ADMISSIBILITY OF THE USUAL CONFIDENCE INTERVAL IN LINEAR REGRESSION

SHORT RUNNING TITLE: ADMISSIBILITY IN REGRESSION

PAUL KABAILA[1*], KHAGESWOR GIRI[2] AND HANNES LEEB[3]

1. Department of Mathematics and Statistics
La Trobe University
Bundoora Victoria 3086
Australia

2. Future Farming Systems Research
Department of Primary Industries
600 Sneydes Road
Werribee 3030 Victoria
Australia

3. Department of Statistics
University of Vienna
Universitätsstr. 5/3, A-1010 Vienna
Austria

Summary

Consider a linear regression model with independent and identically normally distributed random errors. Suppose that the parameter of interest is a specified linear combination of the regression parameters. We prove that the usual confidence interval for this parameter is admissible within a broad class of confidence intervals.

Keywords: admissibility; compromise decision theory; confidence interval; decision theory.

* Author to whom correspondence should be addressed. Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia. Tel.: +61 3 9479 2594, fax: +61 3 9479 2466, e-mail: P.Kabaila@latrobe.edu.au

1. Introduction

Consider the linear regression model $Y = X\beta + \varepsilon$, where Y is a random n-vector of responses, X is a known $n \times p$ matrix with linearly independent columns, β is an unknown parameter p-vector and $\varepsilon \sim N(0, \sigma^2 I_n)$ where σ^2 is an unknown positive parameter. Let $\hat{\beta}$ denote the least squares estimator of β. Also, define $\hat{\sigma}^2 = (Y - X\hat{\beta})^T (Y - X\hat{\beta})/(n-p)$.

Suppose that the parameter of interest is $\theta = a^T \beta$ where a is a given p-vector ($a \neq 0$). We seek a $1-\alpha$ confidence interval for θ. Define the quantile $t(m)$ by the requirement that $P\big(-t(m) \le T \le t(m)\big) = 1 - \alpha$ for $T \sim t_m$. Let $\hat{\Theta}$ denote $a^T\hat{\beta}$, i.e. the least squares estimator of θ. Also let v_{11} denote the variance of $\hat{\Theta}$ divided by σ^2. The usual $1-\alpha$ confidence interval for θ is

$$I = \big[\hat{\Theta} - t(m)\sqrt{v_{11}}\hat{\sigma},\; \hat{\Theta} + t(m)\sqrt{v_{11}}\hat{\sigma}\big]$$

where $m = n - p$. Is this confidence interval admissible? The admissibility of a confidence interval is a much more difficult concept than the admissibility of a point estimator, since confidence intervals must satisfy a coverage probability constraint. Also, admissibility of confidence intervals can be defined in either weak or strong forms (Joshi, 1969, 1982).

Kabaila & Giri (2009, Section 3) describe a broad class $\mathcal{D}$ of confidence intervals that includes I. The main result of the present paper, presented in Section 3, is that I is strongly admissible within the class $\mathcal{D}$. An attractive feature of the proof of this result is that, although lengthy, this proof is quite straightforward and elementary. Section 2 provides a brief description of this class $\mathcal{D}$. For completeness, in Section 4 we describe a strong admissibility result, that follows from the results of Joshi (1969), for the usual $1-\alpha$ confidence interval for θ in the somewhat artificial situation that the error variance σ^2 is assumed to be known.

2. Description of the class $\mathcal{D}$

Define the parameter $\tau = c^T\beta - t$ where the vector c and the number t are given and a and c are linearly independent. Let $\hat{\tau}$ denote $c^T\hat{\beta} - t$ i.e. the least squares estimator of τ. Define the matrix V to be the covariance matrix of $(\hat{\Theta}, \hat{\tau})$ divided by σ^2. Let v_{ij} denote the (i, j) th element of V. We use the notation $[a \pm b]$ for the interval $[a - b, a + b]$ $(b > 0)$. Define the following confidence interval for θ

$$J(b, s) = \left[\hat{\Theta} - \sqrt{v_{11}} \hat{\sigma}\, b\left(\frac{\hat{\tau}}{\hat{\sigma}\sqrt{v_{22}}} \right) \pm \sqrt{v_{11}} \hat{\sigma}\, s\left(\frac{|\hat{\tau}|}{\hat{\sigma}\sqrt{v_{22}}} \right) \right] \tag{1}$$

where the functions b and s are required to satisfy the following restrictions. The function $b : \mathbb{R} \to \mathbb{R}$ is an odd function and $s : [0, \infty) \to (0, \infty)$. Both b and s are bounded. These functions are also continuous except, possibly, at a finite number of values. Also, $b(x) = 0$ for all $|x| \geq d$ and $s(x) = t(m)$ for all $x \geq d$ where d is a given positive number. Let $\mathcal{F}(d)$ denote the class of pairs of functions (b, s) that satisfy these restrictions, for given d $(d > 0)$.

