PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 19, Fasc. 1 (1999), pp. 15–21

AN ANCILLARY PARADOX IN TESTING

BY

KUN HE* (LAWRENCE, KANSAS)

Abstract. In multiple linear regression with normally distributed errors, it is shown that a test procedure for a hypothesis about the intercept which is α-admissible when the design matrix is fixed is inadmissible when the design matrix is an ancillary statistic. The result of this paper is a complementary one to Brown's paper [2].

1. Introduction. The purpose of this paper is to show an ancillary paradox in testing which appears in a linear regression. It will be shown that a test procedure for a hypothesis involving the intercept is α-admissible when the design matrix is fixed, but the test procedure is inadmissible when the design matrix is an ancillary statistic.

Consider the usual multiple linear regression

$$Y_i = \mu + V_i^t \beta + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where $Y = (Y_1, \ldots, Y_n)^t$ is the dependent variable vector, $\mu \in R$, $\beta = (\beta_1, \ldots, \beta_p)^t \in R^p$ are unknown parameters, and $V_i = (V_{i1}, \ldots, V_{ip})^t$, $i = 1, \ldots, n$, are the predictor variables. The errors $(\varepsilon_1, \ldots, \varepsilon_n)^t$ are assumed to be normally distributed, i.e.,

$$(\varepsilon_1, \ldots, \varepsilon_n)^t \sim N(0, \sigma^2 I). \tag{1.2}$$

We are interested in testing for a hypothesis about the y-intercept value μ, i.e., the population mean of the dependent variables when the predictor variables are all zero.

The main purpose of this paper is to show that the admissibility of a test procedure for a hypothesis about μ depends on the distribution of the predictor variables, i.e., the test is α-admissible if the predictor variables are preassigned constant values, but it is inadmissible if the predictor variables are independent normal having mean 0 and identity covariance matrix. This result is a complementary to that of Brown [2].

* Partially supported by The University of Kansas General Research Fund.

Fisher [3] introduces the notion of an ancillary statistic partly as a basis for conditioning, which is an old and commonly used tool in statistical inference. Fisher [3] defines an *ancillary statistic* U as one that has a law independent of θ, and together with the m.l.e. $\hat{\theta}$ forms a sufficient statistic. Fisher's rationale for considering ancillarity is as follows: U by itself contains no information about θ, and does not affect $\hat{\theta}$. However, the value of U may tell us something about the precision of $\hat{\theta}$, e.g., $\mathrm{Var}_\theta(\hat{\theta} \mid U = u)$ might depend on u. It is widely believed that the value of ancillary statistic does not affect statistical inferences, i.e., statistical inference should be carried out conditional on the value of any ancillary statistic.

Brown [2] shows that in multiple linear regression the admissibility of the ordinary estimator of the constant term depends on the distribution of the predictor variables, which are ancillary statistics. He [4]–[7] extends Brown's results to various models, and He and Strawderman [8] discuss the estimation in elliptically contoured regression.

We will discuss a test procedure in Section 2 for a fixed design. We prove that a test procedure for a hypothesis about intercept μ is α-admissible when the predictor variables are fixed. In Section 3 we prove that the test procedure is inadmissible when the predictor variables are random with known normal distribution having mean 0 and identity covariance matrix.

2. The case of fixed design: Admissibility of test. We will first consider the case where the predictor variables V are fixed.

Under the model (1.1) and assumption (1.2) we know that

$$Y \sim N_n(\mathbf{1}\mu + V\beta, \sigma^2 I),$$

with an $(n \times p)$-matrix $V = (V_1, \ldots, V_n)^t$, and $\mathbf{1} = (1, \ldots, 1)^t \in R^n$.

