Taylor & Francis
Taylor & Francis Group

Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Objective Bayesian hypothesis testing and estimation for the intraclass model

Duo Zhang, Daojiang He, Xiaoqian Sun, Tao Lu & Min Wang

STATISTICAL THEORY AND RELATED FIELDS
2018, VOL. 2, NO. 1, 37–47
https://doi.org/10.1080/24754269.2018.1481586

Taylor & Francis
Taylor & Francis Group

Check for updates

Objective Bayesian hypothesis testing and estimation for the intraclass model

Duo Zhang [a], Daojiang He [b], Xiaoqian Sun [c], Tao Lu [d] and Min Wang [e]

[a]Department of Mathematical Sciences, Michigan Technological University, Houghton, MI, USA; [b]School of Mathematics and Computer Science, Anhui Normal University, Wuhu, People's Republic of China; [c]Department of Mathematical Sciences, Clemson University, Clemson, SC, USA; [d]Department of Mathematics and Statistics, University of Nevada, Reno, NV, USA; [e]Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX, USA

ABSTRACT

The intraclass correlation coefficient (ICC) plays an important role in various fields of study as a coefficient of reliability. In this paper, we consider objective Bayesian analysis for the ICC in the context of normal linear regression model. We first derive two objective priors for the unknown parameters and show that both result in proper posterior distributions. Within a Bayesian decision-theoretic framework, we then propose an objective Bayesian solution to the problems of hypothesis testing and point estimation of the ICC based on a combined use of the intrinsic discrepancy loss function and objective priors. The proposed solution has an appealing invariance property under one-to-one reparametrisation of the quantity of interest. Simulation studies are conducted to investigate the performance the proposed solution. Finally, a real data application is provided for illustrative purposes.

ARTICLE HISTORY

Received 6 November 2017
Accepted 24 May 2018

KEYWORDS

Intraclass correlation coefficient; Bayesian reference criterion; intrinsic discrepancy; hypothesis testing; intrinsic estimator; reference priors

1. Introduction

Consider the intraclass model of the form

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $\mathbf{Y}_i$ is a $k \times 1$ vector of response variables, $\mathbf{X}_i$ is a $k \times p$ design matrix of $(p-1)$ regressors (assuming the first column is ones) and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown common regression coefficients. We assume that the random error $\boldsymbol{\varepsilon}_i \overset{\text{iid}}{\sim} N(\mathbf{0}_k, \sigma^2\boldsymbol{\Sigma})$, where $\overset{\text{iid}}{\sim}$ stands for 'independent and identically distributed', $\mathbf{0}_k$ is a $k \times 1$ vector of zeros, and $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k$ with $\mathbf{I}_k$ being a $k \times k$ identity matrix and $\mathbf{J}_k$ being a $k \times k$ matrix containing only ones. The parameter ρ is often referred as the intraclass correlation coefficient (ICC). Note that $\rho \in (-(k-1)^{-1}, 1)$ is the necessary and sufficient condition for positive-definiteness of $\boldsymbol{\Sigma}$. When ρ is equal to 0, the intraclass model becomes the classical linear normal model with independent errors.

The ICC has been widely applied in various fields of study as a coefficient of reliability, from epidemiologic research to genetic studies; see, for example Barkto (1966), Fleiss (1986), Lin, Hedayat, Sinha, Yang (2002), among others. The analysis of the ICC transitionally consists of two branches, hypothesis testing and point estimation, and it has received attentions from two main statistical streams of thought: frequentists and Bayesians. From a frequentist viewpoint, Paul (1990) considered the maximum likelihood estimate (MLE) of the ICC in a generalised model setting by solving iteratively a single estimating equation. Paul (1996) developed the score tests for testing the significance of the interclass correlation in familial data. For Bayesian methods, Jelenkowska (1998) studied Bayesian estimation of the ICC in the linear mixed model. Chung and Dey (1998) considered Bayesian analysis of the ICC using the reference prior under a balanced variance components model. Later on, Ghosh and Heo (2003) considered Bayesian credible intervals for ρ based on different objective priors and made comparisons among these priors in terms of matching the corresponding frequentist coverage probabilities.

It deserves mentioning that the problems of hypothesis testing and point estimation for ρ have not yet been studied within a decision-theoretical viewpoint. This motivates us to propose an objective Bayesian solution to these problems based on the Bayesian reference criterion (for short, BRC) (Bernardo & Rueda, 2002). The proposed solution allows the researchers to simultaneously study important inference summaries of the ICC, including point estimation, credible interval estimation and precise hypotheses. In addition, it enjoys various appealing properties: (i) it is invariant under one-to-one reparametrisation of the parameter of interest ρ; (ii) it depends only on the assumed model, appropriate objective priors and the observed data; (iii) it is appropriate to perform the hypothesis test: $H_0 : \rho = \rho_0$ versus $H_1 : \rho \neq \rho_0$ for any $\rho_0 \in (-(k-1)^{-1}, 1)$ and

CONTACT Min Wang min.wang@ttu.edu, minwang@mtu.edu Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX, USA

(iv) it can be easily approximated numerically in most statistical software and can thus be implemented by the practitioners from different fields.

The remainder of the paper is organised as follows. In Section 2, we derive two objective priors of the unknown parameters and discuss the propriety of their corresponding posterior distributions. In Section 3, we propose an objective Bayesian solution to both hypothesis testing and estimation problems of ρ from a decision-theoretical viewpoint. Section 4 investigates the performance of the proposed solution through simulations and a real data application. Some concluding remarks are provided in Section 5, with additional proofs given in the Appendix.

2. Posterior distribution

For notational convenience, let $\mathbf{Y}$ and $\boldsymbol{\varepsilon}$ be $nk \times 1$ vectors and $\mathbf{X}$ is an $nk \times p$ design matrix, and they are given by

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix},$$

respectively. The model in (1) can be expressed in a more compact way as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{\varepsilon}$ follows an nk-dimensional normal distribution with mean vector $\mathbf{0}_{nk}$ and covariance matrix $\sigma^2\boldsymbol{\Phi}$, where $\boldsymbol{\Phi} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}$ is an nk-dimensional matrix and $\otimes$ denotes the Kronecker product. The likelihood function of the intraclass model in (2) is given by

$$\begin{aligned} p(\mathbf{Y} \mid \boldsymbol{\beta}, \sigma^2, \rho) &\propto |\sigma^2\boldsymbol{\Phi}|^{-1/2} \exp \\ &\quad \times \left\{ -\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Phi}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ &\propto (\sigma^2)^{-nk}(1-\rho)^{-n(k-1)/2} \\ &\quad \times (1 + (k-1)\rho)^{-n/2} \exp \\ &\quad \times \left\{ -\frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Phi}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\}, \end{aligned}$$

where $|\mathbf{A}|$ denotes the determinant of a matrix $\mathbf{A}$.

Bayesian analysis begins with prior specification for all the unknown parameters in the model. In the absence of relevant prior knowledge for $(\boldsymbol{\beta}, \sigma^2, \rho)$ in the above model, noninformative priors are often preferred. One of the most popular noninformative priors is the Jeffreys prior, which is proportional to the square root of the determinant of the Fisher information matrix. It can be shown that the Jeffreys prior is given by

$$\begin{aligned} \pi_J(\rho, \sigma^2, \boldsymbol{\beta}) &\propto (\sigma^2)^{-(p+2)/2}(1-\rho)^{-1} \\ &\quad \times (1 + (k-1)\rho)^{-1}|\mathbf{X}'\boldsymbol{\Phi}^{-1}\mathbf{X}|^{1/2}. \end{aligned} \tag{3}$$

Given that the parameter of interest is ρ, we integrate out $\boldsymbol{\beta}$ and σ^2 (i.e., $\pi_J(\rho \mid D) \propto \int\int f(\mathbf{Y} \mid \boldsymbol{\beta}, \sigma^2, \rho)\pi_J(\rho, \sigma^2, \boldsymbol{\beta})\, d\boldsymbol{\beta}\, d\sigma^2$) and obtain the marginal posterior density for ρ, denoted by $\pi_J(\rho \mid D)$, where D represents the observable data. It follows that

$$\begin{aligned} \pi_J(\rho \mid D) &\propto (1-\rho)^{-n(k-1)/2-1} \\ &\quad \times (1 + (k-1)\rho)^{-n/2-1}\mathbf{S}(\rho)^{-nk/2}, \end{aligned} \tag{4}$$

where $\mathbf{S}(\rho) = \mathbf{Y}'(\boldsymbol{\Phi}^{-1} - \boldsymbol{\Phi}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Phi}^{-1})\mathbf{Y}$. Note that when $\mathbf{X}_1 = \cdots = \mathbf{X}_n$, the prior in (4) can be simplified by replacing $\mathbf{S}(\rho)$ with $(\mathbf{Y} - \mathbf{X}'\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}'\hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{Y}}$ and $\bar{\mathbf{Y}} = \sum_{i=1}^n \mathbf{Y}_i/n$. The simplified version is just the Jeffreys prior derived by Ghosh and Heo (2003).

