Annals of the Institute of Statistical Mathematics
Vol 18, No 3, 1966

Cowles Foundation Paper 247

NON-EXISTENCE OF CONSISTENT ESTIMATOR SEQUENCES AND UNBIASED ESTIMATORS: A PRACTICAL EXAMPLE

H. S. Konijn

(Received May 25, 1966)

Summary

Under conditions often occurring in practice the variance of the estimated constant in a regression equation with correlated disturbances cannot be estimated without bias or consistently, although it is identifiable.

Introduction

It is often thought that identifiability for each sample size implies existence of consistent estimator sequences. A rather artificial counter example is given in [7].[1] We here consider a case which often arises in experimental and survey practice. The example concerns a model with unknown intraclass correlation ρ. We show that the variance of the mean of the observations is identified but that it possesses neither an unbiased estimator sequence nor a consistent estimator sequence. More generally, we show that the variance of the estimated constant term in a regression equation has these properties when the disturbances have constant unknown variances σ^2 and correlations ρ, and a single observation is made for each set of values of the regressors.

Related aspects of estimation in models of this sort are considered in [4].

1. Some families of distributions

Let $(y_1, \cdots, y_n)$ be multivariate nonsingular normal with $n>1$, and with

$$\mu=\mathcal{E}y_j, \quad \sigma^2=\mathcal{E}(y_j-\mu)^2, \quad \rho=\mathcal{E}(y_j-\mu)(y_{j'}-\mu)/\sigma^2 \qquad (j\neq j')$$

[1] It is well-known that in the model considered below the mean, μ, has no consistent estimator sequence, however, this parameter does have an unbiased estimator (the sample mean is unbiased for μ).

unknown constants. We shall refer to the corresponding family of distributions, when μ ranges over the entire real line and the ranges of σ^2 and ρ are as given in (2) below, as $\mathcal{N}$.

Such a family arises naturally in the study of experimental results [1] and in sample survey situations [2], [3]. It also seems natural in the study of interdependence among small clusters of individuals (such as arises in ecology, sociometry, and so forth) or of objects in a plane or space, being the simplest of a class of families in which the covariance of two elements is a function of their distance. An extension is considered in section 6.

The logarithm of the joint density function is twice

$$(1)\quad c_1-n\log\sigma^2-m\log(1-\rho)-\log(1+m\rho)$$
$$-\sigma^{-2}[\sum(y_j-\mu)^2(1-\rho)^{-1}-(\bar{y}-\mu)^2n\{(1-\rho)^{-1}-(1+m\rho)^{-1}\}]\ ,$$

where c_1 is a known constant and $m=n-1$.

From (1) it is seen that nonsingularity of the distribution implies

$$(2)\qquad \sigma^2>0\ ,\qquad -m^{-1}<\rho<1\ .$$

An amusing interpretation of the restriction (2) arises when all members of a group of individuals make a conscious attempt to be nonconformists: if the model is appropriate the amount of possible nonconformity as measured by $-\rho$ appears limited by the size of the group.

It is well known (see section 6) that

$$\bar{y}\quad\text{and}\quad V=\sum(y_j-\bar{y})^2$$

are independently distributed and that their density functions have logarithm proportional to

$$(3)\qquad c_2-\log\sigma^2-\log(1+m\rho)-\sigma^{-2}(\bar{y}-\mu)^2n(1+m\rho)^{-1}$$

and

$$(4)\qquad c_3-m\log\sigma^2-m\log(1-\rho)+(m-2)\log V$$
$$-\sigma^{-2}(\sum y_j^2-n\bar{y}^2)(1-\rho)^{-1}\ .$$

Therefore the conditional density of $(y_1,\cdots,y_n)$ given $(\bar{y}, V)$ has logarithm proportional to

$$(5)\qquad c_4-(m-2)\log V\ ,$$

which does not depend on μ, σ^2 or ρ; and so $(\bar{y}, V)$ is sufficient for the family $\mathcal{N}$.

