JOURNAL OF MULTIVARIATE ANALYSIS 9, 304–313 (1979)

On the Maximum Likelihood Estimation of a Linear Structural Relationship When the Intercept Is Known

LAI K. CHAN* AND TAK K. MAK

The University of Western Ontario

Communicated by P. A. P. Moran

This paper considers the Maximum Likelihood (ML) estimation of the five parameters of a linear structural relationship $y = \alpha + \beta x$ when α is known. The parameters are β, the two variances of observation errors on x and y, the mean and variance of x. When the ML estimates of the parameters cannot be obtained by solving a simple simultaneous system of five equations, they are found by maximizing the likelihood function directly. Some asymptotic properties of the estimates are also obtained.

1. INTRODUCTION

Consider a bivariate random variable (x, y) satisfying the linear relation $y = \alpha + \beta x$, β being unknown and to be estimated. Suppose x and y cannot be observed exactly, but instead we observe $\xi = x + \delta$ and $\eta = y + \epsilon$, where the errors δ and ϵ have zero means and unknown variances σ_δ^2 and σ_ϵ^2 respectively.

If α is unknown, x, δ and ϵ are independent and normally distributed, and x has unknown mean μ and variance σ^2, then β is not identifiable and cannot be estimated consistently from n independent observations (ξ_i, η_i), $i = 1,..., n$ (cf. Kendall and Stuart 1973, ch. 29; Moran 1971a). When σ_δ^2 (or σ_ϵ^2) or $\sigma_\epsilon^2/\sigma_\delta^2$ is known, β becomes identifiable and Maximum Likelihood (ML) estimates in these cases have been obtained (Lindley 1947; Birch 1964).

If α is known and μ is only known to be non-zero, then β also becomes identifiable and can be estimated consistently by $(\bar{\eta}. - \alpha)/\bar{\xi}.$, where $\bar{\eta}. = \sum \eta_i/n$,

Received May 1978; revised October 18, 1978.

AMS 1970 subject classifications: Primary 62F10; Secondary 62F25, 62H05.

Key words and phrases: linear structural relationship, known intercept, maximum likelihood estimation.

* Research supported by the National Research Council of Canada Grant No. A3722.

0047-259X/79/020304–10$02.00/0

$\bar{\xi}. = \sum \xi_i/n$, and $\sum$ denotes summation from $i = 1$ to n. Without loss of generality, let α be zero. The model then becomes

$$\xi = x + \delta,$$

$$\eta = \beta x + \epsilon.$$

The ML estimate $(\hat{\mu}, \hat{\beta}, \hat{\sigma}^2, \hat{\sigma}_\delta^2, \hat{\sigma}_\epsilon^2)$ of $(\mu, \beta, \sigma^2, \sigma_\delta^2, \sigma_\epsilon^2)$ is well-known when $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ are non-negative. However, when one of the variance estimates is negative no full solution to the estimation of $(\mu, \beta, \sigma^2, \sigma_\delta^2, \sigma_\epsilon^2)$ was available, as pointed out by Moran (1971a, p. 252) and Zellner (1971, p. 130). Here we provide a complete ML solution and also obtain its asymptotic properties.

2. Maximum Likelihood Solution

The model considered here is

$$\xi_i = x_i + \delta_i,$$

$$\eta_i = \beta x_i + \epsilon_i, \qquad i = 1,\ldots, n,$$

where the $(x_i, \delta_i, \epsilon_i)$ are independent and identically distributed normal variates, and x_i, δ_i and ϵ_i are mutually independent with mean $(\mu, 0, 0)$ and variances σ^2, σ_δ^2 and σ_ϵ^2, respectively. We further assume that the distribution of (ξ_i, η_i) is non-singular, so that each (ξ_i, η_i) has the bivariate normal distribution with mean $(\mu, \beta\mu)$ and positive definite covariance matrix

$$\mathbf{V} = \begin{pmatrix} \sigma^2 + \sigma_\delta^2 & \beta\sigma^2 \\ \beta\sigma^2 & \beta^2\sigma^2 + \sigma_\epsilon^2 \end{pmatrix}.$$

The positive definiteness of $\mathbf{V}$ is equivalent to the condition that at most one of σ^2, σ_δ^2 and σ_ϵ^2 is zero and $\beta \neq 0$ if $\sigma_\epsilon^2 = 0$. The likelihood function L for (ξ_i, η_i), $i = 1,\ldots, n$, is thus the product of the bivariate normal probability functions.

