Model Averaged Confidence Intervals

Paul Kabaila

Department of Mathematics and Statistics, La Trobe University

A.H. Welsh

Mathematical Sciences Institute, The Australian National University

Waruni Abeysekera

Department of Mathematics and Statistics, La Trobe University

September 30, 2014

Running headline: Model Averaged Confidence Intervals

Abstract

We develop an approach to evaluating frequentist model averaging procedures by considering them in a simple situation in which there are two nested linear regression models over which we average. We introduce a general class of model averaged confidence intervals, obtain exact expressions for the coverage and the scaled expected length of the intervals and use these to compute these quantities for the model averaged profile likelihood confidence intervals proposed by Fletcher and Turek (2011) and the model averaged tail area confidence intervals proposed by Turek and Fletcher (2012). We show that the Fletcher-Turek (2011) confidence intervals can have coverage well below the nominal coverage and expected length greater than that of the standard confidence interval with coverage equal to the same minimum coverage. In these situations, the Fletcher-Turek confidence intervals are not better than the standard confidence interval used after model selection but ignoring the model selection process. The Turek-Fletcher (2012) confidence intervals perform better than the Fletcher-Turek and post-model-selection confidence intervals but, for the examples that we consider, offers little over simply using the standard confidence interval for θ under the full model, with the same nominal coverage.

Keywords: Akaike Information Criterion (AIC); confidence interval; coverage probability; expected length; model selection; nominal coverage; profile likelihood, regression models; tail area confidence interval.

1 Introduction

It is common practice in applied statistics to carry out data-based model selection by, for example, using preliminary hypothesis tests or minimizing a criterion such as the Akaike Information Criterion (AIC) and then to use the selected model to construct confidence intervals as if it had been given to us *a priori* as the true model. This procedure can lead to confidence intervals with minimum coverage probabilities far below the nominal coverage probability; see Kabaila (2009) for a review of the literature on this topic.

In recent years, there has been growing interest in using techniques which involve several models to try to incorporate model uncertainty into the inferences. These techniques, loosely referred to as model-averaging, are used in both the Bayesian and the frequentist literature; see, for example, Buckland et al. (1997), Raftery et al. (1997), Volinsky et al. (1997), Hoeting et al. (1999), Burnham and Anderson (2002) and Claeskens and Hjort (2008). In this paper, we focus on frequentist model-averaging techniques for constructing confidence intervals.

The earliest frequentist approach to constructing model-averaged confidence intervals (see Buckland et al, 1997 and Burnham and Anderson, 2002) was to centre the interval on a model-averaged estimator and determine the width of the interval by an estimate of the standard deviation of this estimator. The distribution theory on which these intervals are based is not (even approximately) correct (Claeskens and Hjort, 2008, p.207) but simulation studies report that these intervals work well in terms of coverage probability in particular cases (Lukacs et al., 2010; Fletcher and Dillingham, 2011). A different approach was proposed by Hjort and Claeskens (2003) but this turns out to be essentially the same as the standard confidence interval based on fitting a full model (Kabaila and Leeb, 2006; Wang and Zou, 2013). More recently, Fletcher and Turek (2011) and Turek and Fletcher (2012) have proposed averaging confidence interval construction procedures from each of the possible models. Fletcher and Turek (2011) averaged the profile likelihood confidence interval procedure and Turek and Fletcher (2012) averaged the tail areas of the distributions of the estimators from each of the possible models.

Given the practical importance of the problem, it is not surprising that considerable hope has been invested in model averaging as a simple, general method for making valid inferences under model uncertainty. In this context, it is important to develop a theoretical understanding of the properties of model averaging procedures so that we can put their increasing use on a firm basis. A good starting point is to explore the properties of procedures in meaningful, tractable scenarios which allow us to evaluate whether they work as expected, to compare different proposals and perhaps to modify and improve current proposals. We make a start on this by developing a general method for studying the theoretical properties of model averaging procedures in a simple scenario that is both meaningful and tractable and then apply it to the Fletcher and Turek (2011) model averaged profile likelihood confidence interval (MPI) procedure and the Turek and Fletcher (2012) model averaged tail area confidence interval (MATA) procedure.

We obtain a $1 - \alpha$ level profile likelihood confidence interval for a parameter θ in a model $\mathcal{M}_j$ by computing the signed-root log-likelihood ratio for θ under $\mathcal{M}_j$ and then solving for the lower and upper endpoints of the interval the two equations obtained by equating the normal cumulative distribution function evaluated at the signed-root log likelihood ratio to $1 - \alpha/2$ and $\alpha/2$, respectively. We obtain tail area confidence intervals in the same way by replacing the the signed-root log-likelihood ratio by the t ratio $T_j(\theta) = (\widehat{\theta}_j - \theta)/\text{se}(\widehat{\theta}_j)$ and solving the two equations obtained by equating $G_{\nu_j}(T_j(\theta))$ to $1 - \alpha/2$ and $\alpha/2$, where G_{ν_j} is the cumulative distribution function of the distribution of $T_j(\theta)$ under model $\mathcal{M}_j$ (i.e. the Student t distribution with ν_j degrees of freedom). When we have models $\{\mathcal{M}_1, \ldots, \mathcal{M}_R\}$ for a fixed, finite R, MPI and MATA confidence intervals for θ, with nominal coverage $1-\alpha$, are obtained by solving for the endpoints a weighted average of the respective endpoint equations for each model. There are various ways to choose the weights; we follow Fletcher and Turek (2011) and Turek and Fletcher (2012) and focus on weights derived by exponentiating the Akaike Information Criterion (AIC) for each model.

The only evaluation of MPI and MATA to date has been by simulation; Fletcher and Turek (2011) and Turek and Fletcher (2012) showed that these procedures perform well in particular settings. It is natural to use simulations to evaluate different confidence intervals, but simulation methods have weaknesses for evaluating performance criteria. First, simulations cover only a limited set of particular settings (particularly, values of the unknown nuisance parameters) and the conclusions apply only to these settings. They may therefore not consider settings where the coverage is low or the expected length is large. We can improve the situation by evaluating minimum coverage probabilities and maximum expected lengths to characterise performance over unknown nuisance parameters. Secondly, the variability in simulation results complicates finding bounds on coverage or expected length, particularly when there are a large number of parameters to vary in the underlying distribution. We therefore use exact calculations to evaluate the properties of the confidence intervals both in particular settings and uniformly over unknown nuisance parameters.

For simplicity, we consider a scenario with only two possible models, a linear regression model with independent and identically distributed normal errors ($\mathcal{M}_2$) and the same model with a linear constraint on the regression parameters ($\mathcal{M}_1$). We evaluate the properties of model averaged confidence intervals, with nominal coverage $1-\alpha$, for a parameter of interest θ that is common to both models. This scenario is simple but, nonetheless, includes practically important problems. For example, in the comparison of two treatments for a given value of the single covariate in a one-way analysis of covariance, the parameter of interest θ is the treatment effect for a given value of the covariate and the two models $\mathcal{M}_2$ and $\mathcal{M}_1$ are distinguished by whether τ, the difference in the coefficients of the covariate, is unconstrained or constrained to equal zero (so the fitted models have parallel mean functions). In general, θ and τ can be any linearly independent linear functions of the regression parameter and we obtain general results for any given model matrix, so allowing any possible set of nuisance regression parameters. We focus on two properties, the coverage probability and the scaled expected length, where the scaling is with respect to the expected length of the standard confidence interval at the minimum coverage level. We derive computationally convenient, exact expressions for the coverage probability and the scaled expected length of model averaged confidence intervals for θ, so that we do not need to resort to simulations.

