Decision Theory for the Variance Ratio in One-Way ANOVA with Random Effects

Teoría de la decisión para la relación de las varianzas en el análisis de la varianza de un factor con efectos aleatorios

Nicholas T. Longford[1,a], Mercedes Andrade[2,b]

[1]SNTL and Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain

[2]School of Statistics, Universidad del Valle, Cali, Colombia

---

Abstract

Estimating a variance component in the model of analysis of variance with random effects and testing the hypothesis that the variance vanishes are important issues in many applications. Such inferences are beyond the confines of the standard (asymptotic) theory because a zero variance is on the boundary of the parameter space and the maximum likelihood or another reasonable estimator of variance has a non-trivial probability of zero in many settings. We derive decision rules regarding the variance ratio in balanced one-way analysis of variance, in both the frequentist and Bayesian perspectives. We argue that this approach is superior to hypothesis testing because it incorporates the consequences of the two kinds of error (incorrect choice) that may be committed. An application to a track athlete's training performance is presented.

Key words: Analysis of Variance with Random Effects, Decision, Equilibrium, Expected Loss, Variance Ratio.

Resumen

La estimación de una de las varianzas en el modelo de análisis de la varianza con efectos aleatorios y la prueba de hipótesis de que la varianza se anula, son temas importantes en muchas aplicaciones. Tales inferencias están fuera de los confines de la teoría asintótica estándar porque una varianza cero está en la frontera del espacio paramétrico y la máxima verosimilitud u otro estimador razonable de una varianza tiene una probabilidad no trivial de cero en muchos contextos. Nosotros derivamos una regla de decisión sobre la

---

[a]Professor. E-mail: sntlnick@sntl.co.uk

[b]Professor. E-mail: mercedes.andrade@correounivalle.edu.co

razón de varianzas en un análisis de varianza de un factor balanceado tanto para la perspectiva frecuentista como la Bayesiana. Argumentamos que este enfoque es superior a la prueba de hipótesis porque incorpora las consecuencias de los dos tipos de error (elección incorrecta) que pueden cometerse. Se presenta una aplicación sobre los rendimientos de los entrenamientos de un atleta de pista.

Palabras clave: análisis de varianza con efectos aleatorios, decisión, equilibrio, pérdida esperada, razón de la varianza.

1. Introduction

Testing the hypothesis that between-cluster variance vanishes in the mixed model of one-way analysis of variance (ANOVA) and its extensions has received considerable attention recently (Crainiceanu & Ruppert 2004, Greven, Crainiceanu, Küchenhoff & Peters 2008, Giampaoli & Singer 2009, Andrade, Longford & Tovar 2014). The principal findings in these references are that asymptotic theory, or its adaptation for the non-standard nature of the inferential problem, provides a poor approximation for small and moderately large samples, and that the likelihood ratio test statistic has a distribution well approximated by a mixture of the constant zero and one or several χ^2 distributions. The mixture probabilities are specific to the setting (design).

We regard hypothesis testing as problematic in general, because it has no means of incorporating the consequences of the two kinds of error (bad choice) that may be committed (Longford 2005, 2012b). We follow up on this criticism by solving the decision-theoretical version of the problem, in which we choose whether to act as if the ratio of between- and within-cluster variances, $\omega = \sigma_\mathrm{B}^2/\sigma_\mathrm{W}^2$, were smaller or greater than a given positive constant ω_0, called the threshold. We specify a loss function (and later a set of loss functions) which quantify the consequences (ramifications) of two kinds of bad decision. We choose the course of action, that is, one of the two verdicts, $\omega \leq \omega_0$ and $\omega > \omega_0$, for which the expected loss is smaller. We use asymmetric loss functions that reflect the dependence of the loss on both the size of the error $|\omega - \omega_0|$ and its sign, when the inappropriate action is chosen. Choosing the appropriate action is associated with no loss. The outcome of our analysis is the preferred action. In contrast, the outcome of a hypothesis has to be interpreted, and the consequences of the two kinds of error considered *ad hoc*.

We address the uncertainty about the loss function by considering a range of plausible loss functions and, in effect, solving the problem for every one of them. Owing to some monotonicity properties, it suffices to solve the problem for the loss functions that delimit the plausible set. This can be regarded as a form of sensitivity analysis. The outcome may be an inferential impasse, but its threat is an incentive for more detailed elicitation and declaration of a narrower range of loss functions. Solutions are developed in the Bayesian paradigm, using prior densities with analytically convenient functional forms. They have a natural frequentist interpretation in terms of additional observations (degrees of freedom).

For background to decision theory we refer to DeGroot (1970, Part 3), Berger (1985) and Lindley (1985) , and for a new perspective to Longford (2013). An application of our approach to compare two normally distributed random samples is presented in Longford (2012a).

The next section presents a condensed background to one-way ANOVA with random effects. For a comprehensive treatment of the topic, including numerous generalisations, see Searle, Casella & McCulloch (2006). The following section deals with decisions about the variance ratio, namely whether it is greater or smaller than a given positive constant. The decision is based on the comparison of the expected losses associated with the contemplated courses of action. Central to our approach is the evaluation of configurations of values of estimates of the variance ratio and of a parameter involved in the loss function for which the two expected losses coincide. These configurations (equilibria), described by a function, divide the space of loss functions to two subsets corresponding to the preference for either action. Section 4 extends the results to unbalanced one-way designs by approximations. An example from athletics training is presented in Section 5. Apart from addressing a substantive issue, we use it to highlight the inappropriateness of hypothesis testing as a basis for decision-making. Technical derivations and related details are collected in a set of appendices.

2. Variance Ratio in One-Way ANOVA

In this section, we derive the ML estimator of ω and show that its distribution is a linear transformation of an F distribution. This simplifies the discussion of the properties of $\hat{\omega}$ and prepares the main development in Section 3.

We consider the balanced one-way ANOVA design

$$y_{ik} = \mu + \delta_k + \varepsilon_{ik}, \tag{1}$$

with clusters $k = 1, \ldots, K$ of m observations each; δ_k and ε_{ik}, $i = 1, \ldots, m$, are two mutually independent random samples from centred normal distributions with respective variances σ_{B}^2 and σ_{W}^2. Denote the variance ratio $\omega = \sigma_{\mathrm{B}}^2/\sigma_{\mathrm{W}}^2$ and the overall sample size $n = Km$.

By setting the partial derivatives of the loglikelihood for the model in (1) to zero, we obtain the estimator $\hat{\mu} = \sum_k \sum_i y_{ik}/n$ and the equations

$$n\hat{\sigma}_{\mathrm{W}}^2 = \mathbf{e}^\top \mathbf{e} - \frac{m^2\hat{\omega}}{1 + m\hat{\omega}} \sum_{k=1}^{K} \bar{e}_k^2$$

$$\frac{n}{1 + m\hat{\omega}} = \frac{1}{\hat{\sigma}_{\mathrm{W}}^2} \frac{m^2}{(1 + m\hat{\omega})^2} \sum_{k=1}^{K} \bar{e}_k^2, \tag{2}$$

where $\mathbf{e}$ is the $(n \times 1)$ vector of residuals $e_{ik} = y_{ik} - \hat{\mu}$, composed of the K within-cluster subvectors $\mathbf{e}_k = (e_{1k}, \ldots, e_{mk})^\top$; $\bar{e}_k$ is the within-cluster average residual, $\bar{e}_k = \frac{1}{m}\mathbf{e}_k^\top \mathbf{1}_m$, where $\mathbf{1}_m$ is the vector of unities of length m. See Appendix A for the derivation of (2).

We decompose $\mathbf{e}^\top \mathbf{e}$ into the within- and between-cluster sum of squares,

$$\mathbf{e}^\top \mathbf{e} = S_{\mathrm{W}} + S_{\mathrm{B}} \,,$$

where $S_{\mathrm{W}} = \sum_{k=1}^{K}(\mathbf{e}_k - \bar{e}_k \mathbf{1}_m)^\top(\mathbf{e}_k - \bar{e}_k \mathbf{1}_m)$ and $S_{\mathrm{B}} = m\sum_{k=1}^{K}\bar{e}_k^2$. For the balanced design, these two statistics are independent and have scaled χ^2 distributions with $n - K$ and $K - 1$ degrees of freedom and respective scaling by σ_{W}^2 and $m\sigma_{\mathrm{B}}^2 + \sigma_{\mathrm{W}}^2$:

$$\frac{S_{\mathrm{W}}}{\sigma_{\mathrm{W}}^2} \sim \chi_{n-K}^2$$

$$\frac{S_{\mathrm{B}}}{m\sigma_{\mathrm{B}}^2 + \sigma_{\mathrm{W}}^2} \sim \chi_{K-1}^2 \,.$$

By simple manipulation of (2) we obtain the identity

$$\hat{\omega} = \frac{m-1}{m}\frac{S_{\mathrm{B}}}{S_{\mathrm{W}}} - \frac{1}{m}\,. \tag{3}$$

Therefore $\hat{\omega} > 0$ when $S_{\mathrm{W}} < (m-1)S_{\mathrm{B}}$. When $S_{\mathrm{W}} > (m-1)S_{\mathrm{B}}$, we truncate $\hat{\omega}$ at zero, to conform with $\omega \geq 0$. To avoid any ambiguity, we denote by $\hat{\omega}_0$ the version of $\hat{\omega}$ given by (3), but truncated at zero. The random variable

$$X = \frac{1}{1 + m\omega}\frac{n-K}{K-1}\frac{S_{\mathrm{B}}}{S_{\mathrm{W}}}$$

has F distribution with $K - 1$ and $n - K$ degrees of freedom, and (3) is equivalent to

$$\hat{\omega} = \frac{X - u}{mu}\,, \tag{4}$$

where $u = K/(K-1)/(1 + m\omega)$. Hence $\mathrm{P}(\hat{\omega}_0 = 0) = \mathrm{P}(\hat{\omega} \leq 0) = \mathrm{P}(X \leq u)$, but this probability depends on ω.

