Spring 2015

Small sample UMPU equivalence testing based on saddlepoint approximations

Renren Zhao

SMALL SAMPLE UMPU EQUIVALENCE TESTING BASED ON SADDLEPOINT APPROXIMATIONS

by

RENREN ZHAO

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

MATHEMATICS AND STATISTICS

2015

Approved by:

Dr. Robert Paige, Advisor
Dr. V.A. Samaranayake
Dr. Xuerong Wen
Dr. Gayla Olbricht
Dr. Gelles Gregory

ABSTRACT

In the first section, we consider small sample equivalence tests for exponentiality. Statistical inference in this setting is particularly challenging since equivalence testing procedures typically require a much larger sample size, in comparison to classical "difference tests", to perform well. We make use of Butler's marginal likelihood for the shape parameter of a gamma distribution in our development of equivalence tests for exponentiality. We consider two procedures using the principle of confidence interval inclusion, four Bayesian methods, and the uniformly most powerful unbiased (UMPU) test where a saddlepoint approximation to the intractable distribution of a canonical sufficient statistic is used. We perform simulation studies to assess the bias of various tests and show that all of the Bayes' posteriors we consider are integrable. Our simulation studies show that the saddlepoint-approximated UMPU method performs remarkably well for small sample sizes and is the only method which consistently exhibits an empirical significance level close to the nominal five percent rate.

In the second section, we consider small sample equivalence tests for mean-to-variance ratio from two normal populations. In general, optimal equivalence tests for the means of two homoskedastic normal populations do not exist unless the common population variance is known. However, we show that if one considers the mean-to-variance ratio then there does exist a uniformly most powerful unbiased (UMPU) equivalence testing procedure. Furthermore, our procedure involves an intractable conditional distribution which we reproduce to a high degree of accuracy using saddlepoint approximations. We also develop six competing equivalence testing procedures for the mean-to-variance ratio. Four of these procedures are Bayesian and the remaining two are based upon the principle of confidence interval inclusion. Small sample simulation studies show that our UMPU method outperforms all competing methods by exhibiting an empirical significance level which is not statistically significantly different from the nominal five percent rate, for all simulation settings.

ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Prof. Robert Paige for his support and guidance through my dissertation, as well as my graduate committee. I appreciate all the help and adivice given to me by the faculty members of the Department of Mathematics and Statistics, Missouri University of Science and Technology. Finally, I would like to thank my parents, for their continued support and encouragement. It is no doubt that I could not finish this effort without their assistance.

DEDICATION

I would like to dedicate this Doctoral dissertation to my parents, Yue Zhao and Xinguang Liu. Without their continued support and counsel, I could not have completed this process.

TABLE OF CONTENTS

Page

LIST OF TABLES

1. SMALL SAMPLE EQUIVALENCE TESTS FOR EXPONENTIALITY

With a test for exponentiality one would like to provide evidence that data comes from a distribution which is at least close to exponential. Existing tests for exponentiality are designed to provide evidence that data comes from a distribution which is not exponential, and lack of such evidence from these tests is usually interpreted as meaning that it is fine to assume the data follows an exponential distribution; see for instance Henze and Meintanis (2005). Wellek (2010; sec. 1.2) points out that interpreting a nonsignificant p-value as evidence in support of the null hypothesis generally fails to yield a valid test procedure. This idea can be formulated as the truism "absence of evidence is not evidence of absence"; see for instance Altman and Bland (1995).

With this in mind, it is worthwhile to consider "goodness of fit" rather than the traditional "lack of fit" tests for exponentiality. Wellek (2010; sec. 1.2) makes the point that a bonafide goodness of fit test should be formulated as an equivalence test where the alternative hypothesis states that the data are consistent with the distribution of interest modulo a minor difference which he refers to as tolerable difference.

We develop seven small sample (goodness of fit) equivalence tests for exponentiality of three different types. These types correspond to the three general small sample approaches for constructing an equivalence test; see Wellek (2010; ch. 3). The first approach relies upon the principle of confidence interval inclusion and involves the construction of a $100\,(1-2\alpha)\,\%$ confidence interval, where α is the nominal significance level for the test. The second approach is Bayesian in nature and the third approach involves the construction of a uniformly most powerful unbiased (UMPU) test for equivalence.

All of the approaches we develop depend upon Butler's marginal likelihood for the shape parameter in a gamma distribution (Butler, 2007, sec. 5.4.4) and two of these make use of saddlepoint approximations which are remarkably accurate in approximating nonnormal distributions (Butler 2007). In particular, the UMPU equivalence test, which we will discuss next, makes use of the Luganani and Rice (1980) saddlepoint approximation to the cumulative distribution function (CDF) of a canonical sufficient statistic is used to obtain an approximate UMPU test possessing a significance level that is consistently close to the nominal 5% level.

1.1. SADDLEPOINT-APPROXIMATED UMPU EQUIVALENCE TEST

The likelihood for a random sample $X_1, \ldots, X_n$ from gamma distribution with shape parameter $\theta > 0$ and rate parameter $\lambda > 0$ is

$$\mathcal{L}(\theta, \lambda) \propto \exp\left\{-\lambda \sum x_i + (\theta - 1) \sum \ln(x_i) + n\left[\theta \ln \lambda - \ln \Gamma(\theta)\right]\right\}.$$

This corresponds to a regular exponential family with canonical sufficient statistics $T_1 = \sum \ln(x_i)$ and $T_2 = \sum x_i$ and canonical parameters $-\lambda$ and θ, respectively. In this setting an equivalence test for exponentiality can be formulated in terms of the following null and alternative hypotheses:

$$H_0 : \theta \leq \theta_1 \text{ or } \theta \geq \theta_2 \quad \text{and} \quad H_a : \theta_1 < \theta < \theta_2 \tag{1.1}$$

where $\theta_1 = 1 - \varepsilon_1$ and $\theta_2 = 1 + \varepsilon_2$ for tolerable deviations $\varepsilon_1, \varepsilon_2 > 0$.

The optimal UMPU test for the above hypotheses is constructed from the conditional distribution of T_1 given the observed value of T_2; see Lehmann (1986, sec. 4.4) and Wellek (2010; sec. 3.3). However, Butler (2007, sec. 5.4.4) notes that

$$\begin{aligned} P(T_1 \leq t_1 | T_2 = t_2, \theta) &= P(T_1 - n\ln(T_2) \leq t_1 - n\ln(t_2) | T_2 = t_2, \theta) \\ &= P(T \leq t | T_2 = t_2, \theta) \\ &= P(T \leq t | \theta) \end{aligned}$$

where $T = \sum \ln(x_i / \sum x_i)$ and the fact that T is independent of T_2 is used. As a result, an UMPU equivalence test for exponentiality can be constructed from the unconditional distribution of T. For $n > 1$, Butler (2007, sec. 5.4.4) develops a marginal likelihood for θ of the form

$$\mathcal{L}_M(\theta) \propto \exp\left\{(\theta-1)T + \ln\Gamma(n\theta) - n\ln\Gamma(\theta)\right\} \tag{1.2}$$

which is also the likelihood for a regular exponential family with canonical sufficient statistic T. Therefore, the level α UMPU equivalence test for exponentiality based upon Butler's marginal likelihood has a rejection region of the form:

$$C_1 < T < C_2$$

where

$$P(C_1 < T < C_2 | \theta = \theta_i) = \alpha \tag{1.3}$$

for $i = 1$ and 2 (Lehmann, 1986, sec. 3.7). The determination of cut-off values C_1 and C_2 is hindered by the intractable distribution of T. Fortunately, Butler's marginal likelihood also provides a closed-form expression for the cumulant generating function (CGF) of T:

$$K_T(s) = n\ln\left\{\Gamma(s+\theta)\backslash\Gamma(\theta)\right\} - \ln\left\{\Gamma\left[n(s+\theta)\backslash\Gamma(n\theta)\right]\right\}.$$

This transform, in turn, provides easy access to highly accurate saddlepoint approximations to distribution of T. In particular, the Luganani and Rice (1980) saddlepoint approximation to the CDF of T is given as

$$\hat{P}(T \leq t;\theta) = \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})\left[\hat{w}^{-1} - \hat{u}^{-1}\right], & \text{if } t \neq E(T) \\ \frac{1}{2} + K_T^{(3)}(0)\left[72\pi K_T^{(2)}(0)^3\right]^{-1/2}, & \text{if } t = E(T) \end{cases} \tag{1.4}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal CDF and PDF functions respectively, $K_T^{(i)}(s)$ is the ith derivative of this CGF for $i = 1,2,3$, $\hat{w} = sgn(\hat{s})\sqrt{2\left[\hat{s}t - K_T(\hat{s})\right]}$,

$\hat{u} = \hat{s}\sqrt{K_T^{(2)}(\hat{s})}$ and saddlepoint $\hat{s}$ is the solution to saddlepoint equation $K_T^{(1)}(\hat{s}) = t$. This saddlepoint approximation is used to determine approximate (C_1, C_2) values.

1.2. BAYESIAN EQUIVALENCE TESTS FOR EXPONENTIALITY

Butler's marginal likelihood is also the starting point for our Bayesian exponentiality tests. Wellek (2010; sec. 3.2) considers the nominal level α Bayesian equivalence test for which

$$P(\theta_1 < \theta < \theta_2 | x_1, \ldots, x_n) \geq 1 - \alpha \tag{1.5}$$

leads to the rejection of the nonequivalence null in (1.1) as well as the double one-sided Bayesian test where this condition is replaced with

$$P(\theta_1 < \theta | x_1, \ldots, x_n) \geq 1 - \alpha \quad \text{and} \quad P(\theta < \theta_2 | x_1, \ldots, x_n) \geq 1 - \alpha. \tag{1.6}$$

With each type of Bayesian test we considered two prior distributions on θ; a flat prior, $\pi(\theta) \propto 1$, (Box and Tiao, 1973), and the objective Jeffreys' prior (Berger, 1985);

$$\pi(\theta) \propto \sqrt{I(\theta)} = \sqrt{n\psi'(\theta) - n^2\psi'(n\theta)}$$

where $I(\theta)$ denotes the expected Fisher information for θ and $\psi'(\theta)$ is the trigamma function which is defined as the second derivative of the log-gamma function. These prior distributions were chosen in hopes that they would have a minimal impact on the posterior distribution. It is shown in the appendix that both yield proper posterior distributions for all $n > 1$.

1.3. PRINCIPLE OF CONFIDENCE INTERVAL INCLUSION

We also consider two methods which make use of the principle of confidence interval inclusion; see Wellek (2010; sec. 3.1). This principle is equivalent to the intersection-union test principle applied to equivalence null hypotheses; see Berger (1982). For methods of this type, a $(1-2\alpha)\,100\%$ confidence interval $\left(\hat{\theta}_L, \hat{\theta}_U\right)$ for θ of is generated. The nominal level α test based on the confidence interval inclusion principle then rejects nonequivalence null in (1.1) if $\left(\hat{\theta}_L, \hat{\theta}_U\right)$ is contained in (θ_1, θ_2), the region corresponding to the equivalence alternative hypothesis in (1.1). We consider two confidence interval methods and take $\alpha = 0.05$ for the sake of concreteness.

1.3.1. Large Sample Confidence Interval. First the 90% classical marginal likelihood-based confidence of the form

$$\hat{\theta} \pm 1.645 \left[I\left(\hat{\theta}\right)\right]^{-1/2} \tag{1.7}$$

where marginal maximum likelihood estimates (MMLE) $\hat{\theta}$ is the maximizer of the marginal likelihood for θ.

1.3.2. Pivotal Confidence Interval. We also consider a 90% pivotal confidence interval where the pivotal quantity is the CDF of canonical sufficient statistic T in (1.2). For a further discussion of the pivotal CDF see Berger and Casella (2002, sec. 9.2.3). Here, we determine a confidence interval for θ through the solution of the following equations:

$$P\left(T \leq t; \hat{\theta}_L\right) = 0.95 \quad \text{and} \quad P\left(T \leq t; \hat{\theta}_U\right) = 0.05. \tag{1.8}$$

This confidence interval has exact coverage under the assumptions of that the family of approximated CDFs $\{P\left(T \leq t; \theta\right)\}$ is stochastically decreasing in θ (Berger and Casella 2002, sec. 9.2.3). In practice, however, we use the saddlepoint approximation in (1.4) in place of the true but intractable CDF $P\left(T \leq t; \theta\right)$. Pivotal CDF confidence intervals often yield lengths and coverage probabilities that compare favorably with those from basically any competing method; see Paige and Trindade (2008), and Paige, Trindade and Fernando (2009).

1.4. MONTE CARLO STUDIES

In our simulation studies we took $n = 10, 20$ and 30, let tolerable differences $\varepsilon_1 = \varepsilon_2 = \varepsilon = 0.1, 0.2, 0.3, 0.4$ and 0.5, and set θ_{true}, the true value of θ, to be $\theta_1 = 1 - \varepsilon$ or $\theta_2 = 1 + \varepsilon$. Note that in Table 1.1 $\theta_{true} = 1 - \varepsilon$ and $\theta_{true} = 1 + \varepsilon$ are represented as "$-\varepsilon$" and "ε", respectively. For each combination of n, ε and θ_{true} values we simulated 100,000 data sets from a gamma distribution with shape parameter θ_{true} and rate parameter $\lambda = 1$. Note that this choice of λ was completely general since all of the equivalence testing methods we consider originate from Butler's marginal likelihood for θ, and its canonical sufficient statistic $T = \sum \ln (x_i / \sum x_i)$ has a distribution which is invariant under scalar transformations of the data. Table 1.1 presents the empirical significance levels for the saddlepoint-approximated optimal UMPU (O) procedure; the four Bayesian procedures, (1.5) with a flat prior (F1), (1.6) also with a flat prior (F2), (1.5) with the Jeffreys' prior (J1) and (1.6) using a Jeffreys' prior (J2); the saddlepoint-based CDF pivot method (CP) in (1.8) and the classical marginal likelihood method in (1.7). Here, empirical significance levels for which the associated 95% Wald confidence interval for proportions contains 0.05 are shown in bold.

We see that the saddlepoint-approximated optimal UMPU (O) procedure is remarkably accurate in terms of significance level even for very small sample sizes. In fact, it is only when $n = 30$ and very wide tolerable differences that any of the competing methods are close to being unbiased. The poor performance of the confidence interval methods is likely due to their wideness for small samples. The poor performance for the Bayesian methods is probably due to the inability of the likelihood, with so little data, to minimize the impact of the prior on the posterior.

