THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

PARAMETRIC ESTIMATION IN A DOUBLY TRUNCATED BIVARIATE NORMAL DISTRIBUTION

by

Danny D. Dyer

Technical Report No. 67
Department of Statistics THEMIS Contract

April 3, 1970

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

PARAMETRIC ESTIMATION IN A DOUBLY TRUNCATED

BIVARIATE NORMAL DISTRIBUTION

A Dissertation Presented to the Faculty of the Graduate School

of

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Statistics

by

Danny D. Dyer
(M.S., Southern Methodist University, 1963)

April 3, 1970

Dyer, Danny D.

B.S., University of Texas at Arlington, 1961
M.S., Southern Methodist University, 1963

<u>Parametric Estimation in a Doubly Truncated Bivariate Normal Distribution</u>

Adviser: Professor Donald B. Owen

Doctor of Philosophy degree conferred May 24, 1970

Dissertation completed April 3, 1970

When sampling from a bivariate normal population, observations may be restricted to certain fixed regions of the two-dimensional plane over which the distribution is defined by truncating the distribution. In this paper, a three-step numerical procedure is proposed to solve the maximum likelihood equations, or equivalently the method of moments equations, for estimating the parameters of a bivariate normal distribution truncated outside a rectangle $\{(x, y): a < x < b, c < y < d\}$. In the first step, an initial set of consistent estimates explicitly given in terms of thirteen sample moments are found. Next, iterants are developed from the method of moments equations, and through the use of the initial estimates as a starting vector, an advancement to a neighborhood of the solution is made with one or more cycles of the functional iterative method. The final estimates are obtained from a single cycle of the Newton-Raphson method which, without any additional computational work, gives an estimate of the asymptotic variance-covariance matrix of the maximum likelihood estimates.

The maximum likelihood equations for estimating the parameters of a bivariate normal distribution truncated outside the infinite strip $\{(x, y): a < x < b, -\infty < y < +\infty\}$ contain the ratios $S(u, v) = [\phi(u)-\phi(v)]/[\Phi(v)-\Phi(u)]$ and $T(u, v) = [u\phi(u)-v\phi(v)]/[\Phi(v)-\Phi(u)]$,

where $\phi(t)$ is the standard normal density function, $\Phi(t)$ is the corresponding cumulative distribution function, and u and v are functions of parameters and truncation points. Thus the maximum likelihood equations are non-linear equations in the parameters. Tiku (1968, <u>Australian Journal of Statistics</u>, <u>10</u>, pp. 64-74) made linear approximations to the ratios $\phi(u)/[\Phi(v)-\Phi(u)]$ and $\phi(v)/[\Phi(v)-\Phi(u)]$, thus simplifying the maximum likelihood equations so that explicit expressions for the estimates could be found. In this paper, using multiple linear regression, a second-order linear model in u and v is fitted to S(u, v) and a third-order linear model in u and v is fitted to T(u, v). The second-order linear model provides a better fit of S(u, v) than does the linear approximation and is valid for any sample size whereas the linear approximation was constructed under the assumption of a large sample size. [An analogous statement may be made for T(u, v).] Using the approximations of this paper in the maximum likelihood equations, explicit expressions for the estimates are obtained which require fewer calculations to evaluate than do those of Tiku.

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to my adviser, Dr. Donald B. Owen, for his constant helpfulness during the writing of this dissertation and for his financial assistance through a THEMIS research assistantship. I wish to thank Dr. Richard P. Bland, Dr. C. H. Kapadia, Dr. Paul D. Minton, Dr. Charles J. Pipes, Jr., and Dr. John T. Webster for their guidance through my graduate studies.

I am very appreciative to Mr. Bill Frawley for his time spent on running the regression programs and to Mrs. Linda White for her excellent job of typing. Finally, my deepest expression of gratitude goes to my wife, Marlene, whose encouragement helped me to achieve this goal.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

When sampling from a multivariate normal population, in many instances observations are restricted to certain regions of the population. For example, in Euclidean 2-space (E^2), suppose rounds are being fired at a rectangular target $\{(x, y): a < x < b, c < y < d\}$. If the number of rounds fired is unknown and the only data come from the measurements of the horizontal and vertical displacements with respect to the origin of the points of impact, then in effect the data constitute a sample from a truncated bivariate normal distribution under the assumption gunfire pattern may be approximated by a bivariate normal distribution. If the number of rounds fired is known with the number of rounds hitting the target a discrete random variable, the data constitute a type I censored sample from a bivariate normal distribution. If the number of rounds fired is known with the number of rounds hitting the target fixed thus requiring the rectangle to have random vertices, the data constitute a type II censored sample from a bivariate normal distribution. Truncation is a characteristic of the population; censoring is a characteristic of the sample. These are two basic ways in which observations are restricted. A complete discussion along with a bibliography of various ways observations may be restricted can be found in Federer (1963).

It is of interest to examine effects of truncation in a multivariate normal distribution. Hotelling (1948) has observed that a set of sufficient statistics for a multivariate normal distribution also serve as a set of sufficient statistics for a multivariate normal distribution truncated outside an arbitrary set of positive probability measure. Smith (1957) has shown that under general conditions if a sufficient statistic has one or more of the properties of completeness, bounded completeness, or minimality before truncation, then it has the same properties after truncation.

An implicit expression for the mn^{th} product moment of the standardized p-variate normal distribution with correlation matrix $R = (r_{ij})$ truncated outside the region $\{(x_1, x_2, \cdots, x_p): x_i \geq t_i, i = 1, 2, \cdots, p\}$ was derived through direct integration by Birnbaum and Meyer (1953). Unfortunately, the general results are left in a somewhat difficult form with which to work if explicit expressions for moments are desired. In an attempt to improve on this difficulty, Tallis (1961) found the moment generating function for the truncated distribution considered by Birnbaum and Meyer. Thus, explicit expressions for moments were obtained through differentiation although this becomes somewhat cumbersome in higher dimensions.

In a paper concerned with estimation, Singh (1960) has derived the maximum likelihood equations for estimating the parameters of an uncorrelated p-variate normal distribution from a type I censored sample with the unmeasured observations lying outside the region $\{(x_1, x_2, \cdots, x_p): a_i \leq x_i \leq b_i, i = 1, 2, \cdots, p\}$. Because of the similarity in the likelihood functions, the maximum likelihood equations for estimating the parameters of an uncorrelated p-variate normal distribution truncated outside the region $\{(x_1, x_2, \cdots, x_p): a_i \leq x_i \leq b_i, i = 1, 2, \cdots, p\}$ were also derived. It was suggested that the type I censored sample estimating equations be solved simultaneously by iteration. The estimating equations for the truncated distribution may be solved pairwise with respect to location and scale parameters for each variate separately. Further, it was apparent that irrespective of the fact whether all or only a subset of the variates were truncated, the estimating equations corresponding to the truncated variates were the same. The estimating equations corresponding to the variates which were not truncated were the usual maximum likelihood equations. It must be pointed out that the above statements pertaining to the estimating equations for the truncated distribution are not true if the variates are correlated.

Because of the complexity of the density for the truncated multivariate normal distribution, practically all results pertaining to expressions for moments as well as to estimation of parameters are left in rather unwieldy forms. For this reason many authors have restricted their attention to problems pertaining to the bivariate normal distribution truncated over some region in E^2.

Weiler (1959) and then Williams and Weiler (1964) obtained explicit expressions for the first and second moments of the standardized bivariate normal distribution singly linearly truncated in both variables, i.e., truncated outside the region $\{(x, y): a < x < +\infty, b < y < +\infty\}$. From these expressions charts were constructed which graphically depict the relationships the truncation points a and b have with the two population means and the proportion of the population retained after truncation for $\rho = 0, \pm .2, \pm .3, \pm .4, \pm .5, \pm .6, \pm .8$. The charts can be used in various ways such as a.) for given population means, the required truncation points as well as the proportion of the population retained can be read; b.) for a given proportion retained from the original population, the most suitable combination of population means can be selected along with the corresponding truncation points; c.) for given truncation points, the population means of the truncated distribution as well as the proportion of the population retained can be read. This last use of the charts along with a relationship given by Tallis (1961) allow graphical determination for given truncation points of the population means of the standardized bivariate normal distribution doubly linearly truncated in both variables, i.e., truncated outside the region $\{(x, y): a_1 < x < a_2, b_1 < y < b_2\}$.

Estimation by the method of maximum likelihood of the parameters of a bivariate normal distribution singly or doubly linearly truncated in one variable has been considered by Raj (1953) and Cohen (1955). Raj, evidently unaware of the work of Hotelling (1948), also noted that the method of moments gave the same results by showing that the estimating equations for both methods were identical. Because of truncation in only one variable, two of the five estimating equations could be solved simultaneously, it was suggested, by using Newton's method. Once these two estimates were found, the other three estimates were easily obtained. Although using a different approach, Cohen essentially did the same thing as Raj. Since the maximum likelihood estimates of the parameters are solutions to nonlinear equations, no explicit expressions for these estimates have been found. However, by replacing population moments with sample moments in a recurrence relation for the population moments of the bivariate normal distribution singly linearly truncated in one variable, Jaiswal and Khatri (1967) were able to find explicit expressions for estimates of the parameter in terms of six sample moments.

Estimation by the method of moments of the parameters of a bivariate normal distribution singly linearly truncated in both variables has been considered by Rosenbaum (1961). He derived by direct integration explicit expressions for the first two population moments as well as the population product moment of the two truncated variables. By equating these expressions to corresponding sample moments, he obtained a system of five nonlinear equations the solution to which yield the estimates of the parameters. A suggested iterative method was given to solve the system of equations. However, the iterants for ρ were obtained by solving a quadratic equation in ρ which could cause considerable difficulty with respect to convergence. Fortunately, Khatri and Jaiswal (1963) have slightly modified Rosenbaum's iterative procedure through the use of a linear equation in ρ. But the main purpose of their work was to obtain explicit expressions for estimates of the parameters of a bivariate normal distribution singly linearly truncated in both variables in terms of eight or nine sample moments.

However, all of the aforementioned moment as well as parametric estimation problems pertaining to a truncated bivariate normal distribution deal with distributions which are special cases of a more general distribution -- the bivariate normal distribution doubly linearly truncated at known points in both variables. Presumably because the number of terms involved in the estimation of the parameters of the general distribution more than quadruples the number of terms involved in the estimation of the parameters of the special cases, no apparent attempt toward estimation in the general distribution has been made. Furthermore, one cannot generalize the results for the special cases to handle the more general case because of the additional terms. Thus, this paper will be concerned with the parametric estimation problem in the bivariate normal distribution doubly linearly truncated at known points in both variables.

CHAPTER II

DERIVATION OF THE ESTIMATING EQUATIONS

1. Density Function

The density function of a two-dimensional random variable $(\underline{\bar{X}}, \underline{\bar{Y}})$ having a bivariate normal distribution doubly linearly truncated in both variables is given by

$$g(x,y) = \frac{\dfrac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\dfrac{1}{2(1-\rho^2)}\left[\left(\dfrac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\dfrac{x-\mu_x}{\sigma_x}\right)\left(\dfrac{y-\mu_y}{\sigma_y}\right) + \left(\dfrac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}}{\displaystyle\int_c^d \int_a^b \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\} dxdy},$$

$$a < x < d$$

$$c < y < d$$

$$= 0, \text{ elsewhere,} \tag{1}$$

where a, b, c, and d are known constants.

So that expressions will be more concise, the origin in the xy-plane is translated to the lower truncation point. The density function of the transformed variable $(\underline{\bar{X}}', \underline{\bar{Y}}') = (\underline{\bar{X}} - a, \underline{\bar{Y}} - c)$ is

$$f(x',y') = \frac{\dfrac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\dfrac{1}{2(1-\rho^2)}\left[\left(\dfrac{x'}{\sigma_x} + \xi\right)^2 - 2\rho\left(\dfrac{x'}{\sigma_x} + \xi\right)\left(\dfrac{y'}{\sigma_y} + \eta\right) + \left(\dfrac{y'}{\sigma_y} + \eta\right)^2\right]\right\}}{R\left(\xi,\ \xi + \dfrac{R_x}{\sigma_x},\ \eta,\ \eta + \dfrac{R_y}{\sigma_y};\ \rho\right)},$$

$$0 < x' < R_x$$

$$0 < y' < R_y$$

$$= 0, \text{ elsewhere,} \tag{2}$$

where

$$R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y}\ ;\ \rho\right) =$$

$$= \int_{\eta}^{\eta+R_y/\sigma_y} \int_{\xi}^{\xi+R_x/\sigma_x} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(s^2-2\rho st+t^2)\right] ds\,dt \tag{3}$$

and

$$\xi = \frac{a-\mu_x}{\sigma_x}\ ,\ \eta = \frac{c-\mu_y}{\sigma_y}\ ,\ R_x = b-a\ ,\ R_y = d-c\ . \tag{4}$$

By making the transformation $u = \dfrac{s-\rho t}{\sqrt{1-\rho^2}}$, we may write

$$R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y}\ ;\ \rho\right) = \int_{\eta}^{\eta+R_y/\sigma_y} \phi(t)\left[\Phi\left(\frac{\xi+\frac{R_x}{\sigma_x}-\rho t}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\xi-\rho t}{\sqrt{1-\rho^2}}\right)\right]dt\ ,$$

or because of symmetry

$$R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y}\ ;\ \rho\right) = \int_{\xi}^{\xi+R_x/\sigma_x} \phi(t)\left[\Phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho t}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\eta-\rho t}{\sqrt{1-\rho^2}}\right)\right]dt\ ,$$

where

$$\phi(t) = \frac{1}{\sqrt{2\pi}}\, e^{-1/2\, t^2}$$

and

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}\, e^{-1/2\, z^2}\, dz\ .$$

2. Maximum Likelihood Equations

The likelihood function for a sample of size N from this population is given by

$$L = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\sum_{i=1}^{N}\left(\frac{x_i'}{\sigma_x}+\xi\right)^2 - 2\rho\sum_{i=1}^{N}\left(\frac{x_i'}{\sigma_x}+\xi\right)\left(\frac{y_i'}{\sigma_y}+\eta\right) + \sum_{i=1}^{N}\left(\frac{y_i'}{\sigma_y}+\eta\right)^2\right]\right\}}{\left[2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}\ R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y}\ ;\ \rho\right)\right]^N} \tag{5}$$

Taking the logarithm of the likelihood function, differentiating partially with respect to ξ, η, σ_x, σ_y, and ρ, and equating the partial derivatives to zero yield the following maximum likelihood equations, the solution to which are the maximum likelihood estimates $\hat{\xi}$, $\hat{\eta}$, $\hat{\sigma}_x$, $\hat{\sigma}_y$, $\hat{\rho}$.

$$\frac{\partial \log L}{\partial \xi} = N\left[(G_1 - G_2) - \frac{(\xi-\rho\eta)}{1-\rho^2} - \frac{m_{1,0}}{\sigma_x(1-\rho^2)} + \frac{\rho m_{0,1}}{\sigma_y(1-\rho^2)}\right] = 0\ ; \tag{6}$$

$$\frac{\partial \log L}{\partial \eta} = N\left[(G_3 - G_4) - \frac{(\eta-\rho\xi)}{1-\rho^2} + \frac{\rho m_{1,0}}{\sigma_x(1-\rho^2)} - \frac{m_{0,1}}{\sigma_y(1-\rho^2)}\right] = 0\ ; \tag{7}$$

$$\frac{\partial \log L}{\partial \sigma_x} = N\left[\frac{R_x}{\sigma_x^2} G_2 - \frac{1}{\sigma_x} + \frac{m_{2,0}}{\sigma_x^3(1-\rho^2)} - \frac{\rho m_{1,1}}{\sigma_x^2\sigma_y(1-\rho^2)} + \frac{(\xi-\rho\eta)m_{1,0}}{\sigma_x^2(1-\rho^2)}\right] = 0\ ; \tag{8}$$

$$\frac{\partial \log L}{\partial \sigma_y} = N\left[\frac{R_y}{\sigma_y^2} G_4 - \frac{1}{\sigma_y} + \frac{m_{0,2}}{\sigma_y^3(1-\rho^2)} - \frac{\rho m_{1,1}}{\sigma_x\sigma_y^2(1-\rho^2)} + \frac{(\eta-\rho\xi)m_{0,1}}{\sigma_y^2(1-\rho^2)}\right] = 0\ ; \tag{9}$$

$$\frac{\partial \log L}{\partial \rho} = N\left[\frac{-g_5}{1-\rho^2} + \frac{\rho}{1-\rho^2} + \frac{\xi\eta\rho^2-(\xi^2+\eta^2)\rho+\xi\eta}{(1-\rho^2)^2} - \frac{\rho m_{2,0}}{\sigma_x^2(1-\rho^2)^2} - \frac{\rho m_{0,2}}{\sigma_y^2(1-\rho^2)^2}\right.$$
$$\left. + \frac{(1+\rho^2)m_{1,1}}{\sigma_x\sigma_y(1-\rho^2)^2} + \frac{(\eta\rho^2-2\xi\rho+\eta)m_{1,0}}{\sigma_x(1-\rho^2)^2} + \frac{(\xi\rho^2-2\eta\rho+\xi)m_{0,1}}{\sigma_y(1-\rho^2)^2}\right] = 0; \tag{10}$$

where

$$G_1 = \phi(\xi)\left[\Phi\left(\frac{\eta + \frac{R_y}{\sigma_y} - \rho\xi}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\eta-\rho\xi}{\sqrt{1-\rho^2}}\right)\right] \Big/ R\left(\xi,\ \xi + \frac{R_x}{\sigma_x},\ \eta,\ \eta + \frac{R_y}{\sigma_y}\ ;\ \rho\right), \tag{11}$$

$$G_2 = \phi\left(\xi + \frac{R_x}{\sigma_x}\right)\left[\Phi\left(\frac{\eta + \frac{R_y}{\sigma_y} - \rho\left(\xi + \frac{R_x}{\sigma_x}\right)}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\eta-\rho\left(\xi + \frac{R_x}{\sigma_x}\right)}{\sqrt{1-\rho^2}}\right)\right] \Big/ R\left(\xi, \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta + \frac{R_y}{\sigma_y};\rho\right), \tag{12}$$

$$G_3 = \phi(\eta)\left[\Phi\left(\frac{\xi + \frac{R_x}{\sigma_x} - \rho\eta}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\xi-\rho\eta}{\sqrt{1-\rho^2}}\right)\right] \Big/ R\left(\xi,\ \xi + \frac{R_x}{\sigma_x},\ \eta,\ \eta + \frac{R_y}{\sigma_y}\ ;\ \rho\right), \tag{13}$$

$$G_4 = \phi\left(\eta + \frac{R_y}{\sigma_y}\right)\left[\Phi\left(\frac{\xi + \frac{R_x}{\sigma_x} - \rho\left(\eta + \frac{R_y}{\sigma_y}\right)}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\xi-\rho\left(\eta + \frac{R_y}{\sigma_y}\right)}{\sqrt{1-\rho^2}}\right)\right] \Big/ R\left(\xi, \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta + \frac{R_y}{\sigma_y};\rho\right), \tag{14}$$

$$g_1 = \phi(\xi)\phi\left(\frac{\eta-\rho\xi}{\sqrt{1-\rho^2}}\right) \Big/ R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\ \rho\right)$$

$$= \phi(\xi)\phi\left(\frac{\xi-\rho\eta}{\sqrt{1-\rho^2}}\right) \Big/ R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\ \rho\right), \tag{15}$$

$$g_2 = \phi(\xi)\phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho\xi}{\sqrt{1-\rho^2}}\right) \Big/ R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\ \rho\right)$$

$$= \phi\left(\eta+\frac{R_y}{\sigma_y}\right)\phi\left(\frac{\xi-\rho\left(\eta+\frac{R_y}{\sigma_y}\right)}{\sqrt{1-\rho^2}}\right) \Big/ R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\rho\right), \tag{16}$$

$$g_3 = \phi\left(\xi+\frac{R_x}{\sigma_x}\right)\phi\left(\frac{\eta-\rho\left(\xi+\frac{R_x}{\sigma_x}\right)}{\sqrt{1-\rho^2}}\right) \Big/ R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\ \rho\right)$$

$$= \phi(\eta)\phi\left(\frac{\xi+\frac{R_x}{\sigma_x}-\rho\eta}{\sqrt{1-\rho^2}}\right) \Big/ R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\rho\right), \tag{17}$$

$$g_4 = \phi\left(\xi+\frac{R_x}{\sigma_x}\right)\phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho\left(\xi+\frac{R_x}{\sigma_x}\right)}{\sqrt{1-\rho^2}}\right) \Big/ R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\ \rho\right)$$

$$= \phi\left(\eta+\frac{R_y}{\sigma_y}\right)\phi\left(\frac{\xi+\frac{R_x}{\sigma_x}-\rho\left(\eta+\frac{R_y}{\sigma_y}\right)}{\sqrt{1-\rho^2}}\right) \Big/ R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\rho\right), \tag{18}$$

$$g_5 = \sqrt{1-\rho^2}\,(g_1 - g_2 - g_3 + g_4), \tag{19}$$

and the sample moments are denoted by

$$m_{r,s} = \frac{1}{N}\sum_{i=1}^{N} x_i'^{\,r}\, y_i'^{\,s}. \tag{20}$$

Note that in the simplification of the expression for $\frac{\partial \log L}{\partial \rho}$, the following result was needed.