Since $\sqrt{n}(\bar{y}-\mu)$ has a normal distribution with zero mean and variance

(6) $$\omega^2=\sigma^2(1+m\rho) ,$$

and since, if

(7) $$\kappa^2=\sigma^2(1-\rho) ,$$

$V\kappa^{-2}$ has a chi-square distribution with m degrees of freedom, it is also convenient to consider a parametrization of $\mathcal{N}$ by μ, κ^2 and ω^2 with (2) replaced by the inequalities

(2′) $$\kappa^2>0 , \qquad \omega^2>0 .$$

Frequently [1] we are really interested in estimating ω^2 and κ^2 rather than σ^2 and ρ.

In some problems it is possible to replace (2) by

(2*) $$\sigma^2>0 , \qquad \rho\geqq 0 ;$$

we shall refer to that subfamily of $\mathcal{N}$ as $\mathcal{N}_0$.

2. Nonexistence of an unbiased estimate of ω^2

If f is a function of the observations and $\mathcal{E}f(y_1, \cdots, y_n)$ exists (as a Lebesgue integral), then the conditional expectation $\mathcal{E}\{f(y_1,\cdots, y_n) \mid \bar{y}, V\}$ exists, and, by the sufficiency of $(\bar{y}, V)$ does not depend on the parameters; call it $g(\bar{y}, V)$. Since the distribution of V does not depend on μ,

$$g_0(\bar{y} \mid \kappa^2)=\mathcal{E}\{g(\bar{y}, V) \mid \bar{y}\}$$

exists and is not a function of μ.

So, if $f(y_1, \cdots, y_n)$ is an unbiased estimate of ω^2, then, for each positive number κ_0^2, $\mathcal{E}g_0(\bar{y} \mid \kappa_0^2)$ equals ω^2 identically in μ and ω^2. Consequently, writing $h(\bar{y}n^{1/2})=g_0(\bar{y} \mid \kappa_0^2)$, z for $\bar{y}n^{1/2}$ and ν for $\mu n^{1/2}$,

$$\mathcal{E}h(z)=(2\pi\omega^2)^{-1/2}\int_{-\infty}^{\infty} h(z) \exp\{-\tfrac{1}{2}(z-\nu)^2\omega^{-2}\}\, dz=\omega^2$$

identically in ν^2 and ω^2. That would mean that there would exist an unbiased estimate of the variance ω^2 of a normal distribution with unknown mean ν based on a single observation. That this is not so is proved in [5].

3. Sequences of families

In discussing asymptotic properties one also has to consider infinite sequences of families $\mathcal{N}$ for $n=2, 3, \cdots$. It should be noted that in such a sequence the second part of (2) is not tenable when ρ is taken to be constant through the sequence, i.e., that case (2) must be replaced by (2*). (Therefore in this case the study of asymptotic properties is without sense, and we have to confine ourselves to the study of fixed sample size properties.)

The logarithm of the characteristic function of $(y_1, \cdots, y_n)$ is

$$(8)\qquad \psi(t_1, \cdots, t_n \mid \mu, \sigma^2, \rho) = \log \mathcal{E} \exp (i \sum y_j t_j) \\ = i\mu n\bar{t} - \tfrac{1}{2}\sigma^2(\sum t_j^2 + \rho \sum\sum t_j t_{j'}) \qquad (j \neq j') .$$

From this it is seen that y_i $(i=1, 2, \cdots)$ defines a stochastic process: for $i_1, \cdots, i_k$ a subset of $(1, 2, \cdots, n)$ and $j_1, \cdots, j_{n-k}$ the remaining elements of $(1, \cdots, n)$,

$$\psi(t_{i_1}, \cdots, t_{i_k} \mid \mu, \sigma^2, \rho) = \psi(t_1, \cdots, t_n \mid \mu, \sigma^2, \rho) \mid_{t_{j_1}=\cdots=t_{j_{n-k}}=0} .$$