1.5. CONCLUSIONS

We developed seven small sample equivalence tests for exponentiality from Butler's marginal likelihood for the shape parameter in a gamma distribution. We considered at

Table 1.1. Empirical significance levels of equivalence tests for exponentiality

Empirical significance levels for $n = 10, 20$ and 30								
n	ε	O	F1	F2	J1	J2	CP	ML
10	−0.1	**4.99**	0.00	0.00	0.00	0.00	0.00	0.00
10	0.1	**4.99**	0.00	0.00	0.00	0.00	0.00	0.00
10	−0.2	**5.09**	0.00	0.00	0.00	0.00	0.00	0.00
10	0.2	5.15	0.00	0.00	0.00	0.00	0.00	0.00
10	−0.3	**4.98**	0.00	0.00	0.00	0.00	0.00	0.00
10	0.3	**4.96**	0.00	0.00	0.00	0.00	0.00	0.00
10	−0.4	**5.05**	0.00	0.00	0.00	0.00	0.00	0.00
10	0.4	**4.87**	0.00	0.00	0.00	0.00	0.00	0.00
10	−0.5	**5.04**	0.00	0.00	0.00	0.00	0.00	0.00
10	0.5	**5.01**	0.00	0.00	0.00	0.00	0.00	0.00
20	−0.1	**5.01**	0.00	0.00	0.00	0.00	0.00	0.00
20	0.1	**5.04**	0.00	0.00	0.00	0.00	0.00	0.00
20	−0.2	**4.93**	0.00	0.00	0.00	0.00	0.00	0.00
20	0.2	**5.00**	0.00	0.00	0.00	0.00	0.00	0.00
20	−0.3	5.14	0.00	0.00	0.00	0.00	0.00	0.00
20	0.3	**4.95**	0.00	0.00	0.00	0.00	0.00	0.00
20	−0.4	**4.97**	0.00	0.00	0.00	0.00	0.00	0.00
20	0.4	**4.94**	0.00	0.00	0.00	0.00	0.00	0.00
20	−0.5	5.15	4.49	6.57	0.00	3.34	3.30	1.25
20	0.5	4.84	1.34	2.39	0.00	3.92	3.90	3.43
30	−0.1	**5.05**	0.00	0.00	0.00	0.00	0.00	0.00
30	0.1	**5.02**	0.00	0.00	0.00	0.00	0.00	0.00
30	−0.2	**5.01**	0.00	0.00	0.00	0.00	0.00	0.00
30	0.2	**5.07**	0.00	0.00	0.00	0.00	0.00	0.00
30	−0.3	**5.03**	0.00	0.00	0.00	0.00	0.00	0.00
30	0.3	**4.92**	0.00	0.00	0.00	0.00	0.00	0.00
30	−0.4	**5.02**	0.00	4.73	0.00	2.51	2.49	1.10
30	0.4	**4.94**	0.00	2.32	0.00	3.08	3.06	2.44
30	−0.5	**5.13**	8.05	8.07	**4.92**	**4.95**	**4.92**	3.10
30	0.5	**4.95**	2.92	3.01	4.75	**4.93**	**4.92**	6.02

least one method from each of the three general small sample approaches for constructing an equivalence test. The saddlepoint-approximated optimal UMPU procedure was virtually unbiased in nearly all settings and is clearly superior to the six competing methods.

2. OPTIMAL EQUIVALENCE TESTING FOR NORMAL POPULATIONS

Tukey (1991) succinctly makes a strong argument for equivalence testing: "It is foolish to ask 'are the effects of A and B different?' They are always different-for some decimal place". Wellek (2010, sec 1.2) makes the argument that optimal equivalence tests are needed since equivalence testing requires much larger sample sizes than "difference" testing which is much more common. We consider the practically important problem of equivalence testing for two independent normal samples.

In general, optimal equivalence tests for the means of two homoskedastic normal populations do not exist unless the common population variance is known; see Romano (2005). If instead one considers standardized means then there exists a uniformly most powerful invariant (UMPI) procedure, as described in Wellek (2010, sec. 6.1). Here the non-equivalence null hypothesis is

$$H_0 : \psi \leq \psi_1 \text{ or } \psi \geq \psi_2$$

with associated equivalence alternative hypothesis

$$H_a : \psi_1 < \psi < \psi_2$$

where ψ is the distributional parameter;

$$\psi = \frac{\mu_1 - \mu_2}{\sigma}$$

and, $\psi_1 < 0$ and $\psi_2 > 0$ are constants which describe to within what tolerance will the standardized means be considered equivalent.

We show in section 2.1 that if one considers the mean-to-variance ratio, with distributional parameter

$$\psi = \frac{\mu_1 - \mu_2}{\sigma^2},$$

then there does in fact exist a uniformly most powerful unbiased (UMPU) equivalence testing procedure. This procedure involves UMPU testing theory for regular exponential families and, as is often the case for tests of this type, is based upon the intractable conditional distribution of one canonical sufficient statistic given the observed values of the others. In section 2.2 we reproduce this conditional distribution to a high degree of accuracy using Skovgaard's saddlepoint approximation to the conditional cumulative distribution function (CDF); see Butler (2007, sec. 5.4.5). The development of the resulting saddlepoint-based equivalence testing procedure involves a non-unique interest parameter preserving (IPP) reparametrization of the likelihood function. However, we show in section 2.1 that the underlying conditional exponential family for ψ is invariant under the choice of IPP transformation and in section 2.2 we show that our saddlepoint-based procedure is also invariant to the choice of IPP reparametrization. We also develop six competing equivalence testing procedures for the mean-to-variance ratio. The four Bayesian methods are discussed in section 2.3. Here we perform all of the required integrations in closed-form up to a univariate integral which is easy to approximate. We also establish the properness of our posterior distributions for the two testing paradigms we consider and the improper flat and Jeffreys' priors that we assume. The two remaining procedures discussed in section 2.4 are based upon the principle of confidence interval inclusion. Here the equivalence test is performed with a $(1-2\alpha)\,100\%$ confidence interval for ψ, where α is the nominal significance level for the equivalence test. In section 2.5 we consider simulation studies which show that our UMPU procedure outperforms all competing methods, for all simulation settings, by exhibiting an empirical significance level which does not differ significantly from the nominal 5% rate. Finally, we present concluding remarks in section **??**.

2.1. CHOICE OF EXPONENTIAL FAMILY

First we discuss the appropriate choice of exponential family structure for the equivalence tests we develop. We assume that we collect two independent random samples from a $N(\mu_1, \sigma^2)$ population and a $N(\mu_2, \sigma^2)$ population, respectively;

$$Y_{1,1}, \ldots, Y_{1,n_1} \sim \text{ i.i.d. } N\left(\mu_1, \sigma^2\right)$$

$$Y_{2,1}, \ldots, Y_{2,n_2} \sim \text{i.i.d. } N\left(\mu_2, \sigma^2\right)$$

where i.i.d. is the abbreviation for "independent and identically distributed". Recall that the likelihood function for a univariate normal random variable $W \sim N(\mu, \sigma^2)$ is

$$\begin{aligned}
\mathcal{L}\left(\mu, \sigma^2\right) &\propto \frac{1}{\sqrt{\sigma^2}} \exp\left\{-\frac{(w-\mu)^2}{2\sigma^2}\right\} \\
&= \exp\left\{-\frac{1}{2\sigma^2}w^2 + \frac{\mu}{\sigma^2}w - \frac{\mu^2}{2\sigma^2} - \frac{\ln\sigma^2}{2}\right\} \\
&= \exp\left\{\theta_0 w^2 + \theta_1 w + c\left(\theta_0, \theta_1\right)\right\}
\end{aligned}$$

where

$$c\left(\theta_0, \theta_1\right) = \frac{\theta_1^2}{4\theta_0} + \frac{1}{2}\ln\left(-2\theta_0\right)$$

and

$$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2\sigma^2} \\ \frac{\mu}{\sigma^2} \end{bmatrix}.$$

The joint likelihood for the two independent normal random samples can be written as

$$\mathcal{L}\left(\mu_1, \mu_2, \sigma^2\right) \propto \exp\left\{\theta_0 S_1 + \theta_1 S_2 + \theta_2 S_3 + n_1 c\left(\theta_0, \theta_1\right) + n_2 c\left(\theta_0, \theta_2\right)\right\} \tag{2.1}$$

where

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2\sigma^2} \\ \frac{\mu_1}{\sigma^2} \\ \frac{\mu_2}{\sigma^2} \end{bmatrix} \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} \sum y_1^2 + \sum y_2^2 \\ \sum y_1 \\ \sum y_2 \end{bmatrix}$$

are the canonical parameters of this likelihood and the associated canonical sufficient statistics.

That leads to the following expression for the likelihood function:

$$\mathcal{L}(\theta_0, \theta_1, \theta_2) \propto \exp\left\{\theta^T \mathbf{S} + n_1 c(\theta_0, \theta_1) + n_2 c(\theta_0, \theta_2)\right\}. \tag{2.2}$$

We are, however, primarily interested in making inference about

$$\psi = \frac{\mu_1}{\sigma^2} - \frac{\mu_2}{\sigma^2} = \theta_1 - \theta_2.$$

Since this parameter is a linear function of θ_1 and θ_2 we can in fact rewrite our likelihood function so that interest parameter ψ is a canonical parameter in the new likelihood function with the resulting nuisance parameters denoted as λ_1 and λ_2. In the process we will implicitly define a new set of canonical sufficient statistics which we shall denote as T_1, T_2 and T_3. One possible reparametrization can be obtained by adding and subtracting a $\theta_2 \sum y_1$ term in the likelihood function and then rearranging the resulting terms to yield

$$\mathcal{L}(\psi, \lambda_1, \lambda_2) \propto \exp\left\{\psi T_1 + \lambda_1 T_2 + \lambda_2 T_3 + c(\psi, \lambda_1, \lambda_2)\right\} \tag{2.3}$$

where

$$\begin{bmatrix} \psi \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \frac{\mu_1}{\sigma^2} - \frac{\mu_2}{\sigma^2} \\ -\frac{1}{2\sigma^2} \\ \frac{\mu_1}{\sigma^2} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = \begin{bmatrix} -\sum y_2 \\ \sum y_1^2 + \sum y_2^2 \\ \sum y_1 + \sum y_2 \end{bmatrix}$$

and

$$c(\psi, \lambda_1, \lambda_2) = n_1 c(\lambda_1, \lambda_2) + n_2 c(\lambda_1, \lambda_2 - \psi).$$

Note that we shall henceforth refer to a reparametrization from original likelihood (2.2) to a likelihood in which ψ is a canonical parameter as a primary reparametrization.

Furthermore, we can write likelihood (2.3) in matrix-vector form as

$$\mathcal{L}(\gamma) \propto \exp\left\{\gamma^T \mathbf{T} + c(\gamma)\right\} \tag{2.4}$$

where

$$\gamma = \begin{bmatrix} \psi \\ \lambda_1 \\ \lambda_2 \end{bmatrix} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}.$$

Optimal UMPU tests for ψ depend upon the conditional distribution of canonical sufficient statistic T_1 given the observed values of statistics T_2 and T_3;

$$f(t_1|t_2, t_3, \psi) \tag{2.5}$$

since it is known that this conditional distribution, which also has an exponential family form, only depends upon interest parameter ψ; see Lehmann (1986, sec. 4.4).

Note however that the likelihood in (2.4) is but one of an uncountably infinite number of likelihoods that can be gotten by reparametrizing the original likelihood (2.2) so as to make ψ a canonical parameter. Butler (2007, sec. 5.1) provides a general procedure for reparametrizing exponential family likelihoods. We use this procedure to develop a characterization of all interest parameter (ψ) preserving (IPP) reparametrizations.

Here after a choice of an appropriately chosen nonsingular matrix B we have for likelihood (2.4) that

$$\gamma^T \mathbf{T} = \gamma^T \mathbf{B}\mathbf{B}^{-1}\mathbf{T} = \left(\mathbf{B}^T\gamma\right)^T \left(\mathbf{B}^{-1}\mathbf{T}\right).$$

To preserve interest parameter ψ as a canonical parameter, this nonsingular B^T matrix must have the following form

$$\mathbf{B}^T = \begin{bmatrix} 1 & 0 & 0 \\ a & b & c \\ d & e & f \end{bmatrix}.$$

Note that the collection of all nonsingular matrices of this type;

$$M = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ a & b & c \\ d & e & f \end{bmatrix} : a, b, c, d, e, f \in \mathbb{R} \text{ and } bf \neq ce \right\}$$

form a group under matrix multiplication. Also, the reparametrizations determined by the elements of M shall henceforth be referred to as secondary or IPP reparametrizations.

Note also that the reparametrized likelihood in (2.4) is obtained from the original likelihood in (2.2) via a transformation determined by a nonsingular matrix A;

$$\theta^T \mathbf{S} = \theta^T \mathbf{A}\mathbf{A}^{-1}\mathbf{S} = \left(\mathbf{A}^T \theta\right)^T \left(\mathbf{A}^{-1}\mathbf{S}\right)$$

where

$$\mathbf{A}^T = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

It follows that any reparametrization from original likelihood (2.2) to a likelihood with a structure like (2.4) in which ψ is a canonical parameter can be generated from A^T by an appropriate choice of B^T matrix. To see this note that for any $B^T \in M$ we have that

$$\begin{bmatrix} 1 & 0 & 0 \\ a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 \\ A & B & C \\ D & E & F \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 \\ Ab+cD & a+Bb+cE & -a+Cb+Fc \\ fD+Ae & d+fE+Be & -d+Ff+Ce \end{bmatrix}.$$

Conversely, suppose you have a general primary reparametrization of the form

$$\mathbf{A}_g^T = \begin{bmatrix} 0 & 1 & -1 \\ A_1 & B_1 & C_1 \\ D_1 & E_1 & F_1 \end{bmatrix}$$

then the answer to the question regarding existence a unique secondary reparametrization which generates this from primary reparametrization

$$\mathbf{A}^T = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

is simply given as

$$\mathbf{B}^T = \begin{bmatrix} 0 & 1 & -1 \\ A_1 & B_1 & C_1 \\ D_1 & E_1 & F_1 \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -C_1 & A_1 & B_1+C_1 \\ -F_1 & D_1 & F_1+E_1 \end{bmatrix}.$$

It turns out that, without lack of generality, we can restrict ourselves to the original primary reparametrization from likelihoods (2.2) to (2.4) for the purpose of developing optimal UMPU tests for ψ. This is because the conditional exponential family for ψ, which is conditional distribution of T_1 given observed values for T_2 and T_3, is invariant under the the group of secondary reparametrizations in M as stated in the following theorem.