Let $m_{\xi\eta} = \sum \xi_i\eta_i/n$, $m_{\xi\xi} = \sum \xi_i^2/n$, $m_{\eta\eta} = \sum \eta_i^2/n$. Then $(\bar{\xi}., \bar{\eta}., m_{\xi\xi} - \bar{\xi}.^2, m_{\eta\eta} - \bar{\eta}.^2, m\xi_\eta - \bar{\xi}.\bar{\eta}.) = (\bar{\xi}., \bar{\eta}., s_{\xi\xi}, s_{\eta\eta}, s_{\xi\eta})$ is the unique ML estimate of the transformed parameters $E(\xi) = \mu$, $E(\eta) = \beta\mu$, $\text{Var}(\xi) = \sigma^2 + \sigma_\delta^2$, $\text{Var}(\eta) = \beta^2\sigma^2 + \sigma_\epsilon^2$, $\text{Cov}(\xi, \eta) = \beta\sigma^2$ when L is considered as a function of the transformed parameters. The transformation is one-to-one if $\beta \neq 0$. Consider first the ML estimation with the restriction that $\beta \neq 0$. It will be shown later that the probability that L is maximized at a point with $\beta = 0$ is zero. By lemma 3.2.3 of Anderson (1958), the solution $(\hat{\mu}, \hat{\beta}, \hat{\sigma}^2, \hat{\sigma}_\delta^2, \hat{\sigma}_\epsilon^2)$ for the equations

$$\bar{\xi}. = \mu, \qquad \bar{\eta}. = \beta\mu,$$

$$s_{\xi\xi} = \sigma^2 + \sigma_\delta^2, \qquad s_{\eta\eta} = \beta^2\sigma^2 + \sigma_\epsilon^2, \qquad s_{\xi\eta} = \beta\sigma^2$$

maximizes L on $\Omega = \{(\mu, \beta, \sigma^2, \sigma_\delta^2, \sigma_\epsilon^2): \mu \neq 0, \beta \neq 0, \mathbf{V}$ is positive definite$\}$. Hence it is the ML estimate of $(\mu, \beta, \sigma^2, \sigma_\delta^2, \sigma_\epsilon^2)$ provided that $\hat{\sigma}^2 \geqslant 0$, $\hat{\sigma}_\delta^2 \geqslant 0$ and $\hat{\sigma}_\epsilon^2 \geqslant 0$. In this case, we have

$$\begin{aligned}
\hat{\mu} &= \bar{\xi}.,\\
\hat{\beta} &= \bar{\eta}./\bar{\xi}.,\\
\hat{\sigma}^2 &= s_{\xi\eta}(\bar{\xi}./\bar{\eta}.) = m_{\xi\eta}(\bar{\xi}./\bar{\eta}.) - \bar{\xi}.^2,\\
\hat{\sigma}_\delta^2 &= s_{\xi\xi} - s_{\xi\eta}(\bar{\xi}./\bar{\eta}.) = m_{\xi\xi} - m_{\xi\eta}(\bar{\xi}./\bar{\eta}.),\\
\hat{\sigma}_\epsilon^2 &= s_{\eta\eta} - s_{\xi\eta}(\bar{\eta}./\bar{\xi}.) = m_{\eta\eta} - m_{\xi\eta}(\bar{\eta}./\bar{\xi}.).
\end{aligned} \tag{2.1}$$

$\bar{\xi}.$, $\bar{\eta}.$, $m_{\xi\xi}$, $m_{\eta\eta}$ and $m_{\xi\eta}$ are also jointly sufficient statistics.

However, a complication arises when one of the $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ is less than zero. Then the likelihood function L has to be maximized directly. L has only one local maximum on the open set Ω at $(\hat{\mu}, \hat{\beta}, \hat{\sigma}^2, \hat{\sigma}_\delta^2, \hat{\sigma}_\epsilon^2)$. If one of the $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ is negative, then when restricted to the set of all admissible values $\omega = \{(\mu, \beta, \sigma^2, \sigma_\delta^2, \sigma_\epsilon^2): \mu \neq 0, \beta \neq 0, \sigma^2 \geqslant 0, \sigma_\delta^2 \geqslant 0, \sigma_\epsilon^2 \geqslant 0, \mathbf{V}$ is positive definite$\}$, L cannot have a local maximum at a point such that all σ^2, σ_δ^2 and σ_ϵ^2 are positive. Thus the problem reduces to maximizing L in each case when $\sigma^2 = 0$, $\sigma_\delta^2 = 0$ or $\sigma_\epsilon^2 = 0$ and taking the one which gives the largest value of L as our ML solution.