4. Identifiability

Consider a collection of specified functions $q, r, \cdots$ of the parameters. Necessary and sufficient for the identifiability of this collection in $\mathcal{N}$ is that for any two sets $(\mu_1, \sigma_1^2, \rho_1)$ and $(\mu_2, \sigma_2^2, \rho_2)$ of values of the parameters the identity over n-space:

$$(9)\qquad \psi(t_1, \cdots, t_n \mid \mu_1, \sigma_1^2, \rho_1) \equiv \psi(t_1, \cdots, t_n \mid \mu_2, \sigma_2^2, \rho_2)$$

can hold if and only if all the functions $q, r, \cdots$ take on the same value for $(\mu_1, \sigma_1^2, \rho_1)$ and $(\mu_2, \sigma_2^2, \rho_2)$.

Suppose, for example, that $q(\mu, \sigma^2, \rho) = \mu$, $r(\mu, \sigma^2, \rho) = \sigma^2$ and $s(\mu, \sigma^2, \rho) = \rho$. For $t_2 = \cdots = t_n = 0$ and $t_1 \neq 0$, the real part of (9) implies that $\sigma_1^2 = \sigma_2^2$, and the imaginary part that $\mu_1 = \mu_2$. This reduces (9) to the identity

$$(10)\qquad (\rho_1 - \rho_2) \sum\sum t_j t_{j'} \equiv 0$$

after division by the common, negative value of $-\frac{1}{2}\sigma^2$. By selecting any nonzero values for t_1 and t_2, and (if $n>2$) setting $t_3 = \cdots = t_n = 0$, this yields $\rho_1 = \rho_2$. So $\{q, r, s\}$ is identifiable in $\mathcal{N}$.

It follows at once from the definition of identifiability that $\{q, r, s\}$ is also identifiable in $\mathcal{N}_0$ and that any collection of functions of (μ, σ^2, ρ) which depends on μ, σ^2, and ρ only through the value of (q, r, s) is identifiable in $\mathcal{N}$ and $\mathcal{N}_0$. Specifically if $t(\mu, \sigma^2, \rho) = \kappa^2$, defined in (7), and $u(\mu, \sigma^2, \rho) = \omega^2$, defined in (6), t and u are functions of r and s alone, and so $\{q, t, u\}$ is identifiable in $\mathcal{N}$. We can also show this directly: The right hand side of (8) can be written as

$$i\mu n\bar{t}-\tfrac{1}{2}\kappa^2(\sum t_j^2-n\bar{t}^2)-\tfrac{1}{2}\omega^2 n\bar{t}^2 .$$

Thus for $t_2=\cdots=t_n=0$ and $t_1\neq 0$, the identity corresponding to (9) yields $\mu_1=\mu_2$ and $\kappa_1^2=\kappa_2^2$, and on substitution of these equalities becomes

$$-\tfrac{1}{2}(\omega_1^2-\omega_2^2)n\bar{t}^2\equiv 0$$

so that also $\omega_1^2=\omega_2^2$.

5. Non-existence of a consistent estimator sequence for ρ in $\mathcal{N}_0$

If σ^2 and ρ are constant, ω^2 depends on n; in particular it equals $\omega_n^2=\kappa^2+n\sigma^2\rho$. By a consistent sequence of estimators of ω_n^2 is meant a sequence of functions f_n' of the observations such that for all $\varepsilon>0$

$$\lim \Pr\{|f_n'(y_1,\cdots,y_n)-\omega_n^2|>\varepsilon\}=0 . \tag{11}$$

If such a sequence exists then there also exists a sequence of functions f_n'' with

$$\lim \Pr\{|f_n''(y_1,\cdots,y_n)-\tau_n^2|>\varepsilon\}=0 ,$$

where

$$\tau_n^2=n^{-1}\omega_n^2 .$$

Moreover, writing

$$\tau^2=\lim \tau_n^2=\sigma^2\rho ,$$

we also have

$$\lim \Pr\{|f_n''(y_1,\cdots,y_n)-\tau^2|>\varepsilon\}=0 . \tag{12}$$

We shall show that no such sequence exists[2]. Note that only κ^2 and τ_n^2, but also κ^2 and τ^2, or κ^2 and ρ are independent parameters; it follows that no consistent estimator sequence exists for ρ.

