Objective Priors for the Bivariate Normal Model with Multivariate Generalizations [1]

James O. Berger
Duke University, Durham, NC 27708, USA
e-mail: berger@stat.duke.edu

and

Dongchu Sun
University of Missouri-Columbia, Columbia, MO 65211, USA
e-mail: dsun@stat.missouri.edu

March 3, 2006

Summary

Study of the bivariate normal distribution raises the full range of issues involving objective Bayesian inference, including the different types of objective priors (e.g., Jeffreys, invariant, reference, matching), the different modes of inference (e.g., Bayesian, frequentist, fiducial), and the criteria involved in deciding on optimal objective priors (e.g., ease of computation, frequentist performance, marginalization paradoxes). Summary recommendations as to optimal objective priors are made for a variety of inferences involving the bivariate normal distribution.

In the course of the investigation, a variety of surprising results were found, including the availability of objective priors that yield exact frequentist inferences for many functions of the bivariate normal parameters, including the correlation coefficient. Several generalizations to the multivariate normal distribution are given.

Some key words: Reference priors, matching priors, Jeffreys priors, right-Haar prior, fiducial inference, frequentist coverage, marginalization paradox, rejection sampling, constructive posterior distributions.

[1] This research was supported by the National Science Foundation, under grants DMS-0103265 and SES-0351523, and the National Institute of Health, under grants R01-CA100760 and R01-MH071418. We are grateful to Fei Liu for performing the numerical frequentist coverage computations, to Xiaoyan Lin for computing the matching priors in Table 6, and to Susie Bayarri for helpful discussions.

Contents

1 Introduction and Prior Distributions

1.1 Notation and Problem Statement

The bivariate normal distribution of (x_1, x_2) has mean parameters (μ_1, μ_2) and covariance matrix $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where ρ is the correlation between x_1 and x_2. The density of (x_1, x_2) is

$$
\begin{aligned}
& f(x_1, x_2 \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \\
= \; & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{\sigma_2^2(x_1-\mu_1)^2 + \sigma_1^2(x_2-\mu_2)^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2)}{2\sigma_1^2\sigma_2^2(1-\rho^2)} \right\}. \quad (1)
\end{aligned}
$$

The data consists of an independent random sample $\boldsymbol{X} = (\boldsymbol{x}_k = (x_{1k}, x_{2k}), k = 1, \cdots, n)$ of size $n \geq 3$, for which the sufficient statistics are

$$\overline{\boldsymbol{x}} = (\overline{x}_1, \overline{x}_2)' \text{ and } \boldsymbol{S} = \sum_{k=1}^{n} (\boldsymbol{x}_k - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})' = \begin{pmatrix} s_{11} & r\sqrt{s_{11}s_{22}} \\ r\sqrt{s_{11}s_{22}} & s_{22} \end{pmatrix}, \tag{2}$$

where, for $i, j = 1, 2$,

$$\overline{x}_i = n^{-1}\sum_{j=1}^{n} x_{ij}, \quad s_{ij} = \sum_{k=1}^{n} (x_{ik} - \overline{x}_i)(x_{jk} - \overline{x}_j), \quad r = \frac{s_{12}}{\sqrt{s_{11}s_{22}}}.$$

We will denote prior densities as $\pi(\mu_1, \mu_2, \sigma_1\ \sigma_2,\ \rho)$, and the corresponding posterior densities as $\pi(\mu_1, \mu_2, \sigma_1\ \sigma_2,\ \rho \mid \mathbf{X})$ (all with respect to $d\mu_1\, d\mu_2\, d\sigma_1\, d\sigma_2\, d\rho$).

We consider objective inference for parameters of the bivariate normal distribution and functions of these parameters, with special focus on development of objective confidence or credible sets. The many results in the paper can be grouped into the following categories.

1. Addressing historical issues and conjectures

2. Obtaining optimal objective Bayesian analyses

3. Obtaining frequentist confidence sets (often exact but, at least, asymptotically optimal)

4. Providing simple computational implementations

The historical issues relate to fiducial inference for the correlation parameter ρ, and conjectures as to whether the fiducial distribution could be derived in a Bayesian fashion. This is addressed in Section 1.2, where it is also observed that the fiducial distribution yields an exact frequentist confidence set for ρ, which seems to have been unrecognized.

The method introduced to show that exact frequentist confidence sets for ρ are available is also utilized to find exact frequentist confidence sets for a variety of other parameters of the bivariate normal distribution. These results are summarized in Section 1.3. It should be stressed that these are exact even for the minimum sample size of $n = 3$ (in contrast to many commonly used confidence procedures that are simply asymptotically correct).

A large number of potential objective prior distributions are considered in the paper. For ease of use, our recommendations as to the priors to use – depending on the desired inference – are summarized in Section 1.4.

Often, the posteriors for the recommended priors are essentially available in computational closed form, allowing direct Monte Carlo simulation. These computational-ready representations of the posteriors (which we call *constructive posterior distributions*) are also given in these introductory sections. Section 2 provides simple accept-reject schemes for computing with the recommended priors when the posteriors are not amenable to direct simulation.

Section 3 and the appendices develop the needed theory (derivation of reference priors, derivation of constructive posterior distributions, proofs of exact frequentist coverage, and proofs of asymptotic frequentist coverage) and also present various simulations that were conducted to enable summary recommendations to be made.

The recommended objective priors for the bivariate normal distribution have extensions to the multivariate normal distribution. These extensions, and algorithms for computing with the resulting posteriors, are given in Section 4.

Technical Issue: We will assume that $|\rho| < 1$ and $|r| < 1$ in virtually all expressions and results that follow. This is because, if either equals 1 in absolute value, then $\rho = \{\text{sign of } r\}$ with probability 1 (either frequentist or Bayesian posterior, as relevant). Indeed, the situation then essentially collapses to the univariate version of the problem in which the randomness is (say) in $f(x_1 \mid \mu_1, \sigma_1)$, with μ_2 and σ_2 (or $\bar{x}_2$ and s_{22}) being determined as $\mu_2 = \bar{x}_2 + (\mu_1 - \bar{x}_1)\sqrt{s_{22}/s_{11}}$ and $\sigma_2 = \sigma_1\sqrt{s_{22}/s_{11}}$. Since the one-dimensional normal problem is standard, we will not explicitly present results for this degenerate case.

1.2 Parameter Marginal Fiducial/Posterior Distributions

The bivariate normal distribution has been extensively studied from frequentist, fiducial and objective Bayesian perspectives. Table 1 summarizes the fiducial distributions for the five parameters of the bivariate normal distribution; these were found in Fisher (1930, 1956) and Pratt (1963). The table also lists objective prior distributions whose marginal posterior distribution for the indicated parameter equals the fiducial distribution. The results for the objective priors are easy consequences of expressions in Geisser and Cornfield (1963), except for that for ρ, which is discussed further below.

Fiducial/posterior distributions are represented in the table as *constructive random distributions*, i.e., by a description of how to simulate from them. Thus to simulate from the fiducial distribution of σ_1, given the data (actually, only s_{11} is needed), one draws independent χ^2_{n-1} random variables and simply computes the corresponding $\sqrt{s_{11}/\chi^2_{n-1}}$; this yields an independent sample from the fiducial/posterior distribution of σ_1. (The * in the notation in the table is to distinguish the constructive random distribution from later expressions

Table 1: Fiducial/posterior distributions and certain objective priors (the right-Haar prior, π_H, and the Jeffreys-rule prior, π_J) that yield these as posteriors. Here Z^* is a standard normal random variable, and χ^{2*}_{n-1} and χ^{2*}_{n-2} are chi-squared random variables with the indicated degrees of freedom, all random variables being independent.

parameter	objective prior	fiducial and posterior distribution
ρ	$\pi_H = \frac{1}{\sigma_1^2(1-\rho^2)}$	$\psi\left(-\frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}} + \frac{\sqrt{\chi^{2*}_{n-2}}}{\sqrt{\chi^{2*}_{n-1}}}\frac{r}{\sqrt{1-r^2}}\right)$, $\psi(x) = \frac{x}{\sqrt{1+x^2}}$
μ_1	$\pi_J = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2}$ and π_H	$\overline{x}_1 + \frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}}\sqrt{\frac{s_{11}}{n}}$
μ_2	π_J and π_H	$\overline{x}_2 + \frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}}\sqrt{\frac{s_{22}}{n}}$
σ_1	π_J and π_H	$\sqrt{\frac{s_{11}}{\chi^{2*}_{n-1}}}$
σ_2	π_J	$\sqrt{\frac{s_{22}}{\chi^{2*}_{n-1}}}$

involving randomness in the data.) We find such constructive random distributions to be much more useful in today's computing environment than are, say, analytic formulas for the corresponding densities. Also, these representations will be seen to be crucial in proving exact frequentist matching, as discussed below.

Table 1 also summarizes frequentist inference concerning these parameters, in the following sense: suppose the indicated fiducial/posterior distribution is used to create one-sided credible intervals $(\theta_L, \theta_{1-\alpha}(\mathbf{X}))$, where θ_L is the lower limit in the relevant parameter space $(-\infty, -1, \text{or}\, 0$ for the five parameters above), and $\theta_{1-\alpha}(\mathbf{X})$ is the posterior quantile of the parameter θ of interest, defined by

$$P(\theta < \theta_{1-\alpha}(\mathbf{X}) \mid \mathbf{X}) = \mathbf{1} - \alpha. \tag{3}$$

(Here θ is the random variable.) Of interest is the frequentist coverage of the corresponding confidence interval, i.e.,

$$C(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = P(\theta < \theta_{1-\alpha}(\mathbf{X}) \mid \boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\sigma_1}, \boldsymbol{\sigma_2}, \rho). \tag{4}$$

(Here $\mathbf{X}$ is the random variable.) The closer $C(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ is to the nominal $1 - \alpha$, the better the procedure (and corresponding objective prior) is judged to be.

The interesting feature of the fiducial/posterior distributions in Table 1 is that they all yield credible sets that have exact frequentist coverage of $1 - \alpha$. When this is the case, the fiducial/objective Bayesian posterior will be called *exact frequentist matching.* This is a very desirable situation (see Bayarri & Berger (2004) for general discussion and the many earlier references), in that one has an exact frequentist procedure that is also known to have good conditional performance (since it also is a Bayesian credible interval). As an example, to obtain an accurate approximation to the exact $100(1-\alpha)\%$ equal-tailed frequentist confidence set for ρ, one can employ the following simple algorithm:

- Draw independent $Z^* \sim N(0, 1)$, χ^{*2}_{n-1} and χ^{*2}_{n-2}.

- Set $\rho = \frac{Y}{\sqrt{1+Y^2}}$, where $Y = -\frac{Z^*}{\sqrt{\chi^{*2}_{n-1}}} + \frac{\sqrt{\chi^{*2}_{n-2}}}{\sqrt{\chi^{*2}_{n-1}}} \frac{r}{\sqrt{1-r^2}}$;
- Repeat the process $10,000$ times.
- Use the $\alpha/2$ upper and lower quantiles of these generated ρ to form the desired confidence limits.

We highlight the results about ρ in Table 1 because they appear to be new, in two senses:

1. We have not been able to find a reference in the literature to the fact that the fiducial/objective posterior distribution of ρ is exact frequentist matching (proved here in Theorem 5). Indeed, standard statistical software utilizes various approximations to arrive at frequentist confidence sets for ρ, missing the fact that a simple exact confidence set exists. (In contrast, it has long been known that the other fiducial/posterior distributions in Table 1 are exact matching.) It was, of course, known that exact frequentist confidence procedures can be constructed (cf. exercise 54, chapter 6 of Lehmann (1986)), but explicit expressions do not seem to be available.

2. Geisser and Cornfield (1963) studied the question of whether the fiducial distribution of ρ could be reproduced as an objective Bayesian posterior, and they concluded that this was most likely not possible. The strongest evidence for this arose from Brillinger (1962), which used results from Lindley (1961) and a difficult analytic argument to show that there does not exist a prior $\pi(\rho)$ such that the fiducial density of ρ equals $f(r \mid \rho)\pi(\rho)$, where $f(r \mid \rho)$ is the density of r given ρ. Since the fiducial distribution of ρ only depends on r, it was certainly reasonable to speculate that, if it were not possible to derive this distribution from the density of r and a prior, then it would not be possible to do so in general. The above result, of course, shows that this speculation was incorrect.

What Brillinger's result does show is that the fiducial/posterior distribution for ρ provides another example of the marginalization paradox (Dawid et al. (1973)). Any proper prior has the property that, if the marginal posterior distribution for a parameter θ depends only on a statistic T – whose distribution in turn depends only on θ – then that posterior can be derived from the distribution of T together with the marginal prior for θ. If this is a basic property of any proper Bayesian procedure, then it would seem paradoxical for the property to be violated, as happens for the fiducial/posterior distribution for ρ. This presumption is not without its own controversies, however. For instance, in the bivariate normal problem, there is probably no proper prior distribution that yields a marginal posterior distribution of ρ which depends only on r, so the relevance of an unattainable property of proper priors could be questioned.

In any case, this situation provides an interesting philosophical conundrum of a type that we have not previously seen: a complete fiducial/objective Bayesian/frequentist unification can be obtained for inference about the usual parameters of the bivariate normal distribution, but only if violation of the marginalization paradox is accepted. We will shortly introduce a prior distribution that avoids the marginalization paradox for ρ, but which is not exactly frequentist matching. We, alas, know of no way to adjudicate between the competing goals of exact frequentist matching and avoidance of the marginalization paradox, and so will simply present both as possible objective Bayesian approaches.

1.3 Other Quantities For Which Exact Matching is Possible

Often, it is functions of the parameters of the bivariate normal distribution that are of interest, and it is natural to ask which other important quantities have objective Bayesian posteriors that are exact frequentist matching. First it is clear that any monotonic function of one of the parameters also has an exact matching posterior obtained by simple transformation of the posterior in Table 1. Thus the exact matching posterior for σ_1^2 (written constructively) is simply s_{11}/χ^{2*}_{n-1}.

Interestingly, the approach that was developed to establish that the fiducial/posterior distribution of ρ is exact frequentist matching also turned out to be applicable to a number of other functions of parameters. Table 2 gives a summary of the exact matching posteriors, of which we are aware, for a variety of functions of parameters of the bivariate normal distribution. (Proofs are given in Section 3.1.) Some of these results are already known, of course.

Table 2 also lists the objective prior distributions that yield the indicated objective posterior. For completeness, we repeat the entries of Table 1, since the new table indicates all the priors that result in the indicated posterior. The notation π_{ab} in the table stands for the important class of prior densities (a subclass of the *generalized Wishart distributions* of Brown et al. (1994))

$$\pi_{ab}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \quad = \quad \frac{1}{\sigma_1^{3-a}\sigma_2^{2-b}(1-\rho^2)^{2-b/2}}. \tag{5}$$

Special cases of this class are $\pi_J = \pi_{10}$, $\pi_H = \pi_{12}$, the *independence Jeffreys* prior $\pi_{IJ} = \pi_{21} = \sigma_1^{-1}\sigma_2^{-1}(1-\rho^2)^{-3/2}$, and π_{RO} which has $a = b = 1$. The independence Jeffreys prior follows from using a constant prior for the means, and then the Jeffreys prior for the covariance matrix with means given. We will see that π_{IJ} is virtually never optimal – while π_J is very often optimal – in contradiction to the common perception that the independence Jeffreys prior is better than the Jeffreys prior. π_{RO} is useful for certain parameters.

There is one other interesting phenomenon here, namely that the posterior in the table for $\xi = \mu_1/\sigma_1$ is known to be subject to a marginalization paradox in the univariate normal case; this was established in Stone & Dawid (1972). Indeed, their same proof shows that the posterior for the bivariate case is also subject to a marginalization paradox (namely, it depends only on the statistic $t = \overline{x}_1/\sqrt{s_{11}}$, whose distribution in turn depends only on ξ, yet there is no prior distribution for ξ which will yield the posterior in the table from just the likelihood of t). Bernardo (1979) showed, in the univariate normal case, that the reference prior is not subject to a marginalization paradox, and the same would almost certainly be true for a one-at-a-time reference prior for ξ (were it to be found). Thus this appears to be another situation in which one must choose between exact matching and avoidance of a marginalization paradox.

Table 2: Parameters with exact matching priors and associated constructive posteriors: Here Z^* is a standard normal random variable, and χ^{2*}_{n-1} and χ^{2*}_{n-2} are chi-squared random variables with the indicated degrees of freedom, all random variables being independent.

parameter	prior	posterior
μ_1	π_{1b}, $\forall\ b$ (including π_J and π_H)	$\overline{x}_1 + \frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}}\sqrt{\frac{s_{11}}{n}}$
μ_2	$\pi_J = \pi_{10}$	$\overline{x}_2 + \frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}}\sqrt{\frac{s_{22}}{n}}$
$\boldsymbol{d}'(\mu_1,\mu_2)'$, $\boldsymbol{d} \in \mathbb{R}^2$	$\pi_J = \pi_{10}$ and π_{H^*} (see Table 4)	$\boldsymbol{d}'(\overline{x}_1,\overline{x}_2)' + \frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}}\sqrt{\frac{\boldsymbol{d'Sd}}{n}}$
σ_1	π_{1b}, $\forall\ b$ (including π_J and π_H)	$\sqrt{\frac{s_{11}}{\chi^{2*}_{n-1}}}$
$\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}$	$\pi_J = \pi_{10}$ and π_{H^*} (see Table 4)	$\sqrt{\frac{\boldsymbol{d'Sd}}{\chi^{2*}_{n-1}}}$
$\frac{\mu_1}{\sigma_1}$	π_{1b}, $\forall\ b$ (including π_J and π_H)	$\frac{Z^*}{\sqrt{n}} + \frac{\overline{x}_1\sqrt{\chi^{2*}_{n-1}}}{\sqrt{s_{11}}}$
ρ	$\pi_H = \pi_{12}$	$\psi\left(-\frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}} + \frac{\sqrt{\chi^{2*}_{n-2}}}{\sqrt{\chi^{2*}_{n-1}}}\frac{r}{\sqrt{1-r^2}}\right)$ $\psi(y) = y/\sqrt{1+y^2}$
$-\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}$	π_{a2}, $\forall\ a$ (including π_H)	$\frac{Z^*}{\sqrt{s_{11}}} - \frac{\sqrt{\chi^{2*}_{n-2}}}{\sqrt{s_{11}}}\frac{r}{\sqrt{1-r^2}}$
$\frac{\rho\sigma_2}{\sigma_1}$	π_{a2}, $\forall\ a$ (including π_H)	$\frac{r\sqrt{s_{22}}}{\sqrt{s_{11}}} - \frac{Z^*}{\sqrt{\chi^{2*}_{n-2}}}\frac{\sqrt{1-r^2}\sqrt{s_{22}}}{\sqrt{s_{11}}}$
$\sigma_2^2(1-\rho^2)$	π_{a2}, $\forall\ a$ (including π_H)	$\frac{s_{22}(1-r^2)}{\sqrt{\chi^{2*}_{n-2}}}$
$\lvert\boldsymbol{\Sigma}\rvert \equiv \sigma_1^2\sigma_2^2(1-\rho^2)$	$\pi_H = \pi_{12}$ and $\pi_{IJ} = \pi_{21}$	$\frac{\lvert\boldsymbol{S}\rvert}{\chi^{2*}_{n-1}\chi^{2*}_{n-2}}$

1.4 General Parameters and Priors

It is actually rare to have exact matching priors for parameters of interest. Also, one is often interested in very complex functions of parameters (e.g., predictive distributions) and/or joint distributions of parameters. For such problems it is important to have a general objective prior that seems to perform reasonably well for all quantities of interest. Furthermore, it is unappealing to many Bayesians to change the prior according to which parameter is declared to be of interest, and an objective prior that performs well overall is often sought.

1.4.1 Recommended priors

In addition to π_J and π_H, we will also recommend the following priors for various purposes:

1. $\pi_{R\rho} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)}$.

2. $\pi_{R\sigma} \propto \frac{\sqrt{1+\rho^2}}{\sigma_1\sigma_2(1-\rho^2)}$.

3. $\pi_{R\lambda} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)\sqrt{\left(\frac{\sigma_1}{\sigma_2}-\frac{\sigma_2}{\sigma_1}\right)^2+4\rho^2}}$.

The first prior was developed in Lindley (1965), using certain notions of transformation to constant information, and was studied extensively in Bayarri (1981), where it was shown to be a reference prior (see the next section). This will be our recommendation for use as a 'general purpose' objective prior. Also, it is a prior that avoids the marginalization paradox for ρ and so is the recommended prior for ρ if one is more concerned with avoiding the marginalization paradox than having exact frequentist matching. The prior $\pi_{R\sigma}$ will be suggested for use with inferences concerning σ_{12}, and $\pi_{R\lambda}$ for use with inferences concerning eigenvalues.

With these definitions, we can make our summary recommendations. Table 3 gives the four objective priors that are recommended for use, and indicates for which parameters (or functions thereof) they are recommended. These recommendations are based on three criteria:

1. The degree of frequentist matching. Having exact matching is viewed as best (according to this criterion). A minimal requirement is 'first order asymptotic matching,' discussed in Section 3.2. Among priors with this latter property, further discrimination is obtained through numerical study of the degree of matching in the most challenging case, namely when $n = 3$; this is discussed in Section 3.3.

2. Being a one-at-a-time reference prior (and hence avoiding the marginalization paradox) is desirable, because of its logical justification.

3. Ease of computation.

Table 3: Recommendations of objective priors for various parameters in the bivariate normal model: $\boxed{}$ indicates that, for this parameter, the posterior will not be exact frequentist matching. (For μ_2 and parameters with σ_1 replaced by σ_2, use the right-Haar prior with the variances interchanged.)

Prior	parameter
$\pi_{R\rho}$	$\boxed{\rho}$, $\boxed{\frac{\sigma_1}{\sigma_2}}$, $\boxed{\textbf{general use}}$
π_H	$\mu_1,\ \sigma_1,\ \frac{\mu_1}{\sigma_1},\ \lvert\boldsymbol{\Sigma}\rvert,\ \frac{\rho\sigma_2}{\sigma_1},\ \sigma_2^2(1-\rho^2),\ \eta_3 = \frac{-\rho}{\sigma_1\sqrt{1-\rho^2}},\ \rho$
$\tilde{\pi}_H$ (see Table 4)	$\boldsymbol{d}'(\mu_1,\mu_2)'$, $\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}$
$\pi_{R\lambda}$	$\boxed{ch_{max}(\boldsymbol{\Sigma})}$
$\pi_{R\sigma}$	$\boxed{\sigma_{12} = \rho\sigma_1\sigma_2}$

The inferences involving the non-boxed parameters in Table 3 are given in essentially closed form in Table 2 (and so are computationally simple), and are exact frequentist matching. Furthermore, with the exception of μ_1/σ_1 and η_3, it will be seen that the non-boxed parameters have the indicated priors as one-at-a-time reference priors, so all three criteria point to the indicated recommendation. (For μ_1/σ_1 and η_3 we provisionally go with the frequentist matching prior, but the marginalization paradox mentioned earlier suggests that this is not as clear a conclusion.)

Computation for the boxed parameters in Table 3 are still quite easy for the indicated priors, as will be seen in Section 2. As these are not exact matching, first order matching and numerically investigated matching are studied in Sections 3.2 and 3.3, respectively. The rationales for the recommended priors for the boxed parameters are also given there.

1.4.2 Reference Priors

This paper began with an effort to derive and catalogue the possible reference priors for the bivariate normal distribution. The reference prior theory (cf. Bernardo, 1979, and Berger and Bernardo, 1992*a*, 1992*b*) has arguably been the most successful technique for deriving objective priors.

Reference priors depend on (i) specification of a parameter of interest; (ii) specification of nuisance parameters; (iii) specification of a grouping of parameters; and (iv) ordering of the groupings. These are all conveyed by the shorthand notation used in Table 4. The following are examples:

- $\{(\mu_1, \mu_2), (\sigma_1, \sigma_2, \rho)\}$ indicates that (μ_1, μ_2) is the parameter of interest, with the remaining parameters being nuisance parameters; and there are two groupings with the indicated ordering. The corresponding reference prior is also the *independence Jeffreys* prior, π_{IJ}, mentioned earlier.

- $\{\rho, \sigma_1, \sigma_2, \mu_1, \mu_2\}$ indicates that ρ is the parameter of interest, with the rest being nuisance parameters; each parameter is it's own group; and the parameters have the indicated order of importance.

- $\{\lambda_1, \lambda_2, \vartheta, \mu_1, \mu_2\}$ introduces the eigenvalues $\lambda_1 > \lambda_2$ of $\boldsymbol{\Sigma}$ as being primarily of interest, with ϑ (the angle defining the orthogonal matrix that diagonalizes $\boldsymbol{\Sigma}$), μ_1, and μ_2 being the nuisance parameters

Based on experience with numerous examples, the reference priors that are typically judged to be best are one-at-a-time reference priors, in which each parameter is listed separately as its own group. Hence we will focus on these priors. It turns out to be the case that, for the one-at-a-time reference priors, the ordering of μ_1 and μ_2 among the variables is irrelevant (i.e., any placement of those two variables in the ordering will result in the same reference prior). Hence if μ_1 and μ_2 are omitted from a listing in Table 4, the resulting reference prior is to be viewed as any one-at-a-time reference prior with the indicated ordering of other variables, with the μ_i being inserted anywhere in the ordering.

We are interested in finding reference priors (preferably one-at-a-time) for the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$, $\theta_1 = \frac{\rho\sigma_2}{\sigma_1}$, $\theta_2 = \sigma_2^2(1-\rho^2)$, $\theta_3 = |\boldsymbol{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2)$, $\sigma_{12} = \rho\sigma_1\sigma_2$, $\theta_4 = \frac{\sigma_2^2}{\sigma_1^2}$, $\eta_3 = -\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}$, $\theta_5 = \frac{\mu_1}{\sigma_1}$, $\theta_6 = \frac{\mu_2}{\sigma_2}$, $\theta_7 = \boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}$ ($\boldsymbol{d}' = (d_1, d_2)$, not proportional to (0,1)), $\boldsymbol{d}'(\mu_1, \mu_2)'$, and $\lambda_1 = ch_{max}(\boldsymbol{\Sigma})$. Table 4 gives one-at-a-time reference priors for all these parameters (i.e., the parameter appears as the first entry in the parameter ordering – recall that the μ_i could go anywhere) except η_3, σ_{12}, and μ_i/σ_i; finding one-at-a-time reference priors for these parameters requires a much more technical analysis than we utilize here. (We do not explicitly list the reference priors for σ_2 in the table, since they can be found by simply switching with σ_1 in the various expressions.)

