Econometric Theory, **14**, 1998, 463–482. Printed in the United States of America.

VALID CONFIDENCE INTERVALS IN REGRESSION AFTER VARIABLE SELECTION

PAUL KABAILA
La Trobe University

We consider a linear regression model with regression parameters $(\theta_1,\ldots,\theta_p)$ and error variance parameter σ^2. Our aim is to find a confidence interval with minimum coverage probability $1-\alpha$ for a parameter of interest θ_1 in the presence of nuisance parameters $(\theta_2,\ldots,\theta_p,\sigma^2)$. We consider two confidence intervals, the first of which is the standard confidence interval for θ_1 with coverage probability $1-\alpha$. The second confidence interval for θ_1 is obtained after a variable selection procedure has been applied to θ_p. This interval is chosen to be as short as possible subject to the constraint that it has minimum coverage probability $1-\alpha$. The confidence intervals are compared using a risk function that is defined as a scaled version of the expected length of the confidence interval. We show that, subject to certain conditions including that [(dimension of response vector) $-\ p$] is small, the second confidence interval is preferable to the first when we anticipate (without being certain) that $|\theta_p|/\sigma$ is small. This comparison of confidence intervals is shown to be mathematically equivalent to a corresponding comparison of prediction intervals.

1. INTRODUCTION

We consider a linear regression model $Y = X\theta + \epsilon$ where Y is a random n-vector of responses, X is a known $n \times p$ matrix, $\theta = [\theta_1,\ldots,\theta_p]^T$ is an unknown parameter vector, and $\epsilon \sim N(0,\sigma^2 I_n)$ where σ^2 is an unknown parameter satisfying $\sigma^2 > 0$. We suppose that the columns of X are linearly independent. We also suppose that θ_1 is the parameter of interest.

The least-squares estimator of θ is $\hat{\Theta} = (X^TX)^{-1}X^TY = [\hat{\Theta}_1,\ldots,\hat{\Theta}_p]^T$. The usual estimator of σ^2 is $S^2 = (Y-X\hat{\Theta})^T(Y-Y\hat{\Theta})/(n-p)$. If $\theta_p = 0$ then the model reduces to $Y = C\psi + \epsilon$ where C is the $n \times (p-1)$ matrix consisting of the first $p-1$ columns of X and $\psi = [\theta_1,\ldots,\theta_{p-1}]^T$. In this case the least-squares estimator of ψ is $\hat{\Psi} = (C^TC)^{-1}C^TY = [\hat{\Psi}_1,\ldots,\hat{\Psi}_{p-1}]^T$ and the estimator of σ^2 *is* $\tilde{S}^2 = (Y - C\hat{\Psi})^T(Y - C\hat{\Psi})/(n-p+1)$.

We will be concerned with inference for θ_1 after variable selection for the variable associated with θ_p. For convenience we henceforth use the phrase "variable selection for θ_p" as shorthand for the phrase "variable selection for the variable associated with θ_p". Let $v_X(p)$ denote the (p,p)th element of $(X^TX)^{-1}$.

I thank the referees for their helpful suggestions and comments. Address correspondence to: Paul Kabaila, Department of Statistical Science, La Trobe University, Bundoora, Victoria, 3083, Australia.

© 1998 Cambridge University Press 0266-4666/98 $9.50

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

By inference for θ_1 after variable selection for θ_p we mean the following. If $|\hat{\Theta}_p|/S\sqrt{v_X(p)} \geq q$ then inference for θ_1 proceeds on the basis of the model $Y = X\theta + \epsilon$, ignoring the fact that the model has been selected by a data-driven procedure. On the other hand, if $|\hat{\Theta}_p|/S\sqrt{v_X(p)} < q$ then inference proceeds differently. Hitherto, if $|\hat{\Theta}_p|/S\sqrt{v_X(p)} < q$ then inference for θ_1 has proceeded on the basis of the model $Y = C\psi + \epsilon$, ignoring the fact that the model has been selected by a data-driven procedure. Later, when the inference of interest is a confidence interval with minimum coverage probability $1 - \alpha$, we introduce a different procedure when $|\hat{\Theta}_p|/S\sqrt{v_X(p)} < q$. Here q is a positive constant chosen by the data analyst.

We now describe a possible motivation for inference for θ_1 after variable selection for θ_p. Suppose that the model $Y = C\psi + \epsilon$ (i.e., the model $Y = X\theta + \epsilon$ with $\theta_p = 0$) is thought to be reasonable a priori. This reasonableness may derive from theory and/or past experience. Suppose, however, that there is some element of doubt concerning the assumption that $\theta_p = 0$. It might be hoped that the use of inference for θ_1 after variable selection for θ_p would lose us little if θ_p differed negligibly from zero, while providing some protection for the analysis if, as a result of unforeseen circumstances, θ_p happened to differ substantially from zero.

The literature on the properties of estimators obtained after the preliminary application of a model selection step is a well-developed field of study (see, e.g., Judge & Bock, 1978; Han, Rao, and Ravichandran, 1988). Consider the case that the inference for θ_1 after variable selection for θ_p is estimation of θ_1. Suppose also that if $|\hat{\Theta}_p|/S\sqrt{v_X(p)} < q$ then estimation of θ_1 proceeds on the basis of the model $y = C\psi + \epsilon$, ignoring the fact that the model has been selected by a data-driven procedure. The resulting estimator is $\hat{\tilde{\Theta}}_1$ defined as follows. If $|\hat{\Theta}_p|/S\sqrt{v_X(p)} \geq q$ then $\hat{\tilde{\Theta}}_1 = \hat{\Theta}_1$; otherwise $\hat{\tilde{\Theta}}_1 = \hat{\Psi}_1$. Suppose that q is fixed and that all the parameters of the model $Y = X\theta + \epsilon$ are fixed, apart from θ_p. The ratio of mean-square errors of estimation $E(\hat{\tilde{\Theta}}_1 - \theta_1)^2/E(\hat{\Theta}_1 - \theta_1)^2$ is less than 1 at $\theta_p = 0$. As θ_p increases, this ratio increases up to a maximum that exceeds 1. The ratio approaches 1 as $\theta_p \to \infty$.

Suppose now that the inference for θ_1 after variable selection for θ_p is a confidence interval for θ_1 with minimum coverage probability $1 - \alpha$. For the model $Y = X\theta + \epsilon$ the standard confidence interval for θ_1 with coverage probability $1 - \alpha$ is

$$\left[\hat{\Theta}_1 - c_{n-p}S\sqrt{v_X(1)},\ \hat{\Theta}_1 + c_{n-p}S\sqrt{v_X(1)}\right], \tag{1}$$

where $v_X(1)$ is the (1,1)th element of $(X^TX)^{-1}$ and c_m satisfies $P(|T| \leq c_m) = 1 - \alpha$ for $T \sim t_m$ where t_m denotes the t-distribution with m degrees of freedom. On the other hand, if $\theta_p = 0$ then the usual confidence interval for θ_1 with coverage probability $1 - \alpha$ is

$$\left[\hat{\Psi}_1 - c_{n-p+1}\tilde{S}\sqrt{v_C(1)},\ \hat{\Psi}_1 + c_{n-p+1}\tilde{S}\sqrt{v_C(1)}\right], \tag{2}$$

where $v_C(1)$ is the (1,1)th element of $(C^TC)^{-1}$. A naive proposal for finding a confidence interval for θ_1 with minimum coverage probability $1 - \alpha$ after vari-

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

able selection for θ_p is as follows. If $|\hat{\Theta}_p|/S\sqrt{v_X(p)} \geq q$ then use the interval (1); otherwise the interval (2). This naive proposal fails because it leads to a confidence interval with minimum coverage probability $<1-\alpha$. However, the naive proposal can be modified to yield a confidence interval that does have minimum coverage probability $1-\alpha$. We call this confidence interval the modified interval. The modified interval is obtained as follows. If $|\hat{\Theta}_p|/S\sqrt{v_X(p)} \geq q$ then use the interval (1); otherwise use the following modification of the interval (2):

$$\left[\hat{\Psi}_1 - c\tilde{S}\sqrt{v_C(1)},\ \hat{\Psi}_1 + c\tilde{S}\sqrt{v_C(1)}\right], \tag{3}$$

where c has been chosen as small as possible subject to the modified interval having minimum coverage probability $1-\alpha$.