To show that no sequence satisfying (12) exists, we first change variables from $(y_1,\cdots,y_n)$ to $(z_1,\cdots,z_n)$ with $z_1=n^{1/2}\bar{y}$, and the vector of the other components denoted by z^*, having a distribution $\varphi_{n-1}^*(z^*|\kappa^2)$ independent of μ and τ_n^2 or τ^2. That this can be done follows from (1) and (2) of section 2 and is shown more explicitly in section 6. In particular denote $n^{-1/2}z_1$ by $\bar{y}_n$ and $n^{-1/2}z^*$ by z_{n-1}^*. Then the joint density, φ_n, of $\bar{y}_n$ and z_{n-1}^* is

[2] The proof is rather technical. Intuitively the essential part of the argument is that τ^2 is identically zero if $\rho=0$ but may take on any positive values otherwise.

$$\varphi_n(\bar{y}_n, z^*_{n-1} \mid \mu, \kappa^2, \tau^2_n) = (2\pi\tau^2_n)^{-1/2} \exp\{-\tfrac{1}{2}(\bar{y}_n - \mu)^2\tau_n^{-2}\}\varphi^*_{n-1}(n^{1/2}z^*_{n-1} \mid \kappa^2)n^{1/2} .$$

Now it is shown in [6] that when τ can take two different values $\tau(1)$ and $\tau(2)$, then, for a sequence of functions f_n over $(\bar{y}_n, z^*_{n-1})$ constituting an estimator[3] sequence of τ to exist, it is necessary that, for $n \to \infty$, the "affinity", A, defined by

$$A(\tau^2_{1n}, \tau^2_{2n}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{\varphi_n(\bar{y}_n, z^*_{n-1} \mid \mu, \kappa^2, \tau^2_{1n})\varphi_n(\bar{y}_n, z^*_{n-1} \mid \mu, \kappa^2, \tau^2_{2n})\}^{1/2} d\bar{y}_n \, dz^*_{n-1}$$

converges to zero. Here

$$\tau^2_{1n} = \tau^2(1) + \kappa^2 n^{-1} , \qquad \tau^2_{2n} = \tau^2(2) + \kappa^2 n^{-1} .$$

Since the integral with respect to z^*_{n-1} gives unity, we have

$$\begin{aligned} A(\tau^2_{1n}, \tau^2_{2n}) &= (2\pi\tau_{1n}\tau_{2n})^{-1/2}\int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}(\bar{y}_n - \mu)^2\bar{\tau}_n^{-2}\}\, d\bar{y}_n \\ &= \bar{\tau}_n(\tau_{1n}\tau_{2n})^{-1/2} \end{aligned}$$

with

$$\bar{\tau}_n^{-2} = \tfrac{1}{2}(\tau_{1n}^{-2} + \tau_{2n}^{-2}) = \tfrac{1}{2}(\tau^2_{1n} + \tau^2_{2n})\tau_{1n}^{-2}\tau_{2n}^{-2} ,$$

so that

$$A^2(\tau^2_{1n}, \tau^2_{2n}) = 2\tau_{1n}\tau_{2n}/(\tau^2_{1n} + \tau^2_{2n}) .$$

So for $\tau(1)$ and $\tau(2)$ both positive, $A(\tau^2_{1n}, \tau^2_{2n})$ does not converge to zero. Thus there exist no consistent estimator sequences for τ when τ can take on any two positive values and consequently none for τ when τ can be any nonnegative number or for ρ when ρ is nonnegative.

6. A transformation and an extension

We have used here the fact that $\bar{y}$ and V are independently distributed according to (3) and (4). This was shown by Walsh [9], but an examination of his proof shows that his argument is valid for ρ independent of n only if $\rho \geqq 0$. Another argument, valid for the range (2), was given by Stuart [8]. It may be desirable, however, to give a more direct proof, and at the same time consider a more general form of problem.

