The Performance of the Turek-Fletcher Model Averaged Confidence Interval

Paul Kabaila[1], A.H. Welsh[2] & Rheanna Mainzer[1]

December 7, 2015

Abstract

We consider the model averaged tail area (MATA) confidence interval proposed by Turek and Fletcher, *CSDA*, 2012, in the simple situation in which we average over two nested linear regression models. We prove that the MATA for *any* reasonable weight function belongs to the class of confidence intervals defined by Kabaila and Giri, *JSPI*, 2009. Each confidence interval in this class is specified by two functions b and s. Kabaila and Giri show how to compute these functions so as to optimize these intervals in terms of satisfying the coverage constraint and minimizing the expected length for the simpler model, while ensuring that the expected length has desirable properties for the full model. These Kabaila and Giri "optimized" intervals provide an upper bound on the performance of the MATA for an arbitrary weight function. This fact is used to evaluate the MATA for a broad class of weights based on exponentiating a criterion related to Mallows' C_P. Our results show that, while far from ideal, this MATA performs surprisingly well, provided that we choose a member of this class that does not put too much weight on the simpler model.

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

[1] *Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia*

[2] *Mathematical Sciences Institute, The Australian National University, Canberra ACT 0200, Australia*

1 Introduction

The Turek and Fletcher (2012) model averaged tail area confidence interval (MATA) endpoints are obtained by solving a weighted average of the tail area equations for the confidence interval endpoints for each model. Turek and Fletcher first consider giving a weight to each model by exponentiating AIC/2, where AIC is the Akaike Information Criterion for the model. This is the earliest form of weight used in frequentist model averaging and was first proposed by Buckland *et al.* (1997). Turek and Fletcher also consider MATA for weights obtained by replacing AIC by either the Akaike Information Criterion corrected for small samples, AIC_c, or the Bayesian Information Criterion, BIC. They provide some comparison of the performance of MATA for these three weights in particular examples using simulation.

We evaluate the performance of the MATA, with nominal coverage $1-\alpha$ and specified weight function, by focusing on the coverage and the scaled expected length, where the expected length is scaled with respect to the length of the standard confidence interval (based on the full model) with coverage equal to the minimum coverage probability of the MATA. We consider a simple situation in which we average over 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 ($\mathcal{M}_1$). Kabaila, Welsh and Abeysekera (2015) derive computationally convenient, exact expressions for the coverage probability and the scaled expected length of the MATA, so that we can readily obtain highly-accurate numerical results without resorting to simulations. In the same simple situation, we consider MATA with *any* reasonable weight function. We prove that the MATA with *any* reasonable weight function belongs to a subclass of the class of confidence intervals, denoted by $J(b,s)$, defined by Kabaila and Giri (2009). Each confidence interval in this class is specified by two functions: b and s. These authors show how to compute these functions so that the scaled expected length is minimized under model $\mathcal{M}_1$, subject to the constraints that (a) the coverage probability of this confidence interval never falls below $1-\alpha$, (b) the maximum scaled expected length under model $\mathcal{M}_2$ is not too large and (c) as the data becomes less consistent with the model $\mathcal{M}_1$, the scaled expected length approaches 1. They found that (to within computational accuracy)

the coverage probability of the resulting confidence interval is $1-\alpha$ throughout the parameter space. These Kabaila and Giri optimized confidence intervals then provide an upper bound on the performance of the MATA for any reasonable weight function. Knowing how the Kabaila and Giri optimized confidence intervals perform enables us to formulate four key scenarios within which we structure our examination of the performance of the MATA with specified weight. These scenarios are used to evaluate the MATA for a broad class of weights based on exponentiating criteria related to Mallows' C_P. Our results show that, while far from ideal, this MATA performs surprisingly well, provided that we choose a member of this class that does not put too much weight on the simpler model $\mathcal{M}_1$.

2 The MATA with a general weight function

2.1 The models

Suppose that 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 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. We write $m = n - p$ throughout the paper. 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 let $\widehat{\sigma}^2 = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})/m$ 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 $v_\theta = \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{a}$ and $v_\tau = \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c}$. Then 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 scaled unknown parameter $\gamma = \tau / \left(\sigma v_\tau^{1/2}\right)$. We denote the estimator of γ by $\widehat{\gamma} = \widehat{\tau} / \left(\widehat{\sigma} v_\tau^{1/2}\right)$.

2.2 The MATA

The MATA is obtained by averaging the equations defining the tail area confidence intervals under the models $\mathcal{M}_2$ and $\mathcal{M}_1$. Suppose that the weight function $w : [0, \infty) \to [0, 1]$ is a decreasing continuous function, such that $w(z)$ approaches 0 as $z \to \infty$. Any reasonable weight function must be of this form. For each a and $\gamma \in \mathbb{R}$, define

$$k(a, \gamma) = w(\gamma^2)\, G_{m+1}\left\{\left(\frac{m+1}{m+\gamma^2}\right)^{1/2} \frac{a - \rho\,\gamma}{(1-\rho^2)^{1/2}}\right\} + \left\{1 - w(\gamma^2)\right\} G_m(a),$$

where G_m denotes the distribution function of the Student t distribution with m degrees of freedom. The MATA, with nominal coverage $1 - \alpha$, is $\left[\widehat{\theta}_\ell,\, \widehat{\theta}_u\right]$, where $\widehat{\theta}_l$ and $\widehat{\theta}_u$ are the solutions for θ of

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

respectively. Equivalently, if we let $a_\ell(\gamma)$ and $a_u(\gamma)$ be the solutions for a of $k(a, \gamma) = 1 - \alpha/2$ and $k(a, \gamma) = \alpha/2$, respectively, then the endpoints of the MATA are given by

$$\begin{aligned} \widehat{\theta}_\ell &= \widehat{\theta} - v_\theta^{1/2} \widehat{\sigma}\, a_\ell(\widehat{\gamma}) \\ \widehat{\theta}_u &= \widehat{\theta} - v_\theta^{1/2} \widehat{\sigma}\, a_u(\widehat{\gamma}). \end{aligned} \tag{1}$$

The notation is slightly different from that used in Kabaila, Welsh and Abeysekara (2015), but the interval is the same.