The modified interval has the following attractive property. Suppose that q is fixed and that all the parameters of the model $Y = X\theta + \epsilon$ are fixed, apart from θ_p. The endpoints of the modified interval approach in probability the endpoints of the standard confidence interval (1) as $\theta_p \to \infty$. In terms of the possible motivation for inference for θ_1 after the usual form of variable selection for θ_p described in paragraph 4 of this section, this property tells us that if, as a result of unforeseen circumstances, θ_p happened to differ very greatly from zero then there is little lost in using the modified interval as opposed to the standard interval. Of course, a full comparison of these intervals requires a consideration of all possible values of θ_p, not just those values that differ very greatly from zero.

In Section 2 we describe a generalization of the modified interval. The purpose of this paper is to compare this interval with the standard interval (1) using a decision-theoretic approach based on the risk function

$$\frac{E\{\text{length of confidence interval}\}}{2\sigma\sqrt{v_X(1)}}. \tag{4}$$

This comparison is greatly eased by the use of the transformed model that is described in Section 2. Our comparison of the generalization of the modified interval with the standard interval (1) is described in Sections 3 and 4. As pointed out in Section 5, this comparison is mathematically equivalent to a corresponding comparison of prediction intervals.

In this paper we adopt the point of view that in practice the amount of prior information concerning $(\theta_1, \ldots, \theta_p, \sigma^2)$ usually lies between two extremes. These extremes are (a) enough prior information to specify a prior distribution for $(\theta_1, \ldots, \theta_p, \sigma^2)$ and (b) no prior information whatsoever concerning $(\theta_1, \ldots, \theta_p, \sigma^2)$. In case (a) a Bayes approach would be appropriate, and in case (b) a minimax criterion would be appropriate. This point of view is very similar to that advocated by Hodges and Lehmann (1952, p. 396). In Sections 3 and 4 it is shown that, subject to certain conditions including that $n-p$ is small, the generalization of the modified interval is preferable to the interval (1) when we anticipate (without being certain) that $|b_{pp}\theta_p|/\sigma$ is small. Here, b_{pp} is a certain function (defined in Section 2) of the known matrix X. In Sections 3 and 4 it is also shown that in the

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

absence of any prior information concerning $(\theta_1,\ldots,\theta_p,\sigma^2)$ that is, when the minimax criterion is appropriate, the generalization of the modified interval performs worse than the standard interval (1).

The study of valid confidence intervals obtained after the preliminary application of a model selection step is a much less well-developed field of study than the study of estimators obtained after the application of such a step. By valid confidence interval we mean a confidence interval with the desired minimum coverage probability. Bickel (1984, Sect. 3) has considered confidence intervals for regression parameters after the preliminary application of a variable selection step. This work differs from the work we present in an number of significant respects that are described in the last part of Section 2.

2. MODEL TRANSFORMATION AND GENERALIZATION OF THE MODIFIED INTERVAL

We suppose that the end result of our statistical analysis is a confidence interval for θ_1 that is required to satisfy the following constraint.

CP Constraint. The infimum over all values of (θ,σ^2) of the coverage probability of the confidence interval $\geq 1-\alpha$.

We are concerned with unconditional properties of the confidence intervals. For a conditional approach see Pötscher (1995).

Before we describe our generalization of the modified interval for θ_1 we transform the model, in the manner described in Appendix A, to a much more convenient form that involves in a simple manner both of the parameters θ_1 and θ_p. The transformed model is

$$[Z_1,\ldots,Z_p]^T = B\theta + [\eta_1,\ldots,\eta_p]^T,$$
$$Z_j = \eta_j \quad \text{for } j = p+1,\ldots,n, \tag{5}$$

where $[Z_1,\ldots,Z_n]^T$ is the vector of responses, B is an upper triangular $p \times p$ matrix (with nonzero diagonal elements), and $\eta = [\eta_1,\ldots,\eta_n]^T$ has independent and identically $N(0,\sigma^2)$ distributed components. Note that $\hat{\Theta} = H[Z_1,\ldots,Z_p]^T$ where $H = B^{-1}$, $S^2 = \sum_{i=p+1}^{n} Z_i^2/(n-p)$, and $\tilde{S}^2 = \sum_{i=p}^{n} Z_i^2/(n-p+1)$. The standard confidence interval is identical to

$$\left[\hat{\Theta}_1 - c_{n-p} S \sqrt{\sum_{j=1}^{p} h_{1j}^2},\ \hat{\Theta}_1 + c_{n-p} S \sqrt{\sum_{j=1}^{p} h_{1j}^2}\right], \tag{6}$$

where $[h_{11},\ldots,h_{1p}]$ is the first row of the matrix H.

Define the estimator $\tilde{\Theta}_1$ of θ_1 by

$$\tilde{\Theta}_1 = \sum_{j=1}^{p-1} h_{1j} Z_j + h_{1p} \times S \times m\left(\frac{Z_p}{S}\right), \tag{7}$$

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

where $m(x) = x$ for $|x| \geq q$ and $m(\cdot)$ is an odd function that is nondecreasing in $x \in [0,q)$ and that satisfies $m(x) < x$ for all $x \in (0,q)$. We also require the restrictions of $m(\cdot)$ to the interval $(-q,q)$ to be a differentiable function with bounded derivative. In Appendix B it is proved that $\tilde{\Theta}_1 = \hat{\tilde{\Theta}}_1$ in the particular case that $m(x) = 0$ for all $x \in (-q,q)$.

Define $d_1 = h_{1p}/(\sum_{j=1}^{p-1} h_{1j}^2)^{1/2}$ and $d_2 = \sqrt{1 + d_1^2}$. Also define $K(\cdot)$ by

$$K(c) = \begin{cases} c_{n-p} S & \text{for } |Z_p|/S \geq k \\ c d_2^{-1} \tilde{S} & \text{otherwise,} \end{cases}$$

where k is a positive constant. Also define c_n^* to be the smallest value of c such that a confidence interval with center $\tilde{\Theta}_1$ and length $2K(c)(\sum_{j=1}^{p} h_{1j}^2)^{1/2}$ satisfies the CP constraint. The existence of c_n^* is proved in Appendix G. The generalization of the modified confidence interval for θ_1 is

$$\left[\tilde{\Theta}_1 - K(c_n^*) \sqrt{\sum_{j=1}^{p} h_{1j}^2}, \tilde{\Theta}_1 + K(c_n^*) \sqrt{\sum_{j=1}^{p} h_{1j}^2} \right]. \tag{8}$$

We call this the alternative confidence interval. As shown in Appendix B, the condition $|\hat{\Theta}_p|/S\sqrt{v_X(p)} \geq q$ is equivalent to $|Z_p|/S \geq q$. Thus, for $k = q$ and $m(x) = 0$ for all $x \in (-q,q)$, the alternative confidence interval reduces to the modified confidence interval described in the introduction.

Define $\bar{\theta}_p = b_{pp}\theta_p/\sigma$. Note that $\hat{\Theta}_p \sim N(\theta_p, b_{pp}^{-2}\sigma^2)$, so that $\bar{\theta}_p$ is equal to $E(\hat{\Theta}_p)$ divided by the standard deviation of $\hat{\Theta}_p$. Suppose that α, $m(\cdot)$, k, and n are fixed. Also suppose that all parameters of the model are fixed, apart from $\bar{\theta}_p$. As shown in Appendix D, $\hat{\Theta}_1 - \tilde{\Theta}_1 \xrightarrow{p} 0$ and $K(c_n^*) - c_{n-p} S \xrightarrow{p} 0$ as $|\bar{\theta}_p| \to \infty$. In other words, the alternative confidence interval has the attractive property that its endpoints approach in probability the endpoints of the standard confidence interval as $|\bar{\theta}_p| \to \infty$.