Case 1. $\sigma^2 = 0$. One can verify that

$$L = 1/((2\pi\sigma_\delta\sigma_\epsilon)^n) \exp\left\{-(1/2)\left[\sum(\xi_i - \mu)^2/\sigma_\delta^2 + \sum(\eta_i - \beta\mu)^2/\sigma_\epsilon^2\right]\right\}.$$

Hence it is clear that L is maximized when

$$\begin{aligned}
\mu &= \bar{\xi}.,\\
\beta &= \bar{\eta}./\bar{\xi}.,\\
\sigma_\delta^2 &= m_{\xi\xi} - \bar{\xi}.^2,\\
\sigma_\epsilon^2 &= m_{\eta\eta} - \bar{\eta}.^2,
\end{aligned} \tag{2.2}$$

and at this point

$$\ln L = -n\ln(2\pi) - (n/2)\ln[(m_{\xi\xi} - \bar{\xi}.^2)(m_{\eta\eta} - \bar{\eta}.^2)] - n. \tag{2.3}$$

Case 2. $\sigma_\delta^2 = 0$. After some algebraic manipulation one finds that

$$L = 1/((2\pi\sigma\sigma_\epsilon)^n) \exp\left\{-(1/2)\left[\sum(\xi_i - \mu)^2/\sigma^2 + \sum(\eta_i - \beta\xi_i)^2/\sigma_\epsilon^2\right]\right\}.$$

Hence L is maximized when

$$\begin{aligned} \mu &= \bar{\xi}., \\ \beta &= m_{\xi\eta}/m_{\xi\xi}, \\ \sigma^2 &= m_{\xi\xi} - \bar{\xi}.^2, \\ \sigma_\epsilon^2 &= m_{\eta\eta} - m_{\xi\eta}^2/m_{\xi\xi}, \end{aligned} \tag{2.4}$$

and at this point

$$\ln L = -n\ln(2\pi) - (n/2)\ln[(m_{\xi\xi} - \bar{\xi}.^2)(m_{\eta\eta} - m_{\xi\eta}^2/m_{\xi\xi})] = n. \tag{2.5}$$

Case 3. $\sigma_\epsilon^2 = 0$. From case 2 we have, by symmetry,

$$L = 1/((2\pi\beta\sigma\sigma_\delta)^n)\exp\left\{-(1/2)\left[\sum(\eta_i - \beta\mu)^2/\beta^2\sigma^2 + \sum(\beta\xi_i - \eta_i)^2/\beta^2\sigma_\delta^2\right]\right\}.$$

which is maximized at

$$\begin{aligned} \mu &= \bar{\eta}.(m_{\xi\eta}/m_{\eta\eta}), \\ \beta &= m_{\eta\eta}/m_{\xi\eta}, \\ \sigma^2 &= (m_{\xi\eta}/m_{\eta\eta})^2(m_{\eta\eta} - \bar{\eta}.^2), \\ \sigma_\delta^2 &= m_{\xi\xi} - m_{\xi\eta}^2/m_{\eta\eta}, \end{aligned} \tag{2.6}$$

and at this point

$$\ln L = -n\ln(2\pi) - (n/2)\ln[(m_{\eta\eta} - \bar{\eta}.^2)(m_{\xi\xi} - m_{\xi\eta}^2/m_{\eta\eta})] - n. \tag{2.7}$$

Therefore we have the following

THEOREM 2.1. *Assume that* $\beta \neq 0$. *If one of* $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ *and* $\hat{\sigma}_\epsilon^2$ *in* (2.1) *is negative, then the ML estimate is given by either* (2.2), (2.4) *or* (2.6) *depending on which of* (2.3), (2.5) *and* (2.7) *gives the largest value of* L.

It is not difficult to see that if $\hat{\sigma}_\delta^2 < 0$, then (2.5) is greater than (2.7) and (2.3). Hence (2.4) gives the ML solution. Similarly if $\hat{\sigma}_\epsilon^2 < 0$, (2.6) gives the ML solution.

Now let us remove the restriction that $\beta \neq 0$. Suppose that L attains its maximum at a point $(\mu', \beta', \sigma'^2, \sigma_\delta'^2, \sigma_\epsilon'^2)$ with $\beta' = 0$. At this point $\mathbf{V}$ becomes a diagonal matrix with elements $\sigma'^2 + \sigma_\delta'^2$ and $\sigma_\epsilon'^2$. However the point $(\mu', 0, \sigma'^2 + \sigma_\delta'^2, 0, \sigma_\epsilon'^2)$ also gives the same maximum. From case 2 we notice that this point is given by (2.4). Hence $m_{\xi\eta}/m_{\xi\xi} = 0$ which has probability zero of occurring.

We now give a geometric interpretation of the estimation situation. Since $y = \beta x$, the true structural relationship must pass through the origin P. The line $l = \{(x, y): y = \hat{\beta}x = (\bar{\eta}./\bar{\xi}.)x\}$ is obtained by joining the sample mean $(\bar{\xi}., \bar{\eta}.)$ to P and asymptotically converges to the true structural relationship. The two sample regression lines l_L and l_U obtained from regressing η on ξ and ξ on η, respectively, also pass through $(\bar{\xi}., \bar{\eta}.)$ and have slopes $\hat{\beta}_L = s_{\xi\eta}/s_{\xi\xi}$ and $\hat{\beta}_U = s_{\eta\eta}/s_{\xi\eta}$, respectively. Now consider co-ordinate axes with origin at $(\bar{\xi}., \bar{\eta}.)$. Since $\hat{\beta}_L\hat{\beta}_U > 0$, the two slopes have the same sign and consequently l_L and l_U must pass through the same two quadrants. From the Schwartz inequality $|\hat{\beta}_L| \leqslant |\hat{\beta}_U|$ we then have the following three situations.