One may argue that, when we aim at a subset of the parameters with the rest treated as nuisance parameters, the direct use of the Jeffreys prior may sometimes be unsatisfactory. To overcome such a pitfall, Bernardo (1979) proposed an algorithm to derive objective priors by maximising some entropy distances. This was further explored by Berger and Bernardo (1992a, 1992b) and named by them the reference priors. We obtain that the one-at-a-time reference prior for the parameter ordering $\{\rho, \sigma^2, \boldsymbol{\beta}\}$ or $\{\rho, \boldsymbol{\beta}, \sigma^2\}$ is given by

$$\pi_R(\rho, \sigma^2, \boldsymbol{\beta}) \propto (\sigma^2)^{-1}(1-\rho)^{-1}(1 + (k-1)\rho)^{-1}, \tag{5}$$

which is exactly the same as the reference prior identified by Ghosh and Heo (2003), because their model is just a special case of model in (1) when we set $\mathbf{X}_1 = \cdots = \mathbf{X}_n$. In addition, it can be shown that the prior in (5) is a second-order matching prior because it achieves approximate frequentist validity of the posterior quantiles of the interest parameter ρ with a margin of error of $o(n^{-1})$. We refer the interested readers to Datta and Ghosh (1995b), Datta and Ghosh (1995a) and Datta and Mukerjee (2004) about the second-order matching criterion in detail. The resulting marginal posterior density of ρ under this prior, denoted by $\pi_R(\rho \mid D)$, is given by

$$\begin{aligned} \pi_R(\rho \mid D) &\propto (1-\rho)^{-n(k-1)/2-1} \\ &\quad \times (1 + (k-1)\rho)^{-n/2-1}|\mathbf{X}'\boldsymbol{\Phi}^{-1}\mathbf{X}|^{-1/2} \\ &\quad \times \mathbf{S}(\rho)^{-(nk-p)/2}. \end{aligned} \tag{6}$$

Given that neither π_J in (3) nor π_R in (5) is proper, it is important to study the propriety of their corresponding posterior distributions, which is summarised in the following theorem with proofs given in the Appendix.

Theorem 2.1: *Consider the intraclass linear model in* (1). *Under either the Jeffreys prior π_J in* (3) *or the reference prior π_R in* (5) *for the unknown parameters, the joint posterior distribution of $(\rho, \sigma^2, \boldsymbol{\beta})$ is proper when $k \geq 2$.*

As commented by Bernardo (2010), the problems of hypothesis testing and point estimation can be viewed as a special decision problem from a Bayesian decision-theoretic point of view. The choice of the loss function plays a central role in the statistical decision theory. There are numerous loss functions, such as the squared error loss, the zero-one loss and the absolute error loss, whereas many of them often lack the invariance property required in practice. For example, the squared error loss is often overused in statistical inference as a measure of the discrepancy between two sampling distributions, heavily depending on the chosen parameterisations (Bernardo, 2005). In this paper, we consider the intrinsic discrepancy as a loss function due to its various appealing properties discussed in the next section.

3. Bayesian reference criterion

In this section, we propose an objective Bayesian solution based on the BRC proposed by Bernardo and Rueda (2002). In Section 3.1, we overview the BRC and derive the intrinsic discrepancy for the hypothesis testing of ρ. We then obtain Bayesian intrinsic statistic in Section 3.2 and Bayesian intrinsic estimator of ρ in Section 3.3.

3.1. Intrinsic discrepancy loss function

Without loss of generality, we assume that the probabilistic behaviour of observable data $\mathbf{y}$ can be appropriately described by the probability model

$$M \equiv \{p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\omega}), \mathbf{y} \in \mathbf{Y}, \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\omega} \in \boldsymbol{\Omega}\}, \qquad (7)$$

where $\boldsymbol{\theta}$ is the parameter of interest and $\boldsymbol{\omega}$ is a nuisance parameter. We aim at deciding whether or not to treat the reduced model $p(\mathbf{y} \mid \boldsymbol{\theta}_0, \boldsymbol{\omega})$ under $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ as a proxy for the general model M. In other words, we decide whether the model under H_0 is compatible with the observable data. Since the Kullback–Leibler (KL) direct divergence is a good measure of discrepancy between two probability distributions (Robert, 1996), Bernardo (1999) developed the logarithmic discrepancy derived by minimising this divergence measure. Given that the logarithmic discrepancy is not symmetric and this feature may be unsuitable in some contexts, Bernardo and Rueda (2002) developed a symmetric version, often called the intrinsic discrepancy given by

$$\delta(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\theta}_0) = \min\{\kappa(\boldsymbol{\theta}_0 \mid \boldsymbol{\omega}, \boldsymbol{\theta}), \kappa(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \boldsymbol{\theta}_0)\},$$

where

$$\kappa(\boldsymbol{\theta}_0 \mid \boldsymbol{\omega}, \boldsymbol{\theta}) = \inf_{\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}} \int p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\omega}) \log \frac{p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\omega})}{p(\mathbf{y} \mid \boldsymbol{\theta}_0, \boldsymbol{\omega}_0)} \, \mathrm{d}\mathbf{y}$$

and

$$\kappa(\boldsymbol{\theta}, \boldsymbol{\omega} \mid \boldsymbol{\theta}_0) = \inf_{\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}} \int p(\mathbf{y} \mid \boldsymbol{\theta}_0, \boldsymbol{\omega}_0) \log \times \frac{p(\mathbf{y} \mid \boldsymbol{\theta}_0, \boldsymbol{\omega}_0)}{p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\omega})} \, \mathrm{d}\mathbf{y}.$$

The unit of the intrinsic discrepancy is the nat of information, while it could be a bit of information if the logarithm was taken in base 2 instead of base e. The intrinsic discrepancy has an invariant property under one-to-one reparametrisation. For a thorough discussion of other properties, see Bernardo and Rueda (2002), Bernardo and Juárez (2003) and Bernardo (2010). In what follows, we provide the intrinsic discrepancy between two intraclass models with its derivations given in the Appendix.

Theorem 3.1: *The intrinsic discrepancy for testing $H_0 : \rho = \rho_0$ versus $H_1 : \rho \neq \rho_0$, for $\rho_0 \in (-(k-1)^{-1}, 1)$ under the intraclass model in* (1) *is given by*

$$\delta(\rho_0, \rho) = \begin{cases} \kappa(\rho_0 \mid \rho) & \text{if } \rho \in \left(-\frac{1}{k-1}, \rho_0\right], \\ \kappa(\rho \mid \rho_0) & \text{if } \rho \in (\rho_0, 1), \end{cases} \qquad (8)$$

where

$$\kappa(\rho \mid \rho_0) = \frac{nk}{2} \log \left\{ \frac{1 + (k-2)\rho - (k-1)\rho_0\rho}{(1 + (k-1)\rho)(1-\rho)} \right\} - \frac{n}{2} \log \left\{ \frac{(1 + (k-1)\rho_0)(1-\rho_0)^{k-1}}{(1 + (k-1)\rho)(1-\rho)^{k-1}} \right\}. \qquad (9)$$

It can be easily verified that $\rho_0 \mapsto \delta(\rho_0, \rho)$ is a continuous convex function with a unique minimum at $\rho = \rho_0$. Figure 1 depicts the curves $\rho_0 \mapsto \delta(\rho_0, \rho)$ for $n = 1$, $k = 4$ and $\rho \in \{-0.3, 0, 0.3\}$. We observe that the corresponding curve of the intrinsic discrepancy always vanishes at $\rho_0 = \rho$.