Theorem 2.1. *The conditional exponential family for ψ is invariant under the choice of IPP transformations in matrix group M or equivalently the choice of canonical sufficient statistics in a reparametrized likelihood for which ψ is a canonical parameter.*

Please see the appendices for the proof.

Next we consider saddlepoint approximations for $f(t_1|t_2, t_3; \psi)$ and the associated conditional CDF $F(t_1|t_2, t_3; \psi)$, and the application of the CDF approximation to UMPU equivalence testing.

2.2. SADDLEPOINT-APPROXIMATED UMPU EQUIVALENCE TEST

From classical UMPU testing theory for regular exponential families (Lehmann, 1986, sec. 4.4) the size α test for hypotheses

$$H_0 : \psi \leq \psi_1 \text{ or } \psi \geq \psi_2 \quad \text{versus} \quad H_a : \psi_1 < \psi < \psi_2 \tag{2.6}$$

rejects null hypothesis H_0 and finds statistical evidence of equivalence if

$$c_1 < t_1 < c_2$$

where cut-offs $c_1$and c_2 satisfy the following equations simultaneously

$$\begin{cases} P\left(c_1 < T_1 < c_2 | T_2 = t_2, T_3 = t_3, \psi = \psi_1\right) = \alpha \\ P\left(c_1 < T_1 < c_2 | T_2 = t_2, T_3 = t_3, \psi = \psi_2\right) = \alpha \end{cases}.$$

Note that (PDF) $f(t_1|t_2, t_3; \psi)$ and CDF $F(t_1|t_2, t_3; \psi)$ are intractable and cannot be evaluated in a closed form due to the intractable surface integral in the normalization constant for $f(t_1|t_2, t_3; \psi)$.

The saddlepoint approximation to $f(t_1|t_2, t_3; \psi)$ is given in Butler (2007, sec 5.4.2) as

$$\hat{f}(t_1|t_2, t_3; \psi) = (2\pi)^{-1/2} \left\{ \frac{|j\left(\hat{\psi}, \hat{\lambda}_1, \hat{\lambda}_2\right)|}{|j_{\lambda\lambda}\left(\psi, \hat{\lambda}_1(\psi), \hat{\lambda}_2(\psi)\right)|} \right\}^{-1/2} \frac{\mathcal{L}\left[\psi, \hat{\lambda}_1(\psi), \hat{\lambda}_2(\psi)\right]}{\mathcal{L}\left(\hat{\psi}, \hat{\lambda}_1, \hat{\lambda}_2\right)} \tag{2.7}$$

where $\hat{\psi}$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are the maximum likelihood estimates (MLEs) for their respective parameters, and $\hat{\lambda}_1(\psi)$ and $\hat{\lambda}_2(\psi)$ are the conditional or constrained MLEs of λ_1 and λ_2 for fixed ψ.

To numerically determine c_1 and c_2, we will use Skovgaard's approximation to conditional CDF $F(t_1|t_2,t_3;\psi)$ described in Butler (2007, sec 5.4.5) as

$$\begin{aligned} \hat{F}(t_1|t_2,t_3;\psi) &= \hat{P}(T_1 < t_1|T_2 = t_2, T_3 = t_3, \psi) \\ &= \Phi(w) + \phi(w)\left(\frac{1}{w} - \frac{1}{u}\right) \end{aligned} \tag{2.8}$$

where

$$w = sgn\left(\hat{\psi} - \psi\right)\sqrt{-2\ln\frac{\mathcal{L}\left(\psi, \hat{\lambda}_1(\psi), \hat{\lambda}_2(\psi)\right)}{\mathcal{L}\left(\hat{\psi}, \hat{\lambda}_1, \hat{\lambda}_2\right)}},$$

$$u = \left(\hat{\psi} - \psi\right)\sqrt{\frac{\left|j\left(\hat{\psi}, \hat{\lambda}_1, \hat{\lambda}_2\right)\right|}{\left|j_{\lambda\lambda}\left(\psi, \hat{\lambda}_1(\psi), \hat{\lambda}_2(\psi)\right)\right|}},$$

$\Phi(\cdot)$ and $\phi(\cdot)$ denote the PDF and CDF for a standard normal random variable, $\lambda = [\lambda_1, \lambda_2]$ and where $sgn(\cdot)$ denotes the sign function.

Next, we derive the likelihood quantities appearing in the above saddlepoint PDF and CDF approximations. Recall the joint log-likelihood function for our setting;

$$\mathcal{L}(\psi, \lambda_1, \lambda_2) \propto \exp\{\psi T_1 + \lambda_1 T_2 + \lambda_2 T_3 + c(\psi, \lambda_1, \lambda_2)\}.$$

To obtain MLEs $\hat{\psi}$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ we compute the score equations as follows

$$\begin{bmatrix} \frac{\partial \mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial \psi} = \frac{1}{2\lambda_1}(\psi n_2 + 2\lambda_1 T_1 - \lambda_2 n_2) \\ \frac{\partial \mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial \lambda_1} = T_2 + \frac{n_1+n_2}{2\lambda_1} - \frac{n_2(\psi-\lambda_2)^2 + n_1\lambda_2^2}{4\lambda_1^2} \\ \frac{\partial \mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial \lambda_2} = \frac{1}{2\lambda_1}(2\lambda_1 T_3 - \psi n_2 + \lambda_2 n_1 + \lambda_2 n_2) \end{bmatrix}.$$

Next we set each equation to zero and solve for the MLEs;

$$\begin{bmatrix} \hat{\psi} \\ \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{bmatrix} = \begin{bmatrix} \frac{\left(\frac{T_1+T_3}{n_1}+\frac{T_1}{n_2}\right)(n_1+n_2)}{T_2-\frac{(T_1+T_3)^2}{n_1}-\frac{T_1^2}{n_2}} \\ -\frac{n_1+n_2}{2\left[T_2-\frac{(T_1+T_3)^2}{n_1}-\frac{T_1^2}{n_2}\right]} \\ \frac{\frac{T_1+T_3}{n_1}(n_1+n_2)}{T_2-\frac{(T_1+T_3)^2}{n_1}-\frac{T_1^2}{n_2}} \end{bmatrix}.$$

Note that by the invariance property for MLEs we can easily obtain the above results using the MLEs for original likelihood (2.1) ;

$$\begin{aligned} \hat{\mu}_1 &= \bar{y}_1 \\ \hat{\mu}_2 &= \bar{y}_2 \\ \hat{\sigma}^2 &= \frac{\sum (y_1-\bar{y}_1)^2 + \sum (y_2-\bar{y}_2)^2}{n_1+n_2} \end{aligned}$$

and

$$\begin{bmatrix} \hat{\psi} \\ \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{bmatrix} = \begin{bmatrix} \frac{\hat{\mu}_1}{\hat{\sigma}^2}-\frac{\hat{\mu}_2}{\hat{\sigma}^2} \\ -\frac{1}{2\hat{\sigma}^2} \\ \frac{\hat{\mu}_1}{\hat{\sigma}^2} \end{bmatrix}$$

Next we need to derive conditional MLEs $\hat{\lambda}_1(\psi)$ and $\hat{\lambda}_2(\psi)$. These are obtained by solving the following set of equations in λ_1 and λ_2 for fixed ψ:

$$\begin{cases} \frac{\partial \mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial \lambda_1} = T_2 - \frac{n_1\lambda_2^2}{4\lambda_1^2} + \frac{n_1}{2\lambda_1} - \frac{n_2(\lambda_2-\psi)^2}{4\lambda_1^2} + \frac{n_2}{2\lambda_1} = 0 \\ \frac{\partial \mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial \lambda_2} = T_3 + \frac{n_1\lambda_2}{2\lambda_1} + \frac{n_2(\lambda_2-\psi)}{2\lambda_1} = 0 \end{cases}.$$

The solutions are

$$\begin{bmatrix} \hat{\lambda}_1(\psi) \\ \hat{\lambda}_2(\psi) \end{bmatrix} = \begin{bmatrix} \frac{-(n_1+n_2)^2-\sqrt{(n_1+n_2)^4+4n_1n_2\psi^2\left[(n_1+n_2)T_2-T_3^2\right]}}{4\left[(n_1+n_2)T_2-T_3^2\right]} \\ \frac{n_2\psi-2\hat{\lambda}_1(\psi)T_3}{n_1+n_2} \end{bmatrix}.$$

Here again details are provided in the appendices.

The Fisher information matrix and the partial information matrix for λ_1 and λ_2 are given as follows;

$$
\begin{aligned}
j(\psi,\lambda_1,\lambda_2) &= \begin{bmatrix} -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\psi^2} & -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\psi\partial\lambda_1} & -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\psi\partial\lambda_2} \\ -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_1\partial\psi} & -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_1^2} & -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_1\partial\lambda_2} \\ -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_2\partial\psi} & -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_2\partial\lambda_1} & -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_2^2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n_2}{2\lambda_1} & -\frac{n_2(\lambda_2-\psi)}{2\lambda_1^2} & \frac{n_2}{2\lambda_1} \\ -\frac{n_2(\lambda_2-\psi)}{2\lambda_1^2} & -\frac{n_1\lambda_2^2+n_2(\lambda_2-\psi)^2-\lambda_1(n_1+n_2)}{2\lambda_1^3} & -\frac{n_2\psi-n_1\lambda_2-n_2\lambda_2}{2\lambda_1^2} \\ \frac{n_2}{2\lambda_1} & -\frac{n_2\psi-n_1\lambda_2-n_2\lambda_2}{2\lambda_1^2} & -\frac{n_1+n_2}{2\lambda_1} \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
j_{\lambda\lambda}(\psi,\lambda_1,\lambda_2) &= \begin{bmatrix} -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_1^2} & -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_1\partial\lambda_2} \\ -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_2\partial\lambda_1} & -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_2^2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n_1\lambda_2^2+n_2(\lambda_2-\psi)^2-\lambda_1(n_1+n_2)}{2\lambda_1^3} & -\frac{n_2\psi-n_1\lambda_2-n_2\lambda_2}{2\lambda_1^2} \\ -\frac{n_2\psi-n_1\lambda_2-n_2\lambda_2}{2\lambda_1^2} & -\frac{n_1+n_2}{2\lambda_1} \end{bmatrix}
\end{aligned}
$$

Theorem 2.2. *The conditional saddlepoint approximations to PDF and CDF are invariant under the choice of IPP transformations in matrix group M or equivalently the choice of canonical sufficient statistics in a reparametrized likelihood for which ψ is a canonical parameter.*

Please see the appendices for the proof. However due to the reparametrization invariance described in theorems 1 and 2, it suffices to simply work with the primary reparametrization given in (2.4) for the purpose of developing saddlepoint-approximated UMPU equivalence test for ψ.

Wellek (2010, sec. 3.3) provides an algorithm for estimating cut-offs c_1 and c_2. However, we developed an alternative approach which we found easier to implement with saddlepoint approximations. Recall that we need to determine c_1 and c_2 that satisfy following two equations simultaneously:

$$\hat{F}_{\psi_1}(c_2) - \hat{F}_{\psi_1}(c_1) = \alpha \tag{2.9}$$

$$\hat{F}_{\psi_2}(c_2) - \hat{F}_{\psi_2}(c_1) = \alpha \tag{2.10}$$

where

$$\hat{F}_{\psi_1}(\cdot) = \hat{F}(\cdot|t_2, t_3; \psi_1)$$

and

$$\hat{F}_{\psi_2}(\cdot) = \hat{F}(\cdot|t_2, t_3; \psi_2).$$

From the first equation we can solve for c_2 as

$$c_2 = \hat{F}_{\psi_1}^{-1}\left[\alpha + \hat{F}_{\psi_1}(c_1)\right]. \tag{2.11}$$

Plugging this solution into the second equation yields

$$\hat{F}_{\psi_2}\left\{\hat{F}_{\psi_1}^{-1}\left[\alpha + \hat{F}_{\psi_1}(c_1)\right]\right\} - \hat{F}_{\psi_2}(c_1) = \alpha.$$

The solution to this equation can be recast as the root of the following function:

$$G(c_1) = \hat{F}_{\psi_2}\left\{\hat{F}_{\psi_1}^{-1}\left[\alpha + \hat{F}_{\psi_1}(c_1)\right]\right\} - \hat{F}_{\psi_2}(c_1) - \alpha$$

To find the root of $G(c_1)$ we perform a grid search followed by the bisection method to generate a $G(c_1)$ value of order 10^{-6}. We then determine c_2 from our estimate for c_1 via equation (2.11).

2.3. BAYESIAN EQUIVALENCE TESTS

Wellek (2010; sec. 3.2) considers two types of nominal α level Bayesian equivalence tests. The first type rejects the nonequivalence null in (2.6) when the posterior probability of the alternative hypothesis is sufficiently large

$$P(\psi_1 < \psi < \psi_2 | x_1, \ldots, x_n) \geq 1 - \alpha. \tag{2.12}$$

The second type is known as the double one-sided Bayesian test and it rejects the nonequivalence null when

$$P(\psi > \psi_1 | x_1, \ldots, x_n) \geq 1 - \alpha \quad \text{and} \quad P(\psi < \psi_2 | x_1, \ldots, x_n) \geq 1 - \alpha. \tag{2.13}$$

In our setting the sample data

$$[x_1, \ldots, x_n] \equiv [y_{1,1}, \ldots, y_{1,n_1}, y_{2,1}, \ldots, y_{2,n_2}].$$

For each type of Bayesian test we considered two different prior distributions on ψ; a flat prior, $\pi(\psi, \lambda_1, \lambda_2) = 1$, (Box and Tiao, 1973), and the objective Jeffreys' prior (Berger, 1985);

$$\begin{aligned}
\pi(\psi, \lambda_1, \lambda_2) &= \sqrt{[\det I(\psi, \lambda_1, \lambda_2)]} \\
&= \sqrt{[\det j(\psi, \lambda_1, \lambda_2)]} \\
&= \sqrt{\frac{(n_1^2 n_2 + n_1 n_2^2)}{8\lambda_1^4}}
\end{aligned}$$

These improper prior distributions were chosen in hopes that they would have a minimal impact on the posterior distribution. Moreover, it is shown in the appendix that both yield proper posterior distributions provided that $n_1 > 1$ or $n_2 > 1$.