Let $\bar{Y} = n^{-1}\mathbf{1}^t Y$ (a scalar), $\bar{V} = n^{-1}\mathbf{1}^t V$ (a $(1 \times p)$ row vector), and $S = (V - \mathbf{1}\bar{V})^t (V - \mathbf{1}\bar{V})$ (a $(p \times p)$-matrix and positive definite with probability 1). The least squared estimators of μ and β are, respectively, the following:

$$\hat{\mu} = \bar{Y} - \bar{V}\hat{\beta}, \tag{2.1}$$

$$\hat{\beta} = S^{-1} V^t (Y - \bar{Y}\mathbf{1}), \tag{2.2}$$

and

$$\begin{pmatrix}\hat{\mu} \\ \hat{\beta}\end{pmatrix} \sim N_{p+1}\left(\begin{pmatrix}\mu \\ \beta\end{pmatrix}, \Sigma(V)\right), \tag{2.3}$$

where

$$\Sigma(V) = \sigma^2 \begin{pmatrix} n^{-1} + \bar{V}S^{-1}\bar{V}^t & -\bar{V}S^{-1} \\ -S^{-1}\bar{V}^t & S^{-1} \end{pmatrix}.$$

We will consider testing the intercept μ in the regression model (1.1). Our hypothesis is

$$H_0\colon \mu \leqslant \mu_1 \text{ or } \mu \geqslant \mu_2 \ (\mu_1 < \mu_2), \qquad H_a\colon \mu_1 < \mu < \mu_2. \tag{2.4}$$

A test ϕ_0 is called *α-admissible* (Lehmann [9], p. 306) if, for any other level-α test ϕ,

$$E_\mu \phi(Y) \geqslant E_\mu \phi_0(Y) \quad \text{for all } \mu \in H_a$$

implies

$$E_\mu \phi(Y) = E_\mu \phi_0(Y) \quad \text{for all } \mu \in H_a.$$

This definition takes no account of the relationship of $E_\mu \phi(Y)$ and $E_\mu \phi_0(Y)$ for $\mu \in H_0$ beyond the requirement that both tests are of level α.

Let $I(A)$ be the indicator function of the set A. We have the following lemma:

LEMMA 2.1. *For given σ^2 and V, the test $\phi_0(\hat{\mu}) = I(c_1 < \hat{\mu} < c_2)$ is α-admissible for the hypothesis* (2.4) *if and only if*

$$E_{\mu_1}\phi_0(\hat{\mu}) = E_{\mu_2}\phi_0(\hat{\mu}) = \alpha, \tag{2.5}$$

where α is the size of the test.

Proof. From formula (2.3) we know that $\hat{\mu} \sim N(\mu, \sigma_\mu^2)$, where $\sigma_\mu^2 = (n^{-1} + \bar{V}S^{-1}\bar{V}^t)\sigma^2$. By Theorem 6 of Lehmann [9], p. 82, the test $\phi_0(\hat{\mu}) = I(c_1 < \hat{\mu} < c_2)$ is the UMP test.

If (2.5) holds, the test $\phi_0(\hat{\mu})$ is the UMP unbiased test, then it is α-admissible.

Suppose $\phi_0(\hat{\mu})$ is α-admissible and (2.5) does not hold; then using the same method as in Example 12 of Lehmann [9], p. 306, we see that $\phi_0(\hat{\mu})$ is not α-admissible. ∎

3. The case of random design: Inadmissibility of test. In this section we will assume that $V = (V_1, \ldots, V_n)^t$ is random with distribution

$$V_i \sim N_p(0, I), \quad i = 1, \ldots, n, \ p \geqslant 3. \tag{3.1}$$

The usual least squared estimator of μ is still

$$\hat{\mu} = \bar{Y} - \bar{V}\hat{\beta}.$$

Following ideas in Brown [2],

$$\tilde{\mu} = \bar{Y} - \bar{V}\tilde{\beta}(\hat{\beta}, S) = \hat{\mu} + \bar{V}(\hat{\beta} - \tilde{\beta}(\hat{\beta}, S)) \tag{3.2}$$

will be used as a competitive estimator of $\hat{\mu}$, where $\tilde{\beta}$ is a certain function of $\hat{\beta}$ and S. Using the above estimators, we construct a competitive test as follows:

$$\phi_1(\tilde{\mu}) = I(c_1 \leqslant \tilde{\mu} \leqslant c_2)$$

for the hypothesis defined in (2.4), which is

$$H_0\colon \mu \leqslant \mu_1 \text{ or } \mu \geqslant \mu_2, \quad H_a\colon \mu_1 < \mu < \mu_2,$$

where $\mu_1 < \mu_2$ are given constants. As in Lemma 2.1, for given σ^2 and fixed V, the test ϕ_0 is α-admissible. However, when V satisfies the assumption (3.1) and when $\phi_0(\hat{\mu}) = I(-c \leqslant \hat{\mu} \leqslant c)$ for c sufficiently small, then ϕ_0 is inadmissible when $\mu_1 = -\mu_2$ is also sufficiently small.

THEOREM 3.1. *In the linear regression model* (1.1), (1.2), (3.1) *for given* $\sigma^2 > 0$, $p \geqslant 3$, *there exist* $\mu_1 = -\mu_2$, *and an estimator* $\tilde{\mu}$ *such that for the hypothesis* (2.4) *and a given* $\beta \neq 0$, *we have*

$$E_{\mu,\beta}\,\phi_1(\tilde{\mu}) > E_{\mu,\beta}\,\phi_0(\hat{\mu}) \quad \text{for } \mu_1 < \mu < \mu_2,$$

where $\phi_1(\tilde{\mu}) = I(-c^* < \tilde{\mu} < c^*)$, *and* c^* *is chosen such that the test has the same size* α *as* ϕ_0, $\alpha = E_{\mu_1}\phi_0(\hat{\mu}) = E_{\mu_2}\phi_0(\hat{\mu})$.

Proof. Note that $E(\bar{Y}\,|\,V) = \mu + \bar{V}\beta$, and $\bar{Y}$ is conditionally independent of $\hat{\beta}$ and S given V, and V is independent of $\hat{\beta}$ and S. Thus, by (3.2),

$$\begin{aligned}
E_{\mu,\beta}\,\phi_1(\tilde{\mu}) &= P_{\mu,\beta}(-c \leqslant \tilde{\mu} \leqslant c)\\
&= P_{\mu,\beta}(-c-\mu \leqslant \tilde{\mu}-\mu \leqslant c-\mu)\\
&= E_{\mu,\beta}\,P_{\mu,\beta}(-c-\mu \leqslant \tilde{\mu}-\mu \leqslant c-\mu \mid \hat{\beta}, S, V)\\
&= E_{\mu,\beta}\,P_{\mu,\beta}(-c-\mu \leqslant \bar{Y}-E(\bar{Y}\,|\,V)-\bar{V}(\tilde{\beta}-\beta) \leqslant c-\mu \mid \hat{\beta}, S, V)\\
&= E_\beta\big(\Phi(\sqrt{n}\,[\bar{V}(\tilde{\beta}-\beta)+c-\mu])-\Phi(\sqrt{n}\,[\bar{V}(\tilde{\beta}-\beta)-c-\mu])\big)\\
&= E_\beta\,E\{\Phi(\sqrt{n}\,[\bar{V}(\tilde{\beta}-\beta)+c-\mu])-\Phi(\sqrt{n}\,[\bar{V}(\tilde{\beta}-\beta)-c-\mu]) \mid \hat{\beta}, S\}\\
&= E_\beta \int_{-\infty}^{\infty} [\Phi(\|\tilde{\beta}-\beta\|t+\sqrt{n}(c-\mu))-\Phi(\|\tilde{\beta}-\beta\|t-\sqrt{n}(c+\mu))]\,f(t)\,dt\\
&= E_\beta\,G(\|\tilde{\beta}-\beta\|, \mu),
\end{aligned}$$

where

$$G(x, \mu) = \int_{-\infty}^{\infty} [\Phi(xt+\sqrt{n}(c-\mu))-\Phi(xt-\sqrt{n}(c+\mu))]\,f(t)\,dt, \qquad x \geqslant 0,$$

and $\Phi(x)$ and $f(x)$ are a standard normal cumulative distribution function and a density function, respectively.