3.2. Bayesian intrinsic statistic

If we select the intrinsic discrepancy as the loss function, then the intrinsic statistic can be defined as the posterior expectation of the intrinsic discrepancy loss, namely,

$$d(\rho_0 \mid D) = \int_{\Theta} \delta(\rho, \rho_0) \pi_\delta(\rho \mid D) \, \mathrm{d}\rho, \qquad (10)$$

where $\pi_\delta(\rho \mid D)$ is the marginal posterior distribution for ρ under the δ-reference prior when the quantity of interest is $\delta(\rho_0, \rho)$ in (8). Because $\delta(\rho_0, \rho)$ is a one-to-one piecewise function of ρ, we follow Proposition 1

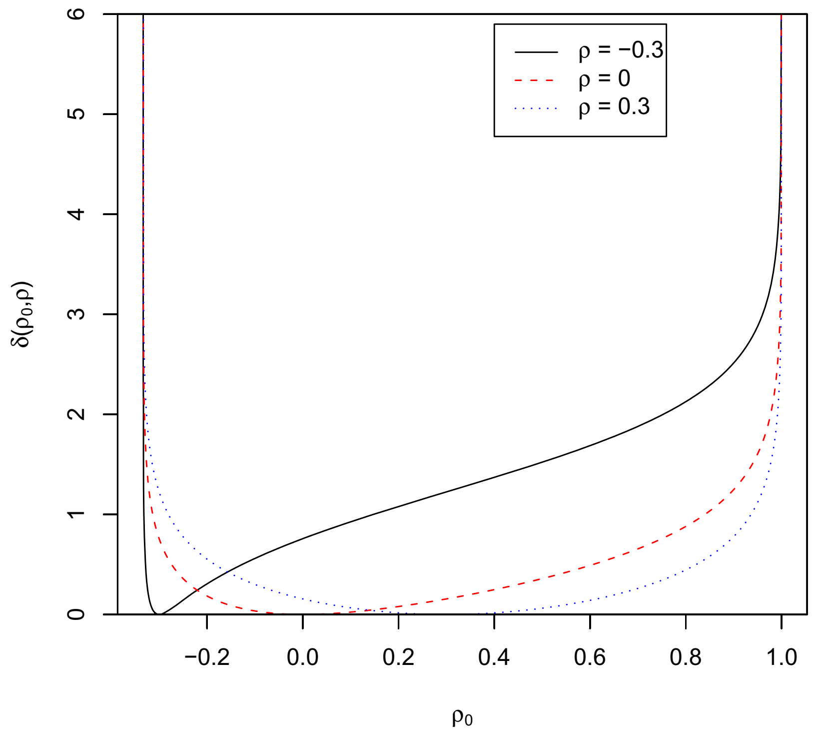


Figure 1. The intrinsic discrepancy $\delta(\rho_0, \rho)$ in (8) as a function of ρ_0 for $n = 1$, $k = 4$ and $\rho \in \{-0.3, 0, 0.3\}$.

of Bernardo (1999) and show that the δ-reference prior corresponding to the parameter of interest $\delta(\rho_0, \rho)$ is exactly the same as the reference prior for ρ corresponding to the parameter of interest ρ. In addition, the posterior distribution of ρ is invariant under this kind of transformations (Bernardo & Smith, 1994, p. 326). The intrinsic statistic in (10) can thus be rewritten as

$$\begin{aligned} d(\rho_0 \mid D) &= \int_\Theta \delta(\rho_0, \rho)\pi_\delta(\rho \mid D)\,\mathrm{d}\rho \\ &= \int_\Theta \delta(\rho_0, \rho)\pi(\rho \mid D)\,\mathrm{d}\rho \\ &= \int_{-1/(k-1)}^{\rho_0} \kappa(\rho_0 \mid \rho)\pi(\rho \mid D)\,\mathrm{d}\rho \\ &\quad + \int_{\rho_0}^{1} \kappa(\rho \mid \rho_0)\pi(\rho \mid D)\,\mathrm{d}\rho, \end{aligned}$$

where $\pi(\rho \mid D)$ is the marginal posterior distribution of ρ under either π_J in (3) or π_R in (5). We observe from Bernardo (2010) that the intrinsic statistic can be interpreted as the expected value of the log-likelihood ratio against the simplified model under H_0. On the other hand, the BRC can be defined as

$$\text{Reject } H_0 : \rho = \rho_0 \quad \text{when} \quad d(\rho_0 \mid D) > d^*$$

for some given utility constant d^*. In this paper, we advocate the conventional choices $d^* \in \{\log(10), \log(100), \log(1000)\}$ for scientific communication. The value of about $\log(10)$ indicates some evidence against H_0; the value of about $\log(100)$ provides rather strong evidence against H_0, while the value of about $\log(1000)$ can be safely used to reject H_0. For further details about these values, we refer the interested readers to Bernardo and Rueda (2002), Bernardo and Juárez (2003), Bernardo and Pérez (2007) and Bernardo (2010).

3.3. Bayesian intrinsic estimator

We follow Bernardo and Juárez (2003) and define the intrinsic estimator of ρ as

$$\rho^* = \rho^*(D) = \arg\min_{\rho_0 \in \Theta} d(\rho_0 \mid D), \qquad (11)$$

which is the value minimising the posterior expectation of the intrinsic discrepancy loss function. The intrinsic estimator inherits the invariance property of the intrinsic statistic under one-to-one piecewise transformation, which means that if $\psi = \psi(\rho)$ is a one-to-one reparametrisation of ρ, then the intrinsic estimator of ψ is simply $\psi^* = \psi(\rho^*)$.

4. Examples

We examine the performance of the proposed solution to both hypothesis testing and point estimation problems of ρ through simulation studies (Section 4.1) and a real data application (Section 4.2).

4.1. Simulation study

We conduct simulation studies to investigate the behaviour of the proposed solution under different scenarios. There are n observations and 2 regressors

($p=3$) and the data are generated from the model in (1). Without loss of generality, we set $\sigma^2 = 1$, $\boldsymbol{\beta} = (1, 1, 1)'$ and $\boldsymbol{\Sigma} = (1 - \rho_T)\mathbf{I}_3 + \rho_T\mathbf{J}_3$, where ρ_T is the prespecified true value of ICC. Each element of $\mathbf{X}_i$ for $i = 1, \ldots, n$ is generated from a uniform density over the interval $(-2, 2)$. To check the variations of the proposed approach, ρ_T is taken to be one of four different values: $-0.3, 0, 0.3, 0.8$ corresponding to the correlation being negative, zero, medium and large, respectively, while considering different sample sizes $n=5$ (small) and $n=20$ (medium). For each simulation setting, we consider $N=10{,}000$ replications. We analyse the averaged estimates along with the mean absolute errors (MAE) given by

$$MAE = \frac{1}{N}\sum_{j=1}^{N} |\hat{\rho}_j - \rho_T|,$$

where $\hat{\rho}_j$ represents the estimate of ρ_T in jth replication.

The MAEs of the Bayesian estimations and the MLE (Paul, 1990) are reported in Tables 1 and 2. Several features can be drawn as follows. (i) The intrinsic estimator under π_R outperforms the one under π_J in most cases, especially when the sample size is small, and they behave similarly as n increases. (ii) The intrinsic estimator under each prior outperforms the posterior mode and is comparable with the posterior median. (iii) When the true value ρ_T is near by 0, the MLE performs the best, whereas when ρ_T is far from 0 (e.g., $\rho_T = 0.8$), the intrinsic estimator performs the best among all the estimators under consideration. (iv) On average, the MAEs of all the estimators decrease significantly with an increasing sample size. In a marked contrast with other estimators, the intrinsic one is invariant under one-to-one transformation, which is not shared by others, such as the posterior mean. Simulations with other choices of ρ have also been conducted, and similar conclusions are achieved and thus not presented here for simplicity.