Our results show that there are situations in which MPI has coverage much lower than the nominal coverage and expected length greater than that of the standard confidence interval with coverage equal to the minimal coverage. In these situations, MPI performs worse than standard confidence intervals used after model selection but ignoring the model selection process. MATA performs better than MPI in these same situations, performing like the standard confidence interval under $\mathcal{M}_2$. This shows the difficulty of improving on the strategy of using complicated models and avoiding any kind of model selection. These results reinforce the need to develop new procedures and highlight the need for careful analysis of new procedures.

We present our theoretical results in Section 2 and illustrate their application to a real data example from a cloud seeding experiment in which the parameter of interest is the effect of cloud seeding in Section 3. We present the coverage probability and the scaled expected length of model averaged confidence intervals for the parameter of interest and show how to interpret these values. We conclude with a brief discussion in Section 4. Theoretical calculations and the proofs of the Theorems are presented in an Appendix.

2 Theoretical details

In this Section, we describe a general class of frequentist model averaged confidence intervals for θ that includes Fletcher and Turek's (2011) MPI and Turek and Fletcher's (2012) MATA procedures. We give exact expressions for the coverage probability and the scaled expected length of these intervals. The proofs are left to the Appendix.

The model $\mathcal{M}_2$ is given by

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{Y}$ is a random n-vector of responses, $\boldsymbol{X}$ is a known $n \times p$ model matrix with $p < n$ linearly independent columns, $\boldsymbol{\beta}$ is an unknown p-vector parameter and $\boldsymbol{\varepsilon} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, with σ^2 an unknown positive parameter. Suppose that we are interested in making inference about the parameter $\theta = \boldsymbol{a}^\top \boldsymbol{\beta}$, where $\boldsymbol{a}$ is a specified nonzero p-vector. Suppose also that we define the parameter $\tau = \boldsymbol{c}^\top \boldsymbol{\beta} - t$, where $\boldsymbol{c}$ is a specified nonzero p-vector that is linearly independent of $\boldsymbol{a}$ and t is a specified number. The model $\mathcal{M}_1$ is $\mathcal{M}_2$ with $\tau = 0$.

Let $\widehat{\boldsymbol{\beta}}$ be the least squares estimator of $\boldsymbol{\beta}$ and $\widehat{\sigma}^2 = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})/(n-p)$ be the usual unbiased estimator of σ^2. Set $\widehat{\theta} = \boldsymbol{a}^\top \widehat{\boldsymbol{\beta}}$ and $\widehat{\tau} = \boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} - t$. Define $v_\theta = \mathrm{Var}(\widehat{\theta})/\sigma^2 = \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{a}$ and $v_\tau = \mathrm{Var}(\widehat{\tau})/\sigma^2 = \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c}$. Two important quantities are the known correlation $\rho = \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c} / (v_\theta v_\tau)^{1/2}$ between $\widehat{\theta}$ and $\widehat{\tau}$ and the unknown parameter $\gamma = \tau / \big(\sigma v_\tau^{1/2}\big)$.

Suppose that under the models $\mathcal{M}_1$ and $\mathcal{M}_2$ confidence intervals $\big[\widehat{\theta}_{1l}, \widehat{\theta}_{1u}\big]$, and

$\left[\widehat{\theta}_{2l},\, \widehat{\theta}_{2u}\right]$ for θ with nominal coverage $1-\alpha$, are found by solving the equations

$$a_1\left\{(\widehat{\theta}-\widehat{\theta}_{1l})/v_\theta^{1/2},\, \widehat{\tau}/v_\tau^{1/2},\, \widehat{\sigma}\right\} = 1-\alpha/2 \quad \text{and} \quad a_1\left\{(\widehat{\theta}-\widehat{\theta}_{1u})/v_\theta^{1/2},\, \widehat{\tau}/v_\tau^{1/2},\, \widehat{\sigma}\right\} = \alpha/2$$

and

$$a_2\left\{(\widehat{\theta}-\widehat{\theta}_{2l})/v_\theta^{1/2},\, \widehat{\sigma}\right\} = 1-\alpha/2 \qquad \text{and} \qquad a_2\left\{(\widehat{\theta}-\widehat{\theta}_{2u})/v_\theta^{1/2},\, \widehat{\sigma}\right\} = \alpha/2,$$

where $a_1(\delta, x, y)$ and $a_2(\delta, y)$ are scale invariant, increasing, continuous and bounded functions of $\delta \in \mathbb{R}$ that approach 1 as $\delta \to \infty$ and approach 0 as $\delta \to -\infty$, for each $y > 0$. Here, scale invariance means that $a_1(\delta, x, y) = a_1(k\delta, kx, ky)$ and $a_2(\delta, y) = a_2(k\delta, ky)$ for all δ, x, $y > 0$ and any $k > 0$.

Suppose that the function $w_1 : [0, \infty) \to [0, 1]$ is a decreasing continuous function, such that $w_1(z)$ approaches 0 as $z \to \infty$. We consider the weight function $w_1(x^2/y^2)$ and define

$$h(\delta, x, y) = w_1(x^2/y^2)\, a_1(\delta, x, y) + \{1 - w_1(x^2/y^2)\}\, a_2(\delta, y). \tag{1}$$

It follows from the assumptions on a_1 and a_2 that $h(\delta, x, y)$ is a scale invariant, increasing continuous function of $\delta \in \mathbb{R}$ that approaches 1 as $\delta \to \infty$ and approaches 0 as $\delta \to -\infty$, for each $x \in \mathbb{R}$ and $y > 0$. We define a frequentist model averaged confidence interval $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$ for θ with nominal coverage $1-\alpha$ by solving

$$h\left\{(\widehat{\theta}-\widehat{\theta}_l)/v_\theta^{1/2}, \widehat{\tau}/v_\tau^{1/2}, \widehat{\sigma}\right\} = 1-\alpha/2 \quad \text{and} \quad h\left\{(\widehat{\theta}-\widehat{\theta}_u)/v_\theta^{1/2}, \widehat{\tau}/v_\tau^{1/2}, \widehat{\sigma}\right\} = \alpha/2$$

for $\widehat{\theta}_l$ and $\widehat{\theta}_u$.

The coverage and expected length properties of $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$ are conveniently expressed in terms of $\delta_u(x, y)$ which, for each $x \in \mathbb{R}$ and $y > 0$, is defined to be the solution in δ of the equation $h(\delta, x, y) = u$.