Let us define

$$\lambda(\mu) = E_\beta[G(\|\tilde{\beta}-\beta\|, \mu)] - E_\beta[G(\|\hat{\beta}-\beta\|, \mu)].$$

We will show first the following two steps: Step (i) $\lambda(0) > 0$ and Step (ii) $\lambda(\mu)$ is a decreasing function of μ for sufficiently small $\mu > 0$.

Step (i). Let $L(x) = 2\Phi(\sqrt{n}c)-1-G(x, 0)$. The function $W(\tilde{\beta}-\beta) = L(\|\tilde{\beta}-\beta\|)$ can be thought of as a loss function for estimating β if we can show that $L(x)$ is an increasing function of $x \geqslant 0$. Let

$$G'_x(x, \mu) = \frac{\partial}{\partial x}\,G(x, \mu).$$

Note that

$$\frac{d}{dx}\,L(x) = -G'_x(x, 0) = \sqrt{\frac{2n}{\pi}}\,cx(x^2+1)^{-3/2}\exp\left\{-\frac{nc^2}{2(x^2+1)}\right\} \geqslant 0,$$

and $L(0) = 0$; then $L(x)$ is strictly increasing in x for $x \geqslant 0$. Furthermore, $L(x)$ is bounded above by $2\Phi(\sqrt{nc})-1$. Note that $L(x)$ is not a convex function, so Theorem 3.3.1 of Brown [1] will be applied.

Since $\hat{\beta} \mid S \sim N_p(\beta, \sigma^2 S^{-1})$, conditional on S, we want to find an estimator $\tilde{\beta} = \tilde{\beta}(\hat{\beta}, S)$ such that

$$E_\beta [W(\tilde{\beta}-\beta) \mid S] < E_\beta [W(\hat{\beta}-\beta) \mid S].$$

By Theorem 3.3.1 of Brown [1], let

$$\tilde{\beta} = \left(I - \frac{A}{a+\|\hat{\beta}\|^2}\right)\hat{\beta}, \tag{3.3}$$

where I is an identity matrix, a is a sufficiently large number, and

$$A = \frac{1}{b}[EXW'(X)]^{-1}, \quad X \sim N_p(0, \sigma^2 S^{-1}); \tag{3.4}$$

here $x = (x_1, \ldots, x_p)^t$, and

$$W'(x) = \left(\frac{\partial}{\partial x_1} W(x), \ldots, \frac{\partial}{\partial x_p} W(x)\right).$$

Since

$$XW'(X) = \frac{-G'_x(\|X\|, 0)}{\|X\|} XX^t$$

is a positive definite matrix, we know that A is positive definite. Therefore, Theorem 3.3.1 of Brown [1] can be applied. This completes the proof of Step (i).

Step (ii). Since

$$\lambda(\mu) = E_\beta [G(\|\tilde{\beta}-\beta\|, \mu)] - E_\beta [G(\|\hat{\beta}-\beta\|, \mu)],$$

from the result of Step (i) we know that $\lambda(0) > 0$.

Let

$$G'_\mu(x, \mu) = \frac{\partial}{\partial \mu} G(x, \mu).$$

We have

$$G'_\mu(x, \mu) = (-\sqrt{n}) \int_{-\infty}^{\infty} [f(xt+\sqrt{n}(c-\mu)) - f(xt-\sqrt{n}(c+\mu))] f(t)\, dt$$

$$= [2\pi(x^2+1)]^{-1/2}\left(\exp\left\{-\frac{n(c+\mu)^2}{2(x^2+1)}\right\} - \exp\left\{-\frac{n(c-\mu)^2}{2(x^2+1)}\right\}\right).$$

Therefore, $G'_\mu(x, 0) = 0$, and $G'_\mu(x, \mu) < 0$ for $\mu > 0$.

Since $G'_\mu(x, 0) = 0$, we have $\lambda'(0) = 0$. To prove Step (ii), it is sufficient to show that $\lambda''(0) < 0$ for sufficiently small $\mu > 0$.