Define $\mathcal{D}$ to be the class of all confidence intervals for θ of the form (1), where c, t, d, b and s satisfy the stated restrictions. Each member of this class is specified by (c, t, d, b, s). Apart from the usual $1 - \alpha$ confidence interval I for θ, the class $\mathcal{D}$ of confidence intervals for θ includes the following:

(a) Suppose that we carry out a preliminary hypothesis test of the null hypothesis $\tau = 0$ against the alternative hypothesis $\tau \neq 0$. Also suppose that we construct a confidence interval for θ with nominal coverage $1-\alpha$ based on the assumption that the selected model had been given to us *a priori* (as the true model). The resulting confidence interval, called the naive $1-\alpha$ confidence interval, belongs to the class $\mathcal{D}$ (Kabaila & Giri, 2009, Section 2).

(b) Confidence intervals for θ that are constructed to utilize (in the particular manner described by Kabaila & Giri, 2009) uncertain prior information that $\tau = 0$.

Let K denote the usual $1-\alpha$ confidence interval for θ based on the assumption that $\tau = 0$. The naive $1 - \alpha$ confidence interval, described in (a), may be expressed in the following form:

$$h\left(\frac{|\hat{\tau}|}{\hat{\sigma}\sqrt{v_{22}}} \right) I + \left(1 - h\left(\frac{|\hat{\tau}|}{\hat{\sigma}\sqrt{v_{22}}} \right) \right) K \tag{2}$$

where $h : [0, \infty) \to [0, 1]$ is the unit step function defined by $h(x) = 0$ for all $x \in [0, q]$ and $h(x) = 1$ for all $x > q$. Now suppose that we replace h by a continuous increasing function satisfying $h(0) = 0$ and $h(x) \to 1$ as $x \to \infty$ (a similar construction is extensively used in the context of point estimation by Saleh, 2006). The confidence interval (2) is also a member of the class $\mathcal{D}$.

3. Main result

As noted in Section 2, each member of the class $\mathcal{D}$ is specified by (c, t, d, b, s). The following result states that the usual $1-\alpha$ confidence interval for θ is strongly admissible within the class $\mathcal{D}$.

Theorem 1. *There does not exist* $(c, t, d, b, s) \in \mathcal{D}$ *such that the following three conditions hold:*

$$(a) \qquad E_{\beta,\sigma^2}\big(\textit{length of } J(b,s)\big) \le E_{\beta,\sigma^2}\big(\textit{length of } I\big) \quad \textit{for all } (\beta, \sigma^2). \tag{3}$$

$$(b) \qquad P_{\beta,\sigma^2}\big(\theta \in J(b,s)\big) \ge P_{\beta,\sigma^2}\big(\theta \in I\big) \quad \textit{for all } (\beta, \sigma^2). \tag{4}$$

(c) *Strict inequality holds in either* (3) *or* (4) *for at least one* (β, σ^2).

The proof of this result is presented in Appendix A.

An illustration of this result is provided by Figure 3 of Kabaila & Giri (2009). Define $\gamma = \tau/(\sigma\sqrt{v_{22}})$. Also define

$$e(\gamma; s) = \frac{\text{expected length of } J(b,s)}{\text{expected length of } I}.$$

We call this the scaled expected length of $J(b,s)$. Theorem 1 tells us that for any confidence interval $J(b,s)$, with minimum coverage probability $1-\alpha$, it cannot be the case that $e(\gamma; s) \le 1$ for all γ, with strict inequality for at least one γ. This fact is illustrated by the bottom panel of Figure 3 of Kabaila & Giri (2009).

Define the class $\widetilde{\mathcal{D}}$ to be the subset of $\mathcal{D}$ in which both b and s are continuous functions. Strong admissibility of the confidence interval I within the class $\mathcal{D}$ implies weak admissibility of this confidence interval within the class $\widetilde{\mathcal{D}}$, as the following result shows. Since $(\hat{\beta}, \hat{\sigma}^2)$ is a sufficient statistic for (β, σ), we reduce the data to $(\hat{\beta}, \hat{\sigma}^2)$.