The aim of the paper is to compare the alternative confidence interval with the standard confidence interval in terms of the risk function (4), which is identical to

$$\frac{E\{\text{length of confidence interval}\}}{2\sigma \sqrt{\sum_{j=1}^{p} h_{1j}^2}}. \tag{9}$$

Fix α, $m(\cdot)$, k, n, and d_1. As shown in Appendix E, for the standard confidence interval the risk function does not depend on the parameters (θ, σ^2); it is denoted R_s. As shown in Appendix G, the risk function for the alternative confidence interval depends on (θ, σ^2) only through $\bar{\theta}_p$. The risk function for this confidence interval is denoted $R_a(\bar{\theta}_p)$. As shown in Appendix H, the limit as $|\bar{\theta}_p| \to \infty$ of $R_a(\bar{\theta}_p)/R_s$ is 1.

Bickel (1984, Sect. 3) has considered confidence intervals for regression parameters after the preliminary application of a variable selection step. This work differs from the work presented here in a number of significant respects. These

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

include the following. (a) In the context considered here, Bickel used the model (5) but with $B\theta$ replaced by a p-dimensional parameter vector γ. We do not do so for the reason given in Appendix A. (b) Bickel assumed that σ^2 is known. We do not make this assumption. (c) Bickel considered confidence intervals that are restricted to having a fixed width. We do not make this restriction. (d) Bickel chose coverage probability to be the risk function. We require the confidence interval to have the desired minimum coverage probability and use a normalized expected length to be the risk function.

3. AN ASYMPTOTIC RESULT FOR THE ALTERNATIVE CONFIDENCE INTERVAL

Fix α, $m(\cdot)$, k, and d_1. Also suppose that $q = k$. The function $R_a^\infty(\cdot)$ is defined in Appendix I. In this appendix it is proved that

$$\lim_{n\to\infty} R_a(\bar{\theta}_p) = R_a^\infty(\bar{\theta}_p),$$

where the convergence is uniform in $\bar{\theta}_p \in \mathbf{R}$. Thus,

$$\lim_{n\to\infty} \inf_{\bar{\theta}_p} R_a(\bar{\theta}_p) = \inf_{\bar{\theta}_p} R_a^\infty(\bar{\theta}_p).$$

From Appendix E, $\lim_{n\to\infty} R_s = c_\infty$ where c_∞ satisfies $P\{|T| \le c_\infty\} = 1 - \alpha$ for $T \sim N(0,1)$. It is proved in Appendix J that

$$R_a^\infty(\bar{\theta}_p) \ge c_\infty \quad \forall \bar{\theta}_p \in \mathbf{R}.$$

Thus

$$\lim_{n\to\infty} \inf_{\bar{\theta}_p} R_a(\bar{\theta}_p) \ge c_\infty. \tag{10}$$

In other words, for n sufficiently large the alternative confidence interval cannot be better than the standard confidence interval by more than a negligibly small amount.

It may be asked whether there is a direct connection between the properties of the alternative confidence interval and the properties of the estimator $\tilde{\Theta}_1$. One piece of evidence in support of a negative answer to this question is as follows. Suppose that $k = q$ and that $m(x) = 0$ for all $x \in (-q, q)$. In this case, $\tilde{\Theta}_1 = \hat{\tilde{\Theta}}_1$. Also suppose that q and d_1 are fixed. Let us compare $\hat{\tilde{\Theta}}_1$ with $\hat{\Theta}_1$ using a normalized mean-square-error criterion. Now $E(\hat{\Theta}_1 - \theta_1)^2/\sigma^2 \sum_{j=1}^p h_{1j}^2 = 1$. It follows from the expression for $(\tilde{\Theta}_1 - \theta_1)/\sigma(\sum_{j=1}^p h_{1j}^2)^{1/2}$ given in Appendix F that $E(\hat{\tilde{\Theta}}_1 - \theta_1)^2/\sigma^2 \sum_{j=1}^p h_{1j}^2$ is a function of $\bar{\theta}_p$. It may be shown that this function converges to a limit as $n \to \infty$, uniformly in $\bar{\theta}_p \in \mathbf{R}$. Further, it may be shown that this limit is less than 1 for $\bar{\theta}_p = 0$ and that the limit function has a maximum that is greater than 1. This comparison of the mean-square errors of $\hat{\tilde{\Theta}}_1$ and $\hat{\Theta}_1$ stands in stark contrast to (10) and the fact that $\lim_{n\to\infty} R_s = c_\infty$.

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

4. MODIFIED INTERVAL, FINITE *n*

In this section we suppose that $q = k$ and that $m(x) = 0$ for all $x \in (-q, q)$. In this case the alternative confidence interval reduces to the modified interval described in the introduction.

Suppose α, k, $n - p$, and d_1 are given. To calculate c_n^* we proceed as follows. For given c we find the coverage probability of the confidence interval with center $\tilde{\Theta}_1$ and length $2K(c)(\sum_{j=1}^{p} h_{1j}^2)^{1/2}$ by evaluating (K.2) and then (K.1) from Appendix K by numerical integration over a range of positive $\bar{\theta}_p$ values. The smallest value of c for which this confidence interval satisfies the CP constraint is, then, c_n^*. For a given $\bar{\theta}_p$ we find $R_a(\bar{\theta}_p)$ by evaluating (K.4) and then (K.3) by numerical integration.

Consider $d_1 = 0$, $\alpha = 0.05$, and $n - p = 1$. The values of c_n^* were found to be 11.73, 8.91, and 6.64 for $k = 3$, 6, and 13, respectively. For each of the cases $(k, c_n^*) = (3, 11.73)$, $(6, 8.91)$, and $(13, 6.64)$ it was found that the coverage probability was close to 0.95 for all $\bar{\theta}_p$. As stated in Appendix G, $c_n^* = c_2 = 4.30$ for the limiting case $k = \infty$.

Figure 1 shows plots of $R_a(\bar{\theta}_p)$ as a function of $\bar{\theta}_p$. The three graphs are for $(k, c_n^*) = (6, 8.91)$, $(13, 6.64)$, and $(\infty, 4.30)$, respectively.

On the basis of these results we now describe a situation in which there are good grounds for choosing the alternative confidence interval over the standard confidence interval. Suppose that we have some prior information concerning the parameters of the model in the form of the reasonable anticipation (which is not a certainty) $|\bar{\theta}_p| \leq 0.15$. We also suppose that we do not have enough prior in-

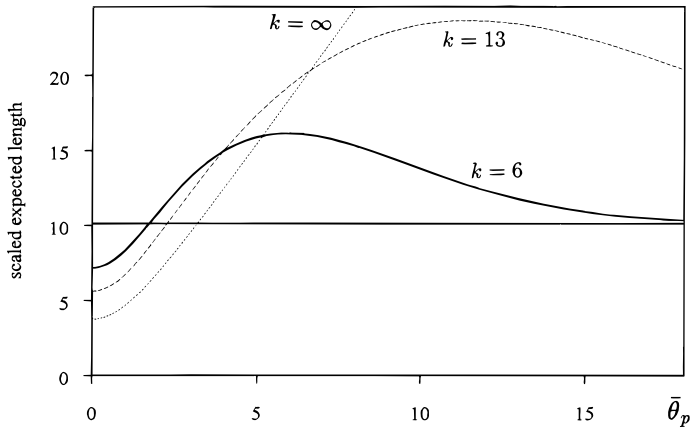


FIGURE 1. Plots of $R_a(\bar{\theta}_p)$, that is, the risk function of the alternative confidence interval as a function of $\bar{\theta}_p$ for $\alpha = 0.05$, $n - p = 1$, and $d_1 = 0$. The three curves are for $k = 6$, 13, and ∞. The horizontal line has ordinate equal to R_s, that is, the risk function of the standard confidence interval.