We further compare the frequentist coverage probability of the posterior distributions of ρ under π_J and π_R. Following Sun and Ye (1996), we let α be the left tail probability and $\rho^{(\alpha)}$ be the corresponding quantile of the marginal posterior distribution $\pi(\rho \mid D)$ under either π_J or π_R. Theoretically, it follows $F(\rho^{(\alpha)}) = \int_{-\infty}^{\rho^{(\alpha)}} \pi(\rho \mid D)d\rho = \alpha$. Letting $P(\alpha \mid \rho_T) = P(\rho < \rho^{(\alpha)} \mid \rho_T, D) = P(F(\rho) < \alpha \mid \rho_T, D) = P(\int_{-\infty}^{\rho} \pi(\rho \mid D)d\rho < \alpha \mid \rho_T, D)$, we observe that $P(\alpha \mid \rho_T)$ should be very close to α if the chosen prior performs well with respect to the probability matching criterion. Table 3 shows the estimated tail probabilities of the posterior distributions between two priors under different scenarios. We observe that the tail probabilities of the posterior distribution of ρ under π_R are closer to the frequentist coverage probabilities than the ones under π_j. This observation is reasonable, because π_R is a second-order matching prior if ρ is the parameter of interest.

In addition to the parameter estimation, the proposed solution can be used to test any value of $\rho = \rho_0 \in (-0.5, 1)$ since $k=3$ in our simulation study. For illustrative purposes, suppose that we are interested in evaluating whether the data are compatible with $H_0 : \rho = 0$. We analyse the frequentist behaviour of the proposed solution under π_R for the hypothesis testing of ρ based on two scenarios discussed below.

First, consider the scenario in which $H_0 : \rho = 0$ is true. We simulate 5000 random samples from the

Table 2. The MAE of the MLE for ρ based on 10,000 replications in the simulation study.

	$n=5$	$n=20$
−0.3	0.137	0.058
0	0.198	0.102
0.3	0.231	0.118
0.8	0.236	0.074

Table 3. The estimated tail probabilities of posterior distributions based on 10,000 replications in the simulation study.

		$n=5$		$n=20$	
ρ_T	Prior	$P(0.05 \mid \rho_T)$	$P(0.90 \mid \rho_T)$	$P(0.05 \mid \rho_T)$	$P(0.90 \mid \rho_T)$
−0.3	π_R	0.0453	0.9127	0.0477	0.9010
	π_J	0.0497	0.9145	0.0425	0.9166
0	π_R	0.0460	0.9069	0.0535	0.8977
	π_J	0.0842	0.8598	0.0617	0.8879
0.3	π_R	0.0439	0.9119	0.0453	0.9054
	π_J	0.1021	0.8357	0.0614	0.8816
0.8	π_R	0.0441	0.9087	0.0484	0.9001
	π_J	0.1341	0.7825	0.0779	0.8587

Table 1. The MAE of the Bayesian estimators for ρ based on 10,000 replications in the simulation study.

		$n=5$				$n=20$			
ρ_T	Prior	Intrinsic	Mean	Median	Mode	Intrinsic	Mean	Median	Mode
−0.3	π_R	0.148	0.155	0.149	0.162	0.058	0.060	0.058	0.059
	π_J	0.164	0.163	0.166	0.185	0.060	0.060	0.060	0.062
0	π_R	0.242	0.213	0.243	0.333	0.108	0.105	0.108	0.115
	π_J	0.294	0.263	0.296	0.379	0.114	0.111	0.114	0.121
0.3	π_R	0.268	0.230	0.268	0.377	0.119	0.115	0.119	0.129
	π_J	0.315	0.276	0.315	0.412	0.124	0.119	0.124	0.133
0.8	π_R	0.148	0.157	0.148	0.151	0.057	0.059	0.057	0.057
	π_J	0.142	0.141	0.141	0.153	0.056	0.056	0.056	0.058

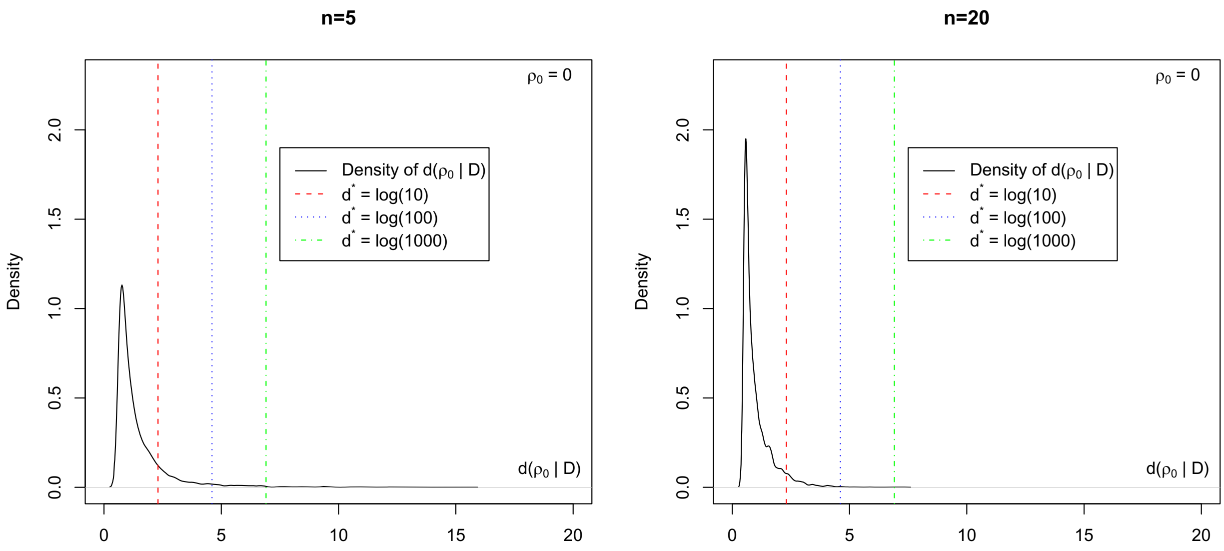


Figure 2. Sampling distribution of $d(\rho \mid D)$ under H_0 obtained from the 5000 simulations with $\rho_T = 0$ for different sample sizes when testing $H_0 : \rho = 0$.

model in (1) with $\rho_T = 0$ based on the simulation setup above. Figure 2 depicts the sampling distribution of $d(\rho \mid D)$ from the 5000 simulations. For $n = 5$, the significance level is around 13.24% for $d^* = \log(10)$ (mild evidence); the significance level is around 3.26% for $d^* = \log(100)$ (strong evidence) and the significance level is around 0.88% for $d^* = \log(1000)$ (safe to reject H_0). We observe that as n increases ($n = 20$), the significance level approximately goes down to 5.20%, 0.26% and 0.06%, respectively. As one would expect, the significance level significantly decreases as n increases from a frequentist viewpoint.

Second, consider the scenario in which $H_0 : \rho = 0$ is not true. We study the behaviour of the sampling distribution of the proposed solution and the relative frequency of the rejection of H_0. We again simulate 5000 random samples from the model in (1) with $\rho_T \in \{-0.3, 0.3, 0.8\}$. Figure 3 shows the sampling distribution of $d(\rho \mid D)$ from the 5000 simulations. Note that the power of the proposed approach increases when ρ_T is far from the testing value $\rho_0 = 0$ or n is larger. For instance, when $H_0 : \rho = 0$ while $\rho_T = 0.8$, for $n = 5$, the relative frequency of rejecting H_0 is approximately equal to 79.46% for $d^* = \log(10)$, to 35.32% for $d^* = \log(100)$, and to 6.56% for $d^* = \log(1000)$; for $n = 20$, this relative frequency significantly increases to 100%, 99.84% and 98.68%, respectively. We may thus conclude that the power of the proposed solution increases with n and that the performance of the proposed solution is quite satisfactory for the problems of hypothesis testing and point estimation of ρ in the intraclass model in (1).