Corollary 1. *There does not exist* $(c, t, d, b, s) \in \widetilde{\mathcal{D}}$ *such that the following three conditions hold:*

$$(a') \qquad \big(\textit{length of } J(b,s)\big) \le \big(\textit{length of } I\big) \quad \textit{for all } (\hat{\beta}, \hat{\sigma}^2). \tag{5}$$

$$(b') \qquad P_{\beta,\sigma^2}\big(\theta \in J(b,s)\big) \ge P_{\beta,\sigma^2}\big(\theta \in I\big) \quad \textit{for all } (\beta, \sigma^2). \tag{6}$$

(c') *Strict inequality holds in either* (5) *or* (13) *for at least one* (β, σ^2).

This corollary is proved in Appendix B.

4. Admissibility result for known error variance

In this section, we suppose that σ^2 is known. Without loss of generality, we assume that $\sigma^2 = 1$. As before, let $\hat{\beta}$ denote the least squares estimator of β. Since $\hat{\beta}$ is a sufficient statistic for β, we reduce the data to $\hat{\beta}$. Assume that the parameter of interest is $\theta = \beta_1 / \sqrt{\text{Var}(\hat{\beta}_1)}$. Thus the least squares estimator of θ is $\hat{\Theta} = \hat{\beta}_1 / \sqrt{\text{Var}(\hat{\beta}_1)}$. Define

$$\hat{\Delta} = \begin{bmatrix} \hat{\beta}_2 - \ell_2 \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p - \ell_p \hat{\beta}_1 \end{bmatrix}$$

where $\ell_2, \dots, \ell_p$ have been chosen such that $\text{Cov}(\hat{\beta}_j - \ell_j \hat{\beta}_1, \hat{\beta}_1) = 0$ for $j = 2, \dots, p$. Now define

$$\delta = \begin{bmatrix} \beta_2 - \ell_2 \beta_1 \\ \vdots \\ \beta_p - \ell_p \beta_1 \end{bmatrix}.$$

Note that $(\hat{\Theta}, \hat{\Delta})$ is obtained by a one-to-one transformation from $\hat{\beta}$. So, we reduce the data to $(\hat{\Theta}, \hat{\Delta})$. Note that $\hat{\Theta}$ and $\hat{\Delta}$ are independent, with $\hat{\Theta} \sim N(\theta, 1)$ and $\hat{\Delta}$ with a multivariate normal distribution with mean δ and known covariance matrix. Define the number z by the requirement that $P(-z \le Z \le z) = 1 - \alpha$ for $Z \sim N(0,1)$. Let $I = \big[\hat{\Theta} - z, \hat{\Theta} + z\big]$. Define

$$\varphi(\hat{\theta}, \theta) = \begin{cases} 1 & \text{if } \theta \in \big[\hat{\theta} - z, \hat{\theta} + z\big] \\ 0 & \text{otherwise} \end{cases}$$

This is the probability that θ is included in the confidence interval I, when $\hat{\theta}$ is the observed value of $\hat{\Theta}$. The length of the confidence interval I is $\int_{-\infty}^{\infty} \varphi(\hat{\theta}, \theta)\, d\theta = 2z$. Let $p_\theta(\cdot)$ denote the probability density function of $\hat{\Theta}$ for given θ. The coverage probability of I is $\int_{-\infty}^{\infty} \varphi(\hat{\theta}, \theta)\, p_\theta(\hat{\theta})\, d\hat{\theta} = 1 - \alpha$.

Now let $\mathcal{C}(\hat{\Theta}, \hat{\Delta})$ denote a confidence set for θ. Define

$$\varphi_\delta(\hat{\theta}, \theta) = P_{\theta,\delta}\big(\theta \in \mathcal{C}(\hat{\theta}, \hat{\Delta})\big),$$

where $\hat{\theta}$ denotes the observed value of $\hat{\Theta}$. For each given $\delta \in \mathbb{R}^{p-1}$, the expected Lebesgue measure of $\mathcal{C}(\hat{\Theta}, \hat{\Delta})$ is $E_{\theta,\delta}\Big(\int_{-\infty}^{\infty} \varphi_\delta(\hat{\Theta}, \theta)\, d\theta\Big)$. For each given $\delta \in \mathbb{R}^{p-1}$, the coverage probability of $\mathcal{C}(\hat{\Theta}, \hat{\Delta})$ is $\int_{-\infty}^{\infty} \varphi_\delta(\hat{\theta}, \theta)\, p_\theta(\hat{\theta})\, d\hat{\theta}$. Theorem 5.1 of Joshi (1969) implies the following strong admissibility result. Suppose that $\varphi_\delta(\hat{\theta}, \theta)$ satisfies the following conditions

(i) $E_{\theta,\delta}\Big(\int_{-\infty}^{\infty} \varphi_\delta(\hat\theta, \theta)\, d\theta \Big) \le E_{\theta,\delta}\Big(\int_{-\infty}^{\infty} \varphi(\hat\theta, \theta)\, d\theta \Big)$ for all $\theta \in \mathbb{R}$.