(1) If P lies between l_L and l_U, the slopes $\hat{\beta}$, $\hat{\beta}_L$ and $\hat{\beta}_U$ all have the same sign and $|\hat{\beta}_L| \leqslant |\hat{\beta}| \leqslant |\hat{\beta}_U|$. Consequently, from (2.1) $\hat{\sigma}^2 \geqslant 0$, $\hat{\sigma}_\delta^2 \geqslant 0$ and $\hat{\sigma}_\epsilon^2 \geqslant 0$ and the normal situation occurs with $(\hat{\mu}, \hat{\beta}, \hat{\sigma}^2, \hat{\sigma}_\delta^2, \hat{\sigma}_\epsilon^2)$ giving the ML estimate.

(2) If P lies in the same quadrants as l_L and l_U but outside the region enclosed by these regression lines, then $\hat{\beta}$, $\hat{\beta}_L$ and $\hat{\beta}_U$ are still of the same sign and either $|\hat{\beta}| \leqslant |\hat{\beta}_L|$ or $|\hat{\beta}_U| \leqslant |\hat{\beta}|$. Thus from (2.1) $\hat{\sigma}^2 = s_{\xi\eta}/\hat{\beta} = s_{\xi\xi}\hat{\beta}_L/\hat{\beta} \geqslant 0$. If P is closer to l_L than l_U, then $|\hat{\beta}| \leqslant |\hat{\beta}_L|$ so that from (2.1) $\hat{\sigma}_\delta^2 \leqslant 0$ and case 2 gives the ML estimate. Otherwise $\hat{\sigma}_\epsilon^2 \leqslant 0$ and case 3 gives the ML estimate. Moran (1971a) discussed these situations intuitively and pointed out that in these cases, the sample variances and covariance of ξ and η should give some information on the slope parameter β. Estimates of β in (2.4) and (2.6) therefore give the necessary adjustment when $\hat{\beta}$ lies outside the bounds $\hat{\beta}_L$ and $\hat{\beta}_U$.

(3) If P lies in the quadrants in which l_L and l_U do not lie, then $\hat{\beta}$ has the opposite sign to $\hat{\beta}_L$. In this case $\hat{\sigma}^2 = s_{\xi\xi}\hat{\beta}_L/\hat{\beta} < 0$.

3. Asymptotic Behaviour of the ML Estimate

Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_5) = (\mu, \beta, \sigma^2, \sigma_\delta^2, \sigma_\epsilon^2)$. Direct computation shows that the information matrix $\mathbf{C} = [c_{ij}]$, where $c_{ij} = -E(\partial^2 \ln L/\partial\theta_i\,\partial\theta_j)$, is given by

$$\mathbf{C} = n\,|\mathbf{V}|^{-2}\begin{pmatrix} \delta_{22}\,|\mathbf{V}|\,\cdot & \delta_{12}\mu\,|\mathbf{V}| & 0 & 0 & 0 \\ \delta_{12}\mu\,|\mathbf{V}| & \sigma_{11}\mu^2|\mathbf{V}|+(\sigma^2)^2(|\mathbf{V}|+2\delta_{12}^2) & \delta_{12}\sigma^2\delta_{22} & -\sigma_{12}\sigma^2\sigma_\epsilon^2 & \delta_{12}\sigma^2\sigma_{11} \\ 0 & \delta_{12}\sigma^2\delta_{22} & \delta_{22}^2/2 & (\sigma_\epsilon^2)^2/2 & \delta_{12}^2/2 \\ 0 & -\sigma_{12}\sigma^2\sigma_\epsilon^2 & (\sigma_\epsilon^2)^2/2 & \sigma_{22}^2/2 & \sigma_{12}^2/2 \\ 0 & \delta_{12}\sigma^2\sigma_{11} & \delta_{12}^2/2 & \sigma_{12}^2/2 & \sigma_{11}^2/2 \end{pmatrix},$$

where $\delta_{12} = \beta\sigma_\delta^2$, $\delta_{22} = \beta^2\sigma_\delta^2 + \sigma_\epsilon^2$, $\sigma_{12} = \beta\sigma^2$, $\sigma_{11} = \sigma^2 + \sigma_\delta^2$ and $\sigma_{22} = \beta^2\sigma^2 + \sigma_\epsilon^2$. After some lengthly algebraic manipulation, it is seen that

