)}}\ . \qquad (4.2.3)$$

4.3 Asymptotic Covariance Matrix

The asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1$, $\hat{\mu}_2$, $\hat{\sigma}_2$ and $\hat{\rho}$ is given by $I^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$, where I is the Fisher information matrix. The Fisher information matrix is given by $(i, j = 1,2,3,4,5)$

$$\boldsymbol{I} = \left[I_{ij}\right] = \left[-E\left(\frac{\partial^2 \ln L}{\partial \tau_i \partial \tau_j}\right)\right],\ \tau_1 = \mu_1,\ \tau_2 = \sigma_1,\ \tau_3 = \mu_2,\ \tau_4 = \sigma_2,\ \tau_5 = \rho\,.$$

If we let $\boldsymbol{I} = n\ \boldsymbol{A}$, the elements of the matrix $\boldsymbol{A}$ are

$$A_{11} = \frac{1}{\sigma_1^2}\left[\frac{b_1}{(b_1+2)} + \frac{b_2}{(b_2+2)}\frac{\rho^2}{(1-\rho^2)}\right],$$

$$A_{12} = \frac{1}{\sigma_1^2}\left[\frac{b_1}{(b_1+2)}\left(\psi(b_1+1) - \psi(2)\right) + \frac{b_2}{(b_2+2)}\frac{\rho^2}{(1-\rho^2)}\left(\psi(b_1) - \psi(1)\right)\right],$$

$$A_{13} = -\frac{b_2}{(b_2+2)}\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)},$$

$$A_{14} = -\frac{b_2}{(b_2+2)}\left[\frac{\rho^2}{\sigma_1\sigma_2(1-\rho^2)}\left(\psi(b_1)-\psi(1)\right)+\frac{\rho}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\left(\psi(b_2+1)-\psi(2)\right)\right],$$

$$A_{15} = -\frac{b_2}{(b_2+2)}\left[\frac{\rho}{\sigma_1(1-\rho^2)}\left(\psi(b_1)-\psi(1)\right)-\frac{\rho^2}{\sigma_1(1-\rho^2)^{3/2}}\left(\psi(b_2+1)-\psi(2)\right)\right],$$

$$A_{22} = \frac{1}{\sigma_1^2}\left[1+\frac{b_1}{(b_1+2)}\left\{\psi'(b_1+1)+\psi'(2)+\left(\psi(b_1+1)-\psi(2)\right)^2\right\}\right.$$

$$\left.+\frac{b_2}{(b_2+2)}\frac{\rho^2}{(1-\rho^2)}\left\{\psi'(b_1)+\psi'(1)+\left(\psi(b_1)-\psi(1)\right)^2\right\}\right],$$

$$A_{23} = -\frac{b_2}{(b_2+2)}\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)}\left(\psi(b_1)-\psi(1)\right),$$

$$A_{24} = -\frac{b_2}{(b_2+2)}\left[\frac{\rho^2}{\sigma_1\sigma_2(1-\rho^2)}\left\{\psi'(b_1)+\psi'(1)+\left(\psi(b_1)-\psi(1)\right)^2\right\}\right.$$

$$\left.+\frac{\rho}{\sigma_1\sigma_2\sqrt{1-\rho^2}}\left(\psi(b_1)-\psi(1)\right)\left(\psi(b_2+1)-\psi(2)\right)\right],$$

$$A_{25} = -\frac{b_2}{(b_2+2)}\left[\frac{\rho}{\sigma_1(1-\rho^2)}\left\{\psi'(b_1)+\psi'(1)+\left(\psi(b_1)-\psi(1)\right)^2\right\}\right.$$

$$\left.-\frac{\rho^2}{\sigma_1(1-\rho^2)^{3/2}}\left(\psi(b_1)-\psi(1)\right)\left(\psi(b_2+1)-\psi(2)\right)\right],$$

$$A_{33} = \frac{b_2}{(b_2+2)}\frac{1}{\sigma_2^2(1-\rho^2)},$$

$$A_{34} = \frac{b_2}{(b_2+2)}\left[\frac{\rho}{\sigma_2^2(1-\rho^2)}\left(\psi(b_1)-\psi(1)\right)+\frac{1}{\sigma_2^2\sqrt{1-\rho^2}}\left(\psi(b_2+1)-\psi(2)\right)\right],$$

$$A_{35} = -\frac{b_2}{(b_2+2)}\left[\frac{\rho}{\sigma_2(1-\rho^2)^{3/2}}\left(\psi(b_2+1)-\psi(2)\right)-\frac{1}{\sigma_2(1-\rho^2)}\left(\psi(b_1)-\psi(1)\right)\right],$$

$$A_{44} = \frac{1}{\sigma_2^2}\Bigg[1 + \frac{b_2}{(b_2+2)}\Bigg(\frac{\rho^2}{(1-\rho^2)} + \left\{\psi'(b_1) + \psi'(1) + \left(\psi(b_1) - \psi(1)\right)^2\right\}$$

$$+ \frac{2\rho}{\sqrt{1-\rho^2}}\left(\psi(b_1) - \psi(1)\right)\left(\psi(b_2+1) - \psi(2)\right)$$

$$+ \left\{\psi'(b_2+1) + \psi'(2) + \left(\psi(b_2+1) - \psi(2)\right)^2\right\}\Bigg)\Bigg],$$

$$A_{45} = -\frac{1}{\sigma_2}\Bigg[\frac{\rho}{(1-\rho^2)} + \frac{b_2}{(b_2+2)}\Bigg(-\frac{\rho}{(1-\rho^2)} + \left\{\psi'(b_1) + \psi'(1) + \left(\psi(b_1) - \psi(1)\right)^2\right\}$$

$$+ \frac{\rho^2}{(1-\rho^2)^{3/2}}\left(\psi(b_1) - \psi(1)\right)\left(\psi(b_2+1) - \psi(2)\right)$$

$$- \frac{1}{\sqrt{1-\rho^2}}\left(\psi(b_1) - \psi(1)\right)\left(\psi(b_2+1) - \psi(2)\right)$$

$$+ \frac{\rho}{(1-\rho^2)}\left\{\psi'(b_2+1) + \psi'(2) + \left(\psi(b_2+1) - \psi(2)\right)^2\right\}\Bigg)\Bigg]$$

and

$$A_{55} = \frac{\rho^2}{(1-\rho^2)^2} + \frac{b_2}{(b_2+2)}\Bigg(\frac{1}{(1-\rho^2)} + \left\{\psi'(b_1) + \psi'(1) + \left(\psi(b_1) - \psi(1)\right)^2\right\}$$

$$- \frac{2\rho}{(1-\rho^2)^{3/2}}\left(\psi(b_1) - \psi(1)\right)\left(\psi(b_2+1) - \psi(2)\right)$$

$$+ \frac{\rho^2}{(1-\rho^2)^2}\left\{\psi'(b_2+1) + \psi'(2) + \left(\psi(b_2+1) - \psi(2)\right)^2\right\}\Bigg). \qquad (4.3.1)$$

The asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1$, $\hat{\mu}_2$, $\hat{\sigma}_2$ and $\hat{\rho}$ is given by $\boldsymbol{\Sigma} \equiv I^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$. See Appendix D for details.

Hypothesis Testing: Our major interest is testing the null hypothesis $H_0 : \rho = 0$. Since the MML estimators are asymptotically equivalent to the ML estimators, the likelihood function L is maximized (asymptotically) by the MML estimators. Thus, the likelihood ratio is (asymptotically)

$$\hat{\lambda} = \frac{\max(L \mid H_0)}{\max(L \mid H_1)}$$

$$= \frac{\dfrac{1}{\hat{\sigma}_{20}^{n}} \dfrac{e^{-\frac{1}{\hat{\sigma}_{20}}\sum_{i=1}^{n}(y_i-\hat{\mu}_{20})}}{\prod_{i=1}^{n}\left(1+e^{-\frac{y_i-\hat{\mu}_{20}}{\hat{\sigma}_{20}}}\right)^{b+1}}}{\dfrac{1}{\hat{\sigma}_{2}^{n}(1-\hat{\rho}^2)^{n/2}} \dfrac{e^{-\frac{1}{\hat{\sigma}_{2.1}}\sum_{i=1}^{n}(y_i-\hat{\mu}_2-\hat{\theta}_1(x_i-\hat{\mu}_1))}}{\prod_{i=1}^{n}\left(1+e^{-\frac{y_i-\hat{\mu}_2-\hat{\theta}_1(x_i-\hat{\mu}_1)}{\hat{\sigma}_{2.1}}}\right)^{b+1}}}. \qquad (4.3.2)$$

It can be shown that for large n, $\hat{\lambda}$ is a monotonic function of $\hat{\rho}^2$. Thus, to test $H_0 : \rho > 0$, we propose the statistic

$$W = \hat{\rho} \Big/ \sqrt{\frac{1}{n}\frac{(b_2+2)}{b_2}\frac{1}{\left(\psi'(b_1)+\psi'(1)\right)}}$$

$$= \hat{\rho}\sqrt{n\frac{b_2}{(b_2+2)}\left(\psi'(b_1)+\psi'(1)\right)}; \qquad (4.3.3)$$

$\frac{1}{n}\frac{(b_2+2)}{b_2}\frac{1}{\left(\psi'(b_1)+\psi'(1)\right)}$ is the asymptotic variance of $\hat{\rho}$ under H_0, obtained from

$$\left\{I^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)\right\}_{\rho=0}. \qquad (4.3.4)$$

The null distribution of W is referred to normal $N(0,1)$. Large values of W lead to rejection of H_0 against $H_1 : \rho > 0$.

Testing the null hypothesis

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{4.3.5}$$

is also of great practical importance.

Since $\hat{\mu}_1$ and $\hat{\mu}_2$ are asymptotically equivalent to the ML estimators, the distribution of the random vector $\sqrt{n}(\hat{\mu}_1, \hat{\mu}_2)$ is bivariate normal with zero mean vector and (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{13} \\ \hat{\sigma}_{13} & \hat{\sigma}_{33} \end{bmatrix} \tag{4.3.6}$$

where, $\hat{\sigma}_{11}$, $\hat{\sigma}_{13}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij} = \mathbf{\Sigma_{ij}} = I_{ij}^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$. Since in these elements $\hat{\sigma}_1$ and $\hat{\sigma}_2$ converge to σ_1 and σ_2, respectively, the asymptotic null distribution of

$$\hat{\mathrm{T}}^2 = n(\hat{\mu}_1, \hat{\mu}_2)\, \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \tag{4.3.7}$$

is chi-square with 2 df. We reject H_0 at the 5% significance level if the value of $\hat{\mathrm{T}}^2$ is greater than $\chi^2_{0.05}(2)$. The nonnull distribution (asymptotic) of $\hat{\mathrm{T}}^2$ is noncentral chi-square with 2 df and noncentrality parameter

$$\lambda^2 = n(\mu_1, \mu_2)\, \mathbf{\Omega}^{-1} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}. \tag{4.3.8}$$

For small n, the null distribution of $\frac{(n-2)}{2(n-1)}\hat{T}^2$ is approximately central-F with $(2, n-2)$ degrees of freedom and the nonnull distribution is approximately noncentral-F with $(2, n-2)$ degrees of freedom and noncentrality parameter λ^2.

Now, we define the Fisher information matrix $I(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$ for estimating $\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}$ and θ_1. If we let $\boldsymbol{I} = n\,\boldsymbol{A}$, the elements of the matrix $\boldsymbol{A}$ are

$$A_{11} = \frac{b_1}{(b_1+2)}\frac{1}{\sigma_1^2},\ A_{12} = \frac{b_1}{(b_1+2)}\frac{\left(\psi(b_1+1)-\psi(2)\right)}{\sigma_1^2},$$

$$A_{12} = \frac{b_1}{(b_1+2)}\frac{\left(\psi(b_1+1)-\psi(2)\right)}{\sigma_1^2},\ A_{13} = 0,\ A_{14} = 0,\ A_{15} = 0,$$

$$A_{22} = \frac{1}{\sigma_1^2}\left(1+\frac{b_1}{(b_1+2)}\left\{\psi'(b_1+1)+\psi'(2)+\left(\psi(b_1+1)-\psi(2)\right)^2\right\}\right),$$

$$A_{23} = -\frac{b_2}{(b_2+2)}\frac{\theta_1}{\sigma_{2.1}^2}\left(\psi(b_1)-\psi(1)\right),\ A_{24} = 0,\ A_{25} = -0,\ A_{33} = \frac{b_2}{(b_2+2)}\frac{1}{\sigma_{2.1}^2},$$

$$A_{34} = \frac{b_2}{(b_2+2)}\frac{\left(\psi(b_2+1)-\psi(2)\right)}{\sigma_{2.1}^2},\ A_{35} = \frac{b_2}{(b_2+2)}\frac{\sigma_1}{\sigma_{2.1}^2}\left(\psi(b_1)-\psi(1)\right),$$

$$A_{44} = \frac{1}{\sigma_{2.1}^2}\left(1+\frac{b_2}{(b_2+2)}\left\{\psi'(b_2+1)+\psi'(2)+\left(\psi(b_2+1)-\psi(2)\right)^2\right\}\right),$$

$$A_{45} = \frac{b_2}{(b_2+2)}\frac{\sigma_1}{\sigma_{2.1}^2}\left(\psi(b_1)-\psi(1)\right)\left(\psi(b_2+1)-\psi(2)\right)$$

and

$$A_{55} = \frac{b_2}{(b_2+2)}\frac{\sigma_1^2}{\sigma_{2.1}^2}\left\{\psi'(b_1)+\psi'(1)+\left(\psi(b_1)-\psi(1)\right)^2\right\}. \tag{4.3.9}$$

Here, the asymptotic covariance matrix of the estimators $\hat{\mu}_1, \hat{\sigma}_1, \hat{\mu}_{2.1}, \hat{\sigma}_{2.1}$ and $\hat{\theta}_1$ is given by $\boldsymbol{\Sigma} \equiv I^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$. See Appendix D for details.

Hypothesis Testing: We also consider the null hypothesis

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_{2.1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{4.3.10}$$

Since $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$ are asymptotically equivalent to the ML estimators, the distribution of the random vector $\sqrt{n}(\hat{\mu}_1, \hat{\mu}_{2.1})$ is bivariate normal with zero mean vector and (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n \begin{bmatrix} \hat{\sigma}_{11} & 0 \\ 0 & \hat{\sigma}_{33} \end{bmatrix} \tag{4.3.11}$$

where $\hat{\sigma}_{11}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij} = \mathbf{\Sigma_{ij}} = I_{ij}^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$. The covariance between $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$ is zero since they are orthogonal components, so there is no need to estimate it. Since in these elements $\hat{\sigma}_1$ and $\hat{\sigma}_{2.1}$ converge to σ_1 and $\sigma_{2.1}$, respectively, the asymptotic null distribution of

$$\begin{aligned} \hat{\mathrm{T}}_1^2 &= n(\hat{\mu}_1, \hat{\mu}_{2.1}) \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_{2.1} \end{pmatrix} \\ &= \hat{\mu}_1{}^2 / \hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2 / \hat{\sigma}_{33} \end{aligned} \tag{4.3.12}$$

is chi-square with 2 df. We reject H_0 at the 5% significance level if the value of $\hat{\mathrm{T}}_1^2$ is greater than $\chi^2_{0.05}(2)$. The nonnull distribution (asymptotic) of $\hat{\mathrm{T}}_1^2$ is noncentral chi-square with 2 df and noncentrality parameter

$$\lambda^2 = n(\mu_1, \mu_{2.1})\,\mathbf{\Omega}^{-1}\begin{pmatrix}\mu_1\\ \mu_{2.1}\end{pmatrix}$$

$$= \mu_1{}^2 \big/ \sigma_{11} + \mu_{2.1}{}^2 / \sigma_{33}. \qquad (4.3.13)$$

For small n, the null distribution of $\frac{(n-2)}{2(n-1)}\hat{T}_1^2$ is approximately central-F with $(2, n-2)$ degrees of freedom and the nonnull distribution is approximately noncentral-F with $(2, n-2)$ degrees of freedom and noncentrality parameter λ^2.

CHAPTER 5

SIMULATION RESULTS AND ILLUSTRATIVE EXAMPLES

Summary: We give the simulated means and variances of the MML and the LS estimators. We show that the MMLE are enormously more efficient and robust than the LSE. We also show that the test statistics based on the MMLE are more powerful and robust than the corresponding test statistics based on the LSE. For conciseness, we only reproduce the results for the situation when the marginal and the conditional distributions both are Generalized Logistic with shape parameters b_1 and b_2, respectively. The simulated values are in perfect agreement with the theoretical results given in earlier chapters.

5.1 Efficiency of the Estimators

To compare the efficiencies of the MMLE and the LSE, we first note that the former are self bias-correcting. If the mean of the random error is not zero, the LSE needs a bias correction. If the variance of the random error is not equal to $\sigma_{2.1}^2$, the LSE $\tilde{\sigma}_{2.1}$ needs a bias correction. The same applies to the LSE calculated from the marginal sample (see Appendix C for details). We give the simulated relative efficiencies of the LSE, namely, 100 times the ratio of the variance (or mean square error) of the MMLE to that of the corresponding values of the LSE. We give the results only for $\rho = 0.5$. The results are similar for other ρ values, other than the estimators of ρ. We consider numerous values of b_1 and b_2 and sample sizes $n = 15, 30, 60$ and 100. The results are based on 10,000

Monte Carlo runs. Without loss of generality, μ_1, σ_1 , μ_2 and σ_2 are taken to be 0, 1, 0 and 1, respectively. The other parameters take values from the relations $\theta_1 = \rho\frac{\sigma_2}{\sigma_1}$, $\mu_{2.1} = \mu_2 - \theta_1\mu_1$ and $\sigma_{2.1} = \sigma_2\sqrt{(1-\rho^2)}$. From these relations, $\mu_{2.1}$, $\sigma_{2.1}$ and θ_1 take the values 0, 0.866 and 0.5, respectively. The computer program to do the simulations is written in Visual Fortran and given in Appendix E.

The simulated values are given in Table 5.1-5.14. The values of the mean, $n*bias^2$ and $n*variance$ of the MML and LS estimators are given. Also given are the relative efficiencies of the LSE, namely, $100*\text{var}(\hat{\tau})/\text{var}(\tilde{\tau})$ and $100*mse(\hat{\tau})/mse(\tilde{\tau})$. The latter is more relevant if at least one of the estimators $\hat{\tau}$ or $\tilde{\tau}$ have substantial bias since $mse(\hat{\tau}) = var(\hat{\tau}) + bias^2$. In Table 5.1, the simulated values are given for $b_1 = 0.5$, $b_2 = 0.5$ and $\rho = 0.5$ and n=15, 30, 60 and 100. This is a situation when the conditional and the marginal distributions are both negatively skewed. Here, all the estimators are almost unbiased, therefore, the relative efficiencies based on the variances of the estimators are relevant. It is seen that for all sample sizes, $n = 15$, 30, 60 and 100, and for all parameters, the MML estimators are more efficient than the LS estimators. This is an expected result since the MMLE have the Fisher efficiency, at any rate for large n. There is another result of interest, namely, the efficiency of the MMLE increases with n. This is also an expected result since the MMLE are asymptotically the MVB estimators. It is seen that the efficiencies of the LSE of ρ and θ_1 which are very important in the context of regression analysis, are only about 80%. Moreover, their efficiency steadily decrease as n increases. In Table 5.2, we consider $b_1 = 0.5$ and $b_2 = 1$ which is a situation when the marginal distribution is skew but the conditional distribution is symmetric. The value of ρ is taken to be 0.5 in the simulations. It is seen that some LSE have substantial bias even for n=100. The LSE $\tilde{\rho}$ of ρ is highly biased and always overestimates ρ. The relative efficiency of $\tilde{\rho}$ (as compared to the MMLE $\hat{\rho}$) is very low - can be

as low as 19% primarily because of the large bias in $\tilde{\rho}$. Also, $\tilde{\mu}_2$ and $\tilde{\sigma}_2$ have some bias for all sample sizes. Moreover, as n gets large, the bias in $\tilde{\mu}_2$ increases although the bias in $\tilde{\sigma}_2$ decreases. On the other hand, the MML estimators have no substantial bias and they are self bias-correcting. For large n, in fact, they have hardly any bias.

Table 5.1 Simulated values for $b_1 = 0.5$, $b_2 = 0.5$ and $\rho = 0.5$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.046	1.025	-0.029	1.049	-0.052	0.881	0.503	0.485
	n*bias2:	0.031	0.010	0.013	0.037	0.041	0.003	0.000	0.003
	n*variance:	5.356	0.833	6.861	0.743	5.689	0.663	0.859	0.628
	n*mse:	5.387	0.843	6.874	0.779	5.730	0.666	0.859	0.632
LS	mean:	-0.056	0.971	-0.127	0.970	-0.058	0.834	0.497	0.489
	n*bias2:	0.047	0.013	0.243	0.013	0.051	0.015	0.000	0.002
	n*variance:	5.548	0.971	7.125	0.797	6.231	0.735	1.023	0.750
	n*mse:	5.595	0.984	7.368	0.811	6.282	0.750	1.023	0.751
	effvar:	96.5	85.8	96.3	93.1	91.3	90.2	84.0	83.8
	effmse:	96.3	85.7	93.3	96.1	91.2	88.9	84.0	84.1
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.036	1.010	0.010	1.025	-0.007	0.875	0.504	0.493
	n*bias2:	0.038	0.003	0.003	0.019	0.001	0.003	0.001	0.001
	n*variance:	5.066	0.796	6.374	0.677	5.197	0.606	0.688	0.556
	n*mse:	5.104	0.799	6.377	0.696	5.199	0.609	0.688	0.558
LS	mean:	-0.019	0.982	-0.042	0.986	-0.011	0.852	0.501	0.496
	n*bias2:	0.011	0.010	0.054	0.006	0.004	0.006	0.000	0.000
	n*variance:	5.460	1.014	6.935	0.829	5.847	0.780	0.836	0.690
	n*mse:	5.471	1.024	6.989	0.834	5.851	0.786	0.836	0.690
	effvar:	92.8	78.5	91.9	81.7	88.9	77.7	82.3	80.7
	effmse:	93.3	78.0	91.2	83.4	88.9	77.5	82.3	80.8
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.021	1.007	0.003	1.011	-0.007	0.869	0.501	0.498
	n*bias2:	0.025	0.003	0.001	0.007	0.003	0.000	0.000	0.000
	n*variance:	5.118	0.786	6.254	0.652	5.062	0.600	0.629	0.563
	n*mse:	5.143	0.789	6.255	0.659	5.065	0.600	0.629	0.563
LS	mean:	-0.007	0.993	-0.024	0.992	-0.010	0.856	0.501	0.500
	n*bias2:	0.003	0.003	0.033	0.004	0.005	0.006	0.000	0.000
	n*variance:	5.572	1.057	6.945	0.842	5.758	0.796	0.803	0.732
	n*mse:	5.575	1.060	6.978	0.845	5.763	0.801	0.803	0.732
	effvar:	91.9	74.4	90.1	77.5	87.9	75.4	78.3	76.9
	effmse:	92.3	74.4	89.6	78.0	87.9	74.9	78.3	77.0
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.013	1.003	0.000	1.006	-0.006	0.868	0.500	0.498
	n*bias2:	0.017	0.001	0.000	0.004	0.004	0.000	0.000	0.000
	n*variance:	5.050	0.751	6.330	0.634	4.993	0.584	0.620	0.566
	n*mse:	5.067	0.752	6.330	0.638	4.997	0.585	0.620	0.566
LS	mean:	-0.004	0.993	-0.016	0.995	-0.008	0.861	0.499	0.498
	n*bias2:	0.002	0.004	0.025	0.003	0.006	0.003	0.000	0.000
	n*variance:	5.550	1.041	7.153	0.855	5.719	0.802	0.797	0.744
	n*mse:	5.551	1.045	7.178	0.858	5.725	0.805	0.797	0.744
	effvar:	91.0	72.1	88.5	74.2	87.3	72.8	77.8	76.1
	effmse:	91.3	71.9	88.2	74.4	87.3	72.6	77.8	76.1

In Table 5.1, the shape parameters b_1 and b_2 are both equal to 0.5. In Table 5.2-5.5, we have given results for numerous other combinations of b_1 and b_2. The same comments apply as for the values in Table 5.1. We conclude, therefore, that the MMLE are enormously more efficient than the LSE.

Table 5.2 Simulated values for $b_1 = 0.5$, $b_2 = 1$ and $\rho = 0.5$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.058	1.021	0.029	1.041	0.001	0.882	0.501	0.491
	n*bias2:	0.051	0.007	0.013	0.025	0.000	0.004	0.000	0.001
	n*variance:	5.154	0.865	4.413	0.640	3.269	0.629	0.494	0.452
	n*mse:	5.205	0.872	4.426	0.665	3.269	0.633	0.494	0.454
LS	mean:	-0.044	0.967	0.090	0.920	0.002	0.840	0.502	0.614
	n*bias2:	0.029	0.017	0.122	0.095	0.000	0.010	0.000	0.196
	n*variance:	5.371	0.978	4.618	0.540	3.495	0.604	0.516	0.548
	n*mse:	5.400	0.994	4.740	0.636	3.495	0.614	0.516	0.744
	effvar:	96.0	88.5	95.6	118.4	93.5	104.2	95.9	82.6
	effmse:	96.4	87.7	93.4	104.5	93.5	103.0	95.9	60.9
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.030	1.009	0.017	1.022	0.002	0.878	0.501	0.494
	n*bias2:	0.028	0.002	0.008	0.015	0.000	0.004	0.000	0.001
	n*variance:	5.174	0.806	4.372	0.591	3.116	0.581	0.417	0.424
	n*mse:	5.201	0.808	4.381	0.606	3.116	0.585	0.417	0.425
LS	mean:	-0.023	0.979	0.113	0.935	0.001	0.857	0.500	0.620
	n*bias2:	0.016	0.013	0.382	0.127	0.000	0.003	0.000	0.429
	n*variance:	5.558	1.001	4.660	0.548	3.354	0.606	0.444	0.513
	n*mse:	5.574	1.014	5.041	0.675	3.354	0.609	0.444	0.942
	effvar:	93.1	80.5	93.8	107.8	92.9	95.8	94.0	82.7
	effmse:	93.3	79.7	86.9	89.8	92.9	96.1	94.0	45.1
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.012	1.006	0.009	1.010	0.003	0.870	0.501	0.499
	n*bias2:	0.008	0.002	0.004	0.006	0.001	0.001	0.000	0.000
	n*variance:	5.202	0.794	4.237	0.547	2.978	0.536	0.373	0.412
	n*mse:	5.210	0.796	4.241	0.552	2.979	0.536	0.373	0.412
LS	mean:	-0.015	0.992	0.124	0.939	0.002	0.859	0.501	0.628
	n*bias2:	0.014	0.004	0.916	0.222	0.000	0.003	0.000	0.987
	n*variance:	5.666	1.059	4.608	0.540	3.221	0.592	0.404	0.503
	n*mse:	5.680	1.063	5.524	0.762	3.222	0.595	0.404	1.490
	effvar:	91.8	75.0	91.9	101.2	92.4	90.5	92.2	81.9
	effmse:	91.7	74.8	76.8	72.5	92.5	90.2	92.2	27.7
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.008	1.002	0.005	1.005	0.001	0.868	0.500	0.499
	n*bias2:	0.007	0.001	0.002	0.003	0.000	0.001	0.000	0.000
	n*variance:	4.966	0.771	4.203	0.544	2.954	0.532	0.372	0.411
	n*mse:	4.973	0.771	4.206	0.547	2.954	0.532	0.372	0.411
LS	mean:	-0.009	0.994	0.128	0.941	0.000	0.862	0.500	0.629
	n*bias2:	0.009	0.004	1.629	0.345	0.000	0.002	0.000	1.669
	n*variance:	5.429	1.041	4.591	0.544	3.240	0.595	0.401	0.501
	n*mse:	5.438	1.045	6.220	0.889	3.240	0.596	0.401	2.170
	effvar:	91.5	74.0	91.6	100.0	91.2	89.4	92.6	82.0
	effmse:	91.5	73.8	67.6	61.5	91.2	89.3	92.6	19.0

Table 5.3 Simulated values for $b_1 = 1$, $b_2 = 0.5$ and $\rho = 0.5$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.008	1.024	-0.045	1.064	-0.049	0.882	0.500	0.460
	n*bias2:	0.001	0.009	0.031	0.062	0.035	0.004	0.000	0.024
	n*variance:	3.079	0.763	4.871	0.927	4.405	0.659	1.673	1.021
	n*mse:	3.080	0.772	4.902	0.989	4.440	0.663	1.673	1.045
LS	mean:	0.009	0.975	-0.027	0.960	-0.046	0.837	0.500	0.373
	n*bias2:	0.001	0.010	0.011	0.024	0.031	0.013	0.000	0.241
	n*variance:	3.344	0.743	5.121	0.883	4.648	0.727	1.993	0.862
	n*mse:	3.346	0.753	5.132	0.908	4.680	0.741	1.993	1.103
	effvar:	92.1	102.7	95.1	104.9	94.8	90.6	83.9	118.5
	effmse:	92.1	102.5	95.5	109.0	94.9	89.5	83.9	94.8

n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.000	1.012	-0.022	1.032	-0.021	0.875	0.498	0.479
	n*bias2:	0.000	0.005	0.014	0.030	0.014	0.003	0.000	0.013
	n*variance:	3.015	0.735	4.788	0.852	4.153	0.617	1.417	0.981
	n*mse:	3.015	0.740	4.802	0.882	4.167	0.620	1.417	0.994
LS	mean:	0.000	0.987	0.010	0.966	-0.021	0.851	0.499	0.377
	n*bias2:	0.000	0.005	0.003	0.036	0.014	0.007	0.000	0.456
	n*variance:	3.298	0.770	5.143	0.903	4.516	0.761	1.742	0.862
	n*mse:	3.298	0.776	5.146	0.939	4.530	0.768	1.742	1.317
	effvar:	91.4	95.4	93.1	94.3	92.0	81.0	81.4	113.8
	effmse:	91.4	95.4	93.3	93.9	92.0	80.6	81.4	75.5

n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.004	1.006	-0.008	1.016	-0.010	0.870	0.501	0.492
	n*bias2:	0.001	0.003	0.004	0.015	0.006	0.001	0.000	0.004
	n*variance:	2.996	0.709	4.804	0.795	4.035	0.605	1.289	0.935
	n*mse:	2.997	0.711	4.808	0.810	4.041	0.606	1.289	0.939
LS	mean:	0.004	0.993	0.031	0.969	-0.011	0.858	0.502	0.380
	n*bias2:	0.001	0.003	0.056	0.059	0.007	0.004	0.000	0.866
	n*variance:	3.299	0.774	5.199	0.911	4.374	0.815	1.617	0.855
	n*mse:	3.300	0.777	5.255	0.970	4.381	0.819	1.617	1.722
	effvar:	90.8	91.5	92.4	87.3	92.3	74.3	79.7	109.3
	effmse:	90.8	91.6	91.5	83.5	92.2	74.0	79.7	54.6

n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.000	1.004	-0.006	1.009	-0.006	0.868	0.501	0.496
	n*bias2:	0.000	0.002	0.004	0.009	0.004	0.000	0.000	0.002
	n*variance:	2.983	0.702	4.577	0.775	3.870	0.587	1.206	0.876
	n*mse:	2.983	0.704	4.581	0.784	3.874	0.587	1.206	0.878
LS	mean:	0.000	0.997	0.035	0.969	-0.007	0.861	0.502	0.379
	n*bias2:	0.000	0.001	0.122	0.094	0.005	0.003	0.000	1.453
	n*variance:	3.264	0.780	5.018	0.921	4.246	0.820	1.560	0.837
	n*mse:	3.264	0.781	5.140	1.015	4.251	0.823	1.561	2.290
	effvar:	91.4	90.0	91.2	84.1	91.1	71.6	77.3	104.6
	effmse:	91.4	90.1	89.1	77.2	91.1	71.4	77.3	38.3