2.3 The Kabaila and Giri optimized confidence intervals

The MATA, with nominal coverage $1 - \alpha$, is a member of the class of confidence intervals $J(b, s)$ defined by Kabaila and Giri (2009). To see this, note that the centre and half-width of the MATA (1) are $\widehat{\theta} - v_\theta^{1/2} \widehat{\sigma} \left\{a_\ell(\widehat{\gamma}) + a_u(\widehat{\gamma})\right\}/2$ and $v_\theta^{1/2} \widehat{\sigma} \left\{a_\ell(\widehat{\gamma}) - a_u(\widehat{\gamma})\right\}/2$, respectively. As shown in Appendix A, $a_\ell(\gamma) + a_u(\gamma)$ and $a_\ell(\gamma) - a_u(\gamma)$ are odd and even functions of γ, respectively, and $a_\ell(\gamma) + a_u(\gamma)$ and $a_\ell(\gamma) - a_u(\gamma)$ approach 0 and $2G_m^{-1}(1 - \alpha/2)$, respectively, as $\gamma \to \infty$. Now define the functions b and s by $b(\gamma) = \{a_\ell(\gamma) + a_u(\gamma)\}/2$ and $s(\gamma) = \{a_\ell(\gamma) - a_u(\gamma)\}/2$. Then the MATA,

with nominal coverage $1-\alpha$, can be written as

$$\Big[\widehat{\theta} - v_\theta^{1/2}\widehat{\sigma}\, b(\widehat{\gamma}) - v_\theta^{1/2}\widehat{\sigma}\, s(|\widehat{\gamma}|),\ \widehat{\theta} - v_\theta^{1/2}\widehat{\sigma}\, b(\widehat{\gamma}) + v_\theta^{1/2}\widehat{\sigma}\, s(|\widehat{\gamma}|)\Big],$$

which is of the form $J(b,s)$ considered by Kabaila and Giri (2009). As proved by these authors, for given $1-\alpha$ and given functions b and s, the coverage probability and the scaled expected length of $J(b,s)$ are functions of the known quantities (m,ρ) and the unknown parameter γ. Since Kabaila and Giri (2009) optimized the choice of b and s separately for each given value of (m,ρ), we cannot expect that, for any given weight function w, the MATA will perform better than the optimized confidence interval of Kabaila and Giri (2009).

3 Weight functions based on Mallows' C_P

For the models $\mathcal{M}_2$ and $\mathcal{M}_1$ the Generalized Information Criteria (GIC; Nishii, 1984, Rao and Wu, 1989) are

$$\mathrm{GIC}_2 = n\log\{m\widehat{\sigma}^2/n\} + dp$$

and

$$\mathrm{GIC}_1 = n\log[\{(\widehat{\tau}^2/v_\tau) + m\widehat{\sigma}^2\}/n] + d(p-1),$$

respectively, where d is a specified nonnegative number. The choices $d=2$ and $d=\log(n)$ yield AIC and BIC, respectively. The corresponding weight function is

$$\begin{aligned} w_*\big(\widehat{\gamma}^2; d\big) &= \frac{\exp\big\{-\frac{1}{2}(\mathrm{GIC}_1 - \mathrm{GIC}_{\min})\big\}}{\exp\big\{-\frac{1}{2}(\mathrm{GIC}_1 - \mathrm{GIC}_{\min})\big\} + \exp\big\{-\frac{1}{2}(\mathrm{GIC}_2 - \mathrm{GIC}_{\min})\big\}} \\ &= \frac{1}{1 + \Big\{1 + \widehat{\gamma}^2/m\Big\}^{n/2}\exp(-d/2)}, \end{aligned}$$

where $\mathrm{GIC}_{\min} = \min(\mathrm{GIC}_1, \mathrm{GIC}_2)$.

We now motivate the use of this weight function, with the power $n/2$ replaced by $m/2$. Define the criteria

$$\mathrm{MIC}_2 = m\log\{m\widehat{\sigma}^2/n\} + dp$$

and

$$\mathrm{MIC}_1 = m\log[\{(\widehat{\tau}^2/v_\tau) + m\widehat{\sigma}^2\}/n] + d(p-1),$$

for the models $\mathcal{M}_2$ and $\mathcal{M}_1$, respectively. As we show in Appendix C, choosing $\mathcal{M}_2$ if and only if (Mallows' C_P for $\mathcal{M}_2$) $\leq$ (Mallows C_P for $\mathcal{M}_1$) is equivalent to choosing $\mathcal{M}_2$ if and only if $\mathrm{MIC}_2 \leq \mathrm{MIC}_1$, for $d = m\log(1 + (2/m))$. The criteria MIC_2 and MIC_1 lead to the weight function

$$\begin{aligned} w(\widehat{\gamma}^2; d) &= \frac{\exp\left\{-\frac{1}{2}(\mathrm{MIC}_1 - \mathrm{MIC}_{\min})\right\}}{\exp\left\{-\frac{1}{2}(\mathrm{MIC}_1 - \mathrm{MIC}_{\min})\right\} + \exp\left\{-\frac{1}{2}(\mathrm{MIC}_2 - \mathrm{MIC}_{\min})\right\}} \\ &= \frac{1}{1 + \left(1 + \widehat{\gamma}^2/m\right)^{m/2}\exp(-d/2)}, \end{aligned}$$

where $\mathrm{MIC}_{\min} = \min(\mathrm{MIC}_1, \mathrm{MIC}_2)$. The criteria MIC_2 and MIC_1 are close to the criteria GIC_2 and GIC_1, respectively, for p fixed and n large, but replacing $n/2$ by $m/2$ achieves a useful simplification; the coverage probability and scaled expected length of the MATA, with nominal coverage $1 - \alpha$, are determined by the known quantities (d, m, ρ) and the unknown parameter γ (as in Kabaila and Giri, 2009) rather than by (d, n, p, ρ) and γ (as shown in Theorems 1 and 2 of Kabaila, Welsh and Abeysekera, 2015). The reduction from the 4 known quantities (d, n, p, ρ) to the 3 known quantities (d, m, ρ) represents a considerable gain in simplicity.