$$\int_\alpha^\beta \phi(t)\,\frac{\partial\Phi\left(\frac{\gamma-\rho t}{\sqrt{1-\rho^2}}\right)}{\partial\rho}\,dt = -\int_\alpha^\beta \left[\frac{t-\rho\gamma}{(1-\rho^2)^{3/2}}\right]\phi(t)\,\phi\left(\frac{\gamma-\rho t}{\sqrt{1-\rho^2}}\right)dt$$

$$= -\,\phi(\gamma)\int_\alpha^\beta \left[\frac{t-\rho\gamma}{(1-\rho^2)^{3/2}}\right]\phi\left(\frac{t-\rho\gamma}{\sqrt{1-\rho^2}}\right)dt$$

$$= \frac{\phi(\gamma)}{\sqrt{1-\rho^2}}\left[\phi\left(\frac{\beta-\rho\gamma}{\sqrt{1-\rho^2}}\right) - \phi\left(\frac{\alpha-\rho\gamma}{\sqrt{1-\rho^2}}\right)\right].$$

The bivariate density $f(x', y')$ belongs to the regular exponential family of distributions since

$$f(x', y') = B(\xi, \eta, \sigma_x, \sigma_y, \rho)h(x', y')\exp\left[\sum_{i=1}^{5} \pi_i(\xi, \eta, \sigma_x, \sigma_y, \rho)t_i(x', y')\right],$$

where

$$B(\xi, \eta, \sigma_x, \sigma_y, \rho) = \frac{\exp\left[-\frac{1}{2(1-\rho^2)}(\xi^2 - 2\rho\xi\eta + \eta^2)\right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}\;R\left(\xi,\ \xi + \frac{R_x}{\sigma_x},\ \eta,\ \eta + \frac{R_y}{\sigma_y};\rho\right)},$$

$$h(x', y') = \begin{cases} 1, & (x', y')\in E \\ 0, & (x', y')\notin E \end{cases} \quad \text{and} \quad E = \left\{(x', y') : \begin{matrix} 0 < x' < R_x, \\ 0 < y' < R_y \end{matrix}\right\},$$

$$\pi_1(\xi, \eta, \sigma_x, \sigma_y, \rho) = \frac{\rho\eta-\xi}{\sigma_x(1-\rho^2)}, \quad t_1(x', y') = x',$$

$$\pi_2(\xi, \eta, \sigma_x, \sigma_y, \rho) = \frac{\rho\xi-\eta}{\sigma_y(1-\rho^2)}, \quad t_2(x', y') = y',$$

$$\pi_3(\xi, \eta, \sigma_x, \sigma_y, \rho) = \frac{-1}{2\sigma_x^2(1-\rho^2)}, \quad t_3(x', y') = x'^2,$$

$$\pi_4(\xi, \eta, \sigma_x, \sigma_y, \rho) = \frac{-1}{2\sigma_y^2(1-\rho^2)}, \quad t_4(x', y') = y'^2,$$

$$\pi_5(\xi, \eta, \sigma_x, \sigma_y, \rho) = \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)}, \quad t_5(x', y') = x'y'.$$

Hence, for a sample of size N we see that

$$\left(\sum_{i=1}^{N} X_i' \, , \, \sum_{i=1}^{N} Y_i' \, , \, \sum_{i=1}^{N} X_{ii}'^2 \, , \, \sum_{i=1}^{N} Y_i'^2 \, , \, \sum_{i=1}^{N} X_i'Y_i' \right) \tag{21}$$

is a minimal set of jointly sufficient statistics for the parameters $(\xi, \eta, \sigma_x, \sigma_y, \rho)$. Furthermore, a convex natural parameter space in E^5 can be easily found so that if $N \geq 5$, then (21) is a jointly complete (minimal) set of sufficient statistics. The existence of such a set ensures that the maximum likelihood equations (6) through (10) have a unique solution, and that this solution maximizes the likelihood function. This result is due to Huzurbazar (1949) who showed that, under regularity conditions, the same was true for any exponential family of distributions admitting a set of jointly sufficient statistics.

Since (21) is a minimal set of jointly sufficient statistics, then by the factorization theorem of Neyman and Pearson (1936) the likelihood function may be written as

$$L = p\left(\sum_{i=1}^{N} x_i', \, \sum_{i=1}^{N} y_i', \, \sum_{i=1}^{N} x_i'^2, \, \sum_{i=1}^{N} y_i'^2, \, \sum_{i=1}^{N} x_i'y_i'; \, \xi, \eta, \sigma_x, \sigma_y, \rho \right) q(\underline{x}', \underline{y}') \ ,$$

where $q(\underline{x}', \underline{y}') = \prod_{i=1}^{N} h(x_i', y_i')$ does not contain $\xi, \eta, \sigma_x, \sigma_y, \rho$. Clearly, the $\hat{\xi}, \hat{\eta}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho}$ which maximize L also maximize p, and since the maximum likelihood equations have a unique solution, the maximum likelihood estimators must be single-valued functions of the minimal sufficient statistics. Using this fact and keeping in mind the form of the minimal sufficient statistics, one might suspect that the estimates of the parameters of f(x', y') found by the method of moments are equivalent to the maximum likelihood estimates. That this, in fact, is the case was evidently shown by Hotelling (1948).

3. Method of Moments Equations

We shall find it necessary to work with not only the system of nonlinear equations given by the method of maximum likelihood but also the equivalent system of nonlinear equations given by the method of moments. We begin the derivation of the method of moments estimating equations by first finding the population moments about the origin from the moment generating function of $f(x', y')$. Denoting the moment generating function of $f(x', y')$ by $M_{X',Y'}(t_1, t_2)$ and noting that

$$M_{X',Y'}(t_1, t_2) = M_{X'/\sigma_x+\xi,\ Y'/\sigma_y+\eta}(\sigma_x t_1, \sigma_y t_2)\exp(-\sigma_x \xi t_1 - \sigma_y \eta t_2), \quad (22)$$

we need only find the moment generating function of the standardized distribution.

The joint density of $(U, V) = (X'/\sigma_x + \xi, Y'/\sigma_y + \eta)$ is

$$h(u, v) = \frac{\dfrac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\dfrac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right]}{R\left(\xi, \xi + \dfrac{R_x}{\sigma_x}, \eta, \eta + \dfrac{R_y}{\sigma_y}; \rho\right)}, \quad \begin{array}{l} \xi < u < \xi + R_x/\sigma_x \\ \eta < v < \eta + R_y/\sigma_y \end{array}$$

$$= 0 \quad , \quad \text{elsewhere}. \quad (23)$$

Thus,

$$M_{U,V}(t_1, t_2) = \frac{\displaystyle\int_{\eta}^{\eta+R_y/\sigma_y} \int_{\xi}^{\xi+R_x/\sigma_x} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp(t_1 u + t_2 v) \exp\left[-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right] du\,dv}{R\left(\xi, \xi + \dfrac{R_x}{\sigma_x}, \eta, \eta + \dfrac{R_y}{\sigma_y}; \rho\right)}. \quad (24)$$

However,

$$u^2 - 2\rho uv + v^2 - 2(1 - \rho^2)(t_1 u + t_2 v)$$

$$= u^2 - 2(1 - \rho^2)t_1 u - 2\rho uv + v^2 - 2(1 - \rho^2)t_2 v$$

$$= u^2 - 2(t_1 + \rho t_2)u + 2\rho(t_2 + \rho t_1)u - 2\rho uv + v^2 - 2(t_2 + \rho t_1)v + 2\rho(t_1 + \rho t_2)v$$

$$= u^2 - 2(t_1 + \rho t_2)u - 2\rho[uv - (t_2 + \rho t_1)u - (t_1 + \rho t_2)v] + v^2 - 2(t_2 + \rho t_1)v$$

$$= u^2 - 2(t_1 + \rho t_2)u + (t_1 + \rho t_2)^2 - (t_1 + \rho t_2)^2$$

$$- 2\rho[uv - (t_2 + \rho t_1)u - (t_1 + \rho t_2)v + (t_2 + \rho t_1)(t_1 + \rho t_2) - (t_2 + \rho t_1)(t_1 + \rho t_2)]$$

$$+ v^2 - 2(t_2 + \rho t_1)v + (t_2 + \rho t_1)^2 - (t_2 + \rho t_1)^2$$

$$= [u - (t_1 + \rho t_2)]^2 - 2\rho[u - (t_1 + \rho t_2)][v - (t_2 + \rho t_1)]$$

$$+ [v - (t_2 + \rho t_1)]^2 - (1 - \rho^2)(t_1^2 + 2\rho t_1 t_2 + t_2^2) .$$

So the double integral in the numerator of (24) becomes

$$\exp[1/2(t_1^2 + 2\rho t_1 t_2 + t_2^2)] \int_{\eta}^{\eta + R_y/\sigma_y} \int_{\xi}^{\xi + R_x/\sigma_x} \frac{1}{2\pi\sqrt{1-\rho^2}} \tag{25}$$

$$* \exp\left\{- \frac{1}{2(1-\rho^2)}\left[(u - t_1 - \rho t_2)^2 - 2\rho(u - t_1 - \rho t_2)(v - t_2 - \rho t_1) + (v - t_2 - \rho t_1)^2\right]\right\} du dv.$$

By making the transformation

$$z = \frac{(u - t_1 - \rho t_2) - \rho(v - t_2 - \rho t_1)}{\sqrt{1-\rho^2}}$$

and integrating iteratively with respect to z and then $(v - t_2 - \rho t_1)$, the double integral in (25) reduces to

$$\int_{\eta}^{\eta + R_y/\sigma_y} \phi(v - t_2 - \rho t_1)\left[\Phi\left(\frac{\xi + \frac{R_x}{\sigma_x} - \rho v - (1-\rho^2)t_1}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\xi - \rho v - (1-\rho^2)t_1}{\sqrt{1-\rho^2}}\right)\right] dv \tag{26}$$

$$* R\left(\xi - t_1 - \rho t_2,\ \xi + \frac{R_x}{\sigma_x} - t_1 - \rho t_2,\ \eta - t_2 - \rho t_1,\ \eta + \frac{R_y}{\sigma_y} - t_2 - \rho t_1;\ \rho\right).$$

Hence, from (22), (24), (25), and (26)

$$M_{X',Y'}(t_1, t_2) = \exp[-\sigma_x\xi t_1 - \sigma_y\eta t_2 + \tfrac{1}{2}(\sigma_x^2 t_1^2 + 2\rho\sigma_x\sigma_y t_1 t_2 + \sigma_y^2 t_2^2)] \tag{27}$$

$$\times \frac{R\left(\xi-\sigma_x t_1-\rho\sigma_y t_2,\ \xi+\frac{R_x}{\sigma_x}-\sigma_x t_1-\rho\sigma_y t_2,\ \eta-\sigma_y t_2-\rho\sigma_x t_1,\ \eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1;\rho\right)}{R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\ \rho\right)},$$

where

$$R\left(\xi-\sigma_x t_1-\rho\sigma_y t_2,\ \xi+\frac{R_x}{\sigma_x}-\sigma_x t_1-\rho\sigma_y t_2,\ \eta-\sigma_y t_2-\rho\sigma_x t_1,\ \eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1;\rho\rho\right)$$

$$= \int_{\eta-\sigma_y t_2-\rho\sigma_x t_1}^{\eta+R_y/\sigma_y-\sigma_y t_2-\rho\sigma_x t_1} \phi(t)\left[\Phi\left(\frac{\xi+\frac{R_x}{\sigma_x}-\sigma_x t_1-\rho\sigma_y t_2-\rho t}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{\xi-\sigma_x t_1-\rho\sigma_y t_2-\rho t}{\sqrt{1-\rho^2}}\right)\right]dt. \tag{28}$$

The population moments about the origin may be found from the relation

$$\mu'_{r,s} = \int_0^{R_y}\int_0^{R_x} x'^r y'^s f(x', y')\,dx'dy' = \left.\frac{\partial^{r+s}M_{X',Y'}(t_1, t_2)}{\partial t_1^r \partial t_2^s}\right|_{t_1=t_2=0}. \tag{29}$$

The following results will be needed.

$$\frac{\partial}{\partial t_1}\left(\int_{\eta-\sigma_y t_2-\rho\sigma_x t_1}^{\eta+R_y/\sigma_y-\sigma_y t_2-\rho\sigma_x t_1} \phi(t)\,\Phi\left(\frac{\gamma-\sigma_x t_1-\rho\sigma_y t_2-\rho t}{\sqrt{1-\rho^2}}\right)dt\right)$$
$$= -\rho\sigma_x\phi\left(\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1\right)\Phi\left(\frac{\gamma-\rho\left(\eta+\frac{R_y}{\sigma_y}\right)-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)$$

$$+\rho\sigma_x\phi(\eta-\sigma_y t_2-\rho\sigma_x t_1)\Phi\left(\frac{\gamma-\rho\eta-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)-\sigma_x\phi(\gamma-\sigma_x t_1-\rho\sigma_y t_2)$$

$$\times\left[\Phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{\eta-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)\right]; \tag{30}$$

$$\frac{\partial}{\partial t_2}\left(\int_{\eta-\sigma_y t_2-\rho\sigma_x t_1}^{\eta+R_y/\sigma_y-\sigma_y t_2-\rho\sigma_x t_1}\phi(t)\,\Phi\left(\frac{\gamma-\sigma_x t_1-\rho\sigma_y t_2-\rho t}{\sqrt{1-\rho^2}}\right)dt\right)$$

$$=-\sigma_y\phi\left(\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1\right)\Phi\left(\frac{\gamma-\rho\left(\eta+\frac{R_y}{\sigma_y}\right)-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)$$

$$+\sigma_y\phi(\eta-\sigma_y t_2-\rho\sigma_x t_1)\Phi\left(\frac{\gamma-\rho\eta-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)-\rho\sigma_y\phi(\gamma-\sigma_x t_1-\rho\sigma_y t_2)$$

$$\times\left[\Phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{\eta-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)\right]; \tag{31}$$

$$\frac{\partial^2}{\partial t_1^2}\left(\int_{\eta-\sigma_y t_2-\rho\sigma_x t_1}^{\eta+R_y/\sigma_y-\sigma_y t_2-\rho\sigma_x t_1}\phi(t)\,\Phi\left(\frac{\gamma-\sigma_x t_1-\rho\sigma_y t_2-\rho t}{\sqrt{1-\rho^2}}\right)dt\right)$$

$$=-\rho^2\sigma_x^2\left(\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1\right)\phi\left(\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1\right)\Phi\left(\frac{\gamma-\rho\left(\eta+\frac{R_y}{\sigma_y}\right)-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)$$

$$+\rho^2\sigma_x^2(\eta-\sigma_y t_2-\rho\sigma_x t_1)\phi(\eta-\sigma_y t_2-\rho\sigma_x t_1)\Phi\left(\frac{\gamma-\rho\eta-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)+\sigma_x^2(\gamma-\sigma_x t_1-\rho\sigma_y t_2)$$

$$\times\phi(\gamma-\sigma_x t_1-\rho\sigma_y t_2)\left[\Phi\left(\frac{\eta-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)\right]$$

$$+\ \rho\sqrt{1-\rho^2}\ \sigma_x^2\phi\left(\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1\right)\phi\left(\frac{\gamma-\rho\left(\eta+\frac{R_y}{\sigma_y}\right)-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)$$

$$-\ \rho\sqrt{1-\rho^2}\ \sigma_x^2\phi(\eta-\sigma_y t_2-\rho\sigma_x t_1)\phi\left(\frac{\gamma-\rho\eta-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)\ ; \tag{32}$$

$$\frac{\partial^2}{\partial t_2^2}\left(\int_{\eta-\sigma_y t_2-\rho\sigma_x t_1}^{\eta+R_y/\sigma_y-\sigma_y t_2-\rho\sigma_x t_1}\phi(t)\Phi\left(\frac{\gamma-\sigma_x t_1-\rho\sigma_y t_2-\rho t}{\sqrt{1-\rho^2}}\right)dt\right)$$

$$= -\sigma_y^2\left(\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1\right)\phi\left(\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1\right)\Phi\left(\frac{\gamma-\rho\left(\eta+\frac{R_y}{\sigma_y}\right)-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)$$

$$+\sigma_y^2(\eta-\sigma_y t_2-\rho\sigma_x t_1)\phi(\eta-\sigma_y t_2-\rho\sigma_x t_1)\Phi\left(\frac{\gamma-\rho\eta-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)+\ \rho^2\sigma_y^2(\gamma-\sigma_x t_1-\rho\sigma_y t_2)$$

$$\times\ \phi(\gamma-\sigma_x t_1-\rho\sigma_y t_2)\left[\Phi\left(\frac{\eta-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)\right]$$

$$-\ \rho\sqrt{1-\rho^2}\ \sigma_y^2\phi(\gamma-\sigma_x t_1-\rho\sigma_y t_2)\left[\phi\left(\frac{\eta-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)-\phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)\right]; \tag{33}$$

$$\frac{\partial^2}{\partial t_1\partial t_2}\left(\int_{\eta-\sigma_y t_2-\rho\sigma_x t_1}^{\eta+R_y/\sigma_y-\sigma_y t_2-\rho\sigma_x t_1}\phi(t)\Phi\left(\frac{\gamma-\sigma_x t_1-\rho\sigma_y t_2-\rho t}{\sqrt{1-\rho^2}}\right)dt\right)$$

$$= -\rho\sigma_x\sigma_y\left(\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1\right)\phi\left(\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1\right)\Phi\left(\frac{\gamma-\rho\left(\eta+\frac{R_y}{\sigma_y}\right)-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)$$

$$+\rho\sigma_x\sigma_y(\eta-\sigma_y t_2-\rho\sigma_x t_1)\phi(\eta-\sigma_y t_2-\rho\sigma_x t_1)\Phi\left(\frac{\gamma-\rho\eta-(1-\rho^2)\sigma_x t_1}{\sqrt{1-\rho^2}}\right)+\rho\sigma_x\sigma_y(\gamma-\sigma_x t_1-\rho\sigma_y t_2)$$

$$\times\ \phi(\gamma-\sigma_x t_1-\rho\sigma_y t_2)\left[\Phi\left(\frac{\eta-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)-\Phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)\right]$$

$$-\sqrt{1-\rho^2}\ \sigma_x\sigma_y\phi(\gamma-\sigma_x t_1-\rho\sigma_y t_2)\left[\phi\left(\frac{\eta-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)-\phi\left(\frac{\eta+\frac{R_y}{\sigma_y}-\rho\gamma-(1-\rho^2)\sigma_y t_2}{\sqrt{1-\rho^2}}\right)\right]. \tag{34}$$

Differentiating (27) partially with respect to t_1 yields

$$\begin{aligned}\frac{\partial M_{X',Y'}(t_1,t_2)}{\partial t_1} &= (-\sigma_x\xi+\sigma_x^2 t_1+\rho\sigma_x\sigma_y t_2)M_{X',Y'}(t_1,t_2)\\ &+\frac{\exp\left[-\sigma_x\xi t_1-\sigma_y\eta t_2+1/2\left(\sigma_x^2t_1^2+2\rho\sigma_x\sigma_y t_1t_2+\sigma_y^2t_2^2\right)\right]}{R\left(\xi,\ \xi+\frac{R_x}{\sigma_x},\ \eta,\ \eta+\frac{R_y}{\sigma_y};\ \rho\right)}\\ &\times\frac{\partial}{\partial t_1}R\left(\xi-\sigma_x t_1-\rho\sigma_y t_2,\xi+\frac{R_x}{\sigma_x}-\sigma_x t_1-\rho\sigma_y t_2,\eta-\sigma_y t_2-\rho\sigma_x t_1,\right.\\ &\left.\eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1;\ \rho\right).\end{aligned} \tag{35}$$

Using (28), (30), and (29), the first moment about the origin of X' is given by

$$\mu'_{1,0}=\left.\frac{\partial M_{X',Y'}(t_1,t_2)}{\partial t_1}\right|_{t_1=t_2=0}=-\sigma_x\xi+\sigma_x G_1-\sigma_x G_2+\rho\sigma_x G_3-\rho\sigma_x G_4\ ;$$

which we write as

$$\mu'_{1,0}=-\sigma_x\xi+\sigma_x Q_{1,0}\ , \tag{36}$$

where

$$Q_{1,0}=G_1-G_2+\rho(G_3-G_4)\ . \tag{37}$$

Differentiating (27) partially with respect to t_2 yields

$$\frac{\partial M_{X',Y'}(t_1\ ,\ t_2)}{\partial t_2} = (-\sigma_y\eta+\rho\sigma_x\sigma_y t_1+\sigma_y^2 t_2)M_{X',Y'}(t_1\ ,\ t_2)$$

$$+\ \frac{\exp\left[-\sigma_x\xi t_1-\sigma_y\eta t_2 + 1/2\left(\sigma_x^2t_1^2+2\rho\sigma_x\sigma_y t_1t_2+\sigma_y^2t_2^2\right)\right]}{R\left(\xi,\ \xi + \frac{R_x}{\sigma_x}\ ,\ \eta,\ \eta + \frac{R_y}{\sigma_y}\ ;\ \rho\right)}$$

$$\times\ \frac{\partial}{\partial t_2}\ R\Bigg(\xi-\sigma_x t_1-\rho\sigma_y t_2, \xi+\frac{R_x}{\sigma_x} -\sigma_x t_1-\rho\sigma_y t_2, \eta-\sigma_y t_2-\rho\sigma_x t_1,$$

$$\eta+\frac{R_y}{\sigma_y} -\sigma_y t_2-\rho\sigma_x t_1;\ \rho\Bigg)\ . \tag{38}$$