(ii) $\int_{-\infty}^{\infty} \varphi_\delta(\hat\theta, \theta)\, p_\theta(\hat\theta)\, d\hat\theta \ge \int_{-\infty}^{\infty} \varphi(\hat\theta, \theta)\, p_\theta(\hat\theta)\, d\hat\theta$ for all $\theta \in \mathbb{R}$.

Then $\varphi_\delta(\hat\theta, \theta) = \varphi(\hat\theta, \theta)$ for almost all $(\hat\theta, \theta) \in \mathbb{R}^2$. This result is true for each $\delta \in \mathbb{R}^{p-1}$. Using standard arguemnts, this entails that $I \setminus \mathcal{C}(\hat\Theta, \hat\Delta)$ and $\mathcal{C}(\hat\Theta, \hat\Delta) \setminus I$ are Lebesgue-null sets, for (Lebesgue-) almost all values of $(\hat\Theta, \hat\Delta)$.

Appendix A: Proof of Theorem 1

Suppose that c is a given vector (such that c and a are linearly independent), t is a given number and d is a given positive number. The proof of Theorem 1 now proceeds as follows. We present a few definitions and a lemma. We then apply this lemma to prove this theorem.

Define $W = \hat\sigma/\sigma$. Note that W has the same distribution as $\sqrt{Q/m}$ where $Q \sim \chi^2_m$. Let f_W denote the probability density function of W. Also let ϕ denote the $N(0,1)$ probability density function. Now define

$$R_1(b, s; \gamma) = \frac{\text{expected length of } J(b,s)}{\text{expected length of } I} - 1.$$

It follows from (7) of Kabaila & Giri (2009) that

$$R_1(b, s; \gamma) = \frac{1}{t(m)\, E(W)} \int_0^\infty \int_{-d}^{d} \big(s(|x|) - t(m)\big)\, \phi(wx - \gamma)\, dx\, w^2\, f_W(w)\, dw. \qquad (7)$$

Thus, for each $(b, s) \in \mathcal{F}(d)$, $R_1(b, s; \gamma)$ is a continuous function of γ.

Also define $R_2(b, s; \gamma) = P\big(\theta \notin J(b, s)\big) - \alpha$. We make the following definitions, also used by Kabaila & Giri (2009). Define $\rho = v_{12}/\sqrt{v_{11} v_{22}}$ and $\Psi(x, y; \mu, v) = P(x \le Z \le y)$, for $Z \sim N(\mu, v)$. Now define the functions

$$k^\dagger(h, w, \gamma, \rho) = \Psi\big(-t(m)w, t(m)w; \rho(h - \gamma), 1 - \rho^2\big)$$
$$k(h, w, \gamma, \rho) = \Psi\big(b(h/w)w - s(|h|/w)w, b(h/w)w + s(|h|/w)w; \rho(h - \gamma), 1 - \rho^2\big).$$

It follows from (6) of Kabaila & Giri (2009), that

$$R_2(b, s; \gamma) = -\int_0^\infty \int_{-d}^{d} \big(k(wx, w, \gamma, \rho) - k^\dagger(wx, w, \gamma, \rho)\big)\, \phi(wx - \gamma)\, dx\, w\, f_W(w)\, dw. \qquad (8)$$

Thus, for each $(b, s) \in \mathcal{F}(d)$, $R_2(b, s; \gamma)$ is a continuous function of γ.

Now $E(W^2) = 1$ and so

$$\int_0^\infty w^2 \, f_W(w) \, dw = 1.$$

It follows from (7) that

$$\int_{-\infty}^{\infty} R_1(b, s; \gamma) \, d\gamma = \frac{2}{t(m) \, E(W)} \int_0^d \big(s(x) - t(m) \big) \, dx. \tag{9}$$

Thus $\int_{-\infty}^{\infty} R_1(b, s; \gamma) \, d\gamma$ exists for all $(b, s) \in \mathcal{F}(d)$.