Given that there are two objective priors: the reference prior (π_R) or the Jeffreys prior (π_J), which of them is preferable for the proposed solution in practical applications? Numerical evidence from the above simulation studies showed that the Bayesian estimations under π_R outperform the ones under π_R. Additionally, π_R is also a second-order matching prior if ρ is the parameter of interest. We thus have a preference to recommend the use of π_R in the analysis of the ICC.

4.2. *An illustrative example*

We use a real data example to illustrate the practical application of the proposed solution. The orthodontic data set is present in Table 4 and obtained from Chapter 5.2 of Frees (2004): 27 individuals including 16 boys and 11 girls were measured for distances from the pituitary to the pteryomaxillary fissure in millimetres, at ages 8, 10, 12 and 14. We consider the intraclass model of the form

$$\mathbf{y}_i = \beta_0 \mathbf{j}_4 + \beta_1 \mathbf{A}_i + \beta_2 \mathrm{G}_i \mathbf{j}_4 + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, 27,$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4})^{\mathrm{T}}$ with y_{ij} being the distance for individual i measured at age j, $\mathbf{A}_i = (8,\ 10,\ 12,\ 14)^{\mathrm{T}}$ is a 4×1 vector of ages and G_i represents the gender (1 for male and 0 for female), and $\boldsymbol{\varepsilon}_i \overset{\text{iid}}{\sim} N(\mathbf{0}_4, \sigma^2 \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (1 - \rho)\mathbf{I}_4 + \rho \mathbf{J}_4$. We observe from Figure 4(a) that the marginal posterior densities for ρ under two objective priors are quite normal in shape. Table 5 provides the point estimators for ρ under different procedures. We here analyse the results under π_R for simplicity. The intrinsic estimator $\rho^* = 0.622$ is close to the posterior median equal to 0.620, whereas both are slightly different from the MLE equal to 0.597. According to the non-rejection regions with $d^* \in \{\log(10), \log(100), \log(1000)\}$ presented in Figure 4(b), we somehow doubt that the true value of ρ is outside $R_{\log(10)} = (0.423,\ 0.773)$; we seriously

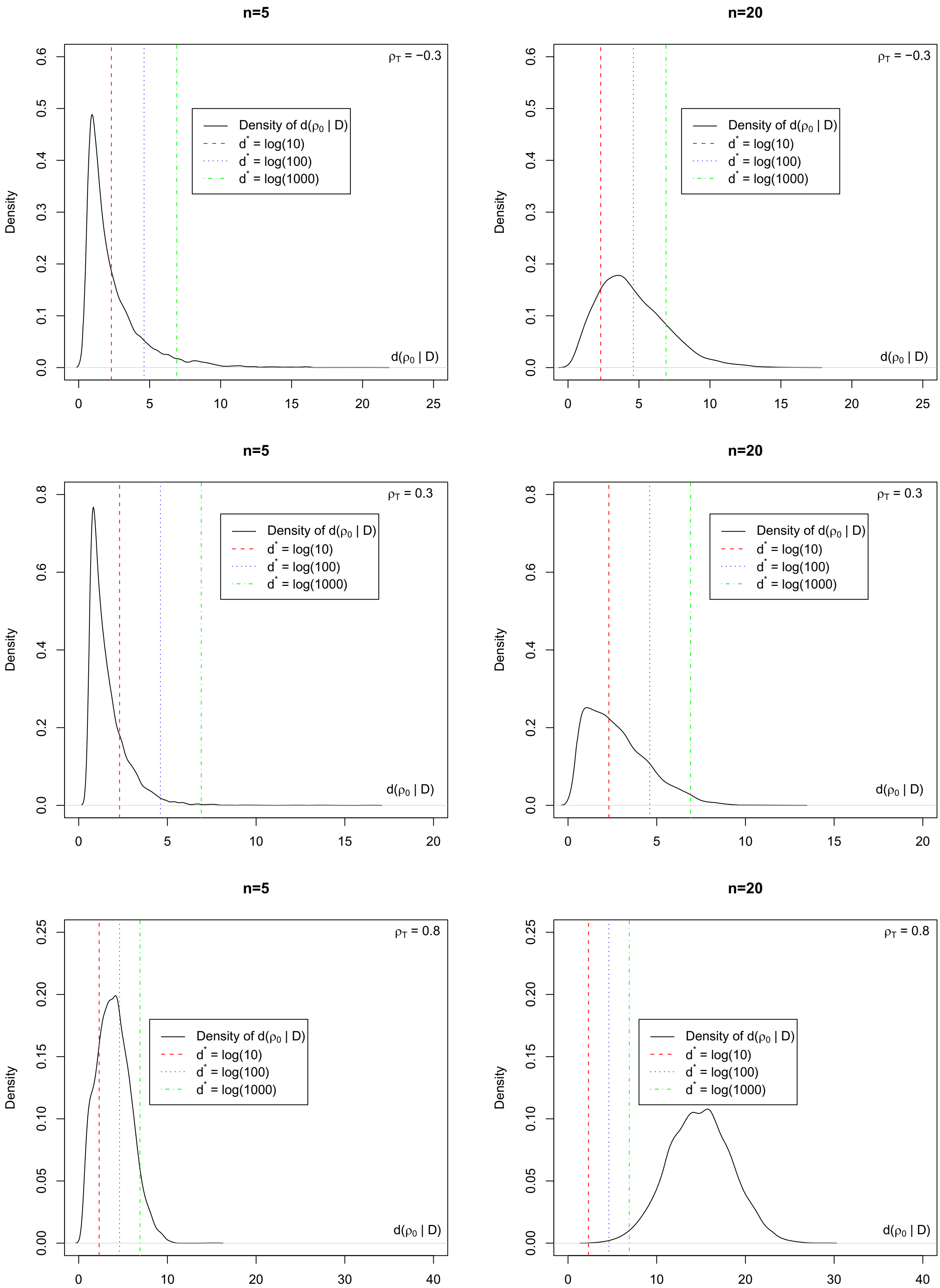


Figure 3. Sampling distribution of $d(\rho \mid D)$ under H_0 obtained from 5000 simulations with $\rho_T \in \{-0.3, 0.3, 0.8\}$ for different sample sizes when testing $H_0 : \rho = 0$.

doubt that ρ is outside $R_{\log(100)} = (0.304,\ 0.833)$, and we are almost sure that the true correlation value ρ is not outside $R_{\log(1000)} = (0.211,\ 0.870)$.

On the other hand, the proposed solution can be used for the hypothesis testing of $\rho = \rho_0 \in (-1/3, 1)$. If we are interested in testing $H_0 : \rho = \rho_0 = 0$ versus

Table 4. The orthodontic data from Frees (2004).

	Age of girls				Age of boys			
Number	8	10	12	14	8	10	12	14
1	21	20	21.5	23	26	25	29	31
2	21	21.5	24	25.5	21.5	22.5	23	26.5
3	20.5	24	24.5	26	23	22.5	24	27.5
4	23.5	24.5	25	26.5	25.5	27.5	26.5	27
5	21.5	23	22.5	23.5	20	23.5	22.5	26
6	20	21	21	22.5	24.5	25.5	27	28.5
7	21.5	22.5	23	25	22	22	24.5	26.5
8	23	23	23.5	24	24	21.5	24.5	25.5
9	20	21	22	21.5	23	20.5	31	26
10	16.5	19	19	19.5	27.5	28	31	31.5
11	24.5	25	28	28	23	23	23.5	25
12					21.5	23.5	24	28
13					17	24.5	26	29.5
14					22.5	25.5	25.5	26
15					23	24.5	26	30
16					22	21.5	23.5	25

$H_1 : \rho \neq \rho_0$, we can numerically verify that the intrinsic statistic under π_R is

$$d(\rho_0 \mid D) = \int_{-1/3}^{1} \delta(\rho_0, \rho)\pi(\rho \mid D)\,\mathrm{d}\rho \approx 14.2747 \approx \log(1582791),$$

which indicates that the expected value of the average of the log likelihood ratio against H_0 is about 14.2747, showing that the likelihood ratio is expected to be about 1,582,791. Thus we may conclude that the data provide very strong evidence against H_0 and that the null hypothesis is opposed to the observable data. Due to the invariance property of the proposed solution, if the parameter of interest is ρ^3, then its intrinsic estimator is simply $(\rho^*)^3 \approx 0.622^3$, and the corresponding non-rejection regions are simply given

Table 5. Estimations of ρ for the orthodontic data from Frees (2004).