$$\mathbf{D} = \mathbf{C}^{-1} = \frac{1}{n}\begin{pmatrix} \sigma_{11} & -\delta_{12}/\mu & r & -r & \beta^2 r \\ -\delta_{12}/\mu & \delta_{22}/\mu^2 & -s & s & -\beta^2 s \\ r & -s & (\sigma_{12}\sigma^2 + \sigma_{11}\sigma_{22})\beta^{-2} + k & (\delta_{12}\sigma_{12} - \sigma_{11}\sigma_\epsilon^2)\beta^{-2} - k & (\sigma^2\sigma_\epsilon^2 - \sigma_\delta^2\sigma_{22}) + \beta^2 k \\ -r & s & (\delta_{12}\sigma_{12} - \sigma_{11}\sigma_\epsilon^2)\beta^{-2} - k & (\sigma_{11}\delta_{22} + \delta_{12}^2)\beta^{-2} + k & (\sigma_\delta^2\sigma_\epsilon^2 - \sigma^2\delta_{22}) - \beta^2 k \\ \beta^2 r & -\beta^2 s & (\sigma^2\sigma_\epsilon^2 - \sigma_\delta^2\sigma_{22}) + \beta^2 k & (\sigma_\delta^2\sigma_\epsilon^2 - \sigma^2\delta_{22}) - \beta^2 k & \delta_{22}\sigma_{22} + (\sigma_\epsilon^2)^2 + \beta^4 k \end{pmatrix}$$

$$= \frac{1}{n}\begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{12}' & \mathbf{D}_{22} \end{pmatrix},$$

where $k = (\sigma^2)^2\delta_{22}/(\beta u)^2$, $r = \sigma^2\sigma_\delta^2/\mu$, $s = \sigma^2\delta_{22}/(\beta\mu^2) = \beta k/\sigma^2$. Also $\mathbf{D}_{11}$, $\mathbf{D}_{12}$ and $\mathbf{D}_{22}$ are matrices of order 2×2, 2×3 and 3×3 respectively. The $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\beta}, \hat{\sigma}^2, \hat{\sigma}_\delta^2, \hat{\sigma}_\epsilon^2)$ of (2.1) is therefore asymptotically normally distributed with mean $\boldsymbol{\theta}$ and covariance matrix $\mathbf{D}$. To investigate the asymptotic behaviour of the ML estimate $\hat{\boldsymbol{\theta}}^M$ of $\boldsymbol{\theta}$, two cases have to be distinguished.

(1) σ^2, σ_δ^2 and σ_ϵ^2 are all positive. In this case, since $\hat{\boldsymbol{\theta}}$ is consistent, the probability that all $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ are positive tends to one as n tends to infinity. Thus we have

THEOREM 3.1. $\Pr(\hat{\boldsymbol{\theta}}^M = \hat{\boldsymbol{\theta}}) \to 1$ *as* $n \to \infty$ *and consequently* $\hat{\boldsymbol{\theta}}^M$ *is asymptotically distributed as* $N(\boldsymbol{\theta}, \mathbf{D})$. *The asymptotic probability that* $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ *and* $\hat{\sigma}_\epsilon^2$ *are all non-negative is given by*

$$\int_0^\infty \int_0^\infty \int_0^\infty f_n(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3,$$

where $f_n(x_1, x_2, x_3)$ *is the probability density function of a trivariate normal distribution with mean* $(\sigma^2, \sigma_\delta^2, \sigma_\epsilon^2)$ *and covariance matrix* $\mathbf{D}_{22}/n$.

(2) One of σ^2, σ_δ^2 and σ_ϵ^2 is zero. We discuss only the case $\sigma_\delta^2 = 0$. The other cases are similar. Chant (1974) and Moran (1971b) discussed ML estimation when some of the true parameters lie on the boundary of the parameter space.

Since $\hat{\sigma}^2 \to^p \sigma^2 > 0$ and $\hat{\sigma}_\epsilon^2 \to^p \sigma_\epsilon^2 > 0$, where "$\to^p$" denotes convergence in probability, $\Pr(\hat{\sigma}^2 > 0 \text{ and } \hat{\sigma}_\epsilon^2 > 0) \to 1$. Therefore with probability tending to one $\hat{\boldsymbol{\theta}}^M$ is given either by $\hat{\boldsymbol{\theta}}$ or $\tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}} = (\tilde{\mu}, \tilde{\beta}, \tilde{\sigma}^2, 0, \tilde{\sigma}_\epsilon^2)$ is defined by (2.4). In fact, we have

THEOREM 3.2. *If* $\sigma_\delta^2 = 0$, *then both* $\Pr(\hat{\boldsymbol{\theta}}^M = \hat{\boldsymbol{\theta}})$ *and* $\Pr(\hat{\boldsymbol{\theta}}^M = \tilde{\boldsymbol{\theta}})$ *tend to* $\frac{1}{2}$ *as* $n \to \infty$.