For notational convenience, in the integrals which follow, we will often use γ, λ and $\mathbf{x}$ to denote $[\psi, \lambda_1, \lambda_2]$, $[\lambda_1, \lambda_2]$ and $[x_1, \ldots, x_n]$ respectively. The posterior distribution of γ is given as

$$f(\psi, \lambda_1, \lambda_2 | \mathbf{x}) = \frac{f(\psi, \lambda_1, \lambda_2, \mathbf{x})}{f(\mathbf{x})} = \frac{f(\mathbf{x} | \psi, \lambda_1, \lambda_2)\, \pi(\psi, \lambda_1, \lambda_2)}{\displaystyle\int_\gamma f(\mathbf{x} | \psi, \lambda_1, \lambda_2)\, \pi(\psi, \lambda_1, \lambda_2)\, d\gamma}$$

where

$$f(\mathbf{x}|\psi,\lambda_1,\lambda_2) \propto \exp\left\{\gamma^T\mathbf{T} + \frac{n_1\lambda_2^2}{4\lambda_1} + \frac{1}{2}(n_1+n_2)\ln(-2\lambda_1) + \frac{n_2(\lambda_2-\psi)^2}{4\lambda_1}\right\}.$$

To make inference about ψ one needs to integrate out nuisance parameters $\lambda = [\lambda_1, \lambda_2]$ to obtain the posterior distribution in ψ alone;

$$f(\psi|\mathbf{x}) = \int_\lambda f(\psi,\lambda_1,\lambda_2|\mathbf{x})\,d\lambda = \frac{\int_\lambda f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\lambda}{\displaystyle\int_\gamma f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\gamma}. \tag{2.14}$$

The posterior probability in (2.12) is obtained by integrating posterior $f(\psi|\mathbf{x})$ over the alternative hypothesis region;

$$P(\psi_1 < \psi < \psi_2|\mathbf{x}) = \int_{\psi_1}^{\psi_2} f(\psi|\mathbf{x})\,d\psi = \frac{\int_{\psi_1}^{\psi_2}\int_\lambda f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\lambda d\psi}{\displaystyle\int_\gamma f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\gamma}.$$

In a similar fashion, the posterior probability in (2.13) is obtained by integrating posterior $f(\psi|\mathbf{x})$ over lower and upper portions of the alternative hypothesis region;

$$P(\psi > \psi_1|\mathbf{x}) = \int_{\psi_1}^{\infty} f(\psi|\mathbf{x})\,d\psi = \frac{\int_{\psi_1}^{\infty}\int_\lambda f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\lambda d\psi}{\displaystyle\int_\gamma f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\gamma}$$

$$P(\psi < \psi_2|\mathbf{x}) = \int_{-\infty}^{\psi_2} f(\psi|\mathbf{x})\,d\psi = \frac{\displaystyle\int_{-\infty}^{\psi_2}\int_\lambda f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\lambda d\psi}{\displaystyle\int_\gamma f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\gamma}.$$

2.3.1. Two-Sided Bayesian Equivalence Procedure. For this procedure we need to compute

$$P(\psi_1 < \psi < \psi_2|\mathbf{x}) = \int_{\psi_1}^{\psi_2} f(\psi|\mathbf{x})\,d\psi = \frac{\int_{\psi_1}^{\psi_2}\int_\lambda f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\lambda d\psi}{\displaystyle\int_\gamma f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\,\pi(\psi,\lambda_1,\lambda_2)\,d\gamma}. \tag{2.15}$$

It is shown in the appendices that under a flat prior this posterior probability is

$$P\left(\psi_1 < \psi < \psi_2|\mathbf{x}\right) = \frac{\int_0^\infty \lambda_1^{\frac{n_1+n_2+2}{2}} \exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\left(\Phi_2-\Phi_1\right)d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2+4}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2+4}{2}}}}$$

and under a Jeffreys' prior it is

$$P\left(\psi_1 < \psi < \psi_2|\mathbf{x}\right) = \frac{\int_0^\infty \lambda_1^{\frac{n_1+n_2-2}{2}} \exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\left(\Phi_2-\Phi_1\right)d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2}{2}}}}$$

where s_p^2 is the pooled sample variance,

$$\Phi_1 = \Phi\left(\frac{\psi_1 + \frac{2\lambda_1(n_1T_1+n_2T_1+n_2T_3)}{n_1n_2}}{\sqrt{-\frac{2\lambda_1(n_1+n_2)}{n_1n_2}}}\right)$$

and

$$\Phi_2 = \Phi\left(\frac{\psi_2 + \frac{2\lambda_1(n_1T_1+n_2T_1+n_2T_3)}{n_1n_2}}{\sqrt{-\frac{2\lambda_1(n_1+n_2)}{n_1n_2}}}\right).$$

For both cases, the single integral in the numerator is easily approximated numerically.

2.3.2. Double One-Sided Bayesian Equivalence Procedure. The two new integrals we need to evaluate here are

$$\int_{\psi_1}^{\infty}\int_{\lambda} f\left(\mathbf{x}|\psi,\lambda_1,\lambda_2\right)\pi\left(\psi,\lambda_1,\lambda_2\right)d\lambda_1 d\lambda_2 d\psi \tag{2.16}$$

and

$$\int_{-\infty}^{\psi_2}\int_{\lambda} f\left(\mathbf{x}|\psi,\lambda_1,\lambda_2\right)\pi\left(\psi,\lambda_1,\lambda_2\right)d\lambda_1 d\lambda_2 d\psi. \tag{2.17}$$

It is shown in the appendices that under a flat prior these pairs of posterior probabilities are

$$P\left(\psi < \psi_2|\mathbf{x}\right) = \frac{\int_0^{\infty} \lambda_1^{\frac{n_1+n_2+2}{2}} \exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\Phi_2 d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2+4}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2+4}{2}}}}$$

and

$$P\left(\psi > \psi_1|\mathbf{x}\right) = \frac{\int_0^{\infty} \lambda_1^{\frac{n_1+n_2+2}{2}} \exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\left(1-\Phi_1\right) d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2+4}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2+4}{2}}}}.$$

Also, it is shown that for a Jeffreys' prior this pair is

$$P\left(\psi < \psi_2|\mathbf{x}\right) = \frac{\int_0^{\infty} \lambda_1^{\frac{n_1+n_2-2}{2}} \exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\Phi_2 d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2}{2}}}}$$

and

$$P\left(\psi > \psi_1|\mathbf{x}\right) = \frac{\int_0^{\infty} \lambda_1^{\frac{n_1+n_2-2}{2}} \exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\left(1-\Phi_1\right) d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2}{2}}}}.$$

2.4. PRINCIPLE OF CONFIDENCE INTERVAL INCLUSION

We also consider two methods which apply the principle of confidence interval inclusion; see Wellek (2010, sec. 3.1). The confidence interval inclusion methods we consider are equivalent to methods based upon the application of intersection-union tests to an equivalence null hypothesis; Berger (1982). For methods of this type, a $(1-2\alpha)\,100\%$ confidence interval $\left(\hat{\psi}_L, \hat{\psi}_U\right)$ for ψ of is generated. The procedures reject the nonequivalence null in (2.6) if confidence interval $\left(\hat{\psi}_L, \hat{\psi}_U\right)$ is contained in (ψ_1, ψ_2); the region corresponding to the equivalence alternative hypothesis in (2.6). We consider two confidence interval methods which condition upon the observed values of canonical sufficient statistics T_2 and T_3.

2.4.1. Conditionally Studentized Confidence Interval. The approximate asymptotic variance of $\hat{\psi}$ given the observed values for T_2 and T_3 is

$$Var_{\psi}\left(\hat{\psi}\right) = \frac{\left|j_{\lambda\lambda}\left(\psi, \hat{\lambda}_1(\psi), \hat{\lambda}_2(\psi)\right)\right|}{\left|j\left(\psi, \hat{\lambda}_1(\psi), \hat{\lambda}_2(\psi)\right)\right|}$$

as described in Butler (2007, sec. 5.4.5). The conditionally studentized statistic for ψ has an asymptotic standard normal distribution under null hypothesis $H_0 : \psi = \psi_0$ and is given as

$$Z_{\psi_0} = \frac{\hat{\psi} - \psi_0}{\sqrt{Var_{\psi_0}\left(\hat{\psi}\right)}}.$$

Note that the invariance of this statistic under the choice of interest parameter preserving transformations in matrix group M is shown in the appendix as part of the proof for Theorem 2.

For the test of $H_0 : \psi = \psi_0$ we fail to reject the null hypothesis at 2α significance if

$$|Z_{\psi_0}| \leq z_{1-2\alpha}$$

where $z_{1-2\alpha}$ denotes $(1-2\alpha)$th quantile of the standard normal distribution. An associated $(1-2\alpha)\,100\%$ confidence interval can be generated from conditionally studentized statistic Z_{ψ_0} by simultaneously solving the following equations:

$$\begin{aligned} Z_{\hat{\psi}_U} &= -z_{1-2\alpha} \\ Z_{\hat{\psi}_L} &= z_{1-2\alpha}. \end{aligned}$$

A grid search followed by the bisection method is used to solve these equations to an error of 10^{-6}.

2.4.2. Pivotal Confidence Interval. We also consider a $(1-2\alpha)\,100\%$ pivotal confidence interval where the pivotal quantity is the conditional CDF of canonical sufficient statistic T_1 given the observed values of T_2 and T_3. For a further discussion of the pivotal CDF method see Berger and Casella (2002, sec. 9.2.3). Here, we determine a confidence interval for ψ through the solution of the following equations:

$$\begin{aligned} F\left(t_1|t_2,t_3;\psi_L\right) &= 1-\alpha \\ F\left(t_1|t_2,t_3;\psi_U\right) &= \alpha \end{aligned} \tag{2.18}$$

This confidence interval has exact coverage under the assumption that the family of CDFs $\{F\left(t_1|t_2,t_3;\psi\right)\}$ is stochastically decreasing in ψ (Berger and Casella 2002, sec. 9.2.3). In practice, however, we use Skovgaard's saddlepoint approximation in place intractable CDF $F\left(t_1|t_2,t_3;\psi\right)$. Pivotal CDF confidence intervals often yield lengths and coverage probabilities that compare favorably with those from basically any competing method; see Paige and Trindade (2008) and Paige, Trindade and Fernando (2009).

2.5. MONTE CARLO STUDIES

In our simulation studies we assume that $\sigma^2 = 1$ and took our common sample size to be $n = 10, 20$ and 30. Here we assumed that Y_1 has zero mean and Y_2 has mean $-\varepsilon$ for

$$\varepsilon = 0.1, 0.2, \ldots, 1.$$

This results in ψ values of

$$\psi = 0.1, 0.2, \ldots, 1.$$

Furthermore, we set $\psi_1 = -\varepsilon$ and $\psi_2 = \varepsilon$ so that the true value of ψ is on the rightmost boundary of the null hypothesis. For each combination of n and ε values we simulated 100,000 data sets. Table 2.1 presents the empirical significance levels for the (i) saddlepoint-approximated optimal UMPU (O) procedure; (ii) the four Bayesian procedures: the two-sided procedure with a flat prior (F_1) and Jeffreys' prior (J_1), and the double one-sided procedure with a flat prior (F_2) and Jeffreys' prior (J_2) and (iii) the saddlepoint-based CDF pivotal confidence interval method ($C.I._1$) and the confidence interval generated from the conditionally studentized statistic for ψ ($C.I._2$). Here, the empirical significance levels for which the associated 95% Wald confidence interval contains nominal rate 0.05 are shown in bold.

We see that the saddlepoint-approximated optimal UMPU (O) procedure is remarkably accurate in terms of significance level even for very small sample sizes. In fact, it is only when $n = 30$ and very wide tolerable differences ε that any of the competing methods are even close to being unbiased. The poor performance of the confidence interval methods is likely due to their wideness for small samples. The poor performance for the Bayesian methods is probably due to the inability of the likelihood, with little data, to dominate the prior distribution.

2.6. CONCLUSIONS

We developed seven small sample equivalence tests from two independent normal samples for distributional parameter

$$\psi = (\mu_1 - \mu_2)/\sigma^2.$$

We considered at least one method from each of the three general small sample approaches for constructing an equivalence test. The saddlepoint-approximated optimal UMPU procedure was virtually unbiased in nearly all settings and is clearly superior to the other six methods.