Table 4: Reference priors for the bivariate normal model (where $\theta_8 = \sqrt{1-\rho^2}\sigma_2/\sigma_1$, $\theta_9 = \sigma_1\sigma_2$, $\tilde{\mu}_1 = \boldsymbol{d}'(\mu_1,\mu_2)'$, $(\tilde{\sigma}_1)^2 = \theta_7$, $\tilde{\rho} = \boldsymbol{d}'\boldsymbol{\Sigma}(0,1)'/(\sigma_1\sqrt{\theta_7})$, $\tilde{\theta}_2 = \sigma_2^2[1-(\tilde{\rho})^2]$, and $\tilde{\theta}_1 = \tilde{\rho}\sigma_2/\tilde{\sigma}_1$); $\{\{\ \}\}$ indicates that any ordering of the parameters would yield the same reference prior.

prior $\pi(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$	for parameter ordering	has form (5) with
$\pi_J \propto \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2}$	$\{(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)\}$	$(a,b)=(1,0)$
$\pi_{IJ} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)^{3/2}}$	$\{(\mu_1,\mu_2),(\sigma_1,\sigma_2,\rho)\}$	$(a,b)=(2,1)$
$\pi_{R\rho} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)}$	$\{\rho,\sigma_1,\sigma_2\}$ $\{\theta_4,\theta_9,\rho\}$	
$\pi_{R\sigma} \propto \frac{\sqrt{1+\rho^2}}{\sigma_1\sigma_2(1-\rho^2)}$	$\{\sigma_1,\sigma_2,\rho\}$	
$\tilde{\pi}_{R\sigma} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)\sqrt{2-\rho^2}}$	$\{\sigma_1,\rho,\sigma_2\}$ $\{\sigma_1,\eta_3,\theta_2\}$	
$\pi_{RO} \propto \frac{1}{\sigma_1^2\sigma_2(1-\rho^2)^{3/2}}$	$\{\sigma_1,\theta_2,\eta_3\}$	$(a,b)=(1,1)$
$\pi_{R\lambda} \propto \frac{\left[\left(\frac{\sigma_1}{\sigma_2}-\frac{\sigma_2}{\sigma_1}\right)^2+4\rho^2\right]^{-1/2}}{\sigma_1\sigma_2(1-\rho^2)}$	$\{\lambda_1,\lambda_2,\vartheta\}$	
$\pi_H \propto \frac{1}{\sigma_1^2(1-\rho^2)}$	$\{\{\sigma_1,\theta_2,\theta_1\}\}$ $\{\{\lvert\boldsymbol{\Sigma}\rvert,\theta_8,\theta_1\}\}$	$(a,b)=(1,2)$
$\tilde{\pi}_H \propto \frac{1}{(\tilde{\sigma}_1)^2[1-(\tilde{\rho})^2]}d\tilde{\mu}_1 d\mu_2 d\tilde{\sigma}_1 d\sigma_2 d\tilde{\rho}$	$\{\{\boldsymbol{d}'(\mu_1,\mu_2)',\mu_2,\theta_7,\tilde{\theta}_2,\tilde{\theta}_1\}\}$	

For most of the parameters, the proofs that these are the indicated reference priors are given in Appendix A. We mention, here, the proof for $\boldsymbol{d}'(\mu_1,\mu_2)'$ and $\theta_7 = \boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}$, because the formulation will be needed later. Indeed, consider the transformation

$$\boldsymbol{Y} = \begin{pmatrix} d_1 & d_2 \\ 0 & 1 \end{pmatrix} \boldsymbol{X} \sim N_2\left(\boldsymbol{d}'(\mu_1,\mu_2)', \begin{pmatrix} \boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d} & \boldsymbol{d}'\boldsymbol{\Sigma}(0,1)' \\ \boldsymbol{d}'\boldsymbol{\Sigma}(0,1)' & \sigma_2^2 \end{pmatrix}\right). \tag{6}$$

For this new bivariate normal problem, we can simply apply the previous result, showing that the right-Haar prior is one-at-a-time matching for the new first coordinate mean $\tilde{\mu}_1 = \boldsymbol{d}'(\mu_1,\mu_2)'$ and first coordinate variance $(\tilde{\sigma}_1)^2 = \theta_7$. (Note that the transformed μ_2 and σ_2 are the same as the originals.)

Note that the non-Jeffreys reference priors are denoted by $\pi_{R\theta}$, with θ referring to the parameter of interest. Discussion of the various choices, further notational details, and derivations are given in the Appendix. Note that $\pi_{R\sigma}$ and $\tilde{\pi}_{R\sigma}$ are almost the same; they differ only by the bounded functions $g_1(\rho) = \sqrt{1+\rho^2}$ and $g_2(\rho) = \sqrt{2}/\sqrt{2-\rho^2}$, which are shown to be very similar in Figure 1.

1.4.3 Other studied priors

Finally, we should mention two other priors that have been suggested, the most common of which is the 'scale prior'

$$\pi_S \propto \frac{1}{\sigma_1\sigma_2}. \tag{7}$$

The motivation that is often given for this prior is that it is 'standard' to use σ_i^{-1} as the prior for a standard deviation, while $-1 < \rho < 1$ is on a bounded set and so one can use a constant prior in ρ. A related prior is

$$\pi_{MS} \propto \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

whose power of $(1-\rho^2)$ is between that of π_S and $\pi_{R\rho}$.

2 Computation

In this paper, a constant prior is always used for (μ_1, μ_2), so that

$$[\mu_1, \mu_2 \,|\, \boldsymbol{\Sigma}, \boldsymbol{X}] \sim N_2((\overline{x}_1, \overline{x}_2), n^{-1}\boldsymbol{\Sigma}). \tag{8}$$

Generation from this conditional posterior distribution is standard, so the challenge of simulation from the posterior distribution requires only sampling from the marginal posterior of $(\sigma_1, \sigma_2, \rho)$ given $\boldsymbol{X}$.

The marginal likelihood of $(\sigma_1, \sigma_2, \rho)$ satisfies

$$L_1(\sigma_1, \sigma_2, \rho) \propto \frac{1}{\{\sigma_1^2\sigma_2^2(1-\rho^2)\}^{(n-1)/2}} etr\Big(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\Big). \tag{9}$$

It is immediate that, under the priors π_J and π_{IJ}, the marginal posteriors of $\boldsymbol{\Sigma}$ are Inverse Wishart $(\boldsymbol{S}^{-1}, n)$ and Inverse Wishart $(\boldsymbol{S}^{-1}, n-1)$, respectively. The following sections deal with the more challenging priors.

2.1 Marginal Posteriors of $(\sigma_1, \sigma_2, \rho)$ under the Priors $\pi_{R\rho}$, $\pi_{R\sigma}$, $\tilde{\pi}_{R\sigma}$, π_S, and π_{MS}

For these priors, an independent sample from their marginal posterior $\pi(\sigma_1, \sigma_2, \rho \,|\, \boldsymbol{X})$ can be obtained by the following acceptance-rejection algorithm (cf. Robert & Casella (2004) for discussion of accept-reject algorithms).

Simulation Step: Generate $(\sigma_1, \sigma_2, \rho)$ from the independence Jeffreys posterior $\pi_{IJ}(\sigma_1, \sigma_2, \rho \,|\, \boldsymbol{X})$ (the Inverse Wishart $(\boldsymbol{S}^{-1}, n-1)$ distribution) and, independently, sample $u \sim \text{Uniform}(0, 1)$.

Rejection Step: Suppose

$$M \equiv \sup_{(\sigma_1,\sigma_2,\rho)} \frac{\pi(\sigma_1,\sigma_2,\rho)}{\pi_{IJ}(\sigma_1,\sigma_2,\rho)} < \infty\,.$$

If $u \le \pi(\sigma_1,\sigma_2,\rho)/[M\,\pi_{IJ}(\sigma_1,\sigma_2,\rho)]$, report (σ_1,σ_2,ρ); otherwise, go back to the *Simulation Step.*

If $M = \infty$, one would instead need to use the generally less efficient Metropolis-Hasting algorithm. Luckily, for the primary priors of interest, M is finite and, indeed, close to 1, which leads to very efficient algorithms.

For each of the priors listed in Table 5, the key ratio, π/π_{IJ}, is listed in the table, along with the upperbound M, the *Rejection Step*, and the resulting acceptance probability for $\rho = 0.80, 0.95, 0.99$.

The rejection algorithm is quite efficient for sampling these posteriors. Indeed, for $\rho \approx 0$, the algorithms accept with probability near one and, even for large $|\rho|$, the acceptance probabilities are very reasonable for the priors $\pi_{R\rho}$, $\pi_{R\sigma}$, and $\tilde{\pi}_{R\sigma}$. For large $|\rho|$, the algorithm is less efficient for the posteriors under the priors π_S and π_{MS}, but even these acceptance rates may well be fine in practice, given the simplicity of the algorithm.

Table 5: Ratio π/π_{IJ}, upper bound M, rejection step and acceptance probability for $\rho = 0.80, 0.95, 0.99$, when $\pi = \pi_{R\rho}$, $\pi_{R\sigma}$, $\tilde{\pi}_{R\sigma}$, π_S, and π_{MS}.

Prior π	Ratio $\frac{\pi}{\pi_{IJ}}$	Bound M	*Rejection Step*	Acceptance Probability		
				$\rho = .80$	$\rho = .95$	$\rho = .99$
$\pi_{R\rho}$	$\sqrt{1-\rho^2}$	1	$u \le \sqrt{1-\rho^2}$	.6000	.3122	.1410
$\pi_{R\sigma}$	$\sqrt{1-\rho^4}$	1	$u \le \sqrt{1-\rho^4}$	.7684	.4307	.1985
$\tilde{\pi}_{R\sigma}$	$\sqrt{\frac{1-\rho^2}{2-\rho^2}}$	$\frac{1}{\sqrt{2}}$	$u \le \sqrt{\frac{2(1-\rho^2)}{2-\rho^2}}$	.7276	.4215	.1975
π_S	$(1-\rho^2)^{3/2}$	1	$u \le (1-\rho^2)^{3/2}$	.2160	.0304	.0028
π_{MS}	$1-\rho^2$	1	$u \le (1-\rho^2)$	.3600	.0975	.0199

2.2 Computation under π_{ab}

The most interesting prior of this form (besides the Jeffreys and independence Jeffreys priors) is the right-Haar prior π_H, although other priors such as π_{11} arise as reference priors and hence are potentially of interest. While Tables 1 and 2 gave an explicit form for the most important marginal posteriors arising from priors of this form, it is of considerable interest that essentially closed form generation from the full posterior of any prior of this form is possible (see, e.g., Brown et al. (1994)). This is briefly reviewed in this section, since the expressions for the resulting constructive posteriors are needed for later results on frequentist coverage.

It is most convenient to work with the transformed variables

$$\begin{cases} \eta_1 &= 1/\sigma_1, \\ \eta_2 &= 1/(\sigma_2\sqrt{1-\rho^2}), \\ \eta_3 &= -\rho/(\sigma_1\sqrt{1-\rho^2}). \end{cases} \tag{10}$$

This parameterization gives a type of Cholesky decomposition of the precision matrix $\mathbf{\Sigma}^{-1}$,

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} \eta_1 & \eta_3 \\ 0 & \eta_2 \end{pmatrix} \begin{pmatrix} \eta_1 & 0 \\ \eta_3 & \eta_2 \end{pmatrix}, \tag{11}$$

which accounts for the simplicity of ensuing computations.

Fact 1 *(a) The inverse transformations are*

$$\sigma_1 = \frac{1}{\eta_1}, \;\; \sigma_2 = \frac{\sqrt{\eta_1^2+\eta_3^2}}{\eta_1\eta_2}, \;\; \rho = -\frac{\eta_3}{\sqrt{\eta_1^2+\eta_3^2}}. \tag{12}$$

(b) The Hessian matrix is

$$\boldsymbol{H} = \frac{\partial(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)}{\partial(\mu_1,\mu_2,\eta_1,\eta_2,\eta_3)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{\eta_1^2} & 0 & 0 \\ 0 & 0 & -\frac{\eta_3^2}{\eta_1^2\eta_2\sqrt{\eta_1^2+\eta_3^2}} & -\frac{\sqrt{\eta_1^2+\eta_3^2}}{\eta_1\eta_2^2} & \frac{\eta_3}{\eta_1\eta_2\sqrt{\eta_1^2+\eta_3^2}} \\ 0 & 0 & \frac{\eta_1\eta_3}{\sqrt{\eta_1^2+\eta_3^2}} & 0 & -\frac{\eta_1^2}{(\eta_1^2+\eta_3^2)^{3/2}} \end{pmatrix}.$$

(c) The Jacobians between the two parameterizations are

$$J \equiv \left|\frac{\partial(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)}{\partial(\mu_1,\mu_2,\eta_1,\eta_2,\eta_3)}\right| = \frac{1}{\eta_1\eta_2^2(\eta_1^2+\eta_3^2)}, \tag{13}$$

$$\frac{1}{J} \equiv \left|\frac{\partial(\mu_1,\mu_2,\eta_1,\eta_2,\eta_3)}{\partial(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)}\right| = \frac{1}{\sigma_1^3\sigma_2^2(1-\rho^2)^2}. \tag{14}$$

(d) The prior π_{ab} of (5) for $(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$ transforms to the extended conjugate class of priors for $(\mu_1,\mu_2,\eta_1,\eta_2,\eta_3)$,

$$\pi_{ab}(\mu_1,\mu_2,\eta_1,\eta_2,\eta_3) = \frac{1}{\eta_1^a\eta_2^b}. \tag{15}$$

The following results are immediate.

Lemma 1 *(a) Under the prior π_{ab}, the conditional posterior of (μ_1, μ_2) given (η_1, η_2, η_3) remains the same as (8) and the marginal posterior density of (η_1, η_2, η_3) (w.r.t. $d\eta_1 d\eta_2 d\eta_3$) is*

$$[\eta_1, \eta_2, \eta_3 \mid \boldsymbol{X}] \propto \eta_1^{n-a-1}\eta_2^{n-b-1} \exp\left\{-\frac{1}{2}\left[\eta_1^2 s_{11} + \eta_2^2 s_{22}(1-r^2) + s_{11}\left(\eta_3 + \eta_2 r\sqrt{\frac{s_{22}}{s_{11}}}\right)^2\right]\right\}. \quad (16)$$

(b) The marginal posterior distribution of η_3 given $(\eta_1, \eta_2; \boldsymbol{X})$ is $N(-\eta_2 r\sqrt{\frac{s_{22}}{s_{11}}}, \frac{1}{s_{11}})$.
(c) The marginal posterior distributions of η_1 and η_2 are independent and

$$\begin{aligned}(\eta_1^2 \mid \boldsymbol{X}) &\sim \mathit{Gamma}\Big(\frac{1}{2}(n-a),\ \frac{1}{2}s_{11}\Big);\\ (\eta_2^2 \mid \boldsymbol{X}) &\sim \mathit{Gamma}\Big(\frac{1}{2}(n-b),\ \frac{1}{2}s_{22}(1-r^2)\Big).\end{aligned}$$

We next present the constructive posteriors of (η_1, η_2, η_3), and from these derive the constructive posteriors of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$; later we consider other functions of these parameters. Recall that these constructive posteriors are not only useful for simulation, but will be the key to proving exact frequentist matching results.

We will use a star to represent a random draw from the implied distribution; thus μ_1^* will represent a random draw from its posterior distribution, Z_1^*, Z_2^*, Z_3^* will be independent draws from the standard normal distribution, and χ_{n-a}^{2*} and χ_{n-b}^{2*} will be independent draws from chi-squared distributions with the indicated degrees of freedom.

Fact 2 *(a) The constructive posterior of (η_1, η_2, η_3) given $\boldsymbol{X}$ can be expressed as*

$$\eta_1^* = \sqrt{\frac{\chi_{n-a}^{2*}}{s_{11}}}, \quad (17)$$

$$\eta_2^* = \sqrt{\frac{\chi_{n-b}^{2*}}{s_{22}(1-r^2)}}, \quad (18)$$

$$\eta_3^* = \frac{Z_3^*}{\sqrt{s_{11}}} - \eta_2^* r\sqrt{\frac{s_{22}}{s_{11}}} = \frac{Z_3^*}{\sqrt{s_{11}}} - \frac{\sqrt{\chi_{n-b}^{2*}}}{\sqrt{s_{11}}}\frac{r}{\sqrt{1-r^2}}. \quad (19)$$

(b) The constructive posterior of $(\sigma_1, \sigma_2, \rho)$ given $\boldsymbol{X}$ can be expressed as

$$\sigma_1^* = \frac{1}{\eta_1^*} = \sqrt{\frac{s_{11}}{\chi_{n-a}^{2*}}}, \quad (20)$$

$$\sigma_2^* = \sqrt{\frac{\eta_1^{*2} + \eta_3^{*2}}{\eta_1^{*2}\eta_2^{*2}}} = \sqrt{s_{22}(1-r^2)}\sqrt{\frac{1}{\chi_{n-b}^{2*}} + \frac{1}{\chi_{n-a}^{2*}}\left(\frac{Z_3^*}{\sqrt{\chi_{n-b}^{2*}}} - \frac{r}{\sqrt{1-r^2}}\right)^2}, \quad (21)$$

$$\rho^* = -\frac{\eta_3^*}{\sqrt{\eta_1^{*2} + \eta_3^{*2}}} = \psi(Y^*), \quad (22)$$

where

$$\psi(x) = \frac{x}{\sqrt{1+x^2}}, \qquad Y^* = -\frac{Z_3^*}{\sqrt{\chi_{n-a}^{2*}}} + \frac{\sqrt{\chi_{n-b}^{2*}}}{\sqrt{\chi_{n-a}^{2*}}}\frac{r}{\sqrt{1-r^2}}.$$

Proof. Part (a) follows immediately from Lemma 1. For Part (b), (20) follows from (17). Clearly

$$\eta_1^{*2} + \eta_3^{*2} = \frac{1}{\sigma_1^{*2}(1-\rho^{*2})} = \frac{1}{s_{11}}\Big[\chi^{2*}_{n-a} + \Big(Z_3^* - \sqrt{\chi^{2*}_{n-b}}\frac{r}{\sqrt{1-r^2}}\Big)^2\Big],$$

so that (21) follows from (17) and (18). For Part (c), the constructive posterior of ρ can be written as

$$\rho^* = -\frac{\eta_3^*}{\sqrt{\eta_1^{*2}+\eta_3^{*2}}} = -\frac{\frac{Z_3^*}{\sqrt{s_{11}}} - \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{s_{11}}}\frac{r}{\sqrt{1-r^2}}}{\left\{\frac{\chi^{2*}_{n-a}}{s_{11}} + \left[\frac{Z_3^*}{\sqrt{s_{11}}} - \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{s_{11}}}\frac{r}{\sqrt{1-r^2}}\right]^2\right\}^{\frac{1}{2}}},$$

which, after some algebra, results in (22). □

Note that the constructive posterior of ρ depends only on r.

Fact 3 *The constructive posterior for μ_1 and μ_2 can be written*

$$\mu_1^* = \overline{x}_1 + \frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}}\sqrt{\frac{s_{11}}{n}} = \overline{x}_1 + t^*_{n-a}\left[\frac{s_{11}}{n(n-a)}\right]^{1/2}, \tag{23}$$

$$\mu_2^* = \overline{x}_2 + \frac{\rho^*\sigma_2^*}{\sigma_1^*}(\mu_1^* - \overline{x}_1) + Z_2^*\sqrt{\frac{\sigma_2^{*2}(1-\rho^{*2})}{n}}. \tag{24}$$

$$= \overline{x}_2 + \frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}}\frac{r\sqrt{s_{22}}}{\sqrt{n}} + \left(\frac{Z_2^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}}\right)\sqrt{\frac{s_{22}(1-r^2)}{n}} \tag{25}$$

$$= \overline{x}_2 + \frac{r\sqrt{s_{22}}\,t^*_{n-a}}{\sqrt{n(n-a)}} + \frac{t^*_{n-b}}{\sqrt{n(n-b)}}\left[\left(1+\frac{t^{*2}_{n-a}}{n-a}\right)s_{22}(1-r^2)\right]^{1/2}, \tag{26}$$

where t^*_{n-a} and t^*_{n-b} are standard t random variables with the indicated degrees of freedom.

Proof. Since $(\mu_1 \mid \mathbf{\Sigma}, \boldsymbol{X}) \sim N(\overline{x}_1, \sigma_1^2/n)$, we have $\mu_1^* = \overline{x}_1 + Z_1^*\sigma_1^*/\sqrt{n}$. Substituting (20) into this expression yields (23). Because

$$(\mu_2 \mid \mu_1, \mathbf{\Sigma}, \boldsymbol{X}) \sim N\Big(\overline{x}_2 + \frac{\rho\sigma_2}{\sigma_1}(\mu_1 - \overline{x}_1), \frac{1}{n}\sigma_2^2(1-\rho^2)\Big),$$

(24) holds. Substituting (23) into (24) yields (25). Expression (26) follows by noting that

$$\frac{Z_2^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}t = \frac{1}{\sqrt{\chi^{2*}_{n-b}}}(Z_2^* - Z_3^* t) = \frac{Z}{\sqrt{\chi^{2*}_{n-b}}}\sqrt{1+t^2}$$

for any fixed t, where Z is also standard normal. The result is immediate. □

2.3 Computation under $\pi_{R\lambda}$

Here is a Metropolis-Hastings algorithm (from Berger et al. (2005)) to generate a (dependent) sample of size m from the marginal posterior $(\sigma_1, \sigma_2, \rho \mid \boldsymbol{X})$ based on the prior $\pi_{R\lambda}$, namely, $(\sigma_{1k}, \sigma_{2k}, \rho_k \mid \boldsymbol{X})$, $k = 1, \cdots, m$. At Stage k, sample

$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \tilde{\sigma}_1^2 & \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 \\ \tilde{\rho}\tilde{\sigma}_1\tilde{\sigma}_2 & \tilde{\sigma}_2^2 \end{pmatrix} \sim \text{Inverse Wishart}(\boldsymbol{S}^{-1}, n-1).$$

Then set

$$(\sigma_{1k}, \sigma_{2k}, \rho_k \mid \boldsymbol{X}) = \begin{cases} (\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\rho}), & \text{with probability } \alpha_k, \\ (\sigma_{1,k-1}, \sigma_{2,k-1}, \rho_{k-1}), & \text{otherwise,} \end{cases}$$

where

$$\alpha_k = \min\left\{1, \frac{\sqrt{\left(\frac{\sigma_{1,k-1}}{\sigma_{2,k-1}} - \frac{\sigma_{2,k-1}}{\sigma_{1,k-1}}\right)^2 + 4\rho_{k-1}^2}}{\sqrt{\left(\frac{\tilde{\sigma}_1}{\tilde{\sigma}_2} - \frac{\tilde{\sigma}_2}{\tilde{\sigma}_1}\right)^2 + 4\tilde{\rho}}} \frac{\sqrt{1-\tilde{\rho}^2}}{\sqrt{1-\rho_{k-1}^2}}\right\}.$$

3 Comparisons of Priors via Frequentist Matching

3.1 Frequentist Coverage Probabilities and Exact Matching

In this subsection we compare the frequentist properties of posterior credible intervals for various quantities under the prior π_{ab}, given in (5) and (15). As is customary in such comparisons, we study one-sided intervals $(\theta_L, q_{1-\alpha}(\mathbf{x}))$ of a parameter θ, where θ_L is the lower bound on the parameter θ (e.g., 0 or $-\infty$) and $q_{1-\alpha}(\mathbf{x})$ is the posterior quantile of θ, defined by

$$P(\theta < q_{1-\alpha}(\boldsymbol{x}) \mid \boldsymbol{x}) = 1 - \alpha.$$

Of interest is the frequentist coverage of the corresponding confidence interval, i.e.,

$$P(\theta < q_{1-\alpha}(\mathbf{X}) \mid \boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\sigma_1}, \boldsymbol{\sigma_2}, \rho).$$

The closer this coverage is to the nominal $1-\alpha$, the better the procedure (and corresponding objective prior) is judged to be. Here is a simple illustration.

Theorem 1 *(a) The posterior $(1-\alpha)$ quantiles of μ_1 and σ_1^2 are, respectively,*

$$(\mu_1^*)_{1-\alpha} = \overline{x}_1 + (t_{n-a}^*)_{1-\alpha}\left[\frac{s_{11}}{n(n-a)}\right]^{1/2}, \quad (\sigma_1^{*2})_{1-\alpha} = \left(\frac{s_{11}}{\chi_{n-a}^{2*}}\right)_{1-\alpha} = \frac{s_{11}}{(\chi_{n-a}^{2*})_\alpha},$$

where $(t_{n-a}^*)_{1-\alpha}$ and $(\chi_{n-1}^{2*})_{1-\alpha}$ are the $1-\alpha$-quantiles of the indicated t and χ^2 distributions. (b) The frequentist coverage probabilities of the one-sided credible intervals $(-\infty, (\mu_1^*)_{1-\alpha})$ and $(0, (\sigma_1^{*2})_{1-\alpha})$ for μ_1 and σ_1^2, respectively, are

$$P\left(t_{n-1} > \left(\frac{n-1}{n-a}\right)^{1/2}(t_{n-a}^*)_\alpha\right), \qquad P(\chi_{n-1}^2 > (\chi_{n-a}^{2*})_\alpha),$$

which do not depend on the parameters and equal $1-\alpha$ *(i.e. are frequentist matching) if and only if* $a=1$.