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

formation to specify a prior distribution for $(\theta_1, \ldots, \theta_p, \sigma^2)$. In other words, we suppose that a Bayes approach is not applicable. On the other hand, we would like to use to advantage the prior information about $\bar{\theta}_p$ in the preparation of the confidence interval. As far as the extent of prior information goes, the situation here is similar to that considered by Hodges and Lehmann (1952, p. 396 and, more particularly, p. 402). Suppose that we use the alternative confidence interval with $k = 13$. Figure 1 shows that if the reasonable anticipation that $|\bar{\theta}_p| \leq 0.15$ turns out to be correct then we have a substantial gain by using the alternative instead of the standard confidence interval in that $R_a(\bar{\theta}_p)$ is substantially less than R_s. However, this anticipated gain comes at a price because if the reasonable anticipation turns out to be incorrect and $|\bar{\theta}_p| \in [5, 20]$ then $R_a(\bar{\theta}_p)$ is very much larger than R_s.

Consider $R_a(\bar{\theta}_p)$ as a function of $\bar{\theta}_p$ ($\bar{\theta}_p \geq 0$) as we vary over k. The graphs shown in Figure 1 and some additional calculations of $R_a(\bar{\theta}_p)$ for $k = 3$ (which are not presented here) indicate that the value of $\bar{\theta}_p$ such that $R_a(\bar{\theta}_p)$ is equal to R_s is an increasing function of k. This suggests that the largest possible value of ξ such that $R_a(\bar{\theta}_p)/R_s \leq 1$ for all $\bar{\theta}_p$ satisfying $|\bar{\theta}_p| \leq \xi$ is approximately 2.25.

Now consider $\alpha = 0.05$ and $n - p = 1$ but $d_1 \neq 0$. For $d_1 \in (0, 0.6)$ and $k \in \{6, 13\}$ the graphs of $R_a(\bar{\theta}_p)$ as a function of $\bar{\theta}_p$ are very similar (though not identical) to those shown in Figure 1.

Finally, consider $\alpha = 0.05$, $d_1 = 0$, and $n - p \geq 1$. Fix $n - p \geq 1$. The following results are highly plausible: (a) for fixed $k \in (0, \infty)$ the risk function $R_a(\bar{\theta}_p)$ is minimized at $\bar{\theta}_p = 0$ and (b) $R_a(\bar{\theta}_p = 0)$ is a decreasing function of $k \in (0, \infty)$. If these results are true then

$$\inf_{k, \bar{\theta}_p} \left[\frac{R_a(\bar{\theta}_p = 0)}{R_s} \right] = \left[\frac{R_a(\bar{\theta}_p = 0)}{R_s} \quad \text{for } k = \infty \right].$$

The right-hand side takes the values 0.376, 0.769, 0.890, and 0.937 for $n - p = 1$, 2, 3, and 4, respectively.

5. PREDICTION INTERVALS

Consider the linear regression model $Y = \widetilde{X}\beta + \epsilon$ where Y is a random n-vector of responses, $\widetilde{X}$ is a known $n \times p$ matrix, $\beta = [\beta_1, \ldots, \beta_p]^T$ is an unknown parameter vector, and $\epsilon \sim N(0, \sigma^2 I_n)$ with σ^2 an unknown parameter satisfying $\sigma^2 > 0$. We suppose that the columns of $\widetilde{X}$ are linearly independent.

Suppose that $a^T\beta$ is the quantity of interest where $a = [a_1, \ldots, a_p]^T$ is a known vector and $a_i \neq 0$ for at least one $i \neq p$. Also suppose that we will apply variable selection for β_p. The model $Y = \widetilde{X}\beta + \epsilon$ may be expressed in the form $Y = X\theta + \epsilon$ where X is a known $n \times p$ matrix and $\theta = [\theta_1, \ldots, \theta_p]^T$ is an unknown parameter vector. Here, $\theta_1 = a^T\beta$ and $\theta_p = \beta_p$. Thus the problem of finding a confidence interval for $a^T\beta$ in the context of the model $Y = \widetilde{X}\beta + \epsilon$ and variable selection for β_p is mathematically equivalent to the problem of finding a confidence interval for θ_1 in the context of the model $Y = X\theta + \epsilon$ and variable selection for θ_p.

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

Now let our aim be to find a prediction interval for $Y^* = a^T\beta + \epsilon^*$ where $a = [a_1, \ldots, a_p]^T$ is a known vector and $a_i \neq 0$ for at least one $i \neq p$. Also suppose that we will apply variable selection for β_p. Express the model $Y = \widetilde{X}\beta + \epsilon$ in the form $Y = X\theta + \epsilon$ where X is a known $n \times p$ matrix and $\theta = [\theta_1, \ldots, \theta_p]^T$ is an unknown parameter vector. Here, $\theta_1 = a^T\beta$ and $\theta_p = \beta_p$. Thus the problem of finding a prediction interval for $Y^* = a^T\beta + \epsilon^*$ in the context of the model $Y = \widetilde{X}\beta + \epsilon$ and variable selection for β_p is equivalent to the problem of finding a prediction interval for $Y^* = \theta_1 + \epsilon^*$ in the context of the model $Y = X\theta + \epsilon$ and variable selection for θ_p. Consider the latter problem and suppose that the end result of our statistical analysis is a prediction interval for Y^* that is required to satisfy the constraint that the infimum over all values of (θ, σ^2) of the coverage probability of the prediction interval $\geq 1 - \alpha$. Now define the standard prediction interval, alternative prediction interval, and the risk function for prediction intervals in the same way as we define the corresponding objects for confidence intervals, except that $(\sum_{j=1}^{p} h_{1j}^2)^{1/2}$ is replaced by $(1 + \sum_{j=1}^{p} h_{1j}^2)^{1/2}$ and d_2 is replaced by $[(1 + \sum_{j=1}^{p} h_{1j}^2)/(1 + \sum_{j=1}^{p-1} h_{1j}^2)]^{1/2}$. As shown in Appendix L, with the appropriately altered definitions, the comparison of the standard and alternative prediction intervals is mathematically equivalent to the comparison of the standard and alternative confidence intervals.

6. DISCUSSION

The results of Sections 3 and 4 suggest that there is no significant gain in using the alternative confidence interval as opposed to the standard confidence interval when $n - p$ is large. Suppose, on the other hand, that $n - p$ is small. Also suppose that we have a reasonable anticipation (which is not a certainty) that $|\bar{\theta}_p|$ is small, less than some known $\xi > 0$. Provided that ξ is not too large and d_1 is not too large, we can find k such that (i) $R_a(\bar{\theta}_p = 0)$ is substantially less than R_s and $R_a(\bar{\theta}_p) < R_s$ for all $|\bar{\theta}_p| \leq \xi$ and (ii) the supremum over all $\bar{\theta}_p$ of $R_a(\bar{\theta}_p)/R_s$ is not too large. (i) and (ii) provide good grounds for preferring the alternative confidence interval to the standard confidence interval.

We should distinguish between the reasonable anticipation that $|\theta_p|/\sigma$ is small and the reasonable anticipation that $|\theta_p|$ is small. It is also of interest to ask which of these anticipations is more likely to be encountered in practice. If the reasonable anticipation that $|\theta_p|/\sigma$ is small is thought to arise fairly often in practice then an important question to ask is whether the decision rule corresponding to the alternative confidence interval is admissible in the class of decision rules satisfying the CP constraint. Here, we take the risk function to be that specified by (4) in the introduction.

In this paper we have restricted k and q to be positive. An analysis of the cases $(k = 0, q > 0)$ and $(k > 0, q = 0)$ is the subject of a paper that is in preparation.

REFERENCES

Bickel, P.J. & K.A. Doksum (1977) *Mathematical Statistics*. Oakland, California: Holden-Day.

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

Bickel, P.J. (1984) Parametric robustness: Small biases can be worthwhile. *Annals of Statistics* 12, 864–879.
Han, C.-P., C.V. Rao, & J. Ravichandran (1988) Inference based on conditional specification: A second bibliography. *Communications in Statistics—Theory and Methods* 17, 1945–1964.
Hodges, J.L. & E.L. Lehmann (1952) The use of previous experience in reaching statistical decisions. *Annals of Mathematical Statistics* 23, 396–407.
Judge, G.G. & M.E. Bock (1978) *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. Amsterdam: North-Holland.
Kabaila, P. (1995) The effect of model selection on confidence regions and prediction regions. *Econometric Theory* 11, 537–549.
Pötscher, B.M. (1995) Comment on "The effect of model selection on confidence regions and prediction regions" by P. Kabaila. *Econometric Theory* 11, 550–559.

