The Canadian Journal of Statistics
Vol. , 2019, Pages 1
La revue canadienne de statistique

Supplementary Material for A Directional Look at F-tests

A. McCormack[1]*, N. Reid [2], N. Sartori [3] and S.-A. Theivendran [2]

[1]*Department of Statistical Science, Duke University*
[2]*Department of Statistical Sciences, University of Toronto*
[3]*Dipartimento di Scienze Statistiche, Università di Padova*

1. INTRODUCTION

The tests examined in McCormack et al. (2019) use a directional argument proposed in Fraser & Massam (1985) and developed further, using higher order approximation theory, in Davison et al. (2014) and Fraser, Reid & Sartori (2016). For completeness we provide a summary of directional tests in §S.1, and provide the formulae needed for the examples in §S.2 - S.4.

S.2 Directional testing

S.2.1 A model on $\mathbb{R}^p$

Suppose our model for $y = (y_1, \ldots, y_n)$ is a linear exponential family

$$f(y; \theta) = \exp[\varphi^{\mathrm{T}}(\theta) u(y) - \kappa\{\varphi(\theta)\}] d(y), \quad \varphi \in \mathbb{R}^p \tag{1}$$

with sufficient statistic u and canonical parameter φ. Inference for φ is based on the marginal distribution of u, which is again an exponential family

$$f(u; \theta) = \exp[\varphi^{\mathrm{T}}(\theta) u - \kappa\{\varphi(\theta)\}] \tilde{d}(u). \tag{2}$$

The function $\tilde{d}(\cdot)$ is obtained by marginalizing (1) and may not be available explicitly, but the saddlepoint approximation to the density of u has relative error $O(n^{-3/2})$ in continuous models:

$$f_{SP}(u; \theta) = \frac{e^{k/n}}{(2\pi)^{p/2}} |\hat{\jmath}|^{-1/2} \exp[\ell\{\varphi(\theta); u\} - \ell\{\varphi(\hat{\theta}); u\}], \tag{3}$$

where $\ell(\varphi; u) = \varphi^{\mathrm{T}} u - \kappa(\varphi)$ is the log-likelihood function, $\hat{\jmath} = -\partial^2 \ell(\hat{\varphi})/\partial\varphi\partial\varphi^{\mathrm{T}}$ is the observed Fisher information, $\hat{\theta}$ is the maximum likelihood estimator, and $\exp(k/n)/(2\pi)^{p/2}$ is an approximation to the normalizing constant.

Directional tests take as their starting point the approximation (3). If the originating model is not in the exponential family, then an approximation to it, the tangent exponential model, is used instead. The construction of the tangent exponential model and its saddlepoint approximation are described in the Appendix of Fraser, Reid & Sartori(2016); see in particular Eq. (A2). Since the examples in the current paper are all exponential family models, this step is not needed.

S.2.2 Nuisance parameters

Suppose that the parameter d-dimensional parameter θ in our original model can be partitioned as $\theta = (\psi, \lambda)$, where ψ is a d-dimensional parameter of interest and λ is a $p - d$-dimensional nuisance parameter. Denote by $\hat{\lambda}_\psi$ the constrained maximum likelihood estimator of λ when ψ is fixed.

In the special case that $\theta = \varphi$, so that the parameter of interest is a sub-vector of the canonical parameter, and if the original model is a full exponential family, then

$$f(u_1, u_2; \psi, \lambda) \propto \exp\{\psi^{\mathrm{T}} u_1 + \lambda^{\mathrm{T}} u_2 - \kappa(\psi, \lambda)\}\tilde{d}(u), \tag{4}$$

and the conditional distribution of u_1 given u_2 is free of λ. The saddlepoint approximation can be used again to give an accurate approximation to the conditional density. Directional testing for models of this form are developed and illustrated in Davison et al. (2014).