As also noted in Appendix C, using the MIC to choose between the models $\mathcal{M}_1$ and $\mathcal{M}_2$ is equivalent to testing the null hypothesis $\tau = 0$ (i.e. $\mathcal{M}_1$) against the alternative hypothesis $\tau \neq 0$ with level of significance

$$2\left(1 - G_m\left[m^{1/2}\left\{\exp\left(\frac{d}{m}\right) - 1\right\}^{1/2}\right]\right).$$

Large values of d correspond to small values of this level of significance and so can be interpreted as putting more weight on the simpler model $\mathcal{M}_1$.

The interpretation of d is quite different in the model selection and model averaging contexts. In the model selection context, setting $d = 0$ means that there is no penalty on the number of parameters and so, for the models $\mathcal{M}_1$ and $\mathcal{M}_2$, we always choose model $\mathcal{M}_2$. By contrast, in the model averaging context, setting $d = 0$ leads to $w(\widehat{\gamma}^2; 0) = 1/\left\{1 + (1 + \widehat{\gamma}^2/m)^{m/2}\right\}$ so we still average over the two models.

4 How well can we expect MATA to perform?

The performance of the Kabaila and Giri (2009, 2013) optimized confidence interval relative to the standard $1-\alpha$ Student t confidence interval under model $\mathcal{M}_2$ can be described under four different scenarios defined by the values of m and ρ, as set out in Table 1. The MATA cannot perform better than the Kabaila and Giri optimized confidence interval so, for each of the four scenarios, we compare the coverage and scaled expected length properties of the MATA against the best we can hope for from the Kabaila and Giri optimized confidence interval. Details of the methods used to compute the coverage and scaled expected length of the MATA are given in Appendix A.

	$\lvert\rho\rvert$ is small	$\lvert\rho\rvert$ is not small
m is not small	**Scenario 1** Cannot do better than the standard $1-\alpha$ t interval for θ	**Scenario 2** Some improvement over the standard $1-\alpha$ t interval for θ
m is small (i.e. 1, 2 or 3)	**Scenario 3** Some improvement over the standard $1-\alpha$ t interval for θ	**Scenario 4** Maximum improvement over standard $1-\alpha$ t interval for θ

Table 1: Performance of the optimized confidence interval of Kabaila and Giri (2009) for various values of the known quantities m and ρ.

In all our numerical work, we fix the nominal coverage $1-\alpha = 0.95$ and vary the values of ρ and m according to the different scenarios (other than the first which we are able to treat theoretically). As in Kabaila, Welsh and Abeysekera (2015), for each scenario we can compute and plot the coverage and scaled expected length of the MATA for fixed d as functions of γ. Typically, for fixed d, the coverage is greater than 0.95, decreases to a minimum value and then increases to asymptote at 0.95 as γ increases; the scaled expected length is less than one for $\gamma = 0$, increases to a maximum and then decreases to an asymptote (often greater than one) as γ

increases. Some examples of these kinds of figures are included in Kabaila, Welsh and Abeysekera (2015). For our present purposes, it is useful to summarise the above results over different choices of $d \in [0, 8]$ by presenting the minimum coverage and maximum scaled expected length over γ for each fixed d and the scaled expected length at $\gamma = 0$ for each fixed d as functions of d. We make these quantities positive with a baseline value of zero by computing the coverage loss (*cov loss*), scaled expected length loss (*sel loss*) and the scaled expected length gain (*sel gain*), defined as follows

$$\begin{aligned} \textit{cov loss} &= (1-\alpha) - (\text{minimum coverage}) \\ \textit{sel loss} &= (\text{maximum scaled expected length}) - 1 \\ \textit{sel gain} &= 1 - (\text{scaled expected length at } \gamma = 0). \end{aligned}$$

Ideally, one would have a high *sel gain* together with small *cov loss* and *sel loss*.

4.1 Scenario 1

This scenario includes the cloud seeding example in Section 3 of Kabaila, Welsh and Abeysekera (2015) where $\rho = 0.2472$ and $m = 11$. For **Scenario 1** we cannot expect the MATA, with nominal coverage $1-\alpha$, to perform any better than the standard $1-\alpha$ confidence interval for θ based on model $\mathcal{M}_2$ so the best hope is that MATA recovers the standard $1-\alpha$ confidence interval for θ based on model $\mathcal{M}_2$. That is, that

$$k(a, \widehat{\gamma}) \approx G_m(a). \tag{2}$$

Indeed, if $|\rho|$ is small, then

$$k(a, \widehat{\gamma}) \approx w(\widehat{\gamma}^2) G_{m+1}\left\{\left(\frac{m+1}{m+\widehat{\gamma}^2}\right)^{1/2} a\right\} + \{1 - w(\widehat{\gamma}^2)\} G_m(a).$$

If, in addition, m is not small, then $G_{m+1} \approx G_m$ and, since w is a decreasing continuous function, such that $w(z)$ approaches 0 as $z \to \infty$,

$$k(a, \widehat{\gamma}) \approx w(\widehat{\gamma}^2) G_m(a) + \{1 - w(\widehat{\gamma}^2)\} G_m(a) = G_m(a),$$

so (2) holds. In Scenario 1, for given ρ and m, as d increases MATA is less likely to recover the standard $1-\alpha$ confidence interval for θ based on model $\mathcal{M}_2$. So, in Scenario 1 a good choice of d is 0.

4.2 Scenario 2

Graphs of *cov loss*, *sel loss* and *sel gain* as functions of $d \in [0, 4]$ for $\rho = 0.8$ and $m \in \{10, 50, 200\}$ are shown in Figure 1. These functions are displayed only for $d \in [0, 4]$ because *cov loss* and *sel loss* are both increasing functions and *sel gain* is a decreasing function of d in $[4, 8]$. In other words, when searching for a good value of d there is no point in considering values of d in $[4, 8]$. For $m = 10$, as we increase d from 0 to 4, the *cov loss* is an increasing function of d, *sel gain* changes slowly and *sel loss* increases. In this circumstance, a good choice of d is 0. Similarly, for $m = 50$ and $m = 200$, a good choice of d is 0. These results show that there is not much gain from using the MATA in Scenario 2.