APPENDIX A

In this appendix we use the QR decomposition to transform the model $Y = X\theta + \epsilon$ to a much more convenient form that involves in a simple manner both of the parameters θ_1 and θ_p. Let x_j denote the jth column of X. Find n-vectors $x_{p+1}, \ldots, x_n$ such that $x_1, \ldots, x_n$ form a basis for $\mathbf{R}^n$ and then apply the Gram–Schmidt orthogonalization process to $x_1, \ldots, x_n$ (in that order) to obtain $g_1, \ldots, g_n$ that are orthonormal, that is, $g_i^T g_j = 1$ for $i = j$ and 0 otherwise. Let G denote the $n \times n$ matrix with ith row equal to g_i^T. Let $Z = GY$, $A = GX$, and $\eta = G\epsilon$ so that $Z = A\theta + \eta$ where $\eta \sim N(0, \sigma^2 I_n)$. Here A is an $n \times p$ matrix whose last $n - p$ rows consist of zeros and whose first p rows form an upper triangular $p \times p$ matrix (with nonzero diagonal elements) that we denote by B. In other words, the transformed model is as described by (5).

There is another transformation of the model $Y = X\theta + \epsilon$ described, for example, in Sections 7.1 and 7.2 of Bickel and Doksum (1977). A particular case of this transformation leads to the model (5) but with $B\theta$ replaced by a p-dimensional parameter vector γ. The transformation described in Sections 7.1 and 7.2 of Bickel and Doksum (1977) does not, in general, result in a model involving both of the parameters θ_1 and θ_p in a simple way. For this reason we do not use this other transformation.

APPENDIX B

Suppose that $m(x) = 0$ for all $x \in (-q, q)$. In this appendix we prove that $\hat{\tilde{\Theta}}_1$ as defined in the introduction is given by (7). Write the matrix B in partitioned form as follows:

$$B = \begin{pmatrix} B_{11} & B_{12} \\ 0 & b_{pp} \end{pmatrix},$$

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

where B_{11} is a $(p-1)\times(p-1)$ matrix. In Section 2 we established the notation that $H = B^{-1}$ and that $[h_{11},\ldots,h_{1p}]$ is the first row of the matrix H. By the rule for the inversion of a partitioned matrix,

$$[h_{11},\ldots,h_{1p}] = [\text{first row of } B_{11}^{-1}, \text{first row of } -B_{11}^{-1}B_{12}b_{pp}^{-1}].$$

Thus

$$\hat{\tilde{\Theta}}_1 = \begin{cases} \sum_{j=1}^{p-1} h_{1j}Z_j + h_{1p}Z_p & \text{for } |\hat{\Theta}_p|/S\sqrt{v_X(p)} \geq q \\ \sum_{j=1}^{p-1} h_{1j}Z_j & \text{otherwise.} \end{cases}$$

Now $Z_p = b_{pp}\theta_p + \eta_p$ so that $\hat{\Theta}_p = b_{pp}^{-1}Z_p = \theta_p + b_{pp}^{-1}\eta_p$. Therefore the condition $|\hat{\Theta}_p|/S\sqrt{v_X(p)} \geq q$ is equivalent to $|Z_p|/S \geq q$.

APPENDIX C

In this and the appendixes that follow we will seek to express the quantities of interest in terms of $\bar{\theta}_p$, k, q, d_2, d_1, U, V, and W. The first five of these quantities were defined in Section 2. We now proceed to define the others. Define

$$U = \bar{\theta}_p + \frac{\eta_p}{\sigma},$$

$$V = \sum_{j=1}^{p-1} h_{1j}\frac{\eta_j}{\sigma}\Big/\sqrt{\sum_{j=1}^{p-1} h_{1j}^2},$$

$$W = \frac{S}{\sigma}.$$

These random variables are independent with $U \sim N(\bar{\theta}_p,1)$, $V \sim N(0,1)$, and W with the same distribution as $\sqrt{Q/(n-p)}$ where $Q \sim \chi^2_{n-p}$.

APPENDIX D

In this appendix we fix α, $m(\cdot)$, k, and n. We suppose that all parameters of the model are fixed, apart from $\bar{\theta}_p$. We will show that $\hat{\Theta}_1 - \tilde{\Theta}_1 \xrightarrow{p} 0$ and $K(c_n^*) - c_{n-p}S \xrightarrow{p} 0$ as $|\bar{\theta}_p| \to \infty$. Note that the existence of c_n^* is proved in Appendix G.

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

From $Z_p = b_{pp}\theta_p + \eta_p$ it follows that

$$\frac{Z_p}{S} = \frac{(b_{pp}\theta_p + \eta_p)/\sigma}{S/\sigma} = \frac{U}{W}.$$

Therefore the conditions $|Z_p|/S \geq q$ is equivalent to $|U|/W \geq q$. Thus $\tilde{\Theta}_1 = \hat{\Theta}_1$ for $|U|/W \geq q$ and $K(c_n^*) = c_{n-p}S$ for $|U|/W \geq k$. The claimed result follows directly from this and the fact that $P\{|U|/W \geq q\} \to 1$ and $P\{|U|/W \geq k\} \to 1$ as $|\bar{\theta}_p| \to \infty$.

APPENDIX E

For the standard confidence interval

$$(9) = c_{n-p}E\{W\}$$

$$= \frac{c_{n-p}}{\sqrt{n-p}}E\{\sqrt{Q}\}$$

$$= c_{n-p}\sqrt{\frac{2}{n-p}} \times \frac{\Gamma\left(\frac{n-p+1}{2}\right)}{\Gamma\left(\frac{n-p}{2}\right)},$$

where $Q \sim \chi^2_{n-p}$.

APPENDIX F

In this appendix we express $(\tilde{\Theta}_1 - \theta_1)/\sigma(\sum_{j=1}^p h_{1j}^2)^{1/2}$ and $K(c)/\sigma$ in terms of $\bar{\theta}_p$, k, q, d_2, d_1, U, V, and W. The first five of these quantities were defined in Section 2 and the remainder were defined in Appendix C.

From $\theta = H\{[Z_1,\ldots,Z_p]^T - [\eta_1,\ldots,\eta_p]^T\}$ it follows that $\theta_1 = \sum_{j=1}^p h_{1j}(Z_j - \eta_j)$. It follows from (7) that

$$\tilde{\Theta}_1 - \theta_1 = \sum_{j=1}^{p-1} h_{1j}\eta_j + h_{1p}\left[Sm\left(\frac{Z_p}{S}\right) - b_{pp}\theta_p\right].$$

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

It now follows that

$$\frac{\tilde{\Theta}_1 - \theta_1}{\sigma} = \sum_{j=1}^{p-1} h_{1j} \frac{\eta_j}{\sigma} + h_{1p}\left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p\right].$$

Thus

$$\begin{aligned}\frac{\tilde{\Theta}_1 - \theta_1}{\sigma\sqrt{\sum_{j=1}^{p} h_{1j}^2}} &= \left\{\frac{\sum_{j=1}^{p-1} h_{1j}^2 \frac{\eta_j}{\sigma}}{\sqrt{\sum_{j=1}^{p-1} h_{1j}^2}} + \frac{h_{1p}}{\sqrt{\sum_{j=1}^{p-1} h_{1j}^2}}\left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p\right]\right\} \Bigg/ \sqrt{\frac{\sum_{j=1}^{p} h_{1j}^2}{\sum_{j=1}^{p-1} h_{1j}^2}} \\ &= \frac{V + d_1\left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p\right]}{d_2}.\end{aligned}$$

It is straightforward to show that

$$\frac{K(c)}{\sigma} = \begin{cases} c_{n-p} W & \text{for } |U|/W \geq k \\ c d_2^{-1} \sqrt{\dfrac{U^2 + (n-p)W^2}{n-p+1}} & \text{otherwise,} \end{cases}$$

where $K(\cdot)$ was defined in Section 2.