If the parameter of interest is not a linear function of the canonical parameter, in which case we write $\psi = \psi(\varphi)$, such a reduction by conditioning is not available. None-the-less, it can be verified that there is a unique variable that measures ψ, and that this variable is obtained by constraining the sufficient statistic u to the d-dimensional sample space obtained by fixing the constrained maximum likelihood estimate of the nuisance parameter to its observed value. The saddlepoint approximation to the density of this variable is

$$h(s; \psi) = \frac{\exp(k'/n)}{(2\pi)^{d/2}} \exp\{\ell(\hat{\varphi}_\psi; s) - \ell(\hat{\varphi}(s)\}|\hat{J}_{\varphi\varphi}|^{-1/2}|\tilde{J}_{(\lambda\lambda)}|^{1/2}, \quad s \in L_\psi, \tag{5}$$

where L_ψ is the plane in the sample space with $\hat{\lambda}_\psi$ fixed, so that h above is a density on $\mathbb{R}^d$. In (5), $s = u - u^0$ is a centred version of the sufficient statistic, and $\ell(\varphi; s) = \varphi^{\mathrm{T}} s + \ell^0(\varphi)$ is an exponential tilt of the observed log-likelihood function $\ell(\theta; y^0)$ in the original model. The centering is described in detail in Davison et al. (2014, §3.1) and assumed in Fraser, Reid & Sartori (2016). The determinants in (5) are:

$$|\hat{J}_{\varphi\varphi}| = |J_{\varphi\varphi}\{\hat{\varphi}(s)\}| = |-\partial^2\ell(\varphi; s)/(\partial\varphi\partial\varphi^{\mathrm{T}})\big|_{\varphi=\hat{\varphi}(s)}, \tag{6}$$

and

$$|\tilde{J}_{(\lambda\lambda)}| = |J_{(\lambda\lambda)}(\hat{\varphi}_\psi)| = \left|-\frac{\partial^2\ell(\hat{\varphi}_\psi; s)}{\partial\lambda\partial\lambda^{\mathrm{T}}}\right| \left|\frac{\partial\varphi(\hat{\theta}_\psi)}{\partial\lambda}\right|^{-2}. \tag{7}$$

The second determinant is not needed when the parameter of interest is linear in φ. However, it turns out to be independent of t in the examples in §2.1, 2.3 and 3.2, even though the parameter of interest is not linear in the canonical parameter. In these cases the canonical parameter is a linear function of the nuisance parameter and so (7) does not depend on t.

S.2.3 Directional testing

The directional test of the hypothesis $\psi(\varphi) = \psi$ is carried out in L_ψ by finding the line that joins s^0 with the value of s, call it s_ψ, that would give $\hat{\varphi}_\psi$ as the maximum likelihood estimate of the parameter. The observed value s^0 gives $\hat{\varphi}$ as the maximum likelihood estimate. The p-value is computed as the probability of s being larger than the observed value s_0, on the line between the two values s_ψ and s^0. Another way to describe it is that we measure the magnitude of the vector $s_0 - s_\psi$, in L_ψ, conditional on its direction. This gives a one-dimensional measure of how much "larger" the observed value s^0 is than would be expected under the hypothesis. We parameterize this line in the sample space by t, and because we center the sufficient statistic so that $s^0 = 0$, the line is simply $s(t) = s_\psi + t(s^0 - s_\psi) = (1-t)s_\psi$. The directional p-value is then

$$p(\psi) = \frac{\int_1^\infty t^{d-1} h(t;\psi)dt}{\int_0^\infty t^{d-1} h(t;\psi)dt}, \tag{8}$$

where $h(t;\psi) = h\{s(t);\psi\}$ using (4), and the inflation factor t^{d-1} comes from the Jacobian of the transformation to polar coordinates.

S.3 Ratio of exponential rates

S.3.1 Finding the integrand of the directional p-value

We consider the directional test for the null hypothesis H_ψ that $\theta_1/\theta_2 = \psi$ where $y_{ij} \sim \exp(\theta_j)$, $j = 1, 2$ $i = 1, ..., n_i$. Note that this test is slightly different than the test performed Davison et al. (2014) as here we are testing the ratio of rates. Under H_ψ the constrained MLE is

$$\hat{\theta}_{1\psi} = \frac{n\psi}{u_1\psi + u_2},$$
$$\hat{\theta}_{2\psi} = \frac{n}{u_1\psi + u_2},$$

where $n = n_1 + n_2$ and $u_j = \Sigma_i y_{ij}$. By solving the score equation it is found that the value of u_j that has $\hat{\theta}_\psi$ as its global MLE is $u_{j\psi} = n_j/\hat{\theta}_{j\psi}$. The line between u_ψ and the observed value of u, $u^0 = (u_1^0, u_2^0)$ is