Theorem 1 *The coverage probability of the frequentist model averaged confidence interval* $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$ *(averaged over* $\mathcal{M}_1$ *and* $\mathcal{M}_2$*), with nominal coverage* $1-\alpha$*, is*

$$\begin{aligned} P\left(\widehat{\theta}_l \le \theta \le \widehat{\theta}_u\right) \;&=\; \int_0^\infty \int_{-\infty}^\infty \left[\Phi\left\{\frac{\delta_{1-\alpha/2}(x,y) - \rho(x-\gamma)}{(1-\rho^2)^{1/2}}\right\}\right. \\ &\qquad \left. -\Phi\left\{\frac{\delta_{\alpha/2}(x,y) - \rho(x-\gamma)}{(1-\rho^2)^{1/2}}\right\}\right] \phi(x-\gamma)\, f_{n-p}(y)\, dx\, dy, \end{aligned}$$

where ϕ *is the probability density function of the standard normal distribution and* $f_\nu(y)$ *is the probability density function of* $(Q/\nu)^{1/2}$*, where* Q *has a* χ^2_ν *distribution.*

Theorem 1 shows that the coverage of $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$ is a function of the nominal coverage $1-\alpha$, the residual degrees of freedom $n-p$, the correlation ρ between $\widehat{\theta}$ and $\widehat{\tau}$, and the unknown parameter $\gamma = \tau/\left(\sigma v_\tau^{1/2}\right)$. The only unknown quantity is γ. We use the minimum coverage over γ to describe the worst case results without having to specify particular values for γ. We can obtain a useful upper bound to the minimum coverage over γ.

Corollary 1 *As* $\gamma \to \infty$*, the coverage probability of the frequentist model averaged confidence interval* $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$ *(averaged over* $\mathcal{M}_1$ *and* $\mathcal{M}_2$*), with nominal coverage* $1-\alpha$*, converges to the coverage probability of the corresponding interval under* $\mathcal{M}_2$*, with nominal coverage* $1-\alpha$*, respectively. That is,*

$$P\left(\widehat{\theta}_l \le \theta \le \widehat{\theta}_u\right) \to P\left(\widehat{\theta}_{2l} \le \theta \le \widehat{\theta}_{2u}\right) \quad as \quad \gamma \to \infty.$$

An immediate consequence is that

$$\inf_\gamma P\left(\widehat{\theta}_l \le \theta \le \widehat{\theta}_u\right) \le P\left(\widehat{\theta}_{2l} \le \theta \le \widehat{\theta}_{2u}\right).$$

Corollary 1 shows that frequentist model averaging cannot increase the minimum coverage probability above that of the interval $\left[\widehat{\theta}_{2l},\, \widehat{\theta}_{2u}\right]$, with nominal coverage $1-\alpha$ under $\mathcal{M}_2$. So, to achieve good coverage, we need to start with intervals with good coverage.

For the expected length of $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$, we obtain the following result.

Theorem 2 *The expected length of the frequentist model averaged confidence interval* $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$ *(averaged over* $\mathcal{M}_1$ *and* $\mathcal{M}_2$*), with nominal level* $1-\alpha$*, is*

$$\mathrm{E}\left(\widehat{\theta}_u - \widehat{\theta}_l\right) = \sigma\, v_\theta^{1/2} \int_0^\infty \int_{-\infty}^\infty \left\{\delta_{1-\alpha/2}(x,y) - \delta_{\alpha/2}(x,y)\right\} \phi(x-\gamma)\, f_{n-p}(y)\, dx\, dy,$$

where ϕ *is the probability density function of the standard normal distribution and* $f_\nu(y)$ *is the probability density function of* $(Q/\nu)^{1/2}$*, where* Q *has a* χ^2_ν *distribution.*

Let $c_{\min}$ denote the minimum coverage probability of $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$. The expected length of the standard interval that has this minimum coverage is

$$2\, G_{n-p}^{-1}((c_{\min}+1)/2)\ \mathrm{E}(\widehat{\sigma})\, v_\theta^{1/2} = 2^{3/2}\sigma\, v_\theta^{1/2}\, G_{n-p}^{-1}((c_{\min}+1)/2) \frac{\Gamma\{(n-p+1)/2\}}{(n-p)^{1/2}\,\Gamma\{(n-p)/2\}},$$

so the scaled expected length of $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$ is

$$(n-p)^{1/2}\,\Gamma\{(n-p)/2\} \frac{\int_0^\infty \int_{-\infty}^\infty \{\delta_{1-\alpha/2}(x,y) - \delta_{\alpha/2}(x,y)\}\phi(x-\gamma)\, f_{n-p}(y)\, dx\, dy}{2^{3/2}\,\Gamma\{(n-p+1)/2\} G_{n-p}^{-1}((c_{\min}+1)/2)}.$$

As with the coverage, the only unknown quantity in this expression is γ, so we study the maximum scaled expected length over γ.

The range of calculations needed to evaluate the coverage probability and the scaled expected length of $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$ are reduced by the following result that shows that, because of symmetry, we need only consider $\gamma \geq 0$ and $\rho \geq 0$.

Theorem 3 *We make the dependence of $\delta_u(x,y)$ on ρ explicit by using the notation $\delta_u(x,y,\rho)$ in place of $\delta_u(x,y)$. Suppose that $\delta_{1-\alpha/2}(-x,y,\rho) = -\delta_{\alpha/2}(x,y,\rho)$ and $\delta_u(x,y,-\rho) = \delta_u(-x,y,\rho)$. The coverage probability and the scaled expected length of the frequentist model averaged confidence interval $\left[\widehat{\theta}_l,\, \widehat{\theta}_u\right]$ (averaged over $\mathcal{M}_1$ and $\mathcal{M}_2$) are both even functions of γ for fixed ρ and even functions of ρ for fixed γ.*

We can apply Theorems 1–3 to a variety of confidence intervals and weight functions, including the profile likelihood and tail area methods of Fletcher and Turek (2011) and Turek and Fletcher (2012). Their recommended weights based on AIC for the models $\mathcal{M}_1$ and $\mathcal{M}_2$ correspond to

$$w_1(z) = \frac{1}{1 + \left\{1 + \frac{z}{n-p}\right\}^{n/2} \exp(-1)}. \tag{2}$$

It is straightforward to incorporate other weights (such as weights based on BIC rather than AIC) but, to save space, we consider only the weights based on AIC.