Using (28), (31), and (29), the first moment about the origin of Y' is given by

$$\mu'_{0,1} = \left.\frac{\partial M_{X',Y'}(t_1\ ,\ t_2)}{\partial t_2}\right|_{t_1=t_2=0} = -\sigma_y\eta+\rho\sigma_y G_1-\rho\sigma_y G_2+\sigma_y G_3-\sigma_y G_4\ ;$$

which we write as

$$\mu'_{0,1} = -\ \sigma_y\eta\ +\ \sigma_y Q_{0,1}\ , \tag{39}$$

where

$$Q_{0,1} = \rho\,(G_1\ -\ G_2)\ +\ G_3\ -\ G_4\ . \tag{40}$$

Differentiating (35) partially with respect to t_1 yields

$$\frac{\partial^2 M_{X',Y'}(t_1\ ,\ t_2)}{\partial t_1^2} = [\sigma_x^2\ -\ (-\sigma_x\xi+\sigma_x^2 t_1+\rho\sigma_x\sigma_y t_2)^2]M_{X',Y'}(t_1\ ,\ t_2)$$

$$+\ 2(-\sigma_x\xi+\sigma_x^2 t_1+\rho\sigma_x\sigma_y t_2)\ \frac{\partial M_{X',Y'}(t_1\ ,\ t_2)}{\partial t_1}$$

$$+ \frac{\exp\left[-\sigma_x \xi t_1 - \sigma_y \eta t_2 + 1/2\left(\sigma_x^2 t_1^2 + 2\rho\sigma_x\sigma_y t_1 t_2 + \sigma_y^2 t_2^2\right)\right]}{R\left(\xi,\ \xi + \frac{R_x}{\sigma_x},\ \eta,\ \eta + \frac{R_y}{\sigma_y}\ ;\ \rho\right)}$$

$$\times \frac{\partial^2}{\partial t_1^2} R\left(\xi - \sigma_x t_1 - \rho\sigma_y t_2, \xi + \frac{R_x}{\sigma_x} - \sigma_x t_1 - \rho\sigma_y t_2, \eta - \sigma_y t_2 - \rho\sigma_x t_1,\right.$$

$$\left.\eta + \frac{R_y}{\sigma_y} - \sigma_y t_2 - \rho\sigma_x t_1 ; \rho\right) . \tag{41}$$

Using (28), (32), and (29), the second moment about the origin of X' is given by

$$\mu'_{2,0} = \left.\frac{\partial^2 M_{X',Y'}(t_1\ ,\ t_2)}{\partial t_2^2}\right|_{t_1=t_2=0} = \sigma_x^2 - \sigma_x^2\xi^2 - 2\sigma_x\xi(-\sigma_x\xi + \sigma_x Q_{1,0})$$
$$+ \sigma_x^2\xi G_1 - \sigma_x^2\left(\xi + \frac{R_x}{\sigma_x}\right)G_2 + \rho^2\sigma_x^2\eta G_3$$
$$- \rho^2\sigma_x^2\left(\eta + \frac{R_y}{\sigma_y}\right)G_4 + \rho\sigma_x^2 g_5 \ ;$$

which we write as

$$\mu'_{2,0} = \sigma_x^2 + \sigma_x^2\xi^2 - 2\sigma_x^2\xi Q_{1,0} + \sigma_x^2 Q_{2,0}\ , \tag{42}$$

where

$$Q_{2,0} = \xi(G_1 - G_2) + \rho^2\eta(G_3 - G_4) - \frac{R_x}{\sigma_x}G_2 - \rho^2\frac{R_y}{\sigma_y}G_4 + \rho g_5\ . \tag{43}$$

Differentiating (38) partially with respect to t_2 yields

$$\frac{\partial^2 M_{X',Y'}(t_1\ ,\ t_2)}{\partial t_2^2} = [\sigma_y^2 - (-\sigma_y\eta + \rho\sigma_x\sigma_y t_1 + \sigma_y^2 t_2)^2]M_{X',Y'}(t_1\ ,\ t_2)$$

$$+ 2(-\sigma_y\eta + \rho\sigma_x\sigma_y t_1 + \sigma_y^2 t_2)\ \frac{\partial M_{X',Y'}(t_1\ ,\ t_2)}{\partial t_2}$$

$$+ \frac{\exp\left[-\sigma_x \xi t_1 - \sigma_y \eta t_2 + 1/2\left(\sigma_x^2 t_1^2 + 2\rho\sigma_x\sigma_y t_1 t_2 + \sigma_y^2 t_2^2\right)\right]}{R\left(\xi,\ \xi + \frac{R_x}{\sigma_x},\ \eta,\ \eta + \frac{R_y}{\sigma_y}\ ;\ \rho\right)}$$

$$\times \frac{\partial^2}{\partial t_2^2} R\left(\xi-\sigma_x t_1-\rho\sigma_y t_2, \xi+ \frac{R_x}{\sigma_x} -\sigma_x t_1-\rho\sigma_y t_2, \eta-\sigma_y t_2-\rho\sigma_x t_1,\right.$$

$$\left.\eta+ \frac{R_y}{\sigma_y} -\sigma_y t_2-\rho\sigma_x t_1 ; \rho\right) . \tag{44}$$

Using (28), (33), and (29), the second moment about the origin of Y' is given by

$$\begin{aligned}\mu'_{0,2} = \left.\frac{\partial^2 M_{X',Y'}(t_1, t_2)}{\partial t_2^2}\right|_{t_1=t_2=0} &= \sigma_y^2 - \sigma_y^2\eta^2 - 2\sigma_y\eta(-\sigma_y\eta + \sigma_y Q_{0,1}) \\ &+ \rho^2\sigma_y^2\xi G_1 - \rho^2\sigma_y^2\left(\xi + \frac{R_x}{\sigma_x}\right)G_2 + \sigma_y^2\eta G_3 \\ &- \sigma_y^2\left(\eta + \frac{R_y}{\sigma_y}\right)G_4 + \rho\sigma_y^2 g_5 ;\end{aligned}$$

which we write as

$$\mu'_{0,2} = \sigma_y^2 + \sigma_y^2\eta^2 - 2\sigma_y^2\eta Q_{0,1} + \sigma_y^2 Q_{0,2} , \tag{45}$$

where

$$Q_{0,2} = \rho^2\xi(G_1 - G_2) + \eta(G_3 - G_4) - \rho^2 \frac{R_x}{\sigma_x} G_2 - \frac{R_y}{\sigma_y} G_4 + \rho g_5 . \tag{46}$$

Differentiating (35) partially with respect to t_2 yields

$$\begin{aligned}\frac{\partial^2 M_{X',Y'}(t_1, t_2)}{\partial t_1 \partial t_2} &= [\rho\sigma_x\sigma_y + (-\sigma_x\xi+\sigma_x^2 t_1+\rho\sigma_x\sigma_y t_2)(-\sigma_y\eta+\rho\sigma_x\sigma_y t_1+\sigma_y^2 t_2)]M_{X',Y'}(t_1, t_2) \\ &+ (-\sigma_x\xi+\sigma_x^2 t_1+\rho\sigma_x\sigma_y t_2) \frac{\partial M_{X',Y'}(t_1, t_2)}{\partial t_2} \\ &+ (-\sigma_y\eta+\rho\sigma_x\sigma_y t_1+\sigma_y^2 t_2) \frac{\partial M_{X',Y'}(t_1, t_2)}{\partial t_1} \\ &+ \frac{\exp\left[-\sigma_x\xi t_1-\sigma_y\eta t_2 + 1/2\left(\sigma_x^2 t_1^2+2\rho\sigma_x\sigma_y t_1 t_2+\sigma_y^2 t_2^2\right)\right]}{R\left(\xi, \xi + \frac{R_x}{\sigma_x}, \eta, \eta + \frac{R_y}{\sigma_y} ; \rho\right)}\end{aligned}$$

$$\times \frac{\partial^2}{\partial t_1 \partial t_2} R\left(\xi-\sigma_x t_1-\rho\sigma_y t_2, \xi+\frac{R_x}{\sigma_x}-\sigma_x t_1-\rho\sigma_y t_2,\right.$$

$$\left.\eta-\sigma_y t_2-\rho\sigma_x t_1, \eta+\frac{R_y}{\sigma_y}-\sigma_y t_2-\rho\sigma_x t_1;\ \rho\right) . \tag{47}$$

Using (28), (34), and (29), the joint moment about the origin is given by

$$\mu'_{1,1} = \left.\frac{\partial^2 M_{X',Y'}(t_1\ ,\ t_2)}{\partial t_1 \partial t_2}\right|_{t_1=t_2=0} = \rho\sigma_x\sigma_y - \sigma_x\sigma_y\xi\eta - \sigma_x\xi(-\sigma_y\eta + \sigma_y Q_{0,1})$$
$$- \sigma_y\eta(-\sigma_x\xi + \sigma_x Q_{1,0}) + \rho\sigma_x\sigma_y\xi G_1$$
$$- \rho\sigma_x\sigma_y\left(\xi + \frac{R_x}{\sigma_x}\right)G_2 + \rho\sigma_x\sigma_y\eta G_3$$
$$- \rho\sigma_x\sigma_y\left(\eta + \frac{R_y}{\sigma_y}\right)G_4 + \sigma_x\sigma_y g_5\ ;$$

which we write as

$$\mu'_{1,1} = \rho\sigma_x\sigma_y + \sigma_x\sigma_y\xi\eta - \sigma_x\sigma_y\xi Q_{0,1} - \sigma_x\sigma_y\eta Q_{1,0} + \sigma_x\sigma_y Q_{1,1}\ ; \tag{48}$$

where

$$Q_{1,1} = \rho\xi(G_1 - G_2) + \rho\eta(G_3 - G_4) - \rho\frac{R_x}{\sigma_x}G_2 - \rho\frac{R_y}{\sigma_y}G_4 + g_5\ . \tag{49}$$

Finally, we obtain the method of moments estimating equations by equating population moments to corresponding sample moments.

$$\mu'_{1,0} = -\sigma_x\xi + \sigma_x Q_{1,0} = m_{1,0}\ ; \tag{50}$$

$$\mu'_{0,1} = -\sigma_y\eta + \sigma_y Q_{0,1} = m_{0,1}\ ; \tag{51}$$

$$\mu'_{2,0} = \sigma_x^2 + \sigma_x^2\xi^2 - 2\sigma_x^2\xi Q_{1,0} + \sigma_x^2 Q_{2,0} = m_{2,0}\ ; \tag{52}$$

$$\mu'_{0,2} = \sigma_y^2 + \sigma_y^2\eta^2 - 2\sigma_y^2\eta Q_{0,1} + \sigma_y^2 Q_{0,2} = m_{0,2}\ ; \tag{53}$$

$$\mu'_{1,1} = \rho\sigma_x\sigma_y + \sigma_x\sigma_y\xi\eta - \sigma_x\sigma_y\xi Q_{0,1} - \sigma_x\sigma_y\eta Q_{1,0} + \sigma_x\sigma_y Q_{1,1} = m_{1,1}\ . \tag{54}$$

That the two systems of nonlinear equations given by the method of maximum likelihood and the methods of moments are equivalent is seen by recalling that the maximum likelihood equations have a unique solution and noting that from (6)-(10) and (50)-(54)

$$\begin{pmatrix} \frac{\partial \log L}{\partial \xi} \\ \frac{\partial \log L}{\partial \eta} \\ \frac{\partial \log L}{\partial \sigma_x} \\ \frac{\partial \log L}{\partial \sigma_y} \\ \frac{\partial \log L}{\partial \rho} \end{pmatrix} = \begin{pmatrix} \frac{N}{\sigma_x(1-\rho^2)} & \frac{-N\rho}{\sigma_y(1-\rho^2)} & 0 & 0 & 0 \\ \frac{-N\rho}{\sigma_x(1-\rho^2)} & \frac{N}{\sigma_y(1-\rho^2)} & 0 & 0 & 0 \\ \frac{-N(\xi-\rho\eta)}{\sigma_x^2(1-\rho^2)} & 0 & \frac{-N}{\sigma_x^3(1-\rho^2)} & 0 & \frac{N\rho}{\sigma_x^2\sigma_y(1-\rho^2)} \\ 0 & \frac{-N(\eta-\rho\xi)}{\sigma_y^2(1-\rho^2)} & 0 & \frac{-N}{\sigma_y^3(1-\rho^2)} & \frac{N\rho}{\sigma_x\sigma_y^2(1-\rho^2)} \\ \frac{-N[(\eta-\rho\xi)-\rho(\xi-\rho\eta)]}{\sigma_x(1-\rho^2)^2} & \frac{-N[(\xi-\rho\eta)-\rho(\eta-\rho\xi)]}{\sigma_y(1-\rho^2)^2} & \frac{N\rho}{\sigma_x^2(1-\rho^2)^2} & \frac{N\rho}{\sigma_y^2(1-\rho^2)^2} & \frac{-N(1+\rho^2)}{\sigma_x\sigma_y(1-\rho^2)^2} \end{pmatrix} \begin{pmatrix} \mu'_{1,0}-m_{1,0} \\ \mu'_{0,1}-m_{0,1} \\ \mu'_{2,0}-m_{2,0} \\ \mu'_{0,2}-m_{0,2} \\ \mu'_{1,1}-m_{1,1} \end{pmatrix}, \tag{55}$$

where the determinant of the square matrix in (55) has a value of $-N^5/\sigma_x^5\sigma_y^5(1-\rho^2)^4 \neq 0$.

CHAPTER III

SOLUTION TO THE ESTIMATING EQUATIONS

1. Proposed Numerical Procedure

The system of nonlinear equations (50)-(54) given by the method of moments (or the equivalent system (6)-(10) given by the method of maximum likelihood) can be solved numerically by either the Newton-Raphson method or the functional iterative method. Both methods are described with conditions for convergence in Isaacson and Keller (1966, pp. 109-122).

The Newton-Raphson method is quite laborious and cumbersome for the method of moments system (or the maximum likelihood system) since at each cycle of the approximation, the first partial derivatives of the $\mu'_{i,j}$ (or the second partial derivatives of the logarithm of the likelihood function) with respect to ξ, η, σ_x, σ_y, and ρ must be evaluated at the approximate solution given by the preceding cycle. Furthermore, since squares, products, and higher powers of the corrections are neglected in the Taylor series expansions of the estimating equations about a vector near the solution, the Newton-Raphson method, if it converges to the solution, tends to do so slowly during the first few cycles when the starting vector [i.e., the initial approximation of $(\hat{\xi}, \hat{\eta}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho})$] is distant from the solution. If the starting vector is in a close neighborhood of the solution, then convergence to the solution is generally assured and is rapid, usually of quadratic order. Obviously, this "closeness" is often impossible to realize in practice.

If the functional iterative method converges to the solution, it tends to do so in an opposite fashion from that of the Newton-Raphson method. The functional iterative method advances more rapidly toward a neighborhood of the solution during the first few cycles than does the Newton-Raphson method, but then advances much slower than does the Newton-Raphson method as the solution is approached. In this method, convergence generally depends on the functional forms of the iterants as well as the starting vector. However, at each cycle of the approximation to the solution only the $\mu'_{i,j}$ (or the first partial derivatives of the logarithm of the likelihood function with respect to ξ, η, σ_x, σ_y, and ρ) must be evaluated.

Because of the computational requirement at each cycle and the convergence behavior of these two methods, the following numerical procedure to find the maximum likelihood estimates is proposed. Through the use of a suitably chosen starting vector as an initial estimate of the solution, we advance to a close neighborhood of the solution with one or more cycles of the functional iterative method, then obtain the final estimates with a single cycle of the Newton-Raphson method. Of great importance is the fact that by using the Newton-Raphson method for the final cycle we not only increase the accuracy of the estimates obtained from the last cycle of the functional iterative method, but also without any additional computational work we obtain an estimate of the asymptotic variance-covariance matrix of the maximum likelihood estimates.

2. Starting Vector

Rather than just choose any starting vector, we look for one which might possibly reduce the number of cycles required by the functional iterative method to reach a close neighborhood of the solution. It is well known that under regularity conditions the joint maximum likelihood estimates converge in probability as a set to the true set of parameter values, i.e., the joint maximum likelihood estimates are consistent. If we consider any other set t_1, t_2, t_3, t_4, t_5 of consistent estimates, then in large samples such estimates will tend to be fairly close to the maximum likelihood estimates.

Wilks (1962, pp. 380-381) has defined the asymptotic efficiency of any set T of consistent estimates having a limiting multivariate normal distribution as the ratio of the asymptotic generalized variance of the maximum likelihood estimates to the asymptotic generalized variance of the set T. Obviously, the maximum likelihood estimates have asymptotic efficiency unity. Furthermore, Geary (1942) has shown that, asymptotically, the joint maximum likelihood estimates minimize the generalized variance of any set of consistent estimates. Hence, any set of consistent estimates having a limiting multivariate normal distribution which differ from the maximum likelihood estimates will have an asymptotic efficiency ratio less than one and will be termed as asymptotically inefficient estimates.

Therefore, we seek a set of asymptotically jointly normal consistent estimates of ξ, η, σ_x, σ_y, and ρ which can be computed without a great deal of effort from the sample data. Most likely, this set of estimates will be asymptotically inefficient. But of course, the higher the asymptotic efficiency ratio, the closer this set of estimates will tend to be to the maximum likelihood estimates in large samples. That such a set of estimates can be found has been suggested by Khatri and Jaiswal (1963).

We first establish recurrence relations for the moments about the origin of $f(x', y')$. Taking the first partial derivatives of $f(x', y')$ with respect to x' and y' yields

$$\frac{\partial f(x',\, y')}{\partial x'} = \left[-\frac{1}{\sigma_x^2(1-\rho^2)}\, x' + \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)}\, y' - \frac{\xi-\rho\eta}{\sigma_x(1-\rho^2)}\right] f(x',\, y') \tag{56}$$

and

$$\frac{\partial f(x',\, y')}{\partial y'} = \left[\frac{\rho}{\sigma_x\sigma_y(1-\rho^2)}\, x' - \frac{1}{\sigma_y^2(1-\rho^2)}\, y' - \frac{\eta-\rho\xi}{\sigma_y(1-\rho^2)}\right] f(x',\, y') \tag{57}$$

Upon multiplying both sides of (56) by $x'^r y'^s$, integrating, and using (29) we obtain

$$\int_0^{R_y}\int_0^{R_x} x'^r y'^s \,\frac{\partial f(x',\, y')}{\partial x'}\, dx'dy' = R^r \frac{1}{\sigma_x^2(1-\rho^2)}\, \mu'_{r+1,s} + \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)}\, \mu'_{r,s+1}$$

$$- \frac{\xi-\rho\eta}{\sigma_x(1-\rho^2)}\, \mu'_{r,s} \quad . \tag{58}$$

Integrating the left-hand side of (58) by parts and assuming $r \geq 1$, we find that

$$\int_0^{R_y}\int_0^{R_x} x'^r y'^s \,\frac{\partial f(x',\, y')}{\partial x'}\, dx'dy' = R_x^r \int_0^{R_y} y'^s f(R_x\, ,\, y')dy' - r\mu'_{r-1,s} \quad . \tag{59}$$

Since

$$\int_0^{R_y} y'^s f(R_x\, ,\, y')dy' = \frac{\dfrac{\sigma_y^s}{\sigma_x}\,\phi\!\left(\xi+\dfrac{R_x}{\sigma_x}\right)\displaystyle\int_{\frac{\eta-\rho\left(\xi+\frac{R_x}{\sigma_x}\right)}{\sqrt{1-\rho^2}}}^{\frac{\eta+\frac{R_y}{\sigma_y}-\rho\left(\xi+\frac{R_x}{\sigma_x}\right)}{\sqrt{1-\rho^2}}} \left[\sqrt{1-\rho^2}\; z-\eta+\rho\left(\xi+\dfrac{R_x}{\sigma_x}\right)\right]^s \phi(z)\,dz}{R\left(\xi,\ \xi+\dfrac{R_x}{\sigma_x}\, ,\ \eta,\ \eta+\dfrac{R_y}{\sigma_y}\ ;\ \rho\right)} \ ,$$

(60)

then (58) may be written as

$$r\mu'_{r-1,s} = \frac{R_x^r \left[\dfrac{\sigma_y^s}{\sigma_x} \phi\left(\xi + \dfrac{R_x}{\sigma_x}\right) \displaystyle\int_{\frac{\eta - \rho\left(\xi + \frac{R_x}{\sigma_x}\right)}{\sqrt{1-\rho^2}}}^{\frac{\eta + \frac{R_y}{\sigma_y} - \rho\left(\xi + \frac{R_x}{\sigma_x}\right)}{\sqrt{1-\rho^2}}} \left[\sqrt{1-\rho^2}\, z - \eta + \rho\left(\xi + \dfrac{R_x}{\sigma_x}\right)\right]^s \phi(z)\,dz \right]}{R\left(\xi,\ \xi + \dfrac{R_x}{\sigma_x},\ \eta,\ \eta + \dfrac{R_y}{\sigma_y};\ \rho\right)}$$

$$+ \frac{1}{\sigma_x^2(1-\rho^2)}\,\mu'_{r+1,s} - \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)}\,\mu'_{r,s+1} + \frac{\xi-\rho\eta}{\sigma_x(1-\rho^2)}\,\mu'_{r,s}\ ;\ r \geq 1,\ s \geq 0\ . \tag{61}$$

Operating on (57) in a similar fashion, we find that

$$s\mu'_{r,s-1} = \frac{R_y^s \left[\dfrac{\sigma_x^r}{\sigma_y} \phi\left(\eta + \dfrac{R_y}{\sigma_y}\right) \displaystyle\int_{\frac{\xi - \rho\left(\eta + \frac{R_y}{\sigma_y}\right)}{\sqrt{1-\rho^2}}}^{\frac{\xi + \frac{R_x}{\sigma_x} - \rho\left(\eta + \frac{R_y}{\sigma_y}\right)}{\sqrt{1-\rho^2}}} \left[\sqrt{1-\rho^2}\, z - \xi + \rho\left(\eta + \dfrac{R_y}{\sigma_y}\right)\right]^r \phi(z)\,dz \right]}{R\left(\xi,\ \xi + \dfrac{R_x}{\sigma_x},\ \eta,\ \eta + \dfrac{R_y}{\sigma_y};\ \rho\right)}$$

$$- \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)}\,\mu'_{r+1,s} + \frac{1}{\sigma_y^2(1-\rho^2)}\,\mu'_{r,s+1} + \frac{\eta-\rho\xi}{\sigma_y(1-\rho^2)}\,\mu'_{r,s}\ ;\ r \geq 0,\ s \geq 1\ . \tag{62}$$

We shall now obtain a system of estimating equations in a considerably different manner than did Khatri and Jaiswal (1963). Their approach was used to solve a special case of the problem considered in this section; thus their recurrence relations for the moments were simpler. This allowed them to develop four different systems of five linearly independent linear equations in five unknowns (giving four different sets of estimates) from six linearly independent linear equations using eight or nine sample moments. We shall develop a single system of five linearly independent linear equations in five unknowns from ten linearly independent linear equations using thirteen sample moments -- $m_{1,0}$, $m_{0,1}$, $m_{2,0}$, $m_{0,2}$, $m_{1,1}$, $m_{3,0}$, $m_{0,3}$, $m_{1,2}$, $m_{2,1}$, $m_{4,0}$, $m_{0,4}$, $m_{1,3}$, and $m_{3,1}$. With the aid of a desk calculator the computation of the required sample moments is not as tedious as at first it might appear. The disadvantage of having to develop a system of five equations from ten equations will be converted into an extremely important advantage in the final analysis in that, because of the way the equations of the system will be constructed, there will exist several relationships between the elements of the coefficient matrix and the constant vector. Thus, we shall be able to simplify considerably all explicit expressions for the elements of the starting vector, and the actual computation of these estimates will not require much more effort than that needed to compute the Khatri and Jaiswal estimates, though we shall be working with more moments and more terms and solving a more general case.