$$\begin{aligned} u_1(t) =& \frac{n_1}{n}(u_1^0 + \frac{u_2^0}{\psi}) + t\big\{u_1^0 - \frac{n_1}{n}(u_1^0 + \frac{u_2^0}{\psi})\big\} = u_{1\psi} + t(u_1^0 - u_{1\psi}), \\ u_2(t) =& \frac{n_2}{n}(u_1^0\psi + u_2^0) + t\big\{u_2^0 - \frac{n_2}{n}(u_1^0\psi + u_2^0)\big\} = u_{2\psi} + t(u_2^0 - u_{2\psi}), \end{aligned}$$

so that

$$\exp[\ell\{\hat{\varphi}_\psi; s(t)\} - \ell\{\hat{\varphi}; s(t)\}] \propto \exp\left\{-\frac{n_1 u_1(t)}{u_{1\psi}} - \frac{n_2 u_2(t)}{u_{2\psi}}\right\} u_1(t)^{n_1} u_2(t)^{n_2}.$$

The determinant of the Hessian of the negative log-likelihood with respect to $\varphi = (\theta_1, \theta_2)$ is the determinant of the Hessian of $-n_1\log(\theta_1) - n_2\log(\theta_2)$. This is easily found to be $|J_{\varphi\varphi}| = n_1 n_2/(\theta_1\theta_2)^2$. We have that $\hat{\theta}_i(t) = n_i/u_i(t)$ and thus

$$|J_{\varphi\varphi}|^{-\frac{1}{2}} \exp[\ell\{\hat{\varphi}_\psi; s(t)\} - \ell\{\hat{\varphi}; s(t)\}] \propto \exp\left\{-\frac{n_1 u_1(t)}{u_{1\psi}} - \frac{n_2 u_2(t)}{u_{2\psi}}\right\} u_1(t)^{n_1-1} u_2(t)^{n_2-1}.$$

Next we check the nuisance parameter adjustment term. Formulating the log-likelihood in terms of nuisance parameter $\theta_2 = \lambda$ we get

$$\ell(\psi, \lambda) = -u_1(t)\psi\lambda - u_2(t)\lambda + \ell^0(\psi, \lambda).$$

It is clear that the second derivative of this function with respect to λ only contains terms that do not involve t. The dimension of our parameter of interest in this case is $d = 1$ so that $t^{d-1} = 1$. The sufficient statistic $u(t)$ is viable as long as it remains positive. As a result, t_{max} is the largest t such that both $u_1(t)$ and $u_2(t)$ are non-negative. Notice that $u_1^0 - u_{1\psi} > 0$ if and only if $\psi\bar{y}_1 \geq$

$\bar{y}_2$ and likewise $u_2^0 - u_{2\psi} > 0$ if and only if $\psi\bar{y}_1 \leq \bar{y}_2$. so that

$$t_{max} = \frac{u_{1\psi}}{u_{1\psi} - u_1^0}\mathrm{I}(\psi\bar{y}_1 \leq \bar{y}_2) + \frac{u_{2\psi}}{u_{2\psi} - u_2^0}\mathrm{I}(\psi\bar{y}_1 \geq \bar{y}_2).$$

We let a_j be equal to the quantity $u_{j\psi}/(u_{j\psi} - u_j^0)$. It can be shown that $(u_{1\psi} - u_1^0)/(u_{2\psi} - u_2^0) = -1/\psi$ and $u_{2\psi}/u_{1\psi} = \psi n_2/n_1$, which shows that $n_1 a_2 + n_2 a_1 = 0$. Now

$$\exp\left\{ -\frac{n_1 u_1(t)}{u_{1\psi}} - \frac{n_2 u_2(t)}{u_{2\psi}} \right\} \propto \exp\left\{ t\big(\frac{n_1}{a_1} + \frac{n_2}{a_2}\big) \right\} = 1,$$

and thus the directional p-value is given by

$$p(\psi) = \frac{\int_1^{t_{max}} u_1(t)^{n_1-1} u_2(t)^{n_2-1} dt}{\int_0^{t_{max}} u_1(t)^{n_1-1} u_2(t)^{n_2-1} dt} = \frac{\int_1^{t_{max}} (1 - \frac{t}{a_1})^{n_1-1}(1 - \frac{t}{a_2})^{n_2-1} dt}{\int_0^{t_{max}} (1 - \frac{t}{a_1})^{n_1-1}(1 - \frac{t}{a_2})^{n_2-1} dt}. \tag{9}$$