Proof. Part (a) follows from (20) and (23). For part (b) in regards to σ_1^2, the frequentist coverage of this interval (now with s_{11} considered random with the other quantities fixed) is clearly

$$C(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) = P\left(\sigma_1^2 < \frac{s_{11}}{(\chi^{2*}_{n-a})_\alpha} \mid \mu_1,\mu_2,\sigma_1,\sigma_2,\rho\right) = P\left(\frac{s_{11}}{\sigma_1^2} > (\chi^{2*}_{n-a})_\alpha \mid \mu_1,\mu_2,\sigma_1,\sigma_2,\rho\right), \quad (27)$$

which equals $1-\alpha$ only when $a=1$, since the frequentist distribution of s_{11}/σ_1^2 is χ^2 with $n-1$ degrees of freedom. The proof of exact matching for the credible intervals for μ_1 is virtually the same. □

Corollary 1 *If one considers the transformed bivariate normal problem in (6), the same results can be applied to the transformed mean (linear contrast)* $\boldsymbol{d}'(\mu_1,\mu_2)'$ *(*$\boldsymbol{d}'=(d_1,d_2)$ *not proportional to* $(0,1)$*) and variance* $\boldsymbol{d}'\boldsymbol{\Sigma}\boldsymbol{d}$.

It follows from Corollary 1 that one obtains exact matching for contrasts or transformed variances from either the transformed right-Haar prior π_{H^*} of Table 4 or the Jeffreys prior (since both have $a=1$). Interestingly, one can use the Jeffreys prior in the original parameterization, since the Jeffreys prior is invariant to transformations. (This result for contrasts and transformed variances was first found for the Jeffreys prior by Geisser & Cornfield (1963).)

3.1.1 Preliminaries

To prove frequentist matching for more complicated θ, the following technical lemmas will be repeatedly utilized. The first lemma is from (3d.2.8) in Rao (1973). The proof of the second lemma is straightforward and is omitted.

Lemma 2 *For* $n \geq 3$ *and given* σ_1,σ_2,ρ*, the following three random variables are independent and have the indicated distributions:*

$$T_2 = \left[\frac{s_{11}}{\sigma_2^2(1-\rho^2)}\right]^{\frac{1}{2}}\left[\frac{r\sqrt{s_{22}}}{\sqrt{s_{11}}} - \frac{\rho\sigma_2}{\sigma_1}\right] \equiv Z_3 \;\; \textit{(standard normal)}, \quad (28)$$

$$T_3 = \frac{s_{22}(1-r^2)}{\sigma_2^2(1-\rho^2)} \equiv \chi^2_{n-2}, \quad (29)$$

$$T_5 = \frac{s_{11}}{\sigma_1^2} \equiv \chi^2_{n-1}. \quad (30)$$

Lemma 3 *Let* $Y_{1-\alpha}$ *denote the* $1-\alpha$ *quantile of any random variable* Y.
(a) If $g(\cdot)$ *is a monotonically increasing function,* $[g(Y)]_{1-\alpha} = g(Y_{1-\alpha})$ *for any* $\alpha \in (0,1)$.
(b) For any $a>0, b\in\mathbb{R}$, $(aY+b)_{1-\alpha} = aY_{1-\alpha}+b$.
(c) If W *is a positive random variable,* $(WY)_{1-\alpha} \geq 0$ *if and only if* $Y_{1-\alpha} \geq 0$.

In the remainder of the chapter we use, without further comment, the notation that * appended to a random variable denotes randomness arising from the constructive posterior, while a random variable without a * refers to randomness arising from the (frequentist) distribution of a statistic. And, again, Z_i denote standard normal random variables and t_m and χ^2_m are, respectively, t and chi-squared random variables with m degrees of freedom. Whenever several of these occur in an expression, they are all independent (except that random variables of the same type and with the same index refer to the same random variable). Finally, we reserve quantile notation for posterior quantiles, with respect to the * distributions. Thus the quantile $\left[(\sigma_1 Z_3^* - r\, Z_3)/\chi^2_{n-1} + \rho\sqrt{s_{11}}\chi^{2*}_{n-b}\right]_{1-\alpha}$ would be computed based on the joint distribution of (Z_3^*, χ^{2*}_{n-b}), while holding $(\sigma_1, \rho, r, s_{11}, Z_3, \chi^2_{n-1})$ fixed.

3.1.2 Credible intervals for $\mu_1 - \mu_2$ under π_{ab}

Although Corollary 1 showed which priors yield optimal confidence sets for contrasts, we will be studying the extent to which other priors fail to yield credible sets having good frequentist coverage. This is important for our later recommendation of a good overall objective prior for the bivariate normal problem. We will specifically study the performance of credible sets for $\mu_1 - \mu_2$ under various priors, and the following theorem provides the needed result about coverage.

Theorem 2 *(a) The posterior $(1-\alpha)$ quantile of $\mu_1 - \mu_2$ is given by*

$$q_{1-\alpha}(\boldsymbol{X}) = (\overline{X}_1 - \overline{X}_2) + \left\{ \frac{t^*_{n-a}\sqrt{s_{11}}}{\sqrt{n(n-a)}} \left(1 - \frac{r\sqrt{s_{22}}}{\sqrt{s_{11}}}\right) - \frac{t^*_{n-b}}{\sqrt{n(n-b)}} \left[\left(1 + \frac{t^{*2}_{n-a}}{n-a}\right) s_{22}(1-r^2)\right]^{\frac{1}{2}} \right\}_{1-\alpha}. \quad (31)$$

(b) The frequentist coverage probability depends only on

$$\tau = \frac{\sigma_1 - \rho\sigma_2}{(\sigma_1^2 + \sigma_2^2 - 2\rho\,\sigma_1\,\sigma_2)^{1/2}} \quad (32)$$

and is given by the expression

$$P(\mu_1 - \mu_2 < q_{1-\alpha}(\mathbf{X}) \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$$
$$= P\left(\frac{Z_1}{\chi^2_{n-1}} < \left[\frac{t^*_{n-a}}{\sqrt{n-a}} \left(\tau - \sqrt{1-\tau^2}\frac{Z_2}{\chi^2_{n-1}}\right) + \frac{t^*_{n-b}\sqrt{1-\tau^2}}{\sqrt{n-b}} \sqrt{\left(1 + \frac{t^{*2}_{n-a}}{n-a}\right)\frac{\chi^2_{n-2}}{\chi^2_{n-1}}} \right]_{1-\alpha} \right), \quad (33)$$

where the probability is computed according to the joint distribution of $(Z_1, Z_2, \chi^2_{n-1}, \chi^2_{n-2})$.

Proof. Part (a) follows from the constructive posteriors for the μ_i given in (23) and (26). Noting that $\sqrt{n}[\mu_1 - \mu_2 - (\overline{X}_1 - \overline{X}_2)]/(\sigma_1^2 + \sigma_2^2 - 2\rho\,\sigma_1\,\sigma_2)^{1/2}$ is standard normal, it follows that

$$P(\mu_1 - \mu_2 < q_{1-\alpha}(\boldsymbol{X}) \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$$
$$= P\left(Z_1 < \sqrt{\frac{1}{\sigma_1^2 + \sigma_2^2 - 2\rho\,\sigma_1\,\sigma_2}} \left\{ \frac{t^*_{n-a}\sqrt{s_{11}}}{\sqrt{n-a}} \left(1 - \frac{r\sqrt{s_{22}}}{\sqrt{s_{11}}}\right) - \frac{t^*_{n-b}}{\sqrt{n-b}} \left[\left(1 + \frac{t^{*2}_{n-a}}{n-a}\right) s_{22}(1-r^2)\right]^{\frac{1}{2}} \right\}_{1-\alpha} \right),$$

where, henceforth, we omit the conditioning on the parameters. The sampling distributions in (28)-(30) together with algebra directly yields the conclusion. □

It is quite surprising that the coverage probability of these confidence intervals depends only on the single parameter τ, but also quite welcome in that coverage probabilities can then be graphed simply as a function of τ over the interval $[0, 1]$. (Because t_{n-a} and $-t_{n-a}$ have the same distribution, it is clear that the coverage is the same for τ and $-\tau$, so that it suffices to graph the coverage over the positive unit interval.)

Corollary 2 *Let F_{n-b} denote the cumulative distribution function of t_{n-b} and f_{n-a} be the density function of t_{n-a}. If $|\tau| \neq 1$, the formula (33) is equivalent to*

$$P(\mu_1 - \mu_2 < q_{1-\alpha}(\boldsymbol{X}) \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \quad = \quad P(g(Z_1, Z_2, \chi^2_{n-1}, \chi^2_{n-2}) < 1 - \alpha), \tag{34}$$

where

$$g(Z_1, Z_2, \chi^2_{n-1}, \chi^2_{n-2}) = \int_{\infty}^{-\infty} F_{n-b}\left(\frac{Z_1 - (\tau\sqrt{\chi^2_{n-1}} - \sqrt{1-\tau^2}Z_2)\frac{t}{\sqrt{n-a}}}{\sqrt{(n-b)\left(1+\frac{t^2}{n-a}\right)(1-\tau^2)\chi^2_{n-2}}} \right) f_{n-a}(t)dt. \tag{35}$$

Proof. It is enough to see that, for any random variables Z_1 and W, if $W_{1-\alpha}$ is the $1-\alpha$ quantile of W, $P(Z_1 < W_{1-\alpha}) = P(P(W < Z_1 \mid Z_1) < 1 - \alpha)$. □

This expression is more convenient for Monte Carlo computation, since one can just generate draws $(Z_1, Z_2, \chi^2_{n-1}, \chi^2_{n-2})$, numerically compute $g(Z_1, Z_2, \chi^2_{n-1}, \chi^2_{n-2})$ for each draw, and record the fraction of the time g is less than $1 - \alpha$. (In contrast, one must determine posterior quantiles at each Monte Carlo step when using (33), a more difficult enterprise.)

3.1.3 Credible intervals for a class of functions of $(\sigma_1, \sigma_2, \rho)$, including σ_1, σ_2, ρ, $\rho\sigma_2/\sigma_1$, $\sigma_2^2(1-\rho^2)$, $|\boldsymbol{\Sigma}|$, $\rho\sigma_1\sigma_2$, σ_2^2/σ_1^2, and $-\rho/(\sigma_1\sqrt{1-\rho^2})$

We consider the one-sided credible intervals of σ_1, σ_2, and ρ and some functions of the form

$$\theta = \sigma_1^{d_1}\sigma_2^{d_2}g(\rho), \tag{36}$$

for $d_1, d_2 \in \mathbb{R}$ and some function $g(\cdot)$. We also consider a class of scale-invariant priors of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$,

$$\pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \quad \propto \quad \frac{h(\rho)}{\sigma_1^{c_1}\sigma_2^{c_2}}, \tag{37}$$

for some $c_1, c_2 \in \mathbb{R}$ and a positive function h. The proof of the following theorem is given in Appendix B.

Theorem 3 *Denote the $1-\alpha$ posterior quantile of θ by $\theta_{1-\alpha}(\boldsymbol{X})$ under the prior (37). For any fixed $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, the frequentist coverage of the credible interval $(\theta_L, \theta_{1-\alpha}(\boldsymbol{X}))$ depends only on ρ. Here θ_L is the lower boundary of the parameter space for θ.*

Remark 1 *Note that the priors π_{ab} (including $\pi_J, \pi_{IJ}, \pi_{RO}$, and π_H) and $\pi_{R\sigma}, \pi_{R\sigma*}, \pi_{R\rho}, \pi_S$ in Table 4 are all of the form (37). From Theorem 3, if we use any of these priors, the frequentist coverage probabilities of the credible intervals for any of the parameters such as $\sigma_1, \sigma_2, \rho, \sigma_1/\sigma_2, \rho\sigma_2/\sigma_1,\ \rho\sigma_1\sigma_2,\ |\mathbf{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2)$ will depend only on ρ. This will be useful in the numerical comparison.*

We next consider the posterior quantiles of σ_2^2 and their coverage probabilities. The proof is given in Appendix B.

Theorem 4 *(a) For any α, the posterior $1-\alpha$ quantile of σ_2^2 has the expression*

$$(\sigma_2^{*2})_{1-\alpha} \;=\; s_{22}(1-r^2)\left[\frac{1}{\chi^{2*}_{n-b}} + \frac{1}{\chi^{2*}_{n-a}}\left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{r}{\sqrt{1-r^2}}\right)^2\right]_{1-\alpha}. \tag{38}$$

(b) For any $\alpha \in (0,1)$, $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$, and $\rho \in (-1,1)$, the frequentist coverage probability of the credible interval $(0, (\sigma_2^{*2})_{1-\alpha})$ is

$$\begin{aligned} & P(\sigma_2^2 < (\sigma_2^{*2})_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \\ =\; & P\left(1 < \chi^2_{n-2}(1-\rho^2)\left[\frac{1}{\chi^{2*}_{n-b}} + \frac{1}{\chi^{2*}_{n-a}}\left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}\right)^2\right]_{1-\alpha} \mid \rho\right). \end{aligned} \tag{39}$$

(c) As $|\rho| \to 1$, the left hand side of (39) goes to

$$P\left(\frac{1}{\chi^2_{n-1}} < \left(\frac{1}{\chi^{*2}_{n-a}}\right)_{1-\alpha}\right), \tag{40}$$

which is $1-\alpha$ if and only if $a = 1$.

Remark 2 *If $\rho = 0$,*

$$P(\sigma_2^2 < (\sigma_2^{*2})_{1-\alpha} \mid \boldsymbol{\xi}, \rho = 0) \;=\; P\left(\frac{1}{\chi^2_{n-2}} < \left[\frac{1}{\chi^{2*}_{n-b}} + \frac{1}{\chi^{2*}_{n-a}}\left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{Z_3}{\sqrt{\chi^2_{n-2}}}\right)^2\right]_{1-\alpha}\right).$$

If $b = 2$ and $\rho = 0$,

$$P(\sigma_2^2 < (\sigma_2^{*2})_{1-\alpha} \mid \boldsymbol{\xi}, \rho = 0) \;>\; P\left(\frac{1}{\chi^2_{n-2}} < \left(\frac{1}{\chi^{2*}_{n-b}}\right)_{1-\alpha}\right) = 1-\alpha.$$

To study the frequentist coverage of the credible intervals of ρ, note that ψ, defined in (23), is invertible, and

$$\psi^{-1}(\rho) = \frac{\rho}{\sqrt{1-\rho^2}}, \quad |\rho| < 1. \tag{41}$$

The following theorem, whose proof is given in Appendix B, is the key result about exact frequentist matching for ρ that was discussed in the introduction.

Theorem 5 *Let $\rho^*_{1-\alpha} = \rho^*_{1-\alpha}(\boldsymbol{X})$ be the $1-\alpha$ posterior quantile of ρ.*
(a) For ψ defined in (23), the posterior $1-\alpha$ quantile of ρ is

$$\rho^*_{1-\alpha} = \psi(Y^*_{1-\alpha}). \tag{42}$$

(b) For any $\alpha \in (0,1)$, $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$, and $\rho \in (-1,1)$,

$$P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) = P\bigg(\bigg(-\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-a}}} + \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{\chi^{2*}_{n-a}}}\frac{r}{\sqrt{1-r^2}}\bigg)_{1-\alpha} > \psi^{-1}(\rho) \;\Big|\; \rho\bigg). \tag{43}$$

Also,

$$P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \;=\; P\bigg(\frac{\sqrt{1-\rho^2}Z_3 + \rho\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} > \bigg(\frac{\sqrt{1-\rho^2}Z_3^* + \rho\sqrt{\chi^{2*}_{n-a}}}{\sqrt{\chi^{2*}_{n-b}}}\bigg)_{\alpha} \;\Big|\; \rho\bigg). \tag{44}$$

(c) For any $\alpha \in (0,1)$, any $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$, and any $\rho \in (-1,1)$,

$$P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \;=\; 1-\alpha. \tag{45}$$

if and only if the right Haar prior is used, i.e., $(a,b) = (1,2)$.

Corollary 3 *(a) For any $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$,*

$$\begin{aligned} P(0 < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho = 0) \;&=\; P\bigg(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} < \Big(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\Big)_{1-\alpha}\bigg) \\ &=\; P\bigg(t_{n-2} < \frac{\sqrt{n-2}}{\sqrt{n-b}}\big(t^*_{n-b}\big)_{1-\alpha}\bigg)\begin{cases} = 1-\alpha, & \text{if } b = 2, \\ < 1-\alpha, & \text{if } b < 2. \end{cases} \end{aligned} \tag{46}$$

(b) As $\rho \to -1$,

$$P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \;\to\; P\bigg(\frac{n-1}{n-2}F_{n-1,n-2} < \frac{n-a}{n-b}\Big(F^*_{n-a,n-b}\Big)_{1-\alpha}\bigg). \tag{47}$$

(c) As $\rho \to 1$,

$$P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \;\to\; P\bigg(\frac{n-1}{n-2}F_{n-1,n-2} > \frac{n-a}{n-b}\Big(F^*_{n-a,n-b}\Big)_{\alpha}\bigg). \tag{48}$$

We next consider six functions of $(\sigma_1, \sigma_2, \rho)$, namely the regression coefficient, conditional variance, the generalized variance or the determinant of the covariance matrix $\boldsymbol{\Sigma}$, the ratio of two variances, covariance, and η_3, i.e.,

$$\begin{cases} \theta_1 = \frac{\rho\sigma_2}{\sigma_1}, \quad \theta_2 = \sigma_2^2(1-\rho^2), \quad \theta_3 = |\boldsymbol{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2), \\ \theta_4 = \frac{\sigma_2^2}{\sigma_1^2}, \quad \sigma_{12} = \rho\sigma_1\sigma_2, \quad \eta_3 = -\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}. \end{cases} \tag{49}$$

Note that all these functions are of the form (36). From Theorem 3, under any of the priors $\pi_J, \pi_{IJ}, \pi_{R\sigma}, \pi_{R\rho}, \pi_{RO}, \pi_H, \pi_S$, the frequentist coverage probabilities of credible intervals for any of the functions in (49) will depend only on ρ. We first give results for θ_1 in the following theorem, whose proof is given in Appendix B.

Theorem 6 *(a) The random posterior of $\theta_1 = \rho\sigma_2/\sigma_1$ has the expression*

$$\theta_1^* = \frac{r\sqrt{s_{22}}}{\sqrt{s_{11}}} - \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\frac{\sqrt{1-r^2}\sqrt{s_{22}}}{\sqrt{s_{11}}}.$$

(b) For any $\alpha \in (0,1)$, let $(\theta_1^*)_{1-\alpha}$ be the $(1-\alpha)$ posterior quantile of θ_1. For any $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$, and any $\rho \in (-1,1)$,

$$P(\theta_1 < (\theta_1^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \;=\; P\Big(t_{n-2} < \sqrt{\frac{n-2}{n-b}}\,(t^*_{n-b})_{1-\alpha}\Big), \tag{50}$$

which does not depend on ρ. Furthermore, (50) equals $1-\alpha$ if and only if $b=2$.

Theorem 7 *(a) The random posterior of $\theta_2 = \sigma_2^2(1-\rho^2) = 1/\eta_2^2$ has the expression*

$$\theta_2^* = \frac{s_{22}(1-r^2)}{\chi^{2*}_{n-b}}. \tag{51}$$

(b) For any $\alpha \in (0,1)$, let $(\theta_2^*)_{1-\alpha}$ be the $(1-\alpha)$ posterior quantile of θ_2. For any $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$, and any $\rho \in (-1,1)$,

$$P(\theta_2 < (\theta_2^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \;=\; P\bigg(\frac{1}{\chi^2_{n-2}} < \Big(\frac{1}{\chi^{2*}_{n-b}}\Big)_{1-\alpha}\bigg), \tag{52}$$

which does not depend on ρ. Furthermore, (52) equals $1-\alpha$ if and only if $b=2$.

Proof. Since $\theta_2 = 1/\eta_2^2$, part (a) follows from (18) directly. From (29),

$$P(\theta_2 < (\theta_2^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \;=\; P\Bigg(\sigma_2^2(1-\rho^2) < \bigg[\frac{\sigma_2^2(1-\rho^2)\chi^2_{n-2}}{\sqrt{\chi^{2*}_{n-b}}}\bigg]_{1-\alpha} \;\Big|\; \boldsymbol{\xi}, \rho\Bigg),$$

which implies part (b). □

The next theorem gives the results for the determinant of $\boldsymbol{\Sigma} = 1/(\eta_1^2\eta_2^2) = \sigma_1^2\sigma_2^2(1-\rho^2)$.

Theorem 8 *(a) The random posterior of $\theta_3 = |\boldsymbol{\Sigma}|$ is*

$$\theta_3^* = \frac{1}{\eta_1^{*2}\eta_2^{*2}} \;=\; \frac{|\boldsymbol{S}|}{\chi^{2*}_{n-a}\chi^{2*}_{n-b}}.$$

(b) For any $\alpha \in (0,1)$, the $(1-\alpha)$ posterior quantile of θ_3 is $(\theta_3^*)_{1-\alpha} = |\boldsymbol{S}|(\chi^{2*}_{n-a}\chi^{2*}_{n-b})_\alpha$.
(c) For any $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$, and any $\rho \in (-1,1)$, the frequentist coverage probability of the $1-\alpha$ credible interval of θ_3 is

$$P(|\boldsymbol{\Sigma}| < |\boldsymbol{S}|(\chi^{2*}_{n-a}\chi^{2*}_{n-b})_\alpha \mid \boldsymbol{\xi}, \rho) \;=\; P\big(\chi^2_{n-1}\chi^2_{n-2} > (\chi^{2*}_{n-a}\chi^{2*}_{n-b})_\alpha\big). \tag{53}$$

(d) The coverage probability in (c) does not depend on ρ, and equals $1-\alpha$ if $(a,b) = (1,2)$ or $(a,b) = (2,1)$.

Proof. Parts (a) and (b) follow from (17) and (18) directly. It follows from (29) and (30) that $|\boldsymbol{S}|/|\boldsymbol{\Sigma}| = \chi^2_{n-1}\chi^2_{n-2}$. Part (c) holds. Part (d) is obvious. $\square$

It is of interesting that both priors π_{IJ} and π_H are exact matching priors for $|\boldsymbol{\Sigma}|$. The next theorem, whose proof is in Appendix B, gives the results for the ratio of two variances.

Theorem 9 *(a) The random posterior of* $\theta_4 = \sigma_2^2/\sigma_1^2$ *is*

$$\theta_4^* = \frac{\sigma_2^{*2}}{\sigma_1^{*2}} = \frac{s_{22}(1-r^2)}{s_{11}}\left[\frac{\chi^{2*}_{n-a}}{\chi^{2*}_{n-b}} + \left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{r}{\sqrt{1-r^2}}\right)^2\right].$$

(b) For any $\alpha \in (0,1)$, *let* $(\theta_4^*)_{1-\alpha}$ *be the posterior* $1-\alpha$ *quantile of* θ_4. *For any* $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$, *and* $\rho \in (-1,1)$, *the frequentist coverage probability of the credible interval* $(0, (\theta_4^*)_{1-\alpha})$ *is*

$$P(\theta_4 < (\theta_4^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) = P(G > 1 \mid \rho), \tag{54}$$

where

$$G = \frac{(1-\rho^2)\chi^2_{n-2}}{\chi^2_{n-1}}\left[\frac{\chi^{2*}_{n-a}}{\chi^{2*}_{n-b}} + \left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}\right)^2\right]_{1-\alpha}. \tag{55}$$

(c) As $|\rho| \to 1$, *the left hand side of (54) goes to*

$$P(\theta_4 < (\theta_4^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \to P((G_2)_{1-\alpha} > 0), \tag{56}$$

where

$$G_2 = \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{Z_3}{\sqrt{\chi^2_{n-2}}}. \tag{57}$$

Furthermore, the right hand side of (56) is $1-\alpha$ *if and only if* $b = 2$.

The following theorems give the results for σ_{12}, the covariance of x_1 and x_2, and the parameter η_3. The proofs can be found in Appendix B.

Theorem 10 *(a) The random posterior of* $\sigma_{12} = \rho\sigma_1\sigma_2$ *has the expression*

$$\sigma_{12}^* = -\left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{r}{\sqrt{1-r^2}}\right)\frac{\sqrt{s_{11}s_{22}(1-r^2)}}{\chi^{2*}_{n-a}}.$$

(b) For any $\alpha \in (0,1)$, *let* $(\sigma_{12}^*)_{1-\alpha}$ *be the* $(1-\alpha)$ *posterior quantile of* σ_{12}. *For any* $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$, *and any* $\rho \in (-1,1)$,

$$\begin{aligned} &P(\sigma_{12} < (\sigma_{12}^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \\ &= P\left(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} - \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} < \left(\frac{Z_1^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{\rho}{\sqrt{1-\rho^2}}\frac{\chi^{2*}_{n-a}}{\sqrt{\chi^2_{n-1}}\sqrt{\chi^2_{n-2}}}\right)_{1-\alpha} \,\Big|\, \rho\right). \end{aligned} \tag{58}$$

Theorem 11 *(a) For any $\alpha \in (0,1)$, let $(\eta_3^*)_{1-\alpha}$ be the $(1-\alpha)$ posterior quantile of η_3. For any $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$, and any $\rho \in (-1,1)$,*

$$P(\eta_3 < (\eta_3^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \;=\; P\Bigg(\frac{Z_3 + \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} < \Bigg(\frac{Z_3^* + \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^{2*}_{n-b}}}\Bigg)_{1-\alpha} \Bigg| \;\rho\Bigg). \quad (59)$$

(b) (59) equals $1-\alpha$ for any $-1 < \rho < 1$ if and only if $b = 2$.

3.1.4 Coverage Probabilities for μ_1/σ_1, μ_2/σ_2, and $\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$ under π_{ab}

There are many other interesting functions of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ not of the form (36). For example,

$$\theta_5 = \frac{\mu_1}{\sigma_1}, \quad \theta_6 = \frac{\mu_2}{\sigma_2}, \quad \theta_7 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2. \quad (60)$$

In the next theorem, whose proof is given in Appendix B, consider $\theta_5 = \mu_1/\sigma_1$.