4.3 Scenario 3

Graphs of *cov loss*, *sel loss* and *sel gain* as functions of $d \in [0, 4]$ for $\rho = 0$ and $m \in \{1, 2, 3\}$ are shown in Figure 2. For $m = 2$ and $m = 3$, the *cov loss* is a decreasing function of d in $[4, 8]$. Also, for $m = 1$, the *cov loss* is initially an increasing function and then a decreasing function of d in $[4, 8]$. However, the *cov loss* remains small for all d in $[0, 8]$ and *sel loss* is an increasing function of d in $[0, 8]$. Therefore, when searching for a good value of d, it seems reasonable to restrict consideration to values of d in $[4, 8]$. For $m = 1$, as we increase d from 0 to 4, the *sel gain* increases slowly and *sel loss* increases much more quickly. In this circumstance, a good choice of d is 0. Similarly, for $m = 2$ and $m = 3$, a good choice of d is 0. In Scenario 3, there is a small gain from using the MATA, provided that we choose d near 0.

4.4 Scenario 4

Graphs of *cov loss*, *sel loss* and *sel gain* as functions of $d \in [0, 4]$ for $\rho = 0.8$ and $m \in \{1, 2, 3\}$ are shown in Figure 3. These functions are displayed only for $d \in [0, 4]$ because *cov loss* and *sel loss* are both increasing functions and *sel gain* is a decreasing function of d in $[4, 8]$. In other words, when searching for a good value of d there is no point in considering values of d in $[4, 8]$. For $m = 1$, as we increase d from 0 to 4, the *cov loss* is a nondecreasing function of d, *sel gain* increases slowly and

sel loss increases much more quickly. In this circumstance, a good choice of d is 0. Similarly, for $m = 2$ and $m = 3$, a good choice of d is 0.

5 Conclusion

We have examined the exact coverage and scaled expected length of the MATA for a parameter θ, with nominal level $1 - \alpha$, in a particular simple situation in which there are two linear regression models (differing in only a single parameter τ) to average over. For weight functions based on Mallows' C_P (specified by $d \in [0, 8]$), we showed that both the coverage and the scaled expected length depend on $m = n-p$, the correlation ρ between the least squares estimators $\widehat{\theta}$ and $\widehat{\tau}$, and the unknown parameter $\gamma = \tau/\left(\sigma v_\tau^{1/2}\right)$. We assess the MATA according to two losses, using the minimum coverage and the maximum scaled expected length over γ and a gain, using the scaled expected length at $\gamma = 0$ i.e. when the simpler model $\mathcal{M}_1$ is true. The results show that, although the MATA is far from ideal, it performs surprisingly well, provided that we do not put too much weight on the model $\mathcal{M}_1$.

References

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

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

Kabaila, P. and 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., Welsh, A.H. and Abeysekera, W. (2015). Model averaged confidence intervals. To appear in *Scandinavian Journal of Statistics*.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics*, **12**, 758–765.

Rao, C.R. and Wu, Y. (1989). A strongly consistent procedure for model selection in regression problems. *Biometrika*, **76**, 369–374.

Turek, D. and Fletcher, D. (2012). Model-averaged Wald confidence intervals. *Computational Statistics and Data Analysis*, **56**, 2809–2815.

Appendix A: The MATA is in the class of confidence intervals $J(b, s)$

We show that $a_\ell(\gamma) + a_u(\gamma)$ and $a_\ell(\gamma) - a_u(\gamma)$ are odd and even functions of γ, respectively, as follows. Recall that $a_u(-\gamma)$ is the solution for a of $k(a, -\gamma) = \alpha/2$ i.e. $a_u(-\gamma)$ is the solution for a of

$$w(\gamma^2)\, G_{m+1}\left\{\left(\frac{m+1}{m+\gamma^2}\right)^{1/2} \frac{a+\rho\,\gamma}{(1-\rho^2)^{1/2}}\right\} + \left\{1 - w(\gamma^2)\right\} G_m(a) = \frac{\alpha}{2}.$$

Using the fact that the probability density function of a Student t distribution is an even function, we can show that

$$\begin{aligned}
&w(\gamma^2)\, G_{m+1}\left\{\left(\frac{m+1}{m+\gamma^2}\right)^{1/2} \frac{a_u(-\gamma)+\rho\,\gamma}{(1-\rho^2)^{1/2}}\right\} + \left\{1 - w(\gamma^2)\right\} G_m\{a_u(-\gamma)\} \\
&\quad = w(\gamma^2)\, \left[1 - G_{m+1}\left\{-\left(\frac{m+1}{m+\gamma^2}\right)^{1/2} \frac{a_u(-\gamma)+\rho\,\gamma}{(1-\rho^2)^{1/2}}\right\}\right] \\
&\qquad + \left\{1 - w(\gamma^2)\right\} \left[1 - G_m\{-a_u(-\gamma)\}\right] \\
&\quad = 1 \; - w(\gamma^2) G_{m+1}\left\{\left(\frac{m+1}{m+\gamma^2}\right)^{1/2} \frac{-a_u(-\gamma)-\rho\,\gamma}{(1-\rho^2)^{1/2}}\right\} \\
&\qquad - \left\{1 - w(\gamma^2)\right\} G_m\{-a_u(-\gamma)\}
\end{aligned}$$

so $a_u(-\gamma) = -a_\ell(\gamma)$. Thus $a_\ell(-\gamma) = -a_u(\gamma)$ and it follows that $a_\ell(\gamma) + a_u(\gamma)$ and $a_\ell(\gamma) - a_u(\gamma)$ are odd and even functions of γ, respectively.

We now show that $a_\ell(\gamma)+a_u(\gamma)$ and $a_\ell(\gamma)-a_u(\gamma)$ approach 0 and $2G_m^{-1}(1-\alpha/2)$, respectively, as $\gamma \to \infty$. As $\gamma \to \infty$, $k(a, \gamma)$ approaches $G_m(a)$. Therefore $a_\ell(\gamma)$ and $a_u(\gamma)$ approach the solutions for a of $G_m(a) = 1 - \alpha/2$ and $G_m(a) = \alpha/2$, respectively. Thus $a_\ell(\gamma) + a_u(\gamma)$ and $a_\ell(\gamma) - a_u(\gamma)$ approach 0 and $2G_m^{-1}(1 - \alpha/2)$, respectively, as $\gamma \to \infty$.