S.3.2 Making a change of variables

Assume that $t_{max} = a_1$, so that $\psi\bar{y}_1/\bar{y}_2 \leq 1$. The numerator of (9) can be written as

$$p_{num} = \int_1^{a_1} (1 - \frac{t}{a_1})^{n_1+n_2} \left(\frac{1 - \frac{t}{a_2}}{1 - \frac{t}{a_1}} \right)^{n_2-1} (1 - \frac{t}{a_1})^{-2} dt.$$

Make the change of variables $x = (1 - t/a_2)/(1 - t/a_1)$, we then get that

$$p_{num} = k_1 \int_{\frac{1-1/a_2}{1-1/a_1}}^{\infty} x^{n_2-1} \left(\frac{a_1 - a_2}{a_1 - a_2 x} \right)^{-n_1-n_2} dx = k_2 \int_{\frac{1-1/a_2}{1-1/a_1}}^{\infty} x^{n_2-1} \big(1 - \frac{a_2 x}{a_1}\big)^{-n_1-n_2} dx.$$

We know from a previous calculation that $a_2/a_1 = -n_2/n_1$. Furthermore,

$$\frac{1 - 1/a_1}{1 - 1/a_2} = \frac{(a_2 - 1)a_1}{(a_1 - 1)a_2} = -\frac{u_2^0(u_{1\psi} - u_1^0)n_1}{u_1^0(u_{2\psi} - u_2^0)n_2} = \frac{\bar{y}_2}{\bar{y}_1\psi}.$$

Therefore we see that

$$p_{num} = k_2 \int_{\frac{\bar{y}_2}{\bar{y}_1\psi}}^{\infty} x^{n_2-1} \big(1 + \frac{n_2}{n_1} x\big)^{-n_1-n_2} dx. \tag{10}$$

Now if we perform the same change of variables on the integral in the denominator of the directional p-value the same normalizing constant k_2 will be produced. The bounds of the integral will change as follows

$$\int_0^{a_1} \Rightarrow \int_1^{\infty}.$$

The integrand in (10) is the density of a $F(2n_2, 2n_1)$ random variable up to a normalizing constant. Thus if $W \sim F(2n_2, 2n_1)$

$$p(\psi) = \frac{\mathrm{P}_W(W > \frac{\bar{y}_2}{\bar{y}_1\psi})}{\mathrm{P}_W(W > 1)}.$$

Similarly when $\bar{y}_1\psi/\bar{y}_2 \geq 1$ we get

$$p(\psi) = \frac{\mathrm{P}_W(W < \frac{\bar{y}_1\psi}{\bar{y}_2})}{\mathrm{P}_W(W < 1)}.$$

Consequently, the directional test is identical to that of the appropriate F-test. In summary, the directional p-value is

$$p(\psi) = \mathrm{I}\big(\bar{y}_1\psi < \bar{y}_2\big)\frac{\mathrm{P}_W(W > \frac{\bar{y}_2}{\bar{y}_1\psi})}{\mathrm{P}_W(W > 1)} + \mathrm{I}\big(\bar{y}_1\psi > \bar{y}_2\big)\frac{\mathrm{P}_W(W < \frac{\bar{y}_1\psi}{\bar{y}_2})}{\mathrm{P}_W(W < 1)}.$$

S.4 Ratio of normal variances

Suppose that $y_{ij} \sim N(\mu_i, \sigma_i^2)$ are independent random variables for $i = 1, 2$ and $j = 1 \ldots, n_i$, and we wish to test $H_\psi : \sigma_1^2/\sigma_2^2 = \psi$. A computation very similar to that given in the previous section shows that the integrand of the directional p-value for testing H_ψ is

$$(1 - tb_1)^{(n_1-3)/2}(1 - tb_2)^{(n_2-3)/2}, \tag{11}$$

with $b_i = (\hat{\sigma}_{i\psi}^2 - v_i^2)/\hat{\sigma}_{i\psi}^2$. The biased within-group sample variances are v_i^2 while $\hat{\sigma}_{i\psi}^2$ is the constrained maximum likelihood estimator for σ_i^2. There is a clear resemblance between the integrands of (9) and (11). The same change of variables used for the exponential rates example can be used here with the only minor difference being that we set $a_i = 1/b_i$ so that (11) equals $(1 - t/a_1)^{(n_1-3)/2}(1 - t/a_2)^{(n_2-3)/2}$. All the bounds of the integrals will change in the same way as in the previous example. In particular, we find that

$$\frac{1 - 1/a_2}{1 - 1/a_1} = \frac{1 - b_2}{1 - b_1} = \frac{\hat{\sigma}_{1\psi}^2 v_2^2}{\hat{\sigma}_{2\psi}^2 v_1^2} = \frac{\psi v_2^2}{v_1^2}.$$