The signed-root log-likelihood ratio statistic used by Fletcher and Turek (2011) is *minus* the usual definition; which definition we adopt makes no essential difference to the results so we follow Fletcher and Turek (2011). We show in the Supplementary

Online Material that the signed-root log-likelihood ratio statistic for $\mathcal{M}_2$ is $r_2\big\{(\widehat{\theta}-\theta)/v_\theta^{1/2}, \widehat{\sigma}\big\}$, where

$$r_2(\delta, y) \quad = \quad \text{sign}(\delta)\left[n \log\left\{1 + \frac{\delta^2}{(n-p)y^2}\right\}\right]^{1/2}, \tag{3}$$

and the signed-root log-likelihood ratio statistic for $\mathcal{M}_1$ is $r_1\big\{(\widehat{\theta}-\theta)/v_\theta^{1/2}, \widehat{\tau}/v_\tau^{1/2}, \widehat{\sigma}\big\}$, where

$$r_1(\delta, x, y) \quad = \quad \text{sign}(\delta - \rho x)\left(n \log\left[1 + \frac{(\delta-\rho x)^2}{(1-\rho^2)\{x^2 + (n-p)y^2\}}\right]\right)^{1/2}. \tag{4}$$

Profile likelihood confidence intervals for θ with nominal coverage $1-\alpha$ under $\mathcal{M}_1$ and $\mathcal{M}_2$ are obtained by solving equations based on

$$a_1(\delta, x, y) = \Phi\{r_1(\delta, x, y)\} \qquad \text{and} \qquad a_2(\delta, y) = \Phi\{r_2(\delta, y)\},$$

where Φ is the standard normal cumulative distribution function. These functions obviously satisfy the conditions of Theorems 1 and 2. We show in the Supplementary Online Material that these functions also satisfy the conditions of Theorem 3. So these Theorems describe the properties of the Fletcher and Turek (2011) MPI and are used to construct Figures 1–2 in the next Section.

Turek and Fletcher (2012) consider tail area confidence intervals for θ with nominal coverage $1-\alpha$ under $\mathcal{M}_1$ and $\mathcal{M}_2$ that are obtained from the t ratios

$$r_2(\delta, y) \quad = \quad \delta/y \qquad \text{and} \qquad r_1(\delta, x, y) = \frac{\delta - \rho x}{\left(\frac{x^2+(n-p)y^2}{n-p+1}\right)^{1/2}(1-\rho^2)^{1/2}}$$

by solving equations based on

$$a_1(\delta, x, y) = G_{n-p+1}\{r_1(\delta, x, y)\} \qquad \text{and} \qquad a_2(\delta, y) = G_{n-p}\{r_2(\delta, y)\},$$

where G_ν is the distribution function of Student's t distribution with ν degrees of freedom. These functions obviously satisfy the conditions of Theorems 1 and 2. We show in the Supplementary Online Material that these functions also satisfy the conditions of Theorem 3. So these Theorems also describe the properties of the

Turek and Fletcher (2012) MATA and are used to construct Figures 3–4 in the next Section.

For MPI, the upper bound for the minimum coverage probability given in Corollary 1 is

$$P\left(\widehat{\theta}_{2l} \le \theta \le \widehat{\theta}_{2u}\right) = 2\,G_{n-p}\left[(n-p)^{1/2}\left\{\exp\left(\frac{z_{1-\alpha/2}^2}{n}\right) - 1\right\}^{1/2}\right] - 1. \quad (5)$$

This upper bound is very easily computed and can be used to provide some guidance as to when MPI should not be used. For MPI, (5) can be well below the nominal coverage $1-\alpha$ because the profile likelihood confidence interval under $\mathcal{M}_2$ can have poor coverage. To see this note that for fixed $p/n = r$, the coverage of the profile likelihood interval under $\mathcal{M}_2$ is

$$\begin{aligned} P\left(\widehat{\theta}_{2l} \le \theta \le \widehat{\theta}_{2u}\right) &= 2\,G_{n(1-r)}\left[(1-r)^{1/2}\left\{z_{1-\alpha/2}^2 + O(n^{-1})\right\}^{1/2}\right] - 1 \\ &\to 2\Phi\left\{(1-r)^{1/2} z_{1-\alpha/2}\right\} - 1\,, \quad \text{as} \;\; n \to \infty, \end{aligned}$$

where $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. Thus the coverage probability of the profile likelihood confidence interval under $\mathcal{M}_2$ (and hence MPI) decreases as $p/n = r$ increases and is substantially less than the nominal coverage $1-\alpha$ unless p/n is small. In contrast, the tail area interval under $\mathcal{M}_2$ with nominal coverage $1-\alpha$ has coverage $1-\alpha$ so the coverage of MATA with nominal coverage $1-\alpha$ approaches $1-\alpha$ as $\gamma \to \infty$. Thus we expect MATA to have better coverage properties than MPI.

3 Cloud seeding example

In this Section, we illustrate how we can use our theoretical results in the context of a real data example from a cloud seeding experiment. The data are presented and analysed by Biondini, Simpson and Woodley (1997), Miller (2002, Section 3.12) and Kabaila (2005). Following Kabaila (2005), we compare the effect of seeding

(TRT=1) against the random control (TRT=2) treatment in the moving echo motion category (CAT=1) subgroup of the data. The response variable is the floating target rainfall volume and the sample size is $n = 33$. In addition to the treatment indicator, there are five other predictor variables, which include seedability. A detailed description of these variables and the units of measurement used, are provided in the Supplementary Online Material. The models considered by Miller (2002, Section 3.12) and Kabaila (2005) included the intercept, treatment indicator, the main effects, squared effects and the interactions between the five predictor variables so that p, the dimension of the regression parameter vector, is 22. All these additional variables can be included in the model or not; variable selection has been carried out by Miller (2002, Section 3.12) and Kabaila (2005) for many variables in this study. For illustration, we consider model averaging over the full model ($p = 22$) and the submodel excluding the squared seedability term whose coefficient we denote by τ. The goal is to construct a 95% confidence interval for θ, the expected response when cloud seeding is used minus the expected response under random control when all the other explanatory variables are the same.

We can construct several confidence intervals for θ with nominal coverage 0.95. The standard (tail area) Student t confidence interval is $[-0.327, 3.421]$ under $\mathcal{M}_2$. AIC selects $\mathcal{M}_1$ so the naive approach of ignoring the model selection process leads to using the standard interval [0.474, 2.650] computed under $\mathcal{M}_1$. The MATA is $[-0.183, 3.370]$. The profile-likelihood confidence interval is [0.554, 2.539] under $\mathcal{M}_2$ and the MPI is [0.618, 2.572].

For MPI, we plot the exact coverage and the scaled expected length in Figures 1 and 2, respectively. We find that the coverage probability of MPI is close to 0.7315 for all γ rather than the nominal 0.95 and the scaled expected length is close to one for all γ. Therefore, MPI is actually similar to the standard 0.7315 confidence interval for θ. The minimum coverage of MPI decreases as $|\rho|$ increases and as p/n increases. For the cloud seeding example, $p/n = 2/3$ which is not small and the correlation between $\widehat{\theta}$ and $\widehat{\tau}$ is $\rho = 0.2472$ which is small and positive. The poor minimum coverage of MPI is due to the value of p/n not being small.

[Figures 1 and 2 near here]

For MATA, we plot the exact coverage and the scaled expected length in Figures 3 and 4, respectively. We find that the coverage probability of MATA is close to 0.95 for all γ with a minimum coverage probability 0.9465 and the scaled expected length is close to one for all γ. Therefore, MATA is similar to and offers no improvement on the standard 0.95 confidence interval for θ under $\mathcal{M}_2$. The minimum coverage of MATA decreases as $|\rho|$ increases but the interval is still better than MPI.

[Figures 3 and 4 near here]

It is interesting to compare the MPI and MATA interval with the confidence interval constructed after selecting between models $\mathcal{M}_1$ and $\mathcal{M}_2$ the model with smaller AIC and ignoring the selection process. The coverage probability of this interval as a function of γ is shown in Figure 5 (Kabaila and Giri, 2009a, b). Comparing this with Figure 1, we see that the coverage probability for this post-model-selection interval is uniformly far better than that of the MPI. In contrast, MATA has slightly better coverage probability than this post-model-selection interval.

