

Confidence intervals centred on bootstrap smoothed estimators: an impossibility result

Paul Kabaila* and Christeen Wijethunga

Department of Mathematics and Statistics, La Trobe University, Melbourne, Australia

arXiv:1910.09695v1 [math.ST] 21 Oct 2019

* Corresponding author. Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia. Tel.: +61 3 9479 2594; fax +61 3 9479 2466. *E-mail address:* P.Kabaila@latrobe.edu.au.

ABSTRACT

Recently, Kabaila and Wijethunga assessed the performance of a confidence interval centred on a bootstrap smoothed estimator, with width proportional to an estimator of Efron's delta method approximation to the standard deviation of this estimator. They used a testbed situation consisting of two nested linear regression models, with error variance assumed known, and model selection using a preliminary hypothesis test. This assessment was in terms of coverage and scaled expected length, where the scaling is with respect to the expected length of the usual confidence interval with the same minimum coverage probability. They found that this confidence interval has scaled expected length that (a) has a maximum value that may be much greater than 1 and (b) is greater than a number slightly less than 1 when the simpler model is correct. We therefore ask the following question. For a confidence interval, centred on the bootstrap smoothed estimator, does there exist a formula for its data-based width such that, in this testbed situation, it has the desired minimum coverage and scaled expected length that (a) has a maximum value that is not too much larger than 1 and (b) is substantially less than 1 when the simpler model is correct? Using a recent decision-theoretic performance bound due to Kabaila and Kong, it is shown that the answer to this question is 'no' for a wide range of scenarios.

KEYWORDS

confidence interval; coverage probability; decision theory; bootstrap smoothed estimator; model selection; nested linear regression models; scaled expected length.

1. Introduction

A bootstrap smoothed (or bagged, Breiman, 1996) estimator is a smoothed version of an estimator found after preliminary data-based model selection. A key result of Efron (2014) is a very convenient and widely applicable formula for a delta method approximation to the standard deviation of this estimator. Efron (2014) also considered a confidence interval, with nominal coverage $1 - \alpha$, centred on the bootstrap estimator and with half-width equal to the $1 - \alpha/2$ quantile of the standard normal distribution multiplied by an estimate of this approximation to the standard deviation. To rigorously assess the performance of this confidence interval, Kabaila & Wijethunga (2019a) use the following testbed situation. They consider two nested normal linear regression models with known error variance and parameter of interest θ , a specified linear combination of the regression parameters. These two nested models are the full model and the simpler model. The simpler model is obtained when τ , a distinct specified linear combination of the regression parameters, is set to 0. The bootstrap smoothed estimator that they consider is a smoothed version of the post-model-selection estimator obtained after a preliminary test of the null hypothesis that $\tau = 0$ against the alternative hypothesis that $\tau \neq 0$.

For the testbed situation they consider, Kabaila & Wijethunga (2019a) derive computationally convenient exact formulas for the ideal (i.e. in the limit as the number of bootstrap replications diverges to infinity) bootstrap smoothed estimator and Efron's delta method approximation to the standard deviation of this estimator. They also assess the performance of the confidence interval, with nominal coverage $1 - \alpha$, centred on the ideal bootstrap estimator and with half-width equal to the $1 - \alpha/2$ quantile of the standard normal distribution multiplied by an estimate of the ideal delta method approximation to the standard deviation. We call this the **sd_{delta} interval**. This confidence interval has the attractive features that (A1) it has endpoints that are continuous functions of the data and (A2) to an excellent approximation it reverts to the usual $1 - \alpha$ confidence interval based on the full model when the data and the simpler model are highly discordant.

For the testbed situation, Kabaila & Wijethunga (2019a) assess the performance of the **sd_{delta} interval** using its coverage probability and scaled expected length, defined as follows. The scaled expected length of a confidence interval is defined to be its expected length divided by the expected length of the usual confidence interval, with the same minimum coverage probability, based on the full model. Let ρ denote the known correlation between the least squares estimators of θ and τ . For $\rho = 0$, the **sd_{delta} interval** is identical to the usual $1 - \alpha$ confidence interval based on the full model. However, as $|\rho|$ increases, these two confidence intervals increasingly differ from each other.

Define the parameter γ to be τ divided by the standard deviation of the least squares estimator of τ , so that γ is unknown. For given nominal coverage $1 - \alpha$ and size of preliminary test, both the coverage probability and the scaled expected length of the **sd_{delta} interval** are functions of $|\rho|$ and $|\gamma|$. Figure 5 of Kabaila & Wijethunga (2019a) and Figures 6–10 of the Supplementary Material for this paper show the following. The **sd_{delta} interval** has scaled expected length that (a) has a maximum value that is an increasing function of $|\rho|$ that can be much larger than 1 for large $|\rho|$ and (b) is greater than a number slightly less than 1 when the simpler model is correct (i.e. when $\gamma = 0$).

In the context of the testbed situation used by Kabaila & Wijethunga (2019a), we have sought a formula for the width of a confidence interval centred on the ideal bootstrap smoothed estimator that leads to this confidence interval having scaled expected length that is substantially less than 1 when the simpler model is correct. We tried five different formulae, including the width being based on the actual standard deviation and the parametric version of the symmetric nonparametric bootstrap confidence described by Hall (1992, Section 3.6). None of these confidence intervals possessed this desired property. This then raises the following question:

In the context of this testbed situation, is there any formula for the width of a confidence interval centred on the ideal bootstrap smoothed estimator that leads to this confidence interval having the attractive features (A1) and (A2), the desired minimum coverage probability $1 - \alpha$ and scaled expected length that (a) has a maximum value that is not too much larger than 1 and (b) is

substantially less than 1 when the simpler model is correct?

We use a performance bound derived by Kabaila & Kong (2016) to answer this question. This bound builds on the earlier work of Blyth (1951), Hodges & Lehmann (1952), Kempthorne (1983, 1987, 1988) and Kabaila & Tuck (2008). As shown in Section 3 of the present paper, the application of this performance bound is carried out by computing two unfavourable discrete distributions, each consisting of a finite number of probability masses. We have programmed this computation in R, using simple code that is based on the theoretical results described in Sections 3 and 5 and Appendix A.3. For the scenarios described in Section 6 the answer to the question above is ‘no’.

2. Mathematical specification of the question we will answer

We consider the testbed situation consisting of two nested linear regression models: the full model \mathcal{M}_2 and the simpler model \mathcal{M}_1 . Suppose that the full model \mathcal{M}_2 is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where \mathbf{y} is a random n -vector of responses, \mathbf{X} is a known $n \times p$ matrix with linearly independent columns ($p < n$), $\boldsymbol{\beta}$ is an unknown p -vector of parameters and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, with σ^2 assumed known. Suppose that $\boldsymbol{\beta} = [\theta, \tau, \boldsymbol{\lambda}^\top]^\top$, where θ is the scalar parameter of interest, τ is a scalar parameter used in specifying the model \mathcal{M}_1 and $\boldsymbol{\lambda}$ is a $(p - 2)$ -dimensional parameter vector. The model \mathcal{M}_1 is \mathcal{M}_2 with $\tau = 0$. As shown by Kabaila & Wijethunga (2019a), this scenario can be obtained by a change of parametrisation from a more general scenario.