If $v_1^2 \leq \psi v_2^2$ so that $t_{max} = 1/a_1$ we get that

$$p_{num} = \int_{\psi v_2^2/v_1^2}^{\infty} x^{\frac{n_2-1}{2}-1}\big(1 + \frac{n_2}{n_1}x\big)^{-\frac{(n_1-1)+(n_2-1)}{2}}\, dt.$$

This integrand is not quite the density of a F-distribution due to the factor of n_2/n_1 appearing instead of $(n_2 - 1)/(n_1 - 1)$. To fix this we make the additional change of variables in both p_{num} and p_{den} from x to
$\{n_1(n_2 - 1)\}/\{n_2(n_1 - 1)\}x$. If $W \sim F(n_2 - 1, n_1 - 1)$ and s_i^2 are the unbiased sample variances we get the desired result that

$$p(\psi) = \frac{P_W\big(W > \frac{\psi s_2^2}{s_1^2}\big)}{P_W\big(W > \frac{n_2(n_1-1)}{n_1(n_2-1)}\big)}.$$

The case where $v_1^2 > \psi v_2^2$ is handled similarly.

S.5 Linear regression with linear constraints

S.5.1 Calculating $\exp[\ell\{\hat{\varphi}(0); s(t)\} - \ell\{\hat{\varphi}(t); s(t)\}]$

Let $y_i \sim N(x_i^{\mathrm{T}}\beta, \sigma^2)$, $i = 1, \ldots, n$, where y_i^0 are realizations of y_i and all of the y_i's are independent. Both x_i and β are $p \times 1$ vectors and σ^2 is an unknown nuisance parameter. We form the

$n \times p$ matrix X by taking i'th row of X to be x_i. Here we wish to test $H_\psi : A\beta = \psi$ using the directional test. The matrix A has dimension $d \times p$ and is of rank d which ensures that the linear constraint is not redundant. The constrained maximum likelihood estimator for β under H_ψ can be found using Lagrange multipliers and is given by

$$\begin{aligned}\hat{\beta}_\psi &= \hat{\beta} - (X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\{A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\}^{-1}(A\hat{\beta} - \psi)\\ &= \hat{\beta} - \frac{1}{2}(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\hat{\lambda}.\end{aligned}$$

The Lagrange multiplier equation used to find this constrained MLE also yields

$$\begin{aligned}\hat{\beta}_\psi^{\mathrm{T}}X^{\mathrm{T}}(y - X\hat{\beta}_\psi) &= \frac{1}{2}\psi^{\mathrm{T}}\hat{\lambda}\\ &= \psi^{\mathrm{T}}\{A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\}^{-1}(A\hat{\beta} - \psi).\end{aligned}$$

The constrained MLE for σ^2 is just the average sum of squared error under $\hat{\beta}_\psi$. The log-likelihood in this situation is

$$\begin{aligned}\ell(\beta, \sigma^2) &= -\frac{y^{\mathrm{T}}y}{2\sigma^2} + \frac{y^{\mathrm{T}}X\beta}{\sigma^2} - \frac{\beta^{\mathrm{T}}X^{\mathrm{T}}X\beta}{2\sigma^2} - \frac{n}{2}\log(\sigma^2)\\ &= \begin{bmatrix} y^{\mathrm{T}}y & y^{\mathrm{T}}X \end{bmatrix}\begin{bmatrix} -\frac{1}{2\sigma^2} \\ \frac{\beta}{\sigma^2} \end{bmatrix} - \kappa(\beta, \sigma^2)\\ &= u^{\mathrm{T}}(y)\varphi - \kappa(\varphi).\end{aligned}$$

