Cowles Foundation Paper 214a

Reprinted from *The Australian Journal of Statistics*, Vol. 5, No. 3,
pages 143–146, November, 1963.

NOTE ON THE NON-EXISTENCE OF A MAXIMUM LIKELIHOOD ESTIMATE[1]

H. S. Konijn

*Cowles Foundation for Research in Economics, Yale University**

1. Summary

It is shown that in a regression model with intraclass correlation ρ, maximum likelihood estimates and minimum variance unbiased linear estimates of the regression coefficients are unaffected by the value of ρ when ρ is given, and that the former do not exist when ρ is not specified. The usual confidence intervals for the regression slopes remain valid, but there are difficulties in giving a confidence interval for the intercept.

2. Introduction and Statement of Result

In Halperin (1951) the following situation is examined. Writing

$$\underset{\sim}{y}=(y_1, \ldots, y_n),\ \underset{\sim}{1}=(1, \ldots, 1),\ \underset{\sim}{x}_p=(x_{p1} \ldots, x_{pn})$$

with $p=1, 2, \ldots, k$ and $0\leq k<n-1$, let $\underset{\sim}{y}$ be multivariate non-singular normal with mean $\mu\underset{\sim}{1}$ if $k=0$ and

$$\mu\underset{\sim}{1}+\sum_{p=1}^{k} C_p\underset{\sim}{x}_p$$

otherwise, and with covariance matrix

$$\underset{\sim}{R}_n=\sigma^2\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ - & - & - & - \\ \rho & \rho & \cdots & 1 \end{bmatrix}.$$

Here the $\underset{\sim}{x}_p$ are fixed and known, linearly independent vectors with $\underset{\sim}{x}_p\underset{\sim}{1}'=0$; and μ, $\underset{\sim}{C}=(C_1, \ldots, C_k)$, $\sigma^2>0$ and ρ are unknown. It is stated in Halperin (1951, p. 574) that "the maximum likelihood equations for the estimation of [the] parameters . . . are of such a formidable character that an explicit solution does not appear possible."

It is shown here that no solution exists due to the fact that the likelihood does not possess a maximum. Nevertheless, if we write

---

[1] Manuscript received July 9, 1963. Presented at the annual September, 1962 meeting of the Institute of Mathematical Statistics.

* Present address: Dept. of Economics, City College, New York 31.

μ^0 and $\underset{\sim}{C}^0$ for the maximum likelihood estimates of μ and $\underset{\sim}{C}$ for the case in which ρ is known to be 0:

$$\mu^0 = n^{-1}\underset{\sim}{y}\underset{\sim}{1}', \quad \underset{\sim}{C}^0 = \underset{\sim}{y}\underset{\sim}{x}'(\underset{\sim}{x}\underset{\sim}{x}')^{-1}, \quad (\underset{\sim}{x}' = [\underset{\sim}{x}_1', \ldots, \underset{\sim}{x}_k'])$$

it turns out that μ^0 and $\underset{\sim}{C}^0$ also maximize the likelihood for *any given* ρ. In fact, $(\mu^0, \underset{\sim}{C}^0)$ is also a minimum variance unbiased linear estimate of $(\mu, \underset{\sim}{C})$ for any given ρ, namely that element in the class of unbiased estimates of $(\mu, \underset{\sim}{C})$ linear in $\underset{\sim}{y}$ which has the smallest generalized variance and has the smallest variance for the estimate of any estimable linear function of $(\mu, \underset{\sim}{C})$, no matter what be the form of the distribution of $\underset{\sim}{y}$ or the values of its first and second moments.

If ρ is known to be non-negative but otherwise unknown, the likelihood equations still do not possess a solution, but the maximum of the likelihood is attained at $\rho = 0$ and so the estimate does not converge stochastically to the true value when $\rho \neq 0$.

3. Proofs

Let $\underset{\sim}{E} = [\underset{\sim}{1}', \ldots, \underset{\sim}{1}']$. Then $R_n = \sigma^2\rho\underset{\sim}{E} + \sigma^2(1-\rho)\underset{\sim}{I}$ and

$$\sigma^2(1-\rho)(1+m\rho)\underset{\sim}{R}_n^{-1} = -\rho\underset{\sim}{E} + (1+m\rho)\underset{\sim}{I}, \quad (m = n-1).$$

That the minimum variance unbiased linear estimates of μ and $\underset{\sim}{C}$ for any given ρ are μ^0 and $\underset{\sim}{C}^0$ now follows by substitution in the formulae for such estimates (Aitken (1934–5)) for the case of a general covariance matrix of the residuals which is fully specified up to a scale factor.