Theorem 12 *(a) The random posterior of $\theta_5 = \mu_1/\sigma_1$ has the expression*

$$\theta_5^* = \frac{Z_1^*}{\sqrt{n}} + \frac{\overline{x}_1\sqrt{\chi^{2*}_{n-a}}}{\sqrt{s_{11}}}.$$

(b) For any $\alpha \in (0,1)$, denote the $(1-\alpha)$ posterior quantile of θ_5 by $(\theta_5^*)_{1-\alpha}$. The frequentist coverage of the credible interval $(-\infty, (\theta_5^*)_{1-\alpha})$ depends on θ_5 only, and equals

$$P(\theta_5 < (\theta_5^*)_{1-\alpha} \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = P\Bigg(\frac{Z_1 - \theta_5\sqrt{n}}{\sqrt{\chi^2_{n-1}}} < \Bigg(\frac{Z_1^* - \theta_5\sqrt{n}}{\sqrt{\chi^{2*}_{n-a}}}\Bigg)_{1-\alpha} \mid \theta_5\Bigg). \quad (61)$$

(c) The coverage probability equals $1-\alpha$ if and only if $a = 1$.

The next theorem, whose proof is given in Appendix B, considers $\theta_6 = \mu_2/\sigma_2$.

Theorem 13 *(a) The random posterior of $\theta_6 = \mu_2/\sigma_2$ can be expressed as*

$$\theta_6^* = \frac{\mu_2^*}{\sigma_2^*} \;=\; \frac{Z_2^*}{\sqrt{n}} + \frac{\overline{x}_2}{\sqrt{s_{22}(1-r^2)}}\left[\frac{1}{\chi^{2*}_{n-b}} + \frac{1}{\chi^{2*}_{n-a}}\left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{r}{\sqrt{1-r^2}}\right)^2\right]^{-1/2}. \quad (62)$$

(b) For any $\alpha \in (0,1)$, denote the $(1-\alpha)$ posterior quantile of θ_6 by $(\theta_6^*)_{1-\alpha}$. The frequentist coverage probability of the credible interval $(-\infty, (\theta_6^*)_{1-\alpha})$ depends on θ_6 and ρ and is given by

$$P(\theta_6 < (\theta_6^*)_{1-\alpha} \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = P\left(\theta_6 < (\theta_6^*)_{1-\alpha} \mid \theta_6, \rho\right), \quad (63)$$

where

$$\theta_6^* = \frac{Z_2^*}{\sqrt{n}} + \left(\theta_6 + \frac{Z_2}{\sqrt{n}}\right)\sqrt{\frac{\chi_{n-a}^{2*}}{\chi_{n-2}^2}}\left[\frac{\chi_{n-a}^{2*}(1-\rho^2)}{\chi_{n-b}^{2*}} + \left(G_2\sqrt{1-\rho^2} + \rho\sqrt{\frac{\chi_{n-1}^2}{\chi_{n-2}^2}}\right)^2\right]^{-\frac{1}{2}} \tag{64}$$

and G_2 *is given by (57).*
(c) as $|\rho| \to 1$,

$$\lim_{|\rho|\to 1} P(\theta_6 < (\theta_6^*)_{1-\alpha} \mid \theta_6, \rho) = P\left(\frac{Z_2 - \theta_6\sqrt{n}}{\sqrt{\chi_{n-1}^2}} < \left(\frac{Z_2^* - \theta_6\sqrt{n}}{\sqrt{\chi_{n-a}^{2*}}}\right)_{1-\alpha} \mid \theta_6\right), \tag{65}$$

which equals $1-\alpha$ *for any* θ_6 *if and only if* $a = 1$.

The next theorem, whose proof is given in Appendix B, considers $\theta_7 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$.

Theorem 14 *(a) The random posterior of* θ_7 *has the expression,*

$$\theta_7^* = \frac{s_{22}(1-r^2)}{\chi_{n-b}^{2*}} + \frac{s_{11}}{\chi_{n-a}^{2*}}\left(1 + \frac{Z_3^*}{\sqrt{\chi_{n-a}^{2*}}}\frac{\sqrt{s_{22}(1-r^2)}}{\sqrt{s_{11}}} - \frac{r\sqrt{s_{22}}}{\sqrt{s_{11}}}\right)^2. \tag{66}$$

(b) For any $\alpha \in (0,1)$, *denote the* $(1-\alpha)$ *posterior quantile of* θ_7 *by* $(\theta_7^*)_{1-\alpha}$. *The frequentist coverage of the* $(1-\alpha)$ *credible interval* $(0, (\theta_7^*)_{1-\alpha})$ *is given by*

$$P(\theta_7 < (\theta_7^*)_{1-\alpha} \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = P\left(1 < (G_4)_{1-\alpha} \mid \tau\right),$$

where τ *is given by (32) and*

$$G_4 = (1-\tau^2)\frac{\chi_{n-2}^2}{\chi_{n-b}^{2*}} + \frac{\chi_{n-1}^2}{\chi_{n-a}^{2*}}\left[\tau - \sqrt{1-\tau^2}\left(\frac{Z_3}{\sqrt{\chi_{n-1}^2}} - \frac{Z_3^*}{\sqrt{\chi_{n-b}^{2*}}}\frac{\sqrt{\chi_{n-2}^2}}{\sqrt{\chi_{n-1}^2}}\right)\right]^2. \tag{67}$$

(c) As $|\tau| \to 1$,

$$\lim_{|\theta|\to 1} P(\theta_7 < (\theta_7^*)_{1-\alpha} \mid \theta) = P\left(\frac{1}{\chi_{n-1}^2} < \left(\frac{1}{\chi_{n-a}^{2*}}\right)_{1-\alpha}\right), \tag{68}$$

which equals $1-\alpha$ *if and only if* $a = 1$.

3.2 First Order Asymptotic Matching

In this subsection, we derive the equations for finding the first order matching priors for each of the parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

Let $\boldsymbol{I}$ be the Fisher information of $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_p) = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ and $\omega = \omega(\boldsymbol{\zeta})$ be a function of $\boldsymbol{\zeta}$. Define the gradient vector by $\nabla\omega = (\frac{\partial}{\partial\zeta_1}\omega, \frac{\partial}{\partial\zeta_2}\omega, \cdots, \frac{\partial}{\partial\zeta_p}\omega)^t$. Define

$$\boldsymbol{\delta} \equiv (\delta_1, \cdots, \delta_p)^t = \frac{\boldsymbol{I}^{-1}\nabla\omega}{\sqrt{(\nabla\omega)^T\boldsymbol{I}^{-1}\nabla\omega}}. \tag{69}$$

Following Datta & Ghosh (1995), if the posterior probability of a one-sided credibility interval for ω and its frequentist coverage probability agree up to $o(n^{-1/2})$, the prior π should satisfy the equation

$$\sum_{i=1}^{p} \frac{\partial}{\partial \zeta_i} (\delta_i \; \pi) \;=\; 0. \tag{70}$$

Any positive solution π of such equation is called a first order matching prior for ω.

3.2.1 The Fisher Information

The Fisher information matrix of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ is given by

$$\boldsymbol{I} \;=\; \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} & 0 & 0 & 0 \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{2-\rho^2}{\sigma_1^2} & \frac{-\rho^2}{\sigma_1\sigma_2} & \frac{-\rho}{\sigma_1} \\ 0 & 0 & \frac{-\rho^2}{\sigma_1\sigma_2} & \frac{2-\rho^2}{\sigma_2^2} & \frac{-\rho}{\sigma_2} \\ 0 & 0 & \frac{-\rho}{\sigma_1} & \frac{-\rho}{\sigma_2} & \frac{1+\rho^2}{1-\rho^2} \end{pmatrix}, \tag{71}$$

The inverse of $\boldsymbol{I}$ is

$$\boldsymbol{V} = \boldsymbol{I}^{-1} \;=\; \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & 0 & 0 & 0 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2}\sigma_1^2 & \frac{1}{2}\rho^2\sigma_1\sigma_2 & \frac{1}{2}\rho(1-\rho^2)\sigma_1 \\ 0 & 0 & \frac{1}{2}\rho^2\sigma_1\sigma_2 & \frac{1}{2}\sigma_2^2 & \frac{1}{2}\rho(1-\rho^2)\sigma_2 \\ 0 & 0 & \frac{1}{2}\rho(1-\rho^2)\sigma_1 & \frac{1}{2}\rho(1-\rho^2)\sigma_2 & (1-\rho^2)^2 \end{pmatrix}. \tag{72}$$

The determinant of $\boldsymbol{I}$ is $1/\{\sigma_1^4\sigma_2^4(1-\rho^2)^4\}$, so the Jeffreys prior is

$$\pi_J \propto \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2}. \tag{73}$$

The independent Jeffreys prior, treating (μ_1, μ_2) and $(\sigma_1, \sigma_2, \rho)$ as being independent, is then

$$\pi_{IJ} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)^{3/2}}. \tag{74}$$

3.2.2 Differential Equations of the First Order Matching Priors

Table 6 gives the quantities $(\delta_1, \cdots, \delta_5)$ in the partial differential equations (69) for each of the indicated parameters. The table also gives the general form of the solutions to these partial differential equations, when they can be solved in closed form. Three illustrations follow.

Illustration 1. (a) To see the differential equation of the first order matching prior π for μ_1, note first that $\nabla\mu_1 = (1, 0, 0, 0, 0)'$. Thus $\boldsymbol{V}\nabla\mu_1 = (\sigma_1^2, \rho\sigma_1\sigma_2, 0, 0, 0)'$ and

$$\boldsymbol{\delta} \;\equiv\; (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)' \;=\; \frac{\boldsymbol{V}\nabla\mu_1}{\sqrt{\nabla\mu_1{}'\boldsymbol{V}\nabla\mu_1}} \;=\; (\sigma_1, \rho\sigma_2, 0, 0, 0)',$$

Table 6: Quantities $(\delta_1, \cdots, \delta_5)$ in partial differential equation (69) for $\omega = \mu_1, \mu_2, \sigma_1, \sigma_2, \rho$, $\boldsymbol{d}'\boldsymbol{\mu} = d_1\mu_1 + d_2\mu_2$, $\theta_1 = \frac{\rho\sigma_2}{\sigma_1}$, $\theta_2 = \sigma_2^2(1-\rho^2)$, $\theta_3 = |\boldsymbol{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2)$, $\theta_4 = \frac{\sigma_2^2}{\sigma_1^2}$, $\sigma_{12} = \rho\sigma_1\sigma_2$, $\eta_3 = -\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}$, $\theta_5 = \frac{\mu_1}{\sigma_1}$, $\theta_6 = \frac{\mu_2}{\sigma_2}$, $\theta_7 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$ and their possible solutions. Here $*$ means a general solution is quite complicated, $g(\cdot,\cdot,\cdot,\cdot)$ is some positive and differentiable function, $h_{11} = (d_1\sigma_1^2 + d_2\rho\sigma_1\sigma_2)/\sqrt{d_1^2\sigma_1^2 + 2d_1d_2\rho\sigma_1\sigma_2 + d_2^2\sigma_2^2}$, $h_{12} = (d_1\rho\sigma_1\sigma_2 + d_2\sigma_2^2)/\sqrt{d_1^2\sigma_1^2 + 2d_1d_2\rho\sigma_1\sigma_2 + d_2^2\sigma_2^2}$, and $h_{35} = (1-\rho^2)[\rho(\sigma_1^2+\sigma_2^2) - 2\sigma_1\sigma_2]/[\sqrt{2}\theta_7]$.

ω	δ_1	δ_2	δ_3	δ_4	δ_5	solution
μ_1	σ_1	$\rho\sigma_2$	0	0	0	$g(\mu_2 - \rho\mu_1\frac{\sigma_2}{\sigma_1}, \sigma_1, \sigma_2, \rho)$
μ_2	$\rho\sigma_1$	σ_2	0	0	0	$g(\mu_1 - \rho\mu_2\frac{\sigma_1}{\sigma_2}, \sigma_1, \sigma_2, \rho)$
σ_1	0	0	$\frac{\sigma_1}{\sqrt{2}}$	$\frac{\sigma_2}{\sqrt{2}}$	$\frac{\rho(1-\rho^2)}{\sqrt{2}}$	$\frac{1}{\rho^2}g(\mu_1, \mu_2, \sigma_2^2(1-\rho^2), \frac{\rho^2}{(1-\rho^2)\sigma_1^2})$
σ_2	0	0	$\frac{\sigma_1}{\sqrt{2}}$	$\frac{\sigma_2}{\sqrt{2}}$	$\frac{\rho(1-\rho^2)}{\sqrt{2}}$	$\frac{1}{\rho^2}g(\mu_1, \mu_2, \sigma_1^2(1-\rho^2), \frac{\rho^2}{(1-\rho^2)\sigma_2^2})$
ρ	0	0	$\frac{\rho\sigma_1}{2}$	$\frac{\rho\sigma_2}{2}$	$1-\rho^2$	$*$
$\boldsymbol{d}'\boldsymbol{\mu}$	h_{11}	h_{12}	0	0	0	$g\left(\mu_2 - \frac{\mu_1\sigma_2(d_1\rho\sigma_1+d_2\sigma_2)}{\sigma_1(d_1\sigma_1+d_2\rho\sigma_2)}, \sigma_1, \sigma_2, \rho\right)$
θ_1	0	0	0	$\rho\sqrt{1-\rho^2}\sigma_2$	$(1-\rho^2)^{3/2}$	$\sigma_2^2 g(\mu_1, \mu_2, \sigma_1, (1-\rho^2)\sigma_2^2)$
θ_2	0	0	0	$\frac{(1-\rho^2)\sigma_2}{\sqrt{2}}$	$-\frac{\rho(1-\rho^2)}{\sqrt{2}}$	$*$
θ_3	0	0	$\frac{1}{2}\sigma_1$	$\frac{1}{2}\sigma_2$	0	$\frac{1}{\sigma_1^2}g(\mu_1, \mu_2, \frac{\sigma_2}{\sigma_1}, \rho)$
θ_4	0	0	$\frac{-\sqrt{1-\rho^2}\sigma_1}{2}$	$\frac{\sqrt{1-\rho^2}\sigma_2}{2}$	0	$g(\mu_1,\ \mu_2,\ \sigma_1\sigma_2,\ \rho)$
σ_{12}	0	0	$\frac{\rho\sigma_1}{\sqrt{1+\rho^2}}$	$\frac{\rho\sigma_2}{\sqrt{1+\rho^2}}$	$\frac{1-\rho^2}{\sqrt{1+\rho^2}}$	$\sqrt{1+\rho^2}g(\mu_1, \mu_2, (1-\rho^2)\sigma_1^2, \frac{\sigma_2}{\sigma_1})$
η_3	0	0	0	$\frac{-\rho(1-\rho^2)\sigma_2}{\sqrt{2}\sqrt{2-\rho^2}}$	$\frac{-(1-\rho^2)\sqrt{2-\rho^2}}{\sqrt{2}}$	$\frac{1}{1-\rho^2}g(\mu_1,\ \mu_2,\ \sigma_1,\ (2-\rho^2)\sigma_2^2)$
θ_5	$\frac{\sqrt{2}\sigma_1}{\sqrt{2+\frac{\mu_1^2}{\sigma_1^2}}}$	$\frac{\sqrt{2}\rho\sigma_2}{\sqrt{2+\frac{\mu_1^2}{\sigma_1^2}}}$	$\frac{-\mu_1}{\sqrt{2}\sqrt{2+\frac{\mu_1^2}{\sigma_1^2}}}$	$\frac{-\frac{\mu_1}{\sigma_1}\rho^2\sigma_2}{\sqrt{2}\sqrt{2+\frac{\mu_1^2}{\sigma_1^2}}}$	$\frac{-\mu_1\rho(1-\rho^2)}{\sqrt{2}\sigma_1\sqrt{2+\frac{\mu_1^2}{\sigma_1^2}}}$	$*$
θ_6	$\frac{\sqrt{2}\rho\sigma_1}{\sqrt{2+\frac{\mu_2^2}{\sigma_2^2}}}$	$\frac{\sqrt{2}\sigma_2}{\sqrt{2+\frac{\mu_2^2}{\sigma_2^2}}}$	$\frac{-\frac{\mu_2}{\sigma_2}\rho^2\sigma_1}{\sqrt{2}\sqrt{2+\frac{\mu_2^2}{\sigma_2^2}}}$	$\frac{-\mu_2}{\sqrt{2}\sqrt{2+\frac{\mu_2^2}{\sigma_2^2}}}$	$\frac{-\mu_2\rho(1-\rho^2)}{\sqrt{2}\sigma_2\sqrt{2+\frac{\mu_2^2}{\sigma_2^2}}}$	$*$
θ_7	0	0	$\frac{\sigma_1(\sigma_1-\rho\sigma_2)^2}{\sqrt{2}\,\theta_7}$	$\frac{\sigma_2(\rho\sigma_1-\sigma_2)^2}{\sqrt{2}\,\theta_7}$	h_{35}	$*$

so equation (69) becomes

$$\sigma_1 \frac{\partial}{\partial \mu_1}\pi + \rho\sigma_2 \frac{\partial}{\partial \mu_2}\pi = 0. \tag{75}$$

(b) Any prior for $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ which does not depend on (μ_1, μ_2) would be a matching prior for μ_1. A general solution of (75) is of the form $g(\mu_2 - \rho\mu_1 \frac{\sigma_2}{\sigma_1}, \sigma_1, \sigma_2, \rho)$ for some positive and differentiable function $g(\cdot, \cdot, \cdot, \cdot)$.

Illustration 2. (a) To see the differential equation of the first order matching prior π for σ_1, note first that $\nabla\sigma_1 = (0, 0, 1, 0, 0)'$. Thus $\boldsymbol{V}\nabla\sigma_1 = (0, 0, \frac{1}{2}\sigma_1^2, \frac{1}{2}\rho^2\sigma_1\sigma_2, \frac{1}{2}\sigma_1\rho(1-\rho^2))'$ and

$$\boldsymbol{\delta} \;\equiv\; (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)' \;=\; \frac{\boldsymbol{V}\nabla\sigma_1}{\sqrt{\nabla\sigma_1{}'\boldsymbol{V}\nabla\sigma_1}} \;=\; \frac{1}{\sqrt{2}}(0, 0, \sigma_1, \rho^2\sigma_2, \rho(1-\rho^2))',$$

so equation (69) becomes

$$\frac{\partial}{\partial \sigma_1}(\sigma_1\pi) + \rho^2 \frac{\partial}{\partial \sigma_2}(\sigma_2\pi) + \frac{\partial}{\partial \rho}\Big(\rho(1-\rho^2)\pi\Big) \;=\; 0. \tag{76}$$

(b) Any prior for $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ which does not depend on $(\sigma_1, \sigma_2, \rho)$ would be a matching prior for σ_1. A general solution of (76) is of the form $\frac{1}{\rho^2} g(\mu_1, \mu_2, \sigma_2^2(1-\rho^2), \frac{\rho^2}{\sigma_1^2(1-\rho^2)})$ for some positive and differentiable function $g(\cdot, \cdot, \cdot, \cdot)$.

Illustration 3. (a) To see the differential equation of the first order matching prior π for ρ, note first that $\nabla\sigma_1 = (0, 0, 0, 0, 1)'$. Thus $\boldsymbol{V}\nabla\rho = (0, 0, \frac{1}{2}\sigma_1\rho(1-\rho^2), \frac{1}{2}\sigma_2\rho(1-\rho^2), (1-\rho^2)^2)'$ and

$$\boldsymbol{\delta} \;\equiv\; (\delta_1, \delta_2, \delta_3, \delta_4, \delta_5)' \;=\; \frac{\boldsymbol{V}\nabla\rho}{\sqrt{\nabla\rho'\boldsymbol{V}\nabla\rho}} \;=\; (0, 0, \frac{1}{2}\rho\sigma_1,\ \frac{1}{2}\rho\sigma_2,\ 1-\rho^2)', \tag{77}$$

so equation (69) becomes

$$\frac{\partial}{\partial \sigma_1}\Big(\frac{1}{2}\rho\sigma_1\pi\Big) + \frac{\partial}{\partial \sigma_2}\Big(\frac{1}{2}\rho\sigma_2\pi\Big) + \frac{\partial}{\partial \rho}\Big((1-\rho^2)\pi\Big) \;=\; 0. \tag{78}$$

The solution to this differential equation is rather complicated and is omitted.

For each of the ten objective priors π_J, π_{IJ}, $\pi_{R\rho}$, $\pi_{R\sigma*}$, π_{RO}, $\pi_{R\lambda}$, π_H, π_S, π_{MS}, and $\pi_{R\sigma}$, we also examine if it is a first order matching prior for each of the parameters such as $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$, $\theta_1 = \frac{\rho\sigma_2}{\sigma_1}$, $\theta_2 = \sigma_2^2(1-\rho^2)$, $\theta_3 = |\boldsymbol{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2)$, $\sigma_{12} = \rho\sigma_1\sigma_2$, $\theta_4 = \frac{\sigma_2^2}{\sigma_1^2}$, $\eta_3 = -\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}$, $\theta_5 = \frac{\mu_1}{\sigma_1}$, $\theta_6 = \frac{\mu_2}{\sigma_2}$, and $\theta_7 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$. The results are listed in Table 7. For example, π_J is a first order matching prior for $\mu_1, \mu_2, \sigma_1, \sigma_2$, $\frac{\sigma_2}{\sigma_1}$, θ_4, θ_5, θ_6, and θ_7, but not for θ_2, θ_3, σ_{12}, and η_3.

Table 7: The first order asymptotic matching of objective priors for $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$, $\mu_1 - \mu_2$, $\theta_1 = \frac{\rho\sigma_2}{\sigma_1}$, $\theta_2 = \sigma_2^2(1-\rho^2)$, $\theta_3 = |\mathbf{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2)$, $\sigma_{12} = \rho\sigma_1\sigma_2$, $\theta_4 = \frac{\sigma_2^2}{\sigma_1^2}$, $\eta_3 = -\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}$, $\theta_5 = \frac{\mu_1}{\sigma_1}$, $\theta_6 = \frac{\mu_2}{\sigma_2}$, $\theta_7 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$. Here a boldface letter indicates exact matching.

prior $\pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$	asymptotic matching: yes	asymptotic matching: no
$\pi_J \propto \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2}$	$\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\sigma_1}, \boldsymbol{\sigma_2}$ $\mu_1-\mu_2, \theta_1,\ \theta_4,\ \boldsymbol{\theta_5},\ \boldsymbol{\theta_6},\ \theta_7$	ρ $\theta_2,\ \theta_3,\ \sigma_{12},\ \eta_3$
$\pi_{IJ} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)^{3/2}}$	μ_1, μ_2 $\mu_1-\mu_2, \theta_1,\ \boldsymbol{\theta_3},\ \theta_4$	σ_1, σ_2, ρ $\theta_2,\ \sigma_{12},\ \eta_3\ \theta_5,\ \theta_6,\ \theta_7$
$\pi_{R\rho} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)}$	μ_1, μ_2, ρ $\mu_1-\mu_2, \theta_3,\ \theta_4$	σ_1, σ_2 $\theta_1,\ \theta_2,\ \sigma_{12},\ \eta_3\ \theta_5,\ \theta_6,\ \theta_7$
$\tilde{\pi}_{R\sigma} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)\sqrt{2-\rho^2}}$	μ_1, μ_2 $\mu_1-\mu_2, \theta_3,\ \theta_4,\ \eta_3$	σ_1, σ_2, ρ $\theta_1,\ \theta_2,\ \sigma_{12},\ \theta_5,\ \theta_6,\ \theta_7$
$\pi_{RO} \propto \frac{1}{\sigma_1^2\sigma_2(1-\rho^2)^{3/2}}$	$\boldsymbol{\mu_1}, \mu_2, \boldsymbol{\sigma_1}$ $\mu_1-\mu_2, \theta_1,\ \boldsymbol{\theta_5}$	σ_2, ρ $\theta_2,\ \theta_3,\ \sigma_{12},\ \theta_4,\ \eta_3,\ \theta_6,\ \theta_7$
$\pi_{R\lambda} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)\sqrt{\left(\frac{\sigma_1}{\sigma_2}-\frac{\sigma_2}{\sigma_1}\right)^2+4\rho^2}}$	μ_1, μ_2 $\mu_1-\mu_2, \theta_3$	σ_1, σ_2, ρ $\theta_1,\ \theta_2,\ \sigma_{12},\ \theta_4,\ \eta_3,\ \theta_5,\ \theta_6,\ \theta_7$
$\pi_H \propto \frac{1}{\sigma_1^2(1-\rho^2)}$	$\boldsymbol{\mu_1}, \mu_2, \boldsymbol{\sigma_1}, \boldsymbol{\rho}$ $\mu_1-\mu_2,\ \boldsymbol{\theta_1},\ \boldsymbol{\theta_2},\ \boldsymbol{\theta_3},\ \boldsymbol{\eta_3},\ \boldsymbol{\theta_5}$	σ_2 $\sigma_{12},\ \theta_4,\ \theta_6,\ \theta_7$
$\pi_S \propto \frac{1}{\sigma_1\sigma_2}$	μ_1, μ_2 $\mu_1-\mu_2, \theta_3,\ \theta_4$	σ_1, σ_2, ρ $\theta_1,\ \theta_2,\ \sigma_{12},\ \eta_3,\ \theta_5,\ \theta_6,\ \theta_7$
$\pi_{MS} \propto \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}}$	μ_1, μ_2 $\mu_1-\mu_2, \theta_3,\ \theta_4$	σ_1, σ_2, ρ $\theta_1,\ \theta_2,\ \sigma_{12},\ \eta_3,\ \theta_5,\ \theta_6,\ \theta_7$
$\pi_{R\sigma} \propto \frac{\sqrt{1+\rho^2}}{\sigma_1\sigma_2(1-\rho^2)}$	μ_1, μ_2 $\mu_1-\mu_2, \theta_3,\ \sigma_{12},\ \theta_4$	σ_1, σ_2, ρ $\theta_1,\ \theta_2,\ \eta_3,\ \theta_5,\ \theta_6,\ \theta_7$

3.3 Numerically Computed Coverage and Summary Recommendations

First order matching is only an asymptotic property, and finite sample performance is more crucial. We thus also implemented a modest numerical study, comparing the numerical values of frequentist coverages of the one-sided credible sets $P(\theta > q_{0.05})$ and $P(\theta < q_{0.95})$ (top four panels) for the nine objective priors $\pi_J, \pi_{IJ}, \pi_{R\rho}, \pi_{R\sigma}$, $\pi_{RO}, \pi_{R\lambda}, \pi_H, \pi_S$, and π_{MS}, when $n = 3$ (the minimal possible sample size and hence the most challenging in terms of obtaining good coverage) and two cases, Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. Here $\theta = \mu_1 - \mu_2, \mu_1, \sigma_1, \sigma_2, \rho, \sigma_1/\sigma_2, \lambda_1, |\boldsymbol{\Sigma}|$ and σ_{12}, given in Figure 2 – 10, respectively. As usual, $q_\alpha = q_\alpha(\boldsymbol{X})$ is the posterior α-quantile of θ, and the coverage probability is computed based on the sampling distribution of $q_\alpha(\boldsymbol{X})$ for the fixed parameter $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ and ρ. Many of the coverage probabilities depend only on ρ, which was thus chosen to be the x-axis in all the graphs.