Let us show that $\lambda''(0) < 0$. We will define a suitable loss function and apply Theorem 3.3.1 of Brown [1] again. Let

$$U(x, \mu) = \frac{\partial^2}{\partial \mu^2} G(x, \mu);$$

then

$$U(x, 0) = -\frac{2nc}{\sqrt{2\pi}}(x^2+1)^{-3/2}\exp\left\{-\frac{nc^2}{2(x^2+1)}\right\}.$$

Defining $W_1(\tilde{\beta}-\beta) = U(\|\tilde{\beta}-\beta\|, 0)$ as a loss function for estimating β, we obtain

$$-\lambda''(0) = E_\beta W_1(\hat{\beta}-\beta) - E_\beta W_1(\tilde{\beta}-\beta).$$

Using results of [1], p. 1131, we have

$$-\lambda''(0) > \frac{E\big(W_1'(X)AX\big)}{a+\|\beta\|^2} + o\left(\frac{1}{b}\right) + o\left(\frac{1}{a+\|\beta\|^2}\right),$$

where a, b and A are defined in (3.3) and (3.4). To show that $\lambda''(0) < 0$, it is sufficient to prove that $bE\big(W_1'(X)AX\big) > \eta > 0$, where η is a positive constant. Note that for small constant c we have

$$U'(x, 0) = \frac{\partial}{\partial x}U(x, 0) = \frac{2nc}{3\sqrt{2\pi}}(x^2+1)^{-7/2}\left(x^2+1-\frac{nc^2}{3}\right) > 0.$$

Since

$$W_1'(X) = U'(\|X\|, 0)X^t/\|X\|,$$

we have

$$bE\big(W_1'(X)AX\big) = E\left[\frac{U'(\|X\|, 0)}{\|X\|}X^t(bA)X\right] > 0.$$

If we let η equal the above number, we prove that $\lambda''(0) < 0$.

Since $G(x, \mu)$ is continuous in c, $G(x, -\mu) = G(x, \mu)$, and $G(x, \mu)$ is decreasing in μ for $\mu > 0$, we can choose $0 < c^* < c$ such that $\phi_1(\tilde{\mu}) = I(-c^* \leqslant \tilde{\mu} \leqslant c^*)$ has size α. Then for $\mu_1 < \mu < \mu_2$ we obtain

$$E_{\mu,\beta}\phi_1(\tilde{\mu}) > E_{\mu,\beta}\phi_0(\hat{\mu}),$$

which completes the proof. ■

Acknowledgment. The author thanks Dr. Kunliang Lu for many discussions and help, and the referee for the helpful comments.

REFERENCES

[1] L. D. Brown, *On the admissibility of invariant estimators of one or more location parameters*, Ann. Math. Statist. 37 (1966), pp. 1087–1135.

[2] — *An ancillarity paradox which appears in multiple linear regression (with discussion)*, Ann. Statist. 18 (1990), pp. 471–538.

[3] R. A. Fisher, *The logic of inductive inference (with discussion)*, J. Roy. Statist. Soc. Ser. A 98 (1935), pp. 39–54.

[4] K. He, *An ancillarity paradox in the estimation of multinomial probabilities*, J. Amer. Statist. Assoc. 85 (1990), pp. 824–828.

[5] — *The estimation of stratum means vector with random sample sizes*, J. Statist. Plann. Inference 37 (1993), pp. 43–50.

[6] — *On estimating a linear combination of stratum means with random sample sizes*, J. Multivariate Analysis 55 (1995), pp. 39–60.

[7] — *On estimating domain totals over a subpopulation*, Ann. Inst. Statist. Math. 47 (1995), pp. 637–644.

[8] — and W. E. Strawderman, *Estimation in the elliptically contoured regression with random design*, manuscript under review, 1995.

[9] E. L. Lehmann, *Testing Statistical Hypotheses*, 2nd edition, Wiley, New York 1986.

5600 Fishers Lane, Room 15B-45
DBI/HFD-710
Food and Drug Administration
Rockville, MD 20857, U.S.A.

Received on 4.11.1997;
revised version on 16.10.1998