Table 2.1. Empirical significance levels of equivalence tests for normal data

Empirical significance levels for $n = 10, 20$ and 30								
n	ε	O	F_1	J_1	F_2	J_2	$C.I._1$	$C.I._2$
10	0.1	**5.125**	0.000	0.000	0.000	0.000	0.000	0.000
10	0.2	**4.929**	0.000	0.000	0.000	0.000	0.000	0.000
10	0.3	**4.945**	0.000	0.000	0.000	0.000	0.000	0.000
10	0.4	**5.042**	0.000	0.000	0.000	0.000	0.000	0.000
10	0.5	**5.047**	0.000	0.000	0.000	0.001	0.001	0.000
10	0.6	**4.973**	0.000	0.008	0.017	0.150	0.170	0.023
10	0.7	**4.985**	0.018	0.195	0.265	0.962	1.064	0.236
10	0.8	**4.908**	0.288	1.102	0.973	2.354	2.569	0.854
10	0.9	**4.906**	0.898	2.420	1.770	3.403	3.769	1.447
10	1.0	**4.915**	1.674	3.449	2.256	3.963	4.425	1.835
20	0.1	**5.193**	0.000	0.000	0.000	0.000	0.000	0.000
20	0.2	**4.988**	0.000	0.000	0.000	0.000	0.000	0.000
20	0.3	**4.989**	0.000	0.000	0.000	0.000	0.000	0.000
20	0.4	**4.921**	0.000	0.000	0.000	0.010	0.008	0.019
20	0.5	**5.068**	0.014	0.061	0.334	0.772	0.819	1.074
20	0.6	**4.953**	0.659	1.431	1.977	3.061	3.191	3.308
20	0.7	**5.060**	2.449	3.721	3.159	4.413	4.655	4.015
20	0.8	**4.999**	3.226	4.554	3.350	4.649	**4.950**	3.866
20	0.9	**4.975**	3.180	4.570	3.205	4.583	**4.973**	3.497
20	1.0	**4.969**	2.991	4.472	2.993	4.474	**4.969**	3.016
30	0.1	**5.036**	0.000	0.000	0.000	0.000	0.000	0.000
30	0.2	**5.035**	0.000	0.000	0.000	0.000	0.000	0.000
30	0.3	**4.962**	0.003	0.000	0.000	0.002	0.000	0.000
30	0.4	**4.964**	0.002	0.007	0.181	0.377	0.390	1.231
30	0.5	**4.913**	0.939	1.626	2.568	3.317	3.421	4.365
30	0.6	**4.959**	3.304	4.257	3.676	4.589	4.777	**4.970**
30	0.7	**4.977**	3.661	4.676	3.690	4.693	**4.970**	4.694
30	0.8	**4.990**	3.518	4.651	3.519	4.652	**4.990**	4.343
30	0.9	**5.024**	3.356	4.608	3.356	4.608	**5.024**	3.939
30	1.0	**5.056**	3.233	4.582	3.233	4.582	**5.056**	3.567

APPENDIX A

PROPERNESS OF POSTERIOR DISTRIBUTIONS FROM SECTION 1

In this section we establish that the posterior distribution for θ is proper for both the flat prior and the objective Jeffreys' prior. For the flat prior, integrability of the posterior from zero to a finite positive constant, call it c, is guaranteed since marginal likelihood (1.7) is bounded and continuous in θ. To establish integrability from c to infinity we need to consider the tail behavior of the posterior. Note that by Gauss's multiplication formula we have that

$$\Gamma(n\theta) = (2\pi)^{\frac{1}{2}(1-n)} n^{n\theta-\frac{1}{2}} \prod_{k=0}^{n-1} \Gamma\left(\theta + \frac{k}{n}\right)$$

and by equation 6.1.47 of Abramowitz and Stegun (1972)

$$\frac{\Gamma(x+a)}{\Gamma(x)} \sim x^a$$

for $a > 0$ where "$\sim$" denotes *asymptotic equivalence* meaning that for large enough x the function on the left is essentially the same as the function on the right. As a result, for $n > 1$,

$$\begin{aligned}\frac{\Gamma(n\theta)}{\Gamma(\theta)^n} &\sim (2\pi)^{\frac{1}{2}(1-n)} n^{n\theta-\frac{1}{2}} \prod_{k=0}^{n-1} \theta^{\frac{k}{n}} \\ &= (2\pi)^{\frac{1}{2}(1-n)} n^{n\theta-\frac{1}{2}} \theta^{\frac{1}{2}n-\frac{1}{2}}\end{aligned}$$

and the posterior is proportional to

$$\begin{aligned}\exp\{(\theta-1)T + \ln\Gamma(n\theta) - n\ln\Gamma(\theta)\} &\sim n^{n\theta-\frac{1}{2}} \theta^{\frac{1}{2}n-\frac{1}{2}} \exp\{(\theta-1)T\} \\ &= \theta^{\frac{1}{2}n-\frac{1}{2}} e^{-\frac{1}{2}\ln n - \sum \ln z_i} e^{\theta[n\ln n + \sum \ln z_i]}\end{aligned}$$

where $z_i = x_i \left(\sum x_i\right)^{-1}$ for $i = 1, \ldots, n$. The well-known inequality of arithmetic and geometric means (Abramowitz and Stegun, 1972, eqn. 3.2.1) states that the mean of nonnegative real numbers is less than or equal to their arithmetic mean (with equality when all of the numbers are equal). Therefore, with probability one

$$\sqrt[n]{z_1 \cdots z_n} < \frac{z_1 + z_2 + ...z_n}{n} = \frac{1}{n}$$

which implies that

$$\sum \ln z_i < -n \ln n$$

and, as a result, the posterior asymptotically equivalent to a finite constant times a gamma density and as such is integrable.

For the Jeffreys' prior posterior integrability from a positive constant c to infinity follows from the fact that Jeffreys' prior approaches zero as $\theta \rightarrow \infty$ and posterior integrability with a flat prior that was proven above. Posterior integrability from zero to a finite constant, however, needs to be investigated since the trigamma function approaches infinity as $\theta \rightarrow 0$.

Differentiation of Gauss's multiplication formula yields equation 6.4.8 of Abramowitz and Stegun (1972);

$$\psi^{'}(n\theta) = n^{-2} \sum_{k=0}^{n-1} \psi^{'}\left(\theta + \frac{k}{n}\right)$$

and, as a result, the following identity:

$$n\psi^{'}(\theta) - n^2 \psi^{'}(n\theta) = (n-1)\,\psi^{'}(\theta) - \sum_{k=1}^{n-1} \psi^{'}\left(\theta + \frac{k}{n}\right).$$

From equation 6.4.10 of Abramowitz and Stegun (1972) we have that

$$\psi^{'}(\theta) = \sum_{i=0}^{\infty} \frac{1}{(\theta+i)^2} = \frac{1}{\theta^2} + \sum_{i=1}^{\infty} \frac{1}{(\theta+i)^2}$$

and

$$\sqrt{n\psi'(\theta) - n^2\psi'(n\theta)} = \frac{1}{\theta}\sqrt{(n-1) + \theta^2\left\{(n-1)\sum_{i=0}^{\infty}\frac{1}{(\theta+i)^2} - \sum_{k=1}^{n-1}\psi'\left(\theta+\frac{k}{n}\right)\right\}}.$$

The reciprocal gamma function is an entire function (Abramowitz and Stegun, 1972, ch. 6) with the following Taylor series expansion around zero

$$\frac{1}{\Gamma(\theta)} = \theta + \gamma\theta^2 + \cdots = \theta + o\left(\theta^2\right)$$

where $\gamma \approx 0.5772$ is Euler's constant. From this we obtain

$$\begin{aligned}\frac{\Gamma(n\theta)}{\Gamma(\theta)^n} &= (2\pi)^{\frac{1}{2}(1-n)}\, n^{n\theta-\frac{1}{2}}\prod_{k=1}^{n-1}\frac{\Gamma\left(\theta+\frac{k}{n}\right)}{\Gamma(\theta)} \\ &= \theta^n (2\pi)^{\frac{1}{2}(1-n)}\, n^{n\theta-\frac{1}{2}}\prod_{k=1}^{n-1}\Gamma\left(\theta+\frac{k}{n}\right) + o\left(\theta^2\right)\end{aligned}$$

and it is now easily verified that the posterior converges to zero as $\theta \to 0$ and is therefore integrable from zero to infinity.

APPENDIX B

PROOFS AND DERIVATIONS FOR SECTION 2

B.1. Proof of Theorem 2.1. A transformation $\mathbf{B}^T$ in matrix group M equates to the following changes in canonical parameters and sufficient statistics:

$$\tilde{\gamma} = \mathbf{B}^T \gamma = \begin{bmatrix} \psi \\ a\psi + b\lambda_1 + c\lambda_2 \\ d\psi + e\lambda_1 + f\lambda_2 \end{bmatrix} = \begin{bmatrix} \psi \\ \tilde{\lambda}_1 \\ \tilde{\lambda}_2 \end{bmatrix}$$

with associated canonical sufficient statistics

$$\tilde{\mathbf{T}} = \mathbf{B}^{-1}\mathbf{T} = \begin{bmatrix} T_1 - T_3 \frac{ae-bd}{ce-bf} + \frac{T_2}{ce-bf}(af - cd) \\ T_3 \frac{e}{ce-bf} - f\frac{T_2}{ce-bf} \\ c\frac{T_2}{ce-bf} - b\frac{T_3}{ce-bf} \end{bmatrix} = \begin{bmatrix} \tilde{T}_1 \\ \tilde{T}_2 \\ \tilde{T}_3 \end{bmatrix}.$$

The associated likelihood function is

$$\begin{aligned} \mathcal{L}(\tilde{\gamma}) &\propto \exp\left\{\psi \tilde{T}_1 + \tilde{\lambda}_1 \tilde{T}_2 + \tilde{\lambda}_2 \tilde{T}_3 + c\left(\psi, \tilde{\lambda}_1, \tilde{\lambda}_2\right)\right\} \\ &= \exp\left\{\tilde{\gamma}^T \tilde{\mathbf{T}} + c\left(\mathbf{B}^{-T}\tilde{\gamma}\right)\right\}. \end{aligned}$$

The optimal UMPU test for ψ depends upon the following conditional distribution:

$$f\left(\tilde{t}_1 | \tilde{t}_2, \tilde{t}_3, \psi\right).$$

Therefore we need to show that

$$f\left(\tilde{t}_1 | \tilde{t}_2, \tilde{t}_3, \psi\right) = f\left(t_1 | t_2, t_3, \psi\right).$$

First note that

$$
\begin{aligned}
f\left(t_1|t_2,t_3;\psi\right) &= \exp\left\{\psi t_1 - c\left(\psi|t_2,t_3\right) - d\left(t_1,t_2,t_3\right)\right\} \\
&= (2\pi)^{-\left(\frac{n_1+n_2}{2}\right)} \exp\left\{\psi t_1 - c\left(\psi|t_2,t_3\right)\right\} \\
&= (2\pi)^{-\left(\frac{n_1+n_2}{2}\right)} \frac{\exp\left(\psi t_1\right)}{\exp\left\{c\left(\psi|t_2,t_3\right)\right\}} \\
&= \frac{(2\pi)^{-\left(\frac{n_1+n_2}{2}\right)} \exp\left(\psi t_1\right)}{\int_{\{t_1:(t_1,t_2,t_3)\in S\}} (2\pi)^{-\left(\frac{n_1+n_2}{2}\right)} \exp\left(\psi t_1\right) dt_1}
\end{aligned} \tag{B.1}
$$

where S is the joint support of (t_1, t_2, t_3).

Note that

$$
\begin{aligned}
t_1 &= -\sum y_2 \\
t_2 &= \sum y_1^2 + \sum y_2^2 \\
t_3 &= \sum y_1 + \sum y_2.
\end{aligned}
$$

Since the joint support S of (t_1, t_2, t_3) is a complicated surface in $\mathbb{R}^3$ it appears that the surface integral in $(B.1)$ cannot be evaluated in closed-form.

Consider a reparametrization from matrix group M as determined by transformation matrix $\mathbf{B}$. We can then reparametrize the joint likelihood function as

$$
\begin{aligned}
\exp\left\{\gamma^T\mathbf{T} + c\left(\gamma\right)\right\} &\propto \exp\left\{\gamma^T\mathbf{B}\mathbf{B}^{-1}\mathbf{T} + c\left(\gamma\right)\right\} \\
&= \exp\left\{\left(\mathbf{B}^T\gamma\right)^T\left(\mathbf{B}^{-1}\mathbf{T}\right) + c\left(\gamma\right)\right\} \\
&= \exp\left\{\tilde{\gamma}^T\tilde{\mathbf{T}} + c\left(\mathbf{B}^{-T}\tilde{\gamma}\right)\right\}
\end{aligned}
$$

where

$$
\gamma = \begin{bmatrix} \psi \\ \lambda_1 \\ \lambda_2 \end{bmatrix}, \mathbf{T} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}, \tilde{\gamma} = \mathbf{B}^T\gamma = \begin{bmatrix} \psi \\ \tilde{\lambda}_1 \\ \tilde{\lambda}_2 \end{bmatrix}, \quad \text{and } \tilde{\mathbf{T}} = \mathbf{B}^{-1}\mathbf{T} = \begin{bmatrix} \tilde{t}_1 \\ \tilde{t}_2 \\ \tilde{t}_3 \end{bmatrix}.
$$

The regular exponential family form for the density is

$$f\left(\tilde{t}_1, \tilde{t}_2, \tilde{t}_3; \psi, \tilde{\lambda}_1, \tilde{\lambda}_2\right) = \exp\left\{ \begin{bmatrix} \tilde{\lambda}_1 \\ \tilde{\lambda}_2 \end{bmatrix}^T \begin{bmatrix} \tilde{t}_2 \\ \tilde{t}_3 \end{bmatrix} + \psi\tilde{t}_1 + c\left(\psi, \tilde{\lambda}_1, \tilde{\lambda}_2\right) \right\}.$$

Then the conditional distribution of $\tilde{T}_1$ given $\tilde{T}_2 = \tilde{t}_2$ and $\tilde{T}_3 = \tilde{t}_3$ can be expressed as

$$\begin{aligned} f\left(\tilde{t}_1 | \tilde{T}_2 = \tilde{t}_2, \tilde{T}_3 = \tilde{t}_3; \psi\right) &= \exp\left\{\psi\tilde{t}_1 + c\left(\psi, \tilde{\lambda}_1, \tilde{\lambda}_2\right)\right\} \\ &= \frac{\exp\left(\psi\tilde{t}_1\right)}{\exp\left\{c\left(\psi, \tilde{\lambda}_1, \tilde{\lambda}_2\right)\right\}} \\ &= \frac{\exp\left(\psi\tilde{t}_1\right)}{\int_{\left\{\tilde{t}_1 : \left(\tilde{t}_1, \tilde{t}_2, \tilde{t}_3\right) \in \tilde{S}\right\}} \exp\left(\psi\tilde{t}_1\right) d\tilde{t}_1}. \end{aligned}$$

Therefore it suffices to show that

$$\frac{\exp\left(\psi\tilde{t}_1\right)}{\int_{\left\{\tilde{t}_1 : \left(\tilde{t}_1, \tilde{t}_2, \tilde{t}_3\right) \in \tilde{S}\right\}} \exp\left(\psi\tilde{t}_1\right) d\tilde{t}_1} = \frac{\exp\left(\psi t_1\right)}{\int_{\{t_1 \,:\, (t_1, t_2, t_3) \in S\}} \exp\left(\psi t_1\right) dt_1}. \tag{B.2}$$

Recall that

$$\tilde{t} = \mathbf{B}^{-1} t$$

so that

$$\tilde{t}_1 = \mathbf{B}^{-1}_{(1)} t$$

where $\mathbf{B}^{-1}_{(1)}$ denotes the first row of $\mathbf{B}^{-1}$.