Table 5.4 Simulated values for $b_1 = 1$, $b_2 = 1$ and $\rho = 0.5$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.006	1.025	-0.002	1.051	0.001	0.883	0.499	0.478
	n*bias2:	0.001	0.009	0.000	0.039	0.000	0.004	0.000	0.007
	n*variance:	3.032	0.786	3.140	0.754	2.530	0.639	0.950	0.680
	n*mse:	3.033	0.795	3.140	0.793	2.530	0.644	0.950	0.688
LS	mean:	-0.005	0.975	-0.001	0.975	0.001	0.841	0.498	0.488
	n*bias2:	0.000	0.009	0.000	0.009	0.000	0.010	0.000	0.002
	n*variance:	3.260	0.763	3.360	0.691	2.704	0.612	0.984	0.720
	n*mse:	3.260	0.772	3.360	0.700	2.704	0.621	0.984	0.722
	effvar:	93.0	102.9	93.5	109.1	93.6	104.6	96.6	94.5
	effmse:	93.0	102.9	93.5	113.3	93.6	103.6	96.6	95.2
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.001	1.013	-0.005	1.025	-0.004	0.875	0.500	0.490
	n*bias2:	0.000	0.005	0.001	0.018	0.000	0.002	0.000	0.003
	n*variance:	3.062	0.749	2.985	0.651	2.330	0.564	0.792	0.627
	n*mse:	3.062	0.754	2.986	0.669	2.330	0.566	0.792	0.630
LS	mean:	0.000	0.987	-0.005	0.988	-0.004	0.853	0.500	0.495
	n*bias2:	0.000	0.005	0.001	0.004	0.001	0.005	0.000	0.001
	n*variance:	3.315	0.773	3.224	0.657	2.519	0.586	0.842	0.674
	n*mse:	3.315	0.778	3.225	0.662	2.519	0.591	0.842	0.675
	effvar:	92.4	97.0	92.6	99.1	92.5	96.2	94.1	93.1
	effmse:	92.4	97.0	92.6	101.2	92.5	95.8	94.1	93.4
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.000	1.008	-0.002	1.014	-0.001	0.873	0.499	0.494
	n*bias2:	0.000	0.004	0.000	0.012	0.000	0.003	0.000	0.003
	n*variance:	3.068	0.729	3.012	0.642	2.274	0.558	0.762	0.625
	n*mse:	3.068	0.733	3.012	0.654	2.274	0.561	0.762	0.627
LS	mean:	0.001	0.996	-0.002	0.996	-0.002	0.862	0.499	0.496
	n*bias2:	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
	n*variance:	3.353	0.808	3.275	0.690	2.479	0.606	0.829	0.683
	n*mse:	3.353	0.809	3.276	0.691	2.479	0.606	0.829	0.684
	effvar:	91.5	90.3	92.0	93.0	91.7	92.2	92.0	91.4
	effmse:	91.5	90.6	92.0	94.6	91.7	92.5	92.0	91.7
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.003	1.005	-0.002	1.005	0.000	0.867	0.499	0.498
	n*bias2:	0.001	0.002	0.000	0.003	0.000	0.000	0.000	0.001
	n*variance:	2.991	0.710	3.051	0.617	2.306	0.517	0.714	0.592
	n*mse:	2.991	0.712	3.051	0.620	2.306	0.517	0.714	0.592
LS	mean:	-0.003	0.997	-0.002	0.995	0.000	0.860	0.499	0.499
	n*bias2:	0.001	0.001	0.000	0.003	0.000	0.003	0.000	0.000
	n*variance:	3.304	0.807	3.308	0.673	2.482	0.573	0.780	0.657
	n*mse:	3.305	0.808	3.308	0.676	2.482	0.577	0.780	0.657
	effvar:	90.5	88.0	92.2	91.6	92.9	90.2	91.5	90.1
	effmse:	90.5	88.1	92.2	91.6	92.9	89.7	91.5	90.2

Table 5.5 Simulated values for $b_1 = 4$, $b_2 = 4$ and $\rho = 0.5$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.029	1.004	0.033	1.032	0.047	0.866	0.508	0.486
	n*bias2:	0.012	0.000	0.016	0.016	0.033	0.000	0.001	0.003
	n*variance:	2.438	0.685	4.925	0.648	4.591	0.554	0.864	0.608
	n*mse:	2.450	0.685	4.942	0.664	4.623	0.554	0.865	0.611
LS	mean:	0.051	0.973	0.129	0.978	0.048	0.842	0.500	0.489
	n*bias2:	0.039	0.011	0.250	0.007	0.034	0.009	0.000	0.002
	n*variance:	2.728	0.850	5.319	0.736	5.361	0.667	1.036	0.735
	n*mse:	2.766	0.861	5.568	0.744	5.395	0.676	1.036	0.737
	effvar:	89.4	80.6	92.6	88.1	85.6	83.0	83.4	82.8
	effmse:	88.6	79.6	88.7	89.3	85.7	82.0	83.5	82.9
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.024	1.003	0.001	1.017	0.013	0.865	0.510	0.498
	n*bias2:	0.017	0.000	0.000	0.009	0.005	0.000	0.003	0.000
	n*variance:	2.267	0.662	4.581	0.601	4.078	0.516	0.710	0.548
	n*mse:	2.283	0.662	4.581	0.610	4.083	0.516	0.713	0.548
LS	mean:	0.027	0.984	0.059	0.988	0.018	0.850	0.506	0.499
	n*bias2:	0.022	0.007	0.105	0.005	0.010	0.007	0.001	0.000
	n*variance:	2.630	0.868	5.277	0.743	4.942	0.665	0.876	0.685
	n*mse:	2.652	0.876	5.381	0.748	4.952	0.672	0.877	0.685
	effvar:	86.2	76.2	86.8	80.9	82.5	77.5	81.1	80.0
	effmse:	86.1	75.6	85.1	81.6	82.5	76.7	81.3	80.1
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.018	1.003	0.000	1.009	0.009	0.867	0.503	0.498
	n*bias2:	0.019	0.000	0.000	0.005	0.005	0.000	0.000	0.000
	n*variance:	2.342	0.628	4.462	0.568	3.879	0.490	0.633	0.513
	n*mse:	2.361	0.629	4.462	0.573	3.884	0.490	0.633	0.513
LS	mean:	0.010	0.994	0.031	0.995	0.013	0.861	0.500	0.497
	n*bias2:	0.006	0.002	0.058	0.002	0.010	0.002	0.000	0.000
	n*variance:	2.849	0.888	5.398	0.760	4.810	0.683	0.800	0.675
	n*mse:	2.855	0.891	5.456	0.761	4.819	0.684	0.800	0.675
	effvar:	82.2	70.7	82.7	74.8	80.6	71.8	79.1	76.1
	effmse:	82.7	70.6	81.8	75.2	80.6	71.6	79.2	76.0
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.009	1.001	0.003	1.004	0.007	0.866	0.501	0.498
	n*bias2:	0.008	0.000	0.001	0.002	0.005	0.000	0.000	0.000
	n*variance:	2.264	0.646	4.381	0.577	3.817	0.502	0.625	0.526
	n*mse:	2.272	0.646	4.382	0.578	3.822	0.502	0.625	0.526
LS	mean:	0.010	0.995	0.023	0.995	0.010	0.862	0.499	0.497
	n*bias2:	0.010	0.002	0.052	0.002	0.010	0.002	0.000	0.001
	n*variance:	2.755	0.911	5.397	0.786	4.801	0.708	0.787	0.689
	n*mse:	2.765	0.914	5.450	0.789	4.811	0.710	0.787	0.690
	effvar:	82.2	70.9	81.2	73.3	79.5	70.9	79.5	76.3
	effmse:	82.2	70.8	80.4	73.3	79.4	70.7	79.4	76.3

Robustness: Another important issue is that of robustness. An estimator is called robust if it is fully efficient (or nearly so) for an assumed model but maintains high efficiency for plausible alternatives. To verify the robustness of the MMLE and the LSE, we consider the situation most favourable to the LSE, namely, $b_1 = 1$ and $b_2 = 1$ (i.e., marginal and the conditional distributions are both logistic). As alternatives, we consider the following:

1) outlier model: $(n-r)GL(\mu_1,\sigma_1)$ and $rGL(\mu_1,4\sigma_1)$, $r=[0.5+0.1n]$

2) mixture model: $0.9GL(\mu_1,\sigma_1)+0.1GL(\mu_1,4\sigma_1)$

3) contamination model: $0.9GL(\mu_1,\sigma_1)+0.1U(\mu_1-\sigma_1/2,\mu_1+\sigma_1/2)$.

In the first model, $(n-r)$ observations (X's) come from $GL(\mu_1,\sigma_1)$ and r (we do not know which) come from $GL(\mu_1,4\sigma_1)$. In the second model, with 0.1 probability X's come from a Generalized Logistic distribution with mean μ_1 and variance $16\sigma_1^2$. In the third model, with 0.1 probability X's come from a uniform distribution with support $\mu_1-\sigma_1/2$ to $\mu_1+\sigma_1/2$. Without loss of generality, μ_1 is taken to be zero and σ_1 is taken to be 1.

For illustration, we give the simulated means, biases, variances and the relative efficiencies of LSE for $\rho=0.2,0.5,0.9$ and $n=15,30,60$ and 100.

It is important to mention here that the efficiencies based on the mse are not valid for the estimators of σ_1 and ρ since the variance of X is not as before for the outlier, mixture or contamination models. Consequently, the true values of σ_1 and ρ are shifted. Therefore, the mean square errors which are calculated by using the true values of the estimators are not valid anymore. The efficiencies based on the variances of the estimators, however, can still give an idea about the merits of the estimators. For the outlier model, the MML estimators are highly efficient as compared to the LSE. This is due to the fact that in the calculation of the MMLE, the extreme observations automatically receive small weights. Thus, the influence of these observations is depleted giving the MMLE the intrinsic robustness property they possess.

Table 5.6 Simulated values for the outlier model for $b_1 = 1$, $b_2 = 1$ and $\rho = 0.2$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.003	1.578	-0.007	1.085	-0.004	1.002	0.200	0.284
	n*bias2:	0.000	5.008	0.001	0.109	0.000	0.007	0.000	0.107
	n*variance:	5.311	4.297	3.156	0.841	3.142	0.802	0.577	0.877
	n*mse:	5.311	9.305	3.157	0.950	3.142	0.810	0.577	0.984
LS	mean:	-0.006	1.618	-0.007	1.015	-0.003	0.955	0.200	0.308
	n*bias2:	0.001	5.722	0.001	0.003	0.000	0.009	0.000	0.175
	n*variance:	9.677	6.122	3.507	0.797	3.357	0.771	0.595	1.023
	n*mse:	9.678	11.844	3.508	0.801	3.358	0.780	0.595	1.198
	effvar:	54.9	70.2	90.0	105.4	93.6	104.1	97.0	85.7
	effmse:	54.9	78.6	90.0	118.6	93.6	103.8	97.0	82.1
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.002	1.395	-0.001	1.043	-0.002	0.989	0.199	0.264
	n*bias2:	0.000	4.676	0.000	0.054	0.000	0.002	0.000	0.124
	n*variance:	4.352	2.984	3.021	0.755	2.916	0.737	0.528	0.829
	n*mse:	4.352	7.660	3.021	0.809	2.916	0.739	0.528	0.953
LS	mean:	0.006	1.511	-0.002	1.015	-0.004	0.965	0.199	0.292
	n*bias2:	0.001	7.833	0.000	0.006	0.000	0.007	0.000	0.253
	n*variance:	8.175	6.208	3.397	0.803	3.159	0.762	0.558	1.050
	n*mse:	8.176	14.042	3.397	0.810	3.159	0.769	0.558	1.303
	effvar:	53.2	48.1	88.9	94.0	92.3	96.6	94.6	79.0
	effmse:	53.2	54.5	88.9	100.0	92.3	96.1	94.6	73.2
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.001	1.362	0.000	1.028	0.000	0.984	0.200	0.265
	n*bias2:	0.000	7.859	0.000	0.046	0.000	0.001	0.000	0.250
	n*variance:	4.214	2.537	3.106	0.709	2.979	0.703	0.450	0.762
	n*mse:	4.214	10.396	3.106	0.755	2.979	0.704	0.450	1.012
LS	mean:	-0.003	1.541	-0.001	1.022	0.001	0.972	0.200	0.299
	n*bias2:	0.000	17.575	0.000	0.028	0.000	0.004	0.000	0.592
	n*variance:	8.157	6.906	3.486	0.810	3.220	0.772	0.483	1.071
	n*mse:	8.157	24.481	3.486	0.838	3.220	0.776	0.483	1.663
	effvar:	51.7	36.7	89.1	87.5	92.5	91.1	93.3	71.1
	effmse:	51.7	42.5	89.1	90.1	92.5	90.8	93.3	60.8
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.003	1.348	0.001	1.023	0.002	0.983	0.200	0.263
	n*bias2:	0.001	12.139	0.000	0.052	0.000	0.001	0.000	0.396
	n*variance:	4.051	2.358	3.072	0.707	2.934	0.697	0.413	0.705
	n*mse:	4.052	14.498	3.072	0.759	2.934	0.698	0.413	1.101
LS	mean:	-0.007	1.559	0.000	1.026	0.002	0.976	0.200	0.302
	n*bias2:	0.005	31.214	0.000	0.066	0.000	0.001	0.000	1.040
	n*variance:	8.137	7.546	3.514	0.831	3.216	0.780	0.442	1.043
	n*mse:	8.142	38.760	3.514	0.896	3.217	0.781	0.442	2.083
	effvar:	49.8	31.3	87.4	85.1	91.2	89.4	93.5	67.6
	effmse:	49.8	37.4	87.4	84.7	91.2	89.4	93.5	52.9

It is also seen from Tables 5.6-5.14 that as the correlation between X and Y gets larger in magnitude, the efficiency of the MMLE increases, particularly for estimating μ_2 and σ_2. The reason for the LSE to have low efficiency for large values of ρ is that the LS estimators of μ_2 and σ_2 use only the $y-$observations but they do not use the $x-$observations. When the correlation

coefficient ρ is large, X contains considerable information about the parameters of Y. Tables 5.6–5.8 show the simulated values of the estimators for three different values of ρ for the outlier model. It can be seen that the MML estimators are superior in terms of efficiency.

Table 5.7 Simulated values for the outlier model for $b_1 = 1$, $b_2 = 1$ and $\rho = 0.5$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.003	1.563	-0.006	1.215	-0.006	0.884	0.500	0.624
	n*bias2:	0.000	4.753	0.001	0.691	0.001	0.005	0.000	0.232
	n*variance:	5.522	4.267	3.785	1.191	2.511	0.634	0.433	0.536
	n*mse:	5.522	9.021	3.786	1.881	2.511	0.639	0.433	0.768
LS	mean:	0.002	1.601	-0.006	1.179	-0.006	0.843	0.500	0.652
	n*bias2:	0.000	5.421	0.001	0.481	0.001	0.008	0.000	0.347
	n*variance:	9.881	6.073	5.011	1.496	2.666	0.609	0.449	0.579
	n*mse:	9.881	11.494	5.012	1.977	2.667	0.617	0.449	0.926
	effvar:	55.9	70.3	75.5	79.6	94.2	104.1	96.5	92.6
	effmse:	55.9	78.5	75.5	95.2	94.2	103.5	96.5	83.0
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.005	1.396	-0.003	1.135	-0.001	0.873	0.501	0.608
	n*bias2:	0.001	4.701	0.000	0.547	0.000	0.002	0.000	0.352
	n*variance:	4.383	2.961	3.381	0.904	2.378	0.569	0.407	0.539
	n*mse:	4.384	7.662	3.382	1.451	2.378	0.571	0.407	0.891
LS	mean:	-0.010	1.516	-0.006	1.152	-0.001	0.853	0.501	0.644
	n*bias2:	0.003	7.982	0.001	0.696	0.000	0.005	0.000	0.626
	n*variance:	8.400	6.116	4.497	1.416	2.548	0.594	0.427	0.646
	n*mse:	8.403	14.098	4.498	2.112	2.548	0.600	0.427	1.272
	effvar:	52.2	48.4	75.2	63.9	93.3	95.7	95.3	83.4
	effmse:	52.2	54.3	75.2	68.7	93.3	95.1	95.4	70.1
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.001	1.364	0.000	1.114	0.001	0.871	0.501	0.610
	n*bias2:	0.000	7.930	0.000	0.780	0.000	0.001	0.000	0.725
	n*variance:	4.188	2.620	3.227	0.810	2.223	0.552	0.342	0.491
	n*mse:	4.188	10.550	3.227	1.589	2.223	0.554	0.343	1.216
LS	mean:	0.002	1.544	0.002	1.163	0.001	0.860	0.502	0.657
	n*bias2:	0.000	17.784	0.000	1.603	0.000	0.002	0.000	1.484
	n*variance:	8.206	7.093	4.392	1.508	2.433	0.600	0.363	0.652
	n*mse:	8.207	24.877	4.392	3.111	2.434	0.602	0.363	2.137
	effvar:	51.0	36.9	73.5	53.7	91.4	92.0	94.3	75.3
	effmse:	51.0	42.4	73.5	51.1	91.4	91.9	94.3	56.9
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.002	1.347	0.000	1.104	-0.001	0.869	0.500	0.609
	n*bias2:	0.000	12.064	0.000	1.086	0.000	0.001	0.000	1.182
	n*variance:	4.099	2.387	3.312	0.762	2.257	0.543	0.320	0.475
	n*mse:	4.099	14.451	3.312	1.847	2.257	0.544	0.320	1.657
LS	mean:	0.004	1.557	0.001	1.167	-0.001	0.863	0.501	0.663
	n*bias2:	0.002	31.076	0.000	2.787	0.000	0.001	0.000	2.642
	n*variance:	8.223	7.579	4.570	1.551	2.488	0.599	0.346	0.676
	n*mse:	8.225	38.655	4.570	4.339	2.489	0.599	0.346	3.318
	effvar:	49.8	31.5	72.5	49.1	90.7	90.7	92.5	70.4
	effmse:	49.8	37.4	72.5	42.6	90.7	90.7	92.5	49.9

Table 5.8 Simulated values for the outlier model $b_1 = 1$, $b_2 = 1$ and $\rho = 0.9$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.002	1.585	0.002	1.506	0.000	0.448	0.900	0.937
	n*bias2:	0.000	5.126	0.000	3.834	0.000	0.002	0.000	0.021
	n*variance:	5.431	4.506	5.039	3.527	0.622	0.159	0.112	0.038
	n*mse:	5.431	9.632	5.039	7.361	0.622	0.161	0.112	0.059
LS	mean:	0.002	1.626	0.002	1.531	0.001	0.427	0.899	0.944
	n*bias2:	0.000	5.876	0.000	4.231	0.000	0.001	0.000	0.029
	n*variance:	9.931	6.460	8.716	5.077	0.670	0.154	0.116	0.037
	n*mse:	9.931	12.336	8.716	9.307	0.670	0.155	0.116	0.066
	effvar:	54.7	69.7	57.8	69.5	92.9	103.0	96.4	104.1
	effmse:	54.7	78.1	57.8	79.1	92.9	103.6	96.4	89.4
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.005	1.398	0.005	1.338	0.001	0.442	0.899	0.935
	n*bias2:	0.001	4.743	0.001	3.423	0.000	0.001	0.000	0.037
	n*variance:	4.461	3.064	4.191	2.395	0.589	0.145	0.103	0.034
	n*mse:	4.462	7.807	4.192	5.818	0.589	0.146	0.103	0.071
LS	mean:	0.008	1.516	0.008	1.435	0.001	0.431	0.899	0.944
	n*bias2:	0.002	8.002	0.002	5.680	0.000	0.001	0.000	0.058
	n*variance:	8.316	6.404	7.366	4.982	0.633	0.151	0.108	0.036
	n*mse:	8.318	14.406	7.368	10.661	0.633	0.152	0.108	0.094
	effvar:	53.6	47.8	56.9	48.1	92.9	95.8	95.9	94.8
	effmse:	53.6	54.2	56.9	54.6	92.9	96.0	95.9	75.7
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.003	1.361	0.001	1.304	−0.001	0.438	0.900	0.938
	n*bias2:	0.000	7.831	0.000	5.539	0.000	0.000	0.000	0.086
	n*variance:	4.214	2.535	3.980	1.963	0.579	0.139	0.087	0.029
	n*mse:	4.215	10.366	3.980	7.502	0.579	0.139	0.087	0.115
LS	mean:	0.001	1.541	0.000	1.456	−0.001	0.432	0.900	0.950
	n*bias2:	0.000	17.580	0.000	12.460	0.000	0.001	0.000	0.147
	n*variance:	8.177	6.892	7.254	5.297	0.622	0.151	0.093	0.031
	n*mse:	8.177	24.471	7.254	17.757	0.622	0.152	0.093	0.179
	effvar:	51.5	36.8	54.9	37.1	93.0	91.9	93.5	91.8
	effmse:	51.5	42.4	54.9	42.3	93.0	91.6	93.5	64.1
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	−0.001	1.345	0.000	1.288	0.001	0.437	0.900	0.938
	n*bias2:	0.000	11.904	0.000	8.315	0.000	0.000	0.000	0.147
	n*variance:	4.172	2.358	3.967	1.820	0.573	0.140	0.079	0.027
	n*mse:	4.172	14.262	3.967	10.135	0.574	0.140	0.079	0.174
LS	mean:	−0.003	1.555	−0.002	1.466	0.001	0.434	0.900	0.952
	n*bias2:	0.001	30.752	0.000	21.732	0.000	0.001	0.000	0.269
	n*variance:	8.187	7.367	7.255	5.630	0.625	0.156	0.085	0.030
	n*mse:	8.188	38.119	7.255	27.362	0.625	0.157	0.085	0.299
	effvar:	51.0	32.0	54.7	32.3	91.7	89.9	92.8	89.4
	effmse:	51.0	37.4	54.7	37.0	91.7	89.7	92.8	58.2

For the mixture model, we have results similar to those for the outlier model. This is evident from the values given in Tables 5.9-5.11. What is true in general is that the MML estimators become more efficient than the LS estimators as n becomes large. In fact, this is a very striking result since it is generally perceived that the LS estimators are highly efficient for large n.

Table 5.9 Simulated values for the mixture model for $b_1 = 1$, $b_2 = 1$ and $\rho = 0.2$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.004	1.416	0.005	1.074	0.004	0.999	0.199	0.258
	n*bias2:	0.000	2.597	0.000	0.081	0.000	0.006	0.000	0.050
	n*variance:	4.775	4.455	3.235	0.814	3.242	0.784	0.732	0.902
	n*mse:	4.775	7.052	3.235	0.895	3.242	0.789	0.732	0.952
LS	mean:	0.003	1.428	0.006	1.000	0.006	0.952	0.199	0.275
	n*bias2:	0.000	2.746	0.001	0.000	0.000	0.011	0.000	0.085
	n*variance:	7.893	5.987	3.553	0.770	3.469	0.757	0.761	1.036
	n*mse:	7.893	8.734	3.554	0.770	3.470	0.768	0.761	1.121
	effvar:	60.5	74.4	91.0	105.7	93.5	103.6	96.2	87.0
	effmse:	60.5	80.7	91.0	116.2	93.5	102.8	96.2	84.9
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.000	1.389	0.000	1.043	0.000	0.989	0.199	0.263
	n*bias2:	0.000	4.528	0.000	0.054	0.000	0.002	0.000	0.117
	n*variance:	4.462	4.138	3.174	0.756	3.101	0.737	0.553	0.847
	n*mse:	4.462	8.667	3.174	0.810	3.101	0.739	0.553	0.964
LS	mean:	0.000	1.493	0.000	1.014	0.000	0.965	0.198	0.287
	n*bias2:	0.000	7.305	0.000	0.006	0.000	0.006	0.000	0.227
	n*variance:	8.236	7.645	3.549	0.807	3.347	0.771	0.587	1.067
	n*mse:	8.236	14.950	3.549	0.812	3.347	0.777	0.587	1.293
	effvar:	54.2	54.1	89.4	93.8	92.7	95.5	94.2	79.4
	effmse:	54.2	58.0	89.4	99.8	92.7	95.1	94.2	74.6
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.002	1.366	0.001	1.029	0.001	0.985	0.200	0.265
	n*bias2:	0.000	8.021	0.000	0.052	0.000	0.002	0.000	0.251
	n*variance:	4.323	3.578	3.051	0.709	2.950	0.698	0.439	0.754
	n*mse:	4.323	11.599	3.051	0.761	2.950	0.700	0.439	1.005
LS	mean:	0.003	1.539	0.001	1.023	0.001	0.973	0.200	0.299
	n*bias2:	0.000	17.437	0.000	0.031	0.000	0.003	0.000	0.583
	n*variance:	8.415	8.461	3.442	0.801	3.194	0.755	0.469	1.050
	n*mse:	8.416	25.898	3.443	0.832	3.194	0.758	0.469	1.633
	effvar:	51.4	42.3	88.6	88.5	92.3	92.4	93.6	71.8
	effmse:	51.4	44.8	88.6	91.5	92.3	92.3	93.6	61.5
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.001	1.347	-0.002	1.023	-0.002	0.983	0.199	0.262
	n*bias2:	0.000	12.068	0.000	0.053	0.000	0.001	0.000	0.386
	n*variance:	4.203	3.391	3.057	0.696	2.894	0.685	0.429	0.766
	n*mse:	4.203	15.459	3.058	0.750	2.895	0.686	0.429	1.152
LS	mean:	-0.001	1.552	-0.002	1.025	-0.002	0.975	0.199	0.301
	n*bias2:	0.000	30.433	0.001	0.064	0.000	0.002	0.000	1.010
	n*variance:	8.314	9.257	3.451	0.827	3.153	0.768	0.458	1.136
	n*mse:	8.314	39.690	3.452	0.891	3.153	0.770	0.458	2.146
	effvar:	50.6	36.6	88.6	84.2	91.8	89.3	93.6	67.4
	effmse:	50.6	38.9	88.6	84.2	91.8	89.2	93.6	53.7

Table 5.10 Simulated values for the mixture model for $b_1 = 1$, $b_2 = 1$ and $\rho = 0.5$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.005	1.424	0.002	1.176	0.000	0.883	0.501	0.585
	n*bias2:	0.000	2.695	0.000	0.465	0.000	0.004	0.000	0.108
	n*variance:	4.913	4.663	3.607	1.235	2.535	0.626	0.629	0.677
	n*mse:	4.914	7.358	3.607	1.700	2.535	0.630	0.629	0.785
LS	mean:	0.007	1.435	0.003	1.129	0.001	0.842	0.501	0.607
	n*bias2:	0.001	2.839	0.000	0.248	0.000	0.009	0.000	0.173
	n*variance:	8.218	6.224	4.576	1.491	2.682	0.607	0.651	0.737
	n*mse:	8.218	9.063	4.576	1.740	2.682	0.616	0.651	0.910
	effvar:	59.8	74.9	78.8	82.8	94.5	103.2	96.6	91.9
	effmse:	59.8	81.2	78.8	97.7	94.5	102.4	96.6	86.3
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.004	1.397	-0.001	1.139	0.000	0.875	0.501	0.604
	n*bias2:	0.000	4.733	0.000	0.581	0.000	0.003	0.000	0.324
	n*variance:	4.406	4.203	3.470	1.050	2.388	0.585	0.439	0.613
	n*mse:	4.406	8.935	3.470	1.631	2.388	0.588	0.439	0.937
LS	mean:	-0.005	1.507	-0.003	1.153	-0.001	0.855	0.501	0.636
	n*bias2:	0.001	7.702	0.000	0.704	0.000	0.004	0.000	0.558
	n*variance:	8.129	7.805	4.597	1.629	2.564	0.610	0.462	0.730
	n*mse:	8.130	15.508	4.598	2.333	2.564	0.614	0.462	1.288
	effvar:	54.2	53.8	75.5	64.4	93.1	95.8	95.0	84.0
	effmse:	54.2	57.6	75.5	69.9	93.1	95.6	95.0	72.7
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.004	1.365	0.002	1.116	0.001	0.872	0.500	0.607
	n*bias2:	0.001	7.998	0.000	0.801	0.000	0.002	0.000	0.683
	n*variance:	4.223	3.698	3.306	0.920	2.280	0.556	0.361	0.578
	n*mse:	4.224	11.696	3.307	1.721	2.280	0.558	0.361	1.260
LS	mean:	0.005	1.539	0.004	1.163	0.001	0.862	0.500	0.651
	n*bias2:	0.002	17.432	0.001	1.598	0.000	0.001	0.000	1.375
	n*variance:	8.135	8.730	4.479	1.716	2.490	0.607	0.382	0.753
	n*mse:	8.137	26.162	4.480	3.314	2.490	0.608	0.382	2.128
	effvar:	51.9	42.4	73.8	53.6	91.6	91.6	94.6	76.8
	effmse:	51.9	44.7	73.8	51.9	91.6	91.8	94.6	59.2
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.003	1.346	0.001	1.103	-0.001	0.868	0.499	0.607
	n*bias2:	0.001	11.956	0.000	1.052	0.000	0.000	0.000	1.140
	n*variance:	4.109	3.383	3.251	0.875	2.231	0.532	0.332	0.558
	n*mse:	4.110	15.339	3.251	1.927	2.231	0.532	0.332	1.698
LS	mean:	0.005	1.550	0.002	1.163	0.000	0.862	0.499	0.659
	n*bias2:	0.003	30.218	0.000	2.673	0.000	0.002	0.000	2.515
	n*variance:	8.057	9.204	4.452	1.771	2.458	0.594	0.355	0.787
	n*mse:	8.060	39.422	4.452	4.444	2.458	0.596	0.355	3.302
	effvar:	51.0	36.8	73.0	49.4	90.8	89.5	93.5	70.9
	effmse:	51.0	38.9	73.0	43.4	90.8	89.3	93.5	51.4

Table 5.11 Simulated values for the mixture model for $b_1 = 1$, $b_2 = 1$ and $\rho = 0.9$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.004	1.433	0.004	1.376	0.000	0.445	0.899	0.923
	n*bias2:	0.000	2.812	0.000	2.119	0.000	0.001	0.000	0.008
	n*variance:	5.103	4.675	4.785	3.591	0.661	0.157	0.148	0.059
	n*mse:	5.104	7.487	4.785	5.710	0.661	0.158	0.148	0.067
LS	mean:	0.005	1.447	0.004	1.378	0.000	0.424	0.899	0.930
	n*bias2:	0.000	3.004	0.000	2.139	0.000	0.002	0.000	0.013
	n*variance:	8.415	6.289	7.520	4.870	0.702	0.151	0.155	0.057
	n*mse:	8.415	9.293	7.520	7.009	0.702	0.154	0.155	0.070
	effvar:	60.6	74.3	63.6	73.7	94.2	103.4	95.7	103.1
	effmse:	60.6	80.6	63.6	81.5	94.2	102.7	95.7	94.9
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.011	1.389	-0.010	1.332	0.000	0.440	0.900	0.932
	n*bias2:	0.004	4.541	0.003	3.308	0.000	0.001	0.000	0.032
	n*variance:	4.519	3.930	4.180	3.009	0.607	0.141	0.111	0.044
	n*mse:	4.523	8.470	4.183	6.317	0.607	0.142	0.111	0.076
LS	mean:	-0.014	1.492	-0.014	1.416	-0.001	0.430	0.900	0.940
	n*bias2:	0.006	7.267	0.005	5.195	0.000	0.001	0.000	0.049
	n*variance:	8.294	7.189	7.248	5.521	0.655	0.148	0.117	0.047
	n*mse:	8.300	14.455	7.254	10.716	0.655	0.149	0.117	0.096
	effvar:	54.5	54.7	57.7	54.5	92.6	95.4	94.7	94.4
	effmse:	54.5	58.6	57.7	59.0	92.6	95.0	94.7	79.0
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.001	1.361	-0.002	1.304	-0.001	0.439	0.900	0.936
	n*bias2:	0.000	7.821	0.000	5.549	0.000	0.000	0.000	0.078
	n*variance:	4.280	3.724	4.031	2.856	0.585	0.140	0.093	0.038
	n*mse:	4.280	11.545	4.031	8.405	0.585	0.140	0.093	0.116
LS	mean:	-0.001	1.532	-0.002	1.449	-0.001	0.433	0.900	0.947
	n*bias2:	0.000	17.001	0.000	12.107	0.000	0.000	0.000	0.133
	n*variance:	8.227	8.875	7.268	6.826	0.639	0.153	0.098	0.042
	n*mse:	8.227	25.876	7.269	18.933	0.639	0.153	0.098	0.175
	effvar:	52.0	42.0	55.5	41.8	91.5	91.3	94.6	89.5
	effmse:	52.0	44.6	55.5	44.4	91.5	91.4	94.6	66.3
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.001	1.352	0.000	1.295	0.000	0.437	0.900	0.938
	n*bias2:	0.000	12.415	0.000	8.706	0.000	0.000	0.000	0.145
	n*variance:	4.179	3.467	3.997	2.640	0.599	0.132	0.083	0.033
	n*mse:	4.179	15.882	3.997	11.346	0.599	0.133	0.083	0.178
LS	mean:	0.000	1.558	0.001	1.470	0.001	0.434	0.900	0.951
	n*bias2:	0.000	31.120	0.000	22.078	0.000	0.000	0.000	0.262
	n*variance:	8.343	9.226	7.381	7.061	0.655	0.145	0.090	0.036
	n*mse:	8.343	40.346	7.381	29.139	0.656	0.146	0.090	0.298
	effvar:	50.1	37.6	54.2	37.4	91.3	91.1	92.8	91.2
	effmse:	50.1	39.4	54.2	38.9	91.3	91.0	92.8	59.7