[3] [6] actually is in terms of tests, namely functions φ_n which range over the unit interval and whose means converge to 0 for all distributions compatible with the null hypothesis and to 1 for all other distributions. Let $A_{n\epsilon i}$ be the set of observations for which $|f_n - \tau(i)| > \varepsilon$ and $\varphi_{n\epsilon i}$ the characteristic function of this set. Then $f_2, f_3, \cdots$ is a consistent estimator sequence of τ if, given $\varepsilon > 0$, there is an N such that for $n > N$ the mean of $\varphi_{n\epsilon i}$ is less than ε if $\tau = \tau(i)$ and more than $1 - \varepsilon$ otherwise. We can now use the $\varphi_{n\epsilon i}$ instead of the φ_n.

For that we change the assumption $\mathcal{E}y_j=\mu$ to $\mathcal{E}y_j=\mu+\sum_{p=1}^{k} C_p x_{pj}$ with $0\leqq k<n-1$. Here the $\boldsymbol{x}_p$ are fixed and known, linearly independent vectors; without loss of generality we assume that for each n the components of $\boldsymbol{x}_p$ add to zero. Let

$$\boldsymbol{y}=(y_1, \cdots, y_n)\ ,\quad \boldsymbol{u}=(u_1, \cdots, u_n)=\boldsymbol{y}-\mathcal{E}\boldsymbol{y}\ ,\quad \mathcal{E}u_j u_{j'}=\begin{cases}\sigma^2 & \text{if } j=j'\\ \sigma^2\rho & \text{if } j\neq j'\end{cases}\ ,$$

$$\boldsymbol{x}_p=(x_{p1}, \cdots, x_{pn})\ ,\qquad \boldsymbol{x}'=[\boldsymbol{x}_1', \cdots, \boldsymbol{x}_k']\ ,\qquad \boldsymbol{C}=(C_1, \cdots, C_k)\ .$$

Like in the case in which ρ is known to vanish, our objectives are attained by using an orthogonal matrix, $\boldsymbol{H}$, in which each element in the first column is equal to $n^{-1/2}$ and the other columns have the sum of elements equal to 0. For $\boldsymbol{z}=\boldsymbol{y}\boldsymbol{H}$, $z_1=n^{1/2}\bar{y}$ and the covariance matrix of z is

$$\begin{aligned}\boldsymbol{H}'\mathcal{E}(\boldsymbol{u}'\boldsymbol{u})\boldsymbol{H}&=\sigma^2\boldsymbol{H}'\{\rho(1 \cdots 1)'(1 \cdots 1)+(1-\rho)\boldsymbol{I}\}\boldsymbol{H}\\ &=\sigma^2\begin{bmatrix}1+m\rho & \boldsymbol{O}\\ \boldsymbol{O}' & (1-\rho)\boldsymbol{I}\end{bmatrix}.\end{aligned}$$

Since the rows of $\boldsymbol{x}$ add to 0, the first column of $\boldsymbol{xH}$ is a zero column; call the remaining columns $\boldsymbol{x}^*$. If we denote $(z_2, \cdots, z_n)$ by $\boldsymbol{z}^*$, we have: z_1 and $\boldsymbol{z}^*$ are independently and normally distributed, the former with mean $\nu=n^{1/2}\mu$ and variance $\omega^2=\sigma^2(1+m\rho)$, the latter with a vector mean $\boldsymbol{Cx}^*$ and covariance matrix $\kappa^2\boldsymbol{I}$ with $\kappa^2=\sigma^2(1-\rho)$; ν and the components of $\boldsymbol{C}$ range over the entire real line, while ω^2 and κ^2 range over the entire positive line.