Denote by f_{k_1,k_2} and F_{k_1,k_2} the respective density and distribution function of the F distribution with k_1 and k_2 degrees of freedom. To establish the properties of the estimators $\hat{\omega}$ and $\hat{\omega}_0$, we use the identities

$$yf_{k_1,k_2}(y) = \frac{k_1}{k_1 + 2}f_{k_1+2,k_2-2}(h_1 y)\,,$$

$$y^2 f_{k_1,k_2}(y) = \frac{k_1}{h_1(k_1+4)}f_{k_1+4,k_2-4}(h_2 y)\,, \tag{5}$$

where

$$h_j = \frac{k_1}{k_2}\frac{k_2 - 2j}{k_1 + 2j}\,, \tag{6}$$

$j = 1, 2$. The identities are derived in Appendix B.

3. Decision Theory for ω

In this section, we derive the posterior distributions of ω for a range of prior distributions and use them to formulate a decision rule for choosing between the actions (options) A and B. Although the derivations refer to the Bayesian paradigm, they retain a frequentist interpretation by equating informative priors to specific data alterations. The decision rule is based on the balance of the expected losses for the two options; we prefer the option for which the expected loss is smaller.

Suppose two courses of action, A and B, are contemplated. They are complementary (one of the actions has to be taken) and exclusive (both actions cannot be taken). Suppose further that A is appropriate when $\omega \leq \omega_0$ and B when $\omega > \omega_0$; ω_0 is a given scalar, but ω is not known, and all the information about it is contained in the vector of outcomes **y** and a prior distribution for ω.

The inverse of the identity in (4) is

$$\omega = \frac{\hat{v}}{X} - \frac{1}{m}, \tag{7}$$

where $\hat{v} = K(\hat{\omega} + 1/m)/(K-1)$. This operation is associated with the so called fiducial argument, originating from Fisher (1935, 1956). Its validity has been extensively discussed (Lindley 1985, Seidenfeld 1992, Hannig 2009) . In fact, (7) is an example of failure of the argument, because the solution ω as a random variable has a positive probability of being negative.

We consider the non-informative improper prior $g(y) = 1$, $y \geq 0$, and a parametric class of proper priors for ω, and base our choice of option (the course of action) on the posterior distribution of ω. The proper prior densities are

$$g(y) = \frac{m(q-1)}{(1+my)^q}, \tag{8}$$

for $y > 0$ and a parameter $q > 1$. A selection of these densities is drawn in Figure 1 for $m = 10$, with the value of the parameter q printed at the left-hand margin. For large q, the densities are highly informative and have large values in the vicinity of $\omega = 0$. We acknowledge that the presence of m in the prior densities may be seen as problematic.

The posterior densities of ω for these priors as well as the (improper) constant prior are derived in Appendix C. They are given by the expression

$$g_q(y; \hat{\omega}) = \frac{H_q m}{1 + m\hat{\omega}} \frac{f_{k_1(q),k_2(q)}\{H_q z(y; \hat{\omega})\}}{1 - F_{k_1(q),k_2(q)}\{H_q z(0; \hat{\omega})\}}, \tag{9}$$

where $z(y; \hat{\omega}) = (1+my)/(1+m\hat{\omega})$, $k_1(q) = n - K - 2q + 2$ and $k_2(q) = K + 2q - 3$ and

$$H_q = \frac{n-K}{K} \frac{k_2(q)}{k_1(q)}.$$

The non-informative prior corresponds to $q = 0$. For $q = 1$, when the prior is not defined, the 'posterior' density in (9) is well defined as

$$g_1(y; \hat{\omega}) = \frac{H_1 m}{1 + m\hat{\omega}} \frac{f_{n-K,K-1}\{H_1 z(y; \hat{\omega})\}}{1 - F_{n-K,K-1}\{H_1/(1 + m\hat{\omega})\}},$$

and $H_1 = (K - 1)/K$. This is the conditional distribution of ω in (7) given that $\omega > 0$. The posterior distribution for $q > 1$ can be interpreted as an addition of $2q$ clusters to the data, with the corresponding reduction of the cluster size m to $n/(K + 2q)$, to keep the overall sample size n unchanged. The estimator $\hat{\omega}$ is also changed, multiplied by H_1/H_q. Thus, a prior in (8) is equivalent to altering the dataset, and the posterior distribution of ω can be treated as the sampling distribution of this altered dataset. Of course, we have to permit fractional within-cluster sample sizes and numbers of clusters. In what follows, we drop q in the arguments of k_1 and k_2 and in the subscript of H.
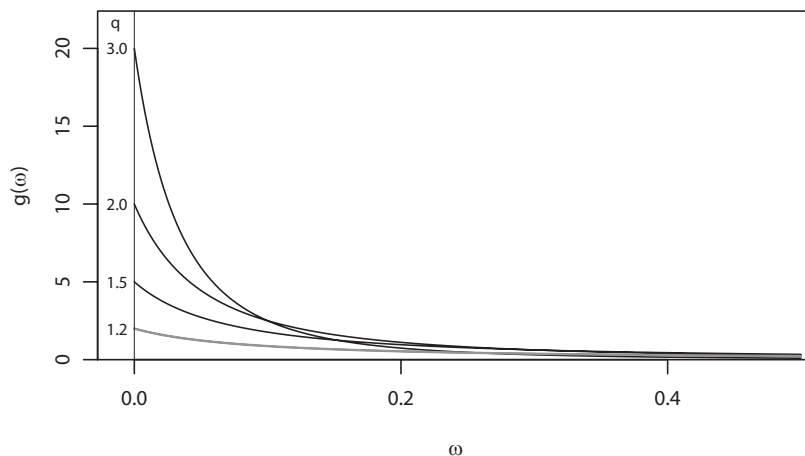

FIGURE 1: A set of prior densities given by (8) for cluster size $m = 10$.

We assume that the appropriate course of action results in no loss. However, when B is chosen (claiming that $\omega > \omega_0$), even though $\omega \leq \omega_0$, we incur unit loss, and when A is chosen even though $\omega > \omega_0$, the loss is equal to R. We refer to R as the penalty ratio. We want to choose the action for which the expected loss is smaller. The choice is based solely on the posterior distribution of ω. Table 1 illustrates the loss function, with values of 0, 1 and R for the subsets of the space of pairs $(\hat{\omega}, \omega)$. It represents the client's perspective, and is therefore subjective. This perspective may be difficult to establish and quantify by specifying the value of R. We address this issue in Section 3.2 by considering a range of plausible values of R, and later we introduce classes of loss functions other than piecewise constant.

TABLE 1: Piecewise constant loss function.

			Reality	
			$\omega \leq \omega_0$	$\omega > \omega_0$
Decision	A	$(\omega \leq \omega_0)$	0	R
	B	$(\omega > \omega_0)$	1	0

The expected loss associated with action B is

$$L_{\mathrm{B}} = \int_0^{\omega_0} g_q(y; \hat{\omega}) \, \mathrm{d}y$$

$$= \frac{Hm}{1 + m\hat{\omega}} \frac{1}{1 - F_{k_1,k_2}\{Hz(0; \hat{\omega})\}} \int_0^{\omega_0} f_{k_1,k_2}\{Hz(y; \hat{\omega})\} \, \mathrm{d}y$$

$$= \frac{F_{k_1,k_2}\{Hz(\omega_0; \hat{\omega})\} - F_{k_1,k_2}\{Hz(0; \hat{\omega})\}}{1 - F_{k_1,k_2}\{Hz(0; \hat{\omega})\}} .$$

Further, $L_{\mathrm{A}} = R(1 - L_{\mathrm{B}})$. We define the balance function as $\Delta L = L_{\mathrm{A}} - L_{\mathrm{B}}$;

$$\Delta L(\hat{\omega}; R, \omega_0) \,=\, R - (R+1)L_{\mathrm{B}} . \tag{10}$$

We choose action A when $\Delta L < 0$ and action B when $\Delta L > 0$. For greater values of $\hat{\omega}$ we should be more inclined toward choosing action B, implying that ΔL, as a function of $\hat{\omega}$, is increasing. We have no analytical proof of this conjecture. The respective limits of ΔL as $\hat{\omega} \to -1/m$ and $\hat{\omega} \to +\infty$ are -1 and R. Therefore there is an *equilibrium* value ω_q^*, for which $\Delta L(\omega_q^*; R, \omega_0) = 0$. In an exhaustive (empirical) search, we have found this equilibrium to be unique. When $\hat{\omega} < \omega_q^*$ we choose A and when $\hat{\omega} > \omega_q^*$ we choose B. Note that the equilibrium may be negative; in fact, $\omega_q^* < 0$ when

$$\Delta L(0) = \frac{F_{k_1,k_2}\{H(1 + m\omega_0)\} - F_{k_1,k_2}(H)}{1 - F_{k_1,k_2}(H)} \,>\, \frac{R}{R+1}.$$

This is equivalent to either of the following conditions

$$R > \frac{F_{k_1,k_2}\{H(1 + m\omega_0)\} - F_{k_1,k_2}(H)}{1 - F_{k_1,k_2}\{H(1 + m\omega_0)\}}$$

$$\omega_0 < \frac{1}{Hm} F_{k_1,k_2}^{-1}\left\{ \frac{R + F_{k_1,k_2}(H)}{R+1} \right\} - \frac{1}{m},$$

that is, when the consequences of choosing action B incorrectly are sufficiently serious (large R) or we are very strict about what we regard as small (small ω_0). The equilibrium value is found by the Newton method or a similar algorithm. Methods that use the derivatives $\partial \Delta L / \partial \hat{\omega}$ are not practical because the expressions involved are quite lengthy.

The decision rule based on the sign of ΔL is substantially different from the outcome of a hypothesis test based on an F distribution. As an aside, we note the complications with ω possibly being at the boundary of the parameter space. The hypothesis in a test is often that $\omega = 0$. However, in such a case we would be satisfied if the hypothesis were not rejected even when ω is positive but small. The value of ω_0 for the decision rule is comparable to the largest value of ω that we would still call 'small' (the largest unimportant value). In this perspective, $\omega_0 = 0$ is not appropriate. For a decision, we always choose $\omega_0 > 0$. In decision theory, the two alternatives, $\omega \leq \omega_0$ and $\omega > \omega_0$, have equal status, whereas in

hypothesis testing the hypothesis defines the default course of action, which is overruled only when there is sufficient evidence against it. In hypothesis testing, the semicontinuous prior for ω, with a mass at zero, can be declared. Such a distribution is a mixture of the mass at zero and a continuous distribution. For a given continuous component, greater prior mass at zero leads to greater posterior mass at $\omega = 0$. Transparency of the declaration, a clear justification for the declared mass, is therefore essential. In the approach we propose, we do not have to resort to such a device.