Appendix B: Computational methods

The computation of the coverage and the scaled expected length of the MATA are implemented using Theorems 1 and 2 of Kabaila, Welsh and Abeysekera (2015). In the following two sections, we establish results which are needed to support the computations.

Truncation of integrals

To compute the coverage and the scaled expected length of the MATA, we need to truncate the integrals in Theorems 1 and 2 of Kabaila, Welsh and Abeysekera (2015). The coverage probability integral in Theorem 1 is relatively easy to handle because cumulative distribution functions are bounded. If we write the integrand in Theorem 1 as $\Delta_m(x,y,\rho,\gamma)$, we have

$$\begin{aligned}
&\left|\Pr(\widehat{\theta}_\ell \le \theta \le \widehat{\theta}_u) - \int_{y_\ell}^{y_u}\int_{x_\ell}^{x_u} \Delta_m(x,y,\rho,\gamma)dxdy\right| \\
&\quad \le \left|\int_0^\infty\int_{-\infty}^\infty \Delta_m(x,y,\rho,\gamma)dxdy - \int_{y_\ell}^{y_u}\int_{-\infty}^\infty \Delta_m(x,y,\rho,\gamma)dxdy\right| \\
&\quad + \left|\int_{y_\ell}^{y_u}\int_{-\infty}^\infty \Delta_m(x,y,\rho,\gamma)dxdy - \int_{y_\ell}^{y_u}\int_{x_\ell}^{x_u} \Delta_m(x,y,\rho,\gamma)dxdy\right| \\
&\quad \le \int_{y_u}^\infty\int_{-\infty}^\infty |\Delta_m(x,y,\rho,\gamma)|dxdy + \int_0^{y_\ell}\int_{-\infty}^\infty |\Delta_m(x,y,\rho,\gamma)|dxdy \\
&\quad + \int_{y_\ell}^{y_u}\int_{x_u}^\infty |\Delta_m(x,y,\rho,\gamma)|dxdy + \int_{y_\ell}^{y_u}\int_{-\infty}^{x_\ell} |\Delta_m(x,y,\rho,\gamma)|dxdy.
\end{aligned}$$

Now $|\Delta_m(x,y,\rho,\gamma)| \le \phi(x-\gamma)g_m(y)$ so

$$\int_{y_u}^\infty\int_{-\infty}^\infty |\Delta_m(x,y,\rho,\gamma)|dxdy \le 1 - G_m(y_u),$$

$$\int_0^{y_\ell}\int_{-\infty}^\infty |\Delta_m(x,y,\rho,\gamma)|dxdy \le G_m(y_\ell),$$

$$\int_{y_\ell}^{y_u}\int_{x_u}^\infty |\Delta_m(x,y,\rho,\gamma)|dxdy \le \{G_m(y_u) - G_m(y_\ell)\}\{1-\Phi(x_u-\gamma)\}$$

and

$$\int_{y_\ell}^{y_u}\int_{-\infty}^{x_\ell} |\Delta_m(x,y,\rho,\gamma)|dxdy \le \{G_m(y_u) - G_m(y_\ell)\}\Phi(x_\ell-\gamma).$$

For any given small positive number ϵ, set $y_\ell = G_m^{-1}(\epsilon/4)$ and $y_u = G_m^{-1}(1-\epsilon/4)$ so the first two terms are both less than or equal to $\epsilon/4$. If we also set $x_\ell = \Phi^{-1}(\epsilon/4)+\gamma$ and $x_u = \Phi^{-1}(1-\epsilon/4)+\gamma$, the last two terms are both less than or equal to

$(1-\epsilon/2)\epsilon/4 \leq \epsilon/4$ and the sum of all the terms is less than or equal to ϵ. Thus the effect of the truncation can be made as small as we want.

The scaled expected length is more difficult to handle because the integrand is unbounded. Nonetheless, $\delta_{1-\alpha/2}(x,y) - \delta_{\alpha/2}(x,y)$ is approximately constant in x and linear in y and we can use this approximation to similarly find truncation values for the integrals in Theorem 2 so that the effect of the truncation is arbitrarily small.

Computation of $\delta_u(x,y)$

The formulae for the coverage probability and the scaled expected length of the MATA, with nominal coverage probability $1-\alpha$, given by Kabaila, Welsh and Abeysekera (2015) require the computation of $\delta_u(x,y)$, which is defined to be the solution for δ of

$$h(\delta,x,y) = u,$$

where $0 < u < 1$ and

$$h(\delta,x,y) = w(x^2/y^2)\, G_{m+1}(r_1(\delta,x,y)) + \{1 - w(x^2/y^2)\}\, G_m(r_2(\delta,y)). \qquad (3)$$

The definitions of $r_1(\delta,x,y)$ and $r_2(\delta,y)$ are given on p.6 of Kabaila, Welsh and Abeysekera (2015). The numerical computation of $\delta_u(x,y)$ typically requires the provision of an interval known to contain $\delta_u(x,y)$. The following result provides just such an interval.