Let $\hat{\theta}$ and $\hat{\tau}$ denote the least squares estimators of θ and τ , respectively. Let $v_\theta = \text{var}(\hat{\theta})/\sigma^2$, $v_\tau = \text{var}(\hat{\tau})/\sigma^2$ and $\rho = \text{corr}(\hat{\theta}, \hat{\tau})$. Note that v_θ , v_τ and ρ are known. Let $\gamma = \tau/(\sigma v_\tau^{1/2})$, which is an unknown parameter, and also let $\hat{\gamma} = \hat{\tau}/(\sigma v_\tau^{1/2})$. Suppose that we carry out a preliminary test of the null hypothesis $\tau = 0$ against the alternative hypothesis $\tau \neq 0$. For given value of the positive number d , suppose that we accept this null hypothesis when $|\hat{\gamma}| \leq d$; otherwise we reject this null hypothesis. Let $\tilde{\alpha}$ denote the

size of this preliminary test. If $|\hat{\gamma}| \leq d$ we choose model \mathcal{M}_1 ; otherwise we choose model \mathcal{M}_2 .

Let $k(\gamma) = \phi(d + \gamma) - \phi(d - \gamma) + \gamma(\Phi(d - \gamma) - \Phi(-d - \gamma))$, where ϕ and Φ denote the $N(0, 1)$ pdf and cdf. Kabaila & Wijethunga (2019a) prove that the ideal (i.e. in the limit as the number of bootstrap replications diverges to infinity) bootstrap smoothed estimator of θ is

$$\tilde{\theta} = \hat{\theta} - \rho \sigma v_{\theta}^{1/2} k(\hat{\gamma}).$$

Efron (2014) derived a delta method approximation to the standard deviation of the bootstrap smoothed estimator. Kabaila & Wijethunga (2019a) show that, in the limit as the number of bootstrap replications diverges to infinity, this approximation is given by $\sigma v_{\theta}^{1/2} r_{\text{delta}}(\gamma)$, where $r_{\text{delta}}(\gamma) = (1 - 2\rho^2 q(\gamma) + \rho^2 q^2(\gamma))^{1/2}$, with $q(\gamma) = \Phi(d - \gamma) - \Phi(-d - \gamma) - d(\phi(d + \gamma) + \phi(d - \gamma))$. Let $z(\alpha) = \Phi^{-1}(1 - \alpha/2)$. The confidence interval for θ , with nominal coverage $1 - \alpha$, centred on $\tilde{\theta}$ and with half-width equal to $z(\alpha)$ multiplied by an estimate of this approximation to the standard deviation is

$$\left[\hat{\theta} - \sigma v_{\theta}^{1/2} \rho k(\hat{\gamma}) \pm \sigma v_{\theta}^{1/2} (z(\alpha) r_{\text{delta}}(\hat{\gamma})) \right].$$

The function $r_{\text{delta}} : \mathbb{R} \rightarrow [0, \infty)$ is a continuous even function and $r_{\text{delta}}(x) \rightarrow 1$ as $x \rightarrow \infty$. As shown in the Supplementary Material, to an excellent approximation, $r_{\text{delta}}(x) = 1$ for all $x \geq c$, where $c = 10$.

We consider confidence intervals of the form

$$\text{CI}(s) = \left[\hat{\theta} - \rho \sigma v_{\theta}^{1/2} k(\hat{\gamma}) \pm \sigma v_{\theta}^{1/2} s(\hat{\gamma}) \right],$$

where $s : \mathbb{R} \rightarrow [0, \infty)$ is an even function. To apply the performance bound of Kabaila & Kong (2016), we suppose that $s(x) = z(\alpha)$ for all $x \geq c$. Let \mathcal{D} denote the class of functions s that satisfy this property. We do not require that the function s is continuous. In other words, a confidence interval $\text{CI}(s)$, where $s \in \mathcal{D}$, does not necessarily possess the attractive feature (A1). In addition, such a confidence interval does not necessarily possess the attractive feature (A2). This is because $\hat{\gamma} \sim N(0, 1)$ under the simpler model, so that the data and this model can be said to be highly discordant when $|\hat{\gamma}| > 4$, say.

We now introduce the following question:

Is there a confidence interval $\text{CI}(s)$, specified by a function s in \mathcal{D} , such that this confidence interval has minimum coverage probability $1 - \alpha$ and scaled expected length that (a) has a maximum value that is not too much larger than 1 and (b) is substantially less than 1 when the simpler model is correct?

If the answer to this question is ‘no’ then the answer to the question posed in the introduction must also be ‘no’.

3. Exact formulas for the coverage probability and scaled expected length of $\text{CI}(s)$

For notational convenience, let $b(x) = \rho k(x)$ for all $x \in \mathbb{R}$, so that

$$\text{CI}(s) = \left[\hat{\theta} - \sigma v_{\hat{\theta}}^{1/2} b(\hat{\gamma}) \pm \sigma v_{\hat{\theta}}^{1/2} s(\hat{\gamma}) \right].$$

Every $s \in \mathcal{D}$ is an even function that satisfies $s(x) = z(\alpha)$ for all $x \geq c$. Hence every $s \in \mathcal{D}$ is specified by s restricted to the domain $[0, c]$. In this section we present exact formulas for the coverage probability and scaled expected length of $\text{CI}(s)$ in terms of the function s restricted to the domain $[0, c]$. The function k is a continuous odd function and $k(x) \rightarrow 0$ as $x \rightarrow \infty$. As shown in the Supplementary Material, to an excellent approximation, $k(x) = 0$ for all $x \geq c$, where $c = 10$. For computational convenience, we approximate $k(x)$, and therefore $b(x)$, by 0 for all $x \geq c$.

The following theorem, proved in Appendix A.1, provides an exact formula for the coverage probability of the confidence interval $\text{CI}(s)$ in terms of the function s restricted to the domain $[0, c]$.

Theorem 1. *For any given ρ and function $s \in \mathcal{D}$, the coverage probability of $\text{CI}(s)$ is a function of γ . We denote this coverage probability by $c(\gamma; s, \rho)$. For any given ρ and $s \in \mathcal{D}$, $c(\gamma; s, \rho)$ is an even function of γ . Also, for any given γ and $s \in \mathcal{D}$, $c(\gamma; s, \rho)$ is an even function of ρ . Let $\ell(h, \gamma; x) = P(b(h) - x \leq \tilde{G} \leq b(h) + x)$ and $\ell^\dagger(h, \gamma) = P(-z(\alpha) \leq \tilde{G} \leq z(\alpha))$, for $\tilde{G} \sim N(\rho(h - \gamma), 1 - \rho^2)$. Now let $R_1(s, \gamma)$ denote*

$$\int_0^c \left((\ell^\dagger(h, \gamma) - \ell(h, \gamma; s(h))) \phi(h - \gamma) + (\ell^\dagger(-h, \gamma) - \ell(-h, \gamma; s(h))) \phi(h + \gamma) \right) dh.$$

Then $c(\gamma; s, \rho) = 1 - \alpha - R_1(s, \gamma)$.