The following equations are obtained from (61).

$$1 = R_x \frac{G_2}{\sigma_x} + \frac{1}{\sigma_x^2(1-\rho^2)} \mu'_{2,0} - \frac{\rho}{\sigma_x \sigma_y (1-\rho^2)} \mu'_{1,1} + \frac{\xi - \rho\eta}{\sigma_x (1-\rho^2)} \mu'_{1,0} \; ; \tag{63}$$

$$2\mu'_{1,0} = R_x^2 \frac{G_2}{\sigma_x} + \frac{1}{\sigma_x^2(1-\rho^2)} \mu'_{3,0} - \frac{\rho}{\sigma_x \sigma_y (1-\rho^2)} \mu'_{2,1} + \frac{\xi - \rho\eta}{\sigma_x (1-\rho^2)} \mu'_{2,0} \tag{64}$$

$$3\mu'_{2,0} = R_x^3 \frac{G_2}{\sigma_x} + \frac{1}{\sigma_x^2(1-\rho^2)} \mu'_{4,0} - \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)} \mu'_{3,1} + \frac{\xi-\rho\eta}{\sigma_x(1-\rho^2)} \mu'_{3,0} \; ; \tag{65}$$

$$\mu'_{0,1} = R_x\left\{\frac{\sigma_y}{\sigma_x}\left[\sqrt{1-\rho^2}\;(g_3 - g_4) - \left(\eta - \rho\left(\xi + \frac{R_x}{\sigma_x}\right)\right)G_2\right]\right\}$$

$$+ \frac{1}{\sigma_x^2(1-\rho^2)} \mu'_{2,1} - \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)} \mu'_{1,2} + \frac{\xi-\rho\eta}{\sigma_x(1-\rho^2)} \mu'_{1,1} \; ; \tag{66}$$

$$2\mu'_{1,1} = R_x^2\left\{\frac{\sigma_y}{\sigma_x}\left[\sqrt{1-\rho^2}\;(g_3 - g_4) - \left(\eta - \rho\left(\xi + \frac{R_x}{\sigma_x}\right)\right)G_2\right]\right\}$$

$$+ \frac{1}{\sigma_x^2(1-\rho^2)} \mu'_{3,1} - \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)} \mu'_{2,2} + \frac{\xi-\rho\eta}{\sigma_x(1-\rho^2)} \mu'_{2,1} \; . \tag{67}$$

The following equations are obtained from (62).

$$1 = R_y \frac{G_4}{\sigma_y} - \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)} \mu'_{1,1} + \frac{1}{\sigma_y^2(1-\rho^2)} \mu'_{0,2} + \frac{\eta-\rho\xi}{\sigma_y(1-\rho^2)} \mu'_{0,1} \; ; \tag{68}$$

$$2\mu'_{0,1} = R_y^2 \frac{G_4}{\sigma_y} - \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)} \mu'_{1,2} + \frac{1}{\sigma_y^2(1-\rho^2)} \mu'_{0,3} + \frac{\eta-\rho\xi}{\sigma_y(1-\rho^2)} \mu'_{0,2} \; ; \tag{69}$$

$$3\mu'_{0,2} = R_y^3 \frac{G_4}{\sigma_y} - \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)} \mu'_{1,3} + \frac{1}{\sigma_y^2(1-\rho^2)} \mu'_{0,4} + \frac{\eta-\rho\xi}{\sigma_y(1-\rho^2)} \mu'_{0,3} \; ; \tag{70}$$

$$\mu'_{1,0} = R_y\left\{\frac{\sigma_x}{\sigma_y}\left[\sqrt{1-\rho^2}\;(g_2 - g_4) - \left(\xi - \rho\left(\eta + \frac{R_y}{\sigma_y}\right)\right)G_4\right]\right\}$$

$$- \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)} \mu'_{2,1} + \frac{1}{\sigma_x^2(1-\rho^2)} \mu'_{1,2} + \frac{\eta-\rho\xi}{\sigma_y(1-\rho^2)} \mu'_{1,1} \; ; \tag{71}$$

$$2\mu'_{1,1} = R_y^2\left\{\frac{\sigma_x}{\sigma_y}\left[\sqrt{1-\rho^2}\;(g_2 - g_4) - \left(\xi - \rho\left(\eta + \frac{R_y}{\sigma_y}\right)\right)G_4\right]\right\}$$

$$- \frac{\rho}{\sigma_x\sigma_y(1-\rho^2)} \mu'_{2,2} + \frac{1}{\sigma_y^2(1-\rho^2)} \mu'_{1,3} + \frac{\eta-\rho\xi}{\sigma_y(1-\rho^2)} \mu'_{1,2} \; . \tag{72}$$

Upon multiplying (63) by R_x and subtracting from (64), multiplying (63) by R_x^2 and subtracting from (65), performing a similar operation using

R_y with (68), (69) and (70), and multiplying (66) by R_x and (71) by $-R_y$, adding these two resulting equations, then subtracting their sum from the difference of (67) and (72), we obtain the following system of equations using only the transformed equations. In matrix form

$$\begin{pmatrix} \mu'_{2,0}-R_x\mu'_{1,0} & \mu'_{3,0}-R_x\mu'_{2,0} & \mu'_{2,1}-R_x\mu'_{1,1} & 0 & 0 \\ \mu'_{3,0}-R_x^2\mu'_{1,0} & \mu'_{4,0}-R_x^2\mu'_{2,0} & \mu'_{3,1}-R_x^2\mu'_{1,1} & 0 & 0 \\ \mu'_{2,1}-R_x\mu'_{1,1} & \mu'_{3,1}-R_x\mu'_{2,1} & R_y\mu'_{2,1}-R_x\mu'_{1,2} & R_y\mu'_{1,1}-\mu'_{1,2} & R_y\mu'_{1,2}-\mu'_{1,3} \\ 0 & 0 & \mu'_{1,2}-R_y\mu'_{1,1} & \mu'_{0,2}-R_y\mu'_{0,1} & \mu'_{0,3}-R_y\mu'_{0,2} \\ 0 & 0 & \mu'_{1,3}-R_y^2\mu'_{1,1} & \mu'_{0,3}-R_y^2\mu'_{0,1} & \mu'_{0,4}-R_y^2\mu'_{0,2} \end{pmatrix} \begin{pmatrix} \dfrac{\xi-\rho\eta}{\sigma_x(1-\rho^2)} \\ \dfrac{1}{\sigma_x^2(1-\rho^2)} \\ \dfrac{-\rho}{\sigma_x\sigma_y(1-\rho^2)} \\ \dfrac{\eta-\rho\xi}{\sigma_y(1-\rho^2)} \\ \dfrac{1}{\sigma_y^2(1-\rho^2)} \end{pmatrix} = \begin{pmatrix} 2\mu'_{1,0}-R_x \\ 3\mu'_{2,0}-R_x^2 \\ R_y\mu'_{1,0}-R_x\mu'_{0,1} \\ 2\mu'_{0,1}-R_y \\ 3\mu'_{0,2}-R_y^2 \end{pmatrix} \tag{73}$$

The population moments $\mu'_{r,s}$ are replaced by the corresponding sample moments $m_{r,s}$ in (73). Using Cramer's rule to solve for the unknown $\left(\dfrac{-\rho}{\sigma_x\sigma_y(1-\rho^2)}\right)$ we obtain, following a considerable amount of simplification,

$$\widehat{\left(\frac{-\rho}{\sigma_x\sigma_y(1-\rho^2)}\right)} = \frac{\begin{aligned} &\alpha_1\alpha_2(R_y m_{1,0}-R_x m_{0,1}) \\ &\quad + \alpha_2[(2m_{1,0}-R_x)\alpha_3 + (3m_{2,0}-2R_x m_{1,0})\alpha_4] \\ &\quad\quad + \alpha_1[(R_y-2m_{0,1})\alpha_5 + (2R_y m_{0,1}-3m_{0,2})\alpha_6] \end{aligned}}{\begin{aligned} &\alpha_1\alpha_2(R_y m_{2,1}-R_x m_{1,2}) \\ &\quad +\alpha_2[(m_{2,1}-R_x m_{1,1})\alpha_3 + (m_{3,1}-R_x m_{2,1})\alpha_4] \\ &\quad\quad + \alpha_1[(R_y m_{1,1}-m_{1,2})\alpha_5 + (R_y m_{1,2}-m_{1,3})\alpha_6] \end{aligned}} , \tag{74}$$

where

$$\alpha_1 = R_x^2(m_{1,0}m_{3,0}-m_{2,0}^2)+ R_x(m_{2,0}m_{3,0}-m_{1,0}m_{4,0})+(m_{2,0}m_{4,0}-m_{3,0}^2) \quad , \tag{75}$$

$$\alpha_2 = R_y^2(m_{0,1}m_{0,3}-m_{0,2}^2)+ R_y(m_{0,2}m_{0,3}-m_{0,1}m_{0,4})+(m_{0,2}m_{0,4}-m_{0,3}^2) \quad , \tag{76}$$

$$\alpha_3 = R_x^2(m_{2,0}m_{2,1}-m_{1,1}m_{3,0})+ R_x(m_{1,1}m_{4,0}-m_{2,0}m_{3,1})+(m_{3,0}m_{3,1}-m_{2,1}m_{4,0}) \quad , \tag{77}$$

$$\alpha_4 = R_x^2(m_{2,0}m_{1,1}-m_{1,0}m_{2,1})+ R_x(m_{1,0}m_{3,1}-m_{1,1}m_{3,0})+(m_{2,1}m_{3,0}-m_{2,0}m_{3,1}) \quad , \tag{78}$$

$$\alpha_5 = R_y^2(m_{0,2}m_{1,2}-m_{1,1}m_{0,3})+ R_y(m_{1,1}m_{0,4}-m_{0,2}m_{1,3})+(m_{0,3}m_{1,3}-m_{1,2}m_{0,4}) \quad , \tag{79}$$

$$\alpha_6 = R_y^2(m_{0,2}m_{1,1}-m_{0,1}m_{1,2})+ R_y(m_{0,1}m_{1,3}-m_{1,1}m_{0,3})+(m_{1,2}m_{0,3}-m_{0,2}m_{1,3}) \quad . \tag{80}$$

Denoting the right-hand side of (74) by h^* and solving simultaneously the first two equations of the system (73), we obtain

$$\widehat{\left(\frac{\xi-\rho\eta}{\sigma_x(1-\rho^2)}\right)} = \frac{\alpha_7 + \alpha_3 h^*}{\alpha_1} \tag{81}$$

and

$$\widehat{\left(\frac{1}{\sigma_x^2(1-\rho^2)}\right)} = \frac{\alpha_8 + \alpha_4 h^*}{\alpha_1} \quad , \tag{82}$$

where

$$\alpha_7 = R_x^2(m_{3,0}-2m_{1,0}m_{2,0})+ R_x(3m_{2,0}^2-m_{4,0})+(2m_{1,0}m_{4,0}-3m_{2,0}m_{3,0}) \quad , \tag{83}$$

$$\alpha_8 = R_x^2(2m_{1,0}^2-m_{2,0})+ R_x(m_{3,0}-3m_{1,0}m_{2,0})+(3m_{2,0}^2-2m_{1,0}m_{3,0}) \quad . \tag{84}$$

Solving simultaneously the last two equations of the system (73), we obtain

$$\widehat{\left(\frac{\eta-\rho\xi}{\sigma_y(1-\rho^2)}\right)} = \frac{\alpha_9 + \alpha_5 h^*}{\alpha_2} \tag{85}$$

and

$$\widehat{\left(\frac{1}{\sigma_y^2(1-\rho^2)}\right)} = \frac{\alpha_{10} + \alpha_6 h^*}{\alpha_2} \quad , \tag{86}$$

where

$$\alpha_9 = R_y^2(m_{0,3}-2m_{0,1}m_{0,2}) + R_y(3m_{0,2}^2-m_{0,4}) + (2m_{0,1}m_{0,4}-3m_{0,2}m_{0,3}) \quad , \tag{87}$$

$$\alpha_{10} = R_y^2(2m_{0,1}^2-m_{0,2}) + R_y(m_{0,3}-3m_{0,1}m_{0,2}) + (3m_{0,2}^2-2m_{0,1}m_{0,3}) \quad . \tag{88}$$

From (74), (81), (82), (85), and (86) the initial estimates of the parameters which serve as the components of the starting vector are found. In the event $\frac{\alpha_1}{\alpha_8 + \alpha_4 h^*} \le 0$ or(and) $\frac{\alpha_2}{\alpha_{10} + \alpha_6 h^*} \le 0$, then we shall use the truncated sample standard deviation(s) and the truncated sample correlation coefficient as initial estimates of the corresponding untruncated population standard deviation(s) and the untruncated population correlation coefficient. A superscript zero in parenthesis will indicate initial estimates.

$$\rho^{(0)} = \begin{cases} -h^*\sqrt{\dfrac{\alpha_1}{\alpha_8+\alpha_4 h^*}}\sqrt{\dfrac{\alpha_2}{\alpha_{10}+\alpha_6 h^*}} \quad , & \text{if } \dfrac{\alpha_1}{\alpha_8+\alpha_4 h^*} > 0 \text{ and } \dfrac{\alpha_2}{\alpha_{10}+\alpha_6 h^*} > 0 \\[2ex] \dfrac{m_{1,1}-m_{1,0}m_{0,1}}{\sigma_x^{(0)}\sigma_y^{(0)}} \quad , & \text{if } \dfrac{\alpha_1}{\alpha_8+\alpha_4 h^*} \le 0 \text{ or } \dfrac{\alpha_2}{\alpha_{10}+\alpha_6 h^*} \le 0 \ ; \end{cases} \tag{89}$$

$$\sigma_x^{(0)} = \begin{cases} \sqrt{\dfrac{1}{1-\rho^{(0)2}}}\sqrt{\dfrac{\alpha_1}{\alpha_8+\alpha_4 h^*}} \quad , & \text{if } \dfrac{\alpha_1}{\alpha_8+\alpha_4 h^*} > 0 \\[2ex] \sqrt{m_{2,0}-m_{1,0}^2} \quad , & \text{if } \dfrac{\alpha_1}{\alpha_8+\alpha_4 h^*} \le 0 \ ; \end{cases} \tag{90}$$

$$\sigma_y^{(0)} = \begin{cases} \sqrt{\dfrac{1}{1-\rho^{(0)2}}}\sqrt{\dfrac{\alpha_2}{\alpha_{10}+\alpha_6 h^*}} \ , & \text{if } \dfrac{\alpha_2}{\alpha_{10}+\alpha_6 h^*} > 0 \\ \sqrt{m_{0,2}-m_{0,1}^2} \ , & \text{if } \dfrac{\alpha_2}{\alpha_{10}+\alpha_6 h^*} \leq 0 \ ; \end{cases} \tag{91}$$

$$\xi^{(0)} = \sigma_x^{(0)}\left(\frac{\alpha_7+\alpha_3 h^*}{\alpha_1}\right) + \rho^{(0)}\sigma_y^{(0)}\left(\frac{\alpha_9+\alpha_5 h^*}{\alpha_2}\right) \ ; \tag{92}$$

$$\eta^{(0)} = \rho^{(0)}\sigma_x^{(0)}\left(\frac{\alpha_7+\alpha_3 h^*}{\alpha_1}\right) + \sigma_y^{(0)}\left(\frac{\alpha_9+\alpha_5 h^*}{\alpha_2}\right) \ . \tag{93}$$

Since sample moments converge in probability to the corresponding population moments [Cramér (1946), page 364)], then by Slutsky's Theorem [Cramér (1946), page 255] any rational function of the sample moments converges in probability to the same rational function of the population moments; hence, the estimates given by (74), (81), (82), (85), and (86) are consistent. When $\dfrac{\alpha_1}{\alpha_8+\alpha_4 h^*} > 0$ and $\dfrac{\alpha_2}{\alpha_{10}+\alpha_6 h^*} > 0$, the components of the starting vector (89)-(93) are consistent estimates by theorems on convergence in probability. For example, from the above discussion, we see that

$$h^* \xrightarrow{P} \frac{-\rho}{\sigma_x\sigma_y(1-\rho^2)} \ ,$$

$$\frac{\alpha_8+\alpha_4 h^*}{\alpha_1} \xrightarrow{P} \frac{1}{\sigma_x^2(1-\rho^2)} \ ,$$

and

$$\frac{\alpha_{10}+\alpha_6 h^*}{\alpha_2} \xrightarrow{P} \frac{1}{\sigma_y^2(1-\rho^2)} \ .$$

Thus,

$$\sqrt{\frac{\alpha_1}{\alpha_8+\alpha_4 h^*}} \xrightarrow{P} \sigma_x\sqrt{1-\rho^2} \quad , \quad \text{when} \quad \frac{\alpha_8+\alpha_4 h^*}{\alpha_1} > 0$$

$$\sqrt{\frac{\alpha_2}{\alpha_{10}+\alpha_6 h^*}} \xrightarrow{P} \sigma_y\sqrt{1-\rho^2} \quad , \quad \text{when} \quad \frac{\alpha_{10}+\alpha_6 h^*}{\alpha_2} > 0$$

and

$$\sqrt{\frac{\alpha_1}{\alpha_8+\alpha_4 h^*}}\sqrt{\frac{\alpha_2}{\alpha_{10}+\alpha_6 h^*}} \xrightarrow{P} \sigma_x\sigma_y(1-\rho^2) \ .$$

Hence,

$$\rho^{(0)} = -h^*\sqrt{\frac{\alpha_1}{\alpha_8+\alpha_4 h^*}}\sqrt{\frac{\alpha_2}{\alpha_{10}+\alpha_6 h^*}} \xrightarrow{P} \rho \ .$$

Since (89)-(93) are continuous functions of sample moments with continuous first and second order partial derivatives with respect to the sample moments, then by a theorem in Cramer (1946, page 366), the components of the starting vector are asymptotically jointly normal.

Finally, we note that since the maximum likelihood estimates are single-valued functions of $m_{1,0}$, $m_{0,1}$, $m_{2,0}$, $m_{0,2}$, and $m_{1,1}$, then certainly the components of the starting vector are asymptotically inefficient since they are functions of higher sample moments. Just how asymptotically inefficient can be determined by finding the asymptotic generalized variance of these estimates and using the asymptotic efficiency ratio. However, these are not the final estimates. On the contrary they are only the initial estimates to be improved upon by the functional iterative method and the Newton-Raphson method. Hence, we shall not bother with finding the asymptotic generalized variance of these estimates.