Minus twice the logarithm of the likelihood of $\underset{\sim}{y}$ is, apart from a constant,

$$n \log \sigma^2 + m \log (1-\rho) + \log (1+m\rho) + \frac{(1+m\rho)\underset{\sim}{u}\underset{\sim}{u}' - \rho(\underset{\sim}{u}\underset{\sim}{1}')^2}{\sigma^2(1-\rho)(1+m\rho)}, \tag{1}$$

where $\underset{\sim}{u} = \underset{\sim}{y} - \mathscr{E}\underset{\sim}{y}$. Here we have substituted $(\underset{\sim}{u}\underset{\sim}{1}')^2 - \underset{\sim}{u}\underset{\sim}{u}'$ for the sum of the cross-products of the elements of $\underset{\sim}{u}$.

From this it is seen that the distribution of $\underset{\sim}{y}$ is non-singular if and only if σ^2 is positive and

$$-\frac{1}{m} < \rho < 1. \tag{2}$$

The partial derivative of (1) with respect to μ is $\utilde{u1}'/\sigma^2(1+m\rho)$; since $\sigma^2>0$ and $m\rho>-1$, this can vanish only if $\utilde{u1}'=0$. For the partial derivative of (1) with respect to $\utilde{C}$ we find (after setting $\utilde{u1}'=0$) $\utilde{ux}'/\sigma^2(1-\rho)$; since $\sigma^2>0$ and $\rho<1$, this can vanish only if $\utilde{ux}'=0$. But $\utilde{u1}'=0$ and $\utilde{ux}'=0$ together constitute the usual normal equations for μ^0 and C^0.

Upon substituting $\utilde{u1}'=0$, the partial derivative of (1) with respect to σ^2 is $n/\sigma^2-\utilde{uu}'/\sigma^4(1-\rho)$, which can vanish only if $\sigma^2(1-\rho)=\utilde{uu}'/n$. Substituting this and $\utilde{u1}'=0$ in (1) we obtain, apart from a constant,

$$-\log(1-\rho)+\log(1+m\rho). \tag{3}$$

As a function of ρ this is steadily increasing over the range (2) and tends to infinite values at the ends of the range. Consequently, the likelihood does not possess a maximum in (2).

If the range of ρ is restricted *a priori* to non-negative values, the maximum of the likelihood is attained at $\rho=0$. Even so, since the derivative, $n(1-\rho)^{-1}(1+m\rho)^{-1}$, of (3) is positive, the likelihood equations fail to have a solution even in this case.

4. Confidence Intervals

The expected value of the sum V of squares of residuals is $\sigma^2(1-\rho)(n-k-1)$. Let $\utilde{D}$ be a $k\times s$ matrix of rank s and $\utilde{CD}$ represent s estimable linear combinations of $C_1,\ldots,C_k$. It is readily shown (see Halperin (1951) and Walsh (1947) for special cases) that if

$$W=(\utilde{C}^0-\utilde{C})\utilde{D}[\utilde{D}'(\utilde{xx}')^{-1}\utilde{D}]^{-1}\utilde{D}'(\utilde{C}^0-\utilde{C})',$$

$W/\sigma^2(1-\rho)$ and $V/\sigma^2(1-\rho)$ are independently distributed like χ^2 with s and $n-k-1$ degrees of freedom. It follows that the usual confidence intervals for any linearly independent set of estimable linear combinations of $C_1,\ldots,C_k$ (and so for any subset) using the F distribution are valid.

Now, writing

$$\Theta=\frac{1-\rho}{1+m\rho},$$

we can also show (Walsh (1947)) that

$$(n\Theta)^{\frac{1}{2}}(\mu^0-\mu)/\sigma(1-\rho)^{\frac{1}{2}}$$

is a standard normal variate distributed independently of W and V, so that the usual confidence interval for μ or combinations of μ and components of $\utilde{C}$ (such as a prediction of $\utilde{y}$ for given $\utilde{x}$) are not valid.

Sometimes, however, it is possible (Konijn (1962)) to obtain a crude estimate of Θ, in which case approximate confidence intervals may be obtained.

References

Aitken, A. C. (1934–5). " On Least Squares and Linear Combination of Observations." *Proc. Roy. Soc. Edinburgh*, 55, 42–48.

Halperin, M. (1951). " Normal Regression Theory in the Presence of Intraclass Correlation." *Ann. Math. Statist.*, 22, 573–580.

Konijn, H. S. (1962). " Regression Analysis in Sample Surveys." *J. Amer. Statist. Assoc.*, 57, 590–606.

Walsh, J. E. (1947). " Concerning the Effect of Intraclass Correlation on Certain Significance Tests." *Ann. Math. Statist.*, 18, 88–96.