It follows from this theorem that, for any given $s \in \mathcal{D}$, $R_1(s, \gamma)$ is an even function of γ . It also follows from this theorem that, for any given γ and $s \in \mathcal{D}$, $R_1(s, \gamma)$ is an even function of ρ .

Let I denote the usual $1 - \alpha$ confidence interval for θ based on the full model. In other words,

$$I = \left[\widehat{\theta} \pm z(\alpha) \sigma v_{\widehat{\theta}}^{1/2} \right].$$

Define the scaled expected length of $\text{CI}(s)$ to be

$$\frac{E(\text{length of CI}(s))}{E(\text{length of } I)} = \frac{E(s(\widehat{\gamma}))}{z(\alpha)}.$$

It follows from (7) that this is a function of γ for given function s . We denote this function by $e(\gamma; s)$. The following theorem, proved in Appendix A.2, provides an exact formula for $e(\gamma; s)$ in terms of the function s restricted to the domain $[0, c]$.

Theorem 2. *For given function s , the scaled expected length of $\text{CI}(s)$ is a function of γ . We denote this scaled expected length by $e(\gamma; s)$. Then $e(\gamma; s) = 1 + R(s, \gamma)$, where*

$$R(s, \gamma) = \int_0^c \left(\frac{s(h)}{z(\alpha)} - 1 \right) (\phi(h - \gamma) + \phi(h + \gamma)) dh. \quad (1)$$

Obviously $e(\gamma; s)$ and $R(s, \gamma)$ are even functions of γ .

The function $R_1(s, \gamma)$ is the probability of non-coverage of θ by $\text{CI}(s)$, for true parameter value γ . The function $R(s, \gamma)$ is the scaled expected length of $\text{CI}(s)$ minus 1, for true parameter value γ . Both $R_1(s, \gamma)$ and $R(s, \gamma)$ are risk functions in the following sense. If both $R_1(s_1, \gamma) < R_1(s_2, \gamma)$ and $R(s_1, \gamma) < R(s_2, \gamma)$ then $\text{CI}(s_1)$ is preferred to $\text{CI}(s_2)$, for true parameter value γ (cf. Kempthorne, 1987, p.172).

4. Application of the performance bound of Kabaila & Kong (2016)

A direct answer to the question raised at the end of Section 2 is provided by finding $\inf_{s \in \mathcal{D}} e(0; s)$, subject to the following two constraints:

(C1) Coverage constraint

The coverage constraint is $c(\gamma; s, \rho) \geq 1 - \alpha$ for all $\gamma \geq 0$. We have used here the fact that $c(\gamma; s, \rho)$ is an even function of γ .

(C2) Maximum Scaled Expected Length constraint

For given $u > 0$, $e(\gamma; s) \leq 1 + u$ for all $\gamma \geq 0$. We have used here the fact that $e(\gamma; s)$ is an even function of γ .

We answer the question raised at the end of Section 2 indirectly by finding a lower bound on $\inf_{s \in \mathcal{D}} e(0; s)$, subject to these two constraints. If this lower bound is greater than 1 then the answer to this question is no.

To apply the performance bound of Kabaila & Kong (2016), we need to express $e(0; s)$ and the constraints (C1) and (C2) in terms of risks and integrated risks. Since $e(\gamma; s) = 1 + R(s, \gamma)$,

$$\text{Subject to the constraints (C1) and (C2), } \inf_{s \in \mathcal{D}} e(0; s) = 1 + \inf_{s \in \mathcal{D}} R(s, 0). \quad (2)$$

Note that

$$R(s, 0) = 2 \int_0^c \left(\frac{s(h)}{z(\alpha)} - 1 \right) \phi(h) dh = \int_{-\infty}^{\infty} R(s, \gamma) d\pi(\gamma),$$

where π is the cumulative distribution function of the distribution with a unit probability point mass at 0, i.e. $\pi(x) = 1$ for $x \geq 0$ and $\pi(x) = 0$ for $x < 0$. Thus $R(s, 0)$ is an integrated risk. To bring the notation in line with that used by Kabaila & Kong (2016), let $R_2(s, \gamma) \equiv R(s, \gamma)$. We now express the two constraints as follows.

(C1) Coverage constraint

The coverage constraint is $R_1(s, \gamma) \leq 0$ for all $\gamma \geq 0$.

(C2) Maximum Scaled Expected Length constraint

For given $u > 0$, $R_2(s, \gamma) \leq u$ for all $\gamma \geq 0$.

We now apply Theorem 2(a) and the method described in Appendix C of Kabaila & Kong (2016). Let m_1 and m_2 be given positive integers. Suppose that $\gamma_1(1), \dots, \gamma_1(m_1)$ and $\gamma_2(1), \dots, \gamma_2(m_2)$ satisfy

$$\begin{aligned} 0 &\leq \gamma_1(1) < \gamma_1(2) < \dots < \gamma_1(m_1) \\ 0 &< \gamma_2(1) < \gamma_2(2) < \dots < \gamma_2(m_2). \end{aligned} \quad (3)$$

Introduce the nonnegative variables $\nu_1(1), \dots, \nu_1(m_1)$ and $\nu_2(1), \dots, \nu_2(m_2)$ and let

$$\begin{aligned}\boldsymbol{\gamma} &= (\gamma_1(1), \dots, \gamma_1(m_1), \gamma_2(1), \dots, \gamma_2(m_2)) \\ \boldsymbol{\nu} &= (\nu_1(1), \dots, \nu_1(m_1), \nu_2(1), \dots, \nu_2(m_2)).\end{aligned}$$

A given value of $(\boldsymbol{\gamma}, \boldsymbol{\nu})$ corresponds to the following two prior distributions:

- (1) A discrete prior distribution that consists of m_1 probability point masses $p_1(1), \dots, p_1(m_1)$ located at $\gamma_1(1), \dots, \gamma_1(m_1)$ respectively, where

$$p_1(j) = \frac{\nu_1(j)}{\sum_{k=1}^{m_1} \nu_1(k)} \quad (j = 1, \dots, m_1).$$

- (2) A discrete prior distribution that consists of m_2 probability point masses $p_2(1), \dots, p_2(m_2)$ located at $\gamma_2(1), \dots, \gamma_2(m_2)$ respectively, where

$$p_2(j) = \frac{\nu_2(j)}{\sum_{k=1}^{m_2} \nu_2(k)} \quad (j = 1, \dots, m_2).$$

Let

$$\tilde{g}(s, \boldsymbol{\gamma}, \boldsymbol{\nu}) = \int_{-\infty}^{\infty} R(s, \gamma) d\pi(\gamma) + \sum_{j=1}^{m_1} \nu_1(j) R_1(s, \gamma_1(j)) + \sum_{j=1}^{m_2} \nu_2(j) R_2(s, \gamma_2(j)).$$

Now let $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$ denote a value of $s \in \mathcal{D}$ that minimizes $\tilde{g}(s, \boldsymbol{\gamma}, \boldsymbol{\nu})$. It follows from Theorem 2(a) and the method described in Appendix C of Kabaila & Kong (2016) that a lower bound for $\inf_{s \in \mathcal{D}} R(s, 0)$, subject to the constraints (C1) and (C2), is

$$\tilde{g}(s(\boldsymbol{\gamma}, \boldsymbol{\nu}), \boldsymbol{\gamma}, \boldsymbol{\nu}) - \sum_{j=1}^{m_2} \nu_2(j) u, \quad (4)$$

where u is the given positive number in the description of the constraint (C2). As proposed by Kabaila & Kong (2016), the lower bound (4) is numerically maximized with respect to $(\boldsymbol{\gamma}, \boldsymbol{\nu})$, to tighten this lower bound. This amounts to computing two unfavorable discrete prior distributions. We then use (2) to obtain the desired lower bound.