Since $k(wx, w, \gamma, \rho)$ and $k^\dagger(wx, w, \gamma, \rho)$ are probabilities,

$$|R_2(b, s; \gamma)| \le \int_0^\infty \int_{-d}^d \phi(wx - \gamma) dx \, w f_W(w) \, dw,$$

so that

$$\int_{-\infty}^{\infty} |R_2(b, s; \gamma)| \, d\gamma \le 2d \int_0^\infty w f_W(w) \, dw = 2d E(W) < \infty.$$

Thus $\int_{-\infty}^{\infty} R_2(b, s; \gamma) \, d\gamma$ exists for all $(b, s) \in \mathcal{F}(d)$.

Thus, we may define

$$g(b, s; \lambda) = \lambda \int_{-\infty}^{\infty} R_1(b, s; \gamma) \, d\gamma + (1 - \lambda) \int_{-\infty}^{\infty} R_2(b, s; \gamma) \, d\gamma,$$

for each $(b, s) \in \mathcal{F}(d)$, where $0 < \lambda < 1$. Kempthorne (1983, 1987, 1988) presents results on what he calls compromise decision theory. Initially, these results were applied only to the solution of some problems of point estimation. Kabaila & Tuck (2008) develop new results in compromise decision theory and apply these to a problem of interval estimation. The following lemma, which will be used in the proof of Theorem 1, is in the style of these compromise decision theory results.

Lemma 1. *Suppose that c is a given vector (such that c and a are linearly independent), t is a given number and d is a given positive number. Also suppose that λ is given and that (b^*, s^*) minimizes $g(b, s; \lambda)$ with respect to $(b, s) \in \mathcal{F}(d)$. Then there does not exist $(b, s) \in \mathcal{F}(d)$ such that*

(a) $R_1(b, s; \gamma) \le R_1(b^*, s^*; \gamma)$ for all γ.

(b) $R_2(b, s; \gamma) \le R_2(b^*, s^*; \gamma)$ for all γ.

(c) Strict inequality holds in either (a) or (b) for at least one γ.

Proof. Suppose that c is a given vector (such that c and a are linearly independent), t is a given number and d is a given positive number. The proof is by contradiction. Suppose that there exist $(b, s) \in \mathcal{F}(d)$ such that (a), (b) and (c) hold. Now,

$$g(b^*, s^*; \lambda) - g(b, s; \lambda) = \lambda \int_{-\infty}^{\infty} \big(R_1(b^*, s^*; \gamma) - R_1(b, s; \gamma)\big)\, d\gamma + (1-\lambda) \int_{-\infty}^{\infty} \big(R_2(b^*, s^*; \gamma) - R_2(b, s; \gamma)\big)\, d\gamma$$

By hypothesis, one of the following 2 cases holds.

<u>Case 1</u> (a) and (b) hold and $R_1(b^*, s^*; \gamma) - R_1(b, s; \gamma) > 0$ for at least one γ. Since $R_1(b^*, s^*; \gamma) - R_1(b, s; \gamma)$ is a continuous function of γ,

$$\int_{-\infty}^{\infty} \big(R_1(b^*, s^*; \gamma) - R_1(b, s; \gamma)\big)\, d\gamma > 0.$$

Thus $g(b^*, s^*; \lambda) > g(b, s; \lambda)$ and we have established a contradiction.

<u>Case 2</u> (a) and (b) hold and $R_2(b^*, s^*; \gamma) - R_2(b, s; \gamma) > 0$ for at least one γ. Since $R_2(b^*, s^*; \gamma) - R_2(b, s; \gamma)$ is a continuous function of γ,

$$\int_{-\infty}^{\infty} \big(R_2(b^*, s^*; \gamma) - R_2(b, s; \gamma)\big)\, d\gamma > 0.$$

Thus $g(b^*, s^*; \lambda) > g(b, s; \lambda)$ and we have established a contradiction.

Lemma 1 follows from the fact that this argument holds for every given vector c (such that c and a are linearly independent), every given number t and every given positive number d.

□

We will first find the (b^*, s^*) that minimizes $g(b, s; \lambda)$ with respect to $(b, s) \in \mathcal{F}(d)$, for given λ. We will then choose λ such that $J(b^*, s^*) = I$, the usual $1 - \alpha$ confidence interval for θ. Theorem 1 is then a consequence of Lemma 1.