Proof. We have

$$\begin{aligned}
\Pr(n^{1/2}\hat{\sigma}_\delta^2 \geqslant 0) \geqslant \Pr(\hat{\boldsymbol{\theta}}^M = \hat{\boldsymbol{\theta}}) &= \Pr(\hat{\sigma}_\delta^2 \geqslant 0, \hat{\sigma}^2 \geqslant 0 \text{ and } \hat{\sigma}_\epsilon^2 \geqslant 0) \\
&= \Pr(\hat{\sigma}_\delta^2 \geqslant 0) + \Pr(\hat{\sigma}^2 \geqslant 0 \text{ and } \hat{\sigma}_\epsilon^2 \geqslant 0) - \Pr(\hat{\sigma}_\delta^2 \geqslant 0 \\
&\qquad \text{or both } \hat{\sigma}^2 \text{ and } \hat{\sigma}_\epsilon^2 \geqslant 0) \\
&\geqslant \Pr(n^{1/2}\hat{\sigma}_\delta^2 \geqslant 0) + \Pr(\hat{\sigma}^2 \geqslant 0 \text{ and } \hat{\sigma}_\epsilon^2 \geqslant 0) - 1.
\end{aligned}$$

Since $\lim_n \Pr(\hat{\sigma}^2 \geqslant 0 \text{ and } \hat{\sigma}_\epsilon^2 \geqslant 0) = 1$, where $\lim_n$ denotes the limit as $n \to \infty$, and $n^{1/2}\hat{\sigma}_\delta^2$ is asymptotically distributed as $N(0, (\mu^2 + \sigma^2)\sigma^2\sigma_\epsilon^2/(\beta\mu)^2)$, we have

$$\tfrac{1}{2} = \Pr(z \geqslant 0) \geqslant \lim_n \Pr(\hat{\boldsymbol{\theta}}^M = \hat{\boldsymbol{\theta}}) \geqslant \Pr(z \geqslant 0) = \tfrac{1}{2},$$

where z is distributed as $N(0, (\mu^2 + \sigma^2)\sigma^2\sigma_\epsilon^2/(\beta\mu)^2)$. Thus $\lim_n \Pr(\hat{\boldsymbol{\theta}}^M = \hat{\boldsymbol{\theta}}) = \frac{1}{2}$. Similarly $\lim_n \Pr(\hat{\boldsymbol{\theta}}^M = \tilde{\boldsymbol{\theta}}) = \frac{1}{2}$.

Now we proceed to find the limiting distribution of $\mathbf{z}_n = n^{1/2}(\hat{\boldsymbol{\theta}}^M - \boldsymbol{\theta})$. For any vectors $\mathbf{a} = (a_1, \ldots, a_p)'$ and $\mathbf{b} = (b_1, \ldots, b_p)'$, we use $\mathbf{a} < \mathbf{b}$ to denote $a_i < b_i$ for all $i = 1, \ldots, p$. Let E_1 be the event that $\hat{\sigma}^2$, $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\epsilon^2$ are non-negative and E_2 be the event that $\hat{\sigma}_\delta^2 < 0$. We have

$$\begin{aligned}
F_n(\mathbf{t}) = \Pr(\mathbf{z}_n < \mathbf{t}) &= \Pr(n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) < \mathbf{t}, E_1) + \Pr(n^{1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) < \mathbf{t}, E_2) \\
&\quad + \Pr(\mathbf{z}_n < \mathbf{t}, E_1^c \cap E_2^c). \qquad (3.1)
\end{aligned}$$

Let the t_3, t_4 and t_5 of $\mathbf{t} = (t_1, \ldots, t_5)$ be positive. If any of t_3, t_4 and t_5 is negative, $F_n(\mathbf{t}) = 0$. We notice immediately that as $n \to \infty$

$$\begin{aligned}
\Pr(\mathbf{z}_n < \mathbf{t}, E_1^c \cap E_2^c) &\leqslant \Pr(E_1^c \cap E_2^c) \\
&= \Pr(\text{one of } \hat{\sigma}^2 \text{ and } \hat{\sigma}_\epsilon^2 \text{ is negative}) \to 0.
\end{aligned}$$

The limit of the second term on the right hand side of (3.1) can be deduced from theorem 2 of Chant (1974) and is equal to $2^{-1}G(t_1, t_2, t_3, t_5)$, where G is the cumulative distribution function of $N(\mathbf{0}, n\mathbf{D}_\delta)$ and $\mathbf{D}_\delta$ is the inverse of the matrix obtained from $\mathbf{C}$ by deleting the fourth row and column and is equal to

$$\begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma_\epsilon^2/(\mu^2+\sigma^2) & 0 & 0 \\ 0 & 0 & 2(\sigma^2)^2 & 0 \\ 0 & 0 & 0 & 2(\sigma_\epsilon^2)^2 \end{pmatrix}.$$