The crucial part of this procedure is the computation of $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$, a value of $s \in \mathcal{D}$ that minimizes $\tilde{g}(s, \boldsymbol{\gamma}, \boldsymbol{\nu})$, for given $(\boldsymbol{\gamma}, \boldsymbol{\nu})$. This computation, which requires some care, is described in the next section.

5. Computation of $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$ for given $(\boldsymbol{\gamma}, \boldsymbol{\nu})$

Throughout this section we suppose that $(\boldsymbol{\gamma}, \boldsymbol{\nu})$ is given. We describe the computation of $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$, a value of $s \in \mathcal{D}$ that minimizes $\tilde{g}(s, \boldsymbol{\gamma}, \boldsymbol{\nu})$. Straightforward manipulations show that

$$\tilde{g}(s, \boldsymbol{\gamma}, \boldsymbol{\nu}) = \int_0^c q(s(h); h, \boldsymbol{\gamma}, \boldsymbol{\nu}) dh, \quad (5)$$

where $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ is defined to be

$$\begin{aligned} & \left(\frac{x}{z(\boldsymbol{\alpha})} - 1 \right) \left(2\phi(h) + \sum_{j=1}^{m_2} \nu_2(j) (\phi(h - \gamma_2(j)) + \phi(h + \gamma_2(j))) \right) \\ & + \sum_{j=1}^{m_1} \nu_1(j) \left[(\ell^\dagger(h, \gamma_1(j)) - \ell(h, \gamma_1(j); x)) \phi(h - \gamma_1(j)) \right. \\ & \quad \left. + (\ell^\dagger(-h, \gamma_1(j)) - \ell(-h, \gamma_1(j); x)) \phi(h + \gamma_1(j)) \right]. \end{aligned}$$

It follows from (5) that a function $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$, defined as a minimizer of $\tilde{g}(s, \boldsymbol{\gamma}, \boldsymbol{\nu})$ over $s \in \mathcal{D}$, may be found as follows. We set $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$, evaluated at any $h \in [0, c]$, to be a minimizer over $x \in [0, \infty)$ of $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$.

Now $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ is a continuous function of $x \in [0, \infty)$ for all $h \in [0, c]$ and every given $(\boldsymbol{\gamma}, \boldsymbol{\nu})$. An examination of some examples of this function of $x \in [0, \infty)$ show that this function may have several local minima, including the possibility of a local minimum at $x = 0$. Consequently, the value of $x \in [0, \infty)$ that minimizes $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ may change discontinuously, as h increases. In other words, the function $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$ of h may have discontinuities. Figure 5 of the Supplementary Material provides some illustrations of functions $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ of $x \in [0, \infty)$ that have two local minima. Figure 4 of the Supplementary Material provides an illustration of a function $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$ of h with discontinuities.

To evaluate the lower bound (4), we need to evaluate

$$\tilde{g}(s(\boldsymbol{\gamma}, \boldsymbol{\nu}), \boldsymbol{\gamma}, \boldsymbol{\nu}) = \int_0^c q\left(s(\boldsymbol{\gamma}, \boldsymbol{\nu}) \text{ evaluated at } h; h, \boldsymbol{\gamma}, \boldsymbol{\nu}\right) dh. \quad (6)$$

Although the function $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$ of h may have discontinuities, the integrand of the integral on the right-hand side of (6) is a continuous function of $h \in [0, c]$. An illustration of this, when the function $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$ of h has discontinuities, is provided by Figure 3 of the Supplementary Material.

To carry out the computation of the function $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$ accurately and effectively, we use the properties of $dq(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx$, considered as a function of x , described in Appendix A.3. Suppose that $h \in [0, c]$ is given. Theorem 3 of Appendix A.3 leads to the procedure described at the end of this appendix for finding an interval $[0, \tilde{x}]$ that must contain a value of $x \geq 0$ that minimizes $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$.

We use the following two step procedure to find the value of $x \in [0, \tilde{x}]$ that minimizes $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$. We find all possible local minima in Step 1 and compare them to find the global minimum in Step 2.

Step 1: By considering $dq(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx$, find all the local minimizers of $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ in the interval $[0, \tilde{x}]$. Define w to be the smallest integer that is greater than or equal to $10\tilde{x}$. We evaluate $dq(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx$ on the evenly-spaced grid $x_1 = 0, x_2 = 0.1, x_3 = 0.2, \dots, x_w$ of values of x . To find the values of $x \in [0, x_w]$ that are local minimizers of $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$, we need to consider the following two cases.

Case 1: $x = 0$

$x = 0$ is a local minimizer of $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ if either $dq(0; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx > 0$ or $dq(0; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx = 0$ and $dq(x_2; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx > 0$; otherwise $x = 0$ is not a local minimizer.

Case 2: $0 < x < x_w$

If $dq(x_i; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx < 0$ and $dq(x_{i+1}; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx > 0$, then $dq(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx$ has a zero in the interval $[x_i, x_{i+1}]$ that is a local minimizer of $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$. We find this zero using the R function `uniroot`, to which we provide the interval $[x_i, x_{i+1}]$. Also, if $dq(x_i; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx = 0$ and $dq(x_{i-1}; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx < 0$ and $dq(x_{i+1}; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx > 0$ then x_i is a zero of $dq(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx$ that is a local minimizer of $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$.

Step 2: Evaluate $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ at the local minimizers of $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ found in Step 1. The global minimum of $q(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ is simply the minimum of the local minima.

6. Numerical results

As noted in Section 3, for any given γ and $s \in \mathcal{D}$, $R_1(s, \gamma)$ is an even function of ρ . Since $R(s, \gamma)$ does not depend on ρ , it is therefore sufficient to consider $\rho \in [0, 1)$. For any given size $\tilde{\alpha}$ of the preliminary test, desired minimum coverage $1 - \alpha$ and $\rho \in [0, 1)$, we

proceed as follows.

Let $Q(m_1, m_2)$ denote the set of possible values of $(\boldsymbol{\gamma}, \boldsymbol{\nu})$, for positive integers m_1 and m_2 . Also let $\text{LB}(u; m_1, m_2, \boldsymbol{\gamma}, \boldsymbol{\nu})$ denote the lower bound (4). We will make use of the following easily-proved, but very useful result.