Table 5.12 Simulated values for the contamination model for $b_1 = 1$, $b_2 = 1$ and $\rho = 0.2$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.006	0.960	-0.001	1.059	-0.002	1.006	0.201	0.179
	n*bias2:	0.000	0.024	0.000	0.051	0.000	0.010	0.000	0.007
	n*variance:	2.575	0.788	3.013	0.863	3.170	0.842	1.387	0.910
	n*mse:	2.576	0.813	3.013	0.915	3.170	0.852	1.388	0.917
LS	mean:	-0.006	0.923	-0.001	0.978	-0.002	0.959	0.201	0.186
	n*bias2:	0.001	0.090	0.000	0.007	0.000	0.007	0.000	0.003
	n*variance:	2.949	0.776	3.184	0.784	3.366	0.812	1.456	1.009
	n*mse:	2.949	0.866	3.184	0.792	3.366	0.819	1.456	1.012
	effvar:	87.3	101.5	94.7	110.0	94.2	103.6	95.3	90.2
	effmse:	87.3	93.8	94.7	115.5	94.2	104.0	95.3	90.6
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.001	0.947	0.003	1.024	0.003	0.989	0.202	0.186
	n*bias2:	0.000	0.084	0.000	0.017	0.000	0.003	0.000	0.006
	n*variance:	2.573	0.757	3.080	0.734	3.092	0.718	1.152	0.894
	n*mse:	2.573	0.841	3.081	0.751	3.092	0.721	1.152	0.901
LS	mean:	0.001	0.936	0.002	0.984	0.002	0.965	0.202	0.191
	n*bias2:	0.000	0.122	0.000	0.007	0.000	0.006	0.000	0.003
	n*variance:	3.059	0.800	3.301	0.745	3.306	0.748	1.224	0.983
	n*mse:	3.059	0.922	3.301	0.753	3.306	0.754	1.224	0.986
	effvar:	84.1	94.6	93.3	98.4	93.5	96.1	94.2	91.0
	effmse:	84.1	91.2	93.3	99.7	93.5	95.6	94.2	91.3
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.001	0.939	0.001	1.008	0.001	0.983	0.201	0.186
	n*bias2:	0.000	0.220	0.000	0.004	0.000	0.001	0.000	0.012
	n*variance:	2.504	0.709	2.950	0.719	2.919	0.705	1.042	0.846
	n*mse:	2.505	0.929	2.950	0.723	2.920	0.706	1.042	0.858
LS	mean:	-0.002	0.943	0.001	0.990	0.002	0.971	0.201	0.191
	n*bias2:	0.000	0.195	0.000	0.006	0.000	0.004	0.000	0.005
	n*variance:	2.969	0.789	3.223	0.784	3.184	0.777	1.116	0.940
	n*mse:	2.969	0.984	3.223	0.791	3.184	0.781	1.116	0.945
	effvar:	84.3	89.9	91.5	91.7	91.7	90.7	93.3	90.0
	effmse:	84.3	94.4	91.5	91.5	91.7	90.3	93.3	90.7
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.000	0.937	-0.003	1.006	-0.004	0.983	0.201	0.187
	n*bias2:	0.000	0.403	0.001	0.003	0.001	0.001	0.000	0.017
	n*variance:	2.514	0.739	2.997	0.707	2.923	0.692	1.018	0.850
	n*mse:	2.514	1.142	2.998	0.710	2.924	0.694	1.018	0.867
LS	mean:	-0.001	0.946	-0.004	0.995	-0.004	0.976	0.201	0.191
	n*bias2:	0.000	0.291	0.001	0.003	0.001	0.001	0.000	0.008
	n*variance:	3.047	0.831	3.317	0.776	3.222	0.765	1.104	0.951
	n*mse:	3.047	1.122	3.319	0.779	3.223	0.767	1.105	0.959
	effvar:	82.5	89.0	90.4	91.1	90.7	90.5	92.2	89.4
	effmse:	82.5	101.8	90.3	91.2	90.7	90.5	92.2	90.4

The contamination model represents mild deviation from the assumed model. Even here, the MMLE are more efficient than the LSE particularly for large n (say, $n > 15$).

Table 5.13 Simulated values for the contamination model for $b_1 = 1$, $b_2 = 1$ and $\rho = 0.5$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	-0.004	0.958	0.004	1.036	0.006	0.884	0.500	0.454
	n*bias2:	0.000	0.026	0.000	0.019	0.001	0.005	0.000	0.032
	n*variance:	2.590	0.788	3.042	0.711	2.592	0.611	1.096	0.704
	n*mse:	2.590	0.814	3.042	0.731	2.592	0.616	1.096	0.736
LS	mean:	-0.003	0.921	0.006	0.963	0.007	0.841	0.500	0.467
	n*bias2:	0.000	0.094	0.000	0.020	0.001	0.009	0.000	0.016
	n*variance:	2.943	0.767	3.265	0.656	2.749	0.587	1.141	0.753
	n*mse:	2.944	0.861	3.265	0.676	2.749	0.596	1.141	0.769
	effvar:	88.0	102.7	93.2	108.5	94.3	104.2	96.1	93.4
	effmse:	88.0	94.5	93.2	108.1	94.3	103.4	96.1	95.6

n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.004	0.945	0.003	1.010	0.000	0.877	0.501	0.463
	n*bias2:	0.000	0.092	0.000	0.003	0.000	0.004	0.000	0.041
	n*variance:	2.494	0.763	2.916	0.647	2.379	0.568	0.913	0.649
	n*mse:	2.494	0.854	2.916	0.650	2.379	0.572	0.913	0.689
LS	mean:	0.003	0.934	0.002	0.977	0.001	0.857	0.501	0.473
	n*bias2:	0.000	0.130	0.000	0.015	0.000	0.003	0.000	0.022
	n*variance:	2.945	0.806	3.200	0.659	2.566	0.593	0.967	0.707
	n*mse:	2.945	0.936	3.200	0.675	2.566	0.596	0.967	0.729
	effvar:	84.7	94.6	91.1	98.1	92.7	95.8	94.4	91.8
	effmse:	84.7	91.3	91.1	96.3	92.7	96.0	94.4	94.6

n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.001	0.939	0.001	0.996	0.001	0.872	0.498	0.468
	n*bias2:	0.000	0.220	0.000	0.001	0.000	0.002	0.000	0.063
	n*variance:	2.511	0.718	2.964	0.628	2.356	0.551	0.820	0.644
	n*mse:	2.511	0.938	2.964	0.629	2.356	0.553	0.821	0.707
LS	mean:	0.002	0.943	0.001	0.982	0.001	0.861	0.498	0.476
	n*bias2:	0.000	0.197	0.000	0.019	0.000	0.001	0.000	0.034
	n*variance:	2.998	0.797	3.319	0.675	2.576	0.595	0.887	0.713
	n*mse:	2.998	0.994	3.319	0.694	2.576	0.596	0.887	0.747
	effvar:	83.7	90.1	89.3	93.1	91.5	92.7	92.5	90.3
	effmse:	83.7	94.4	89.3	90.7	91.5	92.8	92.5	94.6

n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.000	0.935	-0.001	0.991	-0.001	0.869	0.499	0.470
	n*bias2:	0.000	0.422	0.000	0.009	0.000	0.001	0.000	0.090
	n*variance:	2.498	0.724	2.877	0.623	2.265	0.547	0.815	0.630
	n*mse:	2.498	1.147	2.877	0.632	2.265	0.548	0.815	0.721
LS	mean:	-0.001	0.944	0.000	0.984	0.000	0.863	0.499	0.478
	n*bias2:	0.000	0.310	0.000	0.026	0.000	0.001	0.000	0.049
	n*variance:	2.986	0.817	3.219	0.692	2.476	0.609	0.882	0.697
	n*mse:	2.986	1.128	3.219	0.718	2.476	0.610	0.882	0.745
	effvar:	83.7	88.6	89.4	90.1	91.5	89.8	92.3	90.5
	effmse:	83.7	101.7	89.4	88.0	91.5	89.9	92.3	96.7

Table 5.14 Simulated values for the contamination model for $b_1 = 1$, $b_2 = 1$ and $\rho = 0.9$.

n=15		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.000	0.961	0.000	0.982	0.000	0.445	0.898	0.869
	n*bias2:	0.000	0.023	0.000	0.005	0.000	0.001	0.000	0.014
	n*variance:	2.530	0.803	2.666	0.708	0.642	0.156	0.274	0.106
	n*mse:	2.530	0.826	2.666	0.712	0.642	0.157	0.274	0.121
LS	mean:	-0.001	0.923	-0.001	0.935	0.000	0.424	0.898	0.877
	n*bias2:	0.000	0.089	0.000	0.064	0.000	0.002	0.000	0.008
	n*variance:	2.910	0.778	3.021	0.693	0.681	0.151	0.287	0.103
	n*mse:	2.910	0.866	3.021	0.757	0.681	0.153	0.287	0.111
	effvar:	87.0	103.3	88.2	102.1	94.3	103.6	95.6	103.1
	effmse:	87.0	95.3	88.2	94.1	94.3	102.9	95.6	108.7
n=30		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.000	0.945	0.001	0.964	0.000	0.441	0.901	0.879
	n*bias2:	0.000	0.090	0.000	0.040	0.000	0.001	0.000	0.014
	n*variance:	2.545	0.739	2.683	0.648	0.605	0.143	0.235	0.079
	n*mse:	2.545	0.829	2.683	0.688	0.605	0.143	0.235	0.093
LS	mean:	-0.001	0.934	0.000	0.946	0.001	0.430	0.900	0.884
	n*bias2:	0.000	0.130	0.000	0.086	0.000	0.001	0.000	0.008
	n*variance:	2.990	0.786	3.109	0.687	0.654	0.149	0.248	0.080
	n*mse:	2.990	0.917	3.109	0.773	0.654	0.149	0.248	0.088
	effvar:	85.1	93.9	86.3	94.4	92.5	96.0	94.7	99.0
	effmse:	85.1	90.4	86.3	88.9	92.5	95.9	94.7	105.6
n=60		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.001	0.939	0.002	0.955	0.001	0.438	0.901	0.883
	n*bias2:	0.000	0.225	0.000	0.121	0.000	0.000	0.000	0.017
	n*variance:	2.527	0.725	2.582	0.629	0.579	0.137	0.210	0.070
	n*mse:	2.527	0.950	2.582	0.750	0.579	0.138	0.210	0.086
LS	mean:	0.000	0.942	0.001	0.954	0.001	0.433	0.901	0.888
	n*bias2:	0.000	0.200	0.000	0.129	0.000	0.001	0.000	0.009
	n*variance:	3.018	0.793	3.042	0.689	0.636	0.150	0.223	0.071
	n*mse:	3.018	0.993	3.042	0.818	0.636	0.150	0.223	0.080
	effvar:	83.7	91.4	84.9	91.2	91.0	91.8	94.3	98.5
	effmse:	83.7	95.6	84.9	91.6	91.0	91.7	94.3	108.1
n=100		μ_1	σ_1	μ_2	σ_2	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1	ρ
MML	mean:	0.002	0.936	0.002	0.950	0.000	0.437	0.899	0.885
	n*bias2:	0.000	0.413	0.000	0.251	0.000	0.000	0.000	0.023
	n*variance:	2.530	0.704	2.630	0.623	0.579	0.134	0.202	0.067
	n*mse:	2.531	1.116	2.630	0.874	0.579	0.134	0.202	0.090
LS	mean:	0.002	0.946	0.002	0.956	0.000	0.434	0.900	0.889
	n*bias2:	0.001	0.294	0.000	0.196	0.000	0.000	0.000	0.012
	n*variance:	3.029	0.791	3.085	0.706	0.632	0.149	0.218	0.071
	n*mse:	3.030	1.084	3.085	0.902	0.632	0.150	0.218	0.084
	effvar:	83.5	89.0	85.3	88.4	91.7	89.6	92.5	94.6
	effmse:	83.5	103.0	85.3	97.0	91.7	89.5	92.5	108.1

5.2 Simulated Powers of the Hotelling T^2

As we stated earlier, we give results only for the situation when the marginal and the conditional distribution are Generalized Logistic with shape parameters b_1 and

b_2, respectively. In Chapter 4, we defined statistics to test assumed values of the location parameters (μ_1, μ_2). To test

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{5.2.1}$$

in the present situation, we propose the statistic

$$\hat{\mathrm{T}}^2 = n(\hat{\mu}_1, \hat{\mu}_2)\, \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix}. \tag{5.2.2}$$

Here, $\hat{\mathbf{\Omega}}$ is the (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{13} \\ \hat{\sigma}_{13} & \hat{\sigma}_{33} \end{bmatrix} \tag{5.2.3}$$

where, $\hat{\sigma}_{11}$, $\hat{\sigma}_{13}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij} = \mathbf{\Sigma_{ij}} = I_{ij}^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$. The statistic using the LS estimators is calculated just by putting the LS estimators instead of the MML estimators in the test statistic. Unfortunately, the estimated covariance matrix $\tilde{\mathbf{\Omega}}$ of the LS estimators $\sqrt{n}(\tilde{\mu}_1, \tilde{\mu}_2)$ is not known under nonnormal distributions. Thus, we use the simulated covariance matrix. Denote the statistic (5.2.2) by $\hat{T}^2$ and the corresponding statistic based on the LSE by $\tilde{T}^2$.

Alternatively, to test H_0 above we propose the statistic

$$\begin{aligned} \hat{\mathrm{T}}_1^2 &= n(\hat{\mu}_1, \hat{\mu}_{2.1})\, \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_{2.1} \end{pmatrix} \\ &= \hat{\mu}_1{}^2 / \hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2 / \hat{\sigma}_{33}\,. \end{aligned} \tag{5.2.4}$$

Here, $\hat{\mathbf{\Omega}}$ is the (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n\begin{bmatrix} \hat{\sigma}_{11} & 0 \\ 0 & \hat{\sigma}_{33} \end{bmatrix} \tag{5.2.5}$$

where $\hat{\sigma}_{11}$ and $\hat{\sigma}_{33}$ are the estimated elements of the asymptotic covariance matrix $\mathbf{\Sigma}$, more specifically, $\sigma_{ij} = \mathbf{\Sigma_{ij}} = I_{ij}^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$. The statistic based on LS estimators is calculated simply by replacing the MMLE in (5.2.4) by the LSE and, $\hat{\sigma}_{11}$ and $\hat{\sigma}_{33}$ by $\tilde{\sigma}_{11}$ and $\tilde{\sigma}_{33}$. For the same reason as before we use the simulated variances for $\tilde{\sigma}_{11}$ and $\tilde{\sigma}_{33}$. Denote the statistic (5.2.4) by $\hat{T}_1^2$ and the corresponding statistic based on the LSE by $\tilde{T}_1^2$. Testing $(\mu_1, \mu_2) = (0,0)$ is equivalent to testing that $(\mu_1, \mu_{2.1}) = (0,0)$.

Since the covariance matrix does not converge to its expected value fast enough, we will give the simulation results based on the simulated variances and covariances of the estimators. For example, instead of using $\hat{\mathbf{\Omega}}$ in (5.2.2) we use simulated variance of $\hat{\mu}_1$ and $\hat{\mu}_2$ and simulated covariance of $\hat{\mu}_1$ and $\hat{\mu}_2$.

We give the graph plots of the power of $\hat{T}^2$ and $\tilde{T}^2$, and $\hat{T}_1^2$ and $\tilde{T}_1^2$ for testing $(\mu_1, \mu_2) = (0,0)$. We give the plots for $\rho = 0.5$. The relative powers are essentially the same for other values of ρ. We include numerous values of b_1 and b_2 and $n = 15, 30, 60$ and 100. The results are based on 10,000 Monte Carlo runs. First, we give the plots for $\hat{T}_1^2$ and $\tilde{T}_1^2$.

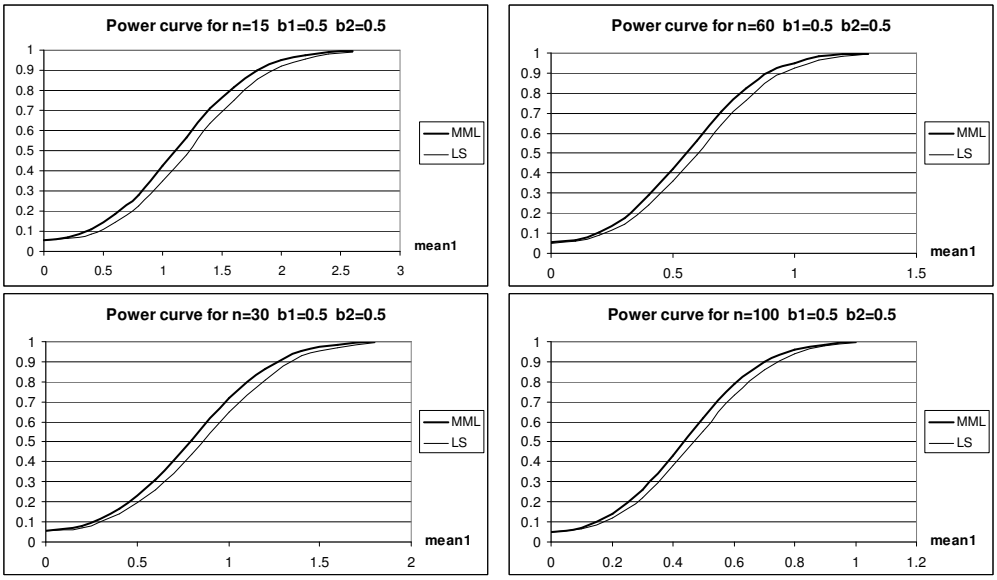

Figure 5.1 Power graphs of $\hat{T}_1^2$ and $\tilde{T}_1^2$ for $b_1 = 0.5$, $b_2 = 0.5$, $\rho = 0.5$ and n=15, 30,60 and 100.

It is seen that $\hat{T}_1^2$ (based on the MMLE) has higher power than $\tilde{T}_1^2$ (based on the LSE) for all sample sizes. The type I error for both of them is almost 0.05, the presumed value.

Figure 5.2 shows the graph plots of power for $b_1 = 0.5$ and $b_2 = 1$. It is seen that $\hat{T}_1^2$ has higher power than $\tilde{T}_1^2$ and the differences increase with increasing n.

Figure 5.3 shows the graph plots of power for $b_1 = 1$, $b_2 = 0.5$. Although the differences between the power values is not very large but again $\hat{T}_1^2$ is ahead of $\tilde{T}_1^2$.

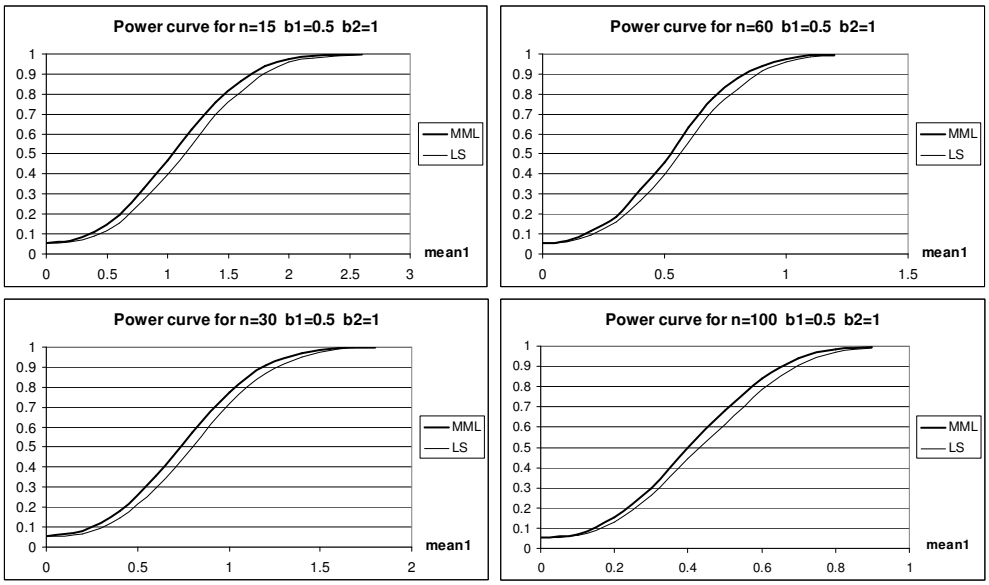

Figure 5.2 Power graphs of $\hat{T}_1^2$ and $\tilde{T}_1^2$ for $b_1 = 0.5$, $b_2 = 1, \rho = 0.5$ and n =15,30,60 and 100.

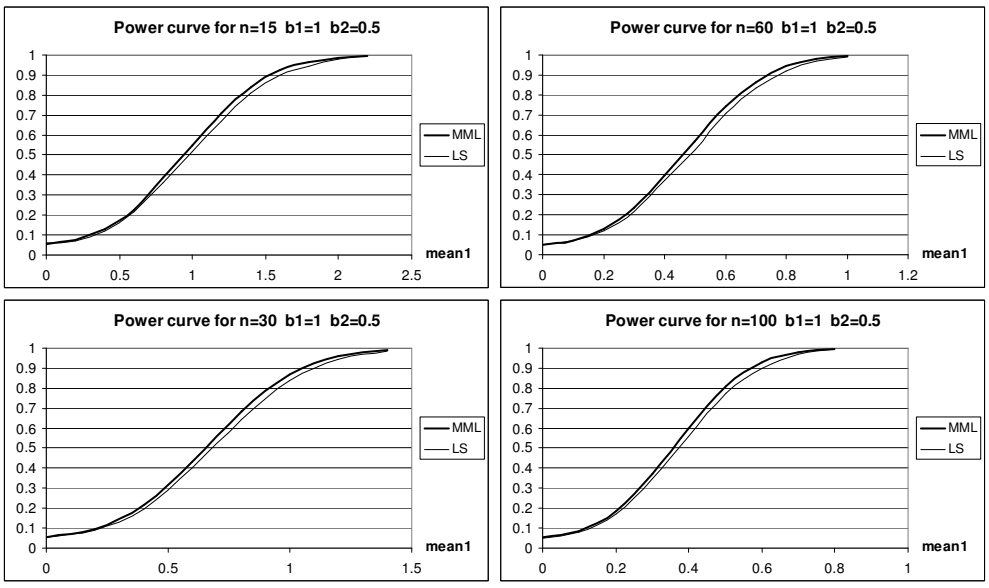

Figure 5.3 Power graphs of $\hat{T}_1^2$ and $\tilde{T}_1^2$ for $b_1 = 1$, $b_2 = 0.5, \rho = 0.5$ and n =15,30,60 and 100.

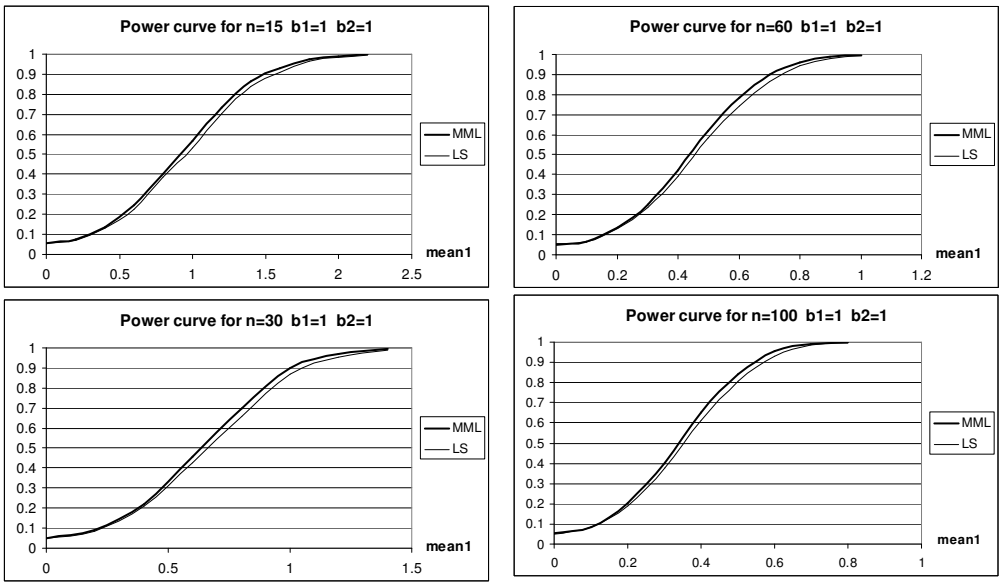

Figure 5.4 Power graphs of $\hat{T}_1^2$ and $\tilde{T}_1^2$ for $b_1 = 1$, $b_2 = 1, \rho = 0.5$ and n =15,30,60 and 100.

In Figure 5.4 we consider the situation when both the marginal and the conditional distributions are symmetric. Like the previous case, the differences between the power of the $\hat{T}_1^2$ and $\tilde{T}_1^2$ tests are not substantial. Nonetheless, $\hat{T}_1^2$ is ahead of the $\tilde{T}_1^2$ test.

Finally, we consider the situation when $b_1 = 4$ and $b_2 = 4$. The graph plots of the power are given in Figure 5.5. For $n \le 30$, there are very little differences between the power values. For $n > 30$, however the differences are more pronounced and $\hat{T}_1^2$ test is more powerful.

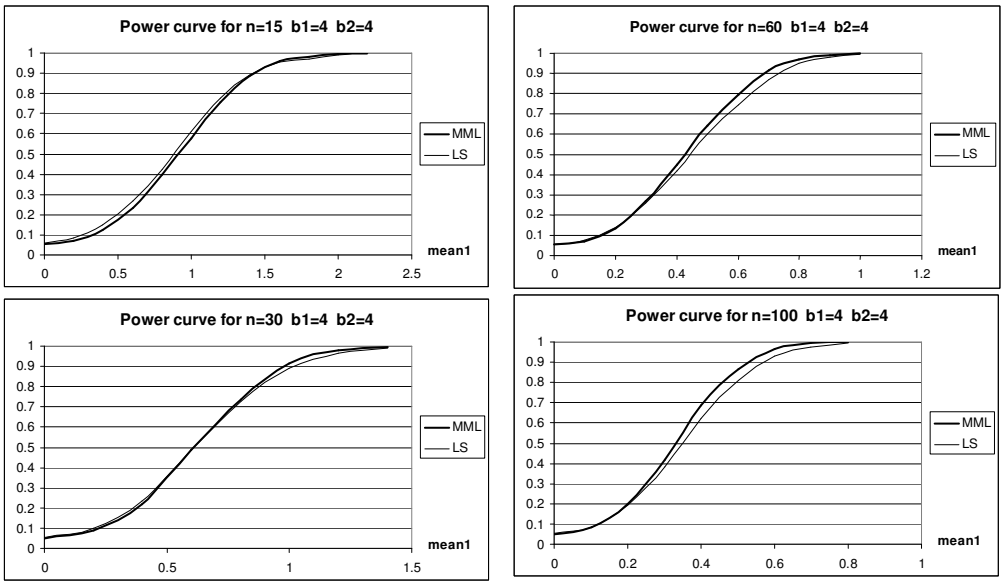

Figure 5.5 Power graphs of $\hat{T}_1^2$ and $\tilde{T}_1^2$ for $b_1 = 4$, $b_2 = 4$, $\rho = 0.5$ and n =15,30,60 and 100.

We now consider the $\hat{T}^2$ and $\tilde{T}^2$ tests. We simulate the values of their power for testing $(\mu_1, \mu_2) = (0,0)$. The graph plots of their power, simulated for $\rho = 0.5$ and $n = 15, 30, 60$ and 100, are given in Figure 5.6 for $b_1 = 0.5$, $b_2 = 0.5$, Figure 5.7 for $b_1 = 0.5$, $b_2 = 1$, Figure 5.8 for $b_1 = 1$, $b_2 = 0.5$, Figure 5.9 for $b_1 = 1$, $b_2 = 1$ and Figure 5.10 for $b_1 = 4$, $b_2 = 4$. The comparison between the power values is essentially the same as for the $\hat{T}^2$ and $\tilde{T}^2$ tests. In the case given in Figure 5.8, $\hat{T}^2$ shows higher power values for all sample sizes. Moreover, as sample size increases, the difference gets bigger in favor of $\hat{T}^2$.

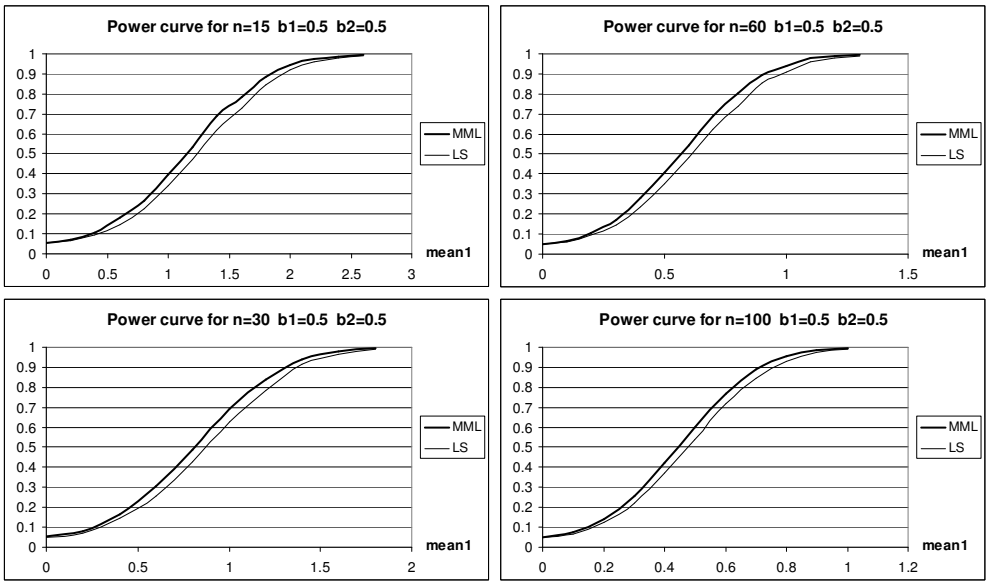

Figure 5.6 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ for $b_1 = 0.5$, $b_2 = 0.5, \rho = 0.5$ and n=15,30,60 and 100.