Instead of $\omega = 0$, we may test the hypothesis that $\omega \leq \omega_0$. However, then the power of the test for values slightly greater than ω_0 is very small, so ω_0 is not a tangible quantity in hypothesis testing as it is in our approach, where it represents a clear boundary between the two options. Note that tests (and decisions) for ω and the intraclass correlation coefficient, $\rho = \sigma_B^2/(\sigma_B^2 + \sigma_2^2)$, are equivalent because $\rho = 1/(1 + 1/\omega)$ is a monotone function of ω. Although the scale of ρ is preferred by some for interpretation, evaluations for $\hat{\omega}$ are easier for both hypothesis testing and decision making because of its relation to F distribution.

Having to set the penalty ratio R might seem as an additional burden to the analyst. However, its importance is obvious from how ω_q^* depends on R. Figure 2 displays the values of ω_0^* as a function of R for the pairs (K, m) set to $(4, 6)$, $(7, 10)$ and $(10, 20)$, and thresholds $\omega_0 = 0.1$ and $\omega_0 = 0.25$ drawn by black and gray colors, respectively. The noninformative prior ($q = 0$) is assumed. The functions are drawn on the original (linear) and log scales, to obtain high resolution for both large and small values of R.
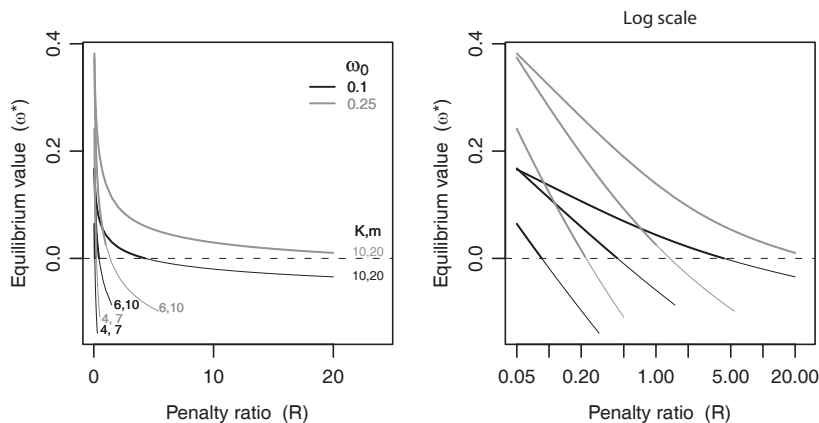

FIGURE 2: The equilibrium values ω^* as functions of the penalty ratio R for number of clusters K and cluster size m indicated in the left-hand panel; non-informative prior for ω.

The functions are decreasing with limits $+\infty$ as $R \to 0$ and $-1/m$ as $R \to +\infty$. They are drawn either for $R < 20$ or up to the value of R at which the equilibrium reaches the minimum of $-1/m$. For $(K, m) = (10, 20)$, the equilibria are above the minimum of -0.05 for both $\omega_0 = 0.1$ and 0.25 even at $R = 20$, but for the other settings the limits are reached for $R < 20$.

For very small penalty ratio R, action B is selected only for extremely large values of $\hat{\omega}$ because the error of incorrectly choosing action A has no serious repercussions. When the equilibrium is at $-1/m$ we choose action B for all values of $\hat{\omega}$. The equilibrium functions $\omega^*(R)$ have steep gradients for small values of R. The equilibria have much less curvature on the log scale. Higher value of ω_0 is associated with uniformly higher equilibrium. The differences of the pairs of equilibrium functions decrease with R.

3.1. Piecewise Linear Loss

In some settings, the loss depends not only on the sign of the error made, but also on its magnitude $|\omega - \omega_0|$. One such class of loss functions are the piecewise linear, defined as $\omega_0 - \omega$ when we choose B inappropriately (when $\omega \leq \omega_0$) and $R(\omega - \omega_0)$ when we choose A inappropriately (when $\omega > \omega_0$).

For a given value of $\hat{\omega}$, we compare the expected losses L_A and L_B, and choose the action with the smaller expected loss. In Appendix D, we derive the following expression for the balance function $\Delta L = L_A - L_B$,

$$
\begin{aligned}
\Delta L = {} & \frac{1 + m\omega_0}{m} \frac{F_{k_1, k_2}\{G(\omega_0)\} - F_{k_1, k_2}\{G(0)\}}{1 - F_{k_1, k_2}\{G(\omega_0)\}} - \frac{K(n - K + 2q - 2)}{(K + 2q - 5)(n - K)} \\
& \times \frac{1 + m\hat{\omega}}{m} \frac{F_{k_1+2, k_2-2}\{h_1 G(\omega_0)\} - F_{k_1+2, k_2-2}\{h_1 G(0)\}}{1 - F_{k_1, k_2}\{G(0)\}},
\end{aligned}
\tag{11}
$$

where $G(y) = H_q z(y; \hat{\omega}) = H_q (1 + my)/(1 + m\hat{\omega})$. This identity holds only when $K > 3 - 2q$. Otherwise L_A is infinite, so action B is chosen, irrespective of $\hat{\omega}$, ω_0 or R. When ΔL is well defined (finite) we choose A if $\Delta L(\hat{\omega}; R, \omega_0) < 0$, and choose B otherwise.

Figure 3 presents the equilibrium functions for a selection of designs (K, m, indicated in the diagram), and ω_0 set to 0.1 and 0.25 (distinct colors), as functions of the penalty ratio R. All six functions in the diagram are non-increasing— greater R makes the choice of B more attractive. Each equilibrium converges to the minimum value of $\hat{\omega}$, equal to $-1/m$. For $\omega_0 = 0.1$, $K = 8$ and $m = 15$, this limit is reached at $R \doteq 5.92 < 20$. For $\omega_0 = 0.1$, $K = 10$ and $m = 14$, the limit is reached at $R \doteq 23.26$, off the horizontal scale. Smaller value of ω_0 is associated with lower equilibrium, and therefore increased preference for action B.

Earlier we conjectured that the balance function for the piecewise constant loss is increasing. This is not the case for the piecewise linear loss in general. A set of examples is drawn in Figure 4 for $K = 8$ and $m = 15$. All the functions in the diagram converge to zero as $\hat{\omega} \to -1/m$. However, for R smaller than a critical value $R_\dagger = R(K, m)$, the balance has a dip before increasing. Without choosing the initial values carefully, a search for the root of ΔL may converge to $-1/m$, even when there is another root. For $R < R_\dagger$, the balance in favor of action A is narrower for $\hat{\omega}$ close to $-1/m$ than for larger values of $\hat{\omega}$ because larger values of $\hat{\omega}$ yield posterior distributions with greater dispersion. This adds strength to the choice of B, although the decision rule remains reasonable:
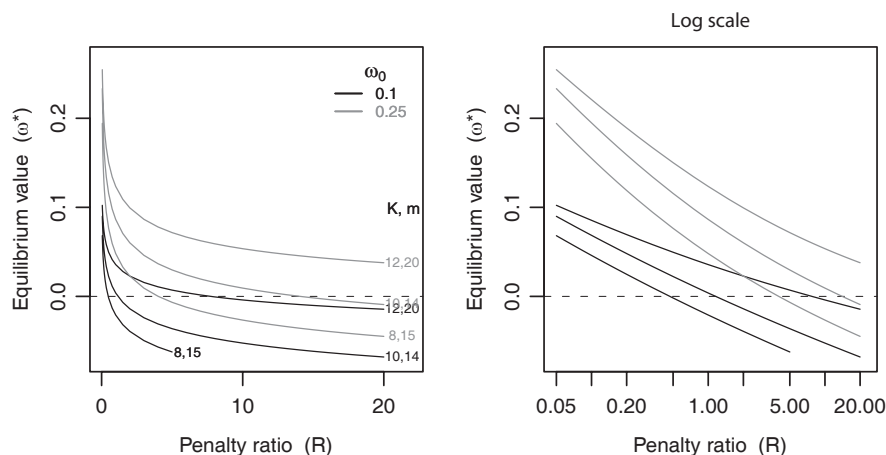
FIGURE 3: The equilibrium values ω^* as functions of the penalty ratio R for number of clusters K and cluster size m indicated in the left-hand panel. Piecewise linear loss and non-informative prior for ω.
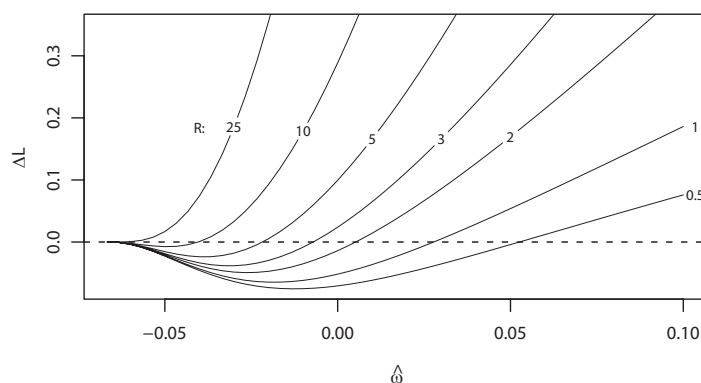

FIGURE 4: The balance functions $\Delta L(\hat{\omega})$ for the piecewise linear loss; $K = 8$, $m = 15$, $\omega_0 = 0.2$ and non-informative prior for ω. The values of R are indicated on the curves.

- for R up to a certain value R^*, choose A for $\hat{\omega}$ smaller than a critical value; otherwise, choose B;

- for R greater than this value, choose B for all values of $\hat{\omega}$.

The borderline value R^* depends on K, m and ω_0.

A more radical dependence of the loss on the magnitude of the error is represented by the quadratic loss function, given as $(\omega - \omega_0)^2$ when we erroneously choose action B, even though $\omega < \omega_0$, and as $R(\omega - \omega_0)^2$ when we choose action A, even though $\omega > \omega_0$. The balance function for this loss is derived in Appendix E.

3.2. Plausible Loss Functions, Priors and Values of ω_0

The values of the penalty ratio R and ω_0 are elicited from the subject-matter expert (the client). Often no single pair of values is arrived at; instead, ranges of plausible values $\mathcal{R} = (R_L, R_U)$ and $\boldsymbol{\Omega} = (\omega_{0L}, \omega_{0U})$ are declared. A range $\mathcal{R}$ (or $\boldsymbol{\Omega}$) is said to be plausible if any value R (or ω_0) outside the range can be ruled out. Thus, if a range is plausible, then so is any (wider) range that contains it.