The bottom four panels in each figure display the frequentist expected values of $q_{0.95}(\mathbf{X})$ and $q_{0.05}(\mathbf{X})$. Among priors with similarly accurate coverage probabilities, those with smallest $E[q_{0.95}(\mathbf{X})]$ and largest $E[q_{0.05}(\mathbf{X})]$ are best, in that the length of the corresponding 90% credible intervals would then be the shortest.

Table 8: Performance of objective priors for each of the parameters

	prior		
parameter	bad	medium	good
μ_1		rest	π_{RO}, π_H, π_J
$\mu_1 - \mu_2$		rest	π_J, π_{RO}
σ_1	π_{IJ}	rest	$\pi_H, \pi_{R\lambda}, \pi_{MS}$
σ_2	$\pi_H, \pi_{RO}, \pi_{IJ}$	rest	π_J
ρ	$\pi_J, \pi_{IJ}, \pi_S, \pi_{RO}$		$\pi_{R\rho}, \pi_{R\sigma}, \pi_{R\lambda}, \pi_H, \pi_{MS}$
λ_1	rest	$\pi_J, \pi_{R\lambda}, \pi_{RO}$	
$\frac{\sigma_1^2}{\sigma_2^2}$	$\pi_H, \pi_J, \pi_{RO}, \pi_{R\lambda}$	rest	
$\lvert\boldsymbol{\Sigma}\rvert$	π_{RO}, π_J	rest	π_{IJ}, π_H
σ_{12}	π_J, π_{IJ} (due to size)	rest	$\pi_H, \pi_{R\rho}, \pi_{R\sigma}$

We are now in a position to justify the recommendations made in Table 3 for the boxed parameters.

- For ρ, we recommend using $\pi_{R\rho}$, since this prior is a one-at-a-time-reference for ρ, first order matching (as shown in Table 7), and has excellent numerical coverage as stated in Table 8. Note that some might prefer to use the right-Haar prior because of its exact matching for ρ (even though it exhibits a marginalization paradox).

- For σ_2/σ_1, the one-at-a-time reference prior was also $\pi_{R\rho}$. As this was first order frequentist matching and among the best in terms of numerical coverage, we also recommend it for this parameter.

- For λ_1, the situation is rather unclear. The one-at-a-time reference prior is $\pi_{R\lambda}$. First order matching results for this parameter are not known, and the numerical coverages of all priors were rather bad. (Note, however, that the sample size was only $n = 3$, and the numerical results strongly suggest that this may be one of the constrained parameter space situations in which simultaneous good Bayesian and frequentist performance is unattainable.) Curiously, π_J had the best numerical coverages; it's simplicity in computation also would argue for this choice. $\pi_{R\lambda}$ has virtually the same coverage, however, and, as the reference prior, is our recommendation.

- For σ_{12}, the only first order matching prior among our candidates is $\pi_{R\sigma}$. It also had the best numerical coverages, and so is a clear recommendation. Note, however, that we were not able to determine if it is a one-at-a-time reference prior for σ_{12}, so the recommendation should be considered tentative.

- The most interesting question is what to recommend for general use, as an all-purpose prior. Looking at Table 3, it might seem that π_H would be a good choice, since it is optimal for so many parameters. Even π_J might appeal, as it also gave exact frequentist coverage for many parameters. However, both these priors can also give quite bad coverages, as indicated in Figures 5 and 7 for π_H, and in Figures 6, 7, 9, and 10 for π_J. Indeed, from Table 8, the only priors that did not have significantly poor performance for at least one parameter (other than λ_1, for which no prior gave good coverages) were $\pi_{R\rho}$, $\pi_{R\sigma}$,and π_{MS}. Computationally, all three were are easy to use via rejection sampling, but the acceptance probability (from Table 5) for the first two is quite a bit larger than for π_{MS}. Also π_{MS} did not arise as a reference prior. The numerical coverages for $\pi_{R\rho}$ and $\pi_{R\sigma}$ are virtually identical for all the parameters, so there is no principled way to choose between them. $\pi_{R\rho}$ is a commonly used prior and somewhat simpler, so it becomes our recommended choice for a general prior.

4 Multivariate Normal

We now consider generalization to the p-variate multivariate normal situation, with density

$$f(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \;=\; (2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big). \tag{79}$$

The most interesting priors discussed in the bivariate case – π_J, π_{IJ} $\pi_{R\rho}$, π_H, and $\pi_{R\lambda}$ – all generalize to the multivariate case. In this section, we record these generalizations and provide simple methodology for computing with the resulting posteriors. Extensive evaluation of these priors in the multivariate case would be too large an enterprise for this paper, but we expect that the recommendations for use of the priors in the bivariate case will extend to their multivariate generalizations.

The Jeffreys prior in the multivariate case is $\pi_J(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-(p+2)/2}$. Interestingly, for $p > 2$, this prior does not have the exact matching performance detailed in Table 2 for the bivariate Jeffreys prior. Indeed, Geisser and Cornfield (1963) show that the prior which is exact matching for all means and variances (and which also yields Fisher's fiducial distribution for these parameters) is $\pi_{GC}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-p}$. It is simple chance that this prior happens to be the Jeffreys prior for $p = 2$ (and perhaps simple chance that it is the independence Jeffreys prior for $p = 1$).

One can thus recommend use of $\pi_{GC}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-p}$ if only individual means (or contrasts) and variances are of interest, but it is likely a very bad prior for correlations, prediction, or other inferences involving a multivariate normal distribution. Also, curiously, when $p > 2$ there would seem to be no justification (from our perspective) for the Jeffreys prior or the independence Jeffreys prior $\pi_{IJ}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-(p+1)/2}$, as they do not seem to be exact matching for any parameters and are not one-at-a-time reference priors. Furthemore, see Berger & Yang (1994) for discussion of the large literature (much of it frequentist) that indicates problems with use of π_J and π_{IJ} in higher dimensions. Surprisingly, the Jeffreys and independence Jeffreys priors are the most commonly used objective priors for the multivariate case, a common practice that is clearly questionable. The alternative generalizations in the next section are thus particularly attractive.

It should be noted that, in the more recent Bayesian literature, aggressive shrinkage of eigenvalues, correlations, or other features of the covariance matrix are entertained; cf. Daniels & Kass (1999), Daniels & Pourahmadi (2002), Liechty et al. (2004) and the references therein. This may well be desirable in many practical situations, but is more aggressive in its prior assumptions than the objective priors we consider next, which are all one-at-a-time reference priors.

4.1 Other Prior Generalizations

The generalization of $\pi_{R\rho}$ to the multivariate case was given in Chang & Eaves (1990), in the sense that they derived the reference prior for a covariance matrix under the ordering $\{\boldsymbol{\Upsilon}, (\sigma_1, \cdots, \sigma_p)\}$, where $\boldsymbol{\Upsilon}$ is the correlation matrix. The resulting prior is

$$\pi_{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma})\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma} = \frac{1}{|\boldsymbol{\Sigma}|^{(p+1)/2}|\boldsymbol{I}_p + \boldsymbol{\Sigma}^* \boldsymbol{\Sigma}^{-1}|^{1/2}}\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma} \tag{80}$$

$$= 2^p \left[\prod_{i=1}^{p} \frac{d\mu_i d\sigma_i}{\sigma_i}\right] \left[\frac{1}{|\boldsymbol{\Upsilon}|^{(p+1)/2}|\boldsymbol{I}_p + \boldsymbol{\Upsilon}^* \boldsymbol{\Upsilon}^{-1}|^{1/2}} \prod_{i<j} d\rho_{ij}\right], \tag{81}$$

where $\boldsymbol{A}^*\boldsymbol{B}$ denotes the Hadamard product of the squared matrices $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$ with entries $c_{ij} = a_{ij}b_{ij}$.

The generalization of $\pi_{R\lambda}$ to the multivariate case was also found in Chang and Eaves (1990), in the sense that they derived the reference prior for a covariance matrix under the ordering $\{(\lambda_1, \cdots, \lambda_p), \boldsymbol{O}\}$, where $\lambda_1 > \cdots > \lambda_p$ are eigenvalues of $\boldsymbol{\Sigma}$, and $\boldsymbol{O}$ is an orthogonal matrix such that $\boldsymbol{\Sigma} = \boldsymbol{O}' diag(\lambda_1, \cdots, \lambda_p)\boldsymbol{O}$. This reference prior was discussed in detail in Berger and Yang (1994), who also showed it to be a one-at-a-time reference prior for the ordering $\{\lambda_1, \cdots, \lambda_p, O_{11}, \cdots\}$. The prior has the form

$$\pi_E(\mu, \boldsymbol{\Sigma})\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma} = \frac{I_{[\lambda_1 > \cdots > \lambda_p]}}{|\boldsymbol{\Sigma}| \prod_{i<j}(\lambda_i - \lambda_j)}\, d\boldsymbol{\mu}\, d\boldsymbol{\Sigma}\,,$$

where $\lambda_1 > \lambda_2 > \ldots > \lambda_p$ are the eigenvalues of $\boldsymbol{\Sigma}$.

The generalization of the right Haar prior has been extensively studied (see, e.g., Eaton & Sudderth (2002)). It is most convenient to express this prior in terms of a lower-triangular matrix $\mathbf{\Psi}$ with positive diagonal elements and such that

$$\mathbf{\Sigma}^{-1} = \mathbf{\Psi}'\mathbf{\Psi}. \tag{82}$$

(Note that there are many such matrices, so that the right Haar prior is not unique.)

The right Haar prior corresponding to this decomposition is given by

$$\pi_H(\mu, \mathbf{\Psi})\, d\boldsymbol{\mu}\, d\mathbf{\Psi} = \prod_{i=1}^{p} \frac{1}{\psi_{ii}^{i}}\, d\boldsymbol{\mu}\, d\mathbf{\Psi}\,. \tag{83}$$

As in Facts 10 and 11 in Appendix A, one can also show that this prior is the reference prior with respect to the ordering $\{\psi_{11}, \cdots, \psi_{pp}, \psi_{ij}/\psi_{jj}, 1 \le j < i \le p\}$.

There is a very large related literature on *generalized Wishart priors*; see Brown (2001) for a review.

4.2 Posterior Computation

Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ be a random sample from $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$. The likelihood function of $(\boldsymbol{\mu}, \mathbf{\Sigma})$ is given by

$$L(\boldsymbol{\mu}, \mathbf{\Sigma}) \;=\; (2\pi)^{-np/2}|\mathbf{\Sigma}|^{-n/2} \exp\Big\{ -\frac{n}{2}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) - \frac{1}{2}tr(\boldsymbol{S}\mathbf{\Sigma}^{-1})\Big\},$$

where $\overline{\boldsymbol{X}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i$ and $\boldsymbol{S} = \sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})'$. Since all the considered priors are constant in $\boldsymbol{\mu}$, the conditional posterior for $\boldsymbol{\mu}$ will be

$$[\boldsymbol{\mu} \mid \mathbf{\Sigma}, \boldsymbol{X}] \sim N_p(\overline{\boldsymbol{x}}, n^{-1}\mathbf{\Sigma}). \tag{84}$$

Generation from this is standard, so the challenge of simulation from the posterior distribution requires only sampling from the marginal posterior of $\mathbf{\Sigma}$ given $\boldsymbol{S}$. Note that the marginal likelihood of $\boldsymbol{S}$ satisfies

$$L_1(\boldsymbol{S}) \propto \frac{1}{\{|\mathbf{\Sigma}|\}^{(n-1)/2}} etr\Big(-\frac{1}{2}\mathbf{\Sigma}^{-1}\boldsymbol{S}\Big). \tag{85}$$

Marginal Posteriors Under π_J and π_{IJ}: It is immediate that these marginal posteriors for $\mathbf{\Sigma}$ are Inverse Wishart $(\boldsymbol{S}^{-1}, n)$ and Inverse Wishart $(\boldsymbol{S}^{-1}, n-1)$, respectively.

Marginal Posterior Under π_{CE}: This marginal posterior distribution is imposing in its complexity. However, rather remarkably there is a simple rejection algorithm (see Section 2.1) that can be used to generate from it:

Step 1. Generate $\mathbf{\Sigma} \sim$ Inverse Wishart$(\boldsymbol{S}^{-1}, n-1)$.

Step 2. Simulate $u \sim$ Uniform$(0, 1)$. If $u \le 2^{p/2}\,|\boldsymbol{I}_p + \mathbf{\Sigma}^{*}\mathbf{\Sigma}^{-1}|^{-1/2}$, report $\mathbf{\Sigma}$. Otherwise go back to *Step 1.*

Note that the acceptance probability $2^{p/2}\,|\boldsymbol{I}_p + \boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}|^{-1/2}$ is equal to one if the proposed $\boldsymbol{\Sigma}$ is diagonal, but is near zero when the proposed $\boldsymbol{\Sigma}$ is nearly singular. That this algorithm is a valid accept-reject algorithm, based on generation of $\boldsymbol{\Sigma}$ from the independence Jeffreys posterior, follows from the following lemma (see also Section 2.1).

Lemma 4 *For the priors π_{IJ} and π_{CE},*

$$\frac{\pi_{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\pi_{IJ}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \le 2^{-p/2}, \quad \forall\ \boldsymbol{\mu}, \boldsymbol{\Sigma}. \tag{86}$$

Proof. First,

$$\frac{\pi_{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\pi_{IJ}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \frac{1}{|\boldsymbol{I}_p + \boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}|^{1/2}}. \tag{87}$$

Clearly,

$$|\boldsymbol{I}_p + \boldsymbol{\Sigma}^*\boldsymbol{\Sigma}^{-1}| = |\boldsymbol{I}_p + \boldsymbol{\Upsilon}^*\boldsymbol{\Upsilon}^{-1}| = |\boldsymbol{\Upsilon}^*(\boldsymbol{I}_p + \boldsymbol{\Upsilon}^{-1})|.$$

Using the fact that, for any positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $|\boldsymbol{A}^*\boldsymbol{B}| \ge |\boldsymbol{B}|\prod_{i=1}^p a_{ii}$, it follows that

$$|\boldsymbol{\Upsilon}^*(\boldsymbol{I}_p + \boldsymbol{\Upsilon}^{-1})| \ge |\boldsymbol{I}_p + \boldsymbol{\Upsilon}^{-1}| = \prod_{i=1}^p \Big(1 + \frac{1}{\eta_i}\Big),$$

where $\eta_1, \cdots, \eta_p$ are the eigenvalues of $\boldsymbol{\Upsilon}$. Note that $\sum_{i=1}^p \eta_i = trace(\boldsymbol{\Upsilon}) = p$, and the minimum of $\prod_{i=1}^p (1 + 1/\eta_i)$ subject to $\eta_i \ge 0$ and $\sum_{i=1}^p \eta_i = p$, is achieved at $\eta_1 = \cdots = \eta_p = 1$. $\square$

Marginal Posterior Under π_E: It is possible to generate from this posterior using the following Metropolis-Hastings algorithm from Berger et. al. (2005).

Step 1. Generate $\boldsymbol{\Sigma}^* \sim$ Inverse Wishart$(\boldsymbol{S}^{-1}, n-1)$.

Step 2. Set $\boldsymbol{\Sigma}' = \begin{cases} \boldsymbol{\Sigma}^* & \text{with probability } \alpha, \\ \boldsymbol{\Sigma} & \text{otherwise,} \end{cases}$

where

$$\alpha = \min\left\{1,\ \frac{\prod_{i<j}(\lambda_i^* - \lambda_j^*)}{\prod_{i<j}(\lambda_i - \lambda_j)} \cdot \frac{|\boldsymbol{\Sigma}|^{(p-1)/2}}{|\boldsymbol{\Sigma}^*|^{(p-1)/2}}\right\}.$$

Constructive Marginal Posterior Under π_H: As in the bivariate case, there is an explicit constructive posterior available for computing with the right Haar prior. Also as in the bivariate case, this can best be written in terms of transformed parameters.

In the following, we write $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_p$, because the dimension is needed for certain useful recursive formulas. Let $\boldsymbol{\psi}_{p-1,p}$ represent the $(p-1)\times 1$ vector of the last column of $\boldsymbol{\Psi}'_p$ excluding ψ_{pp}. Define

$$\boldsymbol{\Psi}_1 = \psi_{11},\ \boldsymbol{\Psi}_2 = \begin{pmatrix} \psi_{11} & 0 \\ \psi_{21} & \psi_{22} \end{pmatrix},\ \cdots, \boldsymbol{\Psi}_p = \begin{pmatrix} \boldsymbol{\Psi}_{p-1} & \boldsymbol{0} \\ \boldsymbol{\psi}'_{p,p-1} & \psi_{pp} \end{pmatrix}.$$

Likewise write $\boldsymbol{S} = \boldsymbol{S}_p$, and let $\boldsymbol{s}_{p,p-1}$ represent the $(p-1) \times 1$ vector of the last column of $\boldsymbol{S}_p$ excluding s_{pp}. Define

$$\boldsymbol{S}_1 = s_{11},\ \boldsymbol{S}_2 = \begin{pmatrix} s_{11} & s_{21} \\ s_{21} & s_{22} \end{pmatrix},\ \cdots, \boldsymbol{S}_p = \begin{pmatrix} \boldsymbol{S}_{p-1} & \boldsymbol{s}_{p,p-1} \\ \boldsymbol{s}'_{p,p-1} & s_{pp} \end{pmatrix}. \tag{88}$$

Finally, define

$$w_i = \begin{cases} s_{11}, & \text{if } i = 1, \\ \frac{|\boldsymbol{S}_i|}{|\boldsymbol{S}_{i-1}|} = s_{ii} - \boldsymbol{s}'_{i,i-1}\boldsymbol{S}_{i-1}^{-1}\boldsymbol{s}_{i,i-1}, & \text{if } i = 2, \cdots, p. \end{cases} \tag{89}$$

Letting χ^{2*}_{n-i} denote independent draws from chi-squared distributions with the indicated degree of freedoms, and $\boldsymbol{z}^*_{i,i-1}$ denote independent draws from $N_{i-1}(\mathbf{0}, \boldsymbol{I}_{i-1})$, the constructive posterior of $(\psi_{11}, \cdots, \psi_{pp}, \boldsymbol{\psi}_{2,1}, \cdots, \boldsymbol{\psi}_{p,p-1})$ given $\boldsymbol{X}$ can be expressed as

$$\psi^*_{ii} = \sqrt{\frac{\chi^{2*}_{n-i}}{w_i}}, \quad i = 1, \cdots, p, \tag{90}$$

$$\begin{aligned} \boldsymbol{\psi}^*_{i,i-1} &= \boldsymbol{S}_{i-1}^{-1/2}\boldsymbol{z}^*_{i,i-1}, -\psi^*_{ii}\boldsymbol{S}_{i-1}^{-1}\boldsymbol{s}_{i,i-1} \\ &= \boldsymbol{S}_{i-1}^{-1/2}\boldsymbol{z}^*_{i,i-1} - \sqrt{\frac{\chi^{2*}_{n-i}}{w_i}}\boldsymbol{S}_{i-1}^{-1}\boldsymbol{s}_{i,i-1}, \quad i = 2, \cdots, p. \end{aligned} \tag{91}$$

Letting $\boldsymbol{\Psi}^* = (\psi^*_{ij})$, the constructive posterior of $\boldsymbol{\Sigma}$ is simply $\boldsymbol{\Sigma}^* = \boldsymbol{\Psi}^{*-1}(\boldsymbol{\Psi}^{*-1})'$.

Appendix A: Derivation of Reference Priors

A.1 When μ_1 Is of Interest

We give the one-at-a-time reference priors for μ_1. Since the Fisher information matrix does not depend on (μ_1, μ_2), the parameters (μ_1, μ_2) can be put in any place in the parameter ordering; without loss of generality, we thus place them last. For the other three parameters, there are six different orderings. Since σ_1 and σ_2 play the same role among the three parameters $(\sigma_1, \sigma_2, \rho)$, we consider only the three orderings: $\{\sigma_1, \sigma_2, \rho\}$, $\{\sigma_1, \rho, \sigma_2\}$ and $\{\rho, \sigma_1, \sigma_2\}$.

In the following derivation, we use the notation from Berger and Bernardo (1992a) by choosing the compact sets:

$$\begin{aligned} \mu_1 &\in [-k_1 i, k_1 i]; \quad \mu_2 \in [-k_2 i, k_2 i]; \\ \sigma_1 &\in [e^{-k_3 i}, e^{k_3 i}]; \quad \sigma_2 \in [e^{-k_4 i}, e^{k_4 i}]; \\ \rho &\in \left[\frac{-1 + e^{-k_5 i}}{1 + e^{-k_5 i}}, \frac{1 - e^{-k_5 i}}{1 + e^{-k_5 i}}\right], \end{aligned}$$

where $k_j > 0, j = 1, \cdots, 5$ are fixed constants.

Fact 4 *(a). The reference priors for $\{\mu_1, \mu_2, \sigma_1, \sigma_2, \rho\}$ and $\{\sigma_1, \sigma_2, \rho, \mu_1, \mu_2\}$ are the same, and given by*

$$\pi_{R\sigma}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \propto \frac{1}{\sigma_1\sigma_2(1 - \rho^2)}\sqrt{1 + \rho^2}\,. \tag{92}$$

(b). The reference priors for $\{\mu_1,\mu_2,\sigma_1,\rho,\sigma_2\}$ and $\{\sigma_1,\rho,\sigma_2,\mu_1,\mu_2\}$ are the same, given by

$$\pi_{R\sigma*}(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) \;\propto\; \frac{1}{\sigma_1\sigma_2(1-\rho^2)}\frac{\sqrt{2}}{\sqrt{2-\rho^2}}. \tag{93}$$

(c). The reference prior for $\{\mu_1,\mu_2,\rho,\sigma_1,\sigma_2\}$ is the same as that of $\{\rho,\sigma_1,\sigma_2,\mu_1,\mu_2\}$, given by

$$\pi_{R\rho}(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) \;\propto\; \frac{1}{\sigma_1\sigma_2(1-\rho^2)}. \tag{94}$$

Proof. The proof follows the algorithm in Berger and Bernardo (1992a) for defining a one-at-a-time reference prior, and we use the notation of that algorithm. For (a), the first part is obvious since the Fisher information matrix is diagonal and does not depend on (μ_1,μ_2). To derive the form of the reference prior for $\{\sigma_1,\sigma_2,\rho\}$, note that

$$h_1 \;\propto\; \frac{1}{\sigma_1^2},\; h_2 \;\propto\; \frac{2}{\sigma_2^2(1-\rho^4)},\; h_3 \;\propto\; \frac{(1+\rho^2)}{(1-\rho^2)^2}.$$

Thus

$$\begin{aligned}
\pi_i(\rho\mid\sigma_1,\sigma_2,\mu_1,\mu_2) &= K_3\sqrt{h_3}, \quad K_3^{-1}=\int_{(-1+e^{-k_5 i})/(1+e^{-k_5 i})}^{(1-e^{-k_5 i})/(1+e^{-k_5 i})}\sqrt{h_3}d\rho,\\
\pi_i(\sigma_2\mid\sigma_1,\mu_1,\mu_2) &= \frac{K_2}{\sigma_2} \;\propto\; \sqrt{h_2}, \quad K_2=\frac{1}{2k_4 i},\\
\pi_i(\sigma_1\mid\mu_1,\mu_2) &= \frac{K_1}{\sigma_1} \;\propto\; \sqrt{h_1}, \quad K_1=\frac{1}{2k_3 i},\\
\pi_i(\mu_1,\mu_2) &= \frac{1}{4k_1k_2 i^2}.
\end{aligned}$$

Choosing $(\sigma_{10},\sigma_{20},\rho_0,\mu_{10},\mu_{20})=(1,1,0,0,0)$, it follows that $\pi(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho)$ equals

$$\begin{aligned}
&\lim_{i\to\infty}\frac{\pi_i(\mu_1,\mu_2)\pi_i(\sigma_1\mid\mu_1,\mu_2)\pi_i(\sigma_2\mid\sigma_1,\mu_1,\mu_2)\pi_i(\rho\mid\sigma_1,\sigma_2,\mu_1,\mu_2)}{\pi_i(\mu_{10},\mu_{20})\pi_i(\sigma_{10}\mid\mu_{10},\mu_{20})\pi_i(\sigma_{20}\mid\sigma_{10},\mu_{10},\mu_{20})\pi_i(\rho_0\mid\sigma_{10},\sigma_{20},\mu_{10},\mu_{20})}\\
&\quad= \frac{\sqrt{1+\rho^2}}{\sigma_1\sigma_2(1-\rho^2)}.
\end{aligned}$$

This proves (a). Note that the result does not depend on the choice of the k_j.