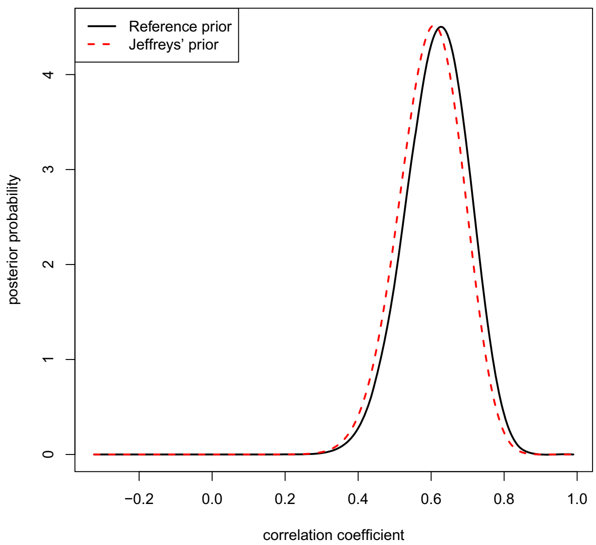
Priors	Intrinsic	Mean	Median	Mode
π_J	0.603	0.598	0.601	0.608
π_R	0.622	0.616	0.620	0.627

by $\tilde{R}_{\log(10)} = (0.076,\ 0.462)$, $\tilde{R}_{\log(100)} = (0.028,\ 0.578)$ and $\tilde{R}_{\log(1000)} = (0.009,\ 0.659)$, respectively.

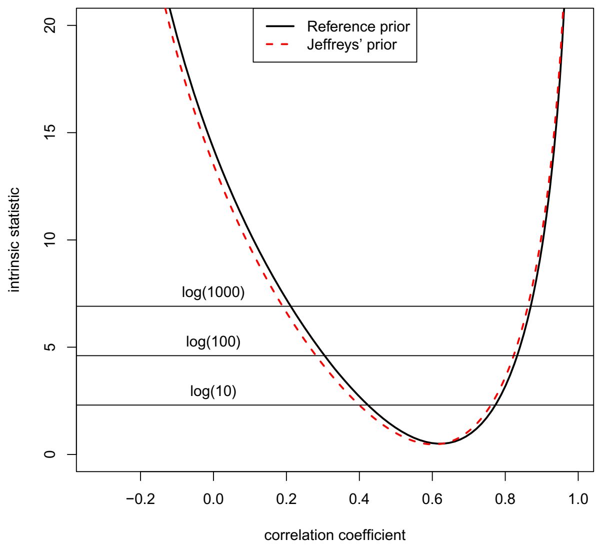
5. Concluding remarks

In this paper, we first derived two objective priors for the unknown parameters in the intraclass model in (1) and proved that both result in proper posterior distributions. Within a Bayesian decision-theoretic framework, we then proposed an objective Bayesian solution to both hypothesis testing and point estimation problems of the ICC ρ. The proposed solution has an appealing invariance property under one-to-one reparametrisation of the quantity of interest, which is not shared by some commonly used estimators, such as the posterior mean.

It deserves mentioning that the proposed solution can be directly applied to the balanced one-way random effect ANOVA model, since it is a special case of the intraclass model in (1) if we let $\sigma^2 = \sigma_a^2 + \sigma_e^2$ and $\rho = \sigma_a^2/\sigma^2 \in (0, 1)$, where σ_a^2 and σ_e^2 stand for the treatment and error variances, respectively. This observation motivates a possible extension of the proposed solution to the unbalanced model with different number of observations in each class, which is currently under investigation and will be reported elsewhere.



(a)



(b)

Figure 4. The marginal posterior density for ρ based on two objective priors (left), and the intrinsic statistic with the non-rejection regions corresponding to the threshold values $d^* \in \{\log(10), \log(100), \log(1000)\}$ (right) for the orthodontic data in Frees (2004): (a) marginal posterior distribution and (b) intrinsic statistic.

Acknowledgements

The authors thank the editor, an associate editor and one referee for their detailed and constructive comments, which have led to a significant improvement of the manuscript. This work was based on the first author's dissertation research which was supervised by the corresponding author. MW and DZ initiated and carried out the study. MW and DZ drafted the manuscript. DH, TL, and XS participated in the discussion and proofread the manuscript. All authors read and approved the final manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Duo Zhang http://orcid.org/0000-0002-0249-1849
Min Wang http://orcid.org/0000-0002-9233-7844

References

Barkto, J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, *19*, 2–11.

Berger, J. O., & Bernardo, J. M. (1992a). On the development of reference priors. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics, 4 (Peñíscola, 1991)* (pp. 35–60). New York: Oxford University Press.

Berger, J. O., & Bernardo, J. M. (1992b). Reference priors in a variance components problem. In P. K. Goel & N. Sreenivas Iyengar (Eds.), *Lecture notes in statistics: Vol. 75. Bayesian analysis in statistics and econometrics (Bangalore, 1988)* (pp. 177–194). New York: Springer.

Bernardo, J. (2010). Integrated objective Bayesian estimation and hypothesis testing. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, & M. West (Eds.), *Bayesian statistics, 9. Proceedings of the ninth valencia international meeting* (pp. 1–68). New York: Oxford University Press.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, *41*, 113–147.

Bernardo, J. M. (1999). Nested hypothesis testing: The Bayesian reference criterion. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics, 6 (Alcoceber, 1998)* (pp. 101–130). New York: Oxford University Press.

Bernardo, J. M. (2005). Reference analysis. In D. K. Dey & C. R. Rao (Eds.), *Handbook of Statistics: Vol. 25. Bayesian thinking: Modeling and computation* (pp. 17–90). Amsterdam: Elsevier/North-Holland.

Bernardo, J. M., & Juárez, M. A. (2003). Intrinsic estimation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, & M. West (Eds.), *Bayesian statistics, 7 (Tenerife, 2002)* (pp. 465–476). New York: Oxford University Press.

Bernardo, J. M., & Pérez, S. (2007). Comparing normal means: New methods for an old problem. *Bayesian Analysis*, *2*, 45–58.

Bernardo, J. M., & Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *International Statistical Review*, *70*, 351–372.

Bernardo, J. M., & Smith, A. F. (1994). *Bayesian theory*. Chichester: Wiley.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.

Chung, Y., & Dey, D. K. (1998). Bayesian approach to estimation of intraclass correlation using reference prior. *Communications in Statistics: Theory and Methods*, *27*, 2241–2255.

Datta, G. S., & Ghosh, J. K. (1995a). Noninformative priors for maximal invariant parameter in group models. *TEST*, *4*, 95–114.

Datta, G. S., & Ghosh, J. K. (1995b). On priors providing frequentist validity for Bayesian inference. *Biometrika*, *82*, 37–45.

Datta, G. S., & Mukerjee, R. (2004). Probability matching priors: Higher order asymptotics. *Lecture notes in statistics: Vol. 178*. New York: Springer-Verlag. .

Fleiss, J. (1986). *The design and analysis of clinical experiments*. New York: Wiley.

Frees, E. W. (2004). *Longitudinal and panel data: Analysis and applications in the social sciences*. New York: Cambridge University Press.

Ghosh, M., & Heo, J. (2003). Noninformative priors, credible sets and Bayesian hypothesis testing for the intraclass model. *Journal of Statistical Planning and Inference*, *112*, 133–146.

Jelenkowska, T. H. (1998). Bayesian estimation of the intraclass correlation coefficients in the mixed linear model. *Applications of Mathematics*, *43*, 103–110.

Lin, L., Hedayat, A. S., Sinha, B., & Yang, M. (2002). Statistical methods in assessing agreement: Models, issues, and tools. *Journal of the American Statistical Association*, *97*, 257–270.

Paul, S. R. (1990). Maximum likelihood estimation of intraclass correlation in the analysis of familial data: Estimating equation approach. *Biometrika*, *77*, 549–555.