With plausible ranges of R and ω_0, we might solve the problem for a grid of plausible values $(R, \omega_0) \in \mathcal{R} \times \boldsymbol{\Omega}$ and choose action A or B if the expected loss with that action is smaller for every plausible pair (R, ω). If action A is preferred for some plausible pairs (R, ω_0) but action B for some others, then an *impasse* results, which can be resolved only by reviewing and reducing the ranges $\mathcal{R}$ and $\boldsymbol{\Omega}$. Uncertainty about the prior parameter q can be dealt with similarly.

We assume that the plausible values of R and ω_0 form a rectangle, $\mathcal{R} \times \boldsymbol{\Omega}$. In practice, it suffices to find the signs of the balance ΔL for the vertices of this rectangle. If $\Delta L < 0$ at all four vertices, then action A is selected; if $\Delta L > 0$ at all four vertices, then action B is selected. In these two cases, we have *unequivocal decisions*; the same action is chosen for all plausible configurations of R and ω_0. Otherwise we reach an impasse. If the plausible ranges of R and ω_0 are reviewed in further elicitation, the reduced ranges have to remain plausible.

In an alternative approach, we split the space of all pairs (R, ω_0) according to the sign of $\Delta L(\hat{\omega}; R, \omega_0)$. For ω_0 close to zero, L_B is very small, and so B is the preferred action. For very large ω_0, action A is preferred. For $\hat{\omega}$ and ω_0 fixed, the balance functions are linear in R. Therefore $\Delta L = 0$ for a unique value $R^* = R(\omega_0; \hat{\omega})$. This function of ω_0 and $\hat{\omega}$ splits the space (R, ω_0) to the regions in which A or B has smaller expected loss. Figure 5 displays these functions for piecewise constant loss and a selection of values $\hat{\omega}$ indicated at the right-hand margin. The curves drawn by black color are for the design $(K = 8, m = 11)$ and those by gray color for $(K = 8, m = 10)$, to explore the dependence of R^* on m. Less information in the data requires smaller penalty ratio R to reach the balance of the two expected losses. This applies also when K is reduced; details are omitted.

If plausible ranges are declared for R and ω_0 and R^* intersects the plausible rectangle, we have an impasse, because for some plausible pairs (R, ω_0) one action, and for other pairs the other action is preferred. The equilibrium value R is an increasing function of ω_0. Therefore, if the same action is preferred for both extremes R_L and R_U, then it is preferred for all $R \in (R_L, R_U)$. If one action is preferred for some plausible values of R, say, action A for $R \in (R_L, R_Q)$, $R_L < R_Q < R_U$, and action B for $R \in (R_Q, R_U)$, then we have an impasse that can be resolved only by resuming elicitation to narrow down the plausible range (R_L, R_U).

Although the standard Bayesian analysis deals with a single prior, it is sometimes practical to consider a range of plausible priors, or prior parameters, (q_L, q_U), reflecting the lack of consensus in the elicitation or uncertainty admitted by the expert. See Longford (2010) for an application and related discussion. An example is given in Figure 6 for the design with $K = 10$ and $m = 10$. Piecewise quadratic
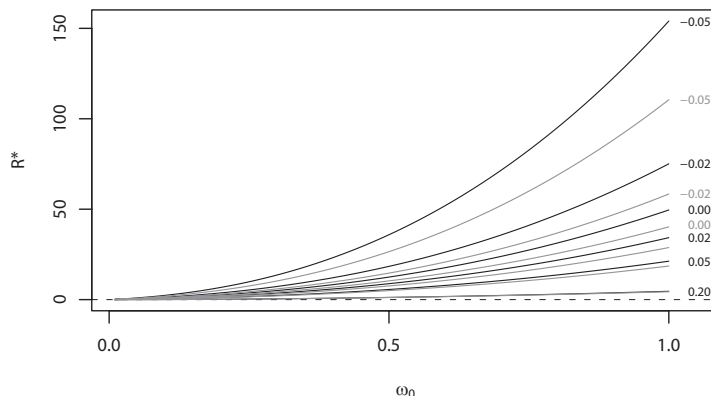
FIGURE 5: The equilibrium functions $R^* = R(\omega_0; \hat{\omega})$ for the piecewise constant loss function and values of $\hat{\omega}$ indicated at the right-hand margin; $K = 8$ and $m = 11$ (black lines) and $m = 10$ (gray).

loss is assumed. The functions drawn in the diagram are the equilibrium functions $R(q)$ which satisfy the identity $\Delta L_q(\hat{\omega}; R(q), \omega_0) = 0$. The solutions for $\hat{\omega} = 0.05$ are drawn by black lines for the values of ω_0 printed at the right-hand margin. The solutions for $\hat{\omega} = 0.055$ and the same values of ω_0 are drawn by gray lines. For $\omega_0 = 0.10$ the two curves are difficult to discern.
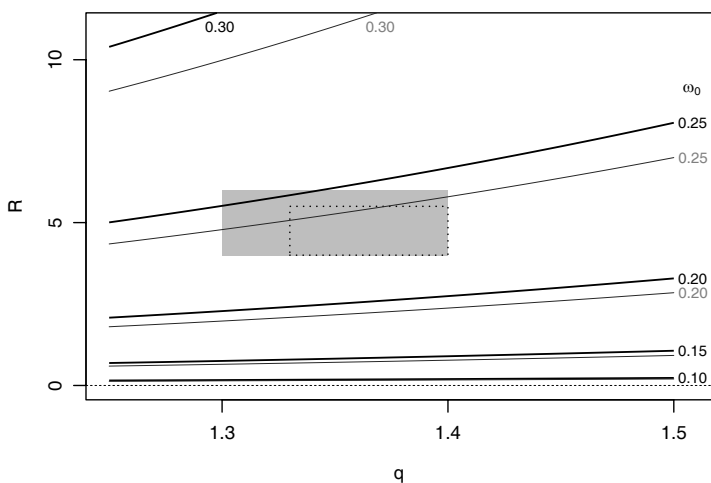

FIGURE 6: The equilibrium functions $R(q)$ for the piecewise quadratic loss function and values of ω_0 indicated at the right-hand margin and at the top; $\hat{\omega} = 0.05$ (black lines) and $\hat{\omega} = 0.055$ (gray), $K = 10$ and $m = 10$. The plausible rectangle is filled by gray color, and its reduced version is delimited by dots.

Suppose the value $\hat{\omega} = 0.05$ was realised and the plausible range $(0.15, 0.20)$ was declared for ω_0. Further, suppose the respective plausible ranges for R and q are $(4, 6)$ and $(1.3, 1.4)$; this rectangle is filled in the diagram by gray color. The

plausible rectangle is above the equilibrium curve $R(q)$ for any plausible value of $\omega_0 \in (0.15, 0.20)$, so action B is selected unequivocally.

Suppose next that the plausible range of ω_0 is $(0.20, 0.25)$. Then nearly the entire plausible rectangle lies between the functions $R(q)$ for $\omega_0 = 0.20$ and $\omega_0 = 0.25$. In this case, we have an impasse—for some plausible values of ω_0 action A and for other values action B is preferred. Finally, suppose the plausible range of ω_0 is $(0.25, 0.30)$. The plausible rectangle intersects the region delimited by the functions $R(q)$ for $\omega_0 = 0.25$ and 0.30, but if the lower limit of plausible values of q could be increased, or the upper limit for R reduced, the reviewed plausible rectangle might be entirely under the curve $R(q)$ for $\omega_0 = 0.25$ and action A would then be preferred unequivocally. Such a reviewed plausible rectangle is delimited in the diagram by dots.

These examples imply a strong incentive to declare as small a set of plausible values as possible for all the parameters involved, to reduce the chances of an equivocal decision. But the client has to be comfortable with the implication that all pairs outside the declared set can be ruled out.

4. Designs Without Balance

For designs without balance we do not have a closed-form expression for the conditional distribution of $\hat{\omega}$ given ω, nor a tractable posterior density of ω. We approximate this distribution by its match among the balanced designs. In the approximation, we use the synthetic number of clusters $K' = n^2 / \sum_k n_k^2$ and the harmonic mean of the cluster sizes $m' = K / \sum_k 1/n_k$ in the respective roles of K and m. These proposals are based on Potthoff, Woodbury & Manton (1992) and Longford (2000). Potthoff et al. (1992) derived a generalisation of the equation for K' by matching the information in a sample with unequal sampling weights with a sample that would have equal weights. The approximation for m' is derived directly from the information about the variance ratio in the likelihood maximisation. We note that these approximations are poor for large ω, such as $\omega > 0.5$.

Figure 7 presents the histograms of the ML estimates $\hat{\omega}$ obtained in 10,000 replications each for the values of ω listed in the titles for the design with two clusters each of size 6, 7 and 8, for which $K' = 5.92$ and $m' = 6.90$. The approximation is not perfect, but in the context of substantial dispersion of $\hat{\omega}$ it is acceptable.