Result A1. Let $\delta_u^{(1)}(x,y)$ denote the solution for δ of

$$G_{m+1}(r_1(\delta,x,y)) = u.$$

Also, let $\delta_u^{(2)}(y)$ denote the solution for δ of

$$G_m(r_2(\delta,y)) = u.$$

The following are explicit expressions for $\delta_u^{(1)}(x,y)$ and $\delta_u^{(2)}(y)$:

$$\delta_u^{(1)}(x,y) = \rho\, x + G_{m+1}^{-1}(u)\, y \left(\frac{m+(x^2/y^2)}{m+1}\right)^{1/2} (1-\rho^2)^{1/2}$$

$$\delta_u^{(2)}(y) = G_m^{-1}(u)\, y.$$

Then

$$\delta_u(x,y) \in \left[\min\left(\delta_u^{(1)}(x,y), \delta_u^{(2)}(y)\right), \max\left(\delta_u^{(1)}(x,y), \delta_u^{(2)}(y)\right)\right] \tag{4}$$

Proof Suppose that (x, y) is given. Since $r_1(\delta, x, y)$ is a continuous increasing function of δ, $G_{m+1}(r_1(\delta, x, y))$ is also a continuous increasing function of δ. Also, since $r_2(\delta, y)$ is a continuous increasing function of δ, $G_m(r_2(\delta, y))$ is also a continuous increasing function of δ. It follows from (3) that $h(\delta, x, y)$ is a continuous increasing function of δ and that

$$\min\left(G_{m+1}(r_1(\delta,x,y)), G_m(r_2(\delta,y))\right) \le h(\delta,x,y) \le \max\left(G_{m+1}(r_1(\delta,x,y)), G_m(r_2(\delta,y))\right).$$

It follows from this that (4) is satisfied.

□

Appendix C: Weights based on Mallows' C_P

Let $\text{RSS} = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = m\widehat{\sigma}^2$. Also, let $\text{RSS}_* = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_*)^T(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_*)$, where $\widehat{\boldsymbol{\beta}}_*$ denotes the least squares estimator of $\boldsymbol{\beta}$ subject to the constraint $\tau = 0$. Note that $\text{RSS}_* = \widehat{\tau}^2/v_\tau + m\widehat{\sigma}^2$. Mallows' C_P for $\mathcal{M}_2$ is

$$\frac{\text{RSS}}{\text{RSS}/m} - n + 2p = n - p - n + 2p = p$$

and Mallows' C_P for $\mathcal{M}_1$ is

$$\frac{\text{RSS}_*}{\text{RSS}/m} - n + 2(p-1) = \frac{(\widehat{\tau}^2/v_\tau) + m\widehat{\sigma}^2}{\widehat{\sigma}^2} - n + 2p - 1.$$

Result C1. Choosing $\mathcal{M}_2$ if and only if (Mallows' C_P for $\mathcal{M}_2$) $\le$ (Mallows C_P for $\mathcal{M}_1$) is equivalent to choosing $\mathcal{M}_2$ if and only if $\text{MIC}_2 \le \text{MIC}_1$, for $d = m\log(1 + (2/m))$.

Proof: The result follows from the fact that (Mallows' C_P for $\mathcal{M}_2$) $\le$ (Mallows C_P for $\mathcal{M}_1$) is equivalent to the following four statements

$$m + 2 \le \frac{(\widehat{\tau}^2/v_\tau) + m\widehat{\sigma}^2}{\widehat{\sigma}^2},$$

$$m\log(\widehat{\sigma}^2/n) + m\log(m+2) \le m\log[\{(\widehat{\tau}^2/v_\tau) + m\widehat{\sigma}^2\}/n],$$

$$m\log(\widehat{\sigma}^2/n) + m\log(m) + m\log(1+2/m) \le m\log[\{(\widehat{\tau}^2/v_\tau) + m\widehat{\sigma}^2\}/n],$$

$$m\log(m\widehat{\sigma}^2/n) \le m\log[\{(\widehat{\tau}^2/v_\tau) + m\widehat{\sigma}^2\}/n] - d,$$

where $d = m\log(1 + (2/m))$, and the last of these is equivalent to $\mathrm{MIC}_2 \leq \mathrm{MIC}_1$.

Result C2. Choosing $\mathcal{M}_2$ if and only if $\mathrm{MIC}_2 \leq \mathrm{MIC}_1$ is equivalent to testing the null hypothesis $\tau = 0$ against the alternative hypothesis $\tau \neq 0$ with level of significance

$$2\left(1 - G_m\left[m\left\{\exp\left(\frac{d}{m}\right) - 1\right\}^{1/2}\right]\right).$$

Proof: The null hypothesis $H_0 : \tau = 0$ corresponds to choosing $\mathcal{M}_1$ so the level of significance of the test is the probability of choosing $\mathcal{M}_2$ under the null hypothesis H_0 and

$$\begin{aligned}
&\Pr(\mathrm{MIC}_1 \geq \mathrm{MIC}_2) \\
&\quad = \Pr\left(m\log[\{(\widehat{\tau}^2/v_\tau) + m\widehat{\sigma}^2\}/n] - d \geq m\log(m\widehat{\sigma}^2/n)\right) \\
&\quad = \Pr\left\{\frac{(\widehat{\tau}^2/v_\tau) + m\widehat{\sigma}^2}{m\widehat{\sigma}^2} \geq \exp\left(\frac{d}{m}\right)\right\} \\
&\quad = \Pr\left[\widehat{\gamma}^2 \geq m\left\{\exp\left(\frac{d}{m}\right) - 1\right\}\right] \\
&\quad = 1 - \Pr\left[-m^{1/2}\left\{\exp\left(\frac{d}{m}\right) - 1\right\}^{1/2} \leq \widehat{\gamma} \leq m^{1/2}\left\{\exp\left(\frac{d}{m}\right) - 1\right\}^{1/2}\right] \\
&\quad = 1 - G_m\left[m^{1/2}\left\{\exp\left(\frac{d}{m}\right) - 1\right\}^{1/2}\right] + G_m\left[-m^{1/2}\left\{\exp\left(\frac{d}{m}\right) - 1\right\}^{1/2}\right] \text{ under } H_0 \\
&\quad = 2\left(1 - G_m\left[m^{1/2}\left\{\exp\left(\frac{d}{m}\right) - 1\right\}^{1/2}\right]\right) \text{ under } H_0.
\end{aligned}$$