Paul, S. R. (1996). Score tests for interclass correlation in familial data. *Biometrics*, *52*, 955–963.

Robert, C. P. (1996). Intrinsic losses. *Theory and Decision*, *40*, 191–214.

Sun, D., & Ye, K. (1996). Frequentist validity of posterior quantiles for a two-parameter exponential family. *Biometrika*, *83*, 55–65.

Appendix

In the appendix, we prove that the posterior distribution is proper under π_R in (5), since the case for π_J is exactly the same and thus omitted for simplicity. We first provide a very useful lemma, which plays an important role in determining the tail behaviour of the key terms of the marginal posterior distribution $\pi_R(\rho \mid D)$.

Lemma A.1: *The marginal posterior distribution $\pi_R(\rho \mid D)$ in (6) is a continuous function in $(-1/(k-1), 1)$ and their terms are such that $|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2} = O((1-\rho)^{p/2})$ and $\mathbf{S}(\rho) = O((1-\rho)^{-1})$ as $\rho \to 1$, and such that $|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2} = O((1+\rho(k-1))^{p/2})$ and $\mathbf{S}(\rho) = O((1+\rho(k-1))^{-1})$ as $\rho \to -1/(k-1)$.*

Proof: Direct inspection shows that $\pi_R(\rho \mid D)$ in (6) is a continuous function in $(-1/(k-1), 1)$. We consider the behaviour of its two key terms as (i) $\rho \to 1$ and (ii) $\rho \to -1/(k-1)$.

(i) Let $\eta_1 = \rho/(1-\rho)$, which tends to infinity as $\rho \to 1$. Given that $\mathbf{\Sigma} = (1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k = (1-\rho)[\mathbf{I}_k + \rho/(1-\rho)\mathbf{J}_k]$, we have

$$\mathbf{\Sigma}^{-1} = (1-\rho)^{-1}\left(\mathbf{I}_k - \frac{\eta_1}{1+\eta_1 k}\mathbf{J}_k\right).$$

Then it follows that

$$\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{\Sigma}^{-1}\mathbf{X}_i = (1-\rho)^{-1}\sum_{i=1}^{n}\left(\mathbf{X}_i'\mathbf{X}_i - \frac{\eta_1\mathbf{X}_i'\mathbf{J}_k\mathbf{X}_i}{1+\eta_1 k}\right). \tag{A1}$$

As $\eta_1 \to \infty$, we have

$$\left|\sum_{i=1}^{n}\left(\mathbf{X}_i'\mathbf{X}_i - \frac{\eta_1\mathbf{X}_i'\mathbf{J}_k\mathbf{X}_i}{1+\eta_1 k}\right)\right| = O(1),$$

which show that $|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}| = O((1-\rho)^{-p})$, and thus

$$|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2} = O((1-\rho)^{p/2}).$$

In addition, as $\eta_1 \to \infty$, we observe that each element of the inverse matrix in the right hand of Equation (A2) becomes $O(1)$. With a little abuse of notation, as $\eta_1 \to \infty$, we denote

$$\left[\sum_{i=1}^{n}\left(\mathbf{X}_i'\mathbf{X}_i - \frac{\eta_1\mathbf{X}_i'\mathbf{J}_k\mathbf{X}_i}{1+\eta_1 k}\right)\right]^{-1} = O(1),$$

which shows that $(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} = O((1-\rho))$. Note also that $\mathbf{\Phi}^{-1} = \mathbf{I}_n \otimes \mathbf{\Sigma}^{-1} = (1-\rho)^{-1}\mathbf{I}_n \otimes (\mathbf{I}_k - \frac{\eta_1}{1+\eta_1 k}\mathbf{J}_k) = (1-\rho)^{-1}\mathbf{\Phi}_1^{-1}$, where

$$\mathbf{\Phi}_1^{-1} = \mathbf{I}_n \otimes \left(\mathbf{I}_k - \frac{\eta_1}{1+\eta_1 k}\mathbf{J}_k\right) \to \mathbf{I}_n \otimes \left(\mathbf{I}_k - \frac{1}{k}\mathbf{J}_k\right),$$

as $\eta_1 \to \infty$. Also, $(\mathbf{X}'\mathbf{\Phi}_1^{-1}\mathbf{X})^{-1} = (1-\rho)^{-1}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} = O(1)$. Thus, as $\rho \to 1$, it follows

$$\begin{aligned}\mathbf{S}(\rho) &= \mathbf{Y}'(\mathbf{\Phi}^{-1} - \mathbf{\Phi}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1})\mathbf{Y} \\ &= \frac{1}{1-\rho}\mathbf{Y}'(\mathbf{\Phi}_1^{-1} - \mathbf{\Phi}_1^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}_1^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}_1^{-1})\mathbf{Y} \\ &= O((1-\rho)^{-1}).\end{aligned}$$

(ii) Let $\eta_2 = \rho/(1+\rho(k-1))$, which tends to infinity as $\rho \to -1/(k-1)$. Given that

$$\begin{aligned}\mathbf{\Sigma}^{-1} &= (1-\rho)^{-1}\left(\mathbf{I}_k - \frac{\rho}{1+\rho(k-1)}\mathbf{J}_k\right) \\ &= (1-\rho)^{-1}(\mathbf{I}_k - \eta_2\mathbf{J}_k),\end{aligned}$$

it follows that

$$\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X} = \sum_{i=1}^{n}\mathbf{X}_i'\mathbf{\Sigma}^{-1}\mathbf{X}_i = (1-\rho)^{-1}\sum_{i=1}^{n}(\mathbf{X}_i'\mathbf{X}_i - \eta_2\mathbf{X}_i'\mathbf{J}_k\mathbf{X}_i). \tag{A2}$$

As $\eta_2 \to \infty$, we have $|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}| = O(\eta_2^p)$, and thus

$$|\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2} = O(\eta_2^{-p/2}) = O((1+\rho(k-1))^{p/2}).$$

In addition, as $\eta_2 \to \infty$, we observe that $(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} = O(1)$ and that $\mathbf{\Phi}^{-1} = \mathbf{I}_n \otimes \mathbf{\Sigma}^{-1} = (1-\rho)^{-1}\mathbf{I}_n \otimes (\mathbf{I}_k - \eta_2\mathbf{J}_k) = \eta_2\mathbf{\Phi}_2^{-1}$, where

$$\mathbf{\Phi}_2^{-1} = \frac{1}{1-\rho}\mathbf{I}_n \otimes \left(\mathbf{J}_k - \frac{1}{\eta_2}\mathbf{I}_k\right) \to \frac{k-1}{k}\mathbf{I}_n \otimes \mathbf{J}_k.$$

As $\eta_2 \to \infty$, we have $(\mathbf{X}'\mathbf{\Phi}_2^{-1}\mathbf{X})^{-1} = \eta_2(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1} = O(1)$, and thus

$$\begin{aligned}\mathbf{S}(\rho) &= \mathbf{Y}'(\mathbf{\Phi}^{-1} - \mathbf{\Phi}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}^{-1})\mathbf{Y} \\ &= \eta_2\mathbf{Y}'(\mathbf{\Phi}_2^{-1} - \mathbf{\Phi}_2^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Phi}_2^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Phi}_2^{-1})\mathbf{Y} \\ &= O(\eta_2) = ((1+\rho(k-1))^{-1}).\end{aligned}$$

■

Proof of Theorem 2.1: We now show that the posterior distribution under π_R is proper. Recall that the corresponding marginal posterior of ρ is given by

$$\pi_R(\rho \mid D) \propto (1-\rho)^{-n(k-1)/2-1}(1+(k-1)\rho)^{-n/2-1} \times |\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2}\mathbf{S}(\rho)^{-(nk-p)/2}. \tag{A3}$$