An alternative is to assume that the statistic S_B is associated with $K - 1$ degrees of freedom and S_W with degrees of freedom in the range (n_l, n_u), where $n_l = (K-1)m_l$ and $n_u = (K-1)m_u$, and m_l and m_u are the sample sizes of the smallest and largest clusters, respectively, 6 and 8 in the example above. Then we solve the problem for $K(m_l - 1)$ and $K(m_u - 1)$ degrees of freedom associated with S_W. If we arrive at the same conclusion in both cases, then it applies also to the original dataset. The method is poorly suited for extremely unbalanced data.
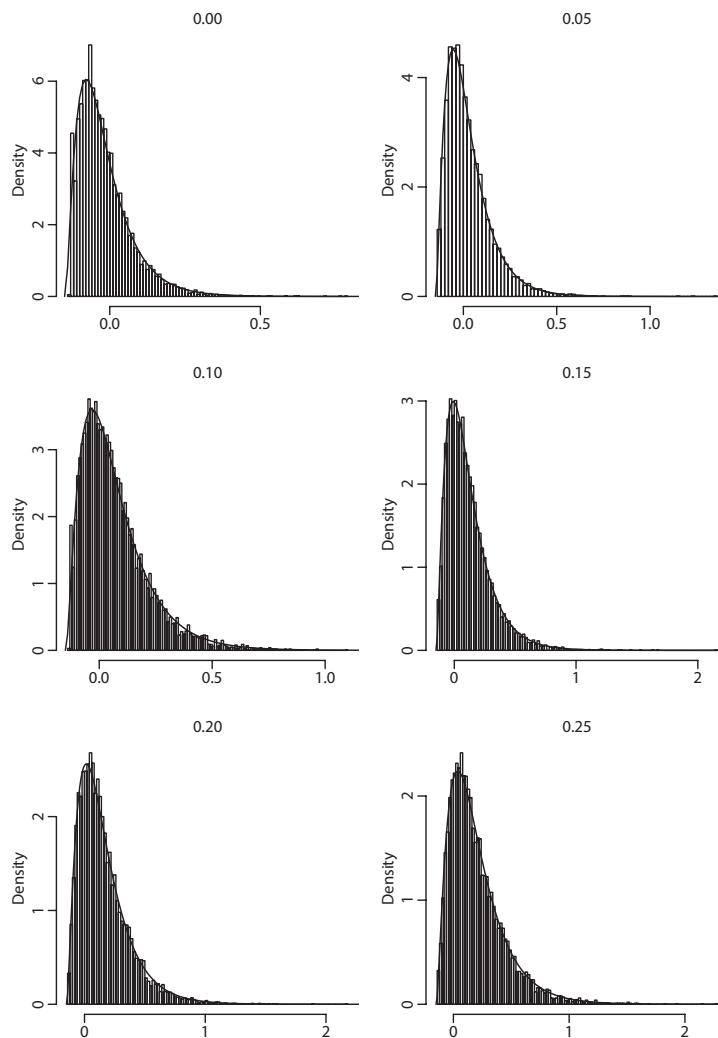
FIGURE 7: The empirical distribution of $\hat{\omega}$ (histogram) for the design with sample sizes 6, 6, 7, 7, 8 and 8 and the fitted distribution based on $K' = 5.92$ and $m' = 6.81$ (solid line), for $\omega = 0.0, 0.05, \ldots, 0.025$ (title). Histograms based on 10,000 replications.

Another alternative is motivated by methods for missing data (Little & Rubin 2002, Rubin 2002). We assume that the (unbalanced) observed dataset is incomplete, and its complete version is balanced, with m_{u} observations in each cluster. The EM algorithm (Dempster, Laird & Rubin 1977) is particularly easy to implement for this setting, but it does not yield the sampling distributions of the estimators. By multiple imputation, we generate a number of plausible completions of the observed data, and then analyse each dataset separately. If in every case we prefer the same action, we have an unequivocal conclusion. Otherwise an

impasse results. The problem with this approach is that the chances of an impasse increase substantially when the sample sizes n_k are in a wide range, because the missing data contains a large fraction of the complete information.

We note that a similar approach to hypothesis testing requires an adjustment because by augmenting the observed (incomplete) dataset to make it balanced we bias the results toward the alternative. See Li, Meng, Raghunathan & Rubin (1991) for a solution.

5. Application

We illustrate the methods on the data from ten training sessions of a track athlete. Each session comprised eight 400 metre runs (one lap of the track), separated by jogging 400 metres for recovery. The athlete ran unaccompanied, and was not informed about any intermediate times (e.g., at 200 metres), nor when completing a lap; he could inspect the eight times only at the end of the session. He aimed to run each lap in 55.0 seconds. The purpose of the sessions was to develop a good judgement of speed, discounting any fatigue and any external factors (weather). The collected data, times in seconds with precision to one decimal place, are presented in Figure 8, with sessions marked as A − J.
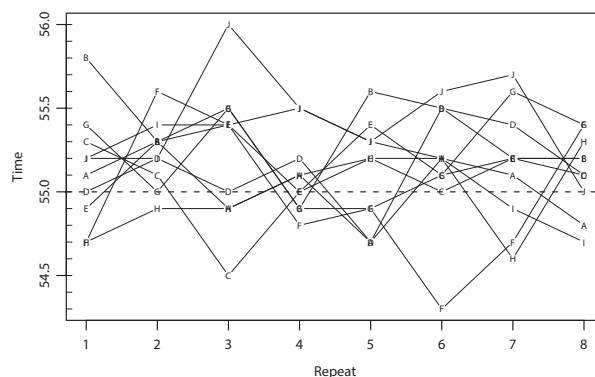

FIGURE 8: The times for running 400 metres (a lap of the track) in sessions A − J comprising eight separate laps (repeats) each.

We ignore the sequential order of the runs and sessions. Running a lap in 55.0 seconds does not require a full effort of the athlete, who could run the distance in well below 50 seconds. Of course, fatigue accumulates over the laps, but there is an equal threat of under- and over-compensating for it. Also, initial data exploration suggests no presence of a trend over the eight laps nor any temporal dependence across the sessions. The sufficient statistics for the random-effects ANOVA are $S_W = 5.595$ and $S_B = 1.797$ and the sample mean is $\hat{\mu} = 55.16$. The maximum likelihood estimates (MLE) of the variance components are $\hat{\sigma}_W^2 = 0.0799$ and $\hat{\sigma}_B^2 = 0.0125$.

If the times achieved do not differ much across the sessions (and are sufficiently close to 55.0 seconds on average), then the athlete can progress to the next stage of training, for which the ability to maintain the particular speed is essential. One aspect of the ability is a sufficiently small value of the variance ratio ω; its MLE is $\hat{\omega} = 0.156$. Elicitation from the coach and his colleagues concluded that $\omega \leq 0.20$ is necessary for progressing to the next stage and that piecewise linear loss functions with $R \in (1/5, 1/3)$ are plausible. Noninformative prior ($q = 0$) is assumed.

Contrary to popular perceptions, (professional) athletes and their coaching and management staff are well aware of uncertainty about future fitness and performance, which they consider in preparing training schedules, planning attendance in competitions and assessing the athlete's prospects. A lot of data is nowadays collected in training, and statistical definitions of distributions and their dispersion are relatively easy to introduce. In the elicitation process, we communicated with the coaching staff mainly through graphs of large (simulated) sets of times, and asked them whether the displayed variation was acceptable or not for progressing to the next stage. We settled first on $\omega \in (0.17, 0.24)$ and later agreed on $\omega = 0.20$. Disagreement and uncertainty persisted about the value of R, which compares the harm done by the two kinds of erroneous choice, and that is why we consider a plausible range for it.

Figure 9 summarises the results of the analysis by the plot of the equilibrium function (marked $q = 0$). The plausible range of R is marked by shading and the value of $\hat{\omega}$ by horizontal dots. Since $\hat{\omega}$ is above the equilibrium function throughout $\mathcal{R}$, the action with smaller expected loss is not to proceed to the next stage; the estimated variance ratio is too large. The equilibrium values ω^* are 0.101 and 0.074 for the respective values $R = \frac{1}{3}$ and $\frac{1}{5}$. For $R = 0.075$, $\omega^* = 0.157$, so the decision would not be affected if values of R much smaller than $\frac{1}{5}$ were regarded as plausible.
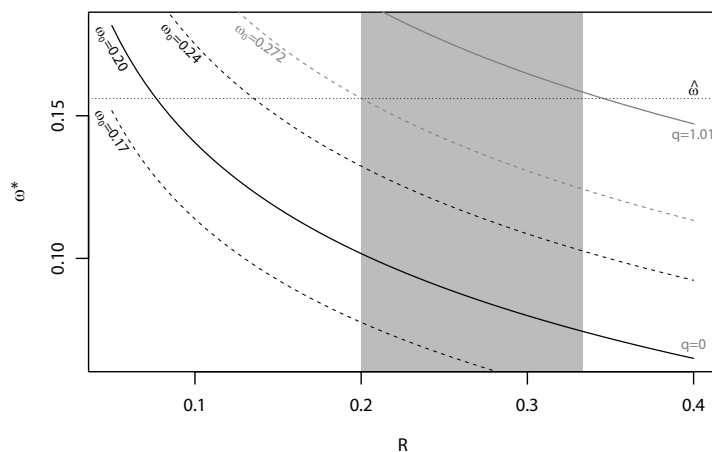

FIGURE 9: The equilibrium values of ω for the study of the track athlete's times for 400 metres.

If instead of a single value, $\omega_0 = 0.20$, the plausible range $\omega_0 = (0.17, 0.24)$ were agreed, as it was in an earlier round of elicitation, the equilibrium functions drawn by black dashes and marked by the values of ω_0 would be obtained. Since both curves are entirely below $\hat{\omega} = 0.156$ for all $R \in (0.20, 0.33)$, the conclusion 'Not to proceed' would not be affected. By trial and error, we found that for $\omega_0 = 0.272$ the equilibrium at $R = 0.2$ is equal to $\hat{\omega}$. Thus, if a standard more lenient than $\omega_0 = 0.272$ for the athlete's consistency were plausible, then the decision would be equivocal (in doubt) because the equilibrium curve would then intersect the horizontal line at $\hat{\omega}$ for a plausible value of R. For $\omega_0 = 0.32$, the equilibrium curve is equal to $\hat{\omega}$ at the upper limit of plausible values of R, 0.33. So, if ω_0 were greater than 0.32 the appropriate decision would be 'To proceed'.

We add a word of caution at this point. Suppose the prior with $q = 1.01$ is adopted. The proximity of q to unity might suggest that this prior is only mildly informative. However, the equilibrium function, drawn by gray color, differs from the equilibrium for $q = 0$ substantially. In fact, with $q = 1.01$, the decision to proceed (that ω is small) would be preferred, because the 'gray' equilibrium function is entirely above $\hat{\omega}$ throughout the plausible range of R.

In an established approach, we would test the hypothesis that $\omega \leq 0.20$. This hypothesis is not rejected; the p value, derived from (4), is 0.48. Commonly, one would conclude that the action appropriate when the hypothesis applies should be taken. Such a decision process is logically incorrect, confusing failure to reject a hypothesis with its acceptance. That is, the analysis started by assuming that the hypothesis is valid, and no contradiction with it was found. The appropriate conclusion is that of ignorance, that we do not know whether $\omega \leq 0.20$ or $\omega > 0.20$, or more precisely, that the data yield sufficient evidence for neither the hypothesis nor the alternative. We note that the hypothesis $\omega > 0.20$ (exchanging the roles of the original hypothesis and alternative) would not be rejected either, further compounding the illogicality of a decision based on the result of a hypothesis test.