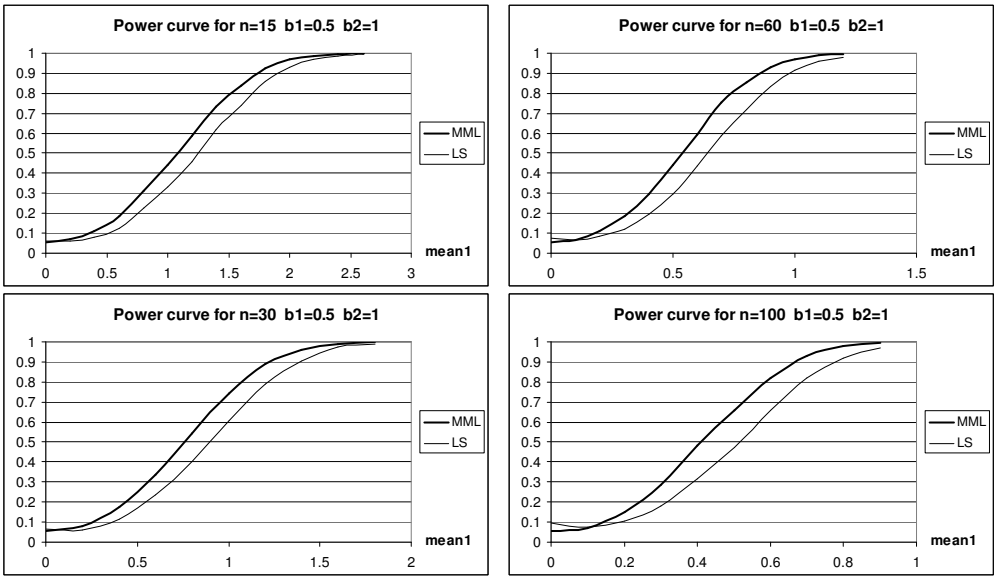

Figure 5.7 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ for $b_1 = 0.5$, $b_2 = 1, \rho = 0.5$ and n=15,30,60 and 100.

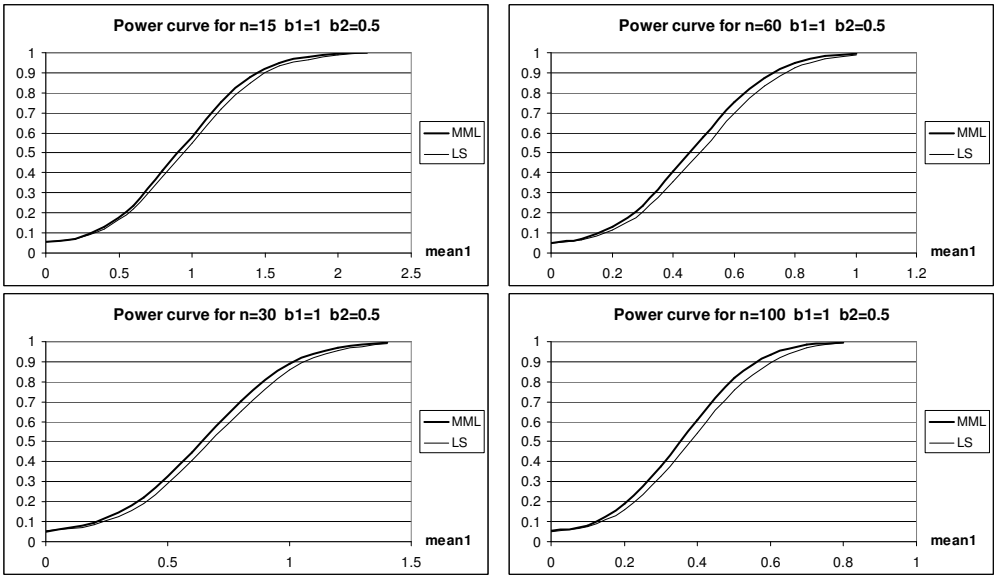

Figure 5.8 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ for $b_1 = 1$, $b_2 = 0.5, \rho = 0.5$ and n=15,30,60 and 100.

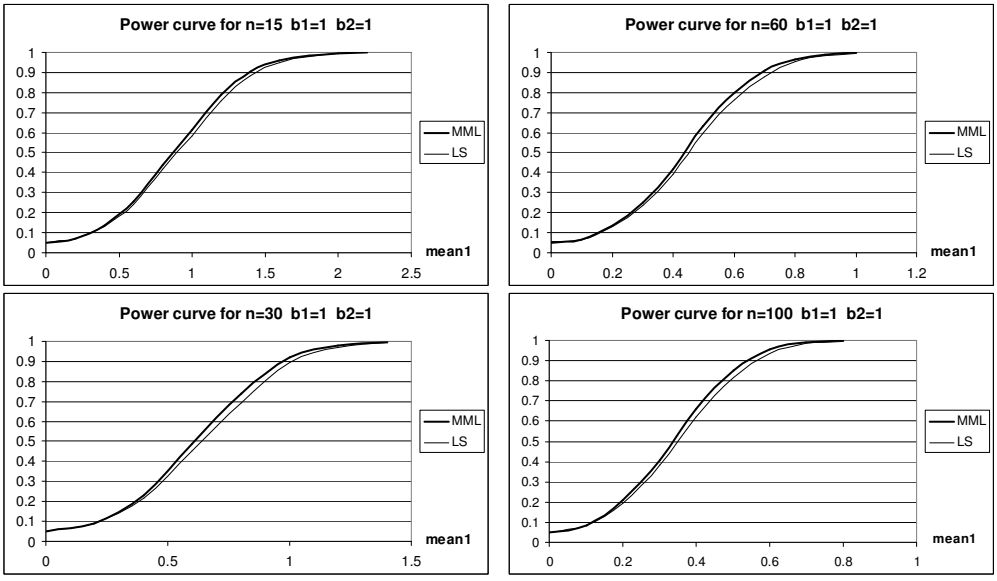

Figure 5.9 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ for $b_1 = 1$, $b_2 = 1, \rho = 0.5$ and n=15,30,60 and 100.

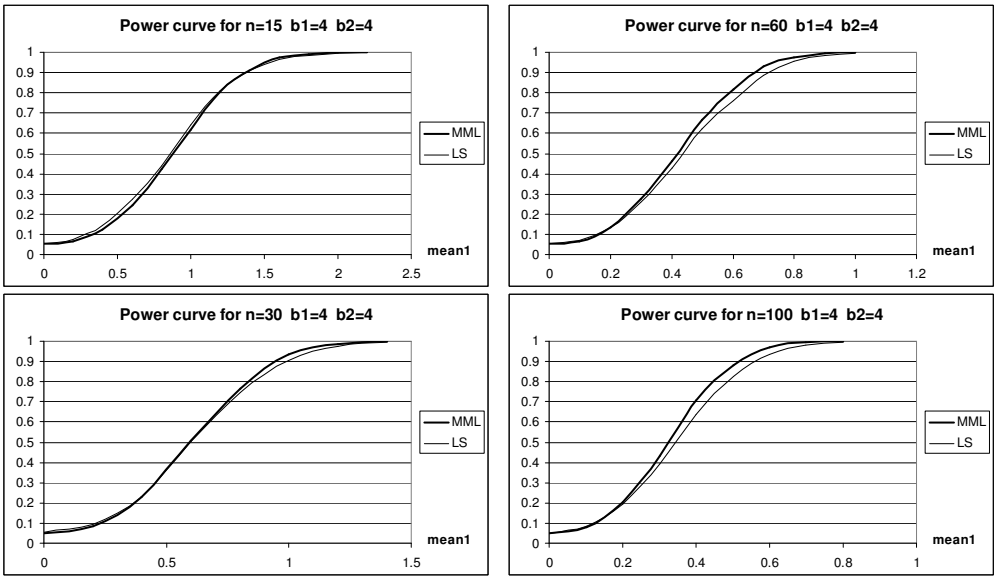

Figure 5.10 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ for $b_1 = 4$, $b_2 = 4$, $\rho = 0.5$ and n =15,30,60 and 100.

Robustness: A test is said to be robust if its Type I error is, for plausible alternatives, never substantially higher than a presumed value and it maintains high power. Consider the situation when the model is the one with $b_1 = 1$ and $b_2 = 1$. The alternatives we consider are the outlier, mixture and contamination models given on page 115. The graph plots of the powers of $\hat{T}_1^2$, $\tilde{T}_1^2$, $\hat{T}^2$ and $\tilde{T}^2$ tests under three alternatives considered above are given in Figure 5.11-5.16. It can be seen that $\hat{T}_1^2$ and $\hat{T}^2$ tests are robust. This is due to the fact that $\hat{T}_1^2$ and $\hat{T}^2$ tests use efficient and robust estimators of the unknown parameters.

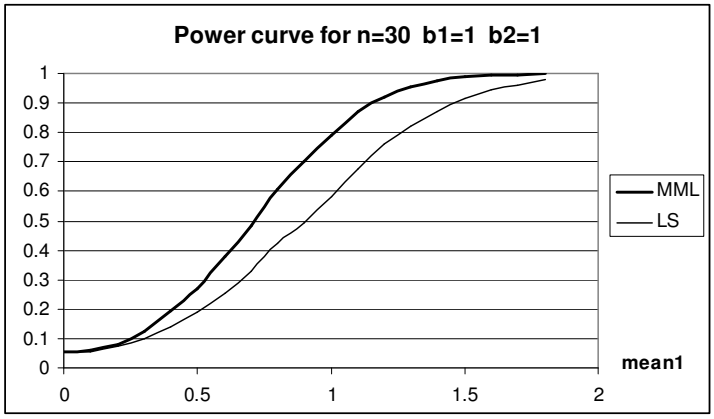

Figure 5.11 Power graphs of $\hat{T}_1^2$ and $\tilde{T}_1^2$ for $b_1 = 1$, $b_2 = 1$, $\rho = 0.5$ and n=30 for the outlier model.

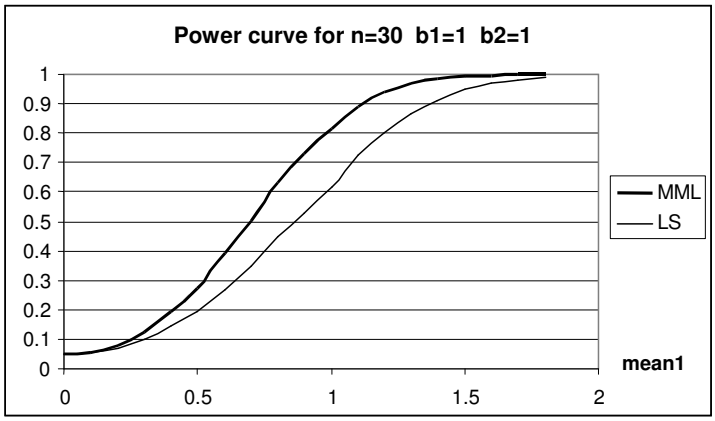

Figure 5.12 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ for $b_1 = 1$, $b_2 = 1$, $\rho = 0.5$ and n=30 for the outlier model.

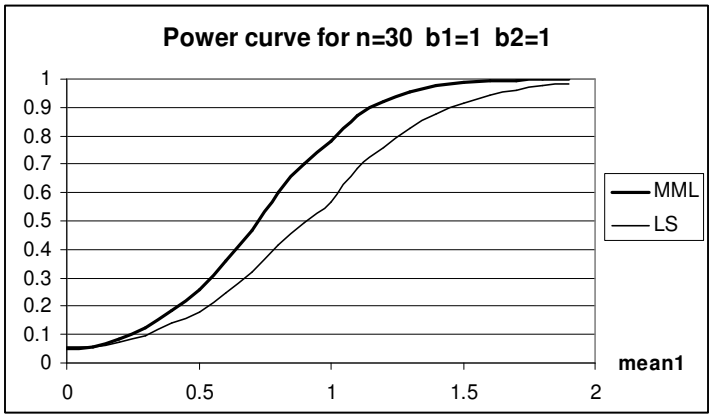

Figure 5.13 Power graphs of $\hat{T}_1^2$ and $\tilde{T}_1^2$ for $b_1 = 1$, $b_2 = 1, \rho = 0.5$ and n=30 for the mixture model.

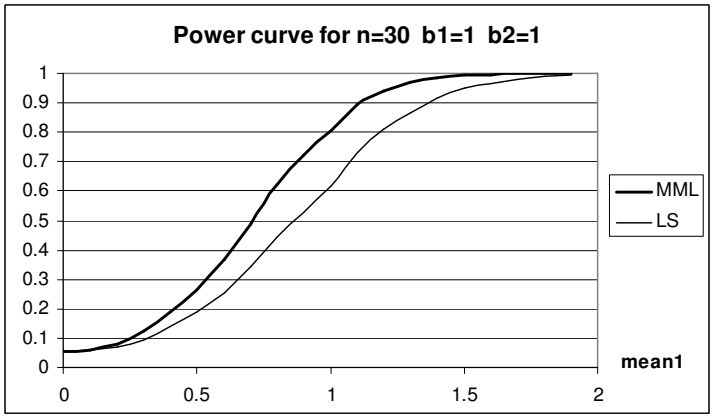

Figure 5.14 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ for $b_1 = 1$, $b_2 = 1, \rho = 0.5$ and n=30 for the mixture model.

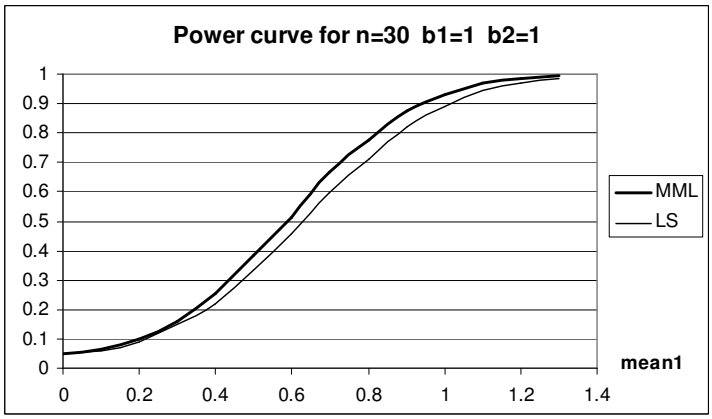

Figure 5.15 Power graphs of $\hat{T}_1^2$ and $\tilde{T}_1^2$ for $b_1 = 1$, $b_2 = 1$, $\rho = 0.5$ and n=30 for the contamination model.

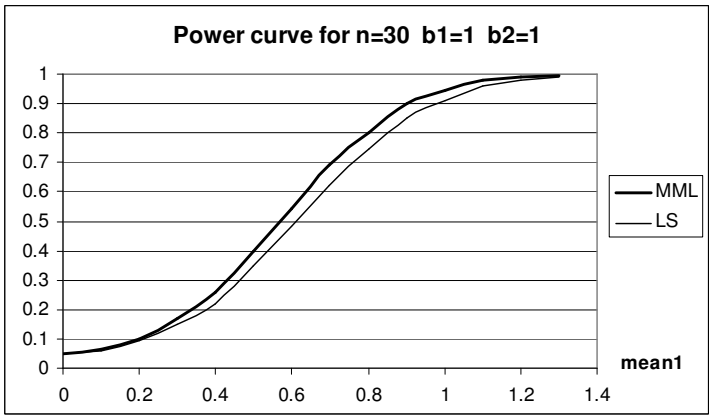

Figure 5.16 Power graphs of $\hat{T}^2$ and $\tilde{T}^2$ for $b_1 = 1$, $b_2 = 1$, $\rho = 0.5$ and n=30 for the contamination model.

5.3 Testing the Correlation Coefficient

As stated in Chapter 4, in order to test $H_0 : \rho = 0$ against $H_1 : \rho < 0$ (or $\rho > 0$), the proposed statistic is

$$W = \hat{\rho} \Big/ \sqrt{\frac{1}{n} \frac{(b_2 + 2)}{b_2} \frac{1}{\left(\psi'(b_1) + \psi'(1)\right)}}$$

$$= \hat{\rho}\sqrt{n\frac{b_2}{(b_2+2)}\left(\psi'(b_1)+\psi'(1)\right)} \tag{5.3.1}$$

Here,

$$\frac{1}{n}\frac{(b_2+2)}{b_2}\frac{1}{\left(\psi'(b_1)+\psi'(1)\right)} \tag{5.3.2}$$

is the asymptotic variance of $\hat{\rho}$ under H_0, obtained from

$$\left\{I^{-1}(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)\right\}_{\rho=0}. \tag{5.3.3}$$

The LSE of ρ is the Pearson sample correlation coefficient, $\tilde{\rho}=\frac{s_{xy}}{s_x s_y}$. The statistic based on the Pearson sample correlation coefficient is found by dividing $\tilde{\rho}$ by the square root of the variance of $\tilde{\rho}$ under H_0. Since we do not know the variance of $\tilde{\rho}$ under nonnormal distributions, we use the simulated variance of $\tilde{\rho}$ under H_0. Denote the test statistic (5.3.1) by W and the corresponding test statistic based on the LSE by W_{LS}. In Section 5.1, we gave the efficiencies of the LS estimator $\tilde{\rho}$ relative to the MML estimator $\hat{\rho}$. In this section we give the simulated power graphs of W and W_{LS} for values of b_1 and b_2, and for $n=15, 30, 60$ and 100. Since the variance of $\hat{\rho}$ does not converge fast enough to the asymptotic value, we use the simulated variance of $\hat{\rho}$ under the null hypothesis $H_0: \rho = 0$.

First, we give the simulated power graphs of W and W_{LS} for $b_1 = 0.5$ and $b_2 = 0.5$ for several sample sizes (Figure 5.17). For all sample sizes, W is more

powerful than W_{LS}. As n gets large, the differences become larger in favor of W due to the asymptotic optimality of the MML estimator $\hat{\rho}$.

In Figure 5.18, the simulated power graphs of W and W_{LS} for $b_1 = 0.5$ and $b_2 = 1$, are given. The difference between the power values is not considerable but again there is a difference in favor of W and it becomes large as the sample size n increases.

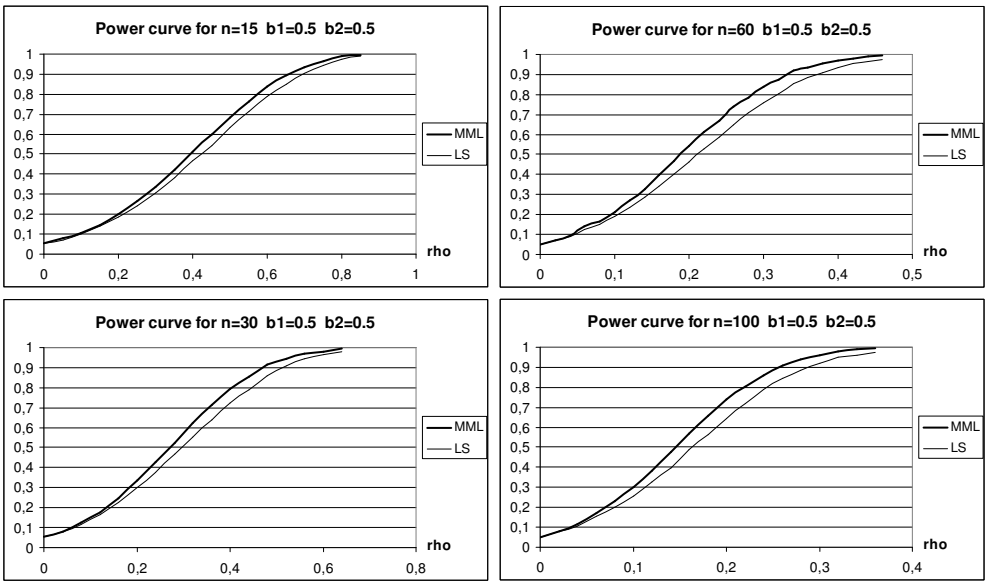

Figure 5.17 Power graphs of W and W_{LS} for $b_1 = 0.5$, $b_2 = 0.5$ and n=15,30,60 and 100.

In Figure 5.19, we give the simulated powers of W and W_{LS} for $b_1 = 1$ and $b_2 = 0.5$. As expected, the W test has higher power for all n.

In Figure 5.20, the power values of the W and W_{LS} tests for $b_1 = 1$ and $b_2 = 1$ are given. The W test is more powerful and and the differences between the values increase with sample size n.

Finally, for $b_1 = 4$ and $b_2 = 4$, the simulated powers of the W and W_{LS} tests are given in Figure 5.21. Now, the W test is considerably more powerful than the W_{LS} test. The differences increase with n.

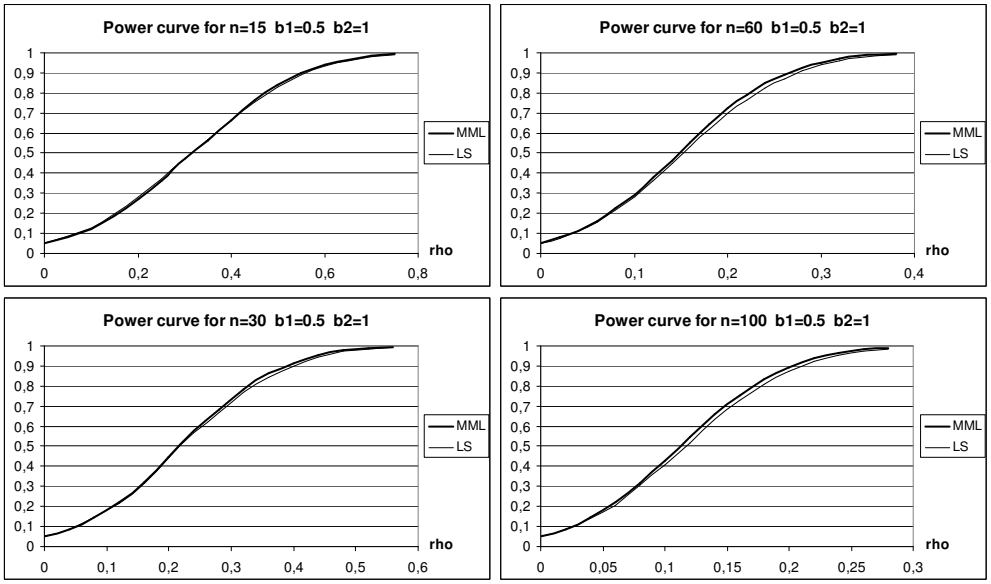

Figure 5.18 Power graphs of W and W_{LS} for $b_1 = 0.5$, $b_2 = 1$ and n=15,30,60 and 100.

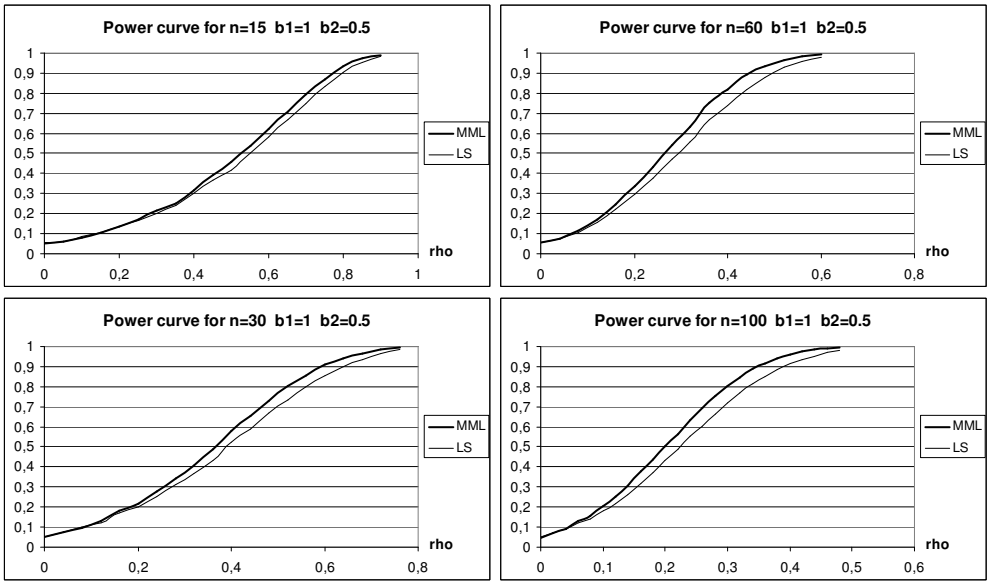

Figure 5.19 Power graphs of W and W_{LS} for $b_1 = 1$, $b_2 = 0.5$ and n=15,30,60 and 100.

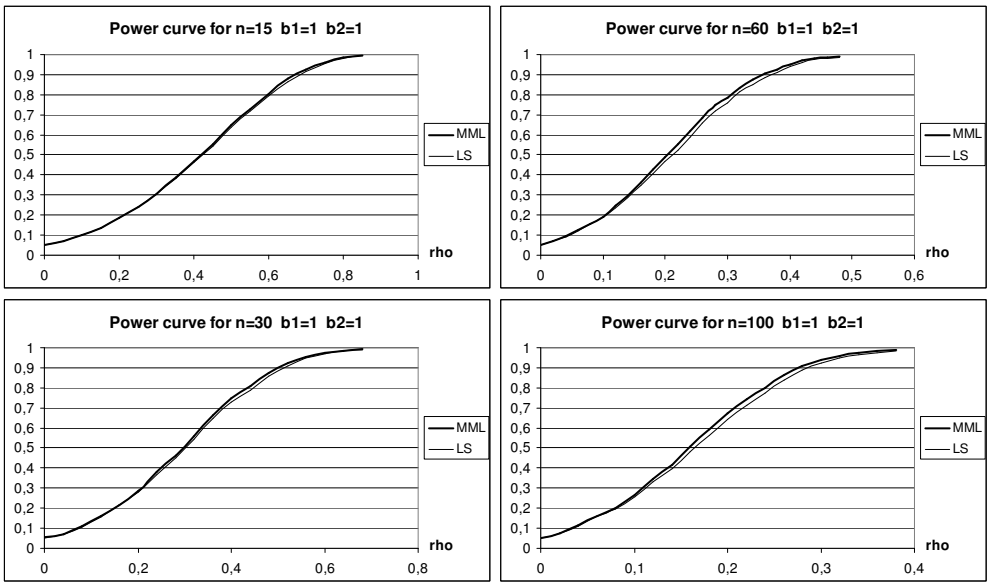

Figure 5.20 Power graphs of W and W_{LS} for $b_1 = 1$, $b_2 = 1$ and n =15,30,60 and 100.

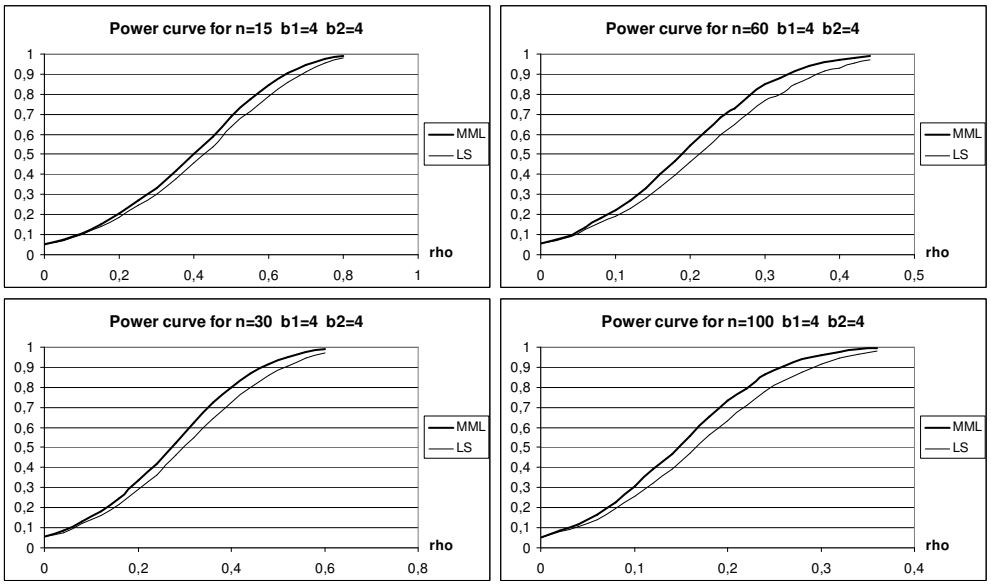

Figure 5.21 Power graphs of W and W_{LS} for $b_1 = 4$, $b_2 = 4$ and n =15,30,60 and 100.

5.4 Proximity of ML and MML Estimators

We know that the MMLE are asymptotically equivalent to the MLE, of course under the regularity conditions. The latter satisfy the likelihood equations of the type $\partial \ln L/\partial \tau = 0$. Since $\partial^2 \ln L/\partial \tau^2 < 0$, the solution of $\partial \ln L/\partial \tau = 0$ in fact maximizes L (or $\ln L$). We now show that the MMLE maximizes L (almost) even for small n. That establishes the numerical proximity of the ML and the corresponding MML estimators.

We generated $N = 10{,}000$ random samples from the bivariate distribution consisting of Generalized Logistic marginal with shape parameter b_1 and Generalized Logistic conditional with shape parameter b_2. We simulate the means and variances of the random variables $\frac{1}{n}\left(\frac{\partial \ln L}{\partial \tau_i} - \frac{\partial \ln L^*}{\partial \tau_i}\right)$ $(i = 1, 8)$. Here, $\tau_1 = \mu_1$, $\tau_2 = \sigma_1$, $\tau_3 = \mu_{2.1}$, $\tau_4 = \sigma_{2.1}$, $\tau_5 = \mu_2$, $\tau_6 = \sigma_2$, $\tau_7 = \theta_1$ and $\tau_8 = \rho$. The values are given in Table 5.15-5.16 for numerous values of b_1 and b_2 and $\rho = 0.5$. The values for other ρ-values are exactly similar. It is concluded that the MMLE are numerically close to the corresponding MLE. This is indeed a very positive result. As such, there is no necessity of chasing the elusive ML estimators.

This result shows that the MML estimators are almost equivalent to ML estimators even for small sample sizes. Additionally, MML estimators are robust to data anomalies.

Table 5.15 Simulated means and variances of $\frac{1}{n}\frac{\partial \ln L}{\partial \tau}$ (evaluated at the MMLE) for $(b_1,b_2)=(0.5,0.5)$, $(0.5,1)$ and $(1,0.5)$, and $\rho = 0.5$.

		$b_1=0.5$ $b_2=0.5$		$b_1=0.5$ $b_2=1$		$b_1=1$ $b_2=0.5$	
		mean	variance	mean	variance	mean	variance
n=15	μ_1	−0.0158	0.0001	−0.0151	0.0001	−0.0007	0.0001
	σ_1	−0.0724	0.0008	−0.0735	0.0007	−0.0874	0.0010
	$\mu_{2.1}$	0.0013	0.0000	0.0000	0.0001	0.0013	0.0000
	$\sigma_{2.1}$	−0.1476	0.0019	−0.1613	0.0022	−0.1471	0.0012
	μ_2	0.0013	0.0000	0.0000	0.0001	0.0013	0.0000
	σ_2	−0.1236	0.0012	−0.1347	0.0012	−0.1213	0.0012
	θ_1	−0.0022	0.0004	0.0003	0.0003	−0.0001	0.0002
	ρ	0.0974	0.0064	0.1053	0.0055	0.1050	0.0107
n=30	μ_1	−0.0082	0.0000	−0.0081	0.0000	0.0000	0.0000
	σ_1	−0.0364	0.0002	−0.0365	0.0002	−0.0448	0.0002
	$\mu_{2.1}$	0.0001	0.0000	−0.0001	0.0000	0.0001	0.0000
	$\sigma_{2.1}$	−0.0722	0.0003	−0.0800	0.0004	−0.0721	0.0003
	μ_2	0.0001	0.0000	−0.0001	0.0000	0.0001	0.0000
	σ_2	−0.0615	0.0002	−0.0681	0.0002	−0.0611	0.0002
	θ_1	−0.0003	0.0001	0.0001	0.0001	0.0000	0.0000
	ρ	0.0448	0.0006	0.0491	0.0006	0.0460	0.0009
n=60	μ_1	−0.0042	0.0000	−0.0042	0.0000	0.0001	0.0000
	σ_1	−0.0181	0.0000	−0.0180	0.0000	−0.0222	0.0000
	$\mu_{2.1}$	−0.0001	0.0000	0.0000	0.0000	−0.0001	0.0000
	$\sigma_{2.1}$	−0.0357	0.0001	−0.0397	0.0001	−0.0355	0.0001
	μ_2	−0.0001	0.0000	0.0000	0.0000	−0.0001	0.0000
	σ_2	−0.0306	0.0000	−0.0342	0.0000	−0.0304	0.0000
	θ_1	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	ρ	0.0215	0.0000	0.0235	0.0001	0.0215	0.0001
n=100	μ_1	−0.0025	0.0000	−0.0026	0.0000	0.0000	0.0000
	σ_1	−0.0106	0.0000	−0.0107	0.0000	−0.0131	0.0000
	$\mu_{2.1}$	−0.0001	0.0000	0.0000	0.0000	−0.0001	0.0000
	$\sigma_{2.1}$	−0.0211	0.0000	−0.0235	0.0000	−0.0211	0.0000
	μ_2	−0.0001	0.0000	0.0000	0.0000	−0.0001	0.0000
	σ_2	−0.0181	0.0000	−0.0203	0.0000	−0.0181	0.0000
	θ_1	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	ρ	0.0126	0.0000	0.0138	0.0000	0.0125	0.0000

Table 5.16 Simulated means and variances of $\frac{1}{n}\frac{\partial \ln L}{\partial \tau}$ (evaluated at the MMLE) for $(b_1,b_2)=(1,1)$ and $(4,4)$, and $\rho = 0.5$.