3. Functional Iterative Method

We shall now determine the iterants for the functional iterative method. These iterants shall have the usual form, namely:

$$\xi = F_1(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)$$

$$\eta = F_2(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)$$

$$\sigma_x = F_3(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)$$

$$\sigma_y = F_4(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)$$

$$\rho = F_5(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)\ .$$

We use the method of moments estimating equations (50)-(54). Squaring both sides of (50), subtracting from (52), and taking the square root of both sides of the resulting equation yield

$$\sigma_x = \frac{\sqrt{m_{2,0}-m_{1,0}^2}}{\sqrt{1+Q_{2,0}-Q_{1,0}^2}} = F_3(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)\ . \tag{94}$$

From (50)

$$\xi = Q_{1,0} - m_{1,0}/\sigma_x = F_1(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)\ . \tag{95}$$

Using (51) and (53) in a similar way, we find

$$\sigma_y = \frac{\sqrt{m_{0,2}-m_{0,1}^2}}{\sqrt{1+Q_{0,2}-Q_{0,1}^2}} = F_4(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho) \tag{96}$$

and

$$\eta = Q_{0,1} - m_{0,1}/\sigma_y = F_2(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)\ . \tag{97}$$

From (54), (95), and (97)

$$\rho = \frac{m_{1,1}}{\sigma_x\sigma_y} - Q_{1,1} - \xi\eta + \eta Q_{1,0} + \xi Q_{0,1}$$

$$= \frac{m_{1,1}}{\sigma_x\sigma_y} - Q_{1,1} - \xi\eta + \eta\left(\xi + \frac{m_{1,0}}{\sigma_x}\right) + \xi\left(\eta + \frac{m_{0,1}}{\sigma_y}\right)$$

$$= \frac{m_{1,1}-m_{1,0}m_{0,1}}{\sigma_x\sigma_y} - Q_{1,1} + \frac{m_{1,0}m_{0,1}}{\sigma_x\sigma_y} + \eta\,\frac{m_{1,0}}{\sigma_x} + \xi\,\frac{m_{0,1}}{\sigma_y} + \xi\eta$$

$$= \frac{m_{1,1}-m_{1,0}m_{0,1}}{\sigma_x\sigma_y} - Q_{1,1} + \left(\frac{m_{1,0}}{\sigma_x} + \xi\right)\left(\frac{m_{0,1}}{\sigma_y} + \eta\right) .$$

Thus,

$$\rho = \frac{m_{1,1}-m_{1,0}m_{0,1}}{\sigma_x\sigma_y} - Q_{1,1} + Q_{1,0}Q_{0,1} = F_5(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho) . \tag{98}$$

An alternate iterant for ρ may be found as follows. Since

$$Q_{2,0}-Q_{1,0}^2 = -\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right] - \rho^2\left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]$$
$$+ \rho g_5 - 2\rho(G_1-G_2)(G_3-G_4)$$

and

$$Q_{0,2}-Q_{0,1}^2 = -\rho^2\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right] - \left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]$$
$$+ \rho g_5 - 2\rho(G_1-G_2)(G_3-G_4) ,$$

then

$$\rho(Q_{2,0}-Q_{1,0}^2)\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right] - \rho(Q_{0,2}-Q_{0,1}^2)\left[(G_3-G_4)(G_3-G_4+\eta) + \frac{R_y}{\sigma_y}G_4\right]$$
$$= -\rho\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right]^2 + \rho\left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]^2 ,$$

$$+ \rho^2[g_5-2(G_1-G_2)(G_3-G_4)]\left\{\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right] - \left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]\right\} . \tag{99}$$

and

$$Q_{0,2}-Q_{0,1}^2-Q_{2,0}+Q_{1,0}^2 = (1-\rho^2)\left\{\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right] - \left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]\right\} . \tag{100}$$

Furthermore,

$$Q_{1,1}-Q_{1,0}Q_{0,1} = -\rho\left\{\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right]+\left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]\right\}$$
$$+ g_5 - (1+\rho^2)(G_1-G_2)(G_3-G_4) ,$$

so that

$$(Q_{1,1}-Q_{1,0}Q_{0,1})\left\{\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right]-\left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]\right\}$$
$$= -\rho\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right]^2 + \rho\left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]^2$$
$$+ [g_5-(1+\rho^2)(G_1-G_2)(G_3-G_4)]\left\{\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right] - \left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]\right\}. \tag{101}$$

Hence, from (99), (100), and (101) we see that

$$\rho(Q_{2,0}-Q_{1,0}^2)\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right] - \rho(Q_{0,2}-Q_{0,1}^2)\left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]$$
$$- (Q_{1,1}-Q_{1,0}Q_{0,1})\left\{\left[(G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x}G_2\right]- \left[(G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y}G_4\right]\right\}$$
$$+ (Q_{0,2}-Q_{0,1}^2-Q_{2,0}+Q_{1,0}^2)[g_5-(G_1-G_2)(G_3-G_4)] = 0 . \tag{102}$$

Upon putting (98) in (102) and solving for ρ we obtain

$$\rho = \frac{(A-B)\left[g_5-(G_1-G_2)(G_3-G_4)\right]+\left(\dfrac{m_{1,1}-m_{1,0}m_{0,1}}{\sigma_x\sigma_y}\right)(C-D)}{AC - BD}$$

$$= F_5(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho) \quad , \tag{103}$$

where

$$A = 1 + Q_{2,0} - Q_{1,0}^2 \qquad C = (G_1-G_2)(G_1-G_2-\xi) + \frac{R_x}{\sigma_x} G_2$$

$$B = 1 + Q_{0,2} - Q_{0,1}^2 \qquad D = (G_3-G_4)(G_3-G_4-\eta) + \frac{R_y}{\sigma_y} G_4 \quad .$$

Let $\gamma^{(k)}$ represent the k^{th} iterant of γ. The $(k+1)^{th}$ iterants to ξ, η, σ_x, σ_y, and ρ are found from the following:

$$\sigma_x^{(k+1)} = \frac{\sqrt{m_{2,0}-m_{1,0}^2}}{\sqrt{1 + Q_{2,0}^{(k)} - \left(Q_{1,0}^{(k)}\right)^2}} \quad , \tag{104}$$

$$\xi^{(k+1)} = Q_{1,0}^{(k)} - m_{1,0}/\sigma_x^{(k+1)} \quad , \tag{105}$$

$$\sigma_y^{(k+1)} = \frac{\sqrt{m_{0,2}-m_{0,1}^2}}{\sqrt{1 + Q_{0,2}^{(k)} - \left(Q_{0,1}^{(k)}\right)^2}} \quad , \tag{106}$$

$$\eta^{(k+1)} = Q_{0,1}^{(k)} - m_{0,1}/\sigma_y^{(k+1)} \quad , \tag{107}$$

$$\rho^{(k+1)} = \frac{m_{1,1}-m_{1,0}m_{0,1}}{\sigma_x^{(k+1)}\ \sigma_y^{(k+1)}} - Q_{1,1}^{(k)} + Q_{1,0}^{(k)}\ Q_{0,1}^{(k)} \tag{108a}$$

or

$$\rho^{(k+1)} = \frac{\left(A^{(k)}-B^{(k)}\right)\left[g_5^{(k)}-\left(G_1^{(k)}-G_2^{(k)}\right)\left(G_3^{(k)}-G_4^{(k)}\right)\right]+\left(\frac{m_{1,1}-m_{1,0}m_{0,1}}{\sigma_x^{(k+1)}\ \sigma_y^{(k+1)}}\right)\left(C^{(k)}-D^{(k)}\right)}{A^{(k)}C^{(k)}-B^{(k)}D^{(k)}}, \tag{108b}$$

where

$$A^{(k)} = 1 + Q_{2,0}^{(k)} - \left(Q_{1,0}^{(k)}\right)^2$$

$$B^{(k)} = 1 + Q_{0,2}^{(k)} - \left(Q_{0,1}^{(k)}\right)^2$$

$$C^{(k)} = \left(G_1^{(k)}-G_2^{(k)}\right)\left(G_1^{(k)}-G_2^{(k)}-\xi^{(k+1)}\right) + \frac{R_x}{\sigma_x^{(k+1)}}\, G_2^{(k)}$$

$$D^{(k)} = \left(G_3^{(k)}-G_4^{(k)}\right)\left(G_3^{(k)}-G_4^{(k)}-\eta^{(k+1)}\right) + \frac{R_y}{\sigma_y^{(k+1)}}\, G_4^{(k)} \quad .$$

Using the $(k+1)^{th}$ iterants, we compute $G_i^{(k+1)}$ and $g_i^{(k+1)}$, $i = 1, 2, 3, 4$. From these values we evaluate the $Q_{i,j}^{(k+1)}$ which are needed to evaluate $\mu_{i,j}'^{(k+1)}$. The values of the $\mu_{i,j}'^{(k+1)}$ are checked for accuracy in the method of moments estimating equations (50)-(54). The functional iterative method will be employed until iterants are obtained which give the desired degree of accuracy. At any cycle we note that G_i and g_i , $i = 1, 2, 3, 4$, may be computed with the aid of existing tables. The numerators may be evaluated through the use of tables of the standard univariate normal distribution and its cumulative distribution function. It is recommended that the tables in National Bureau of Standards (1953) be used to achieve sufficient accuracy. The denominator may be evaluated through the use of tables in National Bureau of Standards (1959).

In working with the functional iterative method the question of whether or not the procedure actually yields iterants which converge to the solution must be considered. If the initial estimates are "sufficiently close" to the solution, then, in general, the procedure will converge. "Sufficiently close" is often measured through the use of a maximum vector norm

$$||(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\rho)-(\hat{\xi},\ \hat{\eta},\ \hat{\sigma}_x,\ \hat{\sigma}_y,\ \hat{\rho})||_\infty \equiv \max\left\{|\xi-\hat{\xi}|,|\eta-\hat{\eta}|,|\sigma_x-\hat{\sigma}_x|,|\sigma_y-\hat{\sigma}_y|,|\rho-\hat{\rho}|\right\}.$$

This defines a metric over the vector space and the concept of a neighborhood is interpreted through this metric.

$$M = \begin{pmatrix} \frac{\partial F_1}{\partial \xi} & \frac{\partial F_1}{\partial \eta} & \frac{\partial F_1}{\partial \sigma_x} & \frac{\partial F_1}{\partial \sigma_y} & \frac{\partial F_1}{\partial \rho} \\[2ex] \frac{\partial F_2}{\partial \xi} & \frac{\partial F_2}{\partial \eta} & \frac{\partial F_2}{\partial \sigma_x} & \frac{\partial F_2}{\partial \sigma_y} & \frac{\partial F_2}{\partial \rho} \\[2ex] \frac{\partial F_3}{\partial \xi} & \frac{\partial F_3}{\partial \eta} & \frac{\partial F_3}{\partial \sigma_x} & \frac{\partial F_3}{\partial \sigma_y} & \frac{\partial F_3}{\partial \rho} \\[2ex] \frac{\partial F_4}{\partial \xi} & \frac{\partial F_4}{\partial \eta} & \frac{\partial F_4}{\partial \sigma_x} & \frac{\partial F_4}{\partial \sigma_y} & \frac{\partial F_4}{\partial \rho} \\[2ex] \frac{\partial F_5}{\partial \xi} & \frac{\partial F_5}{\partial \eta} & \frac{\partial F_5}{\partial \sigma_x} & \frac{\partial F_5}{\partial \sigma_y} & \frac{\partial F_5}{\partial \rho} \end{pmatrix},$$

where $F_i = F_i(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)$; $i = 1, 2, 3, 4, 5$. Under the assumption that the starting vector is "sufficiently close" to the solution, that is,

$$\left|\left|\left(\xi^{(0)},\ \eta^{(0)},\ \sigma_x^{(0)},\ \sigma_y^{(0)}\ ,\ \rho^{(0)}\right)-(\hat{\xi},\ \hat{\eta},\ \hat{\sigma}_x,\ \hat{\sigma}_y,\ \hat{\rho})\right|\right|_\infty < \varepsilon \text{ (arbitrary)}\ , \tag{109}$$

then Isaacson and Keller (1966, pp. 111-112) have shown that a sufficient condition for convergence of the proposed functional iterative method is that the natural norm of the matrix M induced by the maximum vector norm be less than unity, that is,

$$||M||_\infty = \max_{1\le i\le 5}\left\{\left|\frac{\partial F_i}{\partial \xi}\right|+\left|\frac{\partial F_i}{\partial \eta}\right|+\left|\frac{\partial F_i}{\partial \sigma_x}\right|+\left|\frac{\partial F_i}{\partial \sigma_y}\right|+\left|\frac{\partial F_i}{\partial \rho}\right|\right\} < 1 \tag{110}$$

for all points $(\xi, \eta, \sigma_x, \sigma_y, \rho)$ in (109). Looking upon (109) as a neighborhood of the starting vector, a sufficient condition for convergence generally adopted is

$$\max_{1 \le i \le 5} \left\{ \left|\frac{\partial F_i}{\partial \xi}\right|_{(0)} + \left|\frac{\partial F_i}{\partial \eta}\right|_{(0)} + \left|\frac{\partial F_i}{\partial \sigma_x}\right|_{(0)} + \left|\frac{\partial F_i}{\partial \sigma_y}\right|_{(0)} + \left|\frac{\partial F_i}{\partial \rho}\right|_{(0)} \right\} < 1 , \tag{111}$$

where, for example,

$$\left|\frac{\partial F_i}{\partial \xi}\right|_{(0)} = \left|\frac{\partial F_i}{\partial \xi}\right|_{\left(\xi^{(0)},\ \eta^{(0)},\ \sigma_x^{(0)},\ \sigma_y^{(0)},\ \rho^{(0)}\right)} .$$

In order to have rapid convergence, it is necessary that the left-hand side of (111) be much less than one. If the left-hand side of (111) is large, but of course still less than one, then convergence will probably be slow.

Considering the expressions for $F_i(\xi, \eta, \sigma_x, \sigma_y, \rho)$, the number of partial derivatives required and the fact that many times one can tell after a few cycles whether or not the procedure appears to be converging and, if so, the possible rate of convergence, it would not seem practical to attempt to check the validity of (111). Thus, the explicit expressions needed to do so are not given; however, if they are needed, they can be found by differentiation.

In case of slow convergence or even divergence an alteration to the proposed functional iterative method is offered which will generally accelerate the rate of convergence or may even yield a convergent scheme when the basic one diverges. The theory behind this alteration may be found in Isaacson and Keller (1966, pp. 120-123).

In the alteration the $(k+1)^{th}$ iterants are found from the following:

$$\sigma_x^{(k+1)} = \Theta_1^{(k)} F_3\left(\xi^{(k)},\ \eta^{(k)},\ \sigma_x^{(k)},\ \sigma_y^{(k)},\ \rho^{(k)}\right) + \left(1 - \Theta_1^{(k)}\right)\sigma_x^{(k)} , \tag{112}$$

where

$$\Theta_1^{(k)} = \frac{1}{1 - \dfrac{F_3\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}+h, \sigma_y^{(k)}, \rho^{(k)}\right) - F_3\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)}{h}}$$

and

$F_3\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)$ equals the right-hand side of (104);

$$\xi^{(k+1)} = \Theta_2^{(k)} F_1\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right) + \left(1 - \Theta_2^{(k)}\right)\xi^{(k)} \quad , \tag{113}$$

where

$$\Theta_2^{(k)} = \frac{1}{1 - \dfrac{F_1\left(\xi^{(k)}+h, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right) - F_1\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)}{h}}$$

and

$F_1\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)$ equals the right-hand side of (105);

$$\sigma_y^{(k+1)} = \Theta_3^{(k)} F_4\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right) + \left(1 - \Theta_3^{(k)}\right)\sigma_y^{(k)} \quad , \tag{114}$$

where

$$\Theta_3^{(k)} = \frac{1}{1 - \dfrac{F_4\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}+h, \rho^{(k)}\right) - F_4\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)}{h}}$$

and

$F_4\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)$ equals the right-hand side of (106);

$$\eta^{(k+1)} = \Theta_4^{(k)} F_2\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right) + \left(1 - \Theta_4^{(k)}\right)\eta^{(k)} \quad , \tag{115}$$

where

$$\Theta_4^{(k)} = \cfrac{1}{1 - \cfrac{F_2\left(\xi^{(k)}, \eta^{(k)}+h, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right) - F_2\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)}{h}}$$

and

$F_2\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)$ equals the right-hand side of (107);

$$\rho^{(k+1)} = \Theta_5^{(k)} F_5\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right) + \left(1 - \Theta_5^{(k)}\right)\rho^{(k)} \quad , \tag{116}$$

where

$$\Theta_5^{(k)} = \cfrac{1}{1 - \cfrac{F_5\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}+h\right) - F_5\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)}{h}}$$

and

$F_5\left(\xi^{(k)}, \eta^{(k)}, \sigma_x^{(k)}, \sigma_y^{(k)}, \rho^{(k)}\right)$ equals the right-hand side of (108);

h is some suitably chosen small value.

4. Newton-Raphson Method

The Newton-Raphson method is based on a Taylor series expansion of the estimating equations in a neighborhood of the solution.

Let $\xi^{(r)}$, $\eta^{(r)}$, $\sigma_x^{(r)}$, $\sigma_y^{(r)}$, and $\rho^{(r)}$ be the final iterants obtained by the functional iterative method. These estimates should be arbitrarily close to the maximum likelihood estimates since the functional iterative method is carried out until the method of moments estimating equations are very nearly satisfied. Let $\Delta\xi$, $\Delta\eta$, $\Delta\sigma_x$, $\Delta\sigma_y$, and $\Delta\rho$ denote the corrections to these final estimates, that is

$$\xi^{(r)} + \Delta\xi = \hat{\xi}$$

$$\eta^{(r)} + \Delta\eta = \hat{\eta}$$

$$\sigma_x^{(r)} + \Delta\sigma_x = \hat{\sigma}_x$$

$$\sigma_y^{(r)} + \Delta\sigma_y = \hat{\sigma}_y$$

$$\rho^{(r)} + \Delta\rho = \hat{\rho} \quad . \tag{117}$$

Suppose the system of estimating equations, whether given by the method of moments or the method of maximum likelihood, is represented by

$$K_i(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho) = 0 \quad ; \quad i = 1, 2, 3, 4, 5 \ . \tag{118}$$

Expanding the left-hand side of each equation in (118) by Taylor's theorem for a function of five variables about the point $\left(\xi^{(r)},\ \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)},\ \rho^{(r)}\right)$, replacing as is the usual practice, $\xi, \eta, \sigma_x, \sigma_y$, and ρ by $\hat{\xi}, \hat{\eta}, \hat{\sigma}_x, \hat{\sigma}_y$, and $\hat{\rho}$, respectively, and neglecting all squares, products and higher powers of the corrections, then the system of equations (118) may be written as

$$K_i\left(\xi^{(r)},\ \eta^{(r)},\ \sigma_x^{(r)}, \sigma_y^{(r)},\ \rho^{(r)}\right) + \Delta\xi\left(\frac{\partial K_i}{\partial \xi}\right)_{(r)} + \Delta\eta\left(\frac{\partial K_i}{\partial \eta}\right)_{(r)} + \Delta\sigma_x\left(\frac{\partial K_i}{\partial \sigma_x}\right)_{(r)}$$
$$+ \Delta\sigma_y\left(\frac{\partial K_i}{\partial \sigma_y}\right)_{(r)} + \Delta\rho\left(\frac{\partial K_i}{\partial \rho}\right)_{(r)} = 0 \quad ; \quad i = 1, 2, 3, 4, 5 \tag{119}$$

where, for example,

$$\left(\frac{\partial K_i}{\partial \xi}\right)_{(r)} = \left.\frac{\partial K_i(\xi,\ \eta,\ \sigma_x,\ \sigma_y,\ \rho)}{\partial \xi}\right|_{\left(\xi^{(r)},\ \eta^{(r)},\ \sigma_x^{(r)},\ \sigma_y^{(r)},\ \rho^{(r)}\right)} \ .$$

Looking at (119) as a system of equations linear in the corrections, we rewrite (119) in matrix form as

$$\begin{pmatrix} \left(\frac{\partial K_1}{\partial \xi}\right)_{(r)} & \left(\frac{\partial K_1}{\partial \eta}\right)_{(r)} & \left(\frac{\partial K_1}{\partial \sigma_x}\right)_{(r)} & \left(\frac{\partial K_1}{\partial \sigma_y}\right)_{(r)} & \left(\frac{\partial K_1}{\partial \rho}\right)_{(r)} \\ \left(\frac{\partial K_2}{\partial \xi}\right)_{(r)} & \left(\frac{\partial K_2}{\partial \eta}\right)_{(r)} & \left(\frac{\partial K_2}{\partial \sigma_x}\right)_{(r)} & \left(\frac{\partial K_2}{\partial \sigma_y}\right)_{(r)} & \left(\frac{\partial K_2}{\partial \rho}\right)_{(r)} \\ \left(\frac{\partial K_3}{\partial \xi}\right)_{(r)} & \left(\frac{\partial K_3}{\partial \eta}\right)_{(r)} & \left(\frac{\partial K_3}{\partial \sigma_x}\right)_{(r)} & \left(\frac{\partial K_3}{\partial \sigma_y}\right)_{(r)} & \left(\frac{\partial K_3}{\partial \rho}\right)_{(r)} \\ \left(\frac{\partial K_4}{\partial \xi}\right)_{(r)} & \left(\frac{\partial K_4}{\partial \eta}\right)_{(r)} & \left(\frac{\partial K_4}{\partial \sigma_x}\right)_{(r)} & \left(\frac{\partial K_4}{\partial \sigma_y}\right)_{(r)} & \left(\frac{\partial K_4}{\partial \rho}\right)_{(r)} \\ \left(\frac{\partial K_5}{\partial \xi}\right)_{(r)} & \left(\frac{\partial K_5}{\partial \eta}\right)_{(r)} & \left(\frac{\partial K_5}{\partial \sigma_x}\right)_{(r)} & \left(\frac{\partial K_5}{\partial \sigma_y}\right)_{(r)} & \left(\frac{\partial K_5}{\partial \rho}\right)_{(r)} \end{pmatrix} \begin{pmatrix} \Delta\xi \\ \Delta\eta \\ \Delta\sigma_x \\ \Delta\sigma_y \\ \Delta\rho \end{pmatrix} = - \begin{pmatrix} K_1\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \\ K_2\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \\ K_3\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \\ K_4\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \\ K_5\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \end{pmatrix} .$$