Thus

$$G(t_1, t_2, t_3, t_5) = \Phi(t_1\sigma^{-1})\,\Phi(t_2(\mu^2+\sigma^2)^{1/2}\,\sigma_\epsilon^{-1})\,\Phi(t_3 2^{-1/2}\sigma^{-2})\,\Phi(t_5 2^{-1/2}\sigma_\epsilon^{-2}), \quad (3.2)$$

where Φ is the cumulative distribution function of $N(0, 1)$. Alternatively, $\lim_n \Pr(n^{1/2}(\tilde{\boldsymbol\theta} - \boldsymbol\theta) < \mathbf{t}, E_2)$ can be found by writing

$$\Pr(n^{1/2}(\tilde{\boldsymbol\theta} - \boldsymbol\theta) < \mathbf{t}, E_2) = \Pr(n^{1/2}(\tilde{\boldsymbol\theta} - \boldsymbol\theta) < \mathbf{t} \mid E_2)\Pr(E_2)$$

and observing that $\epsilon(\hat\sigma_\delta^2(n)^{1/2}(\tilde\mu - \mu, \tilde\beta - \beta, \tilde\sigma^2 - \sigma^2, \tilde\sigma_\epsilon^2 - \sigma_\epsilon^2)) = \mathbf{0}$, where ϵ denotes expectation with respect to the asymptotic distribution, so that

$$\begin{aligned} &\lim_n \Pr(n^{1/2}(\tilde{\boldsymbol\theta} - \boldsymbol\theta) < \mathbf{t}, E_2) \\ &\quad = \lim_n \{\Pr(n^{1/2}(\tilde\mu - \mu, \tilde\beta - \beta, \tilde\sigma^2 - \sigma^2, \tilde\sigma_\epsilon^2 - \sigma_\epsilon^2) < (t_1, t_2, t_3, t_5))\} \end{aligned}$$

and the limit on the right hand side is equal to $G(t_1, t_2, t_3, t_5)$. Consider the first terms of (3.1). Since

$$\begin{aligned} \Pr(n^{1/2}(\hat{\boldsymbol\theta} - \boldsymbol\theta) < \mathbf{t}, \hat\sigma_\delta^2 \geqslant 0) &= \Pr(n^{1/2}(\hat{\boldsymbol\theta} - \boldsymbol\theta) < \mathbf{t}, E_1) \\ &\quad + \Pr(n^{1/2}(\hat{\boldsymbol\theta} - \boldsymbol\theta) < \mathbf{t}, \hat\sigma_\delta^2 \geqslant 0, \text{at least one of } \hat\sigma^2 \text{ and } \hat\sigma_\epsilon^2 \text{ is negative}), \end{aligned}$$

the second term on the right hand side tends to 0 as $n \to \infty$. Since $n^{1/2}(\hat{\boldsymbol\theta} - \boldsymbol\theta)$ is asymptotically distributed as $N(\mathbf{0}, n\mathbf{D})$, we have

$$\begin{aligned} \lim_n \Pr(n^{1/2}(\hat{\boldsymbol\theta} - \boldsymbol\theta) < \mathbf{t}, E_1) &= \lim_n \Pr(n^{1/2}(\hat{\boldsymbol\theta} - \boldsymbol\theta) < \mathbf{t}, \hat\sigma_\delta^2 \geqslant 0) \\ &= \int_{-\infty}^{t_1}\int_{-\infty}^{t_2}\int_{-\infty}^{t_3}\int_{0}^{t_4}\int_{-\infty}^{t_5} g(x_1, \ldots, x_5)\, dx_1 \cdots dx_5, \end{aligned} \quad (3.3)$$

where g is the probability density function of a normal variate with mean $\mathbf{0}$ and covariance matrix $n\mathbf{D}$. (3.3) is identical to the first component of (19) in theorem I of Moran (1971b). Thus, we have proved

THEOREM 3.3. *If* $\sigma_\delta^2 = 0$, *then*

$$\begin{aligned} \lim_n F_n(\mathbf{t}) &= \lim_n \Pr(n^{1/2}(\hat{\boldsymbol\theta}^M - \boldsymbol\theta) < \mathbf{t}) \\ &= p_1 + p_2, \end{aligned}$$

where p_1 *and* p_2 *are given by* (3.2) *and* (3.3), *respectively.*

4. Some Remarks

Since the parameters are unidentifiable if $\mu = 0$, in practice, before the ML estimation procedure is applied, it is helpful to test the hypothesis $\mu = 0$ by examining whether the sample mean $\bar{\xi}.$ is significantly different from zero $(E(\xi) = E(x) = 0)$. If $\mu = 0$ is rejected, the ML procedure described in section 2 is then carried out to obtain $\hat{\boldsymbol{\theta}}^M$.