To prove part (b), note that

$$h_1 \;=\; \frac{1}{\sigma_1^2};\; h_2 \;=\; \frac{1}{\frac{\sigma_1^2}{2}(1-\rho^2)^2(1-\frac{\rho^2}{2})};\; h_3 \;=\; \frac{2-\rho^2}{\sigma_2^2(1-\rho^2)},$$

and follow the same argument. To prove part (c), note that

$$h_1 \;=\; \frac{1}{(1-\rho^2)^2};\; h_2 \;=\; \frac{1}{\frac{\sigma_1^2}{2}(1-\frac{\rho^2}{2})};\; h_3 \;=\; \frac{1}{\sigma_2^2},$$

and follow the same argument. □

A.2 When ρ Is of Interest

Fact 5 *The reference priors for the orderings* $\{\rho, \sigma_1, \sigma_2, \mu_1, \mu_2\}$, $\{\rho, \sigma_2, \sigma_1, \mu_1, \mu_2\}$, $\{\mu_1, \mu_2, \rho, \sigma_2, \sigma_1\}$, *and* $\{\mu_1, \mu_2, \rho, \sigma_1, \sigma_2\}$ *are all equal to* $\pi_{R\rho}(\cdot)$, *given by (94).*

The result follows from the proof of Fact 4.

A.3 When σ_1 Is of Interest

According to the comments at beginning of Section 4.2, we only consider the two orderings $\{\sigma_1, \sigma_2, \rho, \mu_1, \mu_2\}$ and $\{\sigma_1, \rho, \sigma_2, \mu_1, \mu_2\}$. The corresponding reference priors can easily be shown to be π_1 and $\pi_{R\sigma}$, given by (93) and (94), respectively.

Remark 3 *Note that none of the priors π_1, $\pi_{R\sigma}$ and $\pi_{R\rho}$ satisfy equation (76). Consequently, they are not first order matching priors for σ_1. In the following, we will find matching reference priors when σ_1 is of interest by choosing different nuisance parameters.*

The Fisher information of $(\mu_1, \mu_2, \eta_1, \eta_2, \eta_3)$ is of the form

$$\tilde{\boldsymbol{I}} \;\equiv\; \tilde{\boldsymbol{I}}(\mu_1, \mu_2, \eta_1, \eta_2, \eta_3) \;=\; \boldsymbol{H}'\boldsymbol{I}\boldsymbol{H} = \begin{pmatrix} \eta_1^2 + \eta_3^2 & \eta_2\eta_3 & 0 & 0 & 0 \\ \eta_2\eta_3 & \eta_2^2 & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{\eta_1^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{2\eta_1^2+\eta_3^2}{\eta_1^2\eta_2^2} & -\frac{\eta_3}{\eta_1^2\eta_2} \\ 0 & 0 & 0 & -\frac{\eta_3}{\eta_1^2\eta_2} & \frac{1}{\eta_1^2} \end{pmatrix}.$$

The inverse of the Fisher Information matrix is then

$$\tilde{\boldsymbol{V}} \;=\; \tilde{\boldsymbol{I}}^{-1} \;=\; \begin{pmatrix} \frac{1}{\eta_1^2} & -\frac{\eta_3}{\eta_1^2\eta_2} & 0 & 0 & 0 \\ -\frac{\eta_3}{\eta_1^2\eta_2} & \frac{\eta_1^2+\eta_3^2}{\eta_1^2\eta_2^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{\eta_1^2}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{\eta_2^2}{2} & \frac{\eta_2\eta_3}{2} \\ 0 & 0 & 0 & \frac{\eta_2\eta_3}{2} & \eta_1^2 + \frac{\eta_3^2}{2} \end{pmatrix}.$$

Fact 6 *(a) The one-at-a-time reference prior for the ordering $\{\mu_1, \mu_2, \eta_1, \eta_2, \eta_3\}$ is*

$$\pi_{RO}(\mu_1, \mu_2, \eta_1, \eta_2, \eta_3) \propto \frac{1}{\eta_1\eta_2}. \tag{95}$$

(b) The one-at-a-time reference prior for the ordering $\{\mu_1, \mu_2, \eta_1, \eta_3, \eta_2\}$ is

$$\pi_5(\mu_1, \mu_2, \eta_1, \eta_2, \eta_3) \;\propto\; \frac{1}{\eta_1\eta_2\sqrt{\eta_1^2 + \frac{1}{2}\eta_3^2}}. \tag{96}$$

Proof. For part (a), note that $\pi(\eta_3 \mid \eta_2, \eta_1) \propto 1$, $\pi(\eta_2 \mid \eta_1) \propto 1/\eta_2$, and $\pi(\eta_1) \propto 1/\eta_1$. For part (b), note that $\pi(\eta_2 \mid \eta_3, \eta_1) \propto 1/\eta_2$, $\pi(\eta_3 \mid \eta_1) \propto 1/\sqrt{\eta_1^2 + \frac{1}{2}\eta_3^2}$, and $\pi(\eta_1) \propto 1/\eta_1$. □

Remark 4 *The reference prior for the ordering* $\{\mu_1, \mu_2, \eta_1, (\eta_2, \eta_3)\}$ *is also* π_{RO}. *This prior was considered by Roverato & Consonni (2004) in the context of graphical modeling.*

The next fact, which is straightforward to prove, gives the transformed forms of the priors π_{RO} and π_5, in terms of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

Fact 7 *(a) The reference prior for* σ_1 *(or* η_1*) with ordered nuisance parameters* $\{\mu_1, \mu_2, \eta_2, \eta_3\}$ *is*

$$\pi_{RO}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \quad \propto \quad \frac{1}{\sigma_1^2 \sigma_2 (1-\rho^2)^{3/2}}. \tag{97}$$

(b) π_{RO} *is a matching prior for* σ_1.
(c) The reference prior for σ_1 *with the ordered nuisance parameters* $\{\mu_1, \mu_2, \eta_3, \eta_2\}$ *is the same as* $\pi_{R\sigma}$, *given by (93).*
(d) The reference prior for σ_2, *with nuisance parameters* $(\mu_1, \mu_2, 1/(\sigma_1\sqrt{1-\rho^2}), -\rho/(\sigma_2\sqrt{1-\rho^2}))$, *is*

$$\pi_6(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \quad \propto \quad \frac{1}{\sigma_1 \sigma_2^2 (1-\rho^2)^{3/2}}. \tag{98}$$

(e) π_6 *is a matching prior for* σ_2.

Note that π_{RO} is not symmetric with respect to σ_1 and σ_2. Furthermore, π_{RO} is not a matching prior for any of the parameters (σ_2, ρ). Similarly, π_6 is not symmetric with respect to σ_1 and σ_2 and π_6 is not a matching prior for any of the parameters (σ_1, ρ).

A.4 When Other Parameters Are of Interest

Fact 8 *Define* $\theta_4 = \sigma_2^2/\sigma_1^2$ *and* $\xi = \sigma_1\sigma_2$.

(a) The one-at-a-time reference prior for the ordering $\{\theta_4, \rho, \xi, \mu_1, \mu_2\}$ *is* $\pi_{R\rho}$.

(b) The one-at-a-time reference prior for the ordering $\{\rho, \theta_4, \xi, \mu_1, \mu_2\}$ *is* $\pi_{R\rho}$.

Proof. Note that the Fisher information matrix of $(\theta_4, \rho, \xi, \mu_1, \mu_2)$ is

$$diag\left(\frac{1}{\theta_4^2(1-\rho^2)}, \begin{pmatrix} \frac{1+\rho^2}{1-\rho^2} & \frac{-\rho}{\xi(1-\rho^2)} \\ \frac{-\rho}{\xi(1-\rho^2)} & \frac{1}{\xi^2} \end{pmatrix}, \mathbf{\Sigma}^{-1}\right).$$

Thus the reference prior for $(\theta_4, \rho, \xi, \mu_1, \mu_2)$ is

$$\pi(\theta_4, \rho, \xi, \mu_1, \mu_2) \propto \frac{1}{\theta_4\, \xi\, (1-\rho^2)}.$$

Part (a) follows. Part (b) can be proved similarly. □

Fact 9 *The one-at-a-time reference prior for* $\{\eta_1 = 1/\sigma_1, \eta_2 = 1/\sqrt{\sigma_2^2(1-\rho^2)},\ \theta_1 = -\rho\sigma_2/\sigma_1, \mu_1, \mu_2\}$, *with any ordering, is the right Haar prior* π_H.

Proof. Because $\theta_1 = \eta_3/\eta_2$, the Fisher information matrix of $(\eta_1, \eta_2, \theta_1, \mu_1, \mu_2)$ is

$$\boldsymbol{I} = diag\Big(\frac{2}{\eta_1^2}, \frac{2}{\eta_2^2}, \frac{\eta_2^2}{\eta_1^2}, \boldsymbol{\Sigma}^{-1}\Big), \tag{99}$$

where $\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \eta_1^2 + \eta_2^2\theta_1^2 & \eta_2^2\theta_1 \\ \eta_2^2\theta_1 & \eta_2^2 \end{pmatrix}$. It is straightforward to show that the one-at-a-time reference prior for $\{\eta_1, \eta_2, \theta_1, \mu_1, \mu_2\}$ is $\pi(\eta_1, \eta_2, \theta_1, \mu_1, \mu_2) \propto 1/(\eta_1\eta_2)$, which is equivalent to π_H. □

Fact 10 *Define $\xi_1 = \eta_1\eta_2$ and $\xi_2 = \eta_1/\eta_2$.*

(a) The one-at-a-time reference prior for $\{\xi_1, \xi_2, \theta_1, \mu_1, \mu_2\}$, with any ordering, is the right Haar prior π_H.

(b) Define $\theta_3 = |\boldsymbol{\Sigma}| = \xi_1^2$. The one-at-a-time reference prior for $\{\theta_3, \xi_2, \theta_1, \mu_1, \mu_2\}$, with any ordering, is the right Haar prior π_H.

Proof. The Fisher information matrix of $(\xi_1, \xi_2, \theta_1, \mu_1, \mu_2)$ is

$$\boldsymbol{I} = diag\left(\frac{1}{\xi_1^2}, \frac{1}{\xi_2^2}, \frac{1}{\xi_2^2}, \boldsymbol{\Sigma}^{-1}\right), \tag{100}$$

where $\boldsymbol{\Sigma}^{-1} = \frac{\xi_1}{\xi_2}\begin{pmatrix} \xi_2^2 + \theta_1^2 & \theta_1 \\ \theta_1 & 1 \end{pmatrix}$. It is straightforward to show that the one-at-a-time reference prior for $\{\xi_1, \xi_2, \theta_1, \mu_1, \mu_2\}$ is of the form $\pi(\xi_1, \xi_2, \theta_1, \mu_1, \mu_2) \propto 1/(\xi_1\xi_2)$, which is equivalent to π_H. This proves Part (a). Part (b) is immediate. □

A.5 When the Eigenvalues of $\boldsymbol{\Sigma}$ Are of Interest

Chang and Eaves (1990) derived a reference prior for the multivariate normal covariance matrix, when the eigenvalues are the parameters of interest. This was extensively discussed in Yang and Berger (1994), wherein it was also shown that this was a one-at-a-time reference prior for the ordered eigenvalues. In the bivariate case, let $\lambda_1 > \lambda_2$ be the two eigenvalues of $\boldsymbol{\Sigma}$. The reference prior is given by

$$\pi_{R\lambda}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \;\propto\; \frac{1}{|\boldsymbol{\Sigma}|(\lambda_1 - \lambda_2)}. \tag{101}$$

It is easy to see that

$$\begin{aligned} \lambda_1 &= \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2 + \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2\sigma_1^2\sigma_2^2}\right), \\ \lambda_2 &= \frac{1}{2}\left(\sigma_1^2 + \sigma_2^2 - \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2\sigma_1^2\sigma_2^2}\right), \end{aligned}$$

and $\lambda_1 - \lambda_2 = \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2\sigma_1^2\sigma_2^2}$. The corresponding reference prior for $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ in terms of the original parameters is then

$$\begin{aligned} \pi_{R\lambda}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) &\propto \frac{1}{(1-\rho^2)\sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\sigma_1^2\sigma_2^2\rho^2}} \\ &= \frac{1}{\sigma_1\sigma_2(1-\rho^2)} \frac{1}{\sqrt{\left(\frac{\sigma_1}{\sigma_2} - \frac{\sigma_2}{\sigma_1}\right)^2 + 4\rho^2}}. \end{aligned} \tag{102}$$

Appendix B: Proofs

Proof of Theorem 3. With the constant prior for (μ_1, μ_2), it is easy to see that the marginal likelihood of $(\sigma_1, \sigma_2, \rho)$ given $\boldsymbol{X}$ depends on $\boldsymbol{S}$ and is proportional to

$$|\boldsymbol{\Sigma}|^{-(n-1)/2} f(trace(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})),$$

where $f(t) = \exp(-t^2/2)$. Define

$$\begin{aligned}
\mathcal{D} &= \{(\sigma_1^*, \sigma_2^*, \rho^*) : \sigma_1^{*d_1}\sigma_2^{*d_2} g(\rho^*) < \sigma_1^{d_1}\sigma_2^{d_2} g(\rho)\}, \\
G(\boldsymbol{X}, \sigma_1, \sigma_2, \rho) &= \int_{\mathcal{D}} \pi(\sigma_1^*, \sigma_2^*, \rho^* \mid \boldsymbol{S}) d\sigma_1^* d\sigma_2^* d\rho^*.
\end{aligned}$$

Clearly, the frequentist coverage

$$P\{\theta < \theta_{1-\alpha}(\boldsymbol{X}) \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho\} = P\{G(\boldsymbol{S}, \sigma_1, \sigma_2, \rho) < 1-\alpha \mid \sigma_1, \sigma_2, \rho\}.$$

Under the prior (37),

$$G(\boldsymbol{X}, \sigma_1, \sigma_2, \rho) = \frac{\iiint_{\mathcal{D}} \frac{h(\rho^*)}{\sigma_1^{*(n-1+c_1)}\sigma_2^{*(n-1+c_2)}(1-\rho^{*2})^{(n-1)/2}} f(trace(\boldsymbol{S}\boldsymbol{\Sigma}^{*-1})) d\sigma_1^* d\sigma_2^* d\rho^*}{\iiint \frac{h(\rho^*)}{\sigma_1^{*(n-1+c_1)}\sigma_2^{*(n-1+c_2)}(1-\rho^{*2})^{(n-1)/2}} f(trace(\boldsymbol{S}\boldsymbol{\Sigma}^{*-1})) d\sigma_1^* d\sigma_2^* d\rho^*},$$

where $\boldsymbol{\Sigma}^*$ is the 2×2 symmetric matrix, whose diagonal elements are σ_1^{*2} and σ_2^{*2}, and off-diagonal element is $\sigma_1^*\sigma_2^*\rho^*$. Make transformations

$$\begin{aligned}
\boldsymbol{T} &= \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{pmatrix} \boldsymbol{S} \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{pmatrix} = \begin{pmatrix} \frac{S_{11}}{\sigma_1^2} & \frac{S_{12}}{\sigma_1\sigma_2} \\ \frac{S_{12}}{\sigma_1\sigma_2} & \frac{S_{22}}{\sigma_2^2} \end{pmatrix}, \\
\boldsymbol{\Omega} &= \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{pmatrix} \boldsymbol{\Sigma}^* \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{pmatrix} = \begin{pmatrix} \omega_1^2 & \omega_1\omega_2\rho^* \\ \omega_1\omega_2\rho^* & \omega_2^2 \end{pmatrix}.
\end{aligned}$$

Clearly $trace(\boldsymbol{S}\boldsymbol{\Sigma}^{*-1}) = trace(\boldsymbol{T}\boldsymbol{\Omega}^{-1})$, and the domain of $\mathcal{D}$ becomes

$$\widetilde{\mathcal{D}} = \{(\omega_1, \omega_2, \rho^*) : \omega_1^{d_1}\omega_2^{d_2} g(\rho^*) < g(\rho)\}.$$

Then

$$G(\boldsymbol{X}, \sigma_1, \sigma_2, \rho) = \frac{\iiint_{\widetilde{\mathcal{D}}} \frac{h(\rho^*)}{\omega_1^{n-1+c_1}\omega_2^{n-1+c_2}(1-\rho^{*2})^{(n-1)/2}} f(trace(\boldsymbol{T}\boldsymbol{\Omega}^{-1})) d\omega_1 d\omega_2 d\rho^*}{\iiint \frac{h(\rho^*)}{\omega_1^{n-1+c_1}\omega_2^{n-1+c_2}(1-\rho^{*2})^{(n-1)/2}} f(trace(\boldsymbol{T}\boldsymbol{\Omega}^{-1})) d\omega_1 d\omega_2 d\rho^*}.$$

Since the sampling distribution of $\boldsymbol{T}$ depends only on ρ, so does the sampling distribution of $G(\boldsymbol{X}, \sigma_1, \sigma_2, \rho)$. Also $\widetilde{\mathcal{D}}$ depends on ρ only. The result thus holds. □

Proof of Theorem 4. Part (a) follows from Fact 2 (b) directly. It follows from (28)-(30) that

$$\begin{aligned}
\frac{r}{\sqrt{1-r^2}} &= \frac{s_{12}/\sqrt{s_{11}}}{\sqrt{s_{22}(1-r^2)}} = \frac{\sigma_2\sqrt{1-\rho^2}Z_3 + \frac{\rho\sigma_2}{\sigma_1}\sqrt{s_{11}}}{\sigma_2\sqrt{1-\rho^2}\sqrt{\chi^2_{n-2}}} \\
&= \frac{\sigma_2\sqrt{1-\rho^2}Z_3 + \rho\sigma_2\sqrt{\chi^2_{n-1}}}{\sigma_2\sqrt{1-\rho^2}\sqrt{\chi^2_{n-2}}} = \frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}. \qquad (103)
\end{aligned}$$

Part (b) then follows immediately, as does part (c). □

Proof of Theorem 5. It follows from (22) and Lemma 3 (a) that

$$P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) = P\Big\{\Big[\psi\Big(-\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-a}}} + \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{\chi^{2*}_{n-a}}}\frac{r}{\sqrt{1-r^2}}\Big)\Big]_{1-\alpha} > \rho \;\Big|\; \rho\Big\},$$

which implies (43). It follows from Lemma 3 (b) that

$$\begin{aligned}
P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) &= P\Big(\Big(-\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-a}}} + \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{\chi^{2*}_{n-a}}}\frac{r}{\sqrt{1-r^2}} - \frac{\rho}{\sqrt{1-\rho^2}}\Big)_{1-\alpha} > 0 \;\Big|\; \rho\Big) \\
&= P\Big(\Big(-\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^{2*}_{n-a}}}{\sqrt{\chi^{2*}_{n-b}}}\Big)_{1-\alpha} + \frac{r}{\sqrt{1-r^2}} > 0 \;\Big|\; \rho\Big).
\end{aligned}$$

Using (103), it follows that

$$P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) = P\Big(\frac{Z_3}{\sqrt{\chi^2_{n-1}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} < \Big(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^{2*}_{n-a}}}{\sqrt{\chi^{2*}_{n-b}}}\Big)_{1-\alpha} \;\Big|\; \rho\Big).$$

This completes the proof of Part (a). For part (b), since (45) holds for any $-1 < \rho < 1$, choose $\rho = 0$. Then

$$P\Big(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} < \Big(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\Big)_{1-\alpha}\Big) = 1-\alpha,$$

which implies that $b = 2$. Substituting $b = 2$ into (45) shows that $a = 1$. □

Proof of Theorem 6. Note that

$$\frac{\rho^*\sigma_2^*}{\sigma_1^*} = -\frac{\eta_3^*}{\eta_2^*} = \frac{s_{12}}{s_{11}} - \frac{Z_3^*}{\sqrt{s_{11}}}\frac{\sqrt{|\boldsymbol{S}|}}{\sqrt{\chi^{2*}_{n-b}}\sqrt{s_{11}}}.$$

Part (a) then follows. For part (b), note that

$$\begin{aligned}
P(\theta_1 < (\theta_1^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) &= P\Big(\frac{\rho\sigma_2}{\sigma_1} < \Big(\frac{s_{12}}{s_{11}} - \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\frac{\sqrt{|\boldsymbol{S}|}}{s_{11}}\Big)_{1-\alpha} \;\Big|\; \boldsymbol{\xi}, \rho\Big) \\
&= P\Big(\Big(\frac{\rho\sigma_2}{\sigma_1} - \frac{s_{12}}{s_{11}}\Big)\frac{s_{11}}{\sqrt{|\boldsymbol{S}|}} < \Big(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\Big)_{1-\alpha} \;\Big|\; \boldsymbol{\xi}, \rho\Big).
\end{aligned}$$

Using (28) and (29), it follows that

$$P(\theta_1 < (\theta_1^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \;=\; P\left(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} < \left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\right)_{1-\alpha}\right).$$

Part (b) then follows. □

Proof of Theorem 9. Part (a) follows from Fact 2 (a) and (b) directly. From (29) and (30), we have

$$(\theta_4 < (\theta_4^*)_{1-\alpha}) \;=\; \left(\frac{(1-\rho^2)\chi^2_{n-2}}{\chi^2_{n-1}}\left[\frac{\chi^{2*}_{n-a}}{\chi^{2*}_{n-b}} + \left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{r}{\sqrt{1-r^2}}\right)^2\right]_{1-\alpha} > 1\right).$$

Using (103) establishes part (b). For (c), note that $(\theta_4 < (\theta_4^*)_{1-\alpha}) = (G > 1)$, where

$$G \;=\; \rho^2 + \frac{\chi^2_{n-2}}{\chi^2_{n-1}}\left[(1-\rho^2)\left(\frac{\chi^{2*}_{n-a}}{\chi^{2*}_{n-b}} + G_2^2\right) + \; 2\rho\sqrt{1-\rho^2}G_2\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}\right]_{1-\alpha}.$$

Thus

$$P(G > 1 \mid \rho) \;=\; P\left(\frac{\chi^2_{n-2}}{\chi^2_{n-1}}\left[\sqrt{1-\rho^2}\left(\frac{\chi^{2*}_{n-a}}{\chi^{2*}_{n-b}} + G_2^2\right) + \; 2\rho G_2\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}\right]_{1-\alpha} > \sqrt{1-\rho^2} \mid \rho\right).$$

As $|\rho| \to 1$,

$$P(G > 1 \mid \rho) \;\to\; P\left(\left(2\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}G_2\right)_{1-\alpha} > 0\right),$$

which is the same as the right hand side of (56). It is also equal to

$$P\left(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} < \left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\right)_{1-\alpha}\right) = P\left(t_{n-2} < \frac{\sqrt{n-2}}{\sqrt{n-b}}(t^*_{n-b})_{1-\alpha}\right),$$

which is $1-\alpha$ if and only if $b = 2$. □

Proof of Theorem 10. Because $\sigma_{12} = -\eta_3/(\eta_1^2\eta_2)$, it follows from (17)-(19) that

$$\sigma_{12}^* \;=\; -\frac{\eta_3^*}{(\eta_1^*)^2\eta_2^*} \;=\; -\frac{\frac{Z_3^*}{\sqrt{s_{11}}} - \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{s_{11}}}\frac{r}{\sqrt{1-r^2}}}{\frac{\chi^{2*}_{n-a}}{s_{11}}\frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{s_{22}(1-r^2)}}}.$$

Part (a) follows. For part (b), using (28)-(30) and (103),

$$P(\sigma_{12} < (\sigma_{12}^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho)$$

$$
\begin{aligned}
&= P\bigg(\rho\sigma_1\sigma_2 < \bigg[\bigg(\frac{-Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} + \frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}\bigg)\frac{\sqrt{\chi^2_{n-1}\sigma_1^2\chi^2_{n-2}\sigma_2^2(1-\rho^2)}}{\chi^{2*}_{n-a}}\bigg]_{1-\alpha} \,\Big|\, \boldsymbol{\xi}, \rho\bigg) \\
&= P\bigg(0 < \bigg(- \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} + \frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} - \frac{\rho}{\sqrt{1-\rho^2}}\frac{\chi^{2*}_{n-a}}{\sqrt{\chi^2_{n-1}\chi^2_{n-2}}}\bigg)_{1-\alpha} \,\Big|\, \boldsymbol{\xi}, \rho\bigg) \\
&= P\bigg(- \frac{Z_3}{\sqrt{\chi^2_{n-2}}} - \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} < \bigg(- \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{\rho}{\sqrt{1-\rho^2}}\frac{\chi^{2*}_{n-a}}{\sqrt{\chi^2_{n-1}}\sqrt{\chi^2_{n-2}}}\bigg)_{1-\alpha} \,\Big|\, \rho\bigg).
\end{aligned}
$$

Using the facts that Z_3^* and $-Z_3^*$ have the same distribution, Z_3 and $-Z_3$ have the same distribution, and the independence of $(Z_3^*, \chi^{2*}_{n-a}, \chi^{2*}_{n-b}, Z_3, \chi^2_{n-1}, \chi^2_{n-2})$, part (b) then follows. $\square$

Proof of Theorem 11. It follows from the expression for the random posterior for η_3^* in (19), and from (30) and (103), that

$$
\begin{aligned}
P(\eta_3 < (\eta_3^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho) &= P\bigg(- \frac{\rho}{\sigma_1\sqrt{1-\rho^2}} < \bigg(\frac{Z_3^*}{\sqrt{s_{11}}} - \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{s_{11}}}\frac{r}{\sqrt{1-r^2}}\bigg)_{1-\alpha} \Big| \,\boldsymbol{\xi}, \rho\bigg) \\
&= P\bigg(- \frac{\rho}{\sigma_1\sqrt{1-\rho^2}} < \bigg[\frac{Z_3^*}{\sqrt{\sigma_1^2\chi^2_{n-1}}} - \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{\sigma_1^2\chi^2_{n-1}}}\bigg(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}\bigg)\bigg]_{1-\alpha} \Big| \,\rho\bigg) \\
&= P\bigg(0 < \bigg[\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \bigg(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}\bigg) + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^{2*}_{n-b}}}\bigg]_{1-\alpha} \Big| \,\rho\bigg) \\
&= P\bigg(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} \;<\; \bigg(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^{2*}_{n-b}}}\bigg)_{1-\alpha} \Big| \,\rho\bigg),
\end{aligned}
$$

so that (59) holds.