Then the reference prior π_R leads to a proper posterior distribution if and only if

$$\int_{-1/(k-1)}^{1}\pi_R(\rho \mid D)\,\mathrm{d}\rho < \infty.$$

By following Lemma A.1, we observe that $\rho \to 1$, the tail behaviour of $\pi_R(\rho \mid D)$ follows

$$\begin{aligned}\pi_R(\rho \mid D) &\propto (1-\rho)^{-n(k-1)/2-1}(1+(k-1)\rho)^{-n/2-1} \\ &\quad \times |\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2}\mathbf{S}(\rho)^{-(nk-p)/2} \\ &= O((1-\rho)^{n/2-1}),\end{aligned}$$

and that $\rho \to -1/(k-1)$, the tail behaviour of $\pi_R(\rho \mid D)$ follows

$$\begin{aligned}\pi_R(\rho \mid D) &\propto (1-\rho)^{-n(k-1)/2-1}(1+(k-1)\rho)^{-n/2-1} \\ &\quad \times |\mathbf{X}'\mathbf{\Phi}^{-1}\mathbf{X}|^{-1/2}\mathbf{S}(\rho)^{-(nk-p)/2} \\ &= O((1+\rho(k-1))^{n(k-1)/2-1}).\end{aligned}$$

Given that $\pi_R(\rho \mid D)$ is a continuous function in $(-1/(k-1), 1)$, the posterior distribution under π_R is proper, provided that $k \geq 2$. This completed the proof of Theorem 2.1. ■

Proof of Theorem 3.1: Define $\mathbf{\Sigma} = (1-\rho)\mathbf{I}_k + \rho\mathbf{J}_k$ and $\mathbf{\Sigma}_0 = (1-\rho_0)\mathbf{I}_k + \rho_0\mathbf{J}_k$. It can be easily verified that

$$\mathrm{tr}(\mathbf{\Sigma}_0^{-1}\mathbf{\Sigma}) = \frac{k(1+(k-2)\rho_0 - (k-1)\rho\rho_0)}{(1-\rho_0)(1+(k-1)\rho_0)} \quad \text{and}$$

$$|\mathbf{\Sigma}_0^{-1}\mathbf{\Sigma}| = \frac{(1+(k-1)\rho)(1-\rho)^{k-1}}{(1+(k-1)\rho_0)(1-\rho_0)^{k-1}},$$

where $\mathrm{tr}(\mathbf{M})$ represents the trace of the matrix $\mathbf{M}$.

Consider that the KL divergence measure of a normal linear model $N_{kn}(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \mathbf{\Sigma}_0))$ from another normal linear model $N_{kn}(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \mathbf{\Sigma}))$ is given by

$$\int p(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \mathbf{\Sigma}))\log\frac{p(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \mathbf{\Sigma}))}{p(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \mathbf{\Sigma}_0))}\,\mathrm{d}\mathbf{y}$$

$$= \frac{1}{2}\left\{\frac{R_0}{\sigma_0^2} + \mathrm{tr}\left(\frac{\sigma^2}{\sigma_0^2}(\mathbf{I}_n \otimes \mathbf{\Sigma}_0)^{-1}(\mathbf{I}_n \otimes \mathbf{\Sigma})\right)\right.$$

$$
\begin{aligned}
&- \log\left|\frac{\sigma^2}{\sigma_0^2}(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)^{-1}(\mathbf{I}_n \otimes \boldsymbol{\Sigma})\right| - kn\Big\} \\
&= \frac{1}{2}\left\{\frac{R_0}{\sigma_0^2} + \mathrm{tr}\left(\frac{\sigma^2}{\sigma_0^2}\mathbf{I}_n \otimes (\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma})\right)\right. \\
&\left. - \log\left|\frac{\sigma^2}{\sigma_0^2}\mathbf{I}_n \otimes (\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma})\right| - kn\right\} \\
&= \frac{1}{2}\left\{\frac{R_0}{\sigma_0^2} + n\frac{\sigma^2}{\sigma_0^2}\mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}) - nk\log\left(\frac{\sigma^2}{\sigma_0^2}\right)\right. \\
&\left. - n\log|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}| - kn\right\},
\end{aligned}
$$

where $R_0 = (\boldsymbol{\beta}_0 - \boldsymbol{\beta})'\mathbf{X}'(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0)^{-1}\mathbf{X}(\boldsymbol{\beta}_0 - \boldsymbol{\beta})$. The minimum of the logarithmic divergence above for $\boldsymbol{\beta}_0 \in R^p$ and $\sigma_0 > 0$ is achieved when

$$\boldsymbol{\beta}_0 = \boldsymbol{\beta} \quad \text{and} \quad \sigma_0 = \sigma\sqrt{\frac{\mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma})}{k}},$$

and substitution yields

$$
\begin{aligned}
\kappa(\rho_0 \mid \sigma^2, \boldsymbol{\beta}, \rho) &= \inf_{\boldsymbol{\beta}_0 \in R^p, \sigma_0 > 0} \frac{1}{2}\left\{\frac{R_0}{\sigma_0^2} + n\frac{\sigma^2}{\sigma_0^2}\mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma})\right. \\
&\quad \left. - nk\log\left(\frac{\sigma^2}{\sigma_0^2}\right) - n\log\left|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}\right| - kn\right\} \\
&= \frac{n}{2}\left\{k\log(\mathrm{tr}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}))\right. \\
&\quad \left. - \log(|\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}|) - k\log(k)\right\} \\
&= \frac{nk}{2}\log\left\{\frac{1 + (k-2)\rho_0 - (k-1)\rho\rho_0}{(1 + (k-1)\rho_0)(1-\rho_0)}\right\} \\
&\quad - \frac{n}{2}\log\left\{\frac{(1 + (k-1)\rho)(1-\rho)^{k-1}}{(1 + (k-1)\rho_0)(1-\rho_0)^{k-1}}\right\},
\end{aligned}
$$

which is the same as $\kappa(\rho_0 \mid \rho)$ in (9).

Similarly, the minimum of the logarithmic divergence measure of $N_{kn}(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}))$ from $N_{kn}(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0))$ is given by

$$
\begin{aligned}
&\int p(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0))\log\frac{p(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}_0, \sigma_0^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_0))}{p(\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I}_n \otimes \boldsymbol{\Sigma}))}\,\mathrm{d}\mathbf{y} \\
&= \frac{1}{2}\left\{\frac{R}{\sigma^2} + n\frac{\sigma_0^2}{\sigma^2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0) - nk\log\left(\frac{\sigma_0^2}{\sigma^2}\right)\right. \\
&\quad \left. - n\log|\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0| - kn\right\},
\end{aligned}
$$

where $R = (\boldsymbol{\beta}_0 - \boldsymbol{\beta})'\mathbf{X}'(\mathbf{I}_n \otimes \boldsymbol{\Sigma})^{-1}\mathbf{X}(\boldsymbol{\beta}_0 - \boldsymbol{\beta})$. The minimum of the divergence measure above for $\boldsymbol{\beta}_0 \in R^p$ and $\sigma_0 > 0$ is achieved when

$$\boldsymbol{\beta}_0 = \boldsymbol{\beta} \quad \text{and} \quad \sigma_0 = \sigma\sqrt{\frac{k}{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0)}},$$

and substitution yields

$$
\begin{aligned}
\kappa(\rho, \sigma^2, \boldsymbol{\beta} \mid \rho_0) &= \frac{n}{2}\left\{k\log(\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0))\right. \\
&\quad \left. - \log(|\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0|) - k\log(k)\right\} \\
&= \frac{nk}{2}\log\left\{\frac{1 + (k-2)\rho - (k-1)\rho_0\rho}{(1 + (k-1)\rho)(1-\rho)}\right\} \\
&\quad - \frac{n}{2}\log\left\{\frac{(1 + (k-1)\rho_0)(1-\rho_0)^{k-1}}{(1 + (k-1)\rho)(1-\rho)^{k-1}}\right\} \\
&= \kappa(\rho \mid \rho_0).
\end{aligned}
$$

Therefore, the intrinsic statistic is given by

$$\delta(\rho, \rho_0) = \delta(\rho, \sigma^2, \boldsymbol{\beta}, \rho_0) = \min\{\kappa(\rho_0 \mid \rho), \kappa(\rho \mid \rho_0)\}.$$

It can be easily shown that $\kappa(\rho \mid \rho_0) \geq \kappa(\rho_0 \mid \rho)$ if and only if $\rho \in (-1/(k-1), \rho_0]$. This completed the proof of Theorem 3.1. ■