[Figure 5 near here]

Additional figures and a second example are included in the Supplementary Online Material.

4 Conclusion

We have examined the exact coverage and scaled expected length of a class of model averaged confidence intervals for a parameter θ, with nominal coverage $1 - \alpha$, that includes MPI and MATA in a particular simple situation in which there are two linear regression models (differing in only a single parameter τ) to average over. We showed that both the coverage and the scaled expected length depend on n, $n-p$, the correlation ρ between the least squares estimators $\widehat{\theta}$ and $\widehat{\tau}$, and the unknown true value $\gamma = \tau / \big(\sigma v_{\tau}^{1/2}\big)$. As γ is unknown, it is useful to consider the minimum coverage and the maximum scaled expected length over γ.

Our results show that MPI can perform poorly when p/n is not small or when $|\rho|$ is large, and should not be used in these situations. In these situations, MPI performs no better than than post-model-selection confidence intervals which ignore the selection process. MATA performs better than MPI and post-model-selection confidence intervals but, for the examples that we consider, offers little over simply using the standard confidence interval for θ under $\mathcal{M}_2$ with the same nominal level. An ideal confidence interval should have minimal coverage equal to its nominal coverage and, to show a benefit of model selection, have scaled expected length that (a) is substantially less than 1 under $\mathcal{M}_1$ and (b) has a maximum value that it not much larger than 1 and (c) is close to 1 if the data happens to strongly contradict the model $\mathcal{M}_1$. This is evidently difficult to achieve.

Performing well in the simple situation we have developed in this paper does not mean that a model averaging procedure will always perform well. In particular, we also need to explore other situations, such as other models. For example, MATA is related to the Wald statistic and such statistics often do not perform well in discrete data problems.

References

Biondini, R., Simpson, J. and Woodley, W. (1977). Empirical predictors for natural and seeded rainfall in the Florida area cumulus experiment (FACE), 1970-1975. *Journal of Applied Meteorology* **16**, 585–594.

Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997). Model selection: an integral part of inference. *Biometrics* **53**, 603–618.

Burnham, K.P. and Anderson, D.R. (2002). *Model selection and multimodel in-*

ference, a practical information-theoretic approach, 2nd edition. Springer: New York.

Claeskens, G. and Hjort, N.L. (2008). *Model selection and model averaging.* Cambridge University Press.

Fletcher, D. and Dillingham, P.W. (2011). Model-averaged confidence intervals for factorial experiments. *Comput. Statist. Data Anal.* **55**, 3041–3048.

Fletcher, D. and Turek, D. (2011). Model-averaged profile likelihood intervals. *J. Agric. Biol. Environ. Stat.* **17**, 38–51.

Hoeting, J., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). Bayesian model-averaging: A tutorial (with discussion). *Statist. Sci.* **14**, 382–417.

Hjort, N.L. and Claeskens, G. (2003). Frequentist model average estimators. *J. Amer. Statist. Assoc.* **98**, 879–899.

Kabaila, P. (2005). On the coverage probability of confidence intervals in regression after variable selection. *Aust. N. Z. J. Stat.* **47**, 549–562.

Kabaila, P. (2009). The coverage properties of confidence regions after model selection. *International Statistical Review* **77**, 405–414.

Kabaila, P. and Leeb, H. (2006). On the large-sample minimal coverage probability of confidence intervals after model selection. *J. Amer. Statist. Assoc.* **101**, 619–629.

Kabaila, P. and Giri, K. (2009a). Upper bounds on the minimum coverage probability of confidence intervals in regression after model selection. *Aust. N. Z. J. Stat.* **51**, 271–287.

Kabaila, P. and Giri, K. (2009b). Confidence intervals in regression utilizing uncertain prior information. *J. Statist. Plann. Inference* **139**, 3419–3429.

Lukacs, P.M., Burnham, K.P. and Anderson, D.R. (2010). Model selection bias and Freedman's paradox. *Ann. Inst. Statist. Math.* **62**, 117–125.

Miller, A. (2002). *Subset selection in regression*, 2nd edition. Chapman & Hall/CRC.

Raftery, A.E., Madigan, D. and Hoeting, J.A. (1997). Bayesian model-averaging for linear regression models. *J. Amer. Statist. Assoc.* **92**, 179–191.

Turek, D. and Fletcher, D. (2012). Model-averaged Wald confidence intervals. *Comput. Statist. Data Anal.* **56**, 2809–2815.

Volinsky, C.T., Madigan, D., Raftery, A.E. and Kronmal, R.A. (1997). Bayesian model-averaging in proportional hazard models: Assessing the risk of a stroke. *J. R. Statist. Soc. Ser. C Appl. Stat.* **46**, 433–448.

Wang, H. and Zou, S.Z.F. (2013). Interval estimation by frequentist model averaging. *Commun. Statist. Theory Methods* to appear.

Paul Kabaila, Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia.
E-mail: P.Kabaila@latrobe.edu.au

Appendix

Proof of Theorem 1

The coverage probability of the model averaged confidence interval $\left[\widehat{\theta}_l, \widehat{\theta}_u\right]$ (averaged over $\mathcal{M}_1$ and $\mathcal{M}_2$), with nominal coverage $1-\alpha$, is

$$P\left(\widehat{\theta}_l \le \theta \le \widehat{\theta}_u\right) = 1 - P\left(\theta < \widehat{\theta}_l\right) - P\left(\widehat{\theta}_u > \theta\right).$$

Now $h(\delta, x, y)$ is an increasing function of δ for fixed x and y so

$$\begin{aligned}
&P\left(\theta < \widehat{\theta}_l\right) \\
&\quad = P\left\{(\widehat{\theta}-\theta)/\big(\sigma v_\theta^{1/2}\big) > (\widehat{\theta}-\widehat{\theta}_l)/\big(\sigma v_\theta^{1/2}\big)\right\} \\
&\quad = P\left[h\left\{(\widehat{\theta}-\theta)/\big(\sigma v_\theta^{1/2}\big), \widehat{\tau}/\big(\sigma v_\tau^{1/2}\big), \widehat{\sigma}/\sigma\right\} > 1-\alpha/2\right] \\
&\quad = P\left[(\widehat{\theta}-\theta)/(\sigma v_\theta^{1/2}) > \delta_{1-\alpha/2}\{\widehat{\tau}/(\sigma v_\tau^{1/2}), \widehat{\sigma}/\sigma\}\right] \\
&\quad = \int_0^\infty \int_{-\infty}^\infty P\left[(\widehat{\theta}-\theta)/(\sigma v_\theta^{1/2}) > \delta_{1-\alpha/2}\{\widehat{\tau}/(\sigma v_\tau^{1/2}), \widehat{\sigma}/\sigma\}\,\middle|\,\widehat{\tau}/(\sigma v_\tau^{1/2}) = x, \widehat{\sigma}/\sigma = y\right] \\
&\qquad \times \phi(x-\gamma)\, f_{n-p}(y)\, dx\, dy,
\end{aligned}$$