By changing the variable of integration in the inner integral in (8), it can be shown that $R_2(b, s; \gamma)$ is equal to

$$-\int_0^{\infty} \int_0^d \Big(\big(k(wx, w, \gamma, \rho) - k^{\dagger}(wx, w, \gamma, \rho)\big)\, \phi(wx - \gamma) + \big(k(-wx, w, \gamma, \rho) - k^{\dagger}(-wx, w, \gamma, \rho)\big)\, \phi(wx + \gamma) \Big)\, dx\, w\, f_W(w)\, dw$$

Using this expression and the restriction that b is an odd function, we find that $\int_{-\infty}^{\infty} R_2(b, s; \gamma)\, d\gamma$ is equal to

$$
\begin{aligned}
-\int_0^d \int_0^\infty \int_{-\infty}^{\infty} \Big(& \Psi\big(b(x)w - s(x)w, b(x)w + s(x)w; \rho y, 1-\rho^2\big) \\
& - \Psi\big(-t(m)w, t(m)w; \rho y, 1-\rho^2\big) \\
& + \Psi\big(-b(x)w - s(x)w, -b(x)w + s(x)w; -\rho y, 1-\rho^2\big) \\
& - \Psi\big(-t(m)w, t(m)w; -\rho y, 1-\rho^2\big)\Big)\, \phi(y)\, dy\, w\, f_W(w)\, dw\, dx.
\end{aligned}
$$

Hence, to within an additive constant that does not depend on (b, s), $\int_{-\infty}^{\infty} R_2(b, s; \gamma)\, d\gamma$ is equal to

$$
\begin{aligned}
-\int_0^d \int_0^\infty \int_{-\infty}^{\infty} & \Big(\Psi\big(b(x)w - s(x)w, b(x)w + s(x)w; \rho y, 1-\rho^2\big) \\
& + \Psi\big(-b(x)w - s(x)w, -b(x)w + s(x)w; -\rho y, 1-\rho^2\big)\Big)\, \phi(y)\, dy\, w\, f_W(w)\, dw\, dx.
\end{aligned}
$$

Thus, to within an additive constant that does not depend on (b, s),

$$
g(b, s; \lambda) = \int_0^d q(b, s; x)\, dx,
$$

where $q(b, s; x)$ is equal to

$$
\begin{aligned}
& \frac{2\lambda}{t(m)\, E(W)} s(x) \\
& - (1-\lambda) \int_0^\infty \int_{-\infty}^{\infty} \big(\Psi(b(x)w - s(x)w, b(x)w + s(x)w; \rho y, 1-\rho^2) \\
& \qquad + \Psi(-b(x)w - s(x)w, -b(x)w + s(x)w; -\rho y, 1-\rho^2)\big)\, \phi(y)\, dy\, w\, f_W(w)\, dw.
\end{aligned}
$$

Note that x enters into the expression for $q(b, s; x)$ only through $b(x)$ and $s(x)$. To minimize $g(b, s; \lambda)$ with respect to $(b, s) \in \mathcal{F}(d)$, it is therefore sufficient to minimize $q(b, s; x)$ with respect to $(b(x), s(x))$ for each $x \in [0, d]$. The situation here is similar to the computation of Bayes rules, see e.g. Casella & Berger (2002, pp. 352–353). Therefore, to minimize $g(b, s; \lambda)$ with respect to $(b, s) \in \mathcal{F}(d)$, we simply minimize

$$
\begin{aligned}
\tilde{q}(b, s) = & \frac{2\lambda}{t(m)\, E(W)} s \\
& - (1-\lambda) \int_0^\infty \int_{-\infty}^{\infty} \big(\Psi(bw - sw, bw + sw; \rho y, 1-\rho^2) \\
& \qquad + \Psi(-bw - sw, -bw + sw; -\rho y, 1-\rho^2)\big)\, \phi(y)\, dy\, w\, f_W(w)\, dw
\end{aligned}
$$

with respect to $(b, s) \in \mathbb{R} \times (0, \infty)$, to obtain (b', s') and then set $b(x) = b'$ and $s(x) = s'$ for all $x \in [0, d]$.