APPENDIX G

The coverage probability of the confidence interval with center $\tilde{\Theta}_1$ and length $2K(c)(\sum_{j=1}^{p} h_{1j}^2)^{1/2}$ is equal to

$$P\left\{\frac{|\tilde{\Theta}_1 - \theta_1|}{\sigma\sqrt{\sum_{j=1}^{p} h_{1j}^2}} \leq \frac{K(c)}{\sigma}\right\} = P\left\{\left|V + d_1\left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p\right]\right| \leq d_2 \frac{K(c)}{\sigma}\right\} \qquad \textbf{(G.1)}$$

by the expression for $(\tilde{\Theta}_1 - \theta_1)/\sigma(\sum_{j=1}^{p} h_{1j}^2)^{1/2}$ given in Appendix F.

Fix $\alpha, m(\cdot), k, n,$ and d_1. From the expression for $K(c)/\sigma$ given in Appendix F it follows that this probability is a function of $(c, \bar{\theta}_p)$. Hence c_n^*, if it exists, depends only on $\bar{\theta}_p$. The risk function (9) for the alternative confidence interval is $E\{K(c_n^*)/\sigma\}$. From the expression for $K(c)/\sigma$ given in Appendix F it follows that this risk is a function of $\bar{\theta}_p$.

We now prove the existence of c_n^*. Remember that c_n^* is defined to be the smallest value of c such that (G.1) $\geq 1 - \alpha$ for all $\bar{\theta}_p \in \mathbf{R}$. We present the proof for $k \geq q$. The proof for $k < q$ is based on similar ideas and is omitted for the sake of brevity. Define the events

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

$$D = \left\{ \left| V + d_1 \left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p \right] \right| \leq d_2 \frac{K(c)}{\sigma} \right\},$$

$$E = \{|V + d_1[U - \bar{\theta}_p]| \leq d_2 c_{n-p} W\}.$$

Now $P\{D\} = P\{D \cap \{|U| < kW\}\} + P\{D \cap \{|U| \geq kW\}\}$ and $P\{E\} = P\{E \cap \{|U| < kW\}\} + P\{E \cap \{|U| \geq kW\}\}$. Because $P\{E\} = 1 - \alpha$ and $P\{D \cap \{|U| \geq kW\}\} = P\{E \cap \{|U| \geq kW\}\}$,

$$P\{D\} - (1 - \alpha) = P\{D \cap \{|U| < kW\}\} - P\{E \cap \{|U| < kW\}\}. \quad \textbf{(G.2)}$$

For $c = 0$, $P\{D \cap \{|U| < kW\}\} = 0$ and from $P\{E \cap \{|U| < kW\}\} > 0$ it follows that $P\{D\} < 1 - \alpha$ for all $\bar{\theta}_p \in \mathbf{R}$. Define

$$D_1 = \left\{ \left| V + d_1[U - \bar{\theta}_p] + d_1 \left[Wm\left(\frac{U}{W}\right) - U \right] \right| \leq c \sqrt{\frac{n-p}{n-p+1}}\, W \right\},$$

$$D_2 = \left\{ |V + d_1[U - \bar{\theta}_p]| \leq \left(c \sqrt{\frac{n-p}{n-p+1}} - d_1 k \right) W \right\}.$$

Now $(D \cap \{|U| < kW\}) \supset (D_1 \cap \{|U| < kW\}) \supset (D_2 \cap \{|U| < kW\})$ so that $P\{D \cap \{|U| < kW\}\} \geq P\{D_2 \cap \{|U| < kW\}\}$. It follows from (G.2) that

$$P\{D\} - (1 - \alpha) \geq P\{D_2 \cap \{|U| < kW\}\} - P\{E \cap \{|U| < kW\}\}.$$

For sufficiently large c, $E \subset D_2$ for all $\bar{\theta}_p \in \mathbf{R}$. Thus there exists $c < \infty$ such that $P\{D\} \geq 1 - \alpha$ for all $\bar{\theta}_p \in \mathbf{R}$. For each $\bar{\theta}_p$, $P\{D\}$ is a continuous increasing function of $c (c \geq 0)$. Thus $c(\bar{\theta}_p)$ exists for each $\bar{\theta}_p$ where $c(\bar{\theta}_p)$ is defined to be the value of c such that $P\{D\}$ evaluated at $\bar{\theta}_p$ is equal to $1 - \alpha$. Because there exists $c < \infty$ such that $P\{D\} \geq 1 - \alpha$ for all $\bar{\theta}_p \in \mathbf{R}$, the set $\{c(\bar{\theta}_p) : \bar{\theta}_p \in \mathbf{R}\}$ is bounded above. Hence $\sup\{c(\bar{\theta}_p) : \bar{\theta}_p \in \mathbf{R}\}$ exists. It is easily checked that $c_n^* = \sup\{c(\bar{\theta}_p) : \bar{\theta}_p \in \mathbf{R}\}$. This completes the proof of the existence of c_n^*.

We now present a computationally useful formula for the risk function of the alternative confidence interval when $d_1 = 0$ and k takes the limiting value ∞. For this case $c_n^* = c_{n-p+1}$ and

$$E\{K(c_n^*)/\sigma\} = \frac{c_{n-p+1}}{\sqrt{n-p+1}}\, E\left(\sqrt{\tilde{Q}}\right),$$

where $\tilde{Q} \sim$ noncentral χ^2 with $n - p + 1$ degrees of freedom and noncentrality parameter $\lambda = \bar{\theta}_p^2$. Using the representation of the density function of $\tilde{Q}$ as a mixture of (central) χ^2 density functions we find that

$$E\left(\sqrt{\tilde{Q}}\right) = \sqrt{2} \sum_{k=0}^{\infty} \frac{\Gamma\left[\left(\frac{n-p+1}{2} + k\right) + \frac{1}{2}\right]}{\Gamma\left(\frac{n-p+1}{2} + k\right)} \times \frac{(\frac{1}{2}\lambda)^k e^{-(1/2)\lambda}}{k!}.$$

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

APPENDIX H

In this appendix we show that the limit as $|\bar{\theta}_p| \to \infty$ of $R_a(\bar{\theta}_p)/R_s$ is 1. Observe that $K(c_n^*)/\sigma$ converges to $c_{n-p}W$ in probability as $|\bar{\theta}_p| \to \infty$. For $|U|/W < k$, $U^2 + (n-p)W^2 < (k^2 + n - p)W^2$. Thus, by dominated convergence, $R_a(\bar{\theta}_p) = E\{K(c_n^*)\}/\sigma \to c_{n-p}E\{W\} = R_s$ as $|\bar{\theta}_p| \to \infty$.

APPENDIX I

Fix α, $m(\cdot)$, k, and d_1. Also suppose that $q = k$. Define the random variable

$$F(c) = \begin{cases} c_\infty & \text{for } |U| \geq k \\ cd_2^{-1} & \text{otherwise.} \end{cases}$$

Observe that $K(c)/\sigma \leq W \times \max\{c_{n-p}, cd_2^{-1}[(k^2 + n - p)/(n - p + 1)]^{1/2}\}$. Also observe that when $W \in [1-\epsilon, 1+\epsilon]$ where $\epsilon \in (0,1)$, $(1-\epsilon)m_n(c) \leq K(c)/\sigma \leq (1+\epsilon)\bar{m}_n(c)$ where

$$m_n(c) = \begin{cases} c_{n-p} & \text{for } |U| > k(1+\epsilon) \\ \min(c_{n-p}, cd_2^{-1}\sqrt{(k^2 + n - p)/(n - p + 1)}) & \text{for } k(1-\epsilon) \leq |U| \leq k(1+\epsilon) \\ cd_2^{-1}\sqrt{(n-p)/(n-p+1)} & \text{for } |U| < k(1-\epsilon) \end{cases}$$

and

$$\bar{m}_n(c) = \begin{cases} c_{n-p} & \text{for } |U| > k(1+\epsilon) \\ \max(c_{n-p}, cd_2^{-1}\sqrt{(k^2 + n - p)/(n - p + 1)}) & \text{for } k(1-\epsilon) \leq |U| \leq k(1+\epsilon) \\ cd_2^{-1}\sqrt{(k^2 + n - p)/(n - p + 1)} & \text{for } |U| < k(1-\epsilon). \end{cases}$$

From these inequalities it follows that for each $M \in (0, \infty)$,

$$\lim_{n\to\infty} E\{K(c)/\sigma\} = E\{F(c)\},$$

where the convergence is uniform in $\bar{\theta}_p \in \mathbf{R}$ and $c \in (0, M)$.