The analysis of $\boldsymbol{z}^*$ is an ordinary regression problem (through the origin); for example, the minimum variance linear unbiased estimate of $\boldsymbol{C}$ is

$$\begin{aligned}\boldsymbol{C}^0=\boldsymbol{z}^*\boldsymbol{x}^{*\prime}(\boldsymbol{x}^*\boldsymbol{x}^{*\prime})^{-1}&=[z_1\ \boldsymbol{z}^*][\boldsymbol{O}'\ \boldsymbol{x}^*]'\{[\boldsymbol{O}'\ \boldsymbol{x}^*][\boldsymbol{O}'\ \boldsymbol{x}^*]'\}^{-1}\\ &=\boldsymbol{z}(\boldsymbol{xH})'(\boldsymbol{xHH}'\boldsymbol{x}')^{-1}=\boldsymbol{yx}'(\boldsymbol{xx}')^{-1}\ ,\end{aligned}$$

and the usual estimate of κ^2 is $(m-k)^{-1}V$ with

$$\begin{aligned}V=\|\boldsymbol{z}^*-\boldsymbol{C}^0\boldsymbol{x}^*\|^2&=\boldsymbol{z}\{\boldsymbol{I}-\boldsymbol{x}^{*\prime}(\boldsymbol{x}^*\boldsymbol{x}^{*\prime})^{-1}\boldsymbol{x}^*\}\boldsymbol{z}'-z_1^2\\ &=\boldsymbol{y}\{\boldsymbol{I}-\boldsymbol{x}'(\boldsymbol{xx}')^{-1}\boldsymbol{x}\}\boldsymbol{y}'-n\bar{y}^2=\|\boldsymbol{y}-\bar{y}(1 \cdots 1)-\boldsymbol{C}^0\boldsymbol{x}\|^2\ .\end{aligned}$$

Note that these estimates are precisely the ones obtained for $\boldsymbol{C}$ and κ^2 in the case in which $\rho=0$.

It is now easily seen that z_1, $\boldsymbol{C}^0$ and V are sufficient for the family of distributions of $\boldsymbol{y}$, and that equations (1), (3), (4) and (5) are valid. However, in (4) and (5) m is replaced by $m-k$, and in (1) and (3) μ is replaced by $\mu + \sum_{i=1}^{k} C_p x_{pj}$; and, when $k>0$, the logarithm of the joint density of the components of $\boldsymbol{C}^0$ is proportional to

$$c_5 - k \log \sigma^2(1-\rho) - (\boldsymbol{C}^0 - \boldsymbol{C})\boldsymbol{x}\boldsymbol{x}'(\boldsymbol{C}^0 - \boldsymbol{C})'/\sigma^2(1-\rho) \, .$$

TEL AVIV UNIVERSITY, ISRAEL[4]

REFERENCES

[1] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh, London, 10th ed., XV, p. 354, 1946.
[2] M. Hansen, W. N. Hurwitz and W. G. Madow, *Sample Survey Methods and Theory*, John Wiley & Sons, New York, Chapman Hall, London, 2 vols., 1953.
[3] H. S. Konijn, "Regression analysis in sample surveys," *J. Amer. Statist. Ass.*, 57 (1962), 590-606.
[4] H. S. Konijn, "Note on the nonexistence of a maximum likelihood estimate," *Aust. J. Statist.*, 5 (1963), 143-146.
[5] H. S. Konijn, "On a theorem of Halmos concerning unbiased estimation of moments," *J. Aust. Math. Soc.*, 4 (1964), 229-232.
[6] C. Kraft, "Some conditions for consistency and uniform consistency of statistical procedures," *University of California Publications in Statistics*, 2 (1955), no. 6, 125-142.
[7] C. Kraft and L. LeCam, "A remark on the roots of the maximum likelihood equations," *Ann. Math. Statist.*, 27 (1956), 1174-1177.
[8] A. Stuart, "Equally correlated variates and the multinormal integral," *J. R. Statist. Soc.*, Ser. B, 20 (1958), 373-378.
[9] J. E. Walsh, "Concerning the effect of intraclass correlation on certain significance tests," *Ann. Math. Statist.*, 18 (1947), 88-96.

[4] A draft of this paper was produced at the Cowles Foundation for Research in Economics.