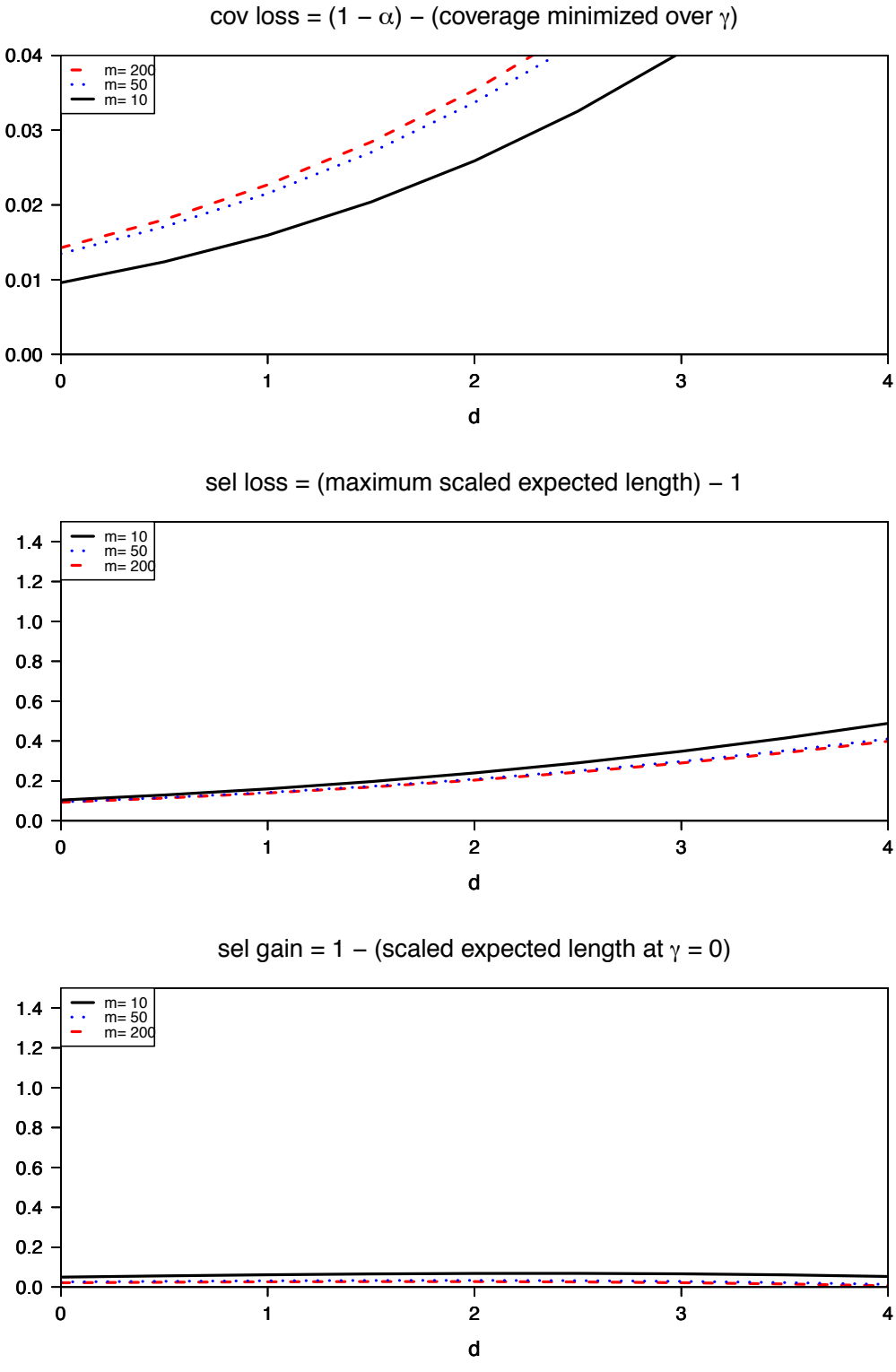

Figure 1: Graphs of the *cov loss*, *sel loss* and *sel gain* of the MATA, with weight determined by d and nominal coverage 95%. Here, $\rho = \mathrm{Corr}(\widehat{\theta}, \widehat{\tau}) = 0.8$ and $m \in \{10, 50, 200\}$.