where $\gamma = \tau/(\sigma v_\tau^{1/2})$. Now the distribution of $(\widehat{\theta} - \theta)/(\sigma v_\theta^{1/2})$ conditional on $\widehat{\tau}/(\sigma v_\tau^{1/2}) = x$ is $N\big(\rho(x-\gamma), 1-\rho^2\big)$, $\widehat{\tau}/(\sigma v_\tau^{1/2}) \sim N(\gamma, 1)$ and $\widehat{\theta}$ and $\widehat{\tau}$ are independent of $\widehat{\sigma}$, so

$$\begin{aligned}
&P\big[(\widehat{\theta}-\theta)/(\sigma v_\theta^{1/2}) > \delta_{1-\alpha/2}\{\widehat{\tau}/(\sigma v_\tau^{1/2}), \widehat{\sigma}/\sigma\} \,\big|\, \widehat{\tau}/(\sigma v_\tau^{1/2}) = x, \, \widehat{\sigma}/\sigma = y\big] \\
&\quad = P\big\{(\widehat{\theta}-\theta)/(\sigma v_\theta^{1/2}) > \delta_{1-\alpha/2}(x,y) \,\big|\, \widehat{\tau}/(\sigma v_\tau^{1/2}) = x\big\} \\
&\quad = 1 - P\big\{(\widehat{\theta}-\theta)/\sigma v_\theta^{1/2} \le \delta_{1-\alpha/2}(x,y) \,\big|\, \widehat{\tau}/(\sigma v_\tau^{1/2}) = x\big\} \\
&\quad = 1 - \Phi\left\{\frac{\delta_{1-\alpha/2}(x,y) - \rho(x-\gamma)}{(1-\rho^2)^{1/2}}\right\}
\end{aligned}$$

and hence

$$P\left(\theta < \widehat{\theta}_l\right) = \int_0^\infty \int_{-\infty}^\infty \left[1 - \Phi\left\{\frac{\delta_{1-\alpha/2}(x,y) - \rho(x-\gamma)}{(1-\rho^2)^{1/2}}\right\}\right] \phi(x-\gamma) f_{n-p}(y)\, dx\, dy.$$

Similarly,

$$\begin{aligned}
1 - P\left(\widehat{\theta}_u > \theta\right) &= P\left(\theta < \widehat{\theta}_u\right) \\
&= P\left[h\{(\widehat{\theta}-\theta)/(\sigma v_\theta^{1/2}), \widehat{\tau}/(\sigma v_\tau^{1/2}), \widehat{\sigma}/\sigma\} > \alpha/2\right] \\
&= P\left[(\widehat{\theta}-\theta)/(\sigma v_\theta^{1/2}) > \delta_{\alpha/2}\{\widehat{\tau}/(\sigma v_\tau^{1/2}), \widehat{\sigma}/\sigma\}\right] \\
&= \int_0^\infty \int_{-\infty}^\infty \left[1 - \Phi\left\{\frac{\delta_{\alpha/2}(x,y) - \rho(x-\gamma)}{(1-\rho^2)^{1/2}}\right\}\right] \phi(x-\gamma)\, f_{n-p}(y)\, dx\, dy.
\end{aligned}$$

Proof of Corollary 1

From the proof of Theorem 1, we can write

$$P\big(\theta < \widehat{\theta}_l\big) = 1 - P\big\{h(G,H,W) \le 1 - \alpha/2\big\},$$

where $G = (\widehat{\theta}-\theta)/\sigma v_\theta^{1/2} \sim N(0,1)$, $H = \widehat{\tau}/\sigma v_\tau^{1/2} \sim N(\gamma, 1)$, $(n-p)W^2 = (n-p)\widehat{\sigma}/\sigma \sim \chi^2_{n-p}$ and (G,H) and W are independent. Note that $w_1(H^2/W^2)$ converges in probability to 0, as $\gamma \to \infty$. Since $0 < a_1(\delta, x, y) < 1$ and $0 < a_2(\delta, y) < 1$ for all $x \in \mathbb{R}$ and $y > 0$, it follows from the definition (1) of h that $h(G,H,W)$ converges in probability to $a_2(G,W)$, as $\gamma \to \infty$. Thus, $h(G,H,W)$ converges in distribution to

$a_2(G, W)$, as $\gamma \to \infty$. The cumulative distribution function of $a_2(G, W)$, evaluated at u, is a continuous function of $u \in \mathbb{R}$. Therefore

$$P\big(\theta < \widehat{\theta}_l\big) \to 1 - P\big\{a_2(G, W) \le 1 - \alpha/2\big\}, \qquad \text{as } \gamma \to \infty.$$

Now consider the confidence interval $\big[\widehat{\theta}_{2l}, \widehat{\theta}_{2u}\big]$, with nominal coverage $1 - \alpha$ under $\mathcal{M}_2$. The lower endpoint of this confidence interval satisfies

$$a_2 \left(\frac{\widehat{\theta} - \widehat{\theta}_{2l}}{\sigma v_\theta^{1/2}}, \frac{\widehat{\sigma}}{\sigma} \right) = 1 - \alpha/2.$$

Note that

$$\begin{aligned} P\big(\theta < \widehat{\theta}_{2l}\big) &= P \left(\frac{\widehat{\theta} - \theta}{\sigma v_\theta^{1/2}} > \frac{\widehat{\theta} - \widehat{\theta}_{2l}}{\sigma v_\theta^{1/2}} \right) \\ &= P \left\{ a_2 \left(\frac{\widehat{\theta} - \theta}{\sigma v_\theta^{1/2}}, \frac{\widehat{\sigma}}{\sigma} \right) > a_2 \left(\frac{\widehat{\theta} - \widehat{\theta}_{2l}}{\sigma v_\theta^{1/2}}, \frac{\widehat{\sigma}}{\sigma} \right) \right\} \\ &= P\big\{a_2(G, W) > 1 - \alpha/2\big\} \\ &= 1 - P\big\{a_2(G, W) \le 1 - \alpha/2\big\}. \end{aligned}$$

So $P\big(\theta < \widehat{\theta}_l\big) \to P\big(\theta < \widehat{\theta}_{2l}\big)$, as $\gamma \to \infty$. Similarly, $P\big(\theta < \widehat{\theta}_u\big) \to P\big(\theta < \widehat{\theta}_{2u}\big)$, as $\gamma \to \infty$ and the Corollary 1 holds.