6. Discussion and Conclusion

The outcome of a hypothesis test is often used to support a decision to continue an analysis as if the hypothesis or the alternative were valid. This is widely acknowledged to be inappropriate when we fail to reject the hypothesis, and yet act as if it did apply, but this is often ignored in practice. In our perspective, such use of a hypothesis test is inappropriate even when the hypothesis is rejected, and thus evidence against it obtained, because the consequences of the type II error are not taken into account. The pragmatic arguments for hypothesis testing, such as relatively simple computational procedures and reference to asymptotic theory, do not have a good foundation in the case of variance estimation, especially in experiments with small or moderate sample sizes, in which the expenditure on the study and the ramifications of the errors of the two kinds are important factors.

In our approach based on decision theory, the consequences of the two kinds of bad decision are represented by a loss function, and the uncertainty about it by a set of plausible loss functions. The loss functions are elicited from the

expert (client), together with the prior information about the variances. The prior distribution of the variances is useful, but specifying it is not an imperative in our analysis. In fact, it can be formulated as an alteration of the realised dataset, so the analyst does not have to subscribe to the Bayesian paradigm. Plausible priors (or prior data) and loss functions have an important role as conduits for sensitivity analysis, exploring how the conclusion is changed as a result of (small) changes in the input. This is greatly simplified by the choice of classes of loss functions, for which the balance is a linear function of the penalty ratio R; see (10), (11) and (16). The priors used in them involve the within-cluster sample size m. They are chosen so that the posterior would have a closed form. An alternative to them is the class of uniform priors on $(0, \omega^\dagger)$, with a value or a plausible range specified for $\omega^\dagger$. The balance functions with these priors are obtained from (11) and (16) by replacing each term $1 - F_{k_1', k_2'}\{hG(\omega_0)\}$ with $F_{k_1', k_2'}\{hG(\omega_0^\dagger)\} - F_{k_1', k_2'}\{hG(\omega_0)\}$, where respectively $k_1' = k_1(q)$, $k_1(q) + 2$ or $k_1(q) + 4$ and similarly for k_2; $h = 1$, h_1 or h_2.

Restricted maximum likelihood (REML; Patterson and Thompson, 1971) is often considered for random effects models. Our method has a simple adaptation for REML. It amounts to setting $u = 1/(1 + m\omega)$ in (4) and

$$H_q = \frac{n - K}{K - 1} \frac{k_1(q)}{k_2(q)}$$

in (9). No other changes are required in the subsequent equations.

The task solved by our approach is to select the action that is appropriate when the unknown value of the variance ratio ω is smaller than a set threshold $\omega_0 > 0$, or the complementary action; ω_0 can be interpreted as the smallest important deviation from zero. Although a typical hypothesis test about ω is for $\omega = 0$, failure to reject it is regarded as appropriate when ω is positive but small. Our approach requires a quantification of what 'small' means, by ω_0, and setting it to zero would not be reasonable. Its magnitude should be informed by the methods (steps in the analysis or in experimentation) contemplated after the decision. Uncertainty about it and the contentious nature of having to specify a single value are dealt with by using a plausible range for ω_0.

Our development is exact only for balanced (one-way) designs; the proposed solutions for unbalanced designs make references to the results for similar balanced designs. Extensions to more complex (multiway) designs are on our research agenda; for such designs the reference to balanced designs may be rather restricting.

Decision theory for some elementary statistical problems, such as estimating a mean (and fixed-effects ANOVA), a variance, and classifying units to two categories, is developed in Longford (2013). With fixed-effects ANOVA, the group-level means μ_k, $k = 1, \dots, K$, are parameters, and $\mu_1 = \dots = \mu_K$ is the commonly tested hypothesis. A measure of the departure from this hypothesis, needed for applying our approach, can be defined through the (finite-sample) variance of the means, $\sigma_G^2 = \sum_k (\mu_k - \mu)^2 / K$, which is similar to, but not the same as σ_B^2. The

variance σ_{G}^2 is estimated by the method of moments, adjusting $\sum_k (\hat{\mu}_k - \hat{\mu})^2 / K$ for its bias.

All computing was carried out in R, and the software developed is available from the first author (NTL). All our evaluations are based on closed-form expressions for the expected losses, and are executed instantly. When a prior is declared for which the posterior density has to be evaluated numerically, Monte Carlo Markov chain (MCMC) calculations can be employed (Robert & Cassella 2004) Large samples from the joint posterior distribution of the parameters are generated and the integrals involved in the expected loss are evaluated from these samples empirically. Much of the calculus, similar to that presented in the Appendices, can be dispensed with, in exchange for approximate results (due to the stochastic nature of MCMC) and concerns related to the convergence of the chain(s). Although MCMC evaluations are computationally much more demanding they should not restrict the scope of sensitivity analysis, which we regard as an integral element of our method.

References

Andrade, M., Longford, N. & Tovar, D. (2014), 'Tests for spatial and temporal correlation in mixed models for climate data', *Revista Brasileira de Biometria* **32**, 229–255.

Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, 2 edn, Springer-Verlag, New York.

Crainiceanu, C. & Ruppert, D. (2004), 'Likelihood ratio tests in linear mixed models with one variance component', *Journal of the Royal Statistical Society. Series B* **66**, 165–185.

DeGroot, M. (1970), *Optimal Statistical Decisions*, McGraw-Hill, New York.

Dempster, A., Laird, N. & Rubin, D. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society. Series B* **39**, 1–38.

Fisher, R. (1935), 'The fiducial argument in statistical inference', *Annals of Eugenics* **6**, 391–398.

Fisher, R. (1956), *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh.

Giampaoli, V. & Singer, J. (2009), 'Likelihood ratio tests for variance components in linear mixed models', *Journal of Statistical Planning and Inference* **139**, 1435–1448.

Greven, S., Crainiceanu, C., Küchenhoff, H. & Peters, H. (2008), 'Restricted likelihood ratio testing for zero variance components in linear mixed models', *Journal of Computational and Graphical Statistics* **17**, 870–891.

Hannig, J. (2009), 'On generalized fiducial inference', *Statistica Sinica* **19**, 491–544.

Li, K., Meng, X.-L., Raghunathan, T. & Rubin, D. (1991), 'Significance levels from repeated p values with multiply-imputed data', *Statistica Sinica* **1**, 65–92.

Lindley, D. (1985), *Making Decisions*, 2 edn, Wiley, Chichester, UK.

Little, R. & Rubin, D. (2002), *Statistical Analysis with Missing Data*, Wiley, New York.

Longford, N. (2000), 'On estimating standard errors in multilevel analysis', *The Statistician* **49**, 389–398.

Longford, N. (2005), 'Editorial: Model selection and efficiency. Is 'Which model ...? the right question?', *Journal of the Royal Statistical Society. Series A* **68**, 469–472.

Longford, N. (2010), 'Bayesian decision making about small binomial rates with uncertainty about the prior', *The American Statistician* **64**, 164–169.

Longford, N. (2012*a*), 'Comparing normal random samples, with uncertainty about the priors and utilities', *Scandinavian Journal of Statistics* **39**, 729–742.

Longford, N. (2012*b*), 'Which model? is the wrong question', *Statistica Neerlandica* **66**, 237–252.

Longford, N. (2013), *Statistical Decision Theory*, Springer-Verlag, Heidelberg, Germany.

Patterson, D. & Thompson, R. (1971), 'Recovery of inter-block information when block sizes are unequal', *Biometrika* **58**, 545–554.

Potthoff, R., Woodbury, M. & Manton, K. (1992), 'Equivalent sample size and equivalent degrees of freedom refinements for inference using survey weights under superpopulation models', *Journal of American Statistical Association* **87**, 383–396.

Robert, C. & Cassella, G. (2004), *Monte Carlo Statistical Methods*, 2 edn, Springer-Verlag, New York.

Rubin, D. (2002), *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York.

Searle, S., Casella, G. & McCulloch, C. (2006), *Variance Components*, 2 edn, Wiley, New York.

Seidenfeld, T. (1992), 'R.A. Fisher's fiducial argument and Bayes' theorem', *Statistical Science* **7**, 358–368.

Appendix A. Maximum Likelihood Estimation

We derive the score function for the model in (1). The joint distribution of the $n \times 1$ vector of outcomes $\mathbf{y}$ is normal with block-diagonal matrix with (identical) blocks $\sigma_W^2 \mathbf{W}_k = \sigma_W^2 \mathbf{W}_1 = \sigma_W^2 (\mathbf{I}_m + \omega \mathbf{J}_m)$ corresponding to clusters $k = 1, \ldots, K$; $\mathbf{I}_m$ is the $m \times m$ identity matrix and $\mathbf{J}_m$ the $m \times m$ matrix of unities. Denote by $\mathbf{1}_m$ the $m \times 1$ vector of unities, so that $\mathbf{J}_m = \mathbf{1}_m \mathbf{1}_m^\top$. Let $\mathbf{e}_k = \mathbf{y}_k - \mu \mathbf{1}_m$ be the vector of residuals in cluster k.

The loglikelihood is $l = l_1 + \cdots + l_K$, with cluster-level contributions

$$
\begin{aligned}
l_k &= -\frac{1}{2} \left\{ m \log \left(2\pi\sigma_W^2 \right) + \log \left(\det \mathbf{W}_1 \right) + \frac{1}{\sigma_W^2} \mathbf{e}_k^\top \mathbf{W}_1^{-1} \mathbf{e}_k \right\} \\
&= -\frac{1}{2} \left[m \log \left(2\pi\sigma_W^2 \right) + \log(1 + m\omega) + \frac{1}{\sigma_W^2} \left\{ \mathbf{e}_k^\top \mathbf{e}_k - \frac{\omega}{1 + m\omega} \left(\mathbf{e}_k^\top \mathbf{1}_m \right)^2 \right\} \right] ,
\end{aligned}
$$

using the identities $\det(\mathbf{W}_1) = 1 + m\omega$ and $\mathbf{W}_1^{-1} = \mathbf{I}_m - \omega/(1 + m\omega)\,\mathbf{J}_m$. The respective score functions for σ_W^2 and ω are

$$
\frac{\partial l}{\partial \sigma_W^2} = -\frac{1}{2\sigma_W^4} \left\{ n\sigma_W^2 - \mathbf{e}^\top \mathbf{e} + \frac{\omega}{1 + m\omega} \sum_{k=1}^{K} \left(\mathbf{e}_k^\top \mathbf{1}_m \right)^2 \right\}
$$

$$
\frac{\partial l}{\partial \omega} = \frac{1}{2} \left\{ -\frac{n}{1 + m\omega} + \frac{1}{\sigma_W^2} \frac{1}{(1 + m\omega)^2} \sum_{k=1}^{K} \left(\mathbf{e}_k^\top \mathbf{1}_m \right)^2 \right\} .
$$

The expressions for the roots of these equations are given in (2).