Theorem 3. *For any given positive integers m_1 and m_2 and $(\boldsymbol{\gamma}, \boldsymbol{\nu}) \in Q(m_1, m_2)$ such that $\sum_{j=1}^{m_2} \nu_2(j) > 0$ the following result is true. The lower bound $\text{LB}(u; m_1, m_2, \boldsymbol{\gamma}, \boldsymbol{\nu})$ is a decreasing function of $u \in (0, \infty)$.*

Proof. Suppose that the positive integers m_1 and m_2 and $(\boldsymbol{\gamma}, \boldsymbol{\nu}) \in Q(m_1, m_2)$, satisfying $\sum_{j=1}^{m_2} \nu_2(j) > 0$, are given. The function $s(\boldsymbol{\gamma}, \boldsymbol{\nu})$ does not depend on $u > 0$. The result now follows from the expression (4). \square

Using the procedure described in Section S3 of the Supplementary Material, we find ‘good values’ of the positive integers m_1^* and m_2^* , $u^* > 0$ and $(\boldsymbol{\gamma}^*, \boldsymbol{\nu}^*) \in Q(m_1^*, m_2^*)$. Now define u^{**} to be the solution for u of

$$1 + \tilde{g}(s(\boldsymbol{\gamma}^*, \boldsymbol{\nu}^*), \boldsymbol{\gamma}^*, \boldsymbol{\nu}^*) - \sum_{j=1}^{m_2} \nu_2^*(j) u = 1.005.$$

In other words,

$$u^{**} = \frac{\tilde{g}(s(\boldsymbol{\gamma}^*, \boldsymbol{\nu}^*), \boldsymbol{\gamma}^*, \boldsymbol{\nu}^*) - 0.005}{\sum_{j=1}^{m_2} \nu_2^*(j)}.$$

If $u^{**} > 0$ then Theorem 3 implies that the answer to the question stated at the end of Section 2 is ‘no’ for all u satisfying $0 < u \leq u^{**}$. Consequently, if $u^{**} > 0$ then for all u satisfying $0 < u \leq u^{**}$ the answer to the question stated in the introduction is ‘no’.

Section S2 of the Supplementary Material describes the details of computations using the R programming language. In Section S4 of the Supplementary Material we provide a method to check the accuracy of the final computed results and describe this procedure using an example. Table 1 describes the values of u^{**} computed using the procedure described in Section S3 of the Supplementary Material, with $\epsilon = 0.05$. All of the results are for nominal coverage $1 - \alpha = 0.95$, $\tilde{\alpha} \in \{0.05, 0.1\}$ and $|\rho| \in \{0.5, 0.6, 0.7, 0.8\}$. This table also lists the values of m_1 and m_2 , the numbers of probability point masses in the discrete prior distributions (described in Section 4) that relate to the coverage probability and scaled expected length, respectively.

Table 1. Values of u^{**} such that $\text{LB}(u; m_1, m_2, \gamma, \nu) > 1$ for all u satisfying $0 < u \leq u^{**}$. These values were computed using the procedure described in Section S3 of the Supplementary Material, with $\epsilon = 0.05$. We consider $\tilde{\alpha} \in \{0.05, 0.1\}$, $|\rho| \in \{0.5, 0.6, 0.7, 0.8\}$ and nominal coverage $1 - \alpha = 0.95$. Here m_1 and m_2 are the number of probability point masses in the two prior distributions described in Section 4.

$\tilde{\alpha}$	$ \rho $	m_1	m_2	u^{**}
0.05	0.5	4	4	0.05268182
	0.6	4	4	0.07535683
	0.7	5	3	0.11375010
	0.8	5	3	0.15037320
0.1	0.5	4	2	0.02763119
	0.6	7	4	0.04430236
	0.7	5	2	0.06345934
	0.8	5	2	0.07750792

In view of the question that we have posed in the introduction and at the end of Section 2, it is reasonable to measure the performance of a confidence interval $\text{CI}(s)$ that satisfies constraints (C1) and (C2) as follows. Define the squared scaled expected length gain (*gain*) to be

$$1 - (e(0; s))^2$$

and the squared scaled expected length loss (*loss*) to be

$$(1 + u)^2 - 1 = u^2 + 2u.$$

Ideally, the *gain* is large and the *loss* is small. If the positive number ℓ is a lower bound on $\inf_{s \in \mathcal{D}} e(0; s)$, subject to the constraints (C1) and (C2), then $1 - \ell^2$ is an upper bound on the *gain* $1 - (e(0; s))^2$ for all $s \in \mathcal{D}$ such that constraints (C1) and (C2) are satisfied. For the values of u listed in Table 2, we used the values of m_1 and m_2 in Table 1 that were used to compute u^{**} , to carry out Step 1 of the procedure described in Section S3 of

the Supplementary Material. Table 2 gives the resulting upper bound on the *gain* for *loss* specified by the given value of $u > 0$. As can be seen from this table, the ratio (upper bound on *gain*)/*loss* is small for all $|\rho| \geq 0.6$.

Table 2. Computed upper bound on the *gain* for *loss* specified by the given value of $u > 0$. These values were computed using the procedure described in Section S3 of the Supplementary Material, with $\epsilon = 0.05$. We consider $\tilde{\alpha} \in \{0.05, 0.1\}$, $|\rho| \in \{0.5, 0.6, 0.7, 0.8\}$ and nominal coverage $1 - \alpha = 0.95$.

$\tilde{\alpha}$	$ \rho $	u	upper bound on <i>gain</i>	<i>loss</i>	(upper bound on <i>gain</i>)/ <i>loss</i>
0.05	0.5	0.079	0.04948720	0.1642	0.3014
		0.105	0.07126956	0.2210	0.3225
	0.6	0.113	0.02959263	0.2387	0.1239
		0.151	0.05424241	0.3248	0.1670
	0.7	0.171	0.02867069	0.3712	0.0772
		0.228	0.05785821	0.5079	0.1139
	0.8	0.226	0.03730076	0.5031	0.0741
		0.301	0.08313421	0.6926	0.1200
0.1	0.5	0.041	0.02654781	0.0837	0.3172
		0.055	0.05058620	0.1130	0.4477
	0.6	0.066	0.02488160	0.1364	0.1824
		0.089	0.05302328	0.1859	0.2852
	0.7	0.095	0.02606301	0.1990	0.1309
		0.127	0.05599555	0.2701	0.2073
	0.8	0.117	0.03431617	0.2477	0.1385
		0.156	0.05958476	0.3363	0.1772

For given nominal coverage $1 - \alpha$, size $\tilde{\alpha}$ of the preliminary test, $|\rho|$, u , m_1 and m_2 ,

it takes roughly 45 minutes to compute the lower bound on $\inf_{s \in \mathcal{D}} e(0; s)$, subject to the constraints (C1) and (C2), using a computer with i7 processor (3.4 GHz) and 32 GB of RAM. In other words, the time taken to complete Step 1 of the procedure described in Section S3 of the Supplementary Material is roughly 45 minutes. As a consequence, the time taken to complete the entire algorithm described in Section S3 of the Supplementary Material for given nominal coverage $1 - \alpha$, size $\tilde{\alpha}$ of the preliminary test, $|\rho|$ and u is about 4 hours. The subsequent computation of u^{**} for the given values of nominal coverage $1 - \alpha$, size $\tilde{\alpha}$ of the preliminary test and $|\rho|$, takes an additional minute or so.