Define the events

$$A_n = \left\{ \left| V + d_1 \left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p \right] \right| \leq d_2 \frac{K(c)}{\sigma} \right\}$$

and $B = \{|V + d_1[m(U) - \bar{\theta}_p]| \leq d_2 F(c)\}$. Our next task is to show that for each $M \in (0, \infty)$, $\lim_{n\to\infty} P\{A_n\} = P\{B\}$ where the convergence is uniform in $\bar{\theta}_p \in \mathbf{R}$ and $c \in (0, M)$. Fix $\epsilon \in (0,1)$. Write

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

$$\begin{aligned}P\{A_n\} = {} & P\{A_n \cap \{W \notin [1-\epsilon, 1+\epsilon]\}\} \\ & + P\{A_n \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{|U| > k(1+\epsilon)\}\} \\ & + P\{A_n \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{k(1-\epsilon) \leq |U| \leq k(1+\epsilon)\}\} \\ & + P\{A_n \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{|U| < k(1-\epsilon)\}\}\end{aligned}$$

and

$$\begin{aligned}P\{B\} = {} & P\{B \cap \{W \notin [1-\epsilon, 1+\epsilon]\}\} \\ & + P\{B \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{|U| > k(1+\epsilon)\}\} \\ & + P\{B \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{k(1-\epsilon) \leq |U| \leq k(1+\epsilon)\}\} \\ & + P\{B \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{|U| < k(1-\epsilon)\}\}.\end{aligned}$$

Observe that both $P\{A_n \cap \{W \notin [1-\epsilon, 1+\epsilon]\}\}$ and $P\{B \cap \{W \notin [1-\epsilon, 1+\epsilon]\}\}$ are bounded above by $P\{W \notin [1-\epsilon, 1+\epsilon]\}$, a function of n converging to zero as $n \to \infty$. Also observe that the magnitude of the difference between $P\{A_n \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{k(1-\epsilon) \leq |U| \leq k(1+\epsilon)\}\}$ and $P\{B \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{k(1-\epsilon) \leq |U| \leq k(1+\epsilon)\}\}$ is bounded above by $P\{k(1-\epsilon) \leq |U| \leq k(1+\epsilon)\} \leq 4k\epsilon/\sqrt{2\pi}$. Now $P\{A_n \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{|U| > k(1+\epsilon)\}\}$ lies between

$$\begin{aligned}P\{\{|V + d_1[U - \bar{\theta}_p]| \leq d_2 c_{n-p}(1-\epsilon)\} \cap \{|U| > k(1+\epsilon)\}\} \\ \times P\{W \in [1-\epsilon, 1+\epsilon]\}\end{aligned} \tag{I.1}$$

and

$$\begin{aligned}P\{\{|V + d_1[U - \bar{\theta}_p]| \leq d_2 c_{n-p}(1+\epsilon)\} \cap \{|U| > k(1+\epsilon)\}\} \\ \times P\{W \in [1-\epsilon, 1+\epsilon]\}.\end{aligned} \tag{I.2}$$

Also, $P\{B \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{|U| > k(1+\epsilon)\}\}$ is equal to

$$P\{\{|V + d_1[U - \bar{\theta}_p]| \leq d_2 c_\infty\} \cap \{|U| > k(1+\epsilon)\}\} \times P\{W \in [1-\epsilon, 1+\epsilon]\}. \tag{I.3}$$

Further, $P\{A_n \cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{|U| < k(1-\epsilon)\}\}$ lies between

$$\begin{aligned}P\left\{\left\{\left|V + d_1\left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p\right]\right| \leq c\sqrt{\frac{n-p}{n-p+1}}\,(1-\epsilon)\right\}\right. \\ \left.\cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{|U| < k(1-\epsilon)\}\right\}\end{aligned} \tag{I.4}$$

and

$$\begin{aligned}P\left\{\left\{\left|V + d_1\left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p\right]\right| \leq c\sqrt{\frac{k^2+n-p}{n-p+1}}\,(1+\epsilon)\right\}\right. \\ \left.\cap \{W \in [1-\epsilon, 1+\epsilon]\} \cap \{|U| < k(1-\epsilon)\}\right\}.\end{aligned} \tag{I.5}$$

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

Also, $P\{B \cap \{W \in [1-\epsilon,1+\epsilon]\} \cap \{|U| < k(1-\epsilon)\}\}$ is equal to

$$P\{\{|V + d_1[m(U) - \bar{\theta}_p]| \le c\} \cap \{W \in [1-\epsilon,1+\epsilon]\} \cap \{|U| < k(1-\epsilon)\}\}. \quad \textbf{(I.6)}$$

Approximate (I.1) and (I.2) by (I.3) and approximate (I.4) and (I.5) by (I.6). In judging the quality of the latter approximation, we consider distributions conditional on $\{W \in [1-\epsilon,1+\epsilon]\} \cap \{|U| < k(1-\epsilon)\}$ and we use the assumption made in Section 2 that the restriction of $m(\cdot)$ to the interval $(-q,q)$ is a differentiable function with bounded derivative. By first choosing ϵ sufficiently small and then choosing n sufficiently large, it may be shown that for each $M \in (0,\infty)$,

$$\lim_{n\to\infty} P\{A_n\} = P\{B\},$$

where the convergence is uniform in $\bar{\theta}_p \in \mathbf{R}$ and $c \in (0,M)$. The importance of the uniformity of convergence in this type of context has been stressed by Kabaila (1995). Thus $\lim_{n\to\infty} c_n^* = \tilde{c}_\infty$ where $\tilde{c}_\infty$ is defined to be the smallest value of c such that

$$P\{|V + d_1[m(U) - \bar{\theta}_p]| \le d_2 F(c)\} \ge 1-\alpha \quad \forall \bar{\theta}_p \in \mathbf{R}. \quad \textbf{(I.7)}$$

Therefore, $\lim_{n\to\infty}\{K(c_n^*)/\sigma\} = E\{F(\tilde{c}_\infty)\}$ where the convergence is uniform in $\bar{\theta}_p \in \mathbf{R}$. In other words, $\lim_{n\to\infty} R_a(\bar{\theta}_p) = R_a^\infty(\bar{\theta}_p)$ where the convergence is uniform in $\bar{\theta}_p \in \mathbf{R}$. Here, we define $R_a^\infty(\bar{\theta}_p) = E\{F(\tilde{c}_\infty)\}$.

APPENDIX J

Fix α, $m(\cdot)$, k, and d_1. In this appendix we show that

$$E\{F(\tilde{c}_\infty)\} \ge c_\infty \quad \forall \bar{\theta}_p \in \mathbf{R}, \quad \textbf{(J.1)}$$

where $\tilde{c}_\infty$ is defined to be the smallest value of c such that (I.7) is satisfied. For $d_1 = 0$ it is straightforward to show that $\tilde{c}_\infty \ge c_\infty$. From $E\{F(\tilde{c}_\infty)\} = c_\infty P\{|U| \ge k\} + \tilde{c}_\infty P\{|U| < k\}$ it follows that (J.1) is satisfied.