Proof of Theorem 2

The expected length of the model averaged confidence interval $\left[\widehat{\theta}_l, \widehat{\theta}_u\right]$ (averaged over $\mathcal{M}_1$ and $\mathcal{M}_2$), with nominal coverage $1 - \alpha$, is

$$\begin{aligned} \mathrm{E}\left(\widehat{\theta}_u - \widehat{\theta}_l\right) &= \sigma\, v_\theta^{1/2}\, \mathrm{E}\left\{(\widehat{\theta} - \widehat{\theta}_l)/(\sigma v_\theta^{1/2}) - (\widehat{\theta} - \widehat{\theta}_u)/(\sigma v_\theta^{1/2})\right\} \\ &= \sigma v_\theta^{1/2}\, \mathrm{E}\left[\delta_{1-\alpha/2}\{\widehat{\tau}/(\sigma v_\tau^{1/2}), \widehat{\sigma}/\sigma\} - \delta_{\alpha/2}\{\widehat{\tau}/(\sigma v_\tau^{1/2}), \widehat{\sigma}/\sigma\}\right] \\ &= \sigma\, v_\theta^{1/2} \int_0^\infty \int_{-\infty}^\infty \left\{\delta_{1-\alpha/2}(x, y) - \delta_{\alpha/2}(x, y)\right\} \phi(x - \gamma)\, f_{n-p}(y)\, dx\, dy. \end{aligned}$$

Proof of Theorem 3

We make the dependence of $\delta_u(x, y)$ on ρ explicit by using the notation $\delta_u(x, y, \rho)$ in place of $\delta_u(x, y)$. Recall that under the conditions of the Theorem $\delta_{1-\alpha/2}(-x, y, \rho) = -\delta_{\alpha/2}(x, y, \rho)$ and $\delta_{1-\alpha/2}(x, y, -\rho) = \delta_{\alpha/2}(-x, y, \rho)$.

Proof: For the coverage.

Let

$$\begin{aligned} C(\gamma,\rho) &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi\left\{ \frac{\delta_{1-\alpha/2}(x,y,\rho) - \rho(x-\gamma)}{(1-\rho^2)^{1/2}} \right\} \right. \\ &\quad \left. -\Phi\left\{ \frac{\delta_{\alpha/2}(x,y,\rho) - \rho(x-\gamma)}{(1-\rho^2)^{1/2}} \right\} \right] \phi(x-\gamma)\, f_{n-p}(y)\, dx\, dy. \end{aligned}$$

For each fixed ρ,

$$\begin{aligned} C(-\gamma,\rho) &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi\left\{ \frac{\delta_{1-\alpha/2}(x,y,\rho) - \rho(x+\gamma)}{(1-\rho^2)^{1/2}} \right\} \right. \\ &\quad \left. -\Phi\left\{ \frac{\delta_{\alpha/2}(x,y,\rho) - \rho(x+\gamma)}{(1-\rho^2)^{1/2}} \right\} \right] \phi(x+\gamma)\, f_{n-p}(y)\, dx\, dy \\ &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi\left\{ \frac{\delta_{1-\alpha/2}(-z,y,\rho) - \rho(-z+\gamma)}{(1-\rho^2)^{1/2}} \right\} \right. \\ &\quad \left. -\Phi\left\{ \frac{\delta_{\alpha/2}(-z,y,\rho) - \rho(-z+\gamma)}{(1-\rho^2)^{1/2}} \right\} \right] \phi(-z+\gamma)\, f_{n-p}(y)\, dz\, dy \\ &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi\left\{ -\frac{\delta_{\alpha/2}(z,y,\rho) - \rho(z-\gamma)}{(1-\rho^2)^{1/2}} \right\} \right. \\ &\quad \left. -\Phi\left\{ -\frac{\delta_{1-\alpha/2}(z,y,\rho) - \rho(z-\gamma)}{(1-\rho^2)^{1/2}} \right\} \right] \phi(z-\gamma)\, f_{n-p}(y)\, dz\, dy \\ &= C(\gamma,\rho). \end{aligned}$$

The second line follows by changing the variable to $z = -x$, the third follows by hypothesis and the fact that the standard normal density is an even function, and the fourth follows from the fact that $\Phi(-x) = 1 - \Phi(x)$.

By hypothesis, $\delta_u(x,y,-\rho) = \delta_u(-x,y,\rho)$. Thus, for each fixed γ,

$$\begin{aligned} C(\gamma,-\rho) &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi\left\{ \frac{\delta_{1-\alpha/2}(-x,y,\rho) + \rho(x-\gamma)}{(1-\rho^2)^{1/2}} \right\} \right. \\ &\quad \left. -\Phi\left\{ \frac{\delta_{\alpha/2}(-x,y,\rho) + \rho(x-\gamma)}{(1-\rho^2)^{1/2}} \right\} \right] \phi(x-\gamma)\, f_{n-p}(y)\, dx\, dy \\ &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi\left\{ -\frac{\delta_{\alpha/2}(x,y,\rho) - \rho(x-\gamma)}{(1-\rho^2)^{1/2}} \right\} \right. \\ &\quad \left. -\Phi\left\{ -\frac{\delta_{1-\alpha/2}(x,y,\rho) - \rho(x-\gamma)}{(1-\rho^2)^{1/2}} \right\} \right] \phi(x-\gamma)\, f_{n-p}(y)\, dx\, dy \\ &= C(\gamma,\rho), \end{aligned}$$

where the second line follows from the Lemma and the third from the fact that $\Phi(-x) = 1 - \Phi(x)$.

Proof: For the expected length.

The expected length and the scaled expected length are proportional to

$$L(\gamma, \rho) = \int_0^\infty \int_{-\infty}^\infty \{\delta_{1-\alpha/2}(x, y, \rho) - \delta_{\alpha/2}(x, y, \rho)\} \, \phi(x - \gamma) \, f_{n-p}(y) \, dx \, dy.$$

For each fixed ρ,

$$\begin{aligned}
L(-\gamma, \rho) &= \int_0^\infty \int_{-\infty}^\infty \{\delta_{1-\alpha/2}(x, y, \rho) - \delta_{\alpha/2}(x, y, \rho)\} \, \phi(x + \gamma) \, f_{n-p}(y) \, dx \, dy \\
&= \int_0^\infty \int_{-\infty}^\infty \{\delta_{1-\alpha/2}(-z, y, \rho) - \delta_{\alpha/2}(-z, y, \rho)\} \, \phi(-z + \gamma) \, f_{n-p}(y) \, dz \, dy \\
&= \int_0^\infty \int_{-\infty}^\infty \{-\delta_{\alpha/2}(x, y, \rho) + \delta_{1-\alpha/2}(z, y, \rho)\} \, \phi(z - \gamma) \, f_{n-p}(y) \, dz \, dy \\
&= L(\gamma, \rho).
\end{aligned}$$

The second line follows by changing the variable to $z = -x$ and the third follows by hypothesis and the fact that the standard normal density is an even function.

It follows from $\delta_u(x, y, -\rho) = \delta_u(-x, y, \rho)$ that

$$\begin{aligned}
L(\gamma, -\rho) &= \int_0^\infty \int_{-\infty}^\infty \{\delta_{1-\alpha/2}(-x, y, \rho) - \delta_{\alpha/2}(-x, y, \rho)\} \, \phi(x - \gamma) \, f_{n-p}(y) \, dx \, dy \\
&= \int_0^\infty \int_{-\infty}^\infty \{-\delta_{\alpha/2}(x, y, \rho) + \delta_{1-\alpha/2}(x, y, \rho)\} \, \phi(x - \gamma) \, f_{n-p}(y) \, dx \, dy \\
&= L(\gamma, \rho),
\end{aligned}$$

where the second line follows by hypothesis.