Let the random variables A and B have the following distribution

$$\begin{bmatrix} A \\ B \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

Note that the distribution of A, conditional on $B = y$, is $N(\rho y, 1 - \rho^2)$. Thus

$$\Psi(bw - sw, bw + sw; \rho y, 1 - \rho^2) = P\big(bw - sw \le A \le bw + sw \,\big|\, B = y\big)$$

Hence

$$\int_0^\infty \int_{-\infty}^\infty \Psi(bw - sw, bw + sw; \rho y, 1 - \rho^2)\, \phi(y)\, dy\, w\, f_W(w)\, dw$$
$$= \int_0^\infty P(bw - sw \le A \le bw + sw)\, w\, f_W(w)\, dw. \tag{10}$$

Let Φ denote the $N(0, 1)$ cumulative distribution function. For every fixed $w > 0$ and $s > 0$,

$$P(bw - sw \le A \le bw + sw) = \Phi(bw + sw) - \Phi(bw - sw)$$

is maximized by setting $b = 0$. Thus, for each fixed $s > 0$, (10) is maximized with respect to $b \in \mathbb{R}$ by setting $b = 0$.

Now let the random variables $\tilde{A}$ and $\tilde{B}$ have the following distribution

$$\begin{bmatrix} \tilde{A} \\ \tilde{B} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \right).$$

Note that the distribution of $\tilde{A}$, conditional on $\tilde{B} = y$, is $N(-\rho y, 1 - \rho^2)$. Thus

$$\Psi(-bw - sw, -bw + sw; -\rho y, 1 - \rho^2) = P\big(-bw - sw \le \tilde{A} \le -bw + sw \,\big|\, \tilde{B} = y\big)$$

Hence

$$\int_0^\infty \int_{-\infty}^\infty \Psi(-bw - sw, -bw + sw; -\rho y, 1 - \rho^2)\, \phi(y)\, dy\, w\, f_W(w)\, dw$$
$$= \int_0^\infty P(-bw - sw \le \tilde{A} \le -bw + sw)\, w\, f_W(w)\, dw. \tag{11}$$

For every fixed $w > 0$ and $s > 0$,

$$P\big(-bw - sw \le \tilde{A} \le -bw + sw\big) = \Phi(-bw + sw) - \Phi(-bw - sw)$$

is maximized by setting $b = 0$. Thus, for each fixed $s > 0$, (11) is maximized with respect to $b \in \mathbb{R}$ by setting $b = 0$.

Therefore, $\tilde{q}(b, s)$ is, for each fixed $s > 0$, minimized with respect to b by setting $b = 0$. Thus $b' = 0$ and so $b^*(x) = 0$ for all $x \in \mathbb{R}$. Hence, to find s' we need to minimize

$$\frac{\lambda}{t(m)E(W)} s - (1-\lambda) \int_0^\infty \big(2\Phi(sw) - 1\big)\, w f_W(w)\, dw$$

with respect to $s > 0$. Therefore, to find s' we may minimize

$$r(s) = \ell(\lambda)\, s - 2\int_0^\infty \Phi(sw)\, w f_W(w)\, dw$$

with respect to $s > 0$, where

$$\ell(\lambda) = \frac{\lambda}{(1-\lambda)t(m)E(W)}.$$

Note that $\ell(\lambda)$ is an increasing function of λ, such that $\ell(\lambda) \downarrow 0$ as $\lambda \downarrow 0$ and $\ell(\lambda) \uparrow \infty$ as $\lambda \uparrow 1$. Choose $\lambda = \lambda^*$, where

$$\ell(\lambda^*) = 2\int_0^\infty \phi\big(t(m)w\big)\, w^2\, f_W(w)\, dw.$$

Note that $0 < \ell(\lambda^*) < \sqrt{2/\pi}$. Now

$$\frac{dr(s)}{ds} = \ell(\lambda^*) - 2\int_0^\infty \phi(sw)\, w^2 f_W(w)\, dw.$$

Since $\int_0^\infty \phi(sw)\, w^2 f_W(w)\, dw$ is a decreasing function of $s > 0$, $dr(s)/ds$ is an increasing function of $s > 0$. Also, for $s = 0$, $\int_0^\infty \phi(sw)\, w^2 f_W(w)\, dw = 1/\sqrt{2\pi}$. Thus, to minimize $r(s)$ with respect to $s > 0$, we need to solve

$$\ell(\lambda^*) - 2\int_0^\infty \phi(sw)\, w^2\, f_W(w)\, dw = 0$$

for $s > 0$. Obviously, this solution in $s = t(m)$. Thus $s^*(x) = t(m)$ for all $x \geq 0$. In other words, $J(b^*, s^*) = I$. By Lemma 1, there does not exist $(b, s) \in \mathcal{F}(d)$ such that

$$(a) \qquad E_{\beta,\sigma^2}\big(\text{length of } J(b,s)\big) \leq E_{\beta,\sigma^2}\big(\text{length of } I\big) \quad \text{for all } (\beta, \sigma^2). \tag{12}$$

$$(b) \qquad P_{\beta,\sigma^2}\big(\theta \in J(b,s)\big) \geq P_{\beta,\sigma^2}\big(\theta \in I\big) \quad \text{for all } (\beta, \sigma^2). \tag{13}$$

(c) Strict inequality holds in either (12) or (13) for at least one (β, σ^2).