Appendix B. A Link Among F Densities

We prove the identities in (5). By Γ_2 we denote the half-gamma function, $\Gamma_2(x) = \Gamma(\frac{1}{2}x)$. The density of the F distribution with k_1 and k_2 degrees of freedom is

$$
f_{k_1,k_2}(y) = \frac{\Gamma_2(k_1 + k_2)}{\Gamma_2(k_1)\,\Gamma_2(k_2)} \left(\frac{k_1}{k_2} \right)^{k_1/2} \frac{y^{k_1/2 - 1}}{\left(1 + \frac{k_1}{k_2} y \right)^{(k_1 + k_2)/2}} . \tag{12}
$$

Hence

$$
\begin{aligned}
y f_{k_1,k_2}(y) &= \frac{\Gamma_2(k_1 + k_2)}{\Gamma_2(k_1 + 2)\Gamma_2(k_2 - 2)}, \left(\frac{k_1 + 2}{k_2 - 2} \right)^{k_1/2 + 1} \frac{(h_1 y)^{k_1/2}}{\left(1 + \frac{k_1 + 2}{k_2 - 2} h_1 y \right)^{(k_1 + k_2)/2}} \\
&= \frac{k_1}{k_1 + 2} f_{k_1 + 2, k_2 - 2}(h_1 y),
\end{aligned} \tag{13}
$$

where

$$
h_1 = \frac{k_1}{k_2} \frac{k_2 - 2}{k_1 + 2} .
$$

The identity in (13) is obtained by matching the expression for $y f_{k_1,k_2}(y)$ with another F density. First, the factor $y^{k_1/2}$ implies $k_1 + 2$ degrees of freedom instead of k_1; then the term $1 + y k_1/k_2$ in the denominator implies the argument $h_1 y$ instead of y, and its exponent $(k_1 + k_2)/2$ implies the change from k_2 degrees of freedom to $k_2 - 2$. The constant factors remaining from the match with $f_{k_1+2,k_2-2}(h_1 y)$ reduce to $k_1/(k_1 + 2)$. By reusing the identity in (13) we obtain

$$
\begin{aligned}
y^2 f_{k_1,k_2}(y) &= \frac{k_1}{k_1 + 2}\, y f_{k_1+2,k_2-2}(h_1 y) \\
&= \frac{k_1}{h_1(k_1 + 4)}\, f_{k_1+4,k_2-4}(h_2 y),
\end{aligned}
\tag{14}
$$

where

$$
h_2 \;=\; h_1 \frac{k_1 + 2}{k_2 - 2}\frac{k_2 - 4}{k_1 + 4} \;=\; \frac{k_1}{k_2}\frac{k_2 - 4}{k_1 + 4}.
$$

Appendix C. Posterior Distribution of ω

Using the terminology associated with the Bayes' theorem, the conditional distribution $(\hat{\omega}\,|\,\omega)$ is derived from (4). Its density is

$$
m u f_{K-1,n-K}(u + m u x) = \frac{K}{K - 1}\frac{m}{1 + m\omega}\, f_{K-1,n-K}\!\left(\frac{K}{K - 1}\frac{1 + mx}{1 + m\omega}\right).
$$

For ω we choose the noninformative prior $g(y) = I(y > 0)$, where I denotes the indicator function; its result is unity when its argument as a statement is correct, and is equal to zero otherwise. The posterior distribution of ω is the standardised product of the densities of $(\hat{\omega}\,|\,\omega)$ and ω:

$$
\begin{aligned}
g(y\,|\,\hat{\omega}) &= \frac{Dm}{1 + my}\frac{1}{B_2(K-1, n-K)}\left(\frac{K}{n-K}\right)^{(K-1)/2}\left(\frac{1 + m\hat{\omega}}{1 + my}\right)^{(K-1)/2 - 1} \\
&\quad \times \left(1 + \frac{K}{n-K}\frac{1 + m\hat{\omega}}{1 + my}\right)^{-(n-1)/2} I(y > 0)\,,
\end{aligned}
$$

where $B_2(k_1, k_2) = \Gamma_2(k_1 + k_2)/\Gamma_2(k_1)/\Gamma_2(k_2)$ and $D = D(\hat{\omega}; K, n)$ is the standardising function (the denominator in the Bayes' theorem). By rearranging the penultimate factor as a power of a linear function of $1 + my$, we obtain

$$
\begin{aligned}
g(y; \hat{\omega}) &= \frac{Dm}{1 + m\hat{\omega}}\frac{I(y > 0)}{B_2(n-K, K-1)}\left(\frac{n-K}{K}\frac{1 + my}{1 + m\hat{\omega}}\right)^{(n-K)/2} \\
&\quad \times \left(1 + \frac{n-K}{K}\frac{1 + my}{1 + m\hat{\omega}}\right)^{-(n-1)/2}.
\end{aligned}
$$

Except for the indicator $I(y > 0)$ and a scalar D', this matches the density

$$
\frac{H_0 m}{1 + m\hat{\omega}}\, f_{n-K+2,K-3}\!\left(H_0 \frac{1 + my}{1 + m\hat{\omega}}\right)^{-(n-1)/2},
\tag{15}
$$

where
$$H_0 = \frac{n-K}{n-K+2}\,\frac{K-3}{K}.$$

Let Y be a random variable with F distribution with $n-K+2$ and $K-3$ degrees of freedom. Then the function in (15) is the density of

$$Z = \frac{1+m\hat{\omega}}{H_0}\,Y - \frac{1}{m}.$$

The posterior distribution of ω has the same distribution as Z, except for the condition that $Z > 0$, that is, $Y > H_0 m/(1+m\hat{\omega})$. Therefore, the posterior density of ω is

$$g_0(y;\hat{\omega}) = \frac{H_0 m}{1+m\hat{\omega}}\,\frac{f_{n-K+2,K-3}\,(H_0 z)}{1 - F_{n-K+2,K-3}\{H_0/(1+m\hat{\omega})\}},$$

where $z = (1+my)/(1+m\hat{\omega})$.

A Class of Informative Priors

A class of proper densities for which the posterior density can be obtained in a closed form is

$$g(y) = \frac{m(q-1)}{(1+my)^q}\,I(y>0)$$

for $q > 1$. For orientation, a selection of densities is drawn in Figure 1. For q close to unity, the prior is highly dispersed (the density for $q = 1.2$ is drawn by a gray line). The densities are decreasing for $y > 0$ and $g(0) = m(q-1)$. For greater q, the density has greater mass around zero and has a steeper slope for small y. Although this is quite a flexible class of functions, the dependence on m is an obvious drawback. Note that q should not be set to a large value (say, $q > 5$), because it corresponds to information comparable to that in a very large sample. In the derivations below, we assume that $q < (n-K)/2 + 1$.

To reduce the typographical length of some subscripts, we introduce the notation $k_1(q) = n-K-2q+2$ and $k_2(q) = K+2q-3$, and drop the argument q when its value is not specified or is obvious from the context. Following the outline for the non-informative prior, we obtain

$$g(y;\hat{\omega},q) = \frac{D_q m}{(1+m\hat{\omega})^{q+1}}\,\frac{I(y>0)}{B_2(n-K,K-1)}\left(\frac{n-K}{K}\,\frac{1+my}{1+m\hat{\omega}}\right)^{(n-K-2q)/2}$$

$$\times \left(1 + \frac{n-K}{K}\,\frac{1+my}{1+m\hat{\omega}}\right)^{-(n-1)/2}$$

$$= \frac{D'_q H_q m}{(1+m\hat{\omega})^{q+1}}\,f_{k_1,k_2}\left(H_q\,\frac{1+my}{1+m\hat{\omega}}\right)^{-(n-1)/2},$$

where
$$H_q = \frac{n-K}{n-K-2q+2}\,\frac{K+2q-3}{K}$$

generalises the definition of H_0. Hence, the posterior density of ω is

$$g_q(y;\hat{\omega}) = \frac{H_q m}{1 + m\hat{\omega}} \frac{f_{k_1,k_2}(H_q z)}{1 - F_{k_1,k_2}\{H_q m/(1 + m\hat{\omega})\}},$$

so long as $q < (n-K)/2 + 1$. In view of Figure 1 this is not a restrictive condition.

Appendix D. The Balance Function for Piecewise Linear Loss

Denote $G(\omega) = H_q z(\omega;\hat{\omega})$. The expected loss with action B is

$$
\begin{aligned}
L_\mathrm{B} &= \int_0^{\omega_0} (\omega_0 - y)\, g(y;\hat{\omega})\, \mathrm{d}y \\
&= \frac{Hm}{1 + m\hat{\omega}} \frac{1}{1 - F_{k_1,k_2}\{G(0)\}} \int_0^{\omega_0} (\omega_0 - y) f_{k_1,k_2}\{G(y)\}\, \mathrm{d}y \\
&= \frac{1}{1 - F_{k_1,k_2}\{G(0)\}} \int_{Hz(0,\hat{\omega})}^{G(\omega_0)} \left\{ \omega_0 + \frac{1}{m} - \frac{u}{Hm}(1 + m\hat{\omega}) \right\} f_{k_1,k_2}(u)\, \mathrm{d}u \\
&= \left(\omega_0 + \frac{1}{m} \right) \frac{F_{k_1,k_2}\{G(\omega_0)\} - F_{k_1,k_2}\{G(0)\}}{1 - F_{k_1,k_2}\{G(0)\}} \\
&\quad - \frac{k_1}{k_1 + 2} \frac{1 + m\hat{\omega}}{Hm} \frac{1}{1 - F_{k_1,k_2}\{G(0)\}} \int_{G(0)}^{G(\omega_0)} f_{k_1+2,k_2-2}(h_1 u)\, \mathrm{d}u \\
&= \left(\omega_0 + \frac{1}{m} \right) \frac{F_{k_1,k_2}\{G(\omega_0)\} - F_{k_1,k_2}\{G(0)\}}{1 - F_{k_1,k_2}\{G(\omega_0)\}} \\
&\quad - \frac{K}{n - K} \frac{n - K + 2q - 2}{K + 2q - 5} \frac{1 + m\hat{\omega}}{m} \\
&\quad \times \frac{F_{k_1+2,k_2-2}\{h_1 G(\omega_0)\} - F_{k_1+2,k_2-2}\{h_1 G(0)\}}{1 - F_{k_1,k_2}\{G(0)\}},
\end{aligned}
$$

where we dropped the index q in G, H, k_1 and k_2.