For part (b), if (59) equals $1-\alpha$ for any $-1 < \rho < 1$, choose $\rho = 0$ and obtain

$$
P\bigg(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} < \bigg(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\bigg)_{1-\alpha} \Big| \,\rho\bigg) = 1-\alpha.
$$

Thus $b = 2$. On the other hand, note that $Z_3^*, \chi^{2*}_{n-b}, Z_3, \chi^2_{n-1}, \chi^2_{n-2}$ are independent of each other. If $b = 2$, for any $-1 < \rho < 1$, the conditional probability of (59) given χ^2_{n-1} should be $1-\alpha$. The result then follows by taking the expectation with respect to χ^2_{n-1}. $\square$

Proof of Theorem 12. Part (a) is obvious. For part (b), since $\overline{x}_1 = \mu_1 + Z_1\sigma_1/\sqrt{n}$ and Z_1 and χ^2_{n-1} are independent, we have

$$
\begin{aligned}
(\theta_5 < (\theta_5^*)_{1-\alpha}) &= \bigg(\frac{\mu_1}{\sigma_1} < \bigg(\frac{Z_1^*}{\sqrt{n}} + \frac{\mu_1 + \sigma_1 Z_1/\sqrt{n}}{\sigma_1\sqrt{\chi^2_{n-1}\chi^{2*}_{n-a}}}\bigg)_{1-\alpha}\bigg) \\
&= \bigg(\bigg[\frac{Z_1^*}{\sqrt{n}} + \frac{\mu_1}{\sigma_1}\bigg(\sqrt{\frac{\chi^{2*}_{n-a}}{\chi^2_{n-1}}} - 1\bigg) + \frac{Z_1}{\sqrt{n}}\sqrt{\frac{\chi^{2*}_{n-a}}{\chi^2_{n-1}}}\bigg]_{1-\alpha} > 0\bigg)
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}} - \theta_5 \left(\frac{\sqrt{n}}{\sqrt{\chi^2_{n-1}}} - \frac{\sqrt{n}}{\sqrt{\chi^{2*}_{n-a}}} \right)_{1-\alpha} > \frac{Z_1}{\sqrt{\chi^2_{n-1}}} \right) \\
&= \left(\frac{Z_1}{\sqrt{\chi^2_{n-1}}} - \theta_5 \frac{\sqrt{n}}{\sqrt{\chi^2_{n-1}}} < \left(\frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}} - \theta_5 \frac{\sqrt{n}}{\sqrt{\chi^{2*}_{n-a}}} \right)_{1-\alpha} \right). \qquad (104)
\end{aligned}
$$

This proves part (b). For part (c), choosing $\theta_5 = 0$, we have

$$
P\left(\frac{Z_1}{\sqrt{\chi^2_{n-1}}} < \left(\frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}} \right)_{1-\alpha} \right) = 1 - \alpha,
$$

which implies that $a = 1$. On the other hand, if $a = 1$, for any fixed μ_1/σ_1,

$$
\frac{Z_1}{\sqrt{\chi^2_{n-1}}} - \frac{\mu_1}{\sigma_1} \frac{\sqrt{n}}{\sqrt{\chi^2_{n-1}}} \text{ and } \frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}} - \frac{\mu_1}{\sigma_1} \frac{\sqrt{n}}{\sqrt{\chi^{2*}_{n-a}}}
$$

have the same distribution. It follows from (104) that the coverage probabilities are exactly $1 - \alpha$. □

Proof of Theorem 13. Using the expression of σ_2^* given in (21), and since $(\mu_2 \mid \sigma_2, \boldsymbol{X}) \sim N(\overline{x}_2, \sigma_2^2/n)$, we could write the random posterior of μ_2 in terms of the random posterior of σ_2,

$$
\mu_2^* = \overline{x}_2 + \frac{\sigma_2^* Z_3^*}{\sqrt{n}}, \qquad (105)
$$

where Z_3^* is $N(0,1)$, independent of σ_2^*, given by the expression (20). Part (a) follows. For part (b), since $\overline{x}_2 = \mu_2 + \sigma_2 Z_2/\sqrt{n}$, we have

$$
\theta_6^* = \frac{Z_2^*}{\sqrt{n}} + \left(\frac{\mu_2}{\sigma_2} + \frac{Z_2}{\sqrt{n}} \right) \sqrt{\frac{\chi^{2*}_{n-a}}{\chi^2_{n-2}(1-\rho^2)}} \left[\frac{\chi^{2*}_{n-a}}{\chi^{2*}_{n-b}} + \left(G_2 + \frac{\rho}{\sqrt{1-\rho^2}} \sqrt{\frac{\chi^2_{n-1}}{\chi^2_{n-2}}} \right)^2 \right]^{-\frac{1}{2}}.
$$

This implies (64). As $|\rho| \to 1$,

$$
\lim_{|\rho| \to 1} P(\theta_6 < (\theta_6^*)_{1-\alpha} \mid \theta_6, \rho) = P\left(\theta_6 < \left[\frac{Z_2^*}{\sqrt{n}} + \left(\theta_6 + \frac{Z_2}{\sqrt{n}} \right) \frac{\chi^{2*}_{n-a}}{\chi^2_{n-1}} \right]_{1-\alpha} \Bigm| \theta_6 \right).
$$

An argument similar to that leading to (104) yields (65). □

Proof of Theorem 14. First note that

$$
\theta_7 = \frac{1}{\eta_1^2} + \frac{\eta_1^2 + \eta_3^2}{\eta_1^2 \eta_2^2} + \frac{2\eta_3}{\eta_1^2 \eta_2} = \frac{1}{\eta_2^2} + \frac{1}{\eta_1^2} \left(1 + \frac{\eta_3}{\eta_2} \right)^2.
$$

Substituting (17)–(19) into this expression yields (66). Rewrite θ_7^* as

$$
\theta_7^* = \frac{\sigma_2^2(1-\rho^2)\chi^2_{n-2}}{\chi^{2*}_{n-b}} + \frac{\sigma_1^2 \chi^2_{n-1}}{\chi^{2*}_{n-a}} \left[1 + \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-a}}} \frac{\sqrt{\sigma_2^2(1-\rho^2)\chi^2_{n-2}}}{\sqrt{\sigma_1^2 \chi^2_{n-1}}} - \sqrt{\frac{\sigma_2^2(1-\rho^2)}{\sigma_1^2 \chi^2_{n-1}}} Z_3 - \frac{r\sigma_2}{\sigma_1} \right]^2
$$

$$
\begin{aligned}
&= \frac{\sigma_2^2(1-\rho^2)\chi^2_{n-2}}{\chi^{2*}_{n-b}} + \frac{\chi^2_{n-1}}{\chi^{2*}_{n-a}}\left[(\sigma_1-\rho\sigma_2) - \sqrt{\sigma_2^2(1-\rho^2)}\left(\frac{Z_3}{\sqrt{\chi^2_{n-1}}} - \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\frac{\sqrt{\chi^2_{n-2}}}{\sqrt{\chi^2_{n-1}}}\right)\right]^2 \\
&= \sigma_2^2(1-\rho^2)\left\{\frac{\chi^2_{n-2}}{\chi^{2*}_{n-b}} + \frac{\chi^2_{n-1}}{\chi^{2*}_{n-a}}\left[\frac{\sigma_1-\rho\sigma_2}{\sqrt{\sigma_2^2(1-\rho^2)}} - \left(\frac{Z_3}{\sqrt{\chi^2_{n-1}}} - \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\frac{\sqrt{\chi^2_{n-2}}}{\sqrt{\chi^2_{n-1}}}\right)\right]^2\right\}.
\end{aligned}
$$

Because

$$
\begin{aligned}
\frac{\sigma_1-\rho\sigma_2}{\sqrt{\sigma_2^2(1-\rho^2)}} &= \frac{\tau}{\sqrt{1-\tau^2}}, \\
\theta_7 = \sigma_2^2(1-\rho^2)\left[1+\left(\frac{\sigma_1-\rho\sigma_2}{\sqrt{\sigma_2^2(1-\rho^2)}}\right)^2\right] &= \frac{\sigma_2^2(1-\rho^2)}{1-\tau^2},
\end{aligned}
$$

it follows that

$$
\begin{aligned}
&(\theta_7 < (\theta_7^*)_{1-\alpha}) \\
&= \left(\frac{1}{1-\tau^2} < \left\{\frac{\chi^2_{n-2}}{\chi^{2*}_{n-b}} + \frac{\chi^2_{n-1}}{\chi^{2*}_{n-a}}\left[\frac{\tau}{\sqrt{1-\tau^2}} - \left(\frac{Z_3}{\sqrt{\chi^2_{n-1}}} - \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\sqrt{\frac{\chi^2_{n-2}}{\chi^2_{n-1}}}\right)\right]^2\right\}_{1-\alpha}\right) \\
&= \left[1 < (G_4)_{1-\alpha}\right].
\end{aligned}
$$

Part (b) follows. For Part (c),

$$
\lim_{|\tau|\to 1} P(\theta_7 < (\theta_7^*)_{1-\alpha} \mid \tau) = P\left(1 < \left(\frac{\chi^2_{n-1}}{\chi^{2*}_{n-a}}\right)_{1-\alpha}\right).
$$

This completes the proof. □

References

Bayarri, M. J. (1981), 'Inferencia bayesiana sobre el coeficiente de correlacin de una poblacin normal bivariante', *Trabajos de Estadistica e Investigacion Operativa* **32**, 18–31.

Bayarri, M. J. & Berger, J. (2004), 'The interplay between bayesian and frequentist analysis', *Statistical Science* **19**, 58–80.

Berger, J. O. & Bernardo, J. M. (1992*a*), On the development of reference priors (with discussion), *in* 'Bayesian Statistics 4', Oxford Univ. Press: London, pp. 35–60.

Berger, J. O. & Bernardo, J. M. (1992*b*), Reference priors in a variance components problem, *in* 'Bayesian analysis in statistics and econometrics', pp. 177–194.

Berger, J. O., Strawderman, W. & Tang, D. (2005), 'Posterior propriety and admissibility of hyperpriors in normal hierarchical models', *Annals of Statistics* **33**, 606–646.

Berger, J. O. & Yang, R. (1994), 'Noninformative priors and Bayesian testing for the AR(1) model', *Econometric Theory* **10**, 461–482.

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

Brillinger, D. R. (1962), 'Examples bearing on the definition of fiducial probability with a bibliography', *Annals of Mathematical Statistics* **33**, 1349–1355.

Brown, P. J. (2001), The generalized inverted wishart distribution, *in* 'Encylopedia of Environmetrics'.

Brown, P., Le, N. & Zidek, J. (1994), Inference for a covariance matrix, *in* 'Aspects of Uncertainty (P.R. Freeman and A.F.M. Smith, eds.)', pp. 77–90.

Chang, T. & Eaves, D. (1990), 'Reference priors for the orbit in a group model', *Annals of Statistics* **18**, 1595–1614.

Daniels, M. & Kass, R. (1999), 'Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models', *Journal of the American Statistical Association* **94**, 1254–1263.

Daniels, M. & Pourahmadi, M. (2002), 'Bayesian analysis of covariance matrices and dynamic models for longitudinal data', *Biometrika* **89**, 553–566.

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

Dawid, A. P., Stone, M. & Zidek, J. V. (1973), 'Marginalization paradoxes in Bayesian and structural inference (with discussion)', *Journal of the Royal Statistical Society, Series B* **35**, 189–233.

Eaton, M. L. & Sudderth, W. (2002), 'Group invariant inference and right haar measure', *Journal of Statistical Planning and Inference* **103**, 87–99.

Fisher, R. A. (1930), 'Inverse probability', *Proceedings of the Cambridge Philosophical Society* **26**, 528–535.

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

Geisser, S. & Cornfield, J. (1963), 'Posterior distributions for multivariate normal parameters', *J. Roy. Statist. Soc. B* **25**, 368–376.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, second edn, John Wiley & Sons Ltd., New York. Wiley Series in Probability and Mathematical Statistics.

Liechty, J., Liechty, M. & Müller, P. (2004), 'Bayesian correlation estimation', *Biometrika* **91**, 1–14.

Lindley, D. V. (1961), The use of prior probability distributions in statistical inference and decisions, pp. 453–468.

Lindley, D. V. (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Cambridge University Press: Cambridge.

Rao, C. R. (1973), *Linear statistical inference and its applications*, John Wiley & Sons.

Robert, C. & Casella, G. (2004), *Monte Carlo Statistical Methods*, Springer-Verlag Inc, New York.

Roverato, A. & Consonni, G. (2004), 'Compatible prior distributions for dag models', *Journal of the Royal Statistical Society, Series B, Methodological* **66**, 47–61.

Stone, M. & Dawid, A. P. (1972), 'Un-Bayesian implications of improper Bayes inference in routine statistical problems', *Biometrika* **59**, 369–375.

Yang, R. & Berger, J. O. (1994), 'Estimation of a covariance matrix using the reference prior', *Annals of Statistics* **22**, 1195–1211.

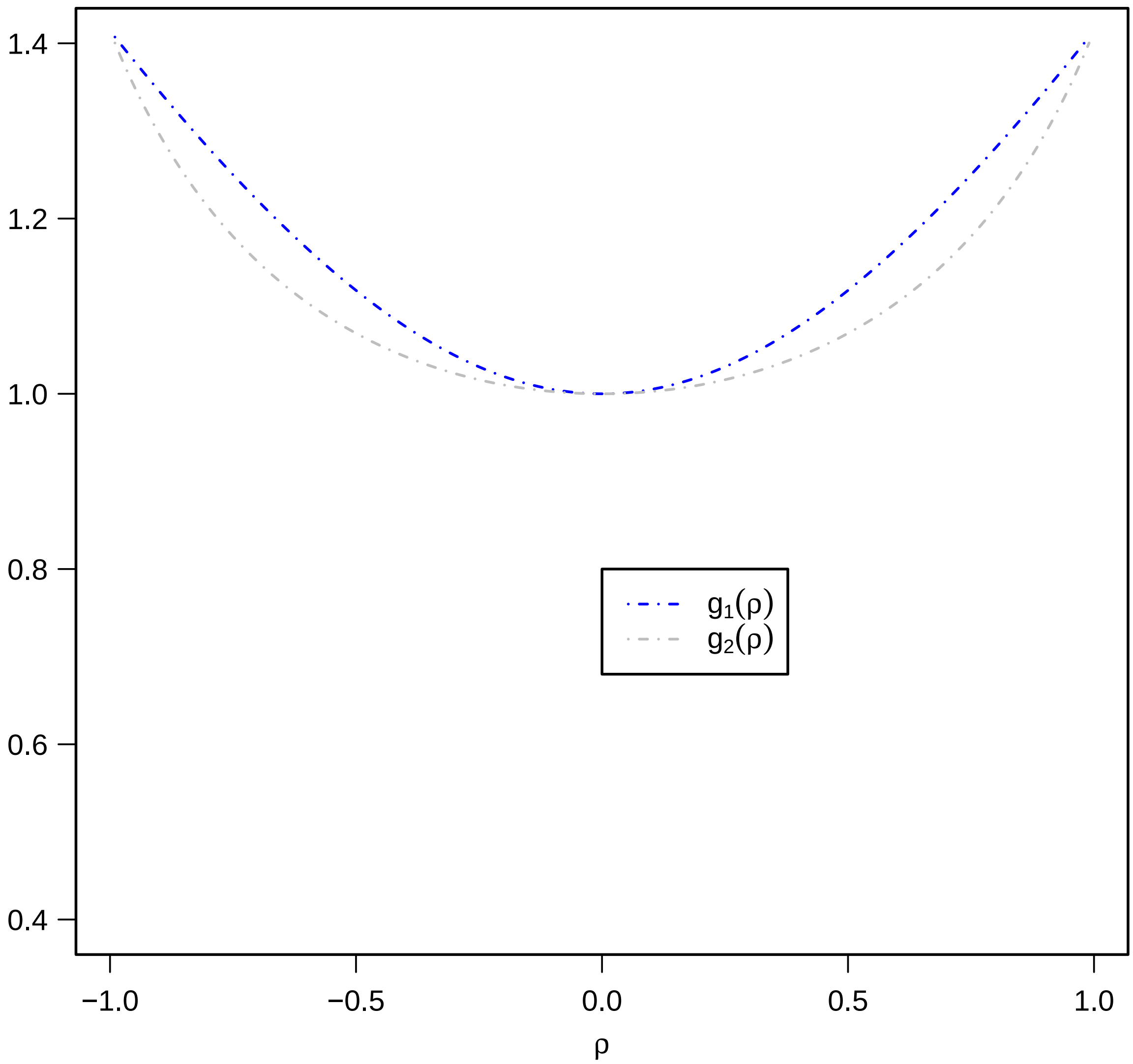

Figure 1: Comparison of the different factors in $\pi_{R\sigma}$ and $\tilde{\pi}_{R\sigma}$: $g_1(\rho) = \sqrt{1+\rho^2}$ and $g_2(\rho) = \sqrt{2}/\sqrt{2-\rho^2}$.

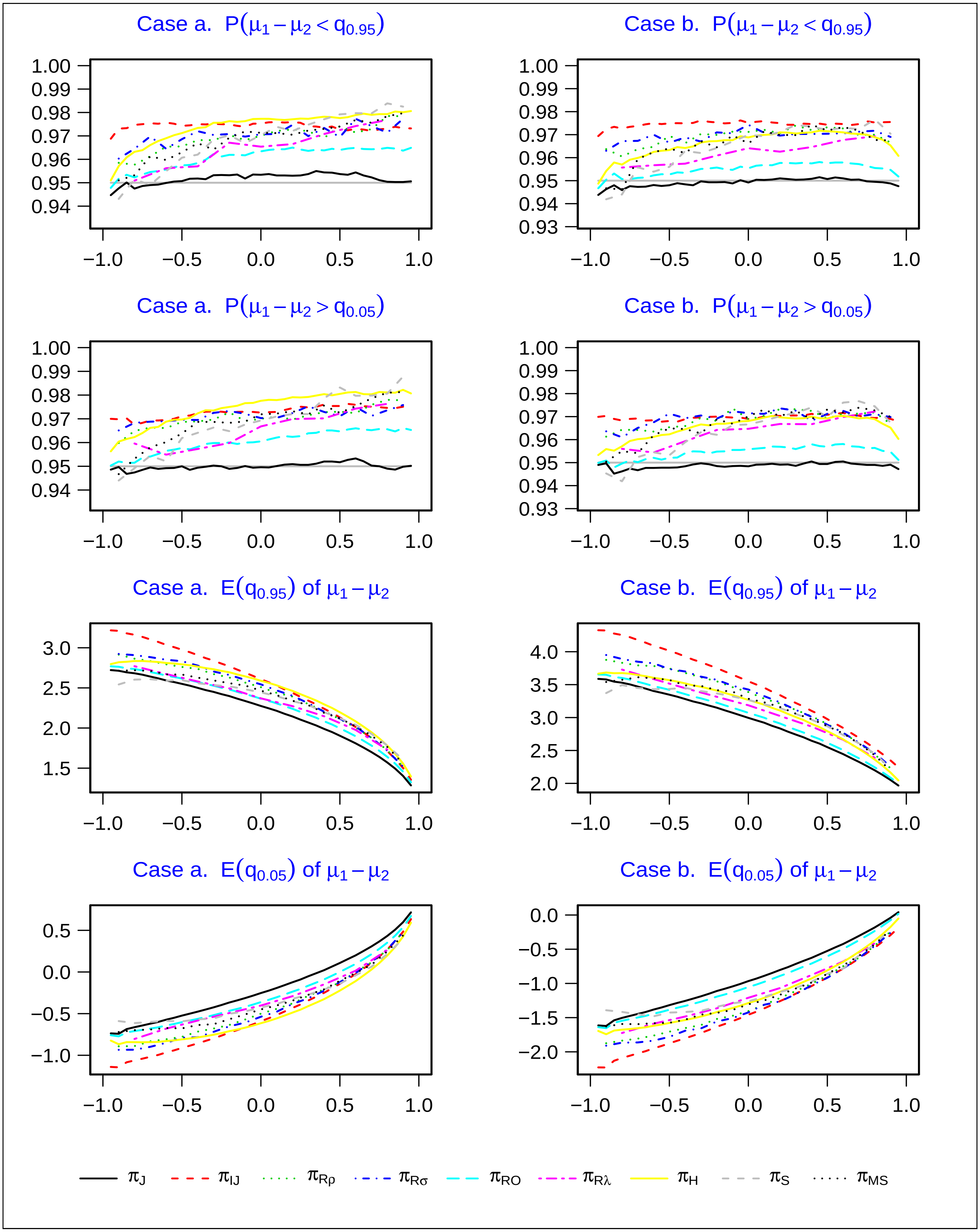

Figure 2: Frequentist coverages and expected posterior quantiles for $\mu_1 - \mu_2$, where Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. The x-axis is for $\rho \in (-1, 1)$.

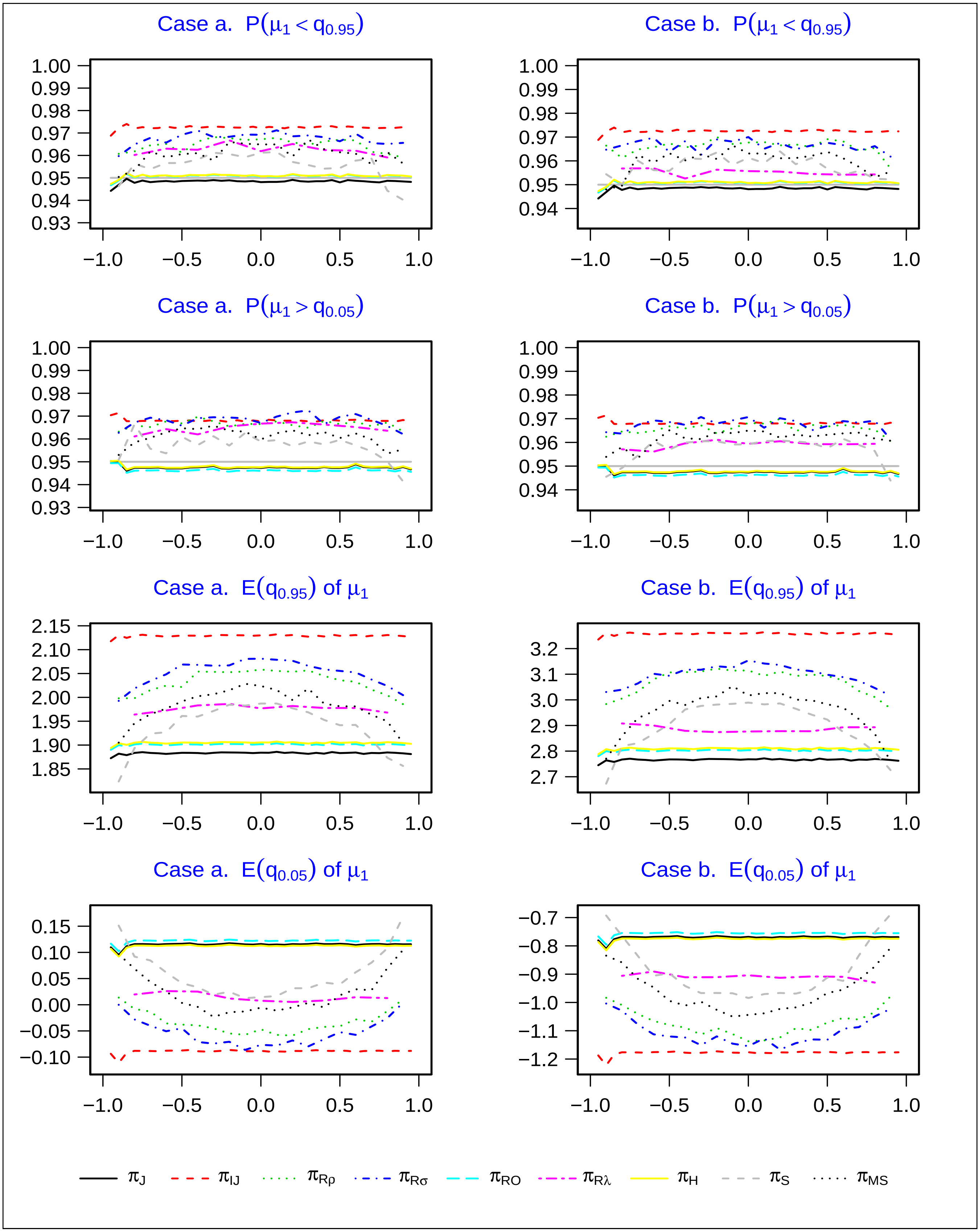

Figure 3: Frequentist coverages and expected posterior quantiles for μ_1, where Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. The x-axis is for $\rho \in (-1, 1)$.