Suppose the main interest is in estimating β. One can use the consistent estimate $\hat{\beta}$. However, the ML estimate $\hat{\beta}^M$ of β, which was originally motivated by the case of negative estimated variances, is already seen to be a refinement of $\hat{\beta}$ using information provided by the second order moments when $\hat{\beta}$ lies outside the bounds formed by $\hat{\beta}_L$ and $\hat{\beta}_U$ (which asymptotically satisfy $\hat{\beta}_L < \beta < \hat{\beta}_U$ if $\beta > 0$ and with inequalities reversed if $\beta < 0$). Therefore $\hat{\beta}^M$ is preferable to $\hat{\beta}$ in finite samples (asymptotically they are identical (theorem 3.1) if σ^2, $\sigma_\delta{}^2$ and $\sigma_\epsilon{}^2$ are positive). If $\sigma_\delta{}^2 = 0$, then with probability tending to $\frac{1}{2}$ (theorem 3.2), the ML estimate of $\sigma_\delta{}^2$ has the value 0 and $\hat{\beta}^M$ could yield a higher precision than $\hat{\beta}$. It would be helpful to examine the accuracy of $\hat{\beta}^M$ or $\hat{\beta}$ by computing the estimated asymptotic variances obtained by replacing the true parameter $\boldsymbol{\theta}$ in the formulas in theorems 3.1 and 3.3 by $\hat{\boldsymbol{\theta}}^M$.

The estimate $\hat{\beta}' = m_{\xi\eta}/m_{\xi\xi}$ in (2.4) converges in probability to $\beta - \beta\sigma_\delta{}^2/(\sigma^2 + \sigma_\delta{}^2 + \mu^2)$ and hence is asymptotically biased (the bias is small when β and $\sigma_\delta{}^2$ are small and when σ^2 is large) but in general has small variance. In finite samples, the consistency of $\hat{\beta}^M$ does not ensure that it is superior to $\hat{\beta}'$. Thus another procedure is to choose adaptively one of $\hat{\beta}^M$ and $\hat{\beta}'$ with smaller estimated asymptotic mean square error (AMSE) (cf. Feldstein 1974). Direct computation shows that

$$\begin{aligned}\text{AMSE}(\hat{\beta}') &= n^{-1}\epsilon(n^{1/2}(\hat{\beta}' - \beta + \beta\sigma_\delta{}^2/(\sigma^2 + \sigma_\delta{}^2 + \mu^2)) + \beta^2(\sigma_\delta{}^2)^2/(\sigma^2 + \sigma_\delta{}^2 + \mu^2)^2 \\ &= \{(\sigma^2 + \sigma_\delta{}^2 + \mu^2)^2[(\sigma^2 + \mu^2)(\beta^2\sigma_\delta{}^2 + \sigma_\epsilon{}^2) + \sigma_\delta{}^2\sigma_\epsilon{}^2] - 2\beta^2(\sigma_\delta{}^2)^2\mu^4\} \\ &\quad \times (\sigma^2 + \sigma_\delta{}^2 + \mu^2)^{-4}n^{-1} + \beta^2(\sigma_\delta{}^2)^2/(\sigma^2 + \sigma_\delta{}^2 + \mu^2)^2,\end{aligned}$$

and can be estimated by replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}^M$.

Acknowledgment

The authors are grateful to the referee for helpful comments.

References

Anderson, T. W. (1958). *An Introduction to Multivariate Analysis.* Wiley, New York.

Birch, M. W. (1964). A Note on the maximum likelihood estimation of a linear structural relationship. *J. Amer. Statist. Assoc.* **59** 1175–1178.

CHANT, D. C. (1974). On asymptotic tests of composite hypotheses in non-standard conditions. *Biometrika* **61** 291–298.

FELDSTEIN, Martin (1974). Errors in variables: A consistent estimator with smaller MSE in finite samples. *J. Amer. Statist. Assoc.* **69** 990–996.

KENDALL, M. G. AND STUART, A. (1973). *The Advanced Theory of Statistics*, Vol. II, 3rd ed. Charles Griffin, London.

LINDLEY, D. V. (1947). Regression lines and the linear functional relationships. *J. Roy. Statist. Soc. Suppl.* **9** 218–244.

MORAN, P. A. P. (1971a). Estimating structural and functional relationships. *J. Multivar. Anal.* **1** 232–255.

MORAN, P. A. P. (1971b). Maximum likelihood estimation in non-standard conditions. *Math. Proc. Camb. Phil. Soc.* **70** 414–450.

ZELLNER, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.