Then the left-hand side of $(B.2)$ can be written as

$$\frac{\exp\left(\psi\tilde{t}_1\right)}{\int_{\left\{\tilde{t}_1 : \left(\tilde{t}_1, \tilde{t}_2, \tilde{t}_3\right) \in \tilde{S}\right\}} \exp\left(\psi\tilde{t}_1\right) d\tilde{t}_1} = \frac{\exp\left(\psi\mathbf{B}^{-1}_{(1)} t\right)}{\int_{\left\{\mathbf{B}^{-1}_{(1)} t \,:\, \mathbf{B}^{-1} t \in \mathbf{B}^{-1} S\right\}} \exp\left(\psi\mathbf{B}^{-1}_{(1)} t\right) d\mathbf{B}^{-1}_{(1)} t}.$$

Now consider the change variable for the right side of $(B.2)$ in which we replace t with $\mathbf{B}t$;

$$\frac{\exp\left(\psi\mathbf{B}_{(1)}^{-1}t\right)}{\int_{\left\{\mathbf{B}_{(1)}^{-1}t\,:\,\mathbf{B}^{-1}t\in\mathbf{B}^{-1}S\right\}}\exp\left(\psi\mathbf{B}_{(1)}^{-1}t\right)d\mathbf{B}_{(1)}^{-1}t}=\frac{\exp\left(\psi\mathbf{B}_{(1)}^{-1}\mathbf{B}t\right)}{\int_{\left\{\mathbf{B}_{(1)}^{-1}\mathbf{B}t\,:\,\mathbf{B}^{-1}\mathbf{B}t\in\mathbf{B}^{-1}\mathbf{B}S\right\}}\exp\left(\psi\mathbf{B}_{(1)}^{-1}\mathbf{B}t\right)d\mathbf{B}_{(1)}^{-1}\mathbf{B}t}.$$

But note that

$$\mathbf{B}_{(1)}^{-1}\mathbf{B}=\begin{bmatrix}1 & 0 & 0\end{bmatrix}$$

so then

$$\frac{\exp\left(\psi\mathbf{B}_{(1)}^{-1}\mathbf{B}t\right)}{\int_{\left\{\mathbf{B}_{(1)}^{-1}\mathbf{B}t\,:\,\mathbf{B}^{-1}\mathbf{B}t\in\mathbf{B}^{-1}\mathbf{B}S\right\}}\exp\left(\psi\mathbf{B}_{(1)}^{-1}\mathbf{B}t\right)d\mathbf{B}_{(1)}^{-1}\mathbf{B}t}=\frac{\exp\left(\psi t_1\right)}{\int_{\{t_1\,:\,(t_1,t_2,t_3)\in S\}}\exp\left(\psi t_1\right)dt_1}.$$

$\square$

B.2. Derivation of Conditional MLEs. Setting the score equations equal to zero, we have

$$\begin{cases}\frac{\partial\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_1}=T_2-\frac{n_1\lambda_2^2}{4\lambda_1^2}+\frac{n_1}{2\lambda_1}-\frac{n_2(\lambda_2-\psi)^2}{4\lambda_1^2}+\frac{n_2}{2\lambda_1}=0\\ \frac{\partial\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\lambda_2}=T_3+\frac{n_1\lambda_2}{2\lambda_1}+\frac{n_2(\lambda_2-\psi)}{2\lambda_1}=0\end{cases}$$

which yields

$$\begin{cases}4T_2\lambda_1^2-n_1\lambda_2^2+2n_1\lambda_1-n_2\left(\lambda_2-\psi\right)^2+2n_2\lambda_1=0\\ 2T_3\lambda_1+n_1\lambda_2+n_2\left(\lambda_2-\psi\right)=0.\end{cases}$$

From the second equation, we obtain

$$\lambda_2=\frac{n_2\psi-2T_3\lambda_1}{n_1+n_2}. \tag{B.3}$$

Plugging this expression into the first equation and simplifying yields

$$4\left(n_1+n_2\right)\left[\left(n_1+n_2\right)T_2-T_3^2\right]\lambda_1^2+2\left(n_1+n_2\right)^3\lambda_1-n_1n_2\psi^2\left(n_1+n_2\right)=0.$$

Solving for λ_1 we obtain two possible solutions

$$\hat{\lambda}_1(\psi) = \frac{-(n_1+n_2)^2 \pm \sqrt{(n_1+n_2)^4 + 4n_1n_2\psi^2\left[(n_1+n_2)T_2 - T_3^2\right]}}{4\left[(n_1+n_2)T_2 - T_3^2\right]}. \tag{B.4}$$

It is easy to show that

$$(n_1+n_2)T_2 - T_3^2 > 0$$

if $n_1 > 1$ or $n_2 > 1$.

Since

$$\lambda_1 = -\frac{1}{2\sigma^2} < 0$$

then the $\hat{\lambda}_1(\psi)$ is the negative solution given in $(B.4)$;

$$\hat{\lambda}_1(\psi) = \frac{-(n_1+n_2)^2 - \sqrt{(n_1+n_2)^4 + 4n_1n_2\psi^2\left[(n_1+n_2)T_2 - T_3^2\right]}}{4\left[(n_1+n_2)T_2 - T_3^2\right]}.$$

Plugging this solution into $(B.3)$ yields

$$\hat{\lambda}_2(\psi) = \frac{n_2\psi - 2\hat{\lambda}_1(\psi)T_3}{n_1+n_2}.$$

B.3. Proof of Theorem 2.2. Consider the likelihood function obtained by applying a transformation $\mathbf{B}^T$ in matrix group M to primary reparametrization (2.4)

$$\begin{aligned} \mathcal{L}(\tilde{\gamma}) &\propto \exp\left\{\psi\tilde{T}_1 + \tilde{\lambda}_1\tilde{T}_2 + \tilde{\lambda}_2\tilde{T}_3 + c\left(\psi, \tilde{\lambda}_1, \tilde{\lambda}_2\right)\right\} \\ &= \exp\left\{\tilde{\gamma}^T\tilde{\mathbf{T}} + c\left(\mathbf{B}^{-T}\tilde{\gamma}\right)\right\} \end{aligned}$$

where

$$\tilde{\mathbf{T}} = \mathbf{B}^{-1}\mathbf{T} = \begin{bmatrix} T_1 - T_3\frac{ae-bd}{ce-bf} + \frac{T_2}{ce-bf}\left(af-cd\right) \\ T_3\frac{e}{ce-bf} - f\frac{T_2}{ce-bf} \\ c\frac{T_2}{ce-bf} - b\frac{T_3}{ce-bf} \end{bmatrix} = \begin{bmatrix} \tilde{T}_1 \\ \tilde{T}_2 \\ \tilde{T}_3 \end{bmatrix}.$$

The associated log-likelihood is

$$\begin{aligned} \ell\left(\tilde{\gamma}\right) &\propto \psi\tilde{T}_1 + \tilde{\lambda}_1\tilde{T}_2 + \tilde{\lambda}_2\tilde{T}_3 + c\left(\psi, \tilde{\lambda}_1, \tilde{\lambda}_2\right) \\ &= \tilde{\gamma}^T\tilde{\mathbf{T}} + c\left(\mathbf{B}^{-T}\tilde{\gamma}\right). \end{aligned} \tag{B.5}$$

The log-likelihood for primary reparametrization in (2.4) is

$$\ell\left(\gamma\right) \propto \gamma^T\mathbf{T} + c\left(\gamma\right).$$

The score equations for this log-likelihood are

$$\mathbf{T}+\nabla c\left(\gamma\right) = 0$$

where ∇ is the gradient symbol which represents the first partial derivative of $c\left(\cdot\right)$ w.r.t. each element in γ. MLE $\hat{\gamma}$ is the solution to above score equation which means that

$$\mathbf{T}+\nabla c\left(\hat{\gamma}\right) = 0.$$

The derivative of $\ell\left(\tilde{\gamma}\right)$ w.r.t. $\tilde{\gamma}$ can be expressed in the following way:

$$\tilde{\mathbf{T}} + \mathbf{B}^{-1}\nabla c\left(\mathbf{B}^{-T}\tilde{\gamma}\right) = \mathbf{B}^{-1}\mathbf{T} + \mathbf{B}^{-1}\nabla c\left(\mathbf{B}^{-T}\tilde{\gamma}\right) = \mathbf{B}^{-1}\left[\mathbf{T}+\nabla c\left(\mathbf{B}^{-T}\tilde{\gamma}\right)\right] \tag{B.6}$$

using matrix derivative formulas from Harville (2000, sec.15.7).

Let $\hat{\tilde{\gamma}}$ denote the MLE of $\tilde{\gamma}$, then

$$\mathbf{B}^{-1}\left[\mathbf{T}+\nabla c\left(\mathbf{B}^{-T}\hat{\tilde{\gamma}}\right)\right] = 0.$$

Since

$$\mathbf{T}+\nabla c\left(\hat{\gamma}\right)=0$$

then it follows that

$$\hat{\gamma}=\mathbf{B}^{-T}\widehat{\tilde{\gamma}}$$

and

$$\widehat{\tilde{\gamma}}=\mathbf{B}^{T}\hat{\gamma}$$

due the uniqueness of MLEs in canonical exponential families. Note also that this result is to be expected due to the invariance of MLEs. A similar argument shows that the conditional MLEs follow the same idea, that is

$$\left[\begin{array}{c}\widehat{\tilde{\lambda}}_1\left(\psi\right)\\ \widehat{\tilde{\lambda}}_2\left(\psi\right)\end{array}\right]=\mathbf{B}_{(2)}^{T}\left[\begin{array}{c}\psi\\ \hat{\lambda}_1\left(\psi\right)\\ \hat{\lambda}_2\left(\psi\right)\end{array}\right]$$

where $\mathbf{B}_{(2)}^{T}$ denotes the last two rows of $\mathbf{B}^{T}$.

Next we consider the likelihood quantities in Skovgaard's CDF approximation (2.8) and show that they are invariant under the interesting parameter ψ preserving reparametrizations induced by transformations in matrix group M. Note that this result will also establish invariance for the saddlepoint conditional PDF approximation (2.7) as well since it depends upon the same likelihood quantities as Skovgaard's approximation.

First we consider likelihood ratio

$$\frac{\mathcal{L}\left(\psi,\hat{\lambda}_1\left(\psi\right),\hat{\lambda}_2\left(\psi\right)\right)}{\mathcal{L}\left(\hat{\psi},\hat{\lambda}_1,\hat{\lambda}_2\right)}$$

which appears in the expression for the w parameter in (2.8) . We first show that the maximized likelihood is invariant under reparametrizations in M. To see this note that

$$\begin{aligned}
\mathcal{L}\left(\widehat{\tilde{\gamma}}\right) &\propto \exp\left\{\widehat{\tilde{\gamma}}^T\tilde{\mathbf{T}} + c\left(\mathbf{B}^{-T}\widehat{\tilde{\gamma}}\right)\right\} \\
&= \exp\left\{\left(\mathbf{B}^T\hat{\gamma}\right)^T\tilde{\mathbf{T}}+c\left(\mathbf{B}^{-T}\mathbf{B}^T\hat{\gamma}\right)\right\} \\
&= \exp\left\{\hat{\gamma}^T\mathbf{T} + c\left(\hat{\gamma}\right)\right\} \\
&= \mathcal{L}\left(\hat{\gamma}\right).
\end{aligned}$$

The profile likelihood has similar invariance properties;

$$\begin{aligned}
\mathcal{L}\left(\psi, \widehat{\tilde{\lambda}}_1(\psi), \widehat{\tilde{\lambda}}_2(\psi)\right) &\propto \exp\left\{\widehat{\tilde{\gamma}}_p^{\,T}\tilde{\mathbf{T}} + c\left(\mathbf{B}^{-T}\widehat{\tilde{\gamma}}_p\right)\right\} \\
&= \exp\left\{\left(\mathbf{B}^T\hat{\gamma}_p\right)^T\tilde{\mathbf{T}}+c\left(\mathbf{B}^{-T}\mathbf{B}^T\hat{\gamma}_p\right)\right\} \\
&= \exp\left\{\hat{\gamma}_p^T\mathbf{T} + c\left(\hat{\gamma}_p\right)\right\} \\
&= \mathcal{L}\left(\psi, \hat{\lambda}_1(\psi), \hat{\lambda}_2(\psi)\right)
\end{aligned}$$

where $\widehat{\tilde{\gamma}}_p$ and $\hat{\gamma}_p$ are augmented MLE vectors defined as

$$\widehat{\tilde{\gamma}}_p = \begin{bmatrix} \psi \\ \widehat{\tilde{\lambda}}_1(\psi) \\ \widehat{\tilde{\lambda}}_2(\psi) \end{bmatrix} \quad \text{and} \quad \hat{\gamma}_p = \begin{bmatrix} \psi \\ \hat{\lambda}_1(\psi) \\ \hat{\lambda}_2(\psi) \end{bmatrix}.$$

Next consider the invariance of the ratio of determinants for the full and partial Fisher information matrices;

$$\frac{\left|j\left(\hat{\psi}, \hat{\lambda}_1, \hat{\lambda}_2\right)\right|}{\left|j_{\lambda\lambda}\left(\psi, \hat{\lambda}_1(\psi), \hat{\lambda}_2(\psi)\right)\right|}$$

which appears in the expression for the u parameter in (2.8).