The expected loss when action A is chosen is obtained by similar steps:

$$
\begin{aligned}
L_\mathrm{A} &= \frac{HmR}{1 + m\hat{\omega}} \frac{1}{1 - F_{k_1,k_2}\{G(0)\}} \int_{\omega_0}^{+\infty} (y - \omega_0) f_{k_1,k_2}\{G(y)\}\, \mathrm{d}y \\
&= \frac{R(1 + m\hat{\omega})}{Hm} \frac{1}{1 - F_{k_1,k_2}\{G(0)\}} \int_{G(\omega_0)}^{+\infty} u f_{k_1,k_2}(u)\, \mathrm{d}u
\end{aligned}
$$

$$= -\frac{R(1+m\omega_0)}{m}\frac{1-F_{k_1,k_2}\{G(\omega_0)\}}{1-F_{k_1,k_2}\{G(0)\}}$$

$$= \frac{R(1+m\hat{\omega})}{Hm}\frac{k_1}{k_1+2}\frac{1}{1-F_{k_1,k_2}\{G(0)\}}\int_{G(\omega_0)}^{+\infty}f_{k_1+2,k_2-2}(h_1 u)\,\mathrm{d}u$$

$$\quad - \frac{R(1+m\omega_0)}{m}\frac{1-F_{k_1,k_2}\{G(\omega_0)\}}{1-F_{k_1,k_2}\{G(0)\}}$$

$$= \frac{RK}{n-K}\frac{n-K+2q-2}{K+2q-5}\frac{1+m\hat{\omega}}{m}\frac{1-F_{k_1+2,k_2-2}\{h_1 G(\omega_0)\}}{1-F_{k_1,k_2}\{G(0)\}}$$

$$\quad - \frac{R(1+m\omega_0)}{m}\frac{1-F_{k_1,k_2}\{G(\omega_0)\}}{1-F_{k_1,k_2}\{G(0)\}},$$

assuming that $k_2 > 2$, that is, $K > 5 - 2q$.

The balance function is

$$\Delta L = \frac{1}{1-F_{k_1,k_2}\{G(0)\}}\left(\frac{K}{n-K}\frac{n-K-2q+2}{K+2q-5}\frac{1+m\hat{\omega}}{m}\right.$$

$$\times\left[R-(R-1)F_{k_1+2,k_2-2}\{h_1 G(\omega_0)\}-F_{k_1+2,k_2-2}\{h_1 G(0)\}\right]$$

$$\left. -\frac{1+m\omega_0}{m}\left[R-(R-1)F_{k_1,k_2}\{G(\omega_0)\}-F_{k_1,k_2}\{G(0)\right]\right).$$

The leading factor can be dropped in a search for the root of $\Delta L(\hat{\omega})$.

Appendix E. The Balance Function for Piecewise Quadratic Loss

When action B is chosen,

$$L_{\mathrm{B}} = \int_0^{\omega_0}(y-\omega_0)^2 g(y;\hat{\omega})\,\mathrm{d}y$$

$$= \frac{Hm}{1+m\hat{\omega}}\frac{1}{1-F_{k_1,k_2}\{G(0)\}}$$

$$\int_0^{\omega_0}\left\{\left(\omega_0+\tfrac{1}{m}\right)^2-2\left(y+\tfrac{1}{m}\right)\left(\omega_0+\tfrac{1}{m}\right)+\left(y+\tfrac{1}{m}\right)^2\right\}f_{k_1,k_2}\{G(y)\}\,\mathrm{d}y$$

$$= \frac{(1+m\omega_0)^2}{m^2}\frac{F_{k_1,k_2}\{G(\omega_0)\}-F_{k_1,k_2}\{G(0)\}}{1-F_{k_1,k_2}\{G(0)\}}$$

$$\quad - \frac{2(1+m\omega_0)(1+m\hat{\omega})}{Hm^2}\frac{1}{1-F_{k_1,k_2}\{G(0)\}}\int_{G(0)}^{G(\omega_0)}u f_{k_1,k_2}(u)\,\mathrm{d}u$$

$$+ \left(\frac{1+m\hat{\omega}}{Hm}\right)^2 \frac{1}{1-F_{k_1,k_2}\{G(0)\}} \int_{G(0)}^{G(\omega_0)} u^2 f_{k_1,k_2}(u)\, \mathrm{d}u$$

$$= \frac{1}{1-F_{k_1,k_2}\{G(0)\}} \left(\frac{(1+m\omega)^2}{m^2} \left[F_{k_1,k_2}\{G(\omega_0)\} - F_{k_1,k_2}\{G(0)\}\right]\right.$$

$$- \frac{2k_2}{k_2-2} \frac{(1+m\omega_0)(1+m\hat{\omega})}{Hm^2}$$

$$\times \left[F_{k_1+2,k_2-2}\{h_1 G(\omega_0)\} - F_{k_1+2,k_2-2}\{h_1 G(0)\}\right]$$

$$+ \frac{k_1}{H^2 h_1 h_2(k_1+4)} \left(\frac{1+m\hat{\omega}}{m}\right)^2$$

$$\times \left. \left[F_{k_1+4,k_2-4}\{h_2 G(\omega_0)\} - F_{k_1+4,k_2-4}\{h_2 G(0)\}\right]\right).$$

The constant factor of the concluding term is

$$\frac{k_1}{H^2 h_1 h_2(k_1+4)} = \frac{k_1+2}{k_1} \frac{k_1(k_1+2)}{(k_2-2)(k_2-4)}.$$

By similar steps we obtain the expression

$$L_{\mathrm{A}} = \frac{R}{1-F_{k_1,k_2}\{G(0)\}} \left(\frac{(1+m\omega)^2}{m^2}\left[1-F_{k_1,k_2}\{G(\omega_0)\}\right]\right.$$

$$- \frac{2k_2}{k_2-2} \frac{(1+m\omega_0)(1+m\hat{\omega})}{Hm^2}\left[1-F_{k_1+2,k_2-2}\{h_1 G(\omega_0)\}\right]$$

$$+ \left.\frac{k_1}{H^2 h_1 h_2(k_1+4)}\left(\frac{1+m\hat{\omega}}{m}\right)^2\left[1-F_{k_1+4,k_2-4}\{h_2 G(\omega_0)\}\right]\right),$$

and hence

$$\Delta L = \frac{1}{1-F_{k_1,k_2}\{G(0)\}}$$

$$\times \left(\frac{(1+m\omega_0)^2}{m^2}\left[R-(R+1)F_{k_1,k_2}\{G(\omega_0)\}+F_{k_1,k_2}\{G(0)\}\right]\right.$$

$$- \frac{2k_2}{k_2-2}\frac{(1+m\omega_0)(1+m\hat{\omega})}{Hm^2} \tag{16}$$

$$\times \left[R-(R+1)F_{k_1+2,k_2-2}\{h_1 G(\omega_0)\}+F_{k_1+2,k_2-2}\{h_1 G(0)\}\right]$$

$$+ \frac{k_1}{h_1 h_2(k_1+4)}\left(\frac{1+m\hat{\omega}}{Hm}\right)^2$$

$$\times \left.\left[R-(R+1)F_{k_1+4,k_2-4}\{h_2 G(\omega_0)\}+F_{k_1+4,k_2-4}\{h_2 G(0)\}\right]\right).$$

It holds only when $K > 5 - 2q$; otherwise L_{A} is infinite and B is the preferred action. The factor $[1-F_{k_1,k_2}\{G(0)\}]^{-1}$ is not relevant for finding the root of ΔL,

and can be dropped. Figure 10 displays the equilibrium values for a selection of designs (K, m) and ω_0 set to 0.1 and 0.25. The equilibrium functions are decreasing and reach their minima of $-1/m$ for finite R. In the diagram, the functions are interrupted at that point. For $K = 9$, $m = 7$ and $\omega_0 = 0.1$, no line is drawn because the equilibrium is equal to $-1/m$ even for very small $\hat{\omega}$.
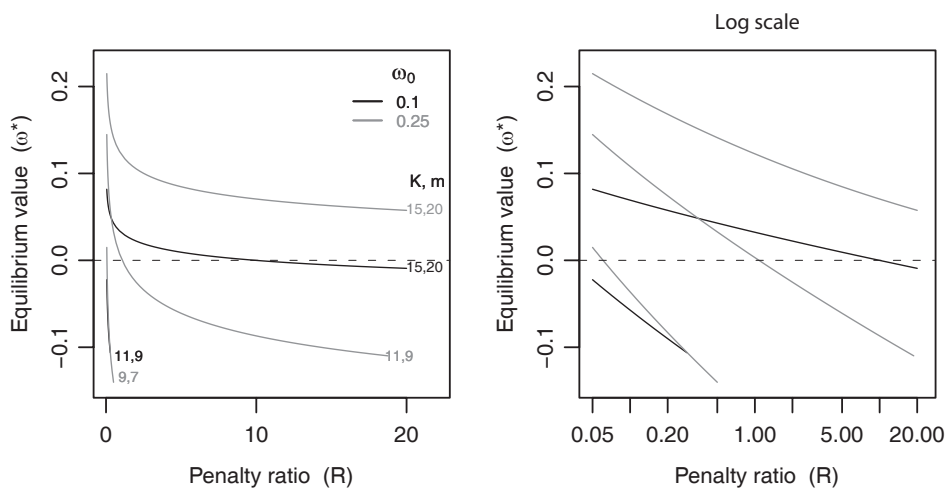

FIGURE 10: The equilibrium values ω^* as functions of the penalty ratio R for number of clusters K and cluster size m indicated in the left-hand panel. Piecewise quadratic loss and non-informative prior for ω.