The sufficient statistics here are $y^{\mathrm{T}}y$ and $y^{\mathrm{T}}X$. These sufficient statistics have unconstrained MLEs that are equal to the constrained MLE when they solve the following equations:

$$\begin{aligned}(X^{\mathrm{T}}X)\hat{\beta}_\psi &= X^{\mathrm{T}}y\\ &\text{and}\\ \frac{1}{n}(y - X\hat{\beta}_\psi)^{\mathrm{T}}(y - X\hat{\beta}_\psi) &= \hat{\sigma}_\psi^2\\ \implies \frac{1}{n}(y^{\mathrm{T}}y - \hat{\beta}_\psi^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}_\psi) &= \hat{\sigma}_\psi^2\\ y^{\mathrm{T}}y &= n\hat{\sigma}_\psi^2 + \hat{\beta}_\psi^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}_\psi.\end{aligned}$$

Thus

$$s_\psi = \begin{bmatrix} n\hat{\sigma}_\psi^2 + \hat{\beta}_\psi^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}_\psi - (y^0)^{\mathrm{T}}y^0 \\ (X^{\mathrm{T}}X)\hat{\beta}_\psi - X^{\mathrm{T}}y^0 \end{bmatrix}.$$

We define $s(t) = (1 - t)s_\psi$. Then let $u(t) = u^0 + s(t)$ so that
$\ell(\varphi; t) = u^{\mathrm{T}}(t)\varphi - \kappa(\varphi)$. We let $u_1(t)$ be the first entry of $u(t)$ and $u_2(t)$ be the remaining

entries of $u(t)$. We see that

$$\begin{aligned}
\frac{\partial}{\partial\sigma^2}\ell(\varphi;t) &= \frac{1}{\sigma^4}\Big\{\frac{u_1(t)}{2} - u_2^{\mathrm{T}}(t)\beta + \frac{\beta^{\mathrm{T}}X^{\mathrm{T}}X\beta}{2}\Big\} - \frac{n}{2\sigma^2} = 0\\
\implies \hat{\sigma}^2(t) &= \frac{1}{n}\big\{u_1(t) - 2u_2^{\mathrm{T}}(t)\hat{\beta}(t) + \hat{\beta}(t)^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}(t)\big\}.\\
\frac{\partial}{\partial\beta}\ell(\varphi;t) &= \frac{1}{\sigma^2}\big\{u_2(t) - X^{\mathrm{T}}X\beta\big\} = 0\\
\implies \hat{\beta}(t) &= (X^{\mathrm{T}}X)^{-1}u_2(t).
\end{aligned}$$

We now find formulas for the log-likelihood terms appearing in the exponent density used for the directional p-value calculation. As the directional p-value takes a ratio of such densities we can ignore all multiplicative factors not involving t in the subsequent calculations:

$$\begin{aligned}
\ell\{\hat{\varphi}(0); s(t)\} &\propto \frac{1}{\hat{\sigma}^2(0)}\big\{-\frac{1}{2}u_1(t) + u_2^{\mathrm{T}}(t)\hat{\beta}(0)\big\}\\
&\propto \frac{1}{\hat{\sigma}^2(0)}\bigg[\frac{t}{2}\big\{n\hat{\sigma}_\psi^2 + \hat{\beta}_\psi^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}_\psi - (y^0)^{\mathrm{T}}y^0\big\} + t\big\{(y^0)^{\mathrm{T}}X - \hat{\beta}_\psi X^{\mathrm{T}}X\big\}\hat{\beta}_\psi\bigg]\\
&\propto \frac{1}{\hat{\sigma}^2(0)}\bigg(\frac{nt}{2}\big[\hat{\sigma}_\psi^2 - \{(y^0)^{\mathrm{T}}y^0 - 2(y^0)^{\mathrm{T}}X\hat{\beta}_\psi + \hat{\beta}_\psi^{\mathrm{T}}X^{\mathrm{T}}X\hat{\beta}_\psi\}\big]\bigg) = 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\ell\{\hat{\varphi}(t); s(t)\} &= \frac{1}{\hat{\sigma}^2(t)}\big\{-\frac{1}{2}u_1(t) + u_2^{\mathrm{T}}(t)\hat{\beta}(t)\big\} - \frac{\hat{\beta}^{\mathrm{T}}(t)X^{\mathrm{T}}X\hat{\beta}(t)}{2\hat{\sigma}^2(t)} - \frac{n}{2}\log\{\hat{\sigma}^2(t)\}\\
&= \frac{1}{\hat{\sigma}^2(t)}\big\{-\frac{1}{2}u_1(t) + \frac{1}{2}u_2^{\mathrm{T}}(t)\hat{\beta}(t)\big\} - \frac{n}{2}\log\{\hat{\sigma}^2(t)\}\\
&= -1 - \frac{n}{2}\log\{\hat{\sigma}^2(t)\}.
\end{aligned}$$