Recall from $(B.6)$ that derivative of $\ell(\tilde{\gamma})$ w.r.t. $\tilde{\gamma}$ can be expressed in the following way:

$$\frac{\partial \ell(\tilde{\gamma})}{\partial \tilde{\gamma}} = \mathbf{B}^{-1}\left[\mathbf{T}+\nabla c\left(\mathbf{B}^{-T}\tilde{\gamma}\right)\right].$$

Therefore, again using results from Harville (2000, sec.15.7), we have that the Hessian matrix for $\ell(\tilde{\gamma})$ is given as

$$\frac{\partial^2 \ell(\tilde{\gamma})}{\partial \tilde{\gamma}\partial \tilde{\gamma}^T} = \mathbf{B}^{-1}\left[Hc\left(\mathbf{B}^{-T}\tilde{\gamma}\right)\right]\mathbf{B}^{-T}$$

where $Hc(\gamma)$ is the Hessian matrix for $c(\gamma)$ and is given as

$$Hc(\gamma) = \frac{\partial^2 \ell(\gamma)}{\partial \gamma \partial \gamma^T}.$$

Therefore

$$\frac{\partial^2 \ell(\tilde{\gamma})}{\partial \tilde{\gamma}\partial \tilde{\gamma}^T} = \mathbf{B}^{-1}\frac{\partial^2 \ell(\tilde{\gamma})}{\partial \tilde{\gamma}\partial \tilde{\gamma}^T}\mathbf{B}^{-T}$$

and

$$j\left(\psi, \tilde{\lambda}_1, \tilde{\lambda}_2\right) = \mathbf{B}^{-1} j(\psi, \lambda_1, \lambda_2)\,\mathbf{B}^{-T}.$$

To consider partial information matrix for $\tilde{\lambda} = \left[\tilde{\lambda}_1, \tilde{\lambda}_2\right]$ we first partition the full information matrix for γ in the following way:

$$j(\psi, \lambda_1, \lambda_2) = \begin{bmatrix} j_{\psi\psi}(\psi, \lambda_1, \lambda_2) & j_{\psi\lambda}(\psi, \lambda_1, \lambda_2) \\ j_{\lambda\psi}(\psi, \lambda_1, \lambda_2) & j_{\lambda\lambda}(\psi, \lambda_1, \lambda_2) \end{bmatrix}$$

where

$$j_{\psi\psi}(\psi, \lambda_1, \lambda_2) = -\frac{\partial^2 \mathcal{L}(\psi, \lambda_1, \lambda_2)}{\partial \psi^2},$$

$$j_{\psi\lambda}(\psi,\lambda_1,\lambda_2) = \begin{bmatrix} -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\psi\partial\lambda_1} & -\frac{\partial^2\mathcal{L}(\psi,\lambda_1,\lambda_2)}{\partial\psi\partial\lambda_2} \end{bmatrix}$$

and

$$j_{\lambda\psi}(\psi,\lambda_1,\lambda_2) = j_{\psi\lambda}^T(\psi,\lambda_1,\lambda_2).$$

Note that

$$\begin{aligned} j\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right) &= \mathbf{B}^{-1} j(\psi,\lambda_1,\lambda_2)\mathbf{B}^{-T} \\ &= \begin{bmatrix} 1 & \mathbf{A}_{1\times 2} \\ \mathbf{0}_{2\times 1} & \mathbf{C}_{2\times 2} \end{bmatrix} \begin{bmatrix} j_{\psi\psi}(\psi,\lambda_1,\lambda_2) & j_{\psi\lambda}(\psi,\lambda_1,\lambda_2) \\ j_{\lambda\psi}(\psi,\lambda_1,\lambda_2) & j_{\lambda\lambda}(\psi,\lambda_1,\lambda_2) \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}_{1\times 2} \\ \mathbf{A}_{2\times 1}^T & \mathbf{C}_{2\times 2}^T \end{bmatrix} \\ &= \begin{bmatrix} j_{\psi\psi}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right) & j_{\psi\tilde{\lambda}}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right) \\ j_{\tilde{\lambda}\psi}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right) & j_{\tilde{\lambda}\tilde{\lambda}}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right) \end{bmatrix} \end{aligned}$$

where

$$j_{\psi\psi}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right) = j_{\psi\psi}(\psi,\lambda_1,\lambda_2) + \mathbf{A} j_{\lambda\psi}(\psi,\lambda_1,\lambda_2) + j_{\psi\lambda}(\psi,\lambda_1,\lambda_2)\mathbf{A}^T + \mathbf{A} j_{\lambda\lambda}(\psi,\lambda_1,\lambda_2)\mathbf{A}^T,$$

$$j_{\psi\tilde{\lambda}}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right) = \left[j_{\psi\lambda}(\psi,\lambda_1,\lambda_2) + \mathbf{A} j_{\lambda\lambda}(\psi,\lambda_1,\lambda_2)\right]\mathbf{C}^T,$$

$$j_{\tilde{\lambda}\psi}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right) = j_{\psi\tilde{\lambda}}^T\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right)$$

and

$$j_{\tilde{\lambda}\tilde{\lambda}}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right) = \mathbf{C} j_{\lambda\lambda}(\psi,\lambda_1,\lambda_2)\mathbf{C}^T.$$

Consider now the following ratio of determinants of the full and partial Fisher information matrices for a secondary reparametrization;

$$\frac{\left|j\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right)\right|}{\left|j_{\tilde{\lambda}\tilde{\lambda}}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right)\right|}.$$

Recall that

$$j\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right)=\mathbf{B}^{-1}j\left(\psi,\lambda_1,\lambda_2\right)\mathbf{B}^{-T}$$

so then

$$\left|j\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right)\right|=\left|j\left(\psi,\lambda_1,\lambda_2\right)\right|\left|\mathbf{B}\right|^{-2}.$$

Also

$$j_{\tilde{\lambda}\tilde{\lambda}}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right)=\mathbf{C}j_{\lambda\lambda}\left(\psi,\lambda_1,\lambda_2\right)\mathbf{C}^{T}$$

which means that

$$\left|j_{\tilde{\lambda}\tilde{\lambda}}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right)\right|=\left|j_{\lambda\lambda}\left(\psi,\lambda_1,\lambda_2\right)\right|\left|\mathbf{C}\right|^{2}.$$

However note that

$$\left|\mathbf{B}\right|=\left|\mathbf{C}\right|^{-1}$$

so then it follows that

$$\frac{\left|j\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right)\right|}{\left|j_{\tilde{\lambda}\tilde{\lambda}}\left(\psi,\tilde{\lambda}_1,\tilde{\lambda}_2\right)\right|}=\frac{\left|j\left(\psi,\lambda_1,\lambda_2\right)\right|}{\left|j_{\lambda\lambda}\left(\psi,\lambda_1,\lambda_2\right)\right|}.$$

This result has a number of ramifications including the reparametrization invariance of the u parameter in Skovgaard's CDF approximation (2.8) and the invariance of the approximate asymptotic conditional variance of $\hat{\psi}$ given that $T_2=t_2$ and $T_3=t_3$ which is given in Butler(2007, sec. 5.4.5) as

$$\begin{aligned}
j_{\psi\psi\cdot\lambda}^{-1} &= \frac{\left|j_{\lambda\lambda}\left(\psi,\hat{\lambda}_1\left(\psi\right),\hat{\lambda}_2\left(\psi\right)\right)\right|}{\left|j\left(\psi,\hat{\lambda}_1\left(\psi\right),\hat{\lambda}_2\left(\psi\right)\right)\right|}\\
&= j_{\psi\psi}\left(\psi,\lambda_1,\lambda_2\right)-j_{\psi\lambda}\left(\psi,\lambda_1,\lambda_2\right)j_{\lambda\lambda}^{-1}\left(\psi,\lambda_1,\lambda_2\right)j_{\lambda\psi}\left(\psi,\lambda_1,\lambda_2\right)\big]_{\psi,\hat{\lambda}_1(\psi),\hat{\lambda}_2(\psi)}.
\end{aligned}$$

This asymptotic conditional variance is used to define the conditionally studentized statistic for ψ in section 2.4.

$\square$

B.4.1. Two-Sided Bayesian Equivalence Procedure: Flat Prior. With a flat improper prior $\pi(\psi, \lambda_1, \lambda_2) = 1$ we are able to evaluate most of our integrals in closed-form. The denominator of (2.15) is given as

$$\int_\gamma f(\mathbf{x}|\psi, \lambda_1, \lambda_2)\, d\gamma = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{0} f(\mathbf{x}|\psi, \lambda_1, \lambda_2)\, d\lambda_1 d\lambda_2 d\psi \tag{B.7}$$

where

$$f(\mathbf{x}|\psi, \lambda_1, \lambda_2) \propto e^{\psi T_1 + \lambda_1 T_2 + \lambda_2 T_3 + \frac{n_1\lambda_2^2 + n_2(\lambda_2-\psi)^2}{4\lambda_1} + \frac{1}{2}(n_1+n_2)\ln(-2\lambda_1)}.$$

One can integrate this function in closed-form over ψ and λ_2 since the associated integrands are proportional to a normal density.

First we integrate with respect to (w.r.t.) λ_2. Separating out the portions of $f(\mathbf{x}|\psi, \lambda_1, \lambda_2)$ which do not depend upon λ_2 yields

$$\int_{-\infty}^{0}\int_{-\infty}^{\infty} (-2\lambda_1)^{\frac{n_1+n_2}{2}} \exp\left\{\psi T_1 + \lambda_1 T_2 + \frac{n_2\psi^2}{4\lambda_1}\right\} I_{\lambda_2}(\psi)\, d\psi d\lambda_1$$

where

$$I_{\lambda_2}(\psi) = \int_{-\infty}^{\infty} \exp\left\{\frac{n_1+n_2}{4\lambda_1}\lambda_2^2 + \left(T_3 - \frac{\psi n_2}{2\lambda_1}\right)\lambda_2\right\} d\lambda_2$$

is the inner integral w.r.t. λ_2. Applying the well-known Gaussian integral identity, i.e.

$$\int_{-\infty}^{\infty} \exp\left\{-ax^2 + bx\right\} dx = \sqrt{\frac{\pi}{a}} \exp\left\{\frac{b^2}{4a}\right\} \text{ for } a > 0,$$

to $I_{\lambda_2}(\psi)$ yields the following closed-form expression:

$$I_{\lambda_2}(\psi) = \sqrt{\frac{-4\pi\lambda_1}{n_1+n_2}} \exp\left\{ \frac{-\lambda_1 \left(T_3 - \frac{\psi n_2}{2\lambda_1}\right)^2}{n_1+n_2} \right\}.$$

This leads to the following integral in ψ and λ_1;

$$\int_{-\infty}^{0}\int_{-\infty}^{\infty} (-2\lambda_1)^{\frac{n_1+n_2}{2}} \sqrt{\frac{-4\pi\lambda_1}{n_1+n_2}} \exp\left\{ \psi T_1 + \lambda_1 T_2 + \frac{n_2\psi^2}{4\lambda_1} - \frac{\lambda_1 \left(T_3 - \frac{\psi n_2}{2\lambda_1}\right)^2}{n_1+n_2} \right\} d\psi d\lambda_1.$$

Separating out the portions of the above integrand which do not depend upon ψ yields

$$\int_{-\infty}^{0} (-2\lambda_1)^{\frac{n_1+n_2}{2}} \sqrt{\frac{-4\pi\lambda_1}{n_1+n_2}} \exp\left\{ \lambda_1 T_2 - \frac{\lambda_1 T_3^2}{n_1+n_2} \right\} I_\psi(\lambda_1)\, d\lambda_1$$

where the inner integral w.r.t. ψ is given as

$$I_\psi(\lambda_1) = \int_{-\infty}^{\infty} \exp\left\{ \frac{n_1 n_2}{4\lambda_1(n_1+n_2)}\psi^2 + \left(T_1 + \frac{n_2 T_3}{n_1+n_2}\right)\psi \right\} d\psi.$$

Applying the Gaussian integral identity to the above inner integral results in the following closed-form expression:

$$I_\psi(\lambda_1) = \sqrt{\frac{-4\pi\lambda_1(n_1+n_2)}{n_1 n_2}} \exp\left\{ \frac{-\lambda_1(n_1+n_2)}{n_1 n_2} \left(T_1 + \frac{n_2 T_3}{n_1+n_2}\right)^2 \right\}.$$

After some simplification we are left with a single integral only in λ_1. Since its integrand is proportional to a gamma density after a change of variable it can be evaluated in closed-form. With the substitution $z = -\lambda_1$ this integral is proportional to

$$\begin{aligned}
&\int_{-\infty}^{0} (-\lambda_1)^{\frac{n_1+n_2+2}{2}} \exp\left\{ -\frac{\lambda_1 (T_1^2 n_1 + T_1^2 n_2 + T_3^2 n_2 + 2T_1T_3n_2 - T_2 n_1 n_2)}{n_1 n_2} \right\} d\lambda_1 \\
&= \int_{0}^{\infty} z^{\alpha-1} e^{-\beta z} dz \\
&= \frac{\Gamma(\alpha)}{\beta^\alpha}
\end{aligned}$$

where

$$\alpha = \frac{n_1 + n_2 + 4}{2}$$

and

$$\beta = -\frac{T_1^2 n_1 + T_1^2 n_2 + T_3^2 n_2 + 2T_1 T_3 n_2 - T_2 n_1 n_2}{n_1 n_2}.$$

Note in the above integration we assumed that $\beta > 0$ which is in fact always the case in practice (provided that $n_1 > 1$ or $n_2 > 1$). To see this recall that

$$\begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = \begin{bmatrix} -\sum y_2 \\ \sum y_1^2 + \sum y_2^2 \\ \sum y_1 + \sum y_2 \end{bmatrix}.$$

It follows that

$$\begin{aligned} \beta &= -\frac{T_1^2 n_1 + T_1^2 n_2 + T_3^2 n_2 + 2T_1 T_3 n_2 - T_2 n_1 n_2}{n_1 n_2} \\ &= \sum (y_1 - \bar{y}_1)^2 + \sum (y_2 - \bar{y}_2)^2 \\ &= (n_1 + n_2 - 2)\, s_p^2. \end{aligned} \tag{B.8}$$

where s_p^2 is the pooled sample variance. Therefore the flat improper prior yields a proper posterior distribution provided that $n_1 > 1$ or $n_2 > 1$.