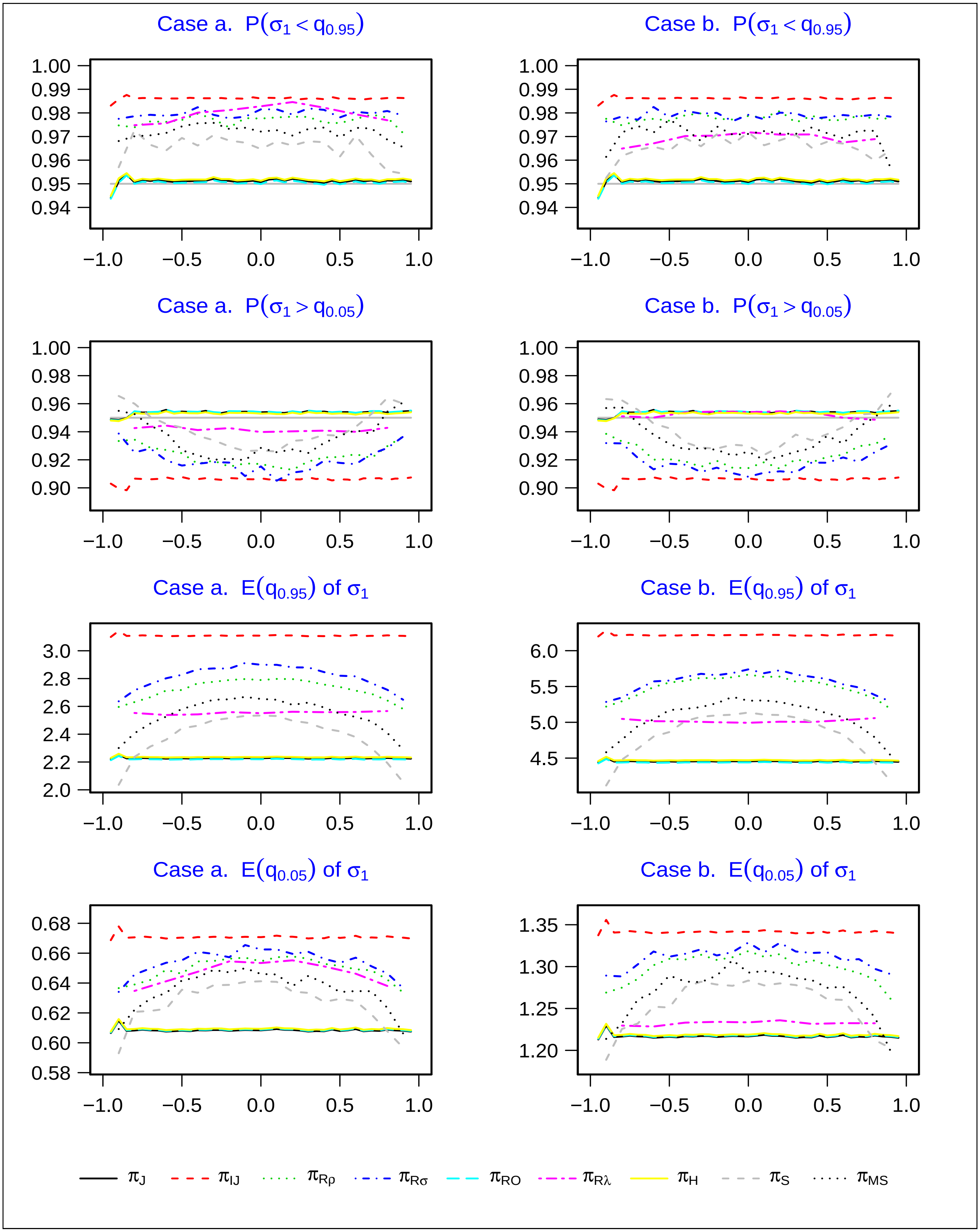

Figure 4: Frequentist coverages and expected posterior quantiles for σ_1, where Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. The x-axis is for $\rho \in (-1, 1)$.

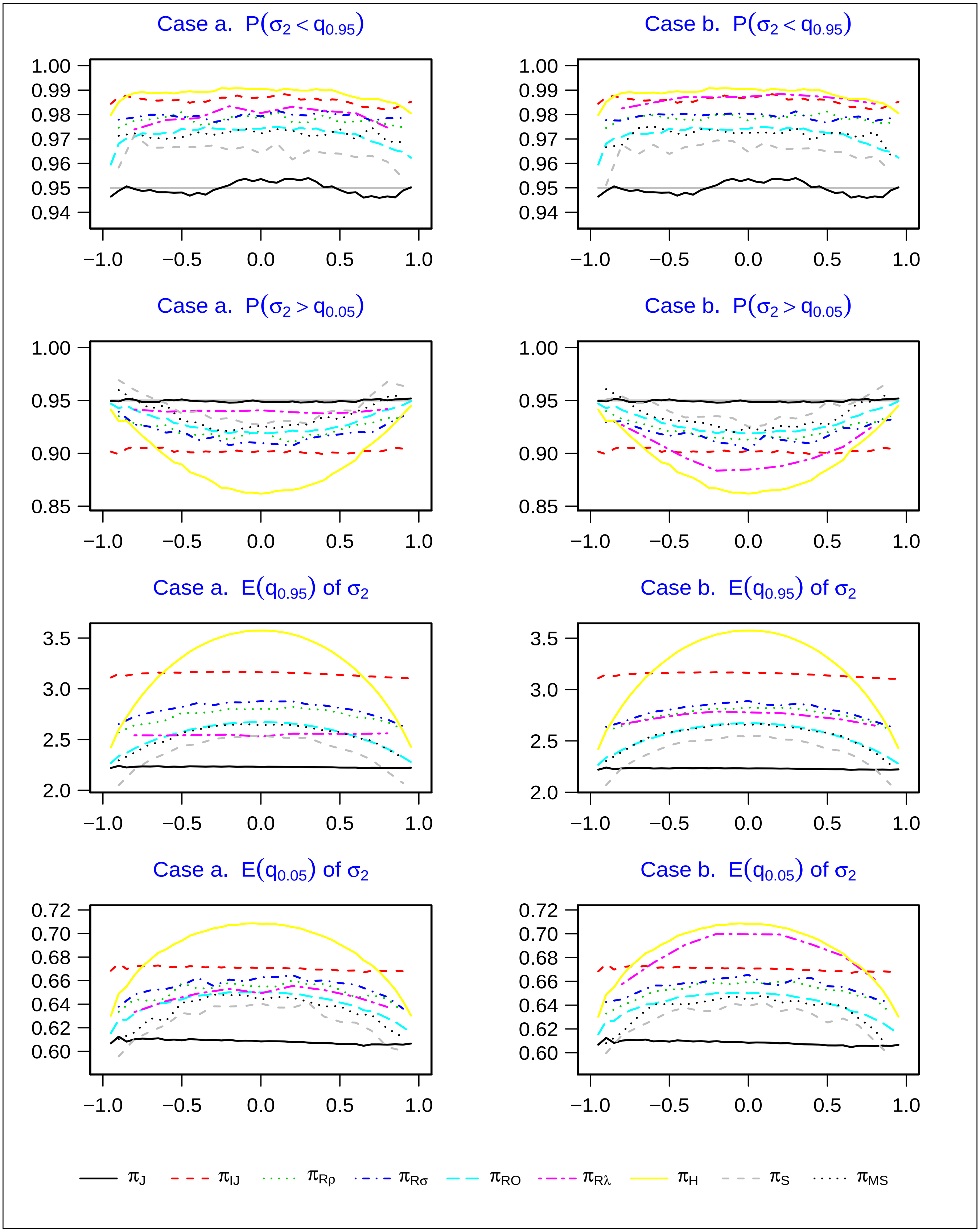

Figure 5: Frequentist coverages and expected posterior quantiles for σ_2, where Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. The x-axis is for $\rho \in (-1, 1)$.

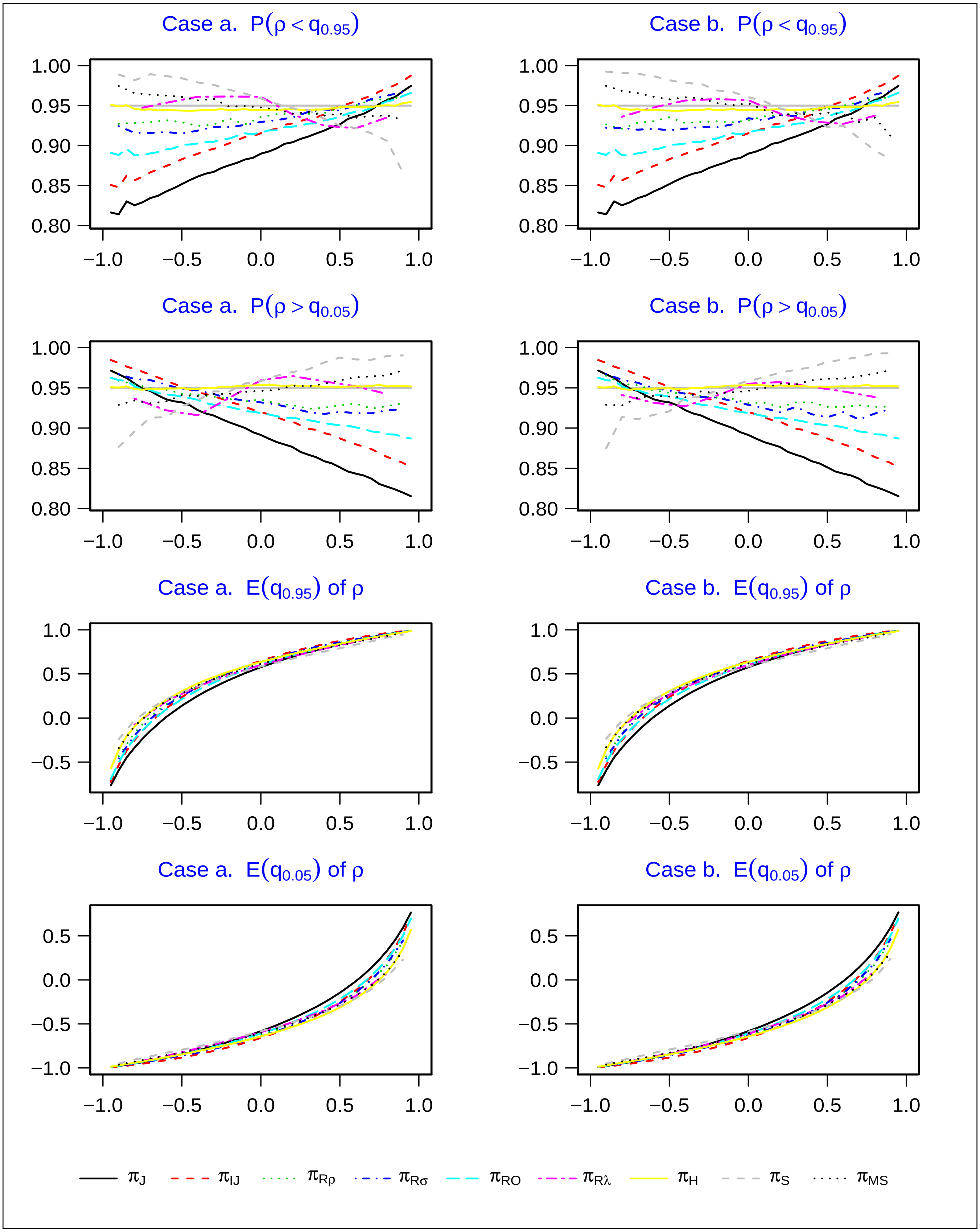

Figure 6: Frequentist coverages and expected posterior quantiles for ρ, where Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. The x-axis is for $\rho \in (-1, 1)$.

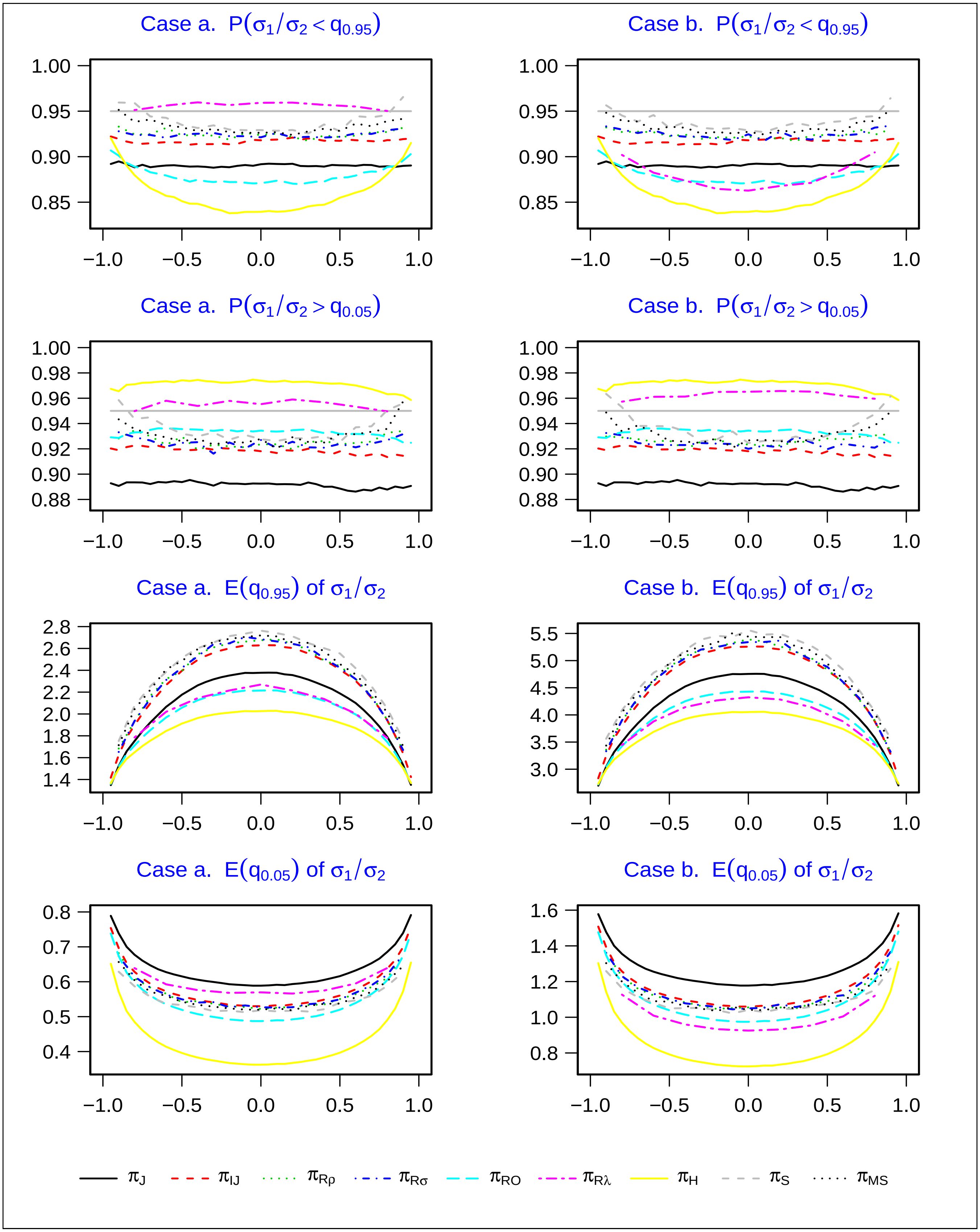

Figure 7: Frequentist coverages and expected posterior quantiles for σ_1/σ_2, where Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. The x-axis is for $\rho \in (-1, 1)$.

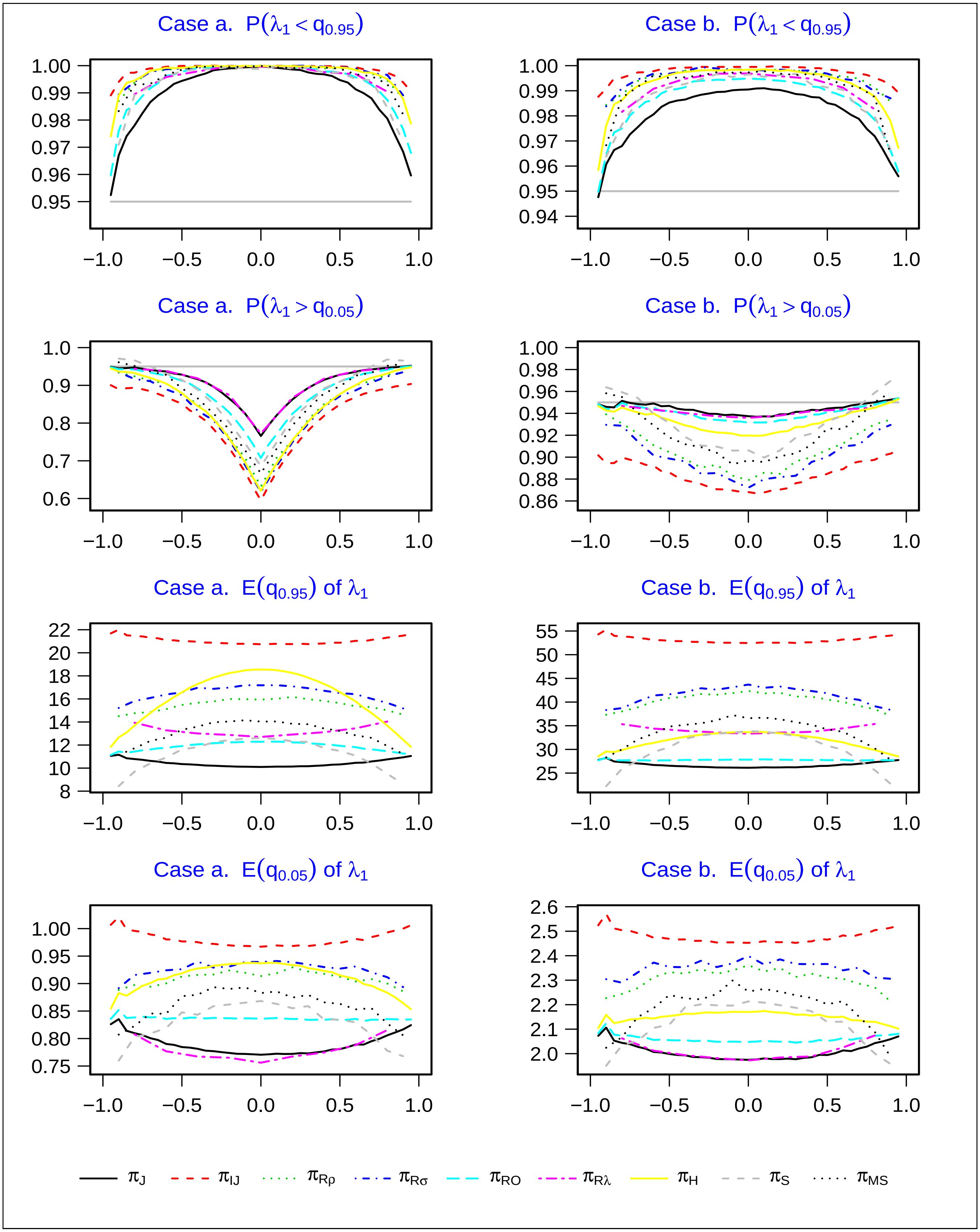

Figure 8: Frequentist coverages and expected posterior quantiles for λ_1, where Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. The x-axis is for $\rho \in (-1, 1)$.

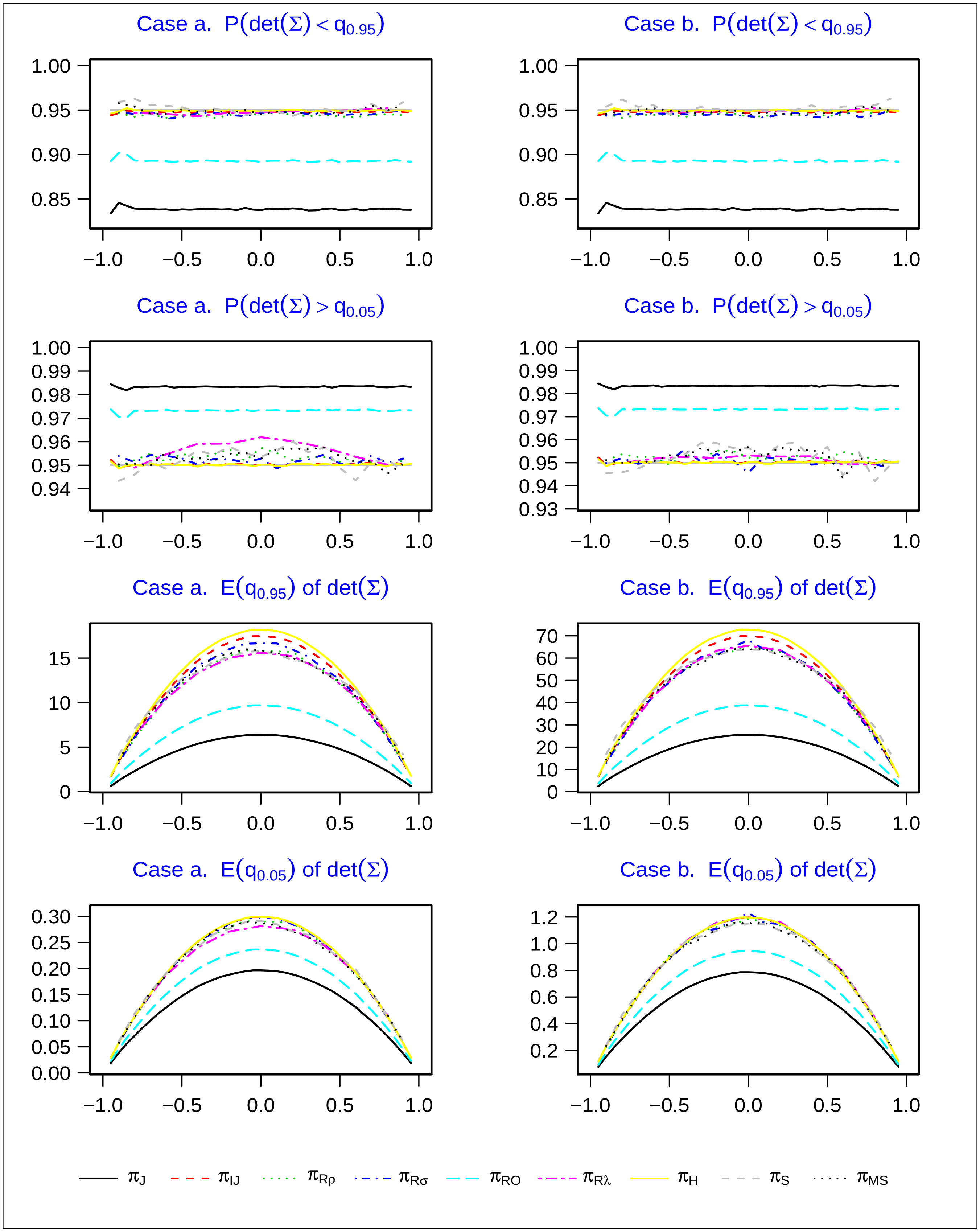

Figure 9: Frequentist coverages and expected posterior quantiles for $|\Sigma|$, where Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. The x-axis is for $\rho \in (-1, 1)$.

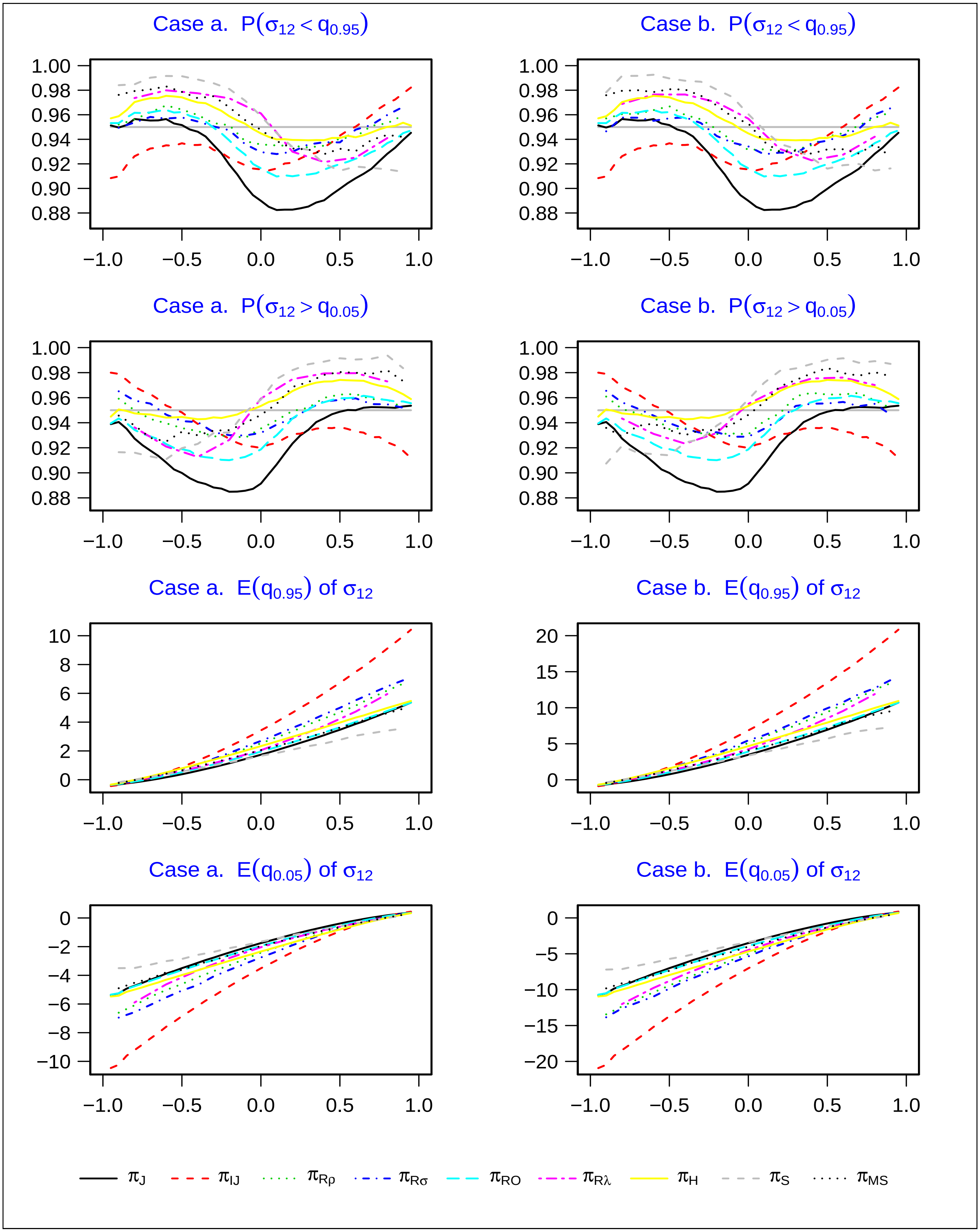

Figure 10: Frequentist coverages and expected posterior quantiles for σ_{12}, where Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. The x-axis is for $\rho \in (-1, 1)$.