Assuming the coefficient matrix is nonsingular, then

$$\begin{pmatrix} \Delta\xi \\ \Delta\eta \\ \Delta\sigma_x \\ \Delta\sigma_y \\ \Delta\rho \end{pmatrix} = \begin{pmatrix} -\left(\frac{\partial K_1}{\partial \xi}\right)_{(r)} & -\left(\frac{\partial K_1}{\partial \eta}\right)_{(r)} & -\left(\frac{\partial K_1}{\partial \sigma_x}\right)_{(r)} & -\left(\frac{\partial K_1}{\partial \sigma_y}\right)_{(r)} & -\left(\frac{\partial K_1}{\partial \rho}\right)_{(r)} \\ -\left(\frac{\partial K_2}{\partial \xi}\right)_{(r)} & -\left(\frac{\partial K_2}{\partial \eta}\right)_{(r)} & -\left(\frac{\partial K_2}{\partial \sigma_x}\right)_{(r)} & -\left(\frac{\partial K_2}{\partial \sigma_y}\right)_{(r)} & -\left(\frac{\partial K_2}{\partial \rho}\right)_{(r)} \\ -\left(\frac{\partial K_3}{\partial \xi}\right)_{(r)} & -\left(\frac{\partial K_3}{\partial \eta}\right)_{(r)} & -\left(\frac{\partial K_3}{\partial \sigma_x}\right)_{(r)} & -\left(\frac{\partial K_3}{\partial \sigma_y}\right)_{(r)} & -\left(\frac{\partial K_3}{\partial \rho}\right)_{(r)} \\ -\left(\frac{\partial K_4}{\partial \xi}\right)_{(r)} & -\left(\frac{\partial K_4}{\partial \eta}\right)_{(r)} & -\left(\frac{\partial K_4}{\partial \sigma_x}\right)_{(r)} & -\left(\frac{\partial K_4}{\partial \sigma_y}\right)_{(r)} & -\left(\frac{\partial K_4}{\partial \rho}\right)_{(r)} \\ -\left(\frac{\partial K_5}{\partial \xi}\right)_{(r)} & -\left(\frac{\partial K_5}{\partial \eta}\right)_{(r)} & -\left(\frac{\partial K_5}{\partial \sigma_x}\right)_{(r)} & -\left(\frac{\partial K_5}{\partial \sigma_y}\right)_{(r)} & -\left(\frac{\partial K_5}{\partial \rho}\right)_{(r)} \end{pmatrix}^{-1} \begin{pmatrix} K_1\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \\ K_2\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \\ K_3\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \\ K_4\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \\ K_5\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right) \end{pmatrix} . \tag{120}$$

For simplicity, in the following discussion we **relabed** the parameters by

$$\underline{\Theta} = \begin{pmatrix} \Theta_1 \\ \Theta_2 \\ \Theta_3 \\ \Theta_4 \\ \Theta_5 \end{pmatrix} = \begin{pmatrix} \xi \\ \eta \\ \sigma_x \\ \sigma_y \\ \rho \end{pmatrix} .$$

It is well known that $(\hat{\xi}, \hat{\eta}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho})$ under regularity conditions are asymptotically jointly normal with

$$\text{mean vector: } \underline{\Theta}$$

$$\text{variance-covariance matrix: } \left(- E\left(\frac{\partial^2 \log L}{\partial\Theta_i \partial\Theta_j}\right)\right)^{-1} . \tag{121}$$

We have shown that $\left(\sum_{i=1}^{N} X_i' , \sum_{i=1}^{N} Y_i' , \sum_{i=1}^{N} X_i'^2 , \sum_{i=1}^{N} Y_i'^2 , \sum_{i=1}^{N} X_i'Y_i'\right)$ is a minimal set of jointly sufficient statistics for the parameters $(\xi, \eta, \sigma_x, \sigma_y, \rho)$. In light of the existence of a set of sufficient statistics, Huzurbazar (1949) has shown through the use of the likelihood function of the most general form of a distribution admitting a set of jointly sufficient statistics that

$$\left(\frac{\partial^2 \log L}{\partial\Theta_i \partial\Theta_j}\right)_{\hat{\underline{\Theta}}=\underline{\Theta}} = E\left(\frac{\partial^2 \log L}{\partial\Theta_i \partial\Theta_j}\right) . \tag{122}$$

The left-hand side of (122) is to be interpreted as the second partial derivative of the logarithm of the likelihood function with the maximum likelihood estimates replaced by the corresponding parameters. That such a replacement can be made and will eliminate all data from the expression for the second partial derivative is guaranteed by the fact that all forms of the data necessarily appear in the second partial derivative of the logarithm of the likelihood function for an exponential family only through functions of the sufficient statistics. Thus, the large sample variance-covariance matrix in (121) may be equivalently written as

$$\left(-E\left(\frac{\partial^2 \log L}{\partial\Theta_i \partial\Theta_j}\right)\right)^{-1} = \left(-\left(\frac{\partial^2 \log L}{\partial\Theta_i \partial\Theta_j}\right)_{\hat{\underline{\Theta}}=\underline{\Theta}}\right)^{-1} . \tag{123}$$

We may obtain an estimate of the large sample variance-covariance matrix of the maximum likelihood estimates by replacing the parameters which appear in the elements of either matrix in (123) by the corresponding maximum likelihood estimates. However, since $\left(\xi^{(r)}, \eta^{(r)}, \sigma_x^{(r)}, \sigma_y^{(r)}, \rho^{(r)}\right)$ is arbitrarily close to $(\hat{\xi}, \hat{\eta}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho})$, then for all practical purposes the inverse matrix of (120) with $K_1 = \frac{\partial \log L}{\partial \xi}$, $K_2 = \frac{\partial \log L}{\partial \eta}$, $K_3 = \frac{\partial \log L}{\partial \sigma_x}$, $K_4 = \frac{\partial \log L}{\partial \sigma_y}$, and $K_5 = \frac{\partial \log L}{\partial \rho}$ could serve as an estimate of the large sample variance-covariance matrix of the maximum likelihood estimates.

One cycle of the Newton-Raphson method yields values for $\Delta\xi'$, $\Delta\eta'$, $\Delta\sigma_x'$, $\Delta\sigma_y'$, and $\Delta\rho'$ from the equations

$$\begin{pmatrix} \Delta\xi' \\ \Delta\eta' \\ \Delta\sigma_x' \\ \Delta\sigma_y' \\ \Delta\rho' \end{pmatrix} = \begin{pmatrix} V_{11} & V_{12} & V_{13} & V_{14} & V_{15} \\ V_{12} & V_{22} & V_{23} & V_{24} & V_{25} \\ V_{13} & V_{23} & V_{33} & V_{34} & V_{35} \\ V_{14} & V_{24} & V_{34} & V_{44} & V_{45} \\ V_{15} & V_{25} & V_{35} & V_{45} & V_{55} \end{pmatrix}^{-1} \begin{pmatrix} \left(\frac{\partial \log L}{\partial \xi}\right)_{(r)} \\ \left(\frac{\partial \log L}{\partial \eta}\right)_{(r)} \\ \left(\frac{\partial \log L}{\partial \sigma_x}\right)_{(r)} \\ \left(\frac{\partial \log L}{\partial \sigma_y}\right)_{(r)} \\ \left(\frac{\partial \log L}{\partial \rho}\right)_{(r)} \end{pmatrix} , \tag{124}$$

where, in terms of the values obtained in the final cycle of the functional iterative method,

$$\left(\frac{\partial \log L}{\partial \xi}\right)_{(r)} = \frac{N}{\sigma_x^{(r)}\sigma_y^{(r)}\left[1-\left(\rho^{(r)}\right)^2\right]}\left[\sigma_y^{(r)}\left(\mu_{1,0}^{\prime\,(r)}-m_{1,0}\right)-\rho^{(r)}\sigma_x^{(r)}\left(\mu_{0,1}^{\prime\,(r)}-m_{0,1}\right)\right],$$

$$\left(\frac{\partial \log L}{\partial \eta}\right)_{(r)} = \frac{N}{\sigma_x^{(r)}\sigma_y^{(r)}\left[1-\left(\rho^{(r)}\right)^2\right]}\left[\sigma_x^{(r)}\left(\mu_{0,1}^{\prime\,(r)}-m_{0,1}\right)-\rho^{(r)}\sigma_y^{(r)}\left(\mu_{1,0}^{\prime\,(r)}-m_{1,0}\right)\right],$$

$$\left(\frac{\partial \log L}{\partial \sigma_x}\right)_{(r)} = \frac{N}{\left(\sigma_x^{(r)}\right)^3\sigma_y^{(r)}\left[1-\left(\rho^{(r)}\right)^2\right]}\left[\rho^{(r)}\sigma_x^{(r)}\left(\mu_{1,1}^{\prime\,(r)}-m_{1,1}\right)-\sigma_y^{(r)}\left(\mu_{2,0}^{\prime\,(r)}-m_{2,0}\right)\right.$$
$$\left.-\sigma_x^{(r)}\sigma_y^{(r)}\left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}\right)\left(\mu_{1,0}^{\prime\,(r)}-m_{1,0}\right)\right],$$

$$\left(\frac{\partial \log L}{\partial \sigma_y}\right)_{(r)} = \frac{N}{\sigma_x^{(r)}\left(\sigma_y^{(r)}\right)^3\left[1-\left(\rho^{(r)}\right)^2\right]}\left[\rho^{(r)}\sigma_y^{(r)}\left(\mu_{1,1}^{\prime\,(r)}-m_{1,1}\right)-\sigma_x^{(r)}\left(\mu_{0,2}^{\prime\,(r)}-m_{0,2}\right)\right.$$
$$\left.-\sigma_x^{(r)}\sigma_y^{(r)}\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}\right)\left(\mu_{0,1}^{\prime\,(r)}-m_{0,1}\right)\right],$$

$$\left(\frac{\partial \log L}{\partial \rho}\right)_{(r)} = \frac{N}{\left(\sigma_x^{(r)}\right)^2\left(\sigma_y^{(r)}\right)^2\left[1-\left(\rho^{(r)}\right)^2\right]^2}\left\{\rho^{(r)}\left(\sigma_y^{(r)}\right)^2\left(\mu_{2,0}^{\prime\,(r)}-m_{2,0}\right)\right.$$
$$+\rho^{(r)}\left(\sigma_x^{(r)}\right)^2\left(\mu_{0,2}^{\prime\,(r)}-m_{0,2}\right)-\sigma_x^{(r)}\sigma_y^{(r)}\left[1+\left(\rho^{(r)}\right)^2\right]\left(\mu_{1,1}^{\prime\,(r)}-m_{1,1}\right)$$
$$-\sigma_x^{(r)}\left(\sigma_y^{(r)}\right)^2\left[\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}\right)-\rho^{(r)}\left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}\right)\right]\left(\mu_{1,0}^{\prime\,(r)}-m_{1,0}\right)$$
$$\left.-\left(\sigma_x^{(r)}\right)^2\sigma_y^{(r)}\left[\left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}\right)-\rho^{(r)}\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}\right)\right]\left(\mu_{0,1}^{\prime\,(r)}-m_{0,1}\right)\right\},$$

and

$$V_{11} = -E\left(\frac{\partial^2 \log L}{\partial \xi^2}\right)\Bigg|_{(r)} = -N\left\{\left(G_1^{(r)}-G_2^{(r)}\right)\left(G_1^{(r)}-G_2^{(r)}-\xi^{(r)}\right)+\frac{R_x}{\sigma_x^{(r)}}G_2^{(r)}-\rho^{(r)}\frac{\left(\rho^{(r)}-g_5^{(r)}\right)}{\left[1-\left(\rho^{(r)}\right)^2\right]}-1\right\},$$

$$V_{12} = -E\left(\frac{\partial^2 \log L}{\partial \xi\,\partial \eta}\right)\Bigg|_{(r)} = -N\left\{\left(G_1^{(r)}-G_2^{(r)}\right)\left(G_3^{(r)}-G_4^{(r)}\right)+\frac{\left(\rho^{(r)}-g_5^{(r)}\right)}{\left[1-\left(\rho^{(r)}\right)^2\right]}\right\},$$

$$V_{13} = -E\left(\frac{\partial^2 \log L}{\partial\xi\,\partial\sigma_x}\right)\Bigg|_{(r)} = -N\left\{\frac{R_x}{\left(\sigma_x^{(r)}\right)^2}\left[G_2^{(r)}\left(G_1^{(r)}-G_2^{(r)}-\xi^{(r)}\right)-\frac{R_x}{\sigma_x^{(r)}}G_2^{(r)} - \rho^{(r)}\frac{\left(g_4^{(r)}-g_3^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right]+\frac{\mu_{1,0}'^{(r)}}{\left(\sigma_x^{(r)}\right)^2\left[1-\left(\rho^{(r)}\right)^2\right]}\right\},$$

$$V_{14} = -E\left(\frac{\partial^2 \log L}{\partial\xi\,\partial\sigma_y}\right)\Bigg|_{(r)} = -N\left\{\frac{R_y}{\left(\sigma_y^{(r)}\right)^2}\left[G_4^{(r)}\left(G_1^{(r)}-G_2^{(r)}\right)+\frac{\left(g_4^{(r)}-g_2^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right] - \frac{\rho^{(r)}\mu_{0,1}'^{(r)}}{\left(\sigma_y^{(r)}\right)^2\left[1-\left(\rho^{(r)}\right)^2\right]}\right\},$$

$$V_{15} = -E\left(\frac{\partial^2 \log L}{\partial\xi\,\partial\rho}\right)\Bigg|_{(r)} = -N\left\{\frac{1}{\left[1-\left(\rho^{(r)}\right)^2\right]}\left[\left(G_3^{(r)}-G_4^{(r)}\right)-\left(g_5^{(r)}+\rho^{(r)}\right)\left(G_1^{(r)}-G_2^{(r)}\right) + \frac{\left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}\right)g_5^{(r)}}{\left[1-\left(\rho^{(r)}\right)^2\right]} - \frac{R_y}{\sigma_y^{(r)}}\rho^{(r)} \times \frac{\left(g_4^{(r)}-g_2^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}} + \frac{R_x}{\sigma_x^{(r)}}\frac{\left(g_4^{(r)}-g_3^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right]\right\},$$

$$V_{22} = -E\left(\frac{\partial^2 \log L}{\partial\eta^2}\right)\Bigg|_{(r)} = -N\left\{\left(G_3^{(r)}-G_4^{(r)}\right)\left(G_3^{(r)}-G_4^{(r)}-\eta^{(r)}\right)+\frac{R_y}{\sigma_y^{(r)}}G_4^{(r)} - \rho^{(r)}\frac{\left(\rho^{(r)}-g_5^{(r)}\right)}{\left[1-\left(\rho^{(r)}\right)^2\right]}-1\right\},$$

$$V_{23} = -E\left(\frac{\partial^2 \log L}{\partial\eta\,\partial\sigma_x}\right)\Bigg|_{(r)} = -N\left\{\frac{R_x}{\left(\sigma_x^{(r)}\right)^2}\left[G_2^{(r)}\left(G_3^{(r)}-G_4^{(r)}\right)+\frac{\left(g_4^{(r)}-g_3^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right] - \frac{\rho^{(r)}\mu_{1,0}'^{(r)}}{\left(\sigma_x^{(r)}\right)^2\left[1-\left(\rho^{(r)}\right)^2\right]}\right\},$$

$$V_{24} = -E\left(\frac{\partial^2 \log L}{\partial\eta\,\partial\sigma_y}\right)\Bigg|_{(r)} = -N\left\{\frac{R_y}{\left(\sigma_y^{(r)}\right)^2}\left[G_4^{(r)}\left(G_3^{(r)}-G_4^{(r)}-\eta^{(r)}\right)-\frac{R_y}{\sigma_y^{(r)}}G_4^{(r)} - \rho^{(r)}\frac{\left(g_4^{(r)}-g_2^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right]+\frac{\mu_{0,1}'^{(r)}}{\left(\sigma_y^{(r)}\right)^2\left[1-\left(\rho^{(r)}\right)^2\right]}\right\},$$

$$
\begin{aligned}
V_{25} = -E\left(\frac{\partial^2 \log L}{\partial\eta\,\partial\rho}\right)\Bigg|_{(r)} = -N\Bigg\{ & \frac{1}{\left[1-\left(\rho^{(r)}\right)^2\right]}\Bigg[\left(G_1^{(r)}-G_2^{(r)}\right)-\left(g_5^{(r)}+\rho^{(r)}\right)\left(G_3^{(r)}-G_4^{(r)}\right) \\
& + \frac{\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}\right)}{\left[1-\left(\rho^{(r)}\right)^2\right]}\, g_5^{(r)} - \frac{R_x}{\sigma_x^{(r)}}\,\rho^{(r)} \\
& \times \frac{\left(g_4^{(r)}-g_3^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}} + \frac{R_y}{\sigma_y^{(r)}}\,\frac{\left(g_4^{(r)}-g_2^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\Bigg]\Bigg\} ,
\end{aligned}
$$

$$
\begin{aligned}
V_{33} = -E\left(\frac{\partial^2 \log L}{\partial\sigma_x^2}\right)\Bigg|_{(r)} = -N\Bigg\{ & \left[\frac{R_x}{\left(\sigma_x^{(r)}\right)^2}\right]^2\left[G_2^{(r)}\left(G_2^{(r)}+\xi^{(r)}+\frac{R_x}{\sigma_x^{(r)}}\right)+\rho^{(r)}\,\frac{\left(g_4^{(r)}-g_3^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right] \\
& + \frac{1}{\left(\sigma_x^{(r)}\right)^2}\left(1-\frac{2R_x}{\sigma_x^{(r)}}\,G_2^{(r)}\right) - \frac{3\mu_{2,0}'^{(r)}}{\left(\sigma_x^{(r)}\right)^4\left[1-\left(\rho^{(r)}\right)^2\right]} \\
& + \frac{2\rho^{(r)}\mu_{1,1}'^{(r)}}{\left(\sigma_x^{(r)}\right)^3\sigma_y^{(r)}\left[1-\left(\rho^{(r)}\right)^2\right]} \\
& - \frac{2\left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}\right)\mu_{1,0}'^{(r)}}{\left(\sigma_x^{(r)}\right)^3\left[1-\left(\rho^{(r)}\right)^2\right]}\Bigg\} ,
\end{aligned}
$$

$$
\begin{aligned}
V_{34} = -E\left(\frac{\partial^2 \log L}{\partial\sigma_x\,\partial\sigma_y}\right)\Bigg|_{(r)} = -N\Bigg\{ & \left[\frac{R_x}{\left(\sigma_x^{(r)}\right)^2}\right]\left[\frac{R_y}{\left(\sigma_y^{(r)}\right)^2}\right]\left(G_2^{(r)}G_4^{(r)}-\frac{g_4^{(r)}}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right) \\
& + \frac{\rho^{(r)}\mu_{1,1}'^{(r)}}{\left(\sigma_x^{(r)}\right)^2\left(\sigma_y^{(r)}\right)^2\left[1-\left(\rho^{(r)}\right)^2\right]}\Bigg\} ,
\end{aligned}
$$

$$
\begin{aligned}
V_{35} = -E\left(\frac{\partial^2 \log L}{\partial\sigma_x\,\partial\rho}\right)\Bigg|_{(r)} = -N\Bigg\{ & \frac{R_x}{\left(\sigma_x^{(r)}\right)^2\left[1-\left(\rho^{(r)}\right)^2\right]}\left[\frac{R_y}{\sigma_y^{(r)}}\,\rho^{(r)}\,\frac{g_4^{(r)}}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right. \\
& \left. - \left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}+\frac{R_x}{\sigma_x^{(r)}}\right)\frac{\left(g_4^{(r)}-g_3^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right] \\
& + \frac{1}{\sigma_x^{(r)}\left[1-\left(\rho^{(r)}\right)^2\right]}\Bigg[\left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}\right)\left(G_3^{(r)}-G_4^{(r)}\right) \\
& + \rho^{(r)}\,\frac{R_y}{\sigma_y^{(r)}}\,G_4^{(r)} ++ \left(\rho^{(r)}-g_5^{(r)}\right)g_5^{(r)}
\end{aligned}
$$

$$- \left(g_5^{(r)}+\rho^{(r)}\right) \frac{R_x}{\sigma_x^{(r)}} G_2^{(r)} \Bigg]\Bigg\} ,$$

$$V_{44} = -E\left(\frac{\partial^2 \log L}{\partial \sigma_y^2}\right)\Bigg|_{(r)} = -N\Bigg\{\left[\frac{R_y}{\left(\sigma_y^{(r)}\right)^2}\right]^2\left[G_4^{(r)}\left(G_4^{(r)}+\eta^{(r)}+\frac{R_y}{\sigma_y^{(r)}}\right)+\rho^{(r)}\frac{\left(g_4^{(r)}-g_2^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right]$$

$$+\frac{1}{\left(\sigma_y^{(r)}\right)^2}\left(1-\frac{2R_y}{\sigma_y^{(r)}}G_4^{(r)}\right)-\frac{3\mu'^{(r)}_{0,2}}{\left(\sigma_y^{(r)}\right)^4\left[1-\left(\rho^{(r)}\right)^2\right]}$$

$$+\frac{2\rho^{(r)}\mu'^{(r)}_{1,1}}{\sigma_x^{(r)}\left(\sigma_y^{(r)}\right)^3\left[1-\left(\rho^{(r)}\right)^2\right]}$$

$$-\frac{2\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}\right)\mu'^{(r)}_{0,1}}{\left(\sigma_y^{(r)}\right)^3\left[1-\left(\rho^{(r)}\right)^2\right]}\Bigg\} ,$$

$$V_{45} = -E\,\frac{\partial^2 \log L}{\partial \sigma_y \partial \rho}\Bigg|_{(r)} = -N\Bigg\{\frac{R_y}{\left(\sigma_y^{(r)}\right)^2\left[1-\left(\rho^{(r)}\right)^2\right]}\left[\frac{R_x}{\sigma_x^{(r)}}\rho^{(r)}\frac{g_4^{(r)}}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right.$$

$$\left.-\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}+\frac{R_y}{\sigma_y^{(r)}}\right)\frac{\left(g_4^{(r)}-g_2^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}}\right]$$

$$+\frac{1}{\sigma_y^{(r)}\left[1-\left(\rho^{(r)}\right)^2\right]}\left[\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}\right)\left(G_1^{(r)}-G_2^{(r)}\right)\right.$$

$$+\rho^{(r)}\frac{R_x}{\sigma_x^{(r)}}G_2^{(r)}+\left(\rho^{(r)}-g_5^{(r)}\right)$$

$$\left.-\left(g_5^{(r)}+\rho^{(r)}\right)\frac{R_y}{\sigma_y^{(r)}}G_4^{(r)}\right]\Bigg\} ,$$

$$V_{55} = -E\left(\frac{\partial^2 \log L}{\partial \rho^2}\right)\Bigg|_{(r)} = -N\Bigg(\frac{1}{\left[1-\left(\rho^{(r)}\right)^2\right]}\left[\frac{R_x}{\sigma_x^{(r)}}G_2^{(r)}+\frac{R_y}{\sigma_y^{(r)}}G_4^{(r)}-\xi^{(r)}\left(G_1^{(r)}-G_2^{(r)}\right)\right.$$

$$\left.-\eta^{(r)}\left(G_3^{(r)}-G_4^{(r)}\right)-1\right]-\frac{1}{\left[1-\left(\rho^{(r)}\right)^2\right]^2}$$

$$\times\Bigg\{\left[\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}\right)-\rho^{(r)}\left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}+\frac{R_x}{\sigma_x^{(r)}}\right)\right] .$$

$$
\begin{aligned}
&\times \frac{R_x}{\sigma_x^{(r)}} \; \frac{\left(g_4^{(r)}-g_3^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}} + \left[\left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}\right)\right. \\
&\left. - \rho^{(r)}\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}+\frac{R_y}{\sigma_y^{(r)}}\right)\right]\frac{R_y}{\sigma_y^{(r)}} \; \frac{\left(g_4^{(r)}-g_2^{(r)}\right)}{\sqrt{1-\left(\rho^{(r)}\right)^2}} \\
&+ \left(\frac{R_x}{\sigma_x^{(r)}}\right)\left(\frac{R_y}{\sigma_y^{(r)}}\right)\left[1+\left(\rho^{(r)}\right)^2\right] \frac{g_4^{(r)}}{\sqrt{1-\left(\rho^{(r)}\right)^2}} \\
&+ 2\left[\rho^{(r)}\right]^2 - 3\rho^{(r)}g_5^{(r)} - \left[g_5^{(r)}\right]^2 \\
&\left. \left. + \frac{\left(\xi^{(r)}-\rho^{(r)}\eta^{(r)}\right)\left(\eta^{(r)}-\rho^{(r)}\xi^{(r)}\right)}{\left[1-\left(\rho^{(r)}\right)^2\right]} \, g_5^{(r)} \right\} \right) .
\end{aligned}
$$

We take as the maximum likelihood estimates

$$
\begin{aligned}
\hat{\xi} &= \xi^{(r)} + \Delta\xi' \\
\hat{\eta} &= \eta^{(r)} + \Delta\eta' \\
\hat{\sigma}_x &= \sigma_x^{(r)} + \Delta\sigma_x' \\
\hat{\sigma}_y &= \sigma_y^{(r)} + \Delta\sigma_y' \\
\hat{\rho} &= \rho^{(r)} + \Delta\rho' \quad .
\end{aligned} \tag{125}
$$

The inverse matrix in (124) is an estimate of the large sample variance-covariance matrix of the maximum likelihood estimates.