In summary the denominator of posterior distribution $(B.7)$ is given in closed-form as

$$\int_\gamma f(\mathbf{x}|\psi, \lambda_1, \lambda_2)\, d\gamma = 2^{\frac{n_1+n_2}{2}} \frac{4\pi}{\sqrt{n_1 n_2}} \frac{\Gamma\left(\frac{n_1+n_2+4}{2}\right)}{\left\{(n_1 + n_2 - 2)\, s_p^2\right\}^{\frac{n_1+n_2+4}{2}}}.$$

The numerator of (2.15) is given as

$$\int_{\psi_1}^{\psi_2} \int_\lambda f(\mathbf{x}|\psi, \lambda_1, \lambda_2)\, d\lambda d\psi = \int_{\psi_1}^{\psi_2} \int_{-\infty}^{\infty} \int_{-\infty}^{0} f(\mathbf{x}|\psi, \lambda_1, \lambda_2)\, d\lambda_1 d\lambda_2 d\psi.$$

We first integrate w.r.t. λ_2 in closed-form as before separating out the portions of the resulting integrand which do not depend upon ψ yields

$$\int_{-\infty}^{0} (-2\lambda_1)^{\frac{n_1+n_2}{2}} \sqrt{\frac{-4\pi\lambda_1}{n_1+n_2}} \exp\left\{\lambda_1 T_2 - \frac{\lambda_1 T_3^2}{n_1+n_2}\right\} I_\psi(\lambda_1)\, d\lambda_1.$$

where the inner integral w.r.t. ψ is given as

$$I_\psi(\lambda_1) = \int_{\psi_1}^{\psi_2} \exp\left\{\frac{n_1 n_2}{4\lambda_1(n_1+n_2)}\psi^2 + \left(T_1 + \frac{n_2 T_3}{n_1+n_2}\right)\psi\right\} d\psi.$$

The integrand in the inner integral is proportional to a normal density in ψ and as such can be evaluated in closed-form. Recall that for $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} P(x_1 \leq X \leq x_2) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\} dx \\ &= \Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right) \end{aligned}$$

where $\Phi(\cdot)$ is the CDF for the standard normal density. As a result

$$\int_{x_1}^{x_2} \exp\left\{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x\right\} dx = \sqrt{2\pi}\sigma \exp\left(\frac{\mu^2}{2\sigma^2}\right)\left\{\Phi\left(\frac{x_2-\mu}{\sigma}\right) - \Phi\left(\frac{x_1-\mu}{\sigma}\right)\right\}.$$

To evaluate inner integral

$$I_\psi(\lambda_1) = \int_{\psi_1}^{\psi_2} \exp\left[\frac{n_1 n_2}{4\lambda_1(n_1+n_2)}\psi^2 + \left(T_1 + \frac{n_2 T_3}{n_1+n_2}\right)\psi\right] d\psi.$$

we set-up the following correspondences:

$$\frac{n_1 n_2}{4\lambda_1(n_1+n_2)} \equiv -\frac{1}{2\sigma^2}$$

and

$$T_1 + \frac{n_2 T_3}{n_1+n_2} \equiv \frac{\mu}{\sigma^2}$$

which yields

$$\sigma^2 \equiv \frac{-2\lambda_1 (n_1+n_2)}{n_1 n_2}$$

and

$$\begin{aligned}\mu &\equiv \frac{-2\lambda_1 (n_1+n_2)}{n_1 n_2}\left(T_1 + \frac{n_2 T_3}{n_1+n_2}\right) \\ &\equiv \frac{-2\lambda_1 (n_1 T_1 + n_2 T_1 + n_2 T_3)}{n_1 n_2}.\end{aligned}$$

Therefore, we have that

$$\begin{aligned}&\int_{\psi_1}^{\psi_2} \exp\left\{\frac{n_1 n_2}{4\lambda_1 (n_1+n_2)}\psi^2 + \left(T_1 + \frac{n_2 T_3}{n_1+n_2}\right)\psi\right\} d\psi \\ &= \sqrt{\frac{-4\pi\lambda_1 (n_1+n_2)}{n_1 n_2}} \exp\left\{\frac{\left[\frac{-2\lambda_1(n_1T_1+n_2T_1+n_2T_3)}{n_1n_2}\right]^2}{-\frac{4\lambda_1(n_1+n_2)}{n_1n_2}}\right\}(\Phi_2 - \Phi_1)\end{aligned}$$

where

$$\Phi_1 = \Phi\left(\frac{\psi_1 + \frac{2\lambda_1(n_1T_1+n_2T_1+n_2T_3)}{n_1n_2}}{\sqrt{-\frac{2\lambda_1(n_1+n_2)}{n_1n_2}}}\right)$$

and

$$\Phi_2 = \Phi\left(\frac{\psi_2 + \frac{2\lambda_1(n_1T_1+n_2T_1+n_2T_3)}{n_1n_2}}{\sqrt{-\frac{2\lambda_1(n_1+n_2)}{n_1n_2}}}\right).$$

Simplifying this expression and doing a sequence of change of variables $z = -\lambda_1$ and then $\lambda_1 = z$ yields :

$$\begin{aligned}&\int_{\psi_1}^{\psi_2}\int_{-\infty}^{\infty}\int_{-\infty}^{0} f(\mathbf{x}|\psi,\lambda_1,\lambda_2)\, d\lambda_1 d\lambda_2 d\psi \\ &= 2^{\frac{n_1+n_2}{2}} \frac{4\pi}{\sqrt{n_1 n_2}} \int_0^{\infty} \lambda_1^{\frac{n_1+n_2+2}{2}} \exp\left\{-\lambda_1 (n_1+n_2-2) s_p^2\right\}(\Phi_2 - \Phi_1)\, d\lambda_1.\end{aligned}$$

Thus the two-sided posterior probability under the flat prior is

$$P\left(\psi_1 < \psi < \psi_2 | \mathbf{x}\right) = \frac{\int_0^\infty \lambda_1^{\frac{n_1+n_2+2}{2}} \exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\left(\Phi_2-\Phi_1\right)d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2+4}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2+4}{2}}}}.$$

B.4.2. Two-Sided Bayesian Equivalence Procedure: Jeffreys' Prior. As before we first consider the denominator in (2.15). Since Jeffereys' prior is flat in ψ and λ_2 integration over these parameters is performed in closed-form like in the previous section. Recall the final univariate integral in λ_1 to calculate the denominator for (2.15) using a flat prior;

$$\int_\gamma f\left(\mathbf{x}|\psi,\lambda_1,\lambda_2\right)d\gamma = 2^{\frac{n_1+n_2}{2}} \frac{4\pi}{\sqrt{n_1 n_2}} \int_{-\infty}^0 \left(-\lambda_1\right)^{\frac{n_1+n_2+2}{2}} \exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}d\lambda_1.$$

For the Jeffreys' prior computation we need to simply perform a sequence of change of variables $z = -\lambda_1$ and then $\lambda_1 = z$ and then multiply the resulting integrand by Jeffereys' prior

$$\pi\left(\psi,\lambda_1,\lambda_2\right) = \sqrt{\frac{\left(n_1^2 n_2 + n_1 n_2^2\right)}{8\lambda_1^4}}.$$

The resulting integral is proportional to the integral of a gamma density and may be evaluated in closed-form to yield

$$\int_\gamma f\left(\mathbf{x}|\psi,\lambda_1,\lambda_2\right)d\gamma = 2^{\frac{n_1+n_2+1}{2}} \sqrt{n_1+n_2}\,\pi \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\left\{\left(n_1+n_2-2\right)s_p^2\right\}^{\frac{n_1+n_2}{2}}}.$$

Here we assumed that the gamma scale parameter as in $(B.8)$ is positive. Hence the Jeffrey's prior also yields a proper posterior distribution when $n_1 > 1$ or $n_2 > 1$.

With regards to the numerator recall the final univariate integral in λ_1 to calculate the numerator for (2.15) using a flat prior;

$$\int_{\psi_1}^{\psi_2}\int_\lambda f\left(\mathbf{x}|\psi,\lambda_1,\lambda_2\right)d\lambda d\psi \propto \int_0^\infty \lambda_1^{\frac{n_1+n_2+2}{2}} \exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\left(\Phi_2-\Phi_1\right)d\lambda_1.$$

Multiplication of the above integrand by Jeffereys' prior results in the following numerator expression

$$\int_{\psi_1}^{\psi_2}\int_{\lambda} f\left(\mathbf{x}|\psi,\lambda_1,\lambda_2\right)\pi\left(\psi,\lambda_1,\lambda_2\right)d\lambda d\psi$$
$$= 2^{\frac{n_1+n_2+1}{2}}\sqrt{n_1+n_2}\pi\int_0^{\infty}\lambda_1^{\frac{n_1+n_2-2}{2}}\exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\left(\Phi_2-\Phi_1\right)d\lambda_1.$$

In summary the posterior probability of the alternative hypothesis under Jeffereys' prior is

$$P\left(\psi_1<\psi<\psi_2|\mathbf{x}\right)=\frac{\int_0^{\infty}\lambda_1^{\frac{n_1+n_2-2}{2}}\exp\left\{-\lambda_1\left(n_1+n_2-2\right)s_p^2\right\}\left(\Phi_2-\Phi_1\right)d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2}{2}}}}.$$

B.4.3. Double One-Sided Bayesian Equivalence Procedure: Flat Prior. To evaluate integrals (2.16) and (2.17) we can first integrate w.r.t. λ_2, as in the two-sided Bayesian equivalence calculations to yield

$$\int_{\psi_1}^{\infty}\int_{\lambda} f\left(\mathbf{x}|\psi,\lambda_1,\lambda_2\right)d\lambda_1 d\lambda_2 d\psi=\int_{-\infty}^{0}\left(-2\lambda_1\right)^{\frac{n_1+n_2}{2}}\sqrt{\frac{-4\pi\lambda_1}{n_1+n_2}}e^{\lambda_1 T_2-\frac{\lambda_1 T_3^2}{n_1+n_2}}I_{\psi}^1\left(\lambda_1\right)d\lambda_1$$

with inner integral

$$I_{\psi}^1\left(\lambda_1\right)=\int_{\psi_1}^{\infty}\exp\left\{\frac{n_1 n_2}{4\lambda_1\left(n_1+n_2\right)}\psi^2+\left(T_1+\frac{n_2 T_3}{n_1+n_2}\right)\psi\right\}d\psi$$

and

$$\int_{-\infty}^{\psi_2}\int_{\lambda} f\left(\mathbf{x}|\psi,\lambda_1,\lambda_2\right)d\lambda_1 d\lambda_2 d\psi=\int_{-\infty}^{0}\left(-2\lambda_1\right)^{\frac{n_1+n_2}{2}}\sqrt{\frac{-4\pi\lambda_1}{n_1+n_2}}e^{\lambda_1 T_2-\frac{\lambda_1 T_3^2}{n_1+n_2}}I_{\psi}^2\left(\lambda_1\right)d\lambda_1$$

with inner integral

$$I_{\psi}^2\left(\lambda_1\right)=\int_{-\infty}^{\psi_2}\exp\left[\frac{n_1 n_2}{4\lambda_1\left(n_1+n_2\right)}\psi^2+\left(T_1+\frac{n_2 T_3}{n_1+n_2}\right)\psi\right]d\psi.$$

Applying the Gaussian identity to the inner integrals yields

$$I_{\psi}^{1}(\lambda_1) = \sqrt{2\pi}\sqrt{\frac{-2\lambda_1(n_1+n_2)}{n_1 n_2}} \exp\left\{\frac{\left[\frac{2\lambda_1(n_1T_1+n_2T_1+n_2T_3)}{n_1n_2}\right]^2}{-\frac{4\lambda_1(n_1+n_2)}{n_1n_2}}\right\}\Phi_2$$

$$I_{\psi}^{2}(\lambda_1) = \sqrt{2\pi}\sqrt{\frac{-2\lambda_1(n_1+n_2)}{n_1 n_2}} \exp\left\{\frac{\left[\frac{2\lambda_1(n_1T_1+n_2T_1+n_2T_3)}{n_1n_2}\right]^2}{-\frac{4\lambda_1(n_1+n_2)}{n_1n_2}}\right\}(1-\Phi_1).$$

After a sequence of change of variables $z = -\lambda_1$ and then $\lambda_1 = z$ we then have

$$P(\psi < \psi_2|\mathbf{x}) = \frac{\int_0^\infty \lambda_1^{\frac{n_1+n_2+2}{2}} \exp\left\{-\lambda_1(n_1+n_2-2)s_p^2\right\}\Phi_2 d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2+4}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2+4}{2}}}}$$

and

$$P(\psi > \psi_1|\mathbf{x}) = \frac{\int_0^\infty \lambda_1^{\frac{n_1+n_2+2}{2}} \exp\left\{-\lambda_1(n_1+n_2-2)s_p^2\right\}(1-\Phi_1) d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2+4}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2+4}{2}}}}.$$

B.4.4. Double One-Sided Bayesian Equivalence Procedure: Jeffreys' Prior. The computations in this setting follow immediately from our previous results to yield

$$P(\psi < \psi_2|\mathbf{x}) = \frac{\int_0^\infty \lambda_1^{\frac{n_1+n_2-2}{2}} \exp\left\{-\lambda_1(n_1+n_2-2)s_p^2\right\}\Phi_2 d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2}{2}}}}$$

and

$$P(\psi > \psi_1|\mathbf{x}) = \frac{\int_0^\infty \lambda_1^{\frac{n_1+n_2-2}{2}} \exp\left\{-\lambda_1(n_1+n_2-2)s_p^2\right\}(1-\Phi_1) d\lambda_1}{\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\left\{(n_1+n_2-2)s_p^2\right\}^{\frac{n_1+n_2}{2}}}}.$$

BIBLIOGRAPHY

[1] Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions*, 9th Edition. New York : Dover.

[2] Altman, D.G. and Bland, J.M. (1995). Absence of Evidence is not Evidence of Absence. *British Medical Journal*. 311, 485.

[3] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Berlin: Springer-Verlag.

[4] Berger, R.L. (1982) . Multiparameter Hypothesis Testing and Acceptance Sampling. *Technometrics*, 24, 295-300.

[5] Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, New York: John Wiley.

[6] Butler, R.W. (2007) . *Saddlepoint Approximations with Applications*. New York: Cambridge University Press.

[7] Casella, G. and Berger, R. L. (2002). *Statistical Inference*, 2nd Edition, . Pacific Grove CA: Duxbury Press.

[8] Harville, D.A. (1997). Matrix Algebra from a Statistician's Perspective, New York: Springer-Verlag.

[9] Henze, N. and Meintanis, S.G. (2005). Recent and Classical Tests for Exponentiality: A Partial Review with Comparisons. *Metrika*, 61, 29–45.

[10] Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, 2nd Edition, New York: John Wiley.

[11] Lugannani, R. and Rice, S.O. (1980). Saddlepoint Approximations for the Distribution of Sums of Independent Random Variables. *Adv. Appl. Prob.* 12, 475-490.

[12] Paige, R. L. and Trindade, A.A. (2008). Practical Small Sample Inference for Single Lag Subset Autoregressive Models. *Journal of Statistical Planning and Inference.* 138, 1934-1949.

[13] Paige, R.L., Trindade, A.A. and Fernando, P. H. (2009). Saddlepoint-Based Bootstrap Inference for Quadratic Estimating Equations. *Scand. J. Stat.*. 36, 98-111.

[14] Romano, J. P. (2005) . Optimal Testing of Equivalence Hypotheses. The Annals of Statistics, 33, 1036-1047.

[15] Tukey, J.W. (1991) .The Philosophy of Multiple Comparisons. Statistical Science, 6, 100-116.

[16] Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*, 2nd Edition, Boca Raton : Chapman & Hall/CRC.

VITA

Renren Zhao attended Chongqing University, China at 2004 and completed his bachelor's degree in Economics and Finance in 2008. Renren entered Missouri University of Science and Technology for graduate studies in 2008. He received his MS degree in Applied Mathematics with Statistics emphasis in December 2011 and received his PhD degree in Statistics in May 2015 at Missouri University of Science and Technology.