Suppose, now, that $d_1 \ne 0$. We prove (J.1) by contradiction. Suppose that $c_\infty > E\{F(\tilde{c}_\infty)\}$ for some $\bar{\theta}_p \in \mathbf{R}$. This implies that $\tilde{c}_\infty d_2^{-1} < c_\infty$. Therefore

$$P\{|V + d_1[m(U) - \bar{\theta}_p]| \le d_2 F(c)\} \le P\{|V + d_1[m(U) - \bar{\theta}_p]| \le d_2 c_\infty\}$$

for all $\bar{\theta}_p \in \mathbf{R}$. We now show that (I.1) is not satisfied.

$$P\{|V + d_1[m(U) - \bar{\theta}_p]| \le d_2 c_\infty\} = \int_{-q}^{q} f(d_1(m(u) - \bar{\theta}_p),\, d_2 c_\infty)\phi(u - \bar{\theta}_p)\, du + \int_{\mathcal{A}} f(d_1(u - \bar{\theta}_p), d_2 c_\infty)\phi(u - \bar{\theta}_p)\, du, \quad \textbf{(J.2)}$$

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

where $\mathcal{A} = (-\infty,-q) \cup (q,\infty)$ and $f(x,y) = P\{|V + x| \le y\} = \Phi(-x + y) - \Phi(-x - y)$. Here Φ and ϕ denote the $N(0,1)$ distribution function and probability density function, respectively. Now

$$1 - \alpha = P\{|V + d_1[U - \bar{\theta}_p]| \le d_2 c_\infty\}$$
$$= \int_{-q}^{q} f(d_1(u - \bar{\theta}_p), d_2 c_\infty)\phi(u - \bar{\theta}_p)\, du + \int_{\mathcal{A}} f(d_1(u - \bar{\theta}_p), d_2 c_\infty)\phi(u - \bar{\theta}_p)\, du$$

so that

$$(1 - \alpha) - (\text{J.2}) = \int_{-q}^{q} [\, f(d_1(u - \bar{\theta}_p), d_2 c_\infty) - f(d_1(m(u) - \bar{\theta}_p), d_2 c_\infty)]\phi(u - \bar{\theta}_p)\, du. \tag{J.3}$$

By using $\int_{-q}^{q} = \int_0^q + \int_{-q}^0$ and changing the variable of integration in $\int_{-q}^0$, it may be shown that (J.3) is equal to

$$\int_0^q \phi(u - \bar{\theta}_p) \times [l(u - \bar{\theta}_p) - l(m(u) - \bar{\theta}_p)] - \phi(-u - \bar{\theta}_p)$$
$$\times [l(m(-u) - \bar{\theta}_p) - l(-u - \bar{\theta}_p)]\, du, \tag{J.4}$$

where $l(x) = f(d_1 x, d_2 c_\infty)$. For $\bar{\theta}_p$ sufficiently large $d^2 l(x - \bar{\theta}_p)/dx^2 > 0$ for all $x \in [-q,q]$. Thus, for $\bar{\theta}_p$ sufficiently large, $\phi(u - \bar{\theta}_p) > \phi(-u - \bar{\theta}_p)$ and $l(u - \bar{\theta}_p) - l(m(U) - \bar{\theta}_p) > l(m(-u) - \bar{\theta}_p) - l(-u - \bar{\theta}_p)$ for all $u \in (0,q)$. Thus (J.4) > 0 for $\bar{\theta}_p$ sufficiently large.

APPENDIX K

Suppose that $q = k$ and $m(x) = 0$ for all $x \in (-q,q)$. We derive easily computed expressions for the coverage probability of the confidence interval with center $\tilde{\Theta}_1 = \hat{\Theta}_1$ and length $2K(c)(\sum_{j=1}^p h_{1j}^2)^{1/2}$ and for $R_a(\bar{\theta}_p)$. From Appendix G, this coverage probability is equal to $P\{D\}$ where

$$D = \left\{ \left| V + d_1 \left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p \right] \right| \le d_2 \frac{K(c)}{\sigma} \right\}.$$

Now

$$P\{D\} = \int_0^\infty P\{D \mid W = w\} p_W(w)\, dw, \tag{K.1}$$

where $P\{D \mid W = w\}$ denotes the probability of D conditional on $W = w$ and $p_W(w)$ denotes the density function of W evaluated at w. Also

$$P\{D \mid W = w\} = I_1(w) + I_2(w), \tag{K.2}$$

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

where

$$I_1(w) = \int_{-kw}^{kw} f\left(-d_1\bar{\theta}_p, c\sqrt{\frac{u^2+(n-p)w^2}{n-p+1}}\right)\phi(u-\bar{\theta}_p)\,du,$$

$$I_2(w) = \int_{\mathcal{A}} f(d_1(u-\bar{\theta}_p),\, d_2 c_{n-p} w)\,\phi(u-\bar{\theta}_p)\,du.$$

Here $\mathcal{A} = (-\infty, -kw) \cup (kw, \infty)$.

Using arguments similar to those leading to (K.1) and (K.2), it may be shown that

$$E\{K(c_n^*)/\sigma\} = \int_0^\infty L(w) p_W(w)\,dw, \tag{K.3}$$

where

$$L(w) = c_n^* d_2^{-1} \int_{-kw}^{kw} \sqrt{\frac{u^2+(n-p)w^2}{n-p+1}}\,\phi(u-\bar{\theta}_p)\,du + c_{n-p} w\{1-\Phi(kw-\bar{\theta}_p)+\Phi(-kw-\bar{\theta}_p)\}. \tag{K.4}$$

APPENDIX L

Suppose that we wish to find a prediction interval for $Y^* = \theta_1 + \epsilon$ in the context of the model $Y = X\theta + \epsilon$ (described in the introduction) and variable selection for θ_p. Replace the definition of the CP constraint by the following definition.

CP Constraint. The infimum over all values of (θ, σ^2) of the coverage probability of the prediction interval $\geq 1-\alpha$.

Define the standard prediction interval to be

$$\left[\hat{\Theta}_1 - c_{n-p} S\sqrt{1+\sum_{j=1}^{p} h_{1j}^2}, \hat{\Theta}_1 + c_{n-p} S\sqrt{1+\sum_{j=1}^{p} h_{1j}^2}\right].$$

Replace the definition of d_2 by $d_2 = [(1+\sum_{j=1}^{p} h_{1j}^2)/(1+\sum_{j=1}^{p-1} h_{1j}^2)]^{1/2}$. Also replace the definition of c_n^* by the following. Define c_n^* to be the smallest value of c such that a prediction interval with center $\tilde{\Theta}_1$ and length $2K(c)(1+\sum_{j=1}^{p} h_{1j}^2)^{1/2}$ satisfies the CP constraint. The alternative prediction interval is defined to be

$$\left[\tilde{\Theta}_1 - K(c_n^*)\sqrt{1+\sum_{j=1}^{p} h_{1j}^2}, \tilde{\Theta}_1 + K(c_n^*)\sqrt{1+\sum_{j=1}^{p} h_{1j}^2}\right].$$

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press

The risk function for a prediction interval satisfying the CP constraint is defined to be

$$\frac{E\{\text{length of prediction interval}\}}{2\sigma\sqrt{1+\sum_{j=1}^{p} h_{1j}^2}}.$$

Replace the definitions of V and d_1 by

$$V = \left(\sum_{j=1}^{p-1} h_{1j}\frac{\eta_j}{\sigma} - \frac{\epsilon^*}{\sigma}\right)\Big/\sqrt{1+\sum_{j=1}^{p-1} h_{1j}^2},$$

$$d_1 = h_{1p}\Big/\sqrt{1+\sum_{j=1}^{p-1} h_{1j}^2}.$$

Using the same sorts of manipulations as in Appendix G it may be shown that

$$\frac{\tilde{\Theta}_1 - \theta_1 - \epsilon^*}{\sigma\sqrt{1+\sum_{j=1}^{p} h_{1j}^2}} = \frac{V + d_1\left[Wm\left(\frac{U}{W}\right) - \bar{\theta}_p\right]}{d_2}.$$

The mathematical equivalence of the comparison of the standard and alternative confidence intervals now follows from this and the fact that $d_2 = \sqrt{1+d_1^2}$ for both the original and altered definitions of d_1 and d_2.

https://doi.org/10.1017/S0266466698144031 Published online by Cambridge University Press