5. Numerical Example

Through the use of tables of correlated random normal deviates in Fieller, Lewis, and Pearson (1955), a random sample of size 60 was obtained from the bivariate normal distribution with parameters $\mu_x = 8.0$, $\mu_y = 5.0$, $\sigma_x = 2.0$, $\sigma_y = 4.0$, $\rho = 0.40$.and truncated outside the region $\{(x, y): 4 < x < 10,\ 1 < y < 11\}$. From the data the sample moments about the lower truncation point are:

$m_{1,0} = 3.763000$	$m_{2,1} = 75.185760$	$m_{3,1} = 347.975866$
$m_{0,1} = 4.340000$	$m_{1,2} = 106.272928$	$m_{1,3} = 735.375806$
$m_{2,0} = 16.293527$	$m_{3,0} = 75.784442$	$m_{4,0} = 370.105383$
$m_{0,2} = 26.060160$	$m_{0,3} = 176.389197$	$m_{0,4} = 1304.256990$
$m_{1,1} = 17.230520$		

Starting with the initial set of approximations (89)-(93), four cycles of the functional iterative method followed by one cycle of the Newton-Raphson method were calculated. The summary of the results is given in Tables Ia and Ib. The first line of each table corresponds to the starting vector; the last line of each table corresponds to the final estimates given by the numerical procedure of this chapter. The accuracy of the successive approximations is measured through the differences $\mu'_{i,j} - m_{i,j}$ which equal zero for the values of the estimates which are the exact solutions to the method of moments equations and consequently to the maximum likelihood equations.

TABLE Ia

SUCCESSIVE APPROXIMATIONS TO THE MAXIMUM LIKELIHOOD ESTIMATES

i	$\xi^{(i)}$	$\eta^{(i)}$	$\sigma_x^{(i)}$	$\sigma_y^{(i)}$	$\rho^{(i)}$
0	-1.96659	-.78073	2.25021	4.89601	.38617
1	-1.97291	-.79115	2.23192	4.77854	.41076
2	-1.94191	-.81998	2.25884	4.72193	.42902
3	-1.91946	-.83563	2.27492	4.66964	.44409
4	-1.91309	-.84517	2.28952	4.65364	.45661
5	-1.92053	-.85126	2.28126	4.64628	.46187

TABLE Ib

ACCURACY OF THE SUCCESSIVE APPROXIMATIONS

i	$\mu'^{(i)}_{1,0}-m_{1,0}$	$\mu'^{(i)}_{0,1}-m_{0,1}$	$\mu'^{(i)}_{2,0}-m_{2,0}$	$\mu'^{(i)}_{0,2}-m_{0,2}$	$\mu'^{(i)}_{1,1}-m_{1,1}$
0	.01662	.05570	.16045	.84609	.04143
1	.02435	.01163	.13333	.12106	-.05451
2	.01388	-.00443	.06307	-.05599	-.13009
3	.00818	-.01222	.02681	-.06696	-.12831
4	.00613	.00412	.01146	-.01999	-.04608
5	.00534	.01540	.00643	-.00875	-.01068

From the last line of Table Ia, the maximum likelihood estimates of the original parameters are

$$\hat{\mu}_x = 8.38123 \qquad \hat{\sigma}_x = 2.28126$$

$$\hat{\mu}_y = 4.95519 \qquad \hat{\sigma}_y = 4.64628$$

$$\hat{\rho} = .46187 \quad .$$

An estimate of the large sample variance-covariance matrix of the maximum likelihood estimates is given by

$$\begin{bmatrix} .181933 & .071225 & .217096 & .343759 & .081593 \\ .071225 & .202458 & .006236 & .559844 & .031019 \\ .217096 & .006236 & .387822 & .342483 & .104244 \\ .343759 & .559844 & .342483 & 2.423728 & .250720 \\ .081593 & .031019 & .104244 & .250720 & .073151 \end{bmatrix} .$$

CHAPTER IV

APPROXIMATE PARAMETRIC ESTIMATION IN SPECIAL CASES

This chapter will be concerned with finding approximations to the maximum likelihood estimates of the parameters of

Case 1: a bivariate normal distribution doubly linearly truncated in one variable, and

Case 2: an uncorrelated bivariate normal distribution doubly linearly truncated in one or both variables.

1. Introduction and Background to the Problem

Case 1. By letting $c = -\infty$ and $d = +\infty$ in (1), the density function of a bivariate normal distribution doubly linearly truncated in the x variable is

$$g(x,y) = \frac{\dfrac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\dfrac{1}{2(1-\rho^2)}\left[\left(\dfrac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\dfrac{x-\mu_x}{\sigma_x}\right)\left(\dfrac{y-\mu_y}{\sigma_y}\right) + \left(\dfrac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}}{\Phi\left(\dfrac{b-\mu_x}{\sigma_x}\right) - \Phi\left(\dfrac{a-\mu_x}{\sigma_x}\right)}, \quad \begin{array}{l} a<x<b \\ -\infty<y<+\infty \end{array}$$

$$= 0, \quad \text{elsewhere.} \tag{126}$$

For a sample of size N from (126), the maximum likelihood equations are

$$\frac{\partial \log L}{\partial \mu_x} = \frac{N}{\sigma_x}\,\frac{\phi\left(\dfrac{b-\mu_x}{\sigma_x}\right)-\phi\left(\dfrac{a-\mu_x}{\sigma_x}\right)}{\Phi\left(\dfrac{b-\mu_x}{\sigma_x}\right)-\Phi\left(\dfrac{a-\mu_x}{\sigma_x}\right)} + \frac{1}{\sigma_x(1-\rho^2)}\left[\sum_{i=1}^{N}\left(\frac{X_i-\mu_x}{\sigma_x}\right) - \rho\sum_{i=1}^{N}\left(\frac{Y_i-\mu_y}{\sigma_y}\right)\right] = 0 , \tag{127}$$

$$\frac{\partial \log L}{\partial \mu_y} = \frac{1}{\sigma_y(1-\rho^2)} \left[\sum_{i=1}^{N} \left(\frac{Y_i-\mu_y}{\sigma_y} \right) - \rho \sum_{i=1}^{N} \left(\frac{X_i-\mu_x}{\sigma_x} \right) \right] = 0 \quad , \tag{128}$$

$$\frac{\partial \log L}{\partial \sigma_x} = \frac{N}{\sigma_x} \left[\frac{\left(\frac{b-\mu_x}{\sigma_x}\right)\phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \left(\frac{a-\mu_x}{\sigma_x}\right)\phi\left(\frac{a-\mu_x}{\sigma_x}\right)}{\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \Phi\left(\frac{a-\mu_x}{\sigma_x}\right)} - 1 \right]$$

$$+ \frac{1}{\sigma_x(1-\rho^2)} \left[\sum_{i=1}^{N} \left(\frac{X_i-\mu_x}{\sigma_x} \right)^2 - \rho \sum_{i=1}^{N} \left(\frac{X_i-\mu_x}{\sigma_x} \right)\left(\frac{Y_i-\mu_y}{\sigma_y} \right) \right] = 0 \quad , \tag{129}$$

$$\frac{\partial \log L}{\partial \sigma_y} = \frac{1}{\sigma_y(1-\rho^2)} \left[\sum_{i=1}^{N} \left(\frac{Y_i-\mu_y}{\sigma_y} \right)^2 - \rho \sum_{i=1}^{N} \left(\frac{X_i-\mu_x}{\sigma_x} \right)\left(\frac{Y_i-\mu_y}{\sigma_y} \right) - N(1-\rho^2) \right] = 0 \quad , \tag{130}$$

$$\frac{\partial \log L}{\partial \rho} = \frac{1}{(1-\rho^2)^2} \left[N\rho(1-\rho^2) - \rho \sum_{i=1}^{N} \left(\frac{X_i-\mu_x}{\sigma_x} \right)^2 - \rho \sum_{i=1}^{N} \left(\frac{Y_i-\mu_y}{\sigma_y} \right)^2 \right.$$

$$\left. + (1+\rho^2) \sum_{i=1}^{N} \left(\frac{X_i-\mu_x}{\sigma_x} \right)\left(\frac{Y_i-\mu_y}{\sigma_y} \right) \right] = 0 \quad . \tag{131}$$

From these equations we obtain the following two nonlinear equations, the solution to which are the maximum likelihood estimates $\hat{\mu}_x$ and $\hat{\sigma}_x$ of the parameters μ_x and σ_x .

$$\frac{1}{N} \sum_{i=1}^{N} \left(\frac{X_i-\mu_x}{\sigma_x} \right) = \frac{\phi\left(\frac{a-\mu_x}{\sigma_x}\right) - \phi\left(\frac{b-\mu_x}{\sigma_x}\right)}{\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \Phi\left(\frac{a-\mu_x}{\sigma_x}\right)} \tag{132}$$

and

$$\frac{1}{N} \sum_{i=1}^{N} \left(\frac{X_i-\mu_x}{\sigma_x} \right)^2 = 1 + \frac{\left(\frac{a-\mu_x}{\sigma_x}\right)\phi\left(\frac{a-\mu_x}{\sigma_x}\right) - \left(\frac{b-\mu_x}{\sigma_x}\right)\phi\left(\frac{b-\mu_x}{\sigma_x}\right)}{\Phi\left(\frac{b-\mu_x}{\sigma_x}\right) - \Phi\left(\frac{a-\mu_x}{\sigma_x}\right)} \quad . \tag{133}$$

Once $\hat{\mu}_x$ and $\hat{\sigma}_x$ have been found, the three remaining maximum likelihood estimates $\hat{\mu}_y$, $\hat{\sigma}_y$, and $\hat{\rho}$ may be found as follows. From (128)

$$(\bar{X}-\hat{\mu}_x)\rho\sigma_y + \hat{\sigma}_x\mu_y = \hat{\sigma}_x\bar{Y} \quad , \tag{134}$$

and from (130) and (131)

$$\frac{1}{N}\sum_{i=1}^{N} (X_i-\hat{\mu}_x)^2\rho\sigma_y + \hat{\sigma}_x(\bar{X}-\hat{\mu}_x)\mu_y = \frac{1}{N}\hat{\sigma}_x \sum_{i=1}^{N} Y_i(X_i-\hat{\mu}_x) \quad . \tag{135}$$

Solving (134) and (135) simultaneously for $\rho\sigma_y$ and μ_y yield

$$\widehat{\rho\sigma_y} = \hat{\sigma}_x \frac{S_{x,y}}{S_{x,x}} \tag{136}$$

and

$$\hat{\mu}_y = \bar{Y} - (\bar{X}-\hat{\mu}_x)\frac{S_{x,y}}{S_{x,x}} \quad ; \tag{137}$$

where

$$S_{x,x} = \frac{1}{N}\sum_{i=1}^{N} (X_i-\bar{X})^2 \quad ,$$

$$S_{x,y} = \frac{1}{N}\sum_{i=1}^{N} (X_i-\bar{X})(Y_i-\bar{Y}) \quad ,$$

and

$$S_{y,y} = \frac{1}{N}\sum_{i=1}^{N} (Y_i-\bar{Y})^2 \quad .$$

Finally, upon adding (135) multiplied by $-N/\sigma_x^2\sigma_y$ and (130) multiplied by $1/\rho$ and using (136) and (137) in the resulting equation, we obtain

$$\hat{\sigma}_y = \sqrt{S_{y,y} + \left(\hat{\sigma}_x^2 - S_{x,x}\right)\left(\frac{S_{x,y}}{S_{x,x}}\right)^2} \tag{138}$$

and

$$\hat{\rho} = \frac{\hat{\sigma}_x \, s_{x,y}}{\hat{\sigma}_y \, s_{x,x}} \quad . \tag{139}$$

Thus the original problem of solving (127)-(131) reduces to the problem of solving (132) and (133). It can be shown that (132) and (133) are equivalent to the two maximum likelihood equations for estimating the parameters μ and σ in a doubly truncated univariate normal distribution whose density is

$$g(t) = \frac{\frac{1}{\sigma}\,\phi\left(\frac{t-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \quad , \quad a < t < b$$

$$= 0 \quad , \quad \text{elsewhere} \quad . \tag{140}$$

This estimation problem was considered by Cohen (1950) who suggested that the modified Newton-Raphson method be used to solve the two estimating equations. In a later paper Cohen (1957) gave a chart from which a graphical solution to the two estimating equations could be obtained. However, the graphical solution has restricted accuracy, so if more precise results are desired, then the graphical solution serves only as a first approximation to be improved through iteration.

In an attempt to avoid solving nonlinear equations, Tiku (1968) made locally linear approximations to the ratios $\frac{\phi(u)}{\Phi(v)-\Phi(u)}$ and $\frac{\phi(v)}{\Phi(v)-\Phi(u)}$, where $u < v$. However, the graphs of $z = \frac{\phi(u)}{\Phi(v)-\Phi(u)}$, $u < v$ with u fixed, as well as $z = \frac{\phi(v)}{\Phi(v)-\Phi(u)}$, $u < v$ with u fixed, are similar to the positive branch of a rectangular hyperbola $z(v-u) = \text{constant} > 0$. Clearly there are intervals over which a straight line might not be a particularly accurate approximation. Furthermore, the computed coefficients in the linear approximations are determined through formulae which are derived under the assumption of a large sample size. Consequently, we shall seek alternate approximations to the ratios which appear in the right-hand sides of (132) and (133).

2. Differential Relationship between $S(u,v)$ and $T(u,v)$

Let

$$S(u, v) = \frac{\phi(u)-\phi(v)}{\Phi(v)-\Phi(u)} \tag{141}$$

and

$$T(u, v) = \frac{u\phi(u)-v\phi(v)}{\Phi(v)-\Phi(u)} , \tag{142}$$

where

$$-\infty < u < v < +\infty .$$

Partial differentiation of $S(u, v)$ with respect to u yields

$$\frac{d\phi(u)}{du} - \frac{\partial S(u, v)}{\partial u}[\Phi(v)-\Phi(u)] + S(u, v)\phi(u) = 0$$

which may be written

$$\frac{u\phi(u)}{\Phi(v)-\Phi(u)} = -\frac{\partial S(u, v)}{\partial u} + \frac{S(u, v)\phi(u)}{\Phi(v)-\Phi(u)} . \tag{143}$$

Partial differentiation of $S(u, v)$ with respect to v yields

$$-\frac{d\phi(v)}{dv} - \frac{\partial S(u, v)}{\partial v}[\Phi(v)-\Phi(u)] - S(u, v)\phi(v) = 0$$

which may be written

$$\frac{-v\phi(v)}{\Phi(v)-\Phi(u)} = -\frac{\partial S(u, v)}{\partial v} - \frac{S(u, v)\phi(v)}{\Phi(v)-\Phi(u)} \quad . \tag{144}$$

Adding (143) and (144), we obtain the relationship

$$T(u, v) + \frac{\partial S(u, v)}{\partial u} + \frac{\partial S(u, v)}{\partial v} = [S(u, v)]^2 \tag{145}$$

which will be useful in achieving the final approximations.

3. Derivation of Bounds for S(u,v) and T(u,v)

In an attempt to gain some insight into the behavior of the functions S(u, v) and T(u,vv), we next establish bounds on these functions. Using ordinary methods of the calculus for finding maximum and minimum values of a function of one variable and noting that

$$\lim_{u \to v} S(u, v) = v \quad , \quad \text{for fixed } v \tag{146}$$

and

$$\lim_{v \to u} S(u, v) = u \quad , \quad \text{for fixed } u \quad ; \tag{147}$$

we find that for fixed v and every u such that $-\infty < u < v < +\infty$,

$$S(u, v) < v \quad , \tag{148}$$

and for fixed u and every v such that $-\infty < u < v < +\infty$,

$$u < S(u, v) \quad . \tag{149}$$

Consequently,

$$u < S(u, v) < v \tag{150}$$

for every u, v such that $-\infty < u < v < +\infty$.

As a sidelight, we note that

$$\lim_{v \to +\infty} S(u, v) = \frac{1}{M(u)} ,$$

where $M(u) = \frac{1-\Phi(u)}{\phi(u)}$, i.e., Mill's ratio. For $v \to +\infty$ in (150)

$$u < \frac{1}{M(u)} , \quad -\infty < u < +\infty .$$

Since $M(u) > 0$, then

$$M(u) < \frac{1}{u} , \quad u > 0$$

which establishes in a different way the upper bound on Mill's ratio given by Gordon (1941).

Squaring both sides of (132) and subtracting from (133) yields

$$T(u, v) + 1 - [S(u, v)]^2 \geq 0 , \quad -\infty < u < v < +\infty . \tag{151}$$

Thus

$$T(u, v) \geq -1 , \quad -\infty < u < v < +\infty \tag{152}$$

with equality holding only when $|u| = |v|$.

An improvement on the lower bound for $T(u, v)$ in (152) can be found through the use of a variation of Wirtinger's inequality given in Shisha (1967, pp. 91-94):

$$\int_u^v w^2(t)\,dt \leq \frac{4}{\pi^2} \max[(\tau-u)^2, (v-\tau)^2] \int_u^v [w'(t)]^2 dt \tag{153}$$

holds for any real-valued function $w(t)$ continuously differentiable on the finite closed interval $u \le t \le v$, and τ is a real number such that $u \le \tau \le v$ and

$$0 \le (v-u)w^2(\tau) - 2w(\tau)\int_u^v w(t)\,dt \quad .$$

For $w(t) = e^{-t^2/4}$, then $w'(t) = -1/2\, te^{-t^2/4}$ and

$$\frac{\int_u^v e^{-t^2/2}\,dt}{\int_u^v \frac{1}{4}t^2 e^{-t^2/2}\,dt} \le \frac{4}{\pi^2}\max[(\tau-u)^2,\ (v-\tau)^2] \quad , \tag{154}$$

where $u \le \tau \le v$ and

$$0 \le (v-u)e^{-\tau^2/2} - 2e^{-\tau^2/4}\int_u^v e^{-t^2/4}\,dt \quad .$$

From the left-hand side of (154) we see that

$$\frac{\int_u^v e^{-t^2/2}\,dt}{\int_u^v \frac{1}{4}t^2 e^{-t^2/2}\,dt} = 4\,\frac{\Phi(v)-\Phi(u)}{u\phi(u)-v\phi(v)+\Phi(v)-\Phi(u)} = 4\,\frac{1}{T(u,\ v)+1} \quad .$$

Thus, for all u, v such that $-\infty < u < v < +\infty$ and $|u| \ne |v|$,

$$T(u,\ v) \ge -1 + \frac{\pi^2}{\max[(\tau-u)^2,\ (v-\tau)^2]} \quad , \tag{155}$$

where $u \le \tau \le v$ and

$$\phi\left(\frac{\tau}{\sqrt{2}}\right) \ge 2\,\frac{\Phi\left(\frac{v}{\sqrt{2}}\right)-\Phi\left(\frac{u}{\sqrt{2}}\right)}{\frac{v}{\sqrt{2}}-\frac{u}{\sqrt{2}}} \quad .$$

Since $\max[(\tau-u)^2, (v-\tau)^2] = \frac{1}{4}[(v-u) + |u+v-2\tau|]^2$, then

$$T(u, v) \begin{cases} \geq -1 + \dfrac{4\pi^2}{(v-u)^2 + 2(v-u)|u+v-2\tau| + (u+v-2\tau)^2}, & -\infty < u < v < +\infty \\ & \text{and } |u| \neq |v| \\ \geq -11 , \; -\infty < u < v < +\infty \text{ and } |u| = |v| , & \end{cases} \tag{156}$$

where $u \leq \tau \leq v$ and

$$\phi\left(\frac{\tau}{\sqrt{2}}\right) \geq 2 \, \frac{\Phi\left(\frac{v}{\sqrt{2}}\right) - \Phi\left(\frac{u}{\sqrt{2}}\right)}{\frac{v}{\sqrt{2}} - \frac{u}{\sqrt{2}}} .$$

4. Series Representations for S(u,v) and T(u,v)

Note that in the lower bound for $T(u, v)$ given by (156),

$$\frac{4\pi^2}{(v-u)^2 + 2(v-u)|u+v-2\tau| + (u+v-2\tau)^2}$$

may be written in the form

$$4\pi^2[1 + d_1(u, v)(v-u) + d_2(u, v)(v-u)^2 + \cdots] .$$

This form suggests a possible series representation of $T(u, v)$ and consequently $S(u, v)$.