7. Remarks

Remark 7.1

It is reasonable to ask whether or not we have, in posing the question described in the introduction, put forward requirements that are excessively restrictive in the sense that these requirements cannot be achieved by any confidence interval whatsoever. In other words, is it possible that, for the testbed scenario of two nested linear regression models and known error variance σ^2 , there does not exist any confidence interval (whose centre is not necessarily the bootstrap smoothed estimator) for which the answer to the question posed in the introduction is ‘yes’?

We know that the requirements that we have put forward are not excessively restrictive because, as shown by Kabaila & Giri (2009), Kabaila & Giri (2013) and Mainzer & Kabaila (2019), it is possible to compute formulas for the centre and width of the confidence interval so that this interval has the attractive features (A1) and (A2), the desired minimum coverage probability $1 - \alpha$ and scaled expected length that (a) has a maximum value that is not too much larger than 1 and (b) is substantially less than 1 when the simpler model is correct. Indeed, the R package `ciuupi`, described by Mainzer & Kabaila (2019), computes confidence intervals that have the attractive features (A1) and (A2), the desired minimum coverage probability $1 - \alpha$ and for which the *gain* is set equal to the *loss*, where *gain* and *loss* are as defined in Section 6.

Remark 7.2

In the present paper we have considered the testbed scenario of two nested linear regression models for known error variance σ^2 . A natural question to ask is the following. Do we obtain similar results when, instead, σ^2 is unknown, so that it must be estimated from the data?

It is highly plausible that, for two nested linear regression models with p fixed, the following is true. The results obtained for the case that the error variance σ^2 is known provide an excellent approximation to the corresponding results obtained for the case that σ^2 is unknown and the residual degrees of freedom $n - p$ is moderately large. Results that support this claim are provided in Appendix B of Kabaila & Wijethunga (2019a) and in Section 3 of Kabaila & Wijethunga (2019b), who consider the case that σ^2 is unknown. Therefore the results of the present paper suggest that for two nested linear regression models and σ^2 unknown the answer to the question posed in the introduction will be ‘no’, provided that $n - p$ is sufficiently large. However, by construction, the results of Kabaila & Wijethunga (2019b) suggest that the answer to this question will be ‘yes’, when $n - p$ is small.

8. Conclusion

In Section 6 we consider a confidence interval centred on the bootstrap smoothed estimator, for preliminary test size either 0.05 or 0.1. This confidence interval is required to have (C1) coverage probability that never falls below 0.95 and (C2) scaled expected length that never exceeds $1 + u$, for given $u > 0$. Table 1 gives values of u^{**} such that for all u satisfying $0 < u \leq u^{**}$, the scaled expected length of this confidence interval must exceed 1 when the simpler model is correct. In other words, this table specifies values of $u > 0$ such that it is impossible to find a formula for the width of this confidence interval such that its scaled expected length is less than 1 when the simpler model is correct.

In Section 6 we also define a *gain* and *loss* for this confidence interval. Table 1 may be viewed as giving values of the *loss* such that it is impossible for this confidence interval to have any *gain*. As Table 2 shows, even for the listed values of $u > u^{**}$, the *gain* cannot be more than a small fraction of the *loss* when $|\rho| \geq 0.6$.

Acknowledgements

This work was supported by an Australian Government Research Training Program Scholarship.

References

- Blyth, C.R. 1951. On minimax statistical decision procedures and their admissibility. *The Annals of Mathematical Statistics*, **22**, 22–42.
- Breiman, L. 1996. Bagging predictors. *Machine Learning*, **24**, 123–140.
- Efron, B. 2014. Estimation and accuracy after model selection. *Journal of the American Statistical Association*, **109**, 991–1007.
- Giri, K. 2008. *Confidence intervals in regression utilizing prior information*. Ph.D. thesis, Department of Mathematics and Statistics, La Trobe University.
- Hall, P. 1992. *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
- Hodges, J.L., & Lehmann, E.L. 1952. The use of previous experience in reaching statistical decisions. *Annals of Mathematical Statistics*, **23**, 396–407.
- Kabaila, P., & Giri, K. 2009. Confidence intervals in regression utilizing prior information. *Journal of Statistical Planning and Inference*, **139**, 3419–3429.
- Kabaila, P., & Giri, K. 2013. Further properties of frequentist confidence intervals in regression that utilize uncertain prior information. *Australian & New Zealand Journal of Statistics*, **55**, 259–270.
- Kabaila, P., & Kong, Y. 2016. Lower bounds on integrated risk, subject to inequality constraints. *Australian & New Zealand Journal of Statistics*, **58**, 293–315.
- Kabaila, P., & Tuck, J. 2008. Confidence intervals utilizing prior information in the Behrens-Fisher problem. *Australian & New Zealand Journal of Statistics*, **50**, 309–328.

- Kabaila, P., & Wijethunga, C. 2019a. Confidence intervals centred on bootstrap smoothed estimators. *Australian & New Zealand Journal of Statistics*, **61**, 19–38.
- Kabaila, P., & Wijethunga, C. 2019b. On confidence intervals centred on bootstrap smoothed estimators. *Stat*, **8**, e233.
- Kempthorne, P.J. 1983. Minimax-Bayes compromise estimators. *Proc. Bus. and Econ. Statist. Sec. Amer. Statist. Assoc.*, 563–573.
- Kempthorne, P.J. 1987. Numerical specification of discrete least favorable 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. *The Annals of Statistics*, **16**, 1594–1608.
- Mainzer, R., & Kabaila, P. 2019. ciuupi: An R package for Computing Confidence Intervals that Utilize Uncertain Prior Information. *R Journal*, **11**, 323–336.

Appendix

We will express all quantities of interest in terms of the random vector $(\hat{\theta}, \hat{\gamma})$, which has the following bivariate normal distribution:

$$\begin{bmatrix} \hat{\theta} \\ \hat{\gamma} \end{bmatrix} \sim N \left(\begin{bmatrix} \theta \\ \gamma \end{bmatrix}, \begin{bmatrix} \sigma^2 v_\theta & \rho \sigma v_\theta^{1/2} \\ \rho \sigma v_\theta^{1/2} & 1 \end{bmatrix} \right). \quad (7)$$

A.1 Proof of Theorem 1

The following proof is based, in part, on the derivations described in Section 4.3 of Giri (2008). The coverage probability of the confidence interval $\text{CI}(s)$ is

$$\begin{aligned} P(\theta \in \text{CI}(s)) &= P\left(\hat{\theta} - \sigma v_\theta^{1/2} b(\hat{\gamma}) - \sigma v_\theta^{1/2} s(\hat{\gamma}) \leq \theta \leq \hat{\theta} - \sigma v_\theta^{1/2} b(\hat{\gamma}) + \sigma v_\theta^{1/2} s(\hat{\gamma})\right) \\ &= P\left(-\hat{\theta} + \sigma v_\theta^{1/2} b(\hat{\gamma}) + \sigma v_\theta^{1/2} s(\hat{\gamma}) \geq -\theta \geq -\hat{\theta} + \sigma v_\theta^{1/2} b(\hat{\gamma}) - \sigma v_\theta^{1/2} s(\hat{\gamma})\right) \\ &= P\left(\sigma v_\theta^{1/2} (b(\hat{\gamma}) - s(\hat{\gamma})) \leq \hat{\theta} - \theta \leq \sigma v_\theta^{1/2} (b(\hat{\gamma}) + s(\hat{\gamma}))\right) \\ &= P\left(b(\hat{\gamma}) - s(\hat{\gamma}) \leq \frac{\hat{\theta} - \theta}{\sigma v_\theta^{1/2}} \leq b(\hat{\gamma}) + s(\hat{\gamma})\right) \\ &= P(b(\hat{\gamma}) - s(\hat{\gamma}) \leq G \leq b(\hat{\gamma}) + s(\hat{\gamma})), \end{aligned}$$

where $G = (\hat{\theta} - \theta)/(\sigma v_{\theta}^{1/2})$. It follows from (7) that

$$\begin{bmatrix} G \\ \hat{\gamma} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ \gamma \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right). \quad (8)$$

For given ρ and function s , the coverage probability of $\text{CI}(s)$ is a function of γ . We denote this coverage probability by $c(\gamma; s, \rho)$.