Theorem 1 follows from the fact that this argument holds for every given vector c (such that c and a are linearly independent), every given number t and every given positive number d.

Appendix B: Proof of Corollary 1

The proof of Corollary 1 is by contradiction. Suppose that c is a given vector (such that c and a are linearly independent), t is a given number and d is a given positive number. Also suppose that there exists $(b, s) \in \mathcal{F}(d)$ such that both b and s are continuous and (a'), (b') and (c'), in the statement of Corollary 1, hold. Now (a') implies that

$$E_{\beta,\sigma^2}\big(\text{length of } J(b,s)\big) \le E_{\beta,\sigma^2}\big(\text{length of } I\big) \quad \text{for all } (\beta, \sigma^2),$$

so that (a) holds. By hypothesis, one of the following two cases holds.

<u>Case 1</u> $\big(\text{length of } J(b,s)\big) < \big(\text{length of } I\big)$ for at least one $(\hat{\beta}, \hat{\sigma}^2)$. Now

$$\big(\text{length of } J(b,s)\big) = 2\sqrt{v_{11}}\hat{\sigma}\, s\left(\frac{|\hat{\tau}|}{\hat{\sigma}\sqrt{v_{22}}}\right),$$

which is a continuous function of $(\hat{\beta}, \hat{\sigma}^2)$. Hence $\big(\text{length of } I\big) - \big(\text{length of } J(b,s)\big)$ is a continuous function of $(\hat{\beta}, \hat{\sigma}^2)$. Thus

$$E_{\beta,\sigma^2}\big(\text{length of } J(b,s)\big) < E_{\beta,\sigma^2}\big(\text{length of } I\big) \quad \text{for at least one } (\beta, \sigma^2).$$

Thus there exists $(b, s) \in \mathcal{F}(d)$ such that (a), (b) and (c), in the statement of Theorem 1, hold. We have established a contradiction.

<u>Case 2</u> There is strict inequality in (b') for at least one (β, σ^2). Thus there exists $(b, s) \in \mathcal{F}(d)$ such that (a), (b) and (c), in the statement of Theorem 1, hold. We have established a contradiction.

Corollary 1 follows from the fact that this argument holds for every given vector c (such that c and a are linearly independent), every given number t and every given positive number d.

References

CASELLA, G. & BERGER, R.L. (2002). *Statistical Inference*, 2nd ed.. Pacific Grove, CA: Duxbury.

JOSHI, V.M. (1969). Admissibility of the usual confidence sets for the mean of a univariate or bivariate normal population. *Annals of Mathematical Statistics*, **40**, 1042–1067.

JOSHI, V.M. (1982). Admissibility. On pp.25–29 of Vol. 1 of *Encyclopedia of Statistical Sciences*, editors-in-chief, Samuel Kotz, Norman L. Johnson ; associate editor, Campbell B. Read. New York: John Wiley.

KABAILA, P. & GIRI, K. (2009). Confidence intervals in regression utilizing prior information. *Journal of Statistical Planning and Inference*, **139**, 3419–3429.

KABAILA, P. & TUCK, J. (2008). Confidence intervals utilizing prior information in the Behrens-Fisher problem. *Australian & New Zealand Journal of Statistics* **50**, 309–328.

KEMPTHORNE, P.J. (1983). Minimax-Bayes compromise estimators. In *1983 Business and Economic Statistics Proceedings of the American Statistical Association*, Washington DC, pp.568–573.

KEMPTHORNE, P.J. (1987). Numerical specification of discrete least favourable prior distributions. SIAM *Journal on Scientific and Statistical Computing* **8**, 171–184.

KEMPTHORNE, P.J. (1988). Controlling risks under different loss functions: the compromise decision problem. *Ann. Statist.* **16**, 1594–1608.

SALEH, A.K.MD.E. (2006) Theory of Preliminary Test and Stein-Type Estimation with Applications. Hoboken, NJ: John Wiley.