Let z_0 represent a fixed complex number. The two functions of a complex variable z given by $\frac{\phi(z)}{\Phi(z)-\Phi(z_0)}$ and $\frac{\phi(z_0)}{\Phi(z)-\Phi(z_0)}$ have a simple pole at the isolated singular point $z = z_0$, and the residue of each function at the isolated singular point has value unity [Churchill (1960), pp. 153-161]. Thus the two Laurentian series which represent $\frac{\phi(z)}{\Phi(z)-\Phi(z_0)}$ and $\frac{\phi(z_0)}{\Phi(z)-\Phi(z_0)}$ when $|z-z_0| > 0$ are given by

$$\frac{\phi(z)}{\Phi(z)-\Phi(z_0)} = \frac{1}{(z-z_0)} + \sum_{n=0}^{\infty} a_n(z_0)(z-z_0)^n \tag{157}$$

and

$$\frac{\phi(z_0)}{\Phi(z)-\Phi(z_0)} = \frac{1}{(z-z_0)} + \sum_{n=0}^{\infty} b_n(z_0)(z-z_0)^n \quad , \tag{158}$$

where the complex-valued functions $a_n(z_0)$ and $b_n(z_0)$ may be found through contour integration.

Consider the reduction of (157) and (158) to the real line with $z_0 = u$, a fixed real number, and $z = v$, a real variable. Since the subtraction of two power series term by term is valid within their common region of convergence, then for fixed u and every $v > u$,

$$S(u, v) = \frac{\phi(u)-\phi(v)}{\Phi(v)-\Phi(u)} = \sum_{n=0}^{\infty} [b_n(u) - a_n(u)](v-u)^n \quad . \tag{159}$$

Upon multiplying both sides of the reduction of (158) to the real line by u and both sides of the reduction of (157) to the real line by $(-v)$ and adding the two resulting power series, we obtain for fixed u and every $v > u$,

$$T(u, v) = \frac{u\phi(u)-v\phi(v)}{\Phi(v)-\Phi(u)} = -1 + \sum_{n=0}^{\infty} [ub_n(u) - va_n(u)](v-u)^n \quad . \tag{160}$$

However, $\lim_{v \to u} S(u, v) = u$, so that $b_0(u)-a_0(u) = u$. Thus (159) and (160) may be written as

$$S(u, v) = u + c_1(u)(v-u) + c_2(u)(v-u)^2 + c_3(u)(v-u)^3 + \cdots \tag{161}$$

and

$$T(u,\ v) = -1 + u^2 + [-a_0(u) + c_1(u)u](v-u) + [-a_1(u) + c_2(u)u](v-u)^2$$

$$+ [-a_2(u) + c_3(u)u](v-u)^3 + \cdots\ , \tag{162}$$

where $c_i(u) = b_i(u) - a_i(u)$, $i = 1, 2, 3, \cdots$.

5. Derivation of Approximate Estimating Equations

For various fixed values of u, graphs of $S(u,\ v)$ were obtained for each of two regions in the uv-plane:

$$\text{region A} = \{(u,\ v):\ -3 < u < v < 3,\ |u| \geq |v|\}$$

and (163)

$$\text{region B} = \{(u,\ v):\ -3 < u < v < 3,\ |u| \leq |v|\}\ .$$

The graphs within each of the regions were very similar, and the shape of the curves suggested that perhaps a second-order or third-order linear model in u and v be fitted in each region to $S(u,\ v)$. We note that since u and v are values assumed by a standard normal random variable, then if we let $u = \frac{a-\mu}{\sigma}$ and $v = \frac{b-\mu}{\sigma}$ with $b > a$, the probability is .005 that a point $(u,\ v)$ falls outside the union of the two regions. Furthermore, from the graph of Cohen (1957) we see that

$$\left(\frac{a-\mu}{\sigma}\ ,\ \frac{b-\mu}{\sigma}\right) \varepsilon \text{ region A if and only if } .5 \leq \frac{\bar{X}-a}{b-a} < 1 \tag{164}$$

and

$$\left(\frac{a-\mu}{\sigma}\ ,\ \frac{b-\mu}{\sigma}\right) \varepsilon \text{ region B if and only if } 0 < \frac{\bar{X}-a}{b-a} \leq .5\ . \tag{165}$$

Using multiple linear regression, a second-order linear model suggested by (161)

$$S(u,\ v) = u + \beta_1(v-u) + \beta_2 u^2 + \beta_3 uv + (-\beta_2-\beta_3)v^2 \tag{166}$$

was fitted to points (u, v) chosen in .04 probabilistic steps in each of two regions. The following table gives the number of points fitted, the maximum absolute residual as well as the average absolute residual for the fitted points, and the values of β_1, β_2, and β_3 for each of the two regions.

TABLE II

RESULTS OF SECOND-ORDER LINEAR MODEL FIT OF S(u, v)

Region	Number pf Points Fitted	max. $\lvert\varepsilon\rvert$	ave. $\lvert\varepsilon\rvert$	β_1	β_2	β_3
A	193	.05	less than .01	.45432	.08872	-.04442
B	156	.05	less than .01	.51802	.05714	.02026

When the approximation to S(u, v) in (166) is put into maximum likelihood equation (132), we obtain upon simplification

$$\mu_x = \frac{\left(\frac{\bar{X}-a}{b-a} - \beta_1\right)\sigma_x + (2\beta_2+\beta_3)b - \beta_2(b-a)}{2\beta_2 + \beta_3} \quad . \tag{167}$$

Using multiple linear regression, a third-order linear model suggested by (162)

$$T(u,v) = -1 + u^2 + \gamma_1(v-u) + \gamma_2 u^2 + \gamma_3 uv + (-\gamma_2-\gamma_3)v^2 + \gamma_4 u(v-u)^2 + \gamma_5(v-u)^3 \tag{168}$$

was fitted to points (u, v) chosen in .04 probabilistic steps in each of the two regions. Table III gives the number of points fitted, the maximum absolute residual as well as the average absolute residual for the fitted points, and the values of γ_1, γ_2, γ_3, and γ_4 for each of the two regions.

TABLE III

RESULTS OF THIRD-ORDER LINEAR MODEL FIT OF T(u, v)

Region	Number of Points Fitted	max.$\lvert\varepsilon\rvert$	ave.$\lvert\varepsilon\rvert$	γ_1	γ_2	γ_3	γ_4	γ_5
A	181	.16	less than .05	.25724	-.98958	.56142	.02099	.02636
B	144	.05	less than .01	.06690	-.59493	.38300	.03570	.01082

When the approximation to T(u, v) in (168) is put into maximum likelihood equation (133), we obtain upon simplification

$$\mu_x = \frac{-\gamma_1\sigma_x^2 + \left[\dfrac{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - a^2}{b-a} + (2\gamma_2+\gamma_3)b - \gamma_2(b-a)\right]\sigma_x - \gamma_4 a(b-a) - \gamma_5(b-a)^2}{\left(2\,\dfrac{\bar{X}-a}{b-a} + 2\gamma_2+\gamma_3\right)\sigma_x - \gamma_4(b-a)} \ . \tag{169}$$

Equating the two expressions for μ_x in (167) and (169) yields the following quadratic equation in σ_x:

$$w_1\sigma_x^2 + w_2\sigma_x + w_3 = 0 \quad , \tag{170}$$

where

$$w_1 = \left(\frac{\bar{X}-a}{b-a} - \beta_1\right)\left(2\,\frac{\bar{X}-a}{b-a} + 2\gamma_2+\gamma_3\right) + \gamma_1(2\beta_2+\beta_3) \quad ,$$

$$w_2 = -(b-a)\left[\gamma_4\left(\frac{\bar{X}-a}{b-a} - \beta_1\right) + \beta_2\left(2\,\frac{\bar{X}-a}{b-a} + 2\gamma_2+\gamma_3\right) - (1+\gamma_2)(2\beta_2+\beta_3)\right]$$
$$- \frac{(2\beta_2+\beta_3)}{b-a}\,[S_{x,x} + (b-\bar{X})^2] \quad ,$$

and

$$w_3 = (b-a)^2[\gamma_4\beta_2 - (\gamma_4 - \gamma_5)(2\beta_2 + \beta_3)] \quad .$$

From (164), (165), Table II, and Table III we see that $w_3 > 0$ for region A and $w_3 < 0$ for region B; whereas, depending on the value of $\bar{X}$, w_1 may be positive or negative in either region. Thus the quadratic equation in (170) may have zero, one, or two positive roots. If there is only one positive root, then this root will be the approximate maximum likelihood estimate of σ_x, say $\tilde{\tilde{\sigma}}_x$. The approximate maximum likelihood estimate of μ_x, say $\tilde{\mu}_x$, may be found from (167).

In the event that (170) has zero or two positive roots, the following alternate procedure is used. Using multiple linear regression, a second-order linear model suggested by (162)

$$T(u, v) = -1 + u^2 + \gamma_1'(v-u) + \gamma_2'u^2 + \gamma_3'uv + (-\gamma_2'-\gamma_3')v^2 \tag{171}$$

was fitted to points (u, v) chosen in .04 probabilistic steps in each of the two regions. Table IV gives the number of points fitted, the maximum absolute residual as well as the average absolute residual for the fitted points, and the values of γ_1', γ_2', and γ_3' for each of the two regions.

TABLE IV

RESULTS OF SECOND-ORDER LINEAR MODEL FIT OF T(u, v)

Region	Number of Points Fitted	max. $\lvert\varepsilon\rvert$	ave. $\lvert\varepsilon\rvert$	γ_1'	γ_2'	γ_3'
A	181	.22	less than .07	.19749	-.94739	.41711
B	144	.06	less than .01	.08442	-.61059	.42211

When the approximation to T(u, v) in (171) is put into maximum likelihood equation (133), we obtain upon simplification

$$\mu_x = \frac{-\gamma_1'\sigma_x + \dfrac{\frac{1}{N}\sum_{i=1}^{N} X_i^2 - a^2}{b-a} + (2\gamma_2' + \gamma_3')b - \gamma_2'(b-a)}{\dfrac{\bar{X}-a}{b-a} + 2\gamma_2' + \gamma_3'} \quad . \tag{172}$$

Equating the two expressions for μ_x in (167) and (172) yields

$$\tilde{\sigma}_x = \frac{(b-a)\left[\beta_2\left(2\,\dfrac{\bar{X}-a}{b-a} + 2\gamma_2'+\gamma_3'\right) - (1+\gamma_2')(2\beta_2+\beta_3)\right] + \dfrac{(2\beta_2+\beta_3)}{b-a}\left[s_{x,x}+(b-\bar{X})^2\right]}{\left(\dfrac{\bar{X}-a}{b-a} - \beta_1\right)\left(2\,\dfrac{\bar{X}-a}{b-a} + 2\gamma_2'+\gamma_3'\right) + \gamma_1'(2\beta_2+\beta_3)} \ . \tag{173}$$

It should be noted that another alternate expression for $\tilde{\sigma}_x$ may be found through the use of the approximation to $S(u, v)$ in (166) and the differential relationship between $S(u, v)$ and $T(u, v)$ in (145). Since

$$S(u,\ v) = u + \beta_1(v-u) + \beta_2 u^2 + \beta_3 uv + (-\beta_2-\beta_3)v^2 \quad ,$$

then

$$\frac{\partial S(u,\ v)}{\partial u} + \frac{\partial S(u,\ v)}{\partial v} = 1 - (2\beta_2 + \beta_3)(v - u) \quad .$$

Hence from (145)

$$T(u,\ v) = -1 + (2\beta_2+\beta_3)(v-u) + [S(u,\ v)]^2 \quad , \tag{174}$$

which is a form very similar to (162). When this approximation to $T(u, v)$ is put into maximum likelihood equation (133), we obtain

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i-\mu_x}{\sigma_x}\right)^2 = (2\beta_2+\beta_3)\left(\frac{b-a}{\sigma_x}\right) + \left[S\left(\frac{a-\mu_x}{\sigma_x}\ ,\ \frac{b-\mu_x}{\sigma_x}\right)\right]^2 \quad . \tag{175}$$

But from (132),

$$\left[S\left(\frac{a-\mu_x}{\sigma_x}, \frac{b-\mu_x}{\sigma_x}\right)\right]^2 = \frac{1}{N^2}\left[\sum_{i=1}^{N}\left(\frac{X_i-\mu_x}{\sigma_x}\right)\right]^2 ,$$

hence we may write (175) as

$$\tilde{\sigma}_x = \frac{S_{x,x}}{(2\beta_2+\beta_3)(b-a)} . \tag{176}$$

Thus, in the alternate procedure, $\tilde{\sigma}_x$ is found from (173) or (176), then $\tilde{\mu}_x$ is found from (167)

Case 2. By letting $\rho = 0$ in (1), the density function of an uncorrelated bivariate normal distribution doubly linearly truncated in both x and y is

$$g(x, y) = \frac{\phi\left(\frac{x-\mu_x}{\sigma_x}\right)\phi\left(\frac{y-\mu_y}{\sigma_y}\right)}{\left[\Phi\left(\frac{b-\mu_x}{\sigma_x}\right)-\Phi\left(\frac{a-\mu_x}{\sigma_x}\right)\right]\left[\Phi\left(\frac{d-\mu_y}{\sigma_y}\right)-\Phi\left(\frac{c-\mu_y}{\sigma_y}\right)\right]} , \quad \begin{array}{l} a < x < b \\ c < y < d \end{array}$$

$$= 0 \quad , \quad \text{elsewhere} . \tag{177}$$

For a sample of size N from (177), the maximum likelihood equations for estimating μ_x and σ_x are the same as (132) and (133). The maximum likelihood equations for estimating μ_y and σ_y are the same as (132) and (133) with X_i, μ_x, σ_x, a, and b replaced by Y_i, μ_y, σ_y, c, and d, respectively. Thus the approximations used for Case 1. may also be used for Case 2.

6. Numerical Example

To illustrate the use of the approximations, we use the data from Des Raj (1953) in which a sample of size 74 was drawn from a bivariate normal distribution doubly linearly truncated in y with parameters $\mu_x = 4$, $\sigma_x = 2$, $\rho = 0$, $\mu_y = 10$, $\sigma_y = 5$, the truncation points being $c = 5.0$ and $d = 17.5$. For the sample taken: $\bar{X} = 3.904297$, $\bar{Y} = 10.584797$, $S_{x,x} = 4.00654$, $S_{y,y} = 10.43671$, $S_{x,y} = -1.21174$.

After several cycles of Newton's iterative method, Des Raj obtained as the maximum likelihood estimates

$$\hat{\mu}_y = 9.5835 \qquad \hat{\mu}_x = 4.0205 \qquad \hat{\rho} = -.29096 \ .$$
$$\hat{\sigma}_y = 5.15 \qquad \hat{\sigma}_x = 2.055$$

Using the linear approximations of Tiku (1968), we obtain

$$\tilde{\mu}_y = 10.17639 \qquad \tilde{\mu}_x = 3.95172 \qquad \tilde{\rho} = -.23349 \ .$$
$$\tilde{\sigma}_y = 4.06655 \qquad \tilde{\sigma}_x = 2.02207$$

Since $\frac{\bar{Y}-c}{d-c} = .44678$, then by (165), the coefficients for the approximations in region B shall be used. Thus (170) becomes

$$.00282\sigma_y^2 + .02406\sigma_y - .20422 = 0 \ .$$

The only positive root to this quadratic equation is $\tilde{\sigma}_y = 5.25532$. From (167) we obtain $\tilde{\mu}_y = 9.40843$. The remaining approximate estimates are found from (137), (138), and (139). Thus using the approximations of this chapter, we have as approximate estimates

$$\tilde{\mu}_y = 99.40843 \qquad \tilde{\mu}_x = 4.02927 \qquad \tilde{\rho} = -.29638 \ ,$$
$$\tilde{\sigma}_y = 5.25532 \qquad \tilde{\sigma}_x = 2.05871$$

which compare very favorably with the actual maximum likelihood estimates.

LIST OF REFERENCES

[1] Birnbaum, Z. W. and Meyer, P. L. (1953). "On the effect of truncation in some or all coordinates of a multinormal population." Journal of the Indian Society of Agricultural Statistics, 5, pp. 17-28.

[2] Churchill, R. V. (1960). Complex Variables and Applications. 2nd edition, McGraw-Hill Book Company, New York.

[3] Cohen, A. C. Jr. (1950). "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples." Annals of Mathematical Statistics, 21, pp. 557-69.

[4] ________. (1955). "Restriction and selection in samples from bivariate normal distributions." Journal of the American Statistical Association, 50, pp. 884-93.

[5] ________. (1957). "On the solution of estimating equations for truncated and censored samples from normal populations." Biometrika, 44, pp. 225-36.

[6] Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton.

[7] Federer, W. T. (1963). "Procedures and designs useful for screening material in selection and allocation, with a bibliography." Biometrics, 19, pp. 553-87.

[8] Fieller, E. C., Lewis, T., and Pearson, E. S. (1955). "Correlated Random Normal Deviates," Tracts for Computers, No. 26. Cambridge University Press, London.

[9] Geary, R. C. (1942). "The estimation of many parameters." Journal of the Royal Statistical Society, Series A, 105, pp. 213-17.

[10] Gordon, R. D. (1941). "Values of Mill's ratio of area to bounding ordinate and of the normal probability integral for large values of the argument." Annals of Mathematical Statistics, 12, pp. 364-66.

[11] Hotelling, H. (1948). "Abstracts of Madison Meeting." Annals of Mathematical Statistics, 19, p. 596.

[12] Huzurbazar, V. S. (1949). "On a property of distributions admitting sufficient statistics." Biometrika, 36, pp. 71-74.

[13] Isaacson, E. and Keller, H. B. (1966). Analysis of Numerical Methods. John Wiley and Sons, Inc., New York.

[14] Jaiswal, M. C. and Khatri, C. G. (1967). "Estimation of parameters for the selected samples from bivariate normal populations." Metron, 26, pp. 326-33.

[15] Khatri, C. G. and Jaiswal, M. C. (1963). "Estimation of parameters of a truncated bivariate normal distribution." Journal of the American Statistical Association, 58, pp. 519-26.

[16] National Bureau of Standards. (1953). Tables of Normal Probability Functions. Applied Mathematics Series 23, U.S. Government Printing Office, Washington, D. C.

[17] ________. (1959). Tables of the Bivariate Normal Distribution Function and Related Functions. Applied Mathematics Series 50, U.S. Government Printing Office, Washington, D. C.

[18] Neyman, J. and Pearson, E. S. (1936). "Sufficient statistics and uniformly most powerful tests of statistical hypotheses." Statistical Research Memoirs, 1, pp. 113-37.

[19] Raj, Des. (1953). "On estimating the parameters of bivariate normal populations from doubly and singly linearly truncated samples." Sankhyā, 12, pp. 277-90.

[20] Rosenbaum, S. (1961). "Moments of a truncated bivariate normal distribution." Journal of the Royal Statistical Society, Series B, 23, pp. 405-08.

[21] Shisha, O. (Editor). (1967). Inequalities, Proceedings of a Symposium held at Wright-Patterson Air Force Base, Ohio, August 19-27, 1965. Academic Press, New York.

[22] Singh, N. (1960). "Estimation of parameters of a multi-variate normal population from truncated and censored samples." Journal of the Royal Statistical Society, Series B, 22, pp. 307-11.

[23] Smith, W. L. (1957). "A note on truncation and sufficient statistics." Annals of Mathematical Statistics, 28, pp. 247-52.

[24] Tallis, G. M. (1961). "The moment generating function of the truncated multi-normal distribution." Journal of the Royal Statistical Society, Series B, 23, pp. 223-29.

[25] Tiku, M. L. (1968). "Estimating the parameters of normal and logistic distributions from censored samples." Australian Journal of Statistics, 10, pp. 64-74.

[26] Weiler, H. (1959). "Means and standard deviations of a truncated normal bivariate distribution." Australian Journal of Statistics, 1, pp. 73-81.

[27] Wilks, S. S. (1962). Mathematical Statistics. John Wiley and Sons, Inc., New York.

[28] Williams, J. M. and Weiler, H. (1964). "Further charts for the means of truncated normal bivariate distributions." Australian Journal of Statistics, 6, pp. 117-29.