		$b_1=1$ $b_2=1$		$b_1=4$ $b_2=4$	
		mean	variance	mean	variance
n=15	μ_1	0.0002	0.0001	0.0446	0.0002
	σ_1	−0.0880	0.0010	−0.0216	0.0006
	$\mu_{2.1}$	−0.0001	0.0001	−0.0125	0.0001
	$\sigma_{2.1}$	−0.1610	0.0021	−0.1680	0.0022
	μ_2	−0.0001	0.0001	−0.0125	0.0001
	σ_2	−0.1340	0.0013	−0.1514	0.0018
	θ_1	0.0001	0.0001	−0.0229	0.0003
	ρ	0.1081	0.0073	0.0888	0.0068
n=30	μ_1	0.0000	0.0000	0.0258	0.0000
	σ_1	−0.0448	0.0002	−0.0081	0.0001
	$\mu_{2.1}$	0.0000	0.0000	−0.0059	0.0000
	$\sigma_{2.1}$	−0.0803	0.0004	−0.0827	0.0003
	μ_2	0.0000	0.0000	−0.0059	0.0000
	σ_2	−0.0683	0.0003	−0.0758	0.0003
	θ_1	0.0000	0.0000	−0.0109	0.0001
	ρ	0.0500	0.0008	0.0402	0.0006
n=60	μ_1	0.0000	0.0000	0.0146	0.0000
	σ_1	−0.0222	0.0000	−0.0029	0.0000
	$\mu_{2.1}$	0.0000	0.0000	−0.0026	0.0000
	$\sigma_{2.1}$	−0.0396	0.0001	−0.0413	0.0001
	μ_2	0.0000	0.0000	−0.0026	0.0000
	σ_2	−0.0341	0.0000	−0.0378	0.0001
	θ_1	0.0000	0.0000	−0.0049	0.0000
	ρ	0.0237	0.0001	0.0199	0.0001
n=100	μ_1	0.0000	0.0000	0.0095	0.0000
	σ_1	−0.0131	0.0000	−0.0012	0.0000
	$\mu_{2.1}$	0.0000	0.0000	−0.0014	0.0000
	$\sigma_{2.1}$	−0.0235	0.0000	−0.0247	0.0000
	μ_2	0.0000	0.0000	−0.0014	0.0000
	σ_2	−0.0203	0.0000	−0.0226	0.0000
	θ_1	0.0000	0.0000	−0.0026	0.0000
	ρ	0.0138	0.0000	0.0119	0.0000

5.5 Illustrative Examples

5.5.1 A Real Life Example

We give the first example using real life data given in Table 5.17 where U represents 100 times the white blood counts and Y represents the survival times (in weeks) of patients who died of acute myelogenous leukemia (Gross and Clark, 1975).

Table 5.17 Gross and Clark data.

i	1	2	3	4	5	6	7	8
U_i:	23	7.5	43	26	60	105	100	170
Y_i:	65	156	100	134	16	108	121	4
i	9	10	11	12	13	14	15	16
U_i:	54	70	94	320	350	1000	1000	520
Y_i:	39	143	56	26	22	1	1	5

Source: Gross and Clark (1975)

Vaughan and Tiku (2000) showed that it is reasonable to regard U as a Weibull random variable with $p = 0.8$. This can be verified by the Q-Q plot of U in Figure 5.22. Since the natural logarithm of a Weibull random variable has an extreme value distribution, if we let $X = \ln U$, it follows that X has an extreme value distribution. This can also be verified by the Q-Q plot of X in Figure 5.23. Vaughan and Tiku (2000) also showed that it is pertinent to regard the conditional distribution of Y given $X = x$ as normal (see Figure 5.24).

We find the MML estimators by taking the marginal distribution (X) as extreme value and the conditional distribution (Y given $X = x$) as normal. The MML and the LS estimates are given in Table 5.18.

Table 5.18 The MML and LS estimates for the Gross and Clark data.

	μ_1	σ_1	μ_2	σ_2	ρ	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1
MML	3.9868	1.3617	83.7533	55.7215	-0.7389	204.2984	38.8660	-30.2358
LS	4.0749	1.0759	82.9921	48.1972	-0.7433	204.2984	38.8660	-30.2358

Realize that since the conditional distribution is normal, the MML and the LS estimates of $\mu_{2.1}$, $\sigma_{2.1}$ and θ_1 which are the parameters belonging to the conditional part are exactly the same.

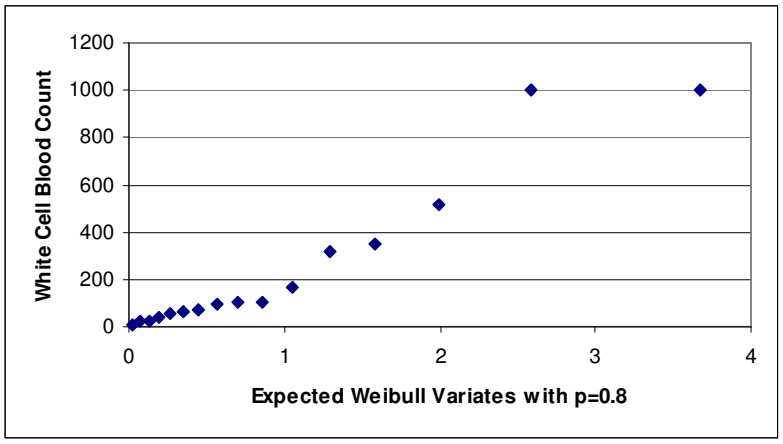

Figure 5.22 Q-Q plot of White Cell Blood Count (*U*).

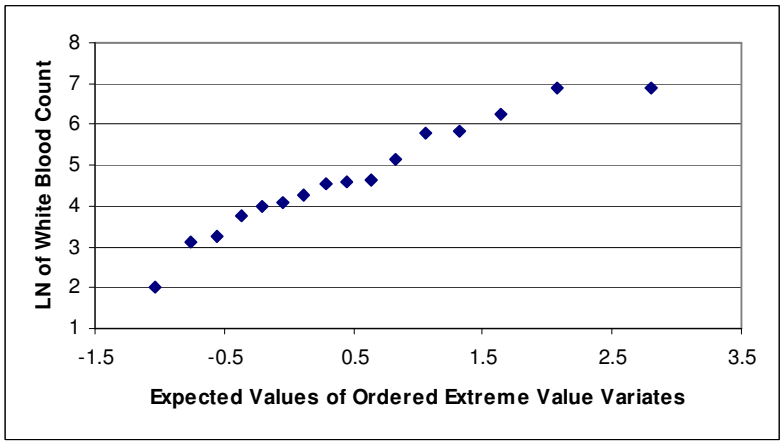

Figure 5.23 Q-Q plot of natural logarithm of White Cell Blood Count (*X*).

Now we derive the test statistics to test the mean vectors and the correlation coefficient. Since the sample size is not large enough to use the asymptotic covariance matrix, we give the results based on the simulated variances and covariances. The simulated variances and covariances of the MML and the LS estimators are given in Table 5.19.

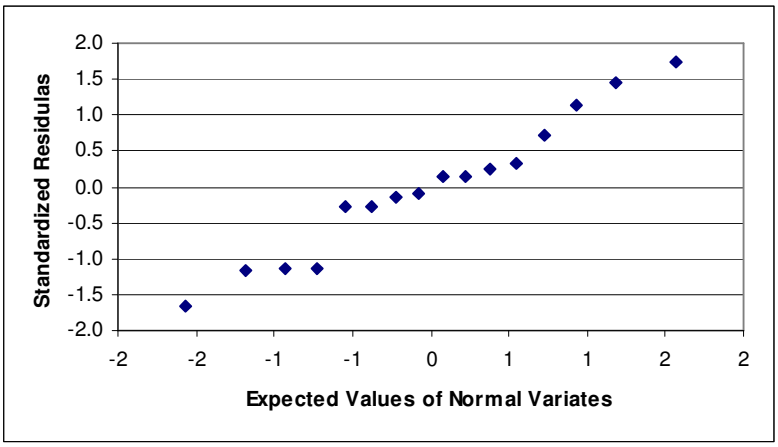

Figure 5.24 Q-Q plot of the standardized residuals.

Table 5.19 Simulated variances and covariances of the MML and the LS estimators for $\rho = -0.74$, and simulated variance of $\hat{\rho}$ under $H_0 : \rho = 0$ for the sample size of $n = 16$.

	$\text{var}(\hat{\mu}_1)$	$\text{var}(\hat{\mu}_2)$	$\text{cov}(\hat{\mu}_1, \hat{\mu}_2)$	$\text{var}(\hat{\mu}_{2.1})$	$\text{var}(\hat{\rho}\ \textit{under}\ H_0 : \rho = 0)$
MML	0.0714	0.0740	-0.0525	0.0366	0.0508
LS	0.0743	0.0749	-0.0549	0.0366	0.0659

To test the mean vector $H_0 : (\mu_1, \mu_2) = (0,0)$, the statistic based on the MML estimators is

$$\hat{\mathrm{T}}^2 = n(\hat{\mu}_1, \hat{\mu}_2)\, \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix}. \qquad (5.5.1.1)$$

Here, $\hat{\mathbf{\Omega}}$ is the (estimated) covariance matrix

$$\hat{\mathbf{\Omega}} = n\begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{13} \\ \hat{\sigma}_{13} & \hat{\sigma}_{33} \end{bmatrix}. \qquad (5.5.1.2)$$

We use the simulated variance of $\hat{\mu}_1$ and $\hat{\mu}_2$ instead of $\hat{\sigma}_{11}$ and $\hat{\sigma}_{33}$, respectively. In the same way, the simulated covariance between $\hat{\mu}_1$ and $\hat{\mu}_2$ is used instead of $\hat{\sigma}_{13}$. The corresponding LS statistic is

$$\tilde{\mathrm{T}}^2 = n(\tilde{\mu}_1, \tilde{\mu}_2)\,\tilde{\mathbf{\Omega}}^{-1}\begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix}. \qquad (5.5.1.3)$$

Here, $\tilde{\mathbf{\Omega}}$ is the (estimated) covariance matrix

$$\tilde{\mathbf{\Omega}} = n\begin{bmatrix} \tilde{\sigma}_{11} & \tilde{\sigma}_{13} \\ \tilde{\sigma}_{13} & \tilde{\sigma}_{33} \end{bmatrix}. \qquad (5.5.1.4)$$

In a similar manner, we replace the elements of $\tilde{\mathbf{\Omega}}$ by the simulated ones from Table 5.19.

Then,

$$\begin{aligned} \hat{\mathrm{T}}^2 &= n(\hat{\mu}_1, \hat{\mu}_2)\,\hat{\mathbf{\Omega}}^{-1}\begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \\ &= 16(3.9868, 83.7533)\left[16\begin{bmatrix} 0.0714 & -0.0525 \\ -0.0525 & 0.0740 \end{bmatrix}\right]^{-1}\begin{pmatrix} 3.9868 \\ 83.7533 \end{pmatrix} \\ &= (3.9868, 83.7533)\begin{bmatrix} 0.0714 & -0.0525 \\ -0.0525 & 0.0740 \end{bmatrix}^{-1}\begin{pmatrix} 3.9868 \\ 83.7533 \end{pmatrix} \end{aligned}$$

$$= (3.9868, 83.7533)\begin{bmatrix} 29.2797 & 20.7727 \\ 20.7727 & 28.2509 \end{bmatrix}\begin{pmatrix} 3.9868 \\ 83.7533 \end{pmatrix}$$

$$= 212{,}507.18\,. \qquad (5.5.1.5)$$

The corresponding LS statistic is

$$\tilde{\mathrm{T}}^2 = n(\tilde{\mu}_1, \tilde{\mu}_2)\,\tilde{\boldsymbol{\Omega}}^{-1}\begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix}$$

$$= 16(4.0749, 82.9921)\left[16\begin{bmatrix} 0.0743 & -0.0549 \\ -0.0549 & 0.0749 \end{bmatrix}\right]^{-1}\begin{pmatrix} 4.0749 \\ 82.9921 \end{pmatrix}$$

$$= (4.0749, 82.9921)\begin{bmatrix} 0.0743 & -0.0549 \\ -0.0549 & 0.0749 \end{bmatrix}^{-1}\begin{pmatrix} 4.0749 \\ 82.9921 \end{pmatrix}$$

$$= (4.0749, 82.9921)\begin{bmatrix} 29.3603 & 21.5205 \\ 21.5205 & 29.1251 \end{bmatrix}\begin{pmatrix} 4.0749 \\ 82.9921 \end{pmatrix}$$

$$= 215{,}648.25. \qquad (5.5.1.6)$$

We either compare them with $\chi^2_{0.05}(2) = 5.99$ by using their asymptotic distribution or calculate $\frac{(n-2)}{2(n-1)}\hat{T}^2$ and $\frac{(n-2)}{2(n-1)}\tilde{T}^2$ and compare them with $F_{0.05}(2, n-2)$. Since both of them are greater than $\chi^2_{0.05}(2) = 5.99$, the null hypothesis is rejected by both of the statistics. To compare them with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}^2 = (14/30)\,212{,}507.18 = 99{,}170.02 \qquad (5.5.1.7)$$

and

$$\frac{(n-2)}{2(n-1)}\tilde{T}^2 = (14/30)\,215{,}648.25 = 100{,}635.85\,. \qquad (5.5.1.8)$$

We again reject the null hypothesis by both methods since $F_{0.05}(2,14) = 3.74$.

Alternatively, to test H_0,

$$\hat{\mathrm{T}}_1^2 = n(\hat{\mu}_1, \hat{\mu}_{2.1})\,\hat{\mathbf{\Omega}}^{-1}\begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_{2.1}\end{pmatrix}$$

$$= \hat{\mu}_1{}^2/\hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2/\hat{\sigma}_{33}$$

$$= (3.9868)^2/0.0714 + (204.2984)^2/0.0366 = 1{,}140{,}600.65. \quad (5.5.1.9)$$

The corresponding LS statistic is

$$\tilde{\mathrm{T}}_1^2 = n(\tilde{\mu}_1, \tilde{\mu}_{2.1})\,\tilde{\mathbf{\Omega}}^{-1}\begin{pmatrix}\tilde{\mu}_1\\ \tilde{\mu}_{2.1}\end{pmatrix}$$

$$= \tilde{\mu}_1{}^2/\tilde{\sigma}_{11} + \tilde{\mu}_{2.1}{}^2/\tilde{\sigma}_{33}$$

$$= (4.0749)^2/0.0743 + (204.2984)^2/0.0366 = 1{,}140{,}601.52. \quad (5.5.1.10)$$

The null hypothesis is rejected since both of them are greater than $\chi^2_{0.05}(2) = 5.99$.

Also,

$$\frac{(n-2)}{2(n-1)}\hat{T}_1^2 = (14/30)\,1{,}140{,}600.65 = 532{,}280.30 \quad (5.5.1.11)$$

and

$$\frac{(n-2)}{2(n-1)}\tilde{T}_1^2 = (14/30)\,1{,}140{,}601.52 = 532{,}280.71. \quad (5.5.1.12)$$

We again reject the null hypothesis by both methods since $F_{0.05}(2,14) = 3.74$.

To test $H_0 : \rho = 0$ against $H_1 : \rho < 0$, the proposed statistic is

$$W = \frac{\hat{\rho}}{\sqrt{6/(n\pi^2)}} = \hat{\rho}\pi\sqrt{\frac{n}{6}}. \qquad (5.5.1.13)$$

Here,

$$6/(n\pi^2) \qquad (5.5.1.14)$$

is the asymptotic variance of $\hat{\rho}$ under H_0. Because of small sample size, we use the simulated variance of $\hat{\rho}$ under H_0 instead of the asymptotic variance. Then the proposed statistic becomes

$$W = \hat{\rho}/\sqrt{0.0508}. \qquad (5.5.1.15)$$

The corresponding statistic based on the LS estimator is

$$W_{LS} = \tilde{\rho}/\sqrt{0.0659}. \qquad (5.5.1.16)$$

Then,

$$W = -0.7389/\sqrt{0.0508} = -3.278 \qquad (5.5.1.17)$$

and

$$W_{LS} = -0.7433/\sqrt{0.0659} = -2.895. \qquad (5.5.1.18)$$

We reject the null hypothesis for both methods since $z_{0.05} = -1.645$.

5.5.2 Examples Using Simulated Data

In this section we give examples using simulated data for various sample sizes. We simulate data for the situation when the marginal and the conditional

distributions both are Generalized Logistic with shape parameters $b_1 = 0.5$ and $b_2 = 1$, respectively. The other parameters are taken as $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 0$, $\sigma_2 = 1$, $\mu_{2.1} = 0$, $\sigma_{2.1} = 0.866$, $\rho = 0.5$ and $\theta_1 = 0.5$. Firstly, we give an illustration for $n = 20$. Since the sample size is not large enough to use the asymptotic covariance matrix, we use only the simulated variances and covariances. The simulated data for $n = 20$ are given in Table 5.20. The MML and LS estimates are given in Table 5.21 using the results given in Chapter 4. The simulated variances and covariances are given in Table 5.22.

Table 5.20 Simulated data for $n = 20$.

i	X_i	Y_i	i	X_i	Y_i
1	0.6111	0.7854	11	-1.3532	-3.4839
2	-1.6487	-5.3303	12	-2.5452	-1.6069
3	0.8257	2.6373	13	-3.4990	-2.9060
4	1.4621	1.0638	14	0.9557	0.1458
5	-0.7485	-1.0418	15	-0.5475	1.3740
6	-0.4574	2.5459	16	-3.4978	-2.6781
7	-0.7913	-1.8961	17	-0.6870	-1.1195
8	-5.2085	-2.6557	18	-0.8852	-0.3033
9	-0.3644	0.7956	19	-0.9960	0.8554
10	-2.2536	-1.4654	20	0.3194	-1.1990

Table 5.21 The MML and LS estimates for $n = 20$.

	μ_1	σ_1	μ_2	σ_2	ρ	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1
MML	-0.0799	0.6663	0.0176	1.0419	0.4737	0.0769	0.9176	0.7407
LS	-0.1603	0.6529	0.0771	0.9659	0.6357	0.0655	0.9077	0.7880

Table 5.22 Simulated variances and covariances of the MML and the LS estimators for $\rho = 0.5$, and simulated variance of $\hat{\rho}$ under $H_0 : \rho = 0$ for the sample size of $n = 20$.

	$\text{var}(\hat{\mu}_1)$	$\text{var}(\hat{\mu}_2)$	$\text{cov}(\hat{\mu}_1, \hat{\mu}_2)$	$\text{var}(\hat{\mu}_{2.1})$	$\text{var}(\hat{\rho}$ *under* $H_0 : \rho = 0)$
MML	0.2641	0.2193	0.1335	0.1563	0.0272
LS	0.2779	0.2311	0.1405	0.1668	0.0530

It is interesting to note that the variances of the MMLE are smaller than those of the LSE as for other data sets presented here.

To test the mean vector $H_0 : (\mu_1, \mu_2) = (0,0)$,

$$\begin{aligned}
\hat{\mathrm{T}}^2 &= n(\hat{\mu}_1, \hat{\mu}_2)\,\hat{\boldsymbol{\Omega}}^{-1}\begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{pmatrix} \\
&= (-0.0799, 0.0176)\begin{bmatrix}0.2641 & 0.1335\\ 0.1335 & 0.2193\end{bmatrix}^{-1}\begin{pmatrix}-0.0799\\ 0.0176\end{pmatrix} \\
&= (-0.0799, 0.0176)\begin{bmatrix}5.4695 & -3.3296\\ -3.3296 & 6.5869\end{bmatrix}\begin{pmatrix}-0.0799\\ 0.0176\end{pmatrix} \\
&= 0.0463 . \qquad (5.5.2.1)
\end{aligned}$$

The corresponding LS statistic is

$$\begin{aligned}
\tilde{\mathrm{T}}^2 &= n(\tilde{\mu}_1, \tilde{\mu}_2)\,\tilde{\boldsymbol{\Omega}}^{-1}\begin{pmatrix}\tilde{\mu}_1\\ \tilde{\mu}_2\end{pmatrix} \\
&= (-0.1603, 0.0771)\begin{bmatrix}0.2779 & 0.1405\\ 0.1405 & 0.2311\end{bmatrix}^{-1}\begin{pmatrix}-0.1603\\ 0.0771\end{pmatrix} \\
&= (-0.1603, 0.0771)\begin{bmatrix}5.1953 & -3.1585\\ -3.1585 & 6.2474\end{bmatrix}\begin{pmatrix}-0.1603\\ 0.0771\end{pmatrix} \\
&= 0.2487 . \qquad (5.5.2.2)
\end{aligned}$$

To compare them with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}^2 = (18/38)0.0463 = 0.0219 \qquad (5.5.2.3)$$

and

$$\frac{(n-2)}{2(n-1)}\tilde{T}^2 = (18/38)\,0.2487 = 0.1178 . \qquad (5.5.2.4)$$

Alternatively, to test H_0,

$$\hat{T}_1^2 = \hat{\mu}_1{}^2/\hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2/\hat{\sigma}_{33}$$
$$= (-0.0799)^2/0.2641 + (0.0769)^2/0.1563 = 0.0620. \qquad (5.5.2.5)$$

The corresponding LS statistic is

$$\tilde{T}_1^2 = \tilde{\mu}_1{}^2/\tilde{\sigma}_{11} + \tilde{\mu}_{2.1}{}^2/\tilde{\sigma}_{33}$$
$$= (-0.1603)^2/0.2779 + (0.0655)^2/0.1668 = 0.1182. \qquad (5.5.2.6)$$

To compare them with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}^2 = (18/38)0.0620 = 0.0294 \qquad (5.5.2.7)$$

and

$$\frac{(n-2)}{2(n-1)}\tilde{T}^2 = (18/38)0.1182 = 0.0560. \qquad (5.5.2.8)$$

To test $H_0 : \rho = 0$ against $H_1 : \rho > 0$ by using the simulated variances, our statistic based on the MML estimator is

$$W = \hat{\rho}/\sqrt{0.0272}. \qquad (5.5.2.9)$$

The corresponding statistic based on the LS estimator is

$$W_{LS} = \tilde{\rho}/\sqrt{0.0530}. \qquad (5.5.2.10)$$

Then,

$$W = 0.4737/\sqrt{0.0272} = 2.872 \tag{5.5.2.11}$$

and

$$W_{LS} = 0.6357/\sqrt{0.0530} = 2.761. \tag{5.5.2.12}$$

Now we give an example for a sample size $n = 30$ which is fairly large. The simulated data are given in Table 5.23. The MML and LS estimates and the simulated variances and covariances are given in Table 5.24 and Table 5.25, respectively.

Table 5.23 Simulated data for $n = 30$.

i	X_i	Y_i	i	X_i	Y_i
1	0.6669	1.1393	16	0.2683	2.8408
2	-1.5802	-0.6026	17	-6.2760	-1.5212
3	-1.5593	-0.3870	18	-3.0977	-1.3426
4	0.0538	0.6795	19	-1.6503	-0.6978
5	-3.1483	-5.7769	20	-2.0622	-1.6132
6	-0.0067	-1.3394	21	-1.3645	0.8223
7	-1.5705	-1.1441	22	-6.9407	-2.3434
8	-2.9500	0.7003	23	-0.8378	-2.6615
9	-2.9839	0.4107	24	-2.2439	-1.3998
10	-2.6043	-3.8578	25	-4.6253	-3.8861
11	-1.5338	-2.7978	26	0.4540	1.0792
12	-0.5314	2.2537	27	0.2349	1.8984
13	0.6660	-2.1259	28	0.4964	-1.6335
14	-2.4359	-1.7708	29	-4.7850	0.4714
15	0.9837	1.5028	30	0.5023	-1.4676

Table 5.24 The MML and LS estimates for $n = 30$.

	μ_1	σ_1	μ_2	σ_2	ρ	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1
MML	-0.3924	0.8393	-0.2767	1.1165	0.2789	-0.1311	1.0722	0.3711
LS	-0.5641	0.8065	-0.2343	1.0001	0.4218	-0.1439	1.0015	0.4014

Table 5.25 Simulated variances and covariances of the MML and the LS estimators for $\rho = 0.5$, and simulated variance of $\hat{\rho}$ under $H_0 : \rho = 0$ for the sample size of $n = 30$.

	$\text{var}(\hat{\mu}_1)$	$\text{var}(\hat{\mu}_2)$	$\text{cov}(\hat{\mu}_1, \hat{\mu}_2)$	$\text{var}(\hat{\mu}_{2.1})$	$\text{var}(\hat{\rho}$ *under* $H_0 : \rho = 0)$
MML	0.1769	0.1441	0.0886	0.1019	0.0169
LS	0.1885	0.1564	0.0956	0.1110	0.0343

To test the mean vector $H_0 : (\mu_1, \mu_2) = (0,0)$,

$$\begin{aligned}
\hat{\mathrm{T}}^2 &= n(\hat{\mu}_1, \hat{\mu}_2)\,\hat{\boldsymbol{\Omega}}^{-1}\begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{pmatrix} \\
&= (-0.3924, -0.2767)\begin{bmatrix}0.1769 & 0.0886\\ 0.0886 & 0.1441\end{bmatrix}^{-1}\begin{pmatrix}-0.3924\\ -0.2767\end{pmatrix} \\
&= (-0.3924, -0.2767)\begin{bmatrix}8.1693 & -5.0223\\ -5.0223 & 10.0276\end{bmatrix}\begin{pmatrix}-0.3924\\ -0.2767\end{pmatrix} \\
&= 0.9349 . \qquad (5.5.2.13)
\end{aligned}$$

The corresponding LS statistic is

$$\begin{aligned}
\tilde{\mathrm{T}}^2 &= n(\tilde{\mu}_1, \tilde{\mu}_2)\,\tilde{\boldsymbol{\Omega}}^{-1}\begin{pmatrix}\tilde{\mu}_1\\ \tilde{\mu}_2\end{pmatrix} \\
&= (-0.5641, -0.2343)\begin{bmatrix}0.1885 & 0.0956\\ 0.0956 & 0.1564\end{bmatrix}^{-1}\begin{pmatrix}-0.5641\\ -0.2343\end{pmatrix} \\
&= (-0.5641, -0.2343)\begin{bmatrix}7.6885 & -4.6996\\ -4.6996 & 9.2665\end{bmatrix}\begin{pmatrix}-0.5641\\ -0.2343\end{pmatrix} \\
&= 1.7130 . \qquad (5.5.2.14)
\end{aligned}$$

To compare them with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}^2 = (28/58)\,0.9349 = 0.4513 \qquad (5.5.2.15)$$

and

$$\frac{(n-2)}{2(n-1)}\tilde{T}^2 = (28/58)\,1.7130 = 0.8269\,. \qquad (5.5.2.16)$$

Alternatively, to test H_0,

$$\begin{aligned}\hat{T}_1^2 &= \hat{\mu}_1{}^2/\hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2/\hat{\sigma}_{33} \\ &= (-0.3924)^2/0.1769 + (-0.1311)^2/0.1019 = 1.0391\,. \end{aligned} \qquad (5.5.2.17)$$

The corresponding LS statistic is

$$\begin{aligned}\tilde{T}_1^2 &= \tilde{\mu}_1{}^2/\tilde{\sigma}_{11} + \tilde{\mu}_{2.1}{}^2/\tilde{\sigma}_{33} \\ &= (-0.5641)^2/0.1885 + (-0.1439)^2/0.1110 = 1.8747\,. \end{aligned} \qquad (5.5.2.18)$$

To compare them with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}^2 = (28/58)\,1.0391 = 0.5016 \qquad (5.5.2.19)$$

and

$$\frac{(n-2)}{2(n-1)}\tilde{T}^2 = (28/58)\,1.8747 = 0.9050\,. \qquad (5.5.2.20)$$

For the sample size $n = 30$, we can also calculate the statistics based on MML estimators by using the asymptotic covariance matrix (estimated by using the MML estimates). We first calculate the components of the estimated Fisher information matrices $\hat{I}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$ and $\hat{I}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$ by using the

results given in Chapter 4. These matrices for the simulated data with the sample size of $n = 30$ are given in Table 5.26 and Table 5.27, respectively.

Table 5.26 The estimated Fisher information matrix, $\hat{I}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$.

	μ_1	σ_1	μ_2	σ_2	ρ
μ_1	9.7157	-4.9510	-3.2280	1.2483	4.9963
σ_1	-4.9510	67.4998	4.4750	-7.6551	-30.6399
μ_2	-3.2280	4.4750	8.6990	-3.3639	-13.4642
σ_2	1.2483	-7.6551	-3.3639	40.1676	11.4105
ρ	4.9963	-30.6399	-13.4642	11.4105	96.1134

Table 5.27 The estimated Fisher information matrix, $\hat{I}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$.

	μ_1	σ_1	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1
μ_1	8.5179	-3.2904	0.0000	0.0000	0.0000
σ_1	-3.2904	57.3162	0.0000	0.0000	0.0000
$\mu_{2.1}$	0.0000	0.0000	8.6990	0.0000	-10.1212
$\sigma_{2.1}$	0.0000	0.0000	0.0000	37.3168	0.0000
θ_1	0.0000	0.0000	-10.1212	0.0000	52.0934

The estimated asymptotic covariance matrices are $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$ and $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$ which are given in Table 5.28 and Table 5.29, respectively.

Table 5.28 The estimated asymptotic covariance matrix, $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$.

	μ_1	σ_1	μ_2	σ_2	ρ
μ_1	0.1201	0.0069	0.0446	0.0007	0.0021
σ_1	0.0069	0.0178	0.0026	0.0018	0.0055
μ_2	0.0446	0.0026	0.1651	0.0070	0.0208
σ_2	0.0007	0.0018	0.0070	0.0263	-0.0016
ρ	0.0021	0.0055	0.0208	-0.0016	0.0151

Table 5.29 The estimated asymptotic covariance matrix, $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$.

	μ_1	σ_1	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1
μ_1	0.1201	0.0069	0.0000	0.0000	0.0000
σ_1	0.0069	0.0178	0.0000	0.0000	0.0000
$\mu_{2.1}$	0.0000	0.0000	0.1485	0.0000	0.0289
$\sigma_{2.1}$	0.0000	0.0000	0.0000	0.0268	0.0000
θ_1	0.0000	0.0000	0.0289	0.0000	0.0248

To test the mean vector $H_0 : (\mu_1, \mu_2) = (0,0)$ by using the asymptotic covariance matrix, we take the components $\hat{\sigma}_{11}$, $\hat{\sigma}_{33}$ and $\hat{\sigma}_{13}$ from the estimated asymptotic covariance matrix $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$ where they correspond to the asymptotic variance of $\hat{\mu}_1$, $\hat{\mu}_2$ and the asymptotic covariance of $(\hat{\mu}_1, \hat{\mu}_2)$, respectively (see Table 5.28). Then,

$$
\begin{aligned}
\hat{\mathrm{T}}^2 &= n(\hat{\mu}_1, \hat{\mu}_2)\, \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \\
&= n(\hat{\mu}_1, \hat{\mu}_2) \left\{ n \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{13} \\ \hat{\sigma}_{13} & \hat{\sigma}_{33} \end{bmatrix} \right\}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \\
&= (\hat{\mu}_1, \hat{\mu}_2) \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{13} \\ \hat{\sigma}_{13} & \hat{\sigma}_{33} \end{bmatrix}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \\
&= (-0.3924, -0.2767) \begin{bmatrix} 0.1201 & 0.0446 \\ 0.0446 & 0.1651 \end{bmatrix}^{-1} \begin{pmatrix} -0.3924 \\ -0.2767 \end{pmatrix} \\
&= (-0.3924, -0.2767) \begin{bmatrix} 9.2548 & -2.5001 \\ -2.5001 & 6.7323 \end{bmatrix} \begin{pmatrix} -0.3924 \\ -0.2767 \end{pmatrix} \\
&= 1.3976 .
\end{aligned}
\tag{5.5.2.21}
$$

To compare with F,

$$
\frac{(n-2)}{2(n-1)} \hat{T}^2 = (28/58)\, 1.3976 = 0.6747 . \tag{5.5.2.22}
$$

Alternatively, to test H_0, we use the following statistic by using the asymptotic covariance matrix with parameters μ_1, σ_1, $\mu_{2.1}$, $\sigma_{2.1}$, θ_1. We take the components $\hat{\sigma}_{11}$ and $\hat{\sigma}_{33}$ and from $\hat{I}^{-1}(\mu_1,\sigma_1,\mu_{2.1},\sigma_{2.1},\theta_1)$ where they correspond to the asymptotic variances of $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$, respectively (see Table 5.29). Then,

$$\begin{aligned}\hat{\mathrm{T}}_1^2 &= \hat{\mu}_1{}^2/\hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2/\hat{\sigma}_{33} \\ &= (-0.3924)^2/0.1201 + (-0.1311)^2/0.1485 = 1.3978 . \end{aligned} \quad (5.5.2.23)$$

To compare with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}_1^2 = (28/58)\,1.3978 = 0.6748 . \quad (5.5.2.24)$$

In fact the values of $\hat{T}^2$ and $\hat{T}_1^2$ are exactly the same since the random sample is generated under $H_0:(\mu_1,\mu_2)=(0,0)$ and the invariance properties of the MML estimators hold. However, they take slightly different values 1.3976 and 1.3978, respectively. This is due to the rounding error.