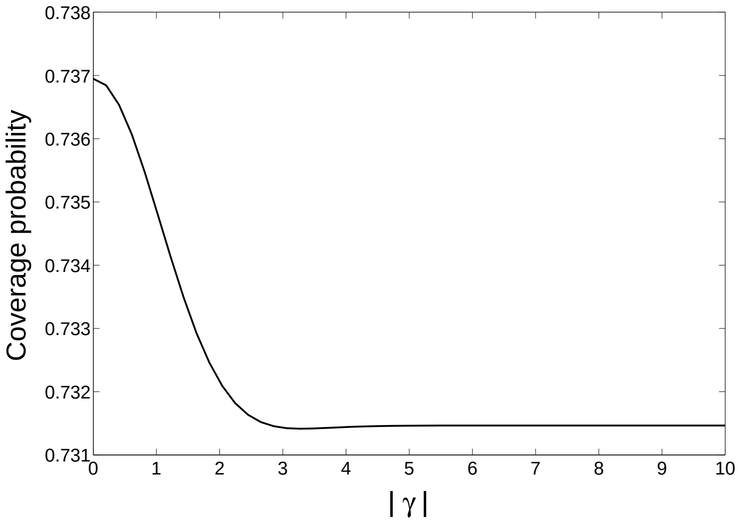

Figure 1: Plot of the coverage probability for MPI, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the submodel is defined by setting the coefficient of the squared seedability equal to zero.

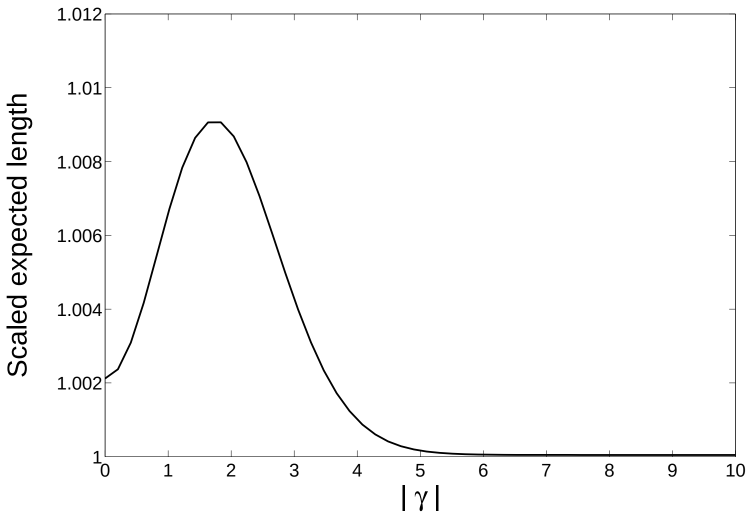

Figure 2: Plot of the scaled expected length for MPI, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the submodel is defined by setting the coefficient of the squared seedability equal to zero.

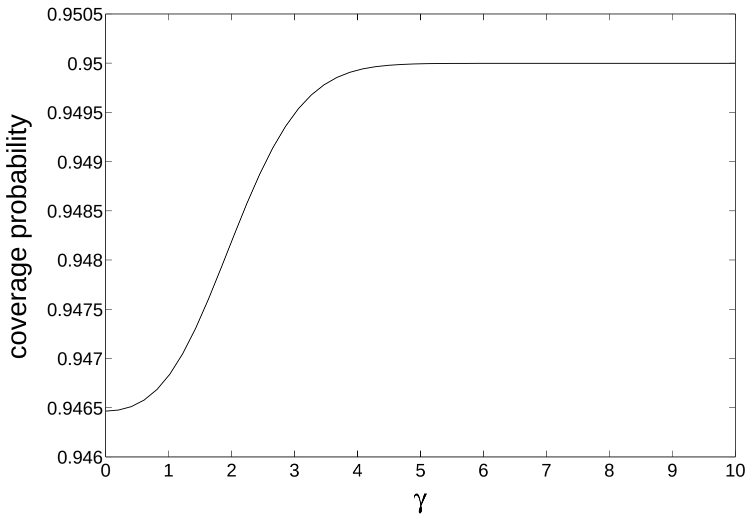

Figure 3: Plot of the coverage probability for MATA, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the submodel is defined by setting the coefficient of the squared seedability equal to zero.

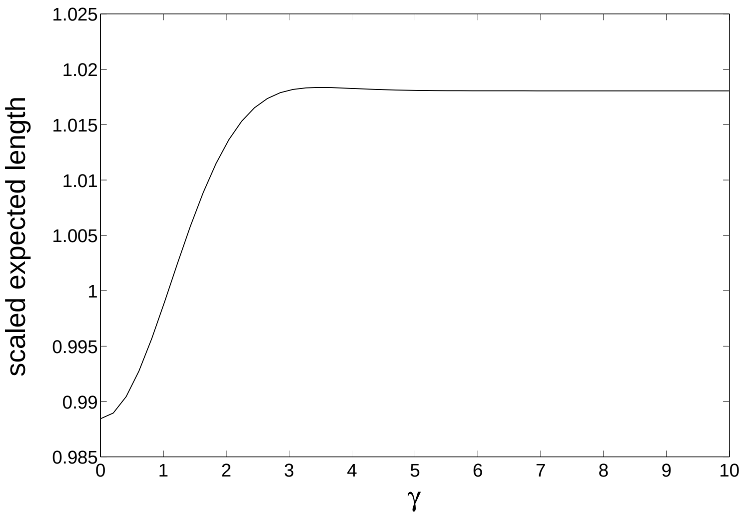

Figure 4: Plot of the scaled expected length for MATA, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the submodel is defined by setting the coefficient of the squared seedability equal to zero.

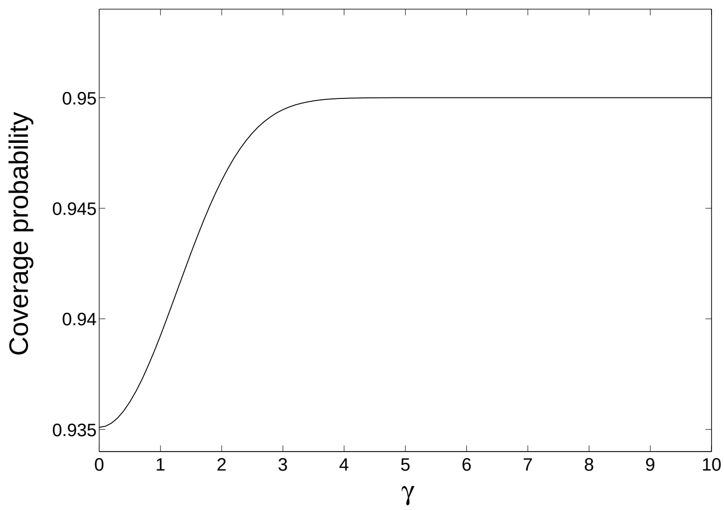

Figure 5: Plot of the coverage probability for the post-model-selection confidence interval, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the possible models are the full model and the submodel defined by setting the coefficient of the squared seedability equal to zero. The model selected is the model with smaller AIC.