Since b and s are odd and even functions, respectively,

$$\begin{aligned} c(\gamma; s, \rho) &= P(-b(-\hat{\gamma}) - s(-\hat{\gamma}) \leq G \leq -b(-\hat{\gamma}) + s(-\hat{\gamma})) \\ &= P(b(-\hat{\gamma}) - s(-\hat{\gamma}) \leq -G \leq b(-\hat{\gamma}) + s(-\hat{\gamma})) \\ &= P(b(\hat{\gamma}') - s(\hat{\gamma}') \leq G' \leq b(\hat{\gamma}') + s(\hat{\gamma}')), \end{aligned}$$

where $G' = -G$ and $\hat{\gamma}' = -\hat{\gamma}$. It follows from (8) that

$$\begin{bmatrix} G' \\ \hat{\gamma}' \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ -\gamma \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

Hence $c(\gamma; s, \rho) = c(-\gamma; s, \rho)$.

Since $b(x) = \rho k(x)$,

$$\begin{aligned} c(\gamma; s, \rho) &= P(\rho k(\hat{\gamma}) - s(\hat{\gamma}) \leq G \leq \rho k(\hat{\gamma}) + s(\hat{\gamma})) \\ &= P(-\rho k(\hat{\gamma}) - s(\hat{\gamma}) \leq -G \leq -\rho k(\hat{\gamma}) + s(\hat{\gamma})) \\ &= P((-\rho)k(\hat{\gamma}) - s(\hat{\gamma}) \leq G' \leq (-\rho)k(\hat{\gamma}) + s(\hat{\gamma})), \end{aligned}$$

where $G' = -G$. It follows from (8) that

$$\begin{bmatrix} G' \\ \hat{\gamma} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ \gamma \end{bmatrix}, \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \right).$$

Hence $c(\gamma; s, \rho) = c(\gamma; s, -\rho)$.

It follows from (8) that the probability distribution of G , conditional on $\hat{\gamma} = h$, is $N(\rho(h - \gamma), 1 - \rho^2)$. Note that

$$\begin{aligned} &P(b(\hat{\gamma}) - s(\hat{\gamma}) \leq G \leq b(\hat{\gamma}) + s(\hat{\gamma})) \\ &= \int_{-\infty}^{\infty} P(b(h) - s(h) \leq G \leq b(h) + s(h) | \hat{\gamma} = h) \phi(h - \gamma) dh \\ &= \int_{-\infty}^{\infty} P(b(h) - s(h) \leq \tilde{G} \leq b(h) + s(h)) \phi(h - \gamma) dh, \end{aligned}$$

where $\tilde{G} \sim N(\rho(h - \gamma), 1 - \rho^2)$. Thus

$$\begin{aligned}
c(\gamma; s, \rho) &= \int_{-\infty}^{\infty} \ell(h, \gamma; s(h)) \phi(h - \gamma) dh \\
&= \int_{-\infty}^{-c} \ell(h, \gamma; s(h)) \phi(h - \gamma) dh + \int_{-c}^c \ell(h, \gamma; s(h)) \phi(h - \gamma) dh + \int_c^{\infty} \ell(h, \gamma; s(h)) \phi(h - \gamma) dh \\
&= \int_{-\infty}^{-c} \ell^\dagger(h, \gamma) \phi(h - \gamma) dh + \int_{-c}^c \ell(h, \gamma; s(h)) \phi(h - \gamma) dh + \int_c^{\infty} \ell^\dagger(h, \gamma) \phi(h - \gamma) dh.
\end{aligned} \tag{9}$$

The usual $1 - \alpha$ confidence interval based on the full model \mathcal{M}_2 has coverage probability $1 - \alpha$. Thus

$$\begin{aligned}
1 - \alpha &= P(-z(\alpha) \leq G \leq z(\alpha)) \\
&= \int_{-\infty}^{\infty} P(-z(\alpha) \leq G \leq z(\alpha) \mid \hat{\gamma} = h) \phi(h - \gamma) dh \\
&= \int_{-\infty}^{\infty} \ell^\dagger(h, \gamma) \phi(h - \gamma) dh.
\end{aligned}$$

Therefore

$$1 - \alpha = \int_{-\infty}^{-c} \ell^\dagger(h, \gamma) \phi(h - \gamma) dh + \int_{-c}^c \ell^\dagger(h, \gamma) \phi(h - \gamma) dh + \int_c^{\infty} \ell^\dagger(h, \gamma) \phi(h - \gamma) dh.$$

It follows from this equality and (9) that

$$\begin{aligned}
c(\gamma; s, \rho) &= 1 - \alpha + \int_{-c}^c (\ell(h, \gamma; s(h)) - \ell^\dagger(h, \gamma)) \phi(h - \gamma) dh \\
&= 1 - \alpha - \left(\int_0^c (\ell(h, \gamma; s(h)) - \ell^\dagger(h, \gamma)) \phi(h - \gamma) dh \right. \\
&\quad \left. + \int_{-c}^0 (\ell(h, \gamma; s(h)) - \ell^\dagger(h, \gamma)) \phi(h - \gamma) dh \right).
\end{aligned}$$

Change the variable of integration to $y = -h$ in the second integral. The result $c(\gamma; s, \rho) = 1 - \alpha - R_1(s, \gamma)$ now follows from the fact that both s and ϕ are even functions.