From Table 5.25, the simulated variances of $\hat{\rho}$ and $\tilde{\rho}$ under $H_0:\rho=0$ are 0.0169 and 0.0343, respectively. Thus, to test $H_0:\rho=0$ against $H_1:\rho>0$ by using the simulated variances, our statistic based on the MML estimator is

$$W = \hat{\rho}/\sqrt{0.0169} . \quad (5.5.2.25)$$

The corresponding statistic based on the LS estimator is

$$W_{LS} = \tilde{\rho}/\sqrt{0.0343} . \quad (5.5.2.26)$$

Thus,

$$W = 0.2789 \Big/ \sqrt{0.0169} = 2.145 \tag{5.5.2.27}$$

and

$$W_{LS} = 0.4218 \Big/ \sqrt{0.0343} = 2.278 . \tag{5.5.2.28}$$

To test $H_0 : \rho = 0$ against $H_1 : \rho > 0$ by using the asymptotic variance, our statistic based on the MML estimator is

$$\begin{aligned} W &= \hat{\rho} \Big/ \sqrt{\frac{1}{n} \frac{(b_2 + 2)}{b_2} \frac{1}{\left(\psi'(b_1) + \psi'(1)\right)}} \\ &= \hat{\rho} \sqrt{n \frac{b_2}{(b_2 + 2)} \left(\psi'(b_1) + \psi'(1)\right)} . \end{aligned} \tag{5.5.2.29}$$

Here,

$$\frac{1}{n} \frac{(b_2 + 2)}{b_2} \frac{1}{\left(\psi'(b_1) + \psi'(1)\right)} \tag{5.5.2.30}$$

is the asymptotic variance of $\hat{\rho}$ under H_0. For our case when $b_1 = 0.5$, $b_2 = 1$ and $n = 30$, the asymptotic variance of $\hat{\rho}$ under H_0 is

$$\frac{1}{n} \frac{(b_2 + 2)}{b_2} \frac{1}{\left(\psi'(b_1) + \psi'(1)\right)} = \frac{1}{30} \frac{(1 + 2)}{1} \frac{1}{\left(\psi'(0.5) + \psi'(1)\right)} = 0.0152 . \tag{5.5.2.31}$$

Then,

$$W = 0.2789 / 0.0152 = 2.262 . \tag{5.5.2.32}$$

Note how close the asymptotic variance (0.0152) and the simulated variance (0.0169) are. This is because of the fact that the asymptotic variance of $\hat{\rho}$ under H_0 does not depend on a parameter other than the shape parameter b_2. Thus, it is not affected by the sample observations. Here, the shape parameter b_2 is accepted to be known, but if it is not known, a plausible value for b_2 can be selected by examining the Q-Q plots of the residuals for various values of b_2. In a neighborhood of the true value of b_2, it does not create any problem because of the robustness features of the MMLE.

We finally give an example for the sample size $n = 50$ which is reasonably large. The simulated data are given in Table 5.30. The MML and LS estimates and the simulated variances and covariances are given in Table 5.31 and Table 5.32, respectively.

Table 5.30 Simulated data for $n = 50$.

i	X_i	Y_i	i	X_i	Y_i
1	0.6835	2.2477	26	-4.2547	-2.5344
2	-0.0916	1.1876	27	-1.0503	-0.7988
3	-2.3285	-2.0603	28	-5.0743	-5.5956
4	-3.3540	-1.9623	29	-3.8151	-0.8886
5	-2.8114	-0.7293	30	0.8401	-0.2699
6	-1.5234	-2.0661	31	-0.4681	-0.2539
7	-4.4988	-2.3772	32	-1.9460	-2.7240
8	0.3929	-0.4555	33	3.5027	0.6134
9	1.3894	0.8483	34	-1.1407	-2.2441
10	-2.6403	-0.3914	35	-3.1005	-1.8102
11	-7.7905	-5.8337	36	-0.0829	0.4360
12	0.3042	0.1504	37	-1.1216	-1.3454
13	-0.4516	-2.0239	38	-6.4154	-3.1029
14	-2.0072	-0.5461	39	-2.3739	-2.0689
15	-1.5964	1.0274	40	-1.8839	-1.1435
16	1.3988	0.7220	41	2.9904	-0.2193
17	-1.9833	-2.4168	42	1.1398	0.6691
18	-1.3818	0.3607	43	-1.9728	-1.6492
19	2.9700	3.4914	44	-0.7201	-2.4401
20	0.4868	0.9639	45	-1.9394	-0.7457
21	2.0186	0.9996	46	-4.4113	-2.7447
22	1.2531	0.5313	47	-0.1953	-0.8101
23	2.0595	2.5948	48	-9.4126	-5.0426
24	-5.4119	-2.3795	49	1.3961	-2.5010
25	-2.0811	0.4135	50	-2.0768	0.7598

Table 5.31 The MML and LS estimates for $n = 50$.

	μ_1	σ_1	μ_2	σ_2	ρ	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1
MML	0.1136	1.1092	-0.0682	0.9215	0.6520	-0.1297	0.6987	0.5416
LS	0.0466	1.0519	-0.0270	0.8331	0.7759	-0.1467	0.6721	0.5500

Table 5.32 Simulated variances and covariances of the MML and the LS estimators for $\rho = 0.5$, and simulated variance of $\hat{\rho}$ under $H_0 : \rho = 0$ for the sample size of $n = 50$.

	var($\hat{\mu}_1$)	var($\hat{\mu}_2$)	cov($\hat{\mu}_1, \hat{\mu}_2$)	var($\hat{\mu}_{2.1}$)	var($\hat{\rho}$ *under* $H_0 : \rho = 0$)
MML	0.1059	0.0867	0.0529	0.0611	0.0098
LS	0.1141	0.0941	0.0572	0.0664	0.0202

To test the mean vector $H_0 : (\mu_1, \mu_2) = (0,0)$,

$$\begin{aligned}
\hat{\mathrm{T}}^2 &= n(\hat{\mu}_1, \hat{\mu}_2)\, \hat{\mathbf{\Omega}}^{-1} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{pmatrix} \\
&= (0.1136, -0.0682) \begin{bmatrix} 0.1059 & 0.0529 \\ 0.0529 & 0.0867 \end{bmatrix}^{-1} \begin{pmatrix} 0.1136 \\ -0.0682 \end{pmatrix} \\
&= (0.1136, -0.0682) \begin{bmatrix} 13.5827 & -8.2875 \\ -8.2875 & 16.5906 \end{bmatrix} \begin{pmatrix} 0.1136 \\ -0.0682 \end{pmatrix} \\
&= 0.3809 . \qquad (5.5.2.33)
\end{aligned}$$

The corresponding LS statistic is

$$\begin{aligned}
\tilde{\mathrm{T}}^2 &= n(\tilde{\mu}_1, \tilde{\mu}_2)\, \tilde{\mathbf{\Omega}}^{-1} \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix} \\
&= (0.0466, -0.0270) \begin{bmatrix} 0.1141 & 0.0572 \\ 0.0572 & 0.0941 \end{bmatrix}^{-1} \begin{pmatrix} 0.0466 \\ -0.0270 \end{pmatrix}
\end{aligned}$$

$$= (0.0466,-0.0270)\begin{bmatrix} 12.6055 & -7.6625 \\ -7.6625 & 15.2847 \end{bmatrix}\begin{pmatrix} 0.0466 \\ -0.0270 \end{pmatrix}$$

$$= 0.0578\,. \qquad (5.5.2.34)$$

To compare them with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}^2 = (48/98)\,0.3809 = 0.1865 \qquad (5.5.2.35)$$

and

$$\frac{(n-2)}{2(n-1)}\tilde{T}^2 = (18/38)\,0.0578 = 0.0283\,. \qquad (5.5.2.36)$$

Alternatively, to test H_0,

$$\hat{\mathrm{T}}_1^2 = \hat{\mu}_1{}^2/\hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2/\hat{\sigma}_{33}$$

$$= (0.1136)^2/0.1059 + (-0.1297)^2/0.0611 = 0.3972\,. \qquad (5.5.2.37)$$

The corresponding LS statistic is

$$\tilde{\mathrm{T}}_1^2 = \tilde{\mu}_1{}^2/\tilde{\sigma}_{11} + \tilde{\mu}_{2.1}{}^2/\tilde{\sigma}_{33}$$

$$= (0.0466)^2/0.1141 + (-0.1467)^2/0.0664 = 0.3431\,. \qquad (5.5.2.38)$$

To compare them with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}^2 = (48/98)\,0.3972 = 0.1945 \qquad (5.5.2.39)$$

and

$$\frac{(n-2)}{2(n-1)}\tilde{T}^2 = (48/98)\,0.3431 = 0.1681\,. \qquad (5.5.2.40)$$

For the sample size $n = 50$, we can also calculate the statistics based on MML estimators by using the asymptotic covariance matrix (estimated by using the MML estimates). We first calculate the components of the estimated Fisher information matrices $\hat{I}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$ and $\hat{I}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$ by using the results given in Chapter 4. These matrices for the simulated data with the sample size $n = 50$ are given in Table 5.33 and Table 5.34.

Table 5.33 The estimated Fisher information matrix, $\hat{I}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$.

	μ_1	σ_1	μ_2	σ_2	ρ
μ_1	18.1432	-17.0241	-18.4916	16.7132	23.6216
σ_1	-17.0241	139.8374	25.6347	-102.4946	-144.8602
μ_2	-18.4916	25.6347	34.141	-30.8577	-43.6126
σ_2	16.7132	-102.4946	-30.8577	207.5803	86.3867
ρ	23.6216	-144.8602	-43.6126	86.3867	338.3961

Table 5.34 The estimated Fisher information matrix, $\hat{I}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$.

	μ_1	σ_1	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1
μ_1	8.1278	-3.1397	0.0000	0.0000	0.0000
σ_1	-3.1397	54.6910	0.0000	0.0000	0.0000
$\mu_{2.1}$	0.0000	0.0000	34.1410	0.0000	-52.4985
$\sigma_{2.1}$	0.0000	0.0000	0.0000	146.4581	0.0000
θ_1	0.0000	0.0000	-52.4985	0.0000	357.1101

The estimated asymptotic covariance matrices are $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$ and $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$ which are given in Table 35 and Table 36, respectively.

To test the mean vector $H_0 : (\mu_1, \mu_2) = (0,0)$ by using the asymptotic covariance matrix, we take the components $\hat{\sigma}_{11}$, $\hat{\sigma}_{33}$ and $\hat{\sigma}_{13}$ from $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$ where they correspond to the asymptotic variance of $\hat{\mu}_1$, $\hat{\mu}_2$ and the asymptotic covariance of $(\hat{\mu}_1, \hat{\mu}_2)$, respectively (see Table 5.35). Then,

$$\hat{\mathrm{T}}^2 = n(\hat{\mu}_1, \hat{\mu}_2)\,\hat{\boldsymbol{\Omega}}^{-1}\begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{pmatrix}$$

$$= (\hat{\mu}_1, \hat{\mu}_2)\begin{bmatrix}\hat{\sigma}_{11} & \hat{\sigma}_{13}\\ \hat{\sigma}_{13} & \hat{\sigma}_{33}\end{bmatrix}^{-1}\begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{pmatrix}$$

$$= (0.1136, -0.0682)\begin{bmatrix}0.1258 & 0.0682\\ 0.0682 & 0.0748\end{bmatrix}^{-1}\begin{pmatrix}0.1136\\ -0.0682\end{pmatrix}$$

$$= (0.1136, -0.0682)\begin{bmatrix}15.7189 & -14.3319\\ -14.3319 & 26.4363\end{bmatrix}\begin{pmatrix}0.1136\\ -0.0682\end{pmatrix}$$

$$= 0.5479\,. \qquad (5.5.2.41)$$

To compare with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}^2 = (48/98)\,0.5479 = 0.2684\,. \qquad (5.5.2.42)$$

Table 5.35 The estimated asymptotic covariance matrix, $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)$.

	μ_1	σ_1	μ_2	σ_2	ρ
μ_1	0.1258	0.0072	0.0682	0.0026	0.0024
σ_1	0.0072	0.0187	0.0039	0.0066	0.0063
μ_2	0.0682	0.0039	0.0748	0.0054	0.0052
σ_2	0.0026	0.0066	0.0054	0.0081	0.0013
ρ	0.0024	0.0063	0.0052	0.0013	0.0058

Table 5.36 The estimated asymptotic covariance matrix, $\hat{I}^{-1}(\mu_1, \sigma_1, \mu_{2.1}, \sigma_{2.1}, \theta_1)$.

	μ_1	σ_1	$\mu_{2.1}$	$\sigma_{2.1}$	θ_1
μ_1	0.1258	0.0072	0.0000	0.0000	0.0000
σ_1	0.0072	0.0187	0.0000	0.0000	0.0000
$\mu_{2.1}$	0.0000	0.0000	0.0378	0.0000	0.0056
$\sigma_{2.1}$	0.0000	0.0000	0.0000	0.0068	0.0000
θ_1	0.0000	0.0000	0.0056	0.0000	0.0036

Alternatively, to test H_0, we use the following statistic by using the estimated asymptotic covariance matrix with parameters μ_1, σ_1, $\mu_{2.1}$, $\sigma_{2.1}$, θ_1. We take the components $\hat{\sigma}_{11}$ and $\hat{\sigma}_{33}$ and from $\hat{I}^{-1}(\mu_1,\sigma_1,\mu_{2.1},\sigma_{2.1},\theta_1)$ where they correspond to the asymptotic variances of $\hat{\mu}_1$ and $\hat{\mu}_{2.1}$, respectively (see Table 5.36). Then,

$$\begin{aligned}\hat{\mathrm{T}}_1^2 &= \hat{\mu}_1{}^2/\hat{\sigma}_{11} + \hat{\mu}_{2.1}{}^2/\hat{\sigma}_{33} \\ &= (0.1136)^2/0.1258 + (-0.1297)^2/0.0378 = 0.5476. \end{aligned} \quad (5.5.2.43)$$

To compare with F,

$$\frac{(n-2)}{2(n-1)}\hat{T}^2 = (28/58)\,0.5476 = 0.2682. \quad (5.5.2.44)$$

From Table 5.32, the simulated variances of $\hat{\rho}$ and $\tilde{\rho}$ under $H_0: \rho = 0$ are 0.0098 and 0.0202, respectively. Thus, to test $H_0: \rho = 0$ against $H_1: \rho > 0$ by using the simulated variances, our statistic based on the MML estimator is

$$W = \hat{\rho}/\sqrt{0.0098}. \quad (5.5.2.45)$$

The corresponding statistic based on the LS estimator is

$$W_{LS} = \tilde{\rho}/\sqrt{0.0202}. \quad (5.5.2.46)$$

Then,

$$W = 0.6520/\sqrt{0.0098} = 6.586 \quad (5.5.2.47)$$

and

$$W_{LS} = 0.7759 \Big/ \sqrt{0.0202} = 5.459 . \qquad (5.5.2.48)$$

To test $H_0 : \rho = 0$ against $H_1 : \rho > 0$ by using the asymptotic variance, our statistic based on the MML estimator is

$$W = \hat{\rho} \Big/ \sqrt{\frac{1}{n}\frac{(b_2+2)}{b_2}\frac{1}{\left(\psi'(b_1)+\psi'(1)\right)}}$$

$$= \hat{\rho}\sqrt{n\frac{b_2}{(b_2+2)}\left(\psi'(b_1)+\psi'(1)\right)} . \qquad (5.5.2.49)$$

Since

$$\frac{1}{n}\frac{(b_2+2)}{b_2}\frac{1}{\left(\psi'(b_1)+\psi'(1)\right)} = \frac{1}{50}\frac{(1+2)}{1}\frac{1}{\left(\psi'(0.5)+\psi'(1)\right)} = 0.0091, \qquad (5.5.2.50)$$

$$W = 0.6520 \Big/ \sqrt{0.0091} = 6.835 . \qquad (5.5.2.51)$$

Since the sample size is reasonably large, the asymptotic variance of $\hat{\rho}$ (0.0091) is closer to the simulated variance of $\hat{\rho}$ (0.0098) as compared to the results given in the earlier example with sample size $n = 30$.

CHAPTER 6

CONCLUSION AND DISCUSSION

In a linear regression model, the design variable X is usually assumed to be nonstochastic and the distribution of the error of measurement e is assumed to be normal. In practice, however, X may be stochastic and e might not be normally distributed. In this thesis, we considered the following four possibilities:

(a) X is normal and $e = Y - \mu_2 - \rho(\sigma_2 / \sigma_1)(x - \mu_1)$ is normal

(b) X is non-normal and e is normal

(c) X is normal and e is non-normal

and the most difficult and important situation when

(d) X and e are both non-normal.

There is no previous work in the areas (c) and (d).

The joint distribution of X and Y involve five parameters which are not functionally related to one another. The five paremeters are the location parameter μ_1 and scale parameter σ_1 of X, the location parameter μ_2 and scale parameter σ_2 of Y, and the correlation coefficient ρ between X and Y. Another parameter of interest is the regression coefficient $\theta_1 = \rho(\sigma_2 / \sigma_1)$ which, of course, is functionally related. Consequently, its estimator can be obtained from those of σ_1, σ_2 and ρ.

In situation (a), the maximum likelihood estimators are the sample means and standard deviations and the sample correlation coefficient. The maximum likelihood estimators are intractable in situations (b)-(d). Thus, we have obtained the modified maximum likelihood estimators. They are explicit functions of sample observations and, therefore, easy to compute. We have shown that they have all the optimal properties of the maximum likelihood estimators. We have also shown that they are robust and enormously more efficient than the commonly used least squares estimators. Finally, we have given a real life application and some examples using simulated data.

In Chapter 1, we review the literature on the estimation and hypothesis testing of the parameters in a linear regression model when the design variable X is nonstochastic and the response variable Y has a location-scale distribution, normal or nonnormal. In this chapter, we also review the work of Vaughan and Tiku (2000) which covers the situation when X is stochastic and the conditional distribution is normal. In Chapter 2, we deal with the situation when X is stochastic. For illustration, we assume that the distribution of X is Weibull which is one of the most important distributions from applications point of view. The conditional distribution of Y given $X = x$ is taken to be normal. As a second situation, we assume that the distribution of X is Generalized Logistic. In Chapter 3, we consider the situation when the marginal distribution of X is normal and the conditional distribution of Y given $X = x$ is nonnormal. In Chapter 4, we deal with the most difficult situation when the distribution of X and the conditional distribution of Y are both nonnormal. In all the situations above, we derive the MML estimators of the five parameters μ_1 and σ_1, μ_2 and σ_2, and the correlation coefficient ρ. Another parameter of interest is the regression coefficient $\theta_1 = \rho(\sigma_2 / \sigma_1)$. We have also obtained its MMLE. We have shown that the MMLE have the following properties:

(i) They are explicit functions of sample observations and, therefore, easy to compute.

(ii) They are asymptotically fully efficient.

(iii) They are highly efficient for small sample sizes since their variances are only marginally bigger than the minimum variance bounds.

(iv) They have the very desirable invariance property, i.e., if $\hat{\theta}$ is the MMLE of θ, then $\tau(\hat{\theta})$ is the MMLE of a one-to-one function $\tau(\theta)$.

We develop T^2 statistics for testing the null hypothesis $H_0 : \underset{\sim}{\mu} = \underset{\sim}{0}$. We show that these statistics are more powerful than the normal theory statistics. We also study the robustness of the procedures to outliers, mixtures and contaminations. Also, we develop a procedure for testing the null hypothesis $H_0 : \rho = 0$ and again show that it is more powerful than the classical test statistic based on the Pearson sample correlation coefficient. Finally, we give one real life example and three computer generated examples.

For a comprehensive study of the MML estimators and their optimality properties (and numerous applications), see Tiku and Akkaya (2004). Akkaya and Tiku (2004) introduce a model for generating inliers and extend the methodology of modified likelihood to analyze univariate samples containing inliers. We did not have an opportunity of discussing this work in this thesis since the work has not yet appeared in print.

REFERENCES

Abramowitz, M., and Stegun, I.A. (1965) *Handbook of Mathematical Functions*, New York, Dover.

Akkaya, A.D. and Tiku, M.L. (2001) Estimating parameters in autoregressive models in non-normal situations; asymmetric innovations, *Commun. Stat.-Theory Meth.* 30, 517-536.

Akkaya, A.D. and Tiku, M.L. (2004) Robust estimation and hypothesis testing under short-tailedness and inliers, *TEST* (to appear).

Barnett, V.D. (1966) Evaluation of the maximum likelihood estimator when the likelihood equation has multiple roots, *Biometrika* 53, 151-165.

Bartlett, M.S. (1953) Approximate confidence intervals, *Biometrika* 40, 12-19.

David, H.A. (1981) *Order Statistics*, 2nd ed. New York, Wiley.

Gross, A.J. and Clark, V.A. (1975) *Survival Distributions*, John Wiley, New York.

Hoeffding, W. (1953) On the distribution of the expected values of the order statistics, *Ann. Math. Stat.* 24, 93-100.

Islam M.Q., Tiku M.L. and Yıldırım, F. (2001) Nonnormal Regression I.Skew Distributions, *Commun. Stat.-Theory Meth.* 30(6), 993-1020.

Kendall, M.G. and Stuart, A. (1979) *The Advanced Theory of Statistics,* Vol.1, Charles Griffin: London.

Lee, K.R., Kapadia C.H. and Dwight B.B. (1980) On estimating the scale parameter of the Rayleigh distribution from doubly censored samples, *Statist. Hefte.* 21, 14-29.

Lieblien, J. (1953) On the exact evaluation of the variances and covariances of order statistics in samples from the extreme distribution, *Ann. Math. Stat.* 24, 282-287.

Puthenpura, S. and Sinha, N.K. (1986) Modified maximum likelihood method for the robust estimation of system parameters from very noisy data. *Automatica* 22, 231-235.

Senoglu and Tiku (2001) Analysis of variance in experimental design with nonnormal error distributions, *Commun. Statist.-Theor. Meth.* 30(7), 1335-1352.

Senoglu and Tiku (2002) Linear contrasts in experimental design with non-identical error distributions, *Biometrical Journal* 44, 3, 359-374.

Tiku, M.L. (1967) Estimating the mean and standard deviation from a censored normal sample, *Biometrika* 54, 155-165.

Tiku, M.L. (1968) Estimating the parameters of normal and logistic distributions from censored samples, *Austral. J. Statist.* 10, 64-74.

Tiku, M.L. (1980) Robustness of MML estimators based on censored samples and robust test statistics, *J. Stat. Plann. Inf.* 4, 123-143.

Tiku, M.L. (1985) Noncentral chi-square and F distribution, *Encyclopedia of Statistical Sciences*, Vol. 6, 276-284, John Wiley&Sons, inc. (Eds., Johnson and Kotz).

Tiku, M.L. and Akkaya, A.D. (2004) *Robust Estimation and Hypothesis Testing* (to be published).

Tiku, M.L. and Kambo, N.S. (1992) Estimation and hypothesis testing for a new family of bivariate nonnormal distributions, *Commun. Stat.-Theory Meth.* 21, 1683-1705.

Tiku, M.L. and Kumra, S. (1981) Expected values and variances and covariances of order statistics for a family of symmetric distributions (Student's t), *Selected Tables in Mathematical Statistics*, Vol. 8. Providence, R.I., *Amer. Math. Soc.* 1985, 141-270.

Tiku, M.L. and Singh, M. (1982) Robust statistics for testing mean vectors of multivariate distributions, *Commun. Statist.-Theor. Meth.* 11(9), 985-1001.

Tiku, M.L. and Suresh, R.P. (1992) A new method of estimation for location and scale parameters, *J. Stat. Plann. Inf.* 30, 281-292.

Tiku, M.L. and Vaughan, D.C. (1999) A new family of short-tailed symmetric distributions, Technical Report, McMaster University, Canada.

Tiku, M.L., Islam, M.Q. and Selcuk, A.S. (2001) Non-normal regression: Part II, symmetric distributions, *Commun. Stat.-Theory Meth.* 30(6), 1021-1045.

Tiku, M.L., Tan, W.Y. and Balakrishnan, N. (1986) *Robust Inference,* Marcel Dekker, New York.

Vaughan, D.C. (1992) On the Tiku-Suresh method of estimation, *Commun. Stat.-Theory Meth.* 21(2), 451-469.

Vaughan, D.C. (2002) The generalized secant hyperbolic distributions and its properties, *Commun. Stat.-Theory Meth. 3*1, 219-238.

Vaughan, D.C. and Tiku, M.L. (2000) Estimation and hypothesis testing for a non-normal bivariate distribution with applications, *J. Mathematical and Computer Modelling* 32, 53-67.

White, J.S. (1969) The moments of the log-Weibull order statistics, *Technometrics* 11, 373-386.

APPENDIX A

Psi Functions

Psi (Digamma) function is defined as

$$\psi(z) = d\left[\ln \Gamma(z)\right]/dz = \Gamma'(z)/\Gamma(z) \tag{A.1}$$

where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.

Values of $\psi(z)$ and its derivative $\psi'(z)$ (Trigamma Function) are given in Abramowitz and Stegun (1965) for $1 < z < 2$. To find the values of $\psi(z)$ and $\psi'(z)$ for $z \geq 2$, the following recurrence relations can be used,

$$\psi(z+1) = \psi(z) + \frac{1}{z} \quad \text{and} \tag{A.2}$$

$$\psi'(z+1) = \psi'(z) - \frac{1}{z^2}. \tag{A.3}$$

To find the value of any polygamma function, the following recurrence relation can be used,

$$\psi^{(n)}(z+1) = \psi^{(n)}(z) + (-1)^n n! z^{-n-1}. \tag{A.4}$$

The values of the psi functions and its derivative for some selected values of b are given in Table A.1 for easy accessibility.

Table A.1 The values of psi functions for selected values of b.

b	$b+1$	$\psi'(b)$	$\psi'(b+1)$	$\psi(b)$	$\psi(b+1)$
0.2	1.2	26.2674	1.2674	-5.2891	-0.2890
0.5	1.5	4.9348	0.9348	-1.9635	0.0365
1	2	1.6449	0.6449	-0.5772	0.4228
2	3	0.6449	0.3949	0.4228	0.9228
3	4	0.3949	0.2838	0.9228	1.2561
4	5	0.2838	0.2213	1.2561	1.5061
6	7	0.1813	0.1535	1.7061	1.8728
8	9	0.1331	0.1175	2.0156	2.1406

APPENDIX B

Bias Correction in the MML Estimators

As a general result, we will give the bias corrections in the estimators $\hat{\mu}_1$, $\hat{\sigma}_1$ and $\hat{\sigma}_{2.1}$ for the situation when the marginal and the conditional distributions are Generalized Logistic with shape parameters b_1 and b_2, respectively. This can be extended to any other distribution. The estimator of σ_1 is

$$\hat{\sigma}_1 = \frac{B_1 + \sqrt{B_1^2 + 4nC_1}}{2n}. \tag{B.1}$$

Since μ_1 is not known and is estimated, we adjust for one degree of freedom. Thus, the bias corrected estimator of σ_1 is

$$\hat{\sigma}_1^* = \frac{\left\{B_1 + \sqrt{B_1^2 + 4nC_1}\right\}}{2\sqrt{n(n-1)}}. \tag{B.2}$$

The estimator of μ_1 is

$$\hat{\mu}_1 = K_1 + D_1\hat{\sigma}_1^* \tag{B.3}$$

where $K_1 = \frac{1}{m_1}\sum_{i=1}^{n}\beta_{1i}x_{(i)}$, $D_1 = \frac{1}{m_1}\sum_{i=1}^{n}\left(\frac{1}{(b_1+1)} - \alpha_{1i}\right)$ and $m_1 = \sum_{i=1}^{n}\beta_{1i}$.

To obtain the bias corrected estimator of μ_1,

$$
\begin{aligned}
E(\hat{\mu}_1) &= E\left(K_1 + D_1\hat{\sigma}_1^*\right) \\
&= \frac{1}{m_1}\sum_{i=1}^{n}\beta_{1i}E\left(x_{(i)}\right) + D_1E\left(\sigma_1^*\right) \\
&= \frac{1}{m_1}\sum_{i=1}^{n}\beta_{1i}(\mu_1 + \sigma_1 t_{1(i)}) + D_1\sigma_1 \\
&= \mu_1 + \frac{\sigma_1}{m_1}\sum_{i=1}^{n}\beta_i t_{1(i)} + D_1\sigma_1
\end{aligned}
$$

since $E(X) = \mu_1 + \sigma_1 t_1$ and $E\left(\sigma_1^*\right) = \sigma_1$. Thus, the bias corrected estimator of μ_1 is

$$\hat{\mu}_1^* = K_1 - \frac{\hat{\sigma}_1^*}{m_1}\left(\sum_{i=1}^{n}\beta_{1i}t_{1(i)}\right). \tag{B.4}$$

The estimator of $\sigma_{2.1}$ is

$$\hat{\sigma}_{2.1} = \frac{-B_2 + \sqrt{B_2^2 + 4nC_2}}{2n}. \tag{B.5}$$

Since $\mu_{2.1}$ and θ_1 are not known and are estimated (note that $w_i = y_i - \theta_1 x_i$ and $\mu_{2.1}$, $\sigma_{2.1}$ and θ_1 are the parameters in the conditional part), the scale parameter in the conditional distribution is adjusted for two degrees of freedom. Thus, the bias corrected estimator of $\sigma_{2.1}$ is

$$\hat{\sigma}_{2.1}^* = \frac{-B_2 + \sqrt{B_2^2 + 4nC_2}}{2\sqrt{n(n-2)}}. \tag{B.6}$$

APPENDIX C

Bias Correction in the Least Square Estimators

LS estimators of location and scale estimate the mean and the variance of the distribution but since we want to estimate the parameters, they need to be adjusted for the bias. As a general case we give the bias corrections for the location and scale estimators of least squares for the case when the marginal and the conditional distribution are both Generalized Logistic with the shape parameters, b_1 and b_2, respectively. It can easily be extended to any location-scale distribution.

For the marginal distribution, the LS estimators, $\bar{x}$ and s_x^2, are estimating the mean $E(X)$ and the variance $V(X)$, respectively. Since $E(X) = \mu_1 + \sigma_1(\psi(b_1) - \psi(1))$ and $V(X) = \sigma_1^2(\psi'(b_1) + \psi'(1))$, the bias corrected LS estimators of μ_1 and σ_1 are, respectively,

$$\tilde{\mu}_1 = \bar{x} - \tilde{\sigma}_1(\psi(b_1) - \psi(1)) \text{ and } \tilde{\sigma}_1 = s_x \Big/ \sqrt{(\psi'(b_1) + \psi'(1))}. \qquad \text{(C.1)}$$

In symmetric distributions, there is no need to adjust for the estimators of the location parameters since the mean of the distribution directly gives the location parameter.