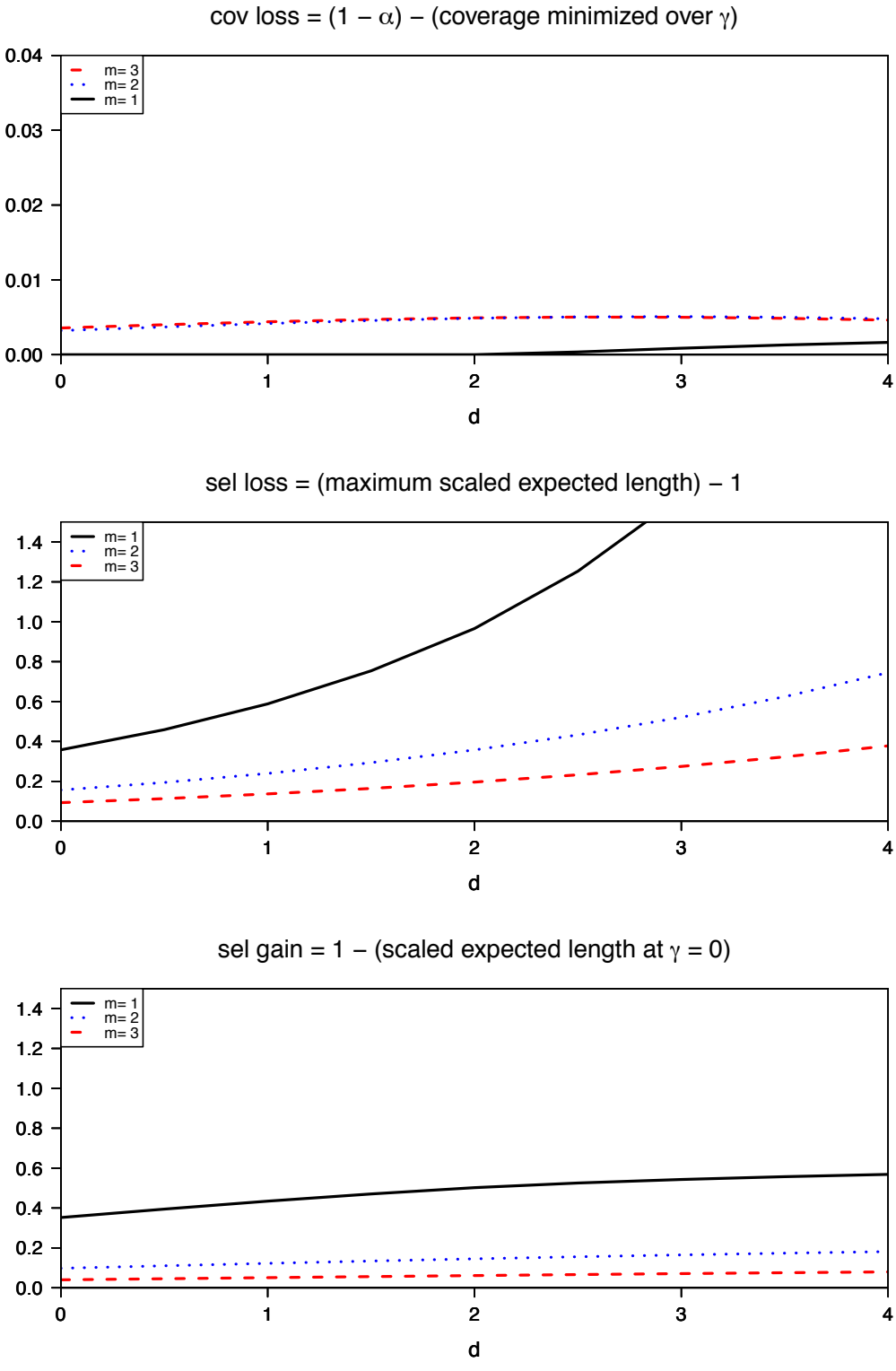

Figure 2: Graphs of the *cov loss*, *sel loss* and *sel gain* of the MATA, with weight determined by d and nominal coverage 95%. Here, $\rho = \mathrm{Corr}(\widehat{\theta}, \widehat{\tau}) = 0$ and $m \in \{1, 2, 3\}$.

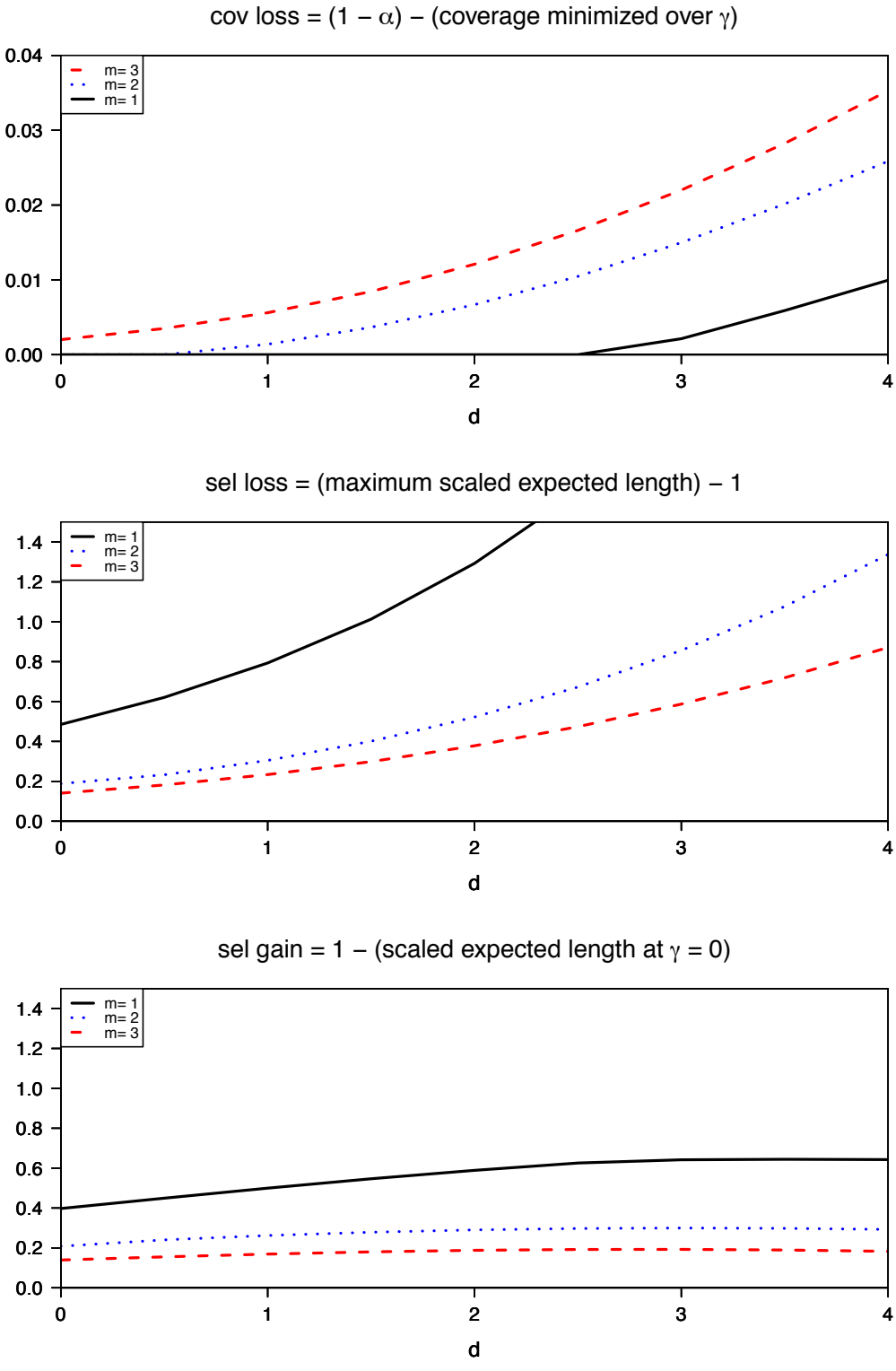

Figure 3: Graphs of the *cov loss*, *sel loss* and *sel gain* of the MATA, with weight determined by d and nominal coverage 95%. Here, $\rho = \mathrm{Corr}(\widehat{\theta}, \widehat{\tau}) = 0.8$ and $m \in \{1, 2, 3\}$.