□

A.2 Proof of Theorem 2

The following proof is based, in part, on the derivations described in Section 4.3 of Giri (2008). Note that

$$\begin{aligned} e(\gamma; s) &= \frac{1}{z(\alpha)} \int_{-\infty}^{\infty} s(h) \phi(h - \gamma) dh \\ &= \int_{-\infty}^{-c} \phi(h - \gamma) dh + \frac{1}{z(\alpha)} \int_{-c}^c s(h) \phi(h - \gamma) dh + \int_c^{\infty} \phi(h - \gamma) dh, \end{aligned} \quad (10)$$

since $s(x) = z(\alpha)$ for all $|x| \geq c$. Obviously,

$$1 = \int_{-\infty}^{-c} \phi(h - \gamma) dh + \int_{-c}^c \phi(h - \gamma) dh + \int_c^{\infty} \phi(h - \gamma) dh.$$

It follows from this equality and (10) that

$$\begin{aligned} e(\gamma; s) &= 1 + \int_{-c}^c \left(\frac{s(h)}{z(\alpha)} - 1 \right) \phi(h - \gamma) dh \\ &= 1 + \int_{-c}^0 \left(\frac{s(h)}{z(\alpha)} - 1 \right) \phi(h - \gamma) dh + \int_0^c \left(\frac{s(h)}{z(\alpha)} - 1 \right) \phi(h - \gamma) dh. \end{aligned}$$

Change the variable of integration to $y = -h$ in the first integral on the right-hand side. The fact that both s and ϕ are even functions implies that (1) is true. □

A.3 Properties of $dq(x; h, \gamma, \nu)/dx$ considered as a function of x

It is straightforward to show that

$$\ell(h, \gamma; x) = \Phi \left(\frac{b(h) + x - \rho(h - \gamma)}{(1 - \rho^2)^{1/2}} \right) - \Phi \left(\frac{b(h) - x - \rho(h - \gamma)}{(1 - \rho^2)^{1/2}} \right)$$

and

$$\ell^\dagger(h, \gamma) = \Phi \left(\frac{z(\alpha) - \rho(h - \gamma)}{(1 - \rho^2)^{1/2}} \right) - \Phi \left(\frac{-z(\alpha) - \rho(h - \gamma)}{(1 - \rho^2)^{1/2}} \right).$$

It follows that

$$\frac{dq(x; h, \gamma, \nu)}{dx} = t_1(h, \gamma, \nu) - t_2(x; h, \gamma, \nu),$$

where $t_1(h, \gamma, \nu)$ is defined to be

$$\frac{1}{z(\alpha)} \left(2\phi(h) + \sum_{j=1}^{m_2} \nu_2(j) (\phi(h - \gamma_2(j)) + \phi(h + \gamma_2(j))) \right)$$

and $t_2(x; h, \gamma, \nu)$ is defined to be

$$\sum_{j=1}^{m_1} \nu_1(j) \left(\phi(h - \gamma_1(j)) \frac{d\ell(h, \gamma_1(j); x)}{dx} + \phi(h + \gamma_1(j)) \frac{d\ell(-h, \gamma_1(j); x)}{dx} \right),$$

with

$$\frac{d\ell(h, \gamma; x)}{dx} = \frac{1}{(1 - \rho^2)^{1/2}} \left(\phi \left(\frac{b(h) + x - \rho(h - \gamma)}{(1 - \rho^2)^{1/2}} \right) + \phi \left(\frac{b(h) - x - \rho(h - \gamma)}{(1 - \rho^2)^{1/2}} \right) \right).$$

Suppose that $h \in [0, c]$ and (γ, ν) are given. Then $t_1(h, \gamma, \nu)$ is a fixed positive number. Observe that $t_2(x; h, \gamma, \nu)$ is a function of $x \in [0, \infty)$ that can only take positive values and $d\ell(h, \gamma; x)/dx$ approaches 0 as $x \rightarrow \infty$. We will use the following theorem to find $\tilde{x} < \infty$, such that $dq(x; h, \gamma, \nu)/dx > 0$ for all $x \geq \tilde{x}$. This implies that a value of x that minimizes $q(x; h, \gamma, \nu)$ cannot belong to the interval $[\tilde{x}, \infty)$.

Theorem 4. *Let $\mu(h, \rho, \gamma) = b(h) - \rho(h - \gamma)$. Then $t_2(x; h, \gamma, \nu)$ is a decreasing function of $x \in [x^*, \infty)$, where*

$$x^* = \max (|\mu(h, \rho, \gamma_1(1))|, |\mu(-h, \rho, \gamma_1(1))|, \dots, |\mu(h, \rho, \gamma_1(m_1))|, |\mu(-h, \rho, \gamma_1(m_1))|).$$

Proof. We first prove that, for every $h \in \mathbb{R}$, $d\ell(h, \gamma; x)/dx$ is a decreasing function of $x \in [|\mu(h, \rho, \gamma)|, \infty)$. Observe that, for all $x \geq 0$,

$$\begin{aligned} & \phi \left(\frac{b(h) + x - \rho(h - \gamma)}{(1 - \rho^2)^{1/2}} \right) + \phi \left(\frac{b(h) - x - \rho(h - \gamma)}{(1 - \rho^2)^{1/2}} \right) \\ &= \phi \left(\frac{b(h) - \rho(h - \gamma) + x}{(1 - \rho^2)^{1/2}} \right) + \phi \left(\frac{-(b(h) - \rho(h - \gamma)) + x}{(1 - \rho^2)^{1/2}} \right), \text{ since } \phi \text{ is an even function,} \\ &= \phi \left(\frac{\mu(h, \rho, \gamma) + x}{(1 - \rho^2)^{1/2}} \right) + \phi \left(\frac{-\mu(h, \rho, \gamma) + x}{(1 - \rho^2)^{1/2}} \right) \\ &= \phi \left(\frac{|\mu(h, \rho, \gamma)| + x}{(1 - \rho^2)^{1/2}} \right) + \phi \left(\frac{-|\mu(h, \rho, \gamma)| + x}{(1 - \rho^2)^{1/2}} \right). \end{aligned}$$

This is a decreasing function of $x \in [|\mu(h, \rho, \gamma)|, \infty)$. Consequently, $d\ell(h, \gamma; x)/dx$ and $d\ell(-h, \gamma; x)/dx$ are decreasing functions of $x \in [|\mu(h, \rho, \gamma)|, \infty)$ and $x \in [|\mu(-h, \rho, \gamma)|, \infty)$, respectively.

Thus $d\ell(h, \gamma_1(j); x)/dx$ is a decreasing function of $x \in [|\mu(h, \rho, \gamma_1(j))|, \infty)$, for $j = 1, \dots, m_1$. Therefore

$$\sum_{j=1}^{m_1} \nu_1(j) \phi(h - \gamma_1(j)) \frac{d\ell(h, \gamma_1(j); x)}{dx}$$

is a decreasing function of $x \in \left[\max_{j=1, \dots, m_1} |\mu(h, \rho, \gamma_1(j))|, \infty \right)$. Similarly, $d\ell(-h, \gamma_1(j); s)/ds$ is a decreasing function of $s \in \left[|\mu(-h, \rho, \gamma_1(j))|, \infty \right)$, for $j = 1, \dots, m_1$. Therefore

$$\sum_{j=1}^{m_1} \nu_1(j) \phi(h + \gamma_1(j)) \frac{d\ell(-h, \gamma_1(j); s)}{ds}$$

is a decreasing function of $s \in \left[\max_{j=1, \dots, m_1} |\mu(-h, \rho, \gamma_1(j))|, \infty \right)$. Therefore $t_2(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})$ is a decreasing function of $x \in [x^*, \infty)$. \square

We use this theorem to find $\tilde{x} < \infty$, such that $dq(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx > 0$ for all $x \geq \tilde{x}$ as follows. First evaluate x^* and then $dq(x^*; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx$. If $dq(x^*; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx > 0$ then set $\tilde{x} = x^*$ and stop; otherwise use the R function `uniroot` to find the solution for $x \in [x^*, \infty)$ of $dq(x; h, \boldsymbol{\gamma}, \boldsymbol{\nu})/dx = 0$ and then set \tilde{x} equal to this solution.