In the same way, for the conditional distribution, the LS estimators $\overline{w} = \bar{y} - \left(s_{xy} \Big/ s_x^2\right)\bar{x}$ and $s_w^2 = \sum_{i=1}^{n} (w_i - \overline{w})^2 \Big/ (n-2)$ are estimating $E(w)$ and

$V(w)$, respectively (Note that $w_i = y_i - \theta_1 x_i$). Since $E(w) = \mu_{2.1} + \sigma_{2.1}(\psi(b_2) - \psi(1))$ and $V(w) = \sigma_{2.1}^2(\psi'(b_2) + \psi'(1))$, the bias corrected LS estimators of $\mu_{2.1}$ and $\sigma_{2.1}$ are, respectively,

$$\tilde{\mu}_{2.1} = \overline{w} - \tilde{\sigma}_{2.1}(\psi(b_2) - \psi(1)) \text{ and } \tilde{\sigma}_{2.1} = s_w \Big/ \sqrt{(\psi'(b_2) + \psi'(1))}\,. \tag{C.2}$$

Also since $y = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1 + \rho\frac{\sigma_2}{\sigma_1}x + e$ (note that $e_i = w_i - \mu_{2.1}$),

$$\begin{aligned} E(Y) &= \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1 + \rho\frac{\sigma_2}{\sigma_1}E(X) + E(e) \\ &= \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1 + \rho\frac{\sigma_2}{\sigma_1}\left(\mu_1 + \sigma_1(\psi(b_1) - \psi(1))\right) + \sigma_2\sqrt{(1-\rho^2)}(\psi(b_2) - \psi(1)) \\ &= \mu_2 + \rho\sigma_2(\psi(b_1) - \psi(1)) + \sigma_2\sqrt{(1-\rho^2)}(\psi(b_2) - \psi(1)) \quad \text{and} \end{aligned}$$

$$\begin{aligned} V(Y) &= \rho^2\frac{\sigma_2^2}{\sigma_1^2}V(X) + V(e) \\ &= \rho^2\frac{\sigma_2^2}{\sigma_1^2}\left(\sigma_1^2(\psi'(b_1) + \psi'(1))\right) + \sigma_2^2(1-\rho^2)(\psi'(b_2) + \psi'(1)) \\ &= \sigma_2^2\left(\rho^2(\psi'(b_1) + \psi'(1)) + (1-\rho^2)(\psi'(b_2) + \psi'(1))\right), \end{aligned}$$

the bias corrected LS estimators of μ_2 and σ_2 are, respectively,

$$\tilde{\mu}_2 = \overline{y} - \tilde{\rho}\tilde{\sigma}_2(\psi(b_1) - \psi(1)) + \tilde{\sigma}_2\sqrt{(1-\tilde{\rho}^2)}(\psi(b_2) - \psi(1)) \tag{C.3}$$

and

$$\tilde{\sigma}_2 = s_y \Big/ \sqrt{\left(\tilde{\rho}^2(\psi'(b_1) + \psi'(1)) + (1-\tilde{\rho}^2)(\psi'(b_2) + \psi'(1))\right)}\,. \tag{C.4}$$

APPENDIX D

The derivation of the Elements of the Fisher Information Matrix

As a general example we give the derivation of one element of the Fisher information matrix for the situation when the marginal and the conditional distribution is Generalized Logistic with shape parameter b_1 and b_2, respectively. We will derive the element $[I_{12}]$ in the Fisher information matrix $I(\mu_1,\sigma_1,\mu_2,\sigma_2,\rho)$. Since

$$[I_{ij}]=\left[-E\left(\frac{\partial^2 \ln L}{\partial\tau_i\partial\tau_j}\right)\right],\ \tau_1=\mu_1,\ \tau_2=\sigma_1,\ \tau_3=\mu_2,\ \tau_4=\sigma_2,\ \tau_5=\rho, \tag{D.1}$$

$$[I_{12}]=-E\left(\frac{\partial^2 \ln L}{\partial\mu_1\partial\sigma_1}\right). \tag{D.2}$$

The loglikelihood equation is

$$\begin{aligned}\ln L = {} & n\ln b_1 - n\ln\sigma_1 - \sum_{i=1}^{n} z_i - (b_1+1)\sum_{i=1}^{n}\ln\left(1+e^{-z_i}\right) \\ & + n\ln b_2 - n\ln\sigma_2 - \frac{n}{2}\ln(1-\rho^2) - \frac{1}{\sigma_2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} e_i \\ & - (b_2+1)\sum_{i=1}^{n}\ln\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right). \quad -\infty<z<\infty,\ -\infty<e<\infty. \end{aligned} \tag{D.3}$$

To evaluate $[I_{12}]$, firstly we take the partial derivative of $\ln L$ w.r.t μ_1 and then w.r.t σ_1:

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n}\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}$$

$$-\frac{n\rho}{\sigma_1\sqrt{(1-\rho^2)}} + \frac{(b_2+1)\rho}{\sigma_1\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}}{\left(1+e^{-\frac{e_i}{\sigma_2\sqrt{(1-\rho^2)}}}\right)}, \text{ which can be written as}$$

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{n}{\sigma_1} - \frac{(b_1+1)}{\sigma_1}\sum_{i=1}^{n}\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}$$

$$-\frac{n\rho}{\sigma_1\sqrt{(1-\rho^2)}} + \frac{(b_2+1)\rho}{\sigma_1\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{e^{-a_i}}{\left(1+e^{-a_i}\right)} \qquad \text{(D.4)}$$

since $a_i = e_i \Big/ \left(\sigma_2\sqrt{(1-\rho^2}\right)$.

$$\frac{\partial^2 \ln L}{\partial \mu_1 \partial \sigma_1} = -\frac{n}{\sigma_1^2} + \frac{(b_1+1)}{\sigma_1^2}\sum_{i=1}^{n}\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)} - \frac{(b_1+1)}{\sigma_1^2}\sum_{i=1}^{n} z_i \frac{e^{-z_i}}{\left(1+e^{-z_i}\right)^2} + \frac{n\rho}{\sigma_1^2\sqrt{(1-\rho^2)}}$$

$$-\frac{(b_2+1)\rho}{\sigma_1^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{e^{-a_i}}{\left(1+e^{-a_i}\right)} - \frac{(b_2+1)\rho^2}{\sigma_1^2(1-\rho^2)}\sum_{i=1}^{n} z_i \frac{e^{-a_i}}{\left(1+e^{-a_i}\right)^2}. \qquad \text{(D.5)}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial \mu_1 \partial \sigma_1}\right) = \frac{n}{\sigma_1^2} - \frac{(b_1+1)}{\sigma_1^2}\sum_{i=1}^{n}E\left(\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}\right) + \frac{(b_1+1)}{\sigma_1^2}\sum_{i=1}^{n}E\left(z_i\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)^2}\right)$$

$$-\frac{n\rho}{\sigma_1^2\sqrt{(1-\rho^2)}} + \frac{(b_2+1)\rho}{\sigma_1^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}E\left(\frac{e^{-a_i}}{\left(1+e^{-a_i}\right)}\right)$$

$$+\frac{(b_2+1)\rho^2}{\sigma_1^2(1-\rho^2)}\sum_{i=1}^{n}E\left(z_i\frac{e^{-a_i}}{\left(1+e^{-a_i}\right)^2}\right)$$

$$= \frac{n}{\sigma_1^2} - \frac{(b_1+1)}{\sigma_1^2}\sum_{i=1}^{n} E\left(\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}\right) + \frac{(b_1+1)}{\sigma_1^2}\sum_{i=1}^{n} E\left(z_i\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)^2}\right)$$

$$- \frac{n\rho}{\sigma_1^2\sqrt{(1-\rho^2)}} + \frac{(b_2+1)\rho}{\sigma_1^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n} E\left(\frac{e^{-a_i}}{\left(1+e^{-a_i}\right)}\right)$$

$$+ \frac{(b_2+1)\rho^2}{\sigma_1^2(1-\rho^2)}\sum_{i=1}^{n} E(z_i)\,E\left(\frac{e^{-a_i}}{\left(1+e^{-a_i}\right)^2}\right), \qquad \text{(D.6)}$$

since the marginal and the conditional parts are independent.

$$-E\left(\frac{\partial^2 \ln L}{\partial\mu_1\partial\sigma_1}\right) = \frac{n}{\sigma_1^2} - \frac{(b_1+1)}{\sigma_1^2}\sum_{i=1}^{n}\frac{1}{(b_1+1)}$$

$$+ \frac{(b_1+1)}{\sigma_1^2}\sum_{i=1}^{n}\frac{b_1}{(b_1+1)(b_2+1)}\left(\psi(b_1+1)-\psi(2)\right)$$

$$- \frac{n\rho}{\sigma_1^2\sqrt{(1-\rho^2)}} + \frac{(b_2+1)\rho}{\sigma_1^2\sqrt{(1-\rho^2)}}\sum_{i=1}^{n}\frac{1}{(b_2+1)}$$

$$+ \frac{(b_2+1)\rho^2}{\sigma_1^2(1-\rho^2)}\sum_{i=1}^{n}\left(\psi(b_1)-\psi(1)\right)\frac{b_2}{(b_2+1)(b_2+2)}. \qquad \text{(D.7)}$$

Finally,

$$-E\left(\frac{\partial^2 \ln L}{\partial\mu_1\partial\sigma_1}\right) = n\left[\frac{1}{\sigma_1^2}\left\{\frac{b_1}{(b_1+2)}\left(\psi(b_1+1)-\psi(2)\right)\right.\right.$$

$$\left.\left. + \frac{b_2}{(b_2+2)}\frac{\rho^2}{(1-\rho^2)}\left(\psi(b_1)-\psi(1)\right)\right\}\right]. \qquad \text{(D.8)}$$

We use the result of some special integrals and functions to find the elements of the Fisher information matrix. We now give them in detail. If X has a Generalized Logistic distribution with location parameter μ_1, scale parameter σ_1 and shape parameter b_1 and we let $z = (x-\mu_1)/\sigma_1$,

$$E(z_i) = \psi(b_1) - \psi(1) \text{ and } E(z_i^2) = \psi'(b_1) + \psi'(1) + \left(\psi(b_1) - \psi(1)\right)^2 \quad \text{(D.9)}$$

and

$$E\left(\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}\right) = \frac{1}{(b_1+1)},\ E\left(\frac{e^{-z_i}}{\left(1+e^{-z_i}\right)^2}\right) = \frac{b_1}{(b_1+1)(b_2+1)},$$

$$E\left(z_i \frac{e^{-z_i}}{\left(1+e^{-z_i}\right)}\right) = \frac{1}{(b_1+1)}\left(\psi(b_1) - \psi(2)\right),$$

$$E\left(z_i \frac{e^{-z_i}}{\left(1+e^{-z_i}\right)^2}\right) = \frac{b_1}{(b_1+1)(b_1+2)}\left(\psi(b_1+1) - \psi(2)\right) \text{ and}$$

$$E\left(z_i^2 \frac{e^{-z_i}}{\left(1+e^{-z_i}\right)^2}\right) = \frac{b_1}{(b_1+1)(b_1+2)}\left\{\psi'(b_1+1) - \psi'(2) + \left(\psi(b_1+1) - \psi(2)\right)^2\right\}. \quad \text{(D.10)}$$

Similar rules apply to a_i but with replacing b_1 with b_2. Also the following recurrence formula is used in finding some elements of the Fisher information matrix:

$$\psi(2) = \psi(1) + 1 \quad \text{(D.11)}$$

which is a particular case of

$$\psi(z+1) = \psi(z) + \frac{1}{z}\ . \quad \text{(D.12)}$$

APPENDIX E

Visual Fortran Program for the Case of Marginal and Conditional Generalized Logistic with Shape Parameters b_1 and b_2

```
c  *** Written by Hakan Savaş Sazak, 2003, Ankara ***

        use numerical_libraries
        real y(100),x(100),xc(100),yc(100),e(100),teta0,teta1
        real xo(100),w(100),wo(100),wols(100),wls(100)
        real z(100),a(100),u1(100),u2(100)
        real uext(100),uextcont(100)
        real mean1(10000),mean2(10000),sigma1(10000),sigma2(10000)
        real rho(10000),bet1(10000)
        real m1,m2,k1,be1,c1
        real mean1ls(10000),mean2ls(10000),sigma1ls(10000)
        real sigma2ls(10000),rhols(10000),bet1ls(10000)
        real sigma2g1ls(10000)
        real mean2g1(10000),sigma2g1(10000),mean2g1ls(10000)
        real t1(100),t2(100),bett1(100),alpha1(100),d2(100)
        real bett2(100),alpha2(100)
        real mean10,mean20,mean2g10
        real asvarmean1(10000),asvarsigma1(10000)
        real asvarmean12(10000),asvarsigma12(10000)
        real asvarmean12ls(10000),asvarsigma12ls(10000)
        real asvarmean2(10000),asvarsigma2(10000)
        real asvarmean2g1(10000),asvarsigma2g1(10000)
        real asvarrho(10000),asvarrho0(10000),ascovmean1mean2(10000)
        real asvarmean1ls(10000),asvarsigma1ls(10000)
        real asvarbet1(10000),ascovmean1mean2g1(10000)
        real psid_b,psid_bplus1,psid_1,psid_2
        real dlnLdmean1overn(10000),dlnLdsigma1overn(10000)
        real dlnLdmean2g1overn(10000),dlnLdsigma2g1overn(10000)
        real dlnLdmean2overn(10000),dlnLdsigma2overn(10000)
        real dlnLdrhoovern(10000),dlnLdbet1overn(10000)


c *** Specifying the Fisher information matrices and their inverses ***

        parameter (LI=5,LIINV=5,KI=5)
        real MI(LI,LI),MIINV(LIINV,LIINV)

        parameter (LI2=5,LI2INV=5,KI2=5)
        real MI2(LI2,LI2),MI2INV(LI2INV,LI2INV)


c *** Specifying the location of the output file ***

        open(unit=1,file='c:\savas\programlar\congenmarggen\out.txt')

        print *, 'enter sample size'
        read *, n


c *** Specifying the number of turns in the simulation ***

        nn=10000
```

```
c *** Specifying the parameter values ***

       b1=0.5
       b2=1.0
       prho=0.5

       pmean1=0.0
       pmean2=0.0
       psigma1=1.0
       psigma2=1.0

       pbet1=prho*(psigma2/psigma1)
       pbet0=pmean2-prho*(psigma2/psigma1)*pmean1
       pmean2g1=pmean2-pbet1*pmean1
       psigma2g1=psigma2*sqrt(1-prho**2)


c *** Means of the null hypothesis ***

       mean10=0.0
       mean20=0.0
       mean2g10=0.0


c *** Specifying the values of the Trigamma functions ***

       psid_1=1.6449
       psid_2=0.6449

       if(b1.eq.0.1) then
       psid_b1=101.4316
       psid_b1plus1=1.4333
       else if(b1.eq.0.2) then
       psid_b1=26.2674
       psid_b1plus1=1.2672
       else if(b1.eq.0.5) then
       psid_b1=4.9348
       psid_b1plus1=0.9348
       else if(b1.eq.1.0) then
       psid_b1=1.6449
       psid_b1plus1=0.6449
       else if(b1.eq.2.0) then
       psid_b1=0.6449
       psid_b1plus1=0.3949
       else if(b1.eq.3.0) then
       psid_b1=0.3949
       psid_b1plus1=0.2838
       else if(b1.eq.4.0) then
       psid_b1=0.2838
       psid_b1plus1=0.2213
       else if(b1.eq.6.0) then
       psid_b1=0.1813
       psid_b1plus1=0.1535
       else if(b1.eq.8.0) then
       psid_b1=0.1331
       psid_b1plus1=0.1175
       endif

       if(b2.eq.0.1) then
       psid_b2=101.4316
       psid_b2plus1=1.4333
       else if(b2.eq.0.2) then
       psid_b2=26.2674
       psid_b2plus1=1.2672
       else if(b2.eq.0.5) then
       psid_b2=4.9348
       psid_b2plus1=0.9348
       else if(b2.eq.1.0) then
       psid_b2=1.6449
       psid_b2plus1=0.6449
```

```
        else if(b2.eq.2.0) then
        psid_b2=0.6449
        psid_b2plus1=0.3949
        else if(b2.eq.3.0) then
        psid_b2=0.3949
        psid_b2plus1=0.2838
        else if(b2.eq.4.0) then
        psid_b2=0.2838
        psid_b2plus1=0.2213
        else if(b2.eq.6.0) then
        psid_b2=0.1813
        psid_b2plus1=0.1535
        else if(b2.eq.8.0) then
        psid_b2=0.1331
        psid_b2plus1=0.1175
        endif


c *** The expected values and variances of zi and ai ***

        expzi=psi(b1)-psi(1.0)
        varzi=psid_b1+psid_1

        expai=psi(b2)-psi(1.0)
        varai=psid_b2+psid_1


c *** The tabulated values ***

        finv=fin(0.95,2.0,1.0*(n-2))
        xki2inv=chiin(0.95,2.0)
        xki1inv=chiin(0.95,1.0)
        xkininv=chiin(0.95,1.0*n)
        xz095=anorin(0.95)


c *** Simulating nn=10,000 random samples ***

        do 300 k=1,nn


c *** Producing random marginal and conditional parts ***

        call rnun(n,u1)
        call rnun(n,u2)
        do 10 i=1,n
        a(i)=-alog(u2(i)**(-1.0/(1.0*b2))-1.0)
        z(i)=-alog(u1(i)**(-1.0/(1.0*b1))-1.0)
10      continue


c *** Creating 10% outlier for x ***

c       r=int(0.1*n+0.5)
c       do 15 i=1,r
c       z(i)=4.0*z(i)
c15     continue


c *** Creating mixture model for x ***

c       call rnun(n,uext)
c       do 15 i=1,n
c       if(uext(i).GT.0.9) then
c       z(i)=4.0*z(i)
c       endif
c15     continue
```

```
c *** Creating contamination model for x ***

c       call rnun(n,uext)
c       call rnun(n,uextcont)
c       do 15 i=1,n
c       if(uext(i).GT.0.9) then
c       z(i)=uextcont(i)-0.5
c       endif
c15     continue


c *** Generating random x and y pairs ***

        do 20 i=1,n
        e(i)=psigma2g1*a(i)
        x(i)=pmean1+psigma1*z(i)
        y(i)=pmean2g1+pbet1*x(i)+e(i)
20      continue


c *** Ordering x's ***

        do 30 i=1,n
        xo(i)=x(i)
30      continue

        do 50 i=1,n
        do 40 j=i+1,n
        if(xo(i).GT.xo(j)) then
        dummy=xo(i)
        xo(i)=xo(j)
        xo(j)=dummy
        endif
40      continue
50      continue


c *** Calculating the MML and the LS estimators ***

        m1=0.0
        m2=0.0
        k1=0.0
        b1t=0.0
        xmeanx=0.0
        xmeany=0.0

        do 60 i=1,n
        xi=(1.0*i)/(1.0*n+1)
        t1(i)=-alog(xi**(-1.0/(1.0*b1))-1.0)
        bett1(i)=exp(-t1(i))/((1.0+exp(-t1(i)))**2.0)
        alpha1(i)=((1+exp(-t1(i))+t1(i))*exp(-t1(i)))/
        &((1.0+exp(-t1(i)))**2.0)
        m1=m1+bett1(i)

        t2(i)=-alog(xi**(-1.0/(1.0*b2))-1.0)
        bett2(i)=exp(-t2(i))/((1.0+exp(-t2(i)))**2.0)
        alpha2(i)=((1+exp(-t2(i))+t2(i))*exp(-t2(i)))/
        &((1.0+exp(-t2(i)))**2.0)
        m2=m2+bett2(i)

        k1=k1+bett1(i)*xo(i)
        d2(i)=alpha2(i)-1.0/(b2+1.0)
        b1t=b1t+bett1(i)*t1(i)
        xmeanx=xmeanx+x(i)
        xmeany=xmeany+y(i)
60      continue

        xmeanx=xmeanx/(1.0*n)
        xmeany=xmeany/(1.0*n)
        k1=k1/m1
```

```
        sx2=0.0
        sy2=0.0
        sxy=0.0

        be1=0.0
        c1=0.0
        sx2=0.0
        sy2=0.0
        sxy=0.0

        do 70 i=1,n
        be1=be1+(b1+1.0)*(1/(b1+1.0)-alpha1(i))*(xo(i)-k1)
        c1=c1+(b1+1.0)*bett1(i)*((xo(i)-k1)**2)
        sx2=sx2+(x(i)-xmeanx)**2
        sy2=sy2+(y(i)-xmeany)**2
        sxy=sxy+(x(i)-xmeanx)*(y(i)-xmeany)
70      continue

        sx2=sx2/(1.0*n-1)
        sy2=sy2/(1.0*n-1)
        sxy=sxy/(1.0*n-1)

        bet1ls(k)=sxy/sx2
        bet1(k)=sxy/sx2
        rhols(k)=sxy/sqrt(sx2*sy2)
        sigma2ls(k)=sqrt(sy2/((rhols(k)**2)*(psid_b1+psid_1)+
        &(1-rhols(k)**2)*(psid_b2+psid_1)))
        mean2ls(k)=xmeany-sigma2ls(k)*
        &(psi(b2)-psi(1.0))*sqrt(1-rhols(k)**2)
        &-sigma2ls(k)*(psi(b1)-psi(1.0))*rhols(k)

        sigma1(k)=(be1+sqrt(be1*be1+4.0*n*c1))/(2.0*sqrt(1.0*n*(n-1)))
        mean1(k)=k1-b1t*sigma1(k)/m1

        sigma1ls(k)=sqrt(sx2/(psid_b1+psid_1))
        mean1ls(k)=xmeanx-sigma1ls(k)*(psi(b1)-psi(1.0))


c *** The iteration # 'ite' ***

        do 150 ite=1,2
        do 80 i=1,n
        w(i)=y(i)-bet1(k)*x(i)
80      continue


c *** Ordering w's and finding the concomitants x and y ***

        do 82 i=1,n
        wo(i)=w(i)
        xc(i)=x(i)
        yc(i)=y(i)
82      continue

        do 87 i=1,n
        do 84 j=i+1,n
        if(wo(i).GT.wo(j)) then
        dummy=wo(i)
        wo(i)=wo(j)
        wo(j)=dummy

        dummy=xc(i)
        xc(i)=xc(j)
        xc(j)=dummy

        dummy=yc(i)
        yc(i)=yc(j)
        yc(j)=dummy

        endif
```

```
84      continue
87      continue

        b2y=0.0
        b2x=0.0
        delt=0.0

        do 115 i=1,n
        b2y=b2y+bett2(i)*yc(i)
        b2x=b2x+bett2(i)*xc(i)
        delt=delt+d2(i)
115     continue

        ycmean=b2y/m2
        xcmean=b2x/m2

        rk2num=0.0
        den=0.0
        rd2num=0.0
        bety2=0.0

        do 120 i=1,n
        rk2num=rk2num+bett2(i)*(xc(i)-xcmean)*yc(i)
        den=den+bett2(i)*((xc(i)-xcmean)**2)
        rd2num=rd2num+d2(i)*(xc(i)-xcmean)
        bety2=bety2+bett2(i)*((yc(i)-ycmean)**2)
120     continue

        rk2=rk2num/den
        rd2=rd2num/den

        be2=0.0

        do 130 i=1,n
        be2=be2+d2(i)*(yc(i)-ycmean-rk2*(xc(i)-xcmean))
130     continue

        be2=(b2+1.0)*be2
        c2=(b2+1.0)*(bety2-rk2*rk2num)

        sigma2g1(k)=(-be2+sqrt(be2**2+4.0*n*c2))/(2.0*sqrt(1.0*n*(n-2)))
        bet1(k)=rk2-rd2*sigma2g1(k)

150     continue

        xmeanwls=0.0

        do 170 i=1,n
        wls(i)=y(i)-bet1ls(k)*x(i)
        xmeanwls=xmeanwls+wls(i)
170     continue

        xmeanwls=xmeanwls/(1.0*n)

        sw2ls=0.0

        do 180 i=1,n
        sw2ls=sw2ls+(wls(i)-xmeanwls)**2
180     continue

        sw2ls=sw2ls/(1.0*(n-2.0))

        sigma2g1ls(k)=sqrt(sw2ls/varai)
        mean2g1ls(k)=xmeanwls-sigma2g1ls(k)*expai
        mean2g1(k)=ycmean-bet1(k)*xcmean-delt*sigma2g1(k)/m2

        mean2(k)=ycmean-bet1(k)*(xcmean-mean1(k))-delt*sigma2g1(k)/m2
        sigma2(k)=sqrt(sigma2g1(k)**2+(bet1(k)*sigma1(k))**2)
        rho(k)=bet1(k)*sigma1(k)/sigma2(k)
```

```
c *** Calculating the derivatives of the ln likelihood functions at MMLE ***

       totz=0.0
       totln_1pluse_z=0.0
       tota=0.0
       totln_1pluse_a=0.0

       totgz=0.0
       totzxgz=0.0
       totga=0.0
       totzxga=0.0
       totaxga=0.0

       do 200 i=1,n
       zhead(i)=(xc(i)-mean1(k))/sigma1(k)
       ahead(i)=(yc(i)-(rho(k)*sigma2(k)/sigma1(k))*xc(i)
       -mean2(k)+(rho(k)*sigma2(k)/sigma1(k))*mean1(k))/
       (sigma2(k)*sqrt(1.0-rho(k)**2))       totz=totz+zhead(i)
       totln_1pluse_z=totln_1pluse_z+alog(1.0+exp(-zhead(i)))
       tota=tota+ahead(i)
       totln_1pluse_a=totln_1pluse_a+alog(1.0+exp(-ahead(i)))

       totgz=totgz+exp(-zhead(i))/(1.0+exp(-zhead(i)))
       totzxgz=totzxgz+zhead(i)*exp(-zhead(i))/(1.0+exp(-zhead(i)))
       totga=totga+exp(-ahead(i))/(1.0+exp(-ahead(i)))
       totzxga=totzxga+zhead(i)*exp(-ahead(i))/(1.0+exp(-ahead(i)))
       totaxga=totaxga+ahead(i)*exp(-ahead(i))/(1.0+exp(-ahead(i)))
200    continue

       dlnLdmean1overn(k)=(1.0/(1.0*n))*(1.0*n/sigma1(k)-
       (b1+1.0)*totgz/sigma1(k)-1.0*n*rho(k)/
       (sigma1(k)*sqrt(1.0-rho(k)**2))+
       (b2+1.0)*rho(k)*totga/(sigma1(k)*sqrt(1.0-rho(k)**2)))

       dlnLdsigma1overn(k)=(1.0/(1.0*n))*(-1.0*n/sigma1(k)+
       &totz/sigma1(k)-(b1+1.0)*totzxgz/sigma1(k)-
       &rho(k)*totz/(sigma1(k)*sqrt(1.0-rho(k)**2))+
       &(b2+1.0)*rho(k)*totzxga/(sigma1(k)*sqrt(1.0-rho(k)**2)))
       dlnLdmean2g1overn(k)=(1.0/(1.0*n))*(1.0*n/sigma2g1(k)-
       &(b2+1.0)*totga/sigma2g1(k))

       dlnLdsigma2g1overn(k)=(1.0/(1.0*n))*(-1.0*n/sigma2g1(k)+
       &tota/sigma2g1(k)-(b2+1.0)*totaxga/sigma2g1(k))

       dlnLdbet1overn(k)=(1.0/(1.0*n))*(sigma1(k)*totz/sigma2g1(k)-
       &(b2+1.0)*sigma1(k)*totzxga/sigma2g1(k))

       dlnLdmean2overn(k)=(1.0/(1.0*n))*(1.0*n/(sigma2(k)*
       &sqrt(1.0-rho(k)**2))-(b2+1.0)*totga/
       &(sigma2(k)*sqrt(1.0-rho(k)**2)))

       dlnLdsigma2overn(k)=(1.0/(1.0*n))*(-1.0*n/sigma2(k)+
       &rho(k)*totz/(sigma2(k)*sqrt(1.0-rho(k)**2))+
       &tota/sigma2(k)-(b2+1.0)*rho(k)*totzxga/
       &(sigma2(k)*sqrt(1.0-rho(k)**2))-(b2+1.0)*totaxga/sigma2(k))

       &dlnLdrhoovern(k)=(1.0/(1.0*n))*(1.0*n*rho(k)/(1.0-rho(k)**2)+
       &totz/sqrt(1.0-rho(k)**2)-rho(k)*tota/(1.0-rho(k)**2)-
       &(b2+1.0)*totzxga/sqrt(1.0-rho(k)**2)+
       &(b2+1.0)*rho(k)*totaxga/(1.0-rho(k)**2))

300    continue
```

```
c *** Simulating the means, variances, biassquares and mean square errors of the
c     MML and LS estimators and the efficiency of the LS estimators as compared to
c     the MML estimators in means of variances and mean square erros ***

c *** Also simulating the means and variances of derivatives of the lnlikelihood
c     functions at MMLE ***

        emean1=0.0
        emean2=0.0
        esigma1=0.0
        esigma2=0.0
        erho=0.0
        ebet1=0.0
        emean2g1=0.0
        esigma2g1=0.0

        emean1ls=0.0
        emean2ls=0.0
        esigma1ls=0.0
        esigma2ls=0.0
        erhols=0.0
        ebet1ls=0.0
        emean2g1ls=0.0
        esigma2g1ls=0.0

        edlnLdmean1overn=0.0
        edlnLdsigma1overn=0.0
        edlnLdmean2g1overn=0.0
        edlnLdsigma2g1overn=0.0
        edlnLdmean2overn=0.0
        edlnLdsigma2overn=0.0
        edlnLdrhoovern=0.0
        edlnLdbet1overn=0.0

        do 350 i=1,nn
        emean1=emean1+mean1(i)
        emean2=emean2+mean2(i)
        esigma1=esigma1+sigma1(i)
        esigma2=esigma2+sigma2(i)
        erho=erho+rho(i)
        ebet1=ebet1+bet1(i)
        emean2g1=emean2g1+mean2g1(i)
        esigma2g1=esigma2g1+sigma2g1(i)

        edlnLdmean1overn=edlnLdmean1overn+dlnLdmean1overn(i)
        edlnLdsigma1overn=edlnLdsigma1overn+dlnLdsigma1overn(i)
        edlnLdmean2g1overn=edlnLdmean2g1overn+dlnLdmean2g1overn(i)
        edlnLdsigma2g1overn=edlnLdsigma2g1overn+dlnLdsigma2g1overn(i)
        edlnLdmean2overn=edlnLdmean2overn+dlnLdmean2overn(i)
        edlnLdsigma2overn=edlnLdsigma2overn+dlnLdsigma2overn(i)
        edlnLdrhoovern=edlnLdrhoovern+dlnLdrhoovern(i)
        edlnLdbet1overn=edlnLdbet1overn+dlnLdbet1overn(i)
350     continue

        emean1=emean1/(1.0*nn)
        emean2=emean2/(1.0*nn)
        esigma1=esigma1/(1.0*nn)
        esigma2=esigma2/(1.0*nn)
        erho=erho/(1.0*nn)
        ebet1=ebet1/(1.0*nn)
        emean2g1=emean2g1/(1.0*nn)
        esigma2g1=esigma2g1/(1*nn)

        emean1ls=emean1ls/(1.0*nn)
        emean2ls=emean2ls/(1.0*nn)
        esigma1ls=esigma1ls/(1.0*nn)
        esigma2ls=esigma2ls/(1.0*nn)
        erhols=erhols/(1.0*nn)
        ebet1ls=ebet1ls/(1.0*nn)
        emean2g1ls=emean2g1ls/(1.0*nn)
```

```
      esigma2g1ls=esigma2g1ls/(1*nn)

      edlnLdmean1overn=edlnLdmean1overn/(1.0*nn)
      edlnLdsigma1overn=edlnLdsigma1overn/(1.0*nn)
      edlnLdmean2g1overn=edlnLdmean2g1overn/(1.0*nn)
      edlnLdsigma2g1overn=edlnLdsigma2g1overn/(1.0*nn)
      edlnLdmean2overn=edlnLdmean2overn/(1.0*nn)
      edlnLdsigma2overn=edlnLdsigma2overn/(1.0*nn)
      edlnLdrhoovern=edlnLdrhoovern/(1.0*nn)
      edlnLdbet1overn=edlnLdbet1overn/(1.0*nn)

      vmean1=0.0
      vmean2=0.0
      vsigma1=0.0
      vsigma2=0.0
      vrho=0.0
      vbet1=0.0
      vmean2g1=0.0
      vsigma2g1=0.0
      covmean1mean2g1=0.0
      covmean1mean2=0.0

      vdlnLdmean1overn=0.0
      vdlnLdsigma1overn=0.0
      vdlnLdmean2g1overn=0.0
      vdlnLdsigma2g1overn=0.0
      vdlnLdmean2overn=0.0
      vdlnLdsigma2overn=0.0
      vdlnLdrhoovern=0.0
      vdlnLdbet1overn=0.0

      do 380 i=1,nn
      vmean1=vmean1+(mean1(i)-emean1)**2
      vmean2=vmean2+(mean2(i)-emean2)**2
      vsigma1=vsigma1+(sigma1(i)-esigma1)**2
      vsigma2=vsigma2+(sigma2(i)-esigma2)**2
      vrho=vrho+(rho(i)-erho)**2
      vbet1=vbet1+(bet1(i)-ebet1)**2
      vmean2g1=vmean2g1+(mean2g1(i)-emean2g1)**2
      vsigma2g1=vsigma2g1+(sigma2g1(i)-esigma2g1)**2
      covmean1mean2g1=covmean1mean2g1+
      &(mean1(i)-emean1)*(mean2g1(i)-emean2g1)
      covmean1mean2=covmean1mean2+(mean1(i)-emean1)*(mean2(i)-emean2)

      vdlnLdmean1overn=vdlnLdmean1overn+
      &(dlnLdmean1overn(i)-edlnLdmean1overn)**2
      vdlnLdsigma1overn=vdlnLdsigma1overn+
      &(dlnLdsigma1overn(i)-edlnLdsigma1overn)**2
      vdlnLdmean2g1overn=vdlnLdmean2g1overn+
      &(dlnLdmean2g1overn(i)-edlnLdmean2g1overn)**2
      vdlnLdsigma2g1overn=vdlnLdsigma2g1overn+
      &(dlnLdsigma2g1overn(i)-edlnLdsigma2g1overn)**2
      vdlnLdmean2overn=vdlnLdmean2overn+
      &(dlnLdmean2overn(i)-edlnLdmean2overn)**2
      vdlnLdsigma2overn=vdlnLdsigma2overn+
      &(dlnLdsigma2overn(i)-edlnLdsigma2overn)**2
      vdlnLdrhoovern=vdlnLdrhoovern+
      &(dlnLdrhoovern(i)-edlnLdrhoovern)**2
      vdlnLdbet1overn=vdlnLdbet1overn+
      &(dlnLdbet1overn(i)-edlnLdbet1overn)**2
380   continue

      vmean1=vmean1/(1.0*nn)
      vmean2=vmean2/(1.0*nn)
      vsigma1=vsigma1/(1.0*nn)
      vsigma2=vsigma2/(1.0*nn)
      vrho=vrho/(1.0*nn)
      vbet1=vbet1/(1.0*nn)
      vmean2g1=vmean2g1/(1.0*nn)
      vsigma2g1=vsigma2g1/(1.0*nn)
```

```
covmean1mean2g1=covmean1mean2g1/(1.0*nn)
covmean1mean2=covmean1mean2/(1.0*nn)

vdlnLdmean1overn=vdlnLdmean1overn/(1.0*nn)
vdlnLdsigma1overn=vdlnLdsigma1overn/(1.0*nn)
vdlnLdmean2g1overn=vdlnLdmean2g1overn/(1.0*nn)
vdlnLdsigma2g1overn=vdlnLdsigma2g1overn/(1.0*nn)
vdlnLdmean2overn=vdlnLdmean2overn/(1.0*nn)
vdlnLdsigma2overn=vdlnLdsigma2overn/(1.0*nn)
vdlnLdrhoovern=vdlnLdrhoovern/(1.0*nn)
vdlnLdbet1overn=vdlnLdbet1overn/(1.0*nn)

vmean1n=vmean1*1.0*n
vmean2n=vmean2*1.0*n
vsigma1n=vsigma1*1.0*n
vsigma2n=vsigma2*1.0*n
vrhon=vrho*1.0*n
vbet1n=vbet1*1.0*n
vmean2g1n=vmean2g1*1.0*n
vsigma2g1n=vsigma2g1*1.0*n
covmean1mean2g1n=covmean1mean2g1*1.0*n
covmean1mean2n=covmean1mean2*1.0*n

vmean1lsn=vmean1ls*1.0*n
vmean2lsn=vmean2ls*1.0*n
vsigma1lsn=vsigma1ls*1.0*n
vsigma2lsn=vsigma2ls*1.0*n
vrholsn=vrhols*1.0*n
vbet1lsn=vbet1ls*1.0*n
vmean2g1lsn=vmean2g1ls*1.0*n
vsigma2g1lsn=vsigma2g1ls*1.0*n
covmean1mean2g1lsn=covmean1mean2g1ls*1.0*n
covmean1mean2lsn=covmean1mean2ls*1.0*n

bias2mean1=(emean1-pmean1)**2
bias2mean2=(emean2-pmean2)**2
bias2sigma1=(esigma1-psigma1)**2
bias2sigma2=(esigma2-psigma2)**2
bias2rho=(erho-prho)**2
bias2bet1=(ebet1-pbet1)**2
bias2mean2g1=(emean2g1-pmean2g1)**2
bias2sigma2g1=(esigma2g1-psigma2g1)**2

bias2mean1ls=(emean1ls-pmean1)**2
bias2mean2ls=(emean2ls-pmean2)**2
bias2sigma1ls=(esigma1ls-psigma1)**2
bias2sigma2ls=(esigma2ls-psigma2)**2
bias2rhols=(erhols-prho)**2
bias2bet1ls=(ebet1ls-pbet1)**2
bias2mean2g1ls=(emean2g1ls-pmean2g1)**2
bias2sigma2g1ls=(esigma2g1ls-psigma2g1)**2

bias2mean1n=bias2mean1*1.0*n
bias2mean2n=bias2mean2*1.0*n
bias2sigma1n=bias2sigma1*1.0*n
bias2sigma2n=bias2sigma2*1.0*n
bias2rhon=bias2rho*1.0*n
bias2bet1n=bias2bet1*1.0*n
bias2mean2g1n=bias2mean2g1*1.0*n
bias2sigma2g1n=bias2sigma2g1*1.0*n

bias2mean1lsn=bias2mean1ls*1.0*n
bias2mean2lsn=bias2mean2ls*1.0*n
bias2sigma1lsn=bias2sigma1ls*1.0*n
bias2sigma2lsn=bias2sigma2ls*1.0*n
bias2rholsn=bias2rhols*1.0*n
bias2bet1lsn=bias2bet1ls*1.0*n
bias2mean2g1lsn=bias2mean2g1ls*1.0*n
bias2sigma2g1lsn=bias2sigma2g1ls*1.0*n
```

```
        xmsemean1=vmean1+bias2mean1
        xmsemean2=vmean2+bias2mean2
        xmsesigma1=vsigma1+bias2sigma1
        xmsesigma2=vsigma2+bias2sigma2
        xmserho=vrho+bias2rho
        xmsebet1=vbet1+bias2bet1
        xmsemean2g1=vmean2g1+bias2mean2g1
        xmsesigma2g1=vsigma2g1+bias2sigma2g1

        xmsemean1ls=vmean1ls+bias2mean1ls
        xmsemean2ls=vmean2ls+bias2mean2ls
        xmsesigma1ls=vsigma1ls+bias2sigma1ls
        xmsesigma2ls=vsigma2ls+bias2sigma2ls
        xmserhols=vrhols+bias2rhols
        xmsebet1ls=vbet1ls+bias2bet1ls
        xmsemean2g1ls=vmean2g1ls+bias2mean2g1ls
        xmsesigma2g1ls=vsigma2g1ls+bias2sigma2g1ls

        xmsemean1n=vmean1n+bias2mean1n
        xmsemean2n=vmean2n+bias2mean2n
        xmsesigma1n=vsigma1n+bias2sigma1n
        xmsesigma2n=vsigma2n+bias2sigma2n
        xmserhon=vrhon+bias2rhon
        xmsebet1n=vbet1n+bias2bet1n
        xmsemean2g1n=vmean2g1n+bias2mean2g1n
        xmsesigma2g1n=vsigma2g1n+bias2sigma2g1n

        xmsemean1lsn=vmean1lsn+bias2mean1lsn
        xmsemean2lsn=vmean2lsn+bias2mean2lsn
        xmsesigma1lsn=vsigma1lsn+bias2sigma1lsn
        xmsesigma2lsn=vsigma2lsn+bias2sigma2lsn
        xmserholsn=vrholsn+bias2rholsn
        xmsebet1lsn=vbet1lsn+bias2bet1lsn
        xmsemean2g1lsn=vmean2g1lsn+bias2mean2g1lsn
        xmsesigma2g1lsn=vsigma2g1lsn+bias2sigma2g1lsn

        effvmean1=100.0*vmean1n/vmean1lsn
        effvsigma1=100.0*vsigma1n/vsigma1lsn
        effvmean2=100.0*vmean2n/vmean2lsn
        effvsigma2=100.0*vsigma2n/vsigma2lsn
        effvrho=100.0*vrhon/vrholsn
        effvmean2g1=100.0*vmean2g1n/vmean2g1lsn
        effvsigma2g1=100.0*vsigma2g1n/vsigma2g1lsn
        effvbet1=100.0*vbet1n/vbet1lsn

        effmsemean1=100.0*xmsemean1n/xmsemean1lsn
        effmsesigma1=100.0*xmsesigma1n/xmsesigma1lsn
        effmsemean2=100.0*xmsemean2n/xmsemean2lsn
        effmsesigma2=100.0*xmsesigma2n/xmsesigma2lsn
        effmserho=100.0*xmserhon/xmserholsn
        effmsemean2g1=100.0*xmsemean2g1n/xmsemean2g1lsn
        effmsesigma2g1=100.0*xmsesigma2g1n/xmsesigma2g1lsn
        effmsebet1=100.0*xmsebet1n/xmsebet1lsn


c *** Output of the simulated means, variances, biassquares and mean square errors
c     of the MMLE and LSE, and the efficiency of the LSE as compared to the MMLE
c     in means of variances and mean square errors ***

450   format(a2,i3,9x,a5,1x,a6,3x,a5,1x,a6,2x,a7,1x,a8,2x,a6,4x,a3)
       write(1,450) 'n=',n,'mean1','sigma1','mean2','sigma2'
       &,'mean2.1','sigma2.1','theta1','rho'

490   format(a5,7x,f7.3,f7.3,1x,f7.3,f7.3,f8.3,f8.3,f10.3,f8.3)
       write(1,490) 'mean:',emean1,esigma1,emean2,esigma2
       &,emean2g1,esigma2g1,ebet1,erho

500   format(a8,4x,f7.3,f7.3,1x,f7.3,f7.3,f8.3,f8.3,f10.3,f8.3)
       write(1,500) 'n*bias2:',bias2mean1n,bias2sigma1n,bias2mean2n,
       &bias2sigma2n,bias2mean2g1n,bias2sigma2g1n,bias2bet1n,bias2rhon
```

```
550   format(a11,1x,f7.3,f7.3,1x,f7.3,f7.3,f8.3,f8.3,f10.3,f8.3)
       write(1,550) 'n*variance:',vmean1n,vsigma1n,vmean2n,vsigma2n,
       &vmean2g1n,vsigma2g1n,vbet1n,vrhon

555   format(a6,6x,f7.3,f7.3,1x,f7.3,f7.3,f8.3,f8.3,f10.3,f8.3,/)
       write(1,555) 'n*mse:',xmsemean1n,xmsesigma1n,xmsemean2n,
       &xmsesigma2n,xmsemean2g1n,xmsesigma2g1n,xmsebet1n,xmserhon

560   format(a7,5x,f7.3,f7.3,1x,f7.3,f7.3,f8.3,f8.3,f10.3,f8.3)
       write(1,560) 'meanls:',emean1ls,esigma1ls,emean2ls
       &,esigma2ls,emean2g1ls,esigma2g1ls,ebet1ls,erhols

570   format(a10,2x,f7.3,f7.3,x,f7.3,f7.3,f8.3,f8.3,f10.3,f8.3)
       write(1,570) 'n*bias2ls:',bias2mean1lsn,bias2sigma1lsn,
       &bias2mean2lsn,bias2sigma2lsn,bias2mean2g1lsn
       &,bias2sigma2g1lsn,bias2bet1lsn,bias2rholsn

580   format(a13,f6.3,f7.3,x,f7.3,f7.3,f8.3,f8.3,f10.3,f8.3)
       write(1,580) 'n*variancels:',vmean1lsn,vsigma1lsn,vmean2lsn
       &,vsigma2lsn,vmean2g1lsn,vsigma2g1lsn,vbet1lsn,vrholsn

585   format(a8,4x,f7.3,f7.3,x,f7.3,f7.3,f8.3,f8.3,f10.3,f8.3,/)
       write(1,585) 'n*msels:',xmsemean1lsn,xmsesigma1lsn,xmsemean2lsn,
       &xmsesigma2lsn,xmsemean2g1lsn
       &,xmsesigma2g1lsn,xmsebet1lsn,xmserholsn

588   format(a6,8x,f5.1,2x,f5.1,3x,f5.1,2x,f5.1,3x
       &,f5.1,3x,f5.1,5x,f5.1,3x,f5.1)
       write(1,588) 'effvar', effvmean1,effvsigma1,effvmean2,effvsigma2,
       &effvmean2g1,effvsigma2g1,effvbet1,effvrho

589   format(a6,8x,f5.1,2x,f5.1,3x,f5.1,2x,f5.1,3x
       &,f5.1,3x,f5.1,5x,f5.1,3x,f5.1,3/)
       write(1,589) 'effmse', effmsemean1,effmsesigma1,
       &effmsemean2,effmsesigma2,
       &effmsemean2g1,effmsesigma2g1,effmsebet1,effmserho


c *** Calculating the little simulated covariance matrix for the hotelling T2
c     to test mean1=mean10 and mean2=mean20 ***

       detom2=vmean1*vmean2-covmean1mean2**2
       detom2ls=vmean1ls*vmean2ls-covmean1mean2ls**2

       simom211=vmean2/detom2
       simom212=-covmean1mean2/detom2
       simom222=vmean1/detom2

       simom2ls11=vmean2ls/detom2ls
       simom2ls12=-covmean1mean2ls/detom2ls
       simom2ls22=vmean1ls/detom2ls


c *** Simulating the powers of the test statistics for nn=10,000 ***

       xpowerast2=0.0
       xpoweras2t2=0.0
       xpowerasw=0.0
       xpowersimw=0.0
       xpowersimwls=0.0
       simpowert2=0.0
       simpowert2ls=0.0
       simpowert22=0.0
       simpowert22ls=0.0

       do 600 k=1,nn
```

```
c *** The estimated Fisher information matrix of mean1, sigma1, mean2g1, sigma2g1
c     and theta1 ***

       MI(1,1)=1.0*n*((b1/(b1+2.0))*(1.0/(sigma1(k)**2)))
       MI(1,2)=1.0*n*((b1/(b1+2.0))*(psi(b1+1.0)-psi(2.0))
       &/(sigma1(k)**2))
       MI(1,3)=0.0
       MI(1,4)=0.0
       MI(1,5)=0.0
       MI(2,1)=1.0*n*((b1/(b1+2.0))*(psi(b1+1.0)-psi(2.0))
       &/(sigma1(k)**2))
       MI(2,2)=1.0*n*((1.0/(sigma1(k)**2))*(1.0+(b1/(b1+2.0))*
       &(psid_b1plus1+psid_2+(psi(b1+1.0)-psi(2.0))**2)))
       MI(2,3)=0.0
       MI(2,4)=0.0
       MI(2,5)=0.0
       MI(3,1)=0.0
       MI(3,2)=0.0
       MI(3,3)=1.0*n*(b2/(b2+2.0))/(sigma2g1(k)**2)
       MI(3,4)=1.0*n*(b2/(b2+2.0))*
       &(psi(b2+1.0)-psi(2.0))/(sigma2g1(k)**2)
       MI(3,5)=1.0*n*(b2/(b2+2.0))*
       &sigma1(k)*(psi(b1)-psi(1.0))/(sigma2g1(k)**2)
       MI(4,1)=0.0
       MI(4,2)=0.0
       MI(4,3)=1.0*n*(b2/(b2+2.0))*
       &(psi(b2+1.0)-psi(2.0))/(sigma2g1(k)**2)
       MI(4,4)=1.0*n*(1.0/(sigma2g1(k)**2))*(1.0+(b2/(b2+2.0))*
       &(psid_b2plus1+psid_2+(psi(b2+1.0)-psi(2.0))**2))
       MI(4,5)=1.0*n*(b2/(b2+2.0))*(sigma1(k)/(sigma2g1(k)**2))*
       &(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))
       MI(5,1)=0.0
       MI(5,2)=0.0
       MI(5,3)=1.0*n*(b2/(b2+2.0))*
       &sigma1(k)*(psi(b1)-psi(1.0))/(sigma2g1(k)**2)
       MI(5,4)=1.0*n*(b2/(b2+2.0))*(sigma1(k)/(sigma2g1(k)**2))*
       &(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))
       MI(5,5)=1.0*n*(b2/(b2+2.0))*((sigma1(k)/sigma2g1(k))**2)*
       &(psid_b1+psid_1+(psi(b1)-psi(1.0))**2)


c *** The estimated Fisher information matrix of mean1, sigma1, mean2, sigma2
c     and rho ***

       MI2(1,1)=1.0*n*(1/(sigma1(k)**2))*
       &((b1/(b1+2.0))+(b2/(b2+2.0))*((rho(k)**2)/(1-rho(k)**2)))
       MI2(1,2)=1.0*n*(1/(sigma1(k)**2))*
       &((b1/(b1+2.0))*(psi(b1+1.0)-psi(2.0))
       &+(b2/(b2+2.0))*((rho(k)**2)*(psi(b1)-psi(1.0))/(1-rho(k)**2)))
       MI2(1,3)=-1.0*n*(b2/(b2+2.0))*
       &(rho(k)/(sigma1(k)*sigma2(k)*(1-rho(k)**2)))
       MI2(1,4)=-1.0*n*(b2/(b2+2.0))*
       &((rho(k)**2)*(psi(b1)-psi(1.0))/
       &(sigma1(k)*sigma2(k)*(1-rho(k)**2))+
       &(rho(k)*(psi(b2+1.0)-psi(2.0)))/
       &(sigma1(k)*sigma2(k)*sqrt(1-rho(k)**2)))
       MI2(1,5)=-1.0*n*(b2/(b2+2.0))*
       &(rho(k)*(psi(b1)-psi(1.0))/
       &(sigma1(k)*(1-rho(k)**2))-
       &((rho(k)**2)*(psi(b2+1.0)-psi(2.0)))/
       &(sigma1(k)*((1-rho(k)**2)**1.5)))
       MI2(2,1)=1.0*n*(1/(sigma1(k)**2))*
       &((b1/(b1+2.0))*(psi(b1+1.0)-psi(2.0))
       &+(b2/(b2+2.0))*((rho(k)**2)*(psi(b1)-psi(1.0))/(1-rho(k)**2)))
       MI2(2,2)=1.0*n*(1/(sigma1(k)**2))*
       &(1.0+(b1/(b1+2.0))*(psid_b1plus1+psid_2+(psi(b1+1.0)-psi(2.0))**2)
       &+(b2/(b2+2.0))*((rho(k)**2)/(1-rho(k)**2))*
       &(psid_b1+psid_1+(psi(b1)-psi(1.0))**2))
       MI2(2,3)=-1.0*n*(b2/(b2+2.0))*(psi(b1)-psi(1.0))*
       &(rho(k)/(sigma1(k)*sigma2(k)*(1-rho(k)**2)))
```

```
MI2(2,4)=-1.0*n*(b2/(b2+2.0))*
&((psid_b1+psid_1+(psi(b1)-psi(1.0))**2)*
&(rho(k)**2)/(sigma1(k)*sigma2(k)*(1-rho(k)**2))+
&(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))*rho(k)/
&(sigma1(k)*sigma2(k)*sqrt(1-rho(k)**2)))
MI2(2,5)=-1.0*n*(b2/(b2+2.0))*
&((psid_b1+psid_1+(psi(b1)-psi(1.0))**2)*
&rho(k)/(sigma1(k)*(1-rho(k)**2))-
&(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))*(rho(k)**2)/
&(sigma1(k)*((1-rho(k)**2)**1.5)))
MI2(3,1)=-1.0*n*(b2/(b2+2.0))*
&(rho(k)/(sigma1(k)*sigma2(k)*(1-rho(k)**2)))
MI2(3,2)=-1.0*n*(b2/(b2+2.0))*(psi(b1)-psi(1.0))*
&(rho(k)/(sigma1(k)*sigma2(k)*(1-rho(k)**2)))
MI2(3,3)=1.0*n*(b2/(b2+2.0))*
&(1.0/((sigma2(k)**2)*(1-rho(k)**2)))
MI2(3,4)=1.0*n*(b2/(b2+2.0))*
&(rho(k)*(psi(b1)-psi(1.0))/((sigma2(k)**2)*(1-rho(k)**2))+
&(psi(b2+1.0)-psi(2.0))/((sigma2(k)**2)*sqrt(1-rho(k)**2)))
MI2(3,5)=-1.0*n*(b2/(b2+2.0))*
&((psi(b2+1.0)-psi(2.0))*
&rho(k)/(sigma2(k)*((1-rho(k)**2)**1.5))-
&(psi(b1)-psi(1.0))/(sigma2(k)*(1-rho(k)**2)))
MI2(4,1)=-1.0*n*(b2/(b2+2.0))*
&((rho(k)**2)*(psi(b1)-psi(1.0))/
&(sigma1(k)*sigma2(k)*(1-rho(k)**2))+
&(rho(k)*(psi(b2+1.0)-psi(2.0)))/
&(sigma1(k)*sigma2(k)*sqrt(1-rho(k)**2)))
MI2(4,2)=-1.0*n*(b2/(b2+2.0))*
&((psid_b1+psid_1+(psi(b1)-psi(1.0))**2)*
&(rho(k)**2)/(sigma1(k)*sigma2(k)*(1-rho(k)**2))+
&(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))*rho(k)/
&(sigma1(k)*sigma2(k)*sqrt(1-rho(k)**2)))
MI2(4,3)=1.0*n*(b2/(b2+2.0))*
&(rho(k)*(psi(b1)-psi(1.0))/((sigma2(k)**2)*(1-rho(k)**2))+
&(psi(b2+1.0)-psi(2.0))/((sigma2(k)**2)*sqrt(1-rho(k)**2)))
MI2(4,4)=1.0*n*(1.0/(sigma2(k)**2))*
&(1.0+(b2/(b2+2.0))*((psid_b1+psid_1+(psi(b1)-psi(1.0))**2)*
&(rho(k)**2)/(1-rho(k)**2)+
&2.0*rho(k)*(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))/
&sqrt(1-rho(k)**2)+
&psid_b2plus1+psid_2+(psi(b2+1.0)-psi(2.0))**2))
MI2(4,5)=-1.0*n*(1.0/sigma2(k))*
&(rho(k)/(1-rho(k)**2)+(b2/(b2+2.0))*(
&(-1.0)*(psid_b1+psid_1+(psi(b1)-psi(1.0))**2)*
&rho(k)/(1-rho(k)**2)+
&(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))*
&(rho(k)**2)/((1-rho(k)**2)**1.5)-
&(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))/
&sqrt(1-rho(k)**2)+
&(psid_b2plus1+psid_2+(psi(b2+1.0)-psi(2.0))**2)*rho(k)/
&(1-rho(k)**2)))
MI2(5,1)=-1.0*n*(b2/(b2+2.0))*
&(rho(k)*(psi(b1)-psi(1.0))/
&(sigma1(k)*(1-rho(k)**2))-
&((rho(k)**2)*(psi(b2+1.0)-psi(2.0)))/
&(sigma1(k)*((1-rho(k)**2)**1.5)))
MI2(5,2)=-1.0*n*(b2/(b2+2.0))*
&((psid_b1+psid_1+(psi(b1)-psi(1.0))**2)*
&rho(k)/(sigma1(k)*(1-rho(k)**2))-
&(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))*(rho(k)**2)/
&(sigma1(k)*((1-rho(k)**2)**1.5)))
MI2(5,3)=-1.0*n*(b2/(b2+2.0))*
&((psi(b2+1.0)-psi(2.0))*
&rho(k)/(sigma2(k)*((1-rho(k)**2)**1.5))-
&(psi(b1)-psi(1.0))/(sigma2(k)*(1-rho(k)**2)))
```

```
        MI2(5,4)=-1.0*n*(1.0/sigma2(k))*
        &(rho(k)/(1-rho(k)**2)+(b2/(b2+2.0))*(
        &(-1.0)*(psid_b1+psid_1+(psi(b1)-psi(1.0))**2)*
        &rho(k)/(1-rho(k)**2)+
        &(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))*
        &(rho(k)**2)/((1-rho(k)**2)**1.5)-
        &(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))/
        &sqrt(1-rho(k)**2)+
        &(psid_b2plus1+psid_2+(psi(b2+1.0)-psi(2.0))**2)*rho(k)/
        &(1-rho(k)**2)))
        MI2(5,5)=1.0*n*((rho(k)**2)/((1-rho(k)**2)**2)+
        &(b2/(b2+2.0))*((psid_b1+psid_1+(psi(b1)-psi(1.0))**2)
        &/(1-rho(k)**2)-2.0*rho(k)*
        &(psi(b1)-psi(1.0))*(psi(b2+1.0)-psi(2.0))/
        &((1-rho(k)**2)**1.5)+
        &((rho(k)**2)/((1-rho(k)**2)**2))*
        &(psid_b2plus1+psid_2+(psi(b2+1.0)-psi(2.0))**2)))


c *** Just taking the inverse of the estimated Fisher information matrices ***

        call LINRG (KI,MI,LI,MIINV,LIINV)
        call LINRG (KI2,MI2,LI2,MI2INV,LI2INV)


c *** Taking the elements from the inverse of the Fisher Information matrices
c     which are the estimated asymptotic covariance matrices ***

        asvarmean1(k)=MIINV(1,1)
        asvarsigma1(k)=MIINV(2,2)
        asvarmean2g1(k)=MIINV(3,3)
        asvarsigma2g1(k)=MIINV(4,4)
        asvarbet1(k)=MIINV(5,5)
        ascovmean1mean2g1(k)=MIINV(1,3)

        asvarmean12(k)=MI2INV(1,1)
        asvarsigma12(k)=MI2INV(2,2)
        asvarmean2(k)=MI2INV(3,3)
        asvarsigma2(k)=MI2INV(4,4)
        asvarrho(k)=MI2INV(5,5)
        ascovmean1mean2(k)=MI2INV(1,3)


c *** Calculating the little estimated asymptotic covariance matrix for
c     the hotelling T2 to test mean1=mean10 and mean2=mean20 ***

        asbnvmean12=asvarmean12(k)*(1.0*n)
        asbnvmean2=asvarmean2(k)*(1.0*n)
        asbncovmean1mean2=ascovmean1mean2(k)*(1.0*n)

        detasom2=asbnvmean12*asbnvmean2-asbncovmean1mean2**2

        asom211=asbnvmean2/detasom2
        asom212=-asbncovmean1mean2/detasom2
        asom222=asbnvmean12/detasom2


c *** The followings are the test statistics based on asymptotic variances ***
c *** ast2 is for mean1=mean10,mean2g1=mean2g10 ***
c *** as2t2 is for mean1=mean10,mean2=mean20 ***
c *** asw is for rho=0 ***

c *** The followings are the test statistics based on simulated variances ***
c *** simt2 is for mean1=mean10,mean2g1=mean2g10 ***
c *** sim2t2 is for mean1=mean10,mean2=mean20 ***
c *** simw is for rho=0 ***
```

```
c *** Calculating the test statistics ***

        ast2=((mean1(k)-mean10)**2)/asvarmean1(k)+
        &((mean2g1(k)-mean2g10)**2)/asvarmean2g1(k)

        ast2f=ast2*((1.0*n-2.0)/(2.0*(1.0*n-1.0)))

        as2t2=1.0*n*(((mean1(k)-mean10)**2)*asom211+
        &((mean2(k)-mean20)**2)*asom222+
        &2.0*(mean1(k)-mean10)*(mean2(k)-mean20)*asom212)

        as2t2f=as2t2*((1.0*n-2.0)/(2.0*(1.0*n-1.0)))

        if(prho.eq.0.0)then
        vrho_underH0=vrho
        vrhols_underH0=vrhols
        endif

        simw=rho(k)/sqrt(vrho_underH0)
        simwls=rhols(k)/sqrt(vrhols_underH0)

        asw=rho(k)*sqrt(1.0*n*(psid_b1+psid_1)*(b2/(b2+2.0)))
        aswls=rhols(k)*sqrt(1.0*n*(psid_b1+psid_1)*(b2/(b2+2.0)))

        simt2=((mean1(k)-mean10)**2)/vmean1+
        &((mean2g1(k)-mean2g10)**2)/vmean2g1

        simt2f=simt2*((1.0*n-2.0)/(2.0*(1.0*n-1.0)))

        simt2ls=((mean1ls(k)-mean10)**2)/vmean1ls+
        &((mean2g1ls(k)-mean2g10)**2)/vmean2g1ls

        simt2lsf=simt2ls*((1.0*n-2.0)/(2.0*(1.0*n-1.0)))

        simt22=((mean1(k)-mean10)**2)*simom211+
        &((mean2(k)-mean20)**2)*simom222+
        &2.0*(mean1(k)-mean10)*(mean2(k)-mean20)*simom212

        simt22f=simt22*((1.0*n-2.0)/(2.0*(1.0*n-1.0)))

        simt22ls=((mean1ls(k)-mean10)**2)*simom2ls11+
        &((mean2ls(k)-mean20)**2)*simom2ls22+
        &2.0*(mean1ls(k)-mean10)*(mean2ls(k)-mean20)*simom2ls12
        simt22lsf=simt22ls*((1.0*n-2.0)/(2.0*(1.0*n-1.0)))


c *** Calculating the simulated powers by comparing with the tabulated values ***

        if(ast2.GT.xki2inv) then
        xpowerast2=xpowerast2+1.0
        endif

        if(as2t2.GT.xki2inv) then
        xpoweras2t2=xpoweras2t2+1.0
        endif

        if(asw.GT.xz095) then
        xpowerasw=xpowerasw+1.0
        endif

        if(simw.GT.xz095) then
        xpowersimw=xpowersimw+1.0
        endif

        if(simwls.GT.xz095) then
        xpowersimwls=xpowersimwls+1.0
        endif
```

```
        if(simt2.GT.xki2inv) then
        simpowert2=simpowert2+1.0
        endif

        if(simt2ls.GT.xki2inv) then
        simpowert2ls=simpowert2ls+1.0
        endif

        if(simt22.GT.xki2inv) then
        simpowert22=simpowert22+1.0
        endif

        if(simt22ls.GT.xki2inv) then
        simpowert22ls=simpowert22ls+1.0
        endif

600   continue

        xpowerast2=xpowerast2/(1.0*nn)
        xpoweras2t2=xpoweras2t2/(1.0*nn)
        xpowerasw=xpowerasw/(1.0*nn)

        xpowersimw=xpowersimw/(1.0*nn)
        xpowersimwls=xpowersimwls/(1.0*nn)

        simpowert2=simpowert2/(1.0*nn)
        simpowert2ls=simpowert2ls/(1.0*nn)

        simpowert22=simpowert22/(1.0*nn)
        simpowert22ls=simpowert22ls/(1.0*nn)


c *** The End ***


        end
```

VITA

Hakan Savaş Sazak was born in Ankara on December 27, 1973. He received his B.S. degree in Statistics from the Middle East Technical University in June 1996. Since then he has been a research assistant in the Department of Statistics. In August 1999, he received his M.S. degree from the same department. His main area of interest is robust estimation and hypothesis testing in stochastic regression.