Consequently, $\exp\big[\ell\{\hat{\varphi}(0); s(t)\} - \ell\{\hat{\varphi}(t); s(t)\}\big] = \{\hat{\sigma}^2(t)\}^{\frac{n}{2}}$.

S.5.2 Finding $|J_{\varphi\varphi}\{\hat{\varphi}(t); s(t)\}|$ and the nuisance parameter adjustment

To start we find all of the second order derivatives of $\kappa(\varphi)$ yielding

$$\begin{aligned}
\frac{\partial^2\kappa}{\partial\varphi_1^2} &= \frac{n}{2\varphi_1^2} + \frac{1}{\varphi_1^3}\sum_{i=1}^{n}(\sum_{j=2}^{p+1}X_{ij}\varphi_j)^2,\\
\frac{\partial^2\kappa}{\partial\varphi_1\partial\varphi_k} &= -\frac{1}{\varphi_1^2}\sum_{i=1}^{n}X_{ik}\sum_{j=2}^{p+1}X_{ij}\varphi_j,\\
\frac{\partial^2\kappa}{\partial\varphi_k\partial\varphi_l} &= \frac{1}{\varphi_1}\sum_{i=1}^{n}X_{ik}X_{il}.
\end{aligned}$$

Define $\bar{\varphi}$ to be the vector containing the last p entries of φ. From the second order derivatives above we find that the Hessian of the negative log-likelihood function has the form

$$J_{\varphi\varphi}\{\varphi; s(t)\} = \frac{1}{\varphi_1}\begin{bmatrix} (\frac{n}{2\varphi_1} + \frac{1}{\varphi_1^2}\bar{\varphi}^{\mathrm{T}}X^{\mathrm{T}}X\bar{\varphi}) & -\frac{1}{\varphi_1}\bar{\varphi}^{\mathrm{T}}X^{\mathrm{T}}X \\ -\frac{1}{\varphi_1}X^{\mathrm{T}}X\bar{\varphi} & X^{\mathrm{T}}X \end{bmatrix}.$$

We multiply the above matrix on the left by the matrix

$$\begin{bmatrix} 1 & 0^{\mathrm{T}} \\ 0 & (X^{\mathrm{T}}X)^{-1} \end{bmatrix}.$$

In doing this the determinant of the resulting matrix only changes by the constant $\det\{(X^{\mathrm{T}}X)^{-1}\}$. We then find that find that:

$$|J_{\varphi\varphi}\{\varphi; s(t)\}| \propto (\frac{1}{\varphi_1})^{p+1}\det\left(\begin{bmatrix} (\frac{n}{2\varphi_1} + \frac{1}{\varphi_1^2}\bar{\varphi}^{\mathrm{T}}X^{\mathrm{T}}X\bar{\varphi}) & -\frac{1}{\varphi_1}\bar{\varphi}^{\mathrm{T}}X^{\mathrm{T}}X \\ -\frac{1}{\varphi_1}\bar{\varphi} & I_p \end{bmatrix}\right).$$

The determinant above can be found by performing a cofactor expansion along the first column of the matrix and then performing a cofactor expansion along first row of the resulting minor. Fortunately, performing this cofactor expansion twice on the i'th entry of the first column and the j'th entry of the first row will produce a minor that has a row of zeros if $i \neq j$. As a result, we only have to be concerned about when $i = j$, but this case is simple as it is just minus one times the i'th entry of the last column times the i'th entry of the last row multiplied by the determinant of the identity. The bottom right entry has to be treated separately, but it clearly just returns itself in the cofactor expansion. In short

$$\begin{aligned} |J\{\varphi; s(t)\}| &\propto (\frac{1}{\varphi_1})^{p+1}\Big\{ -\frac{1}{\varphi_1^2}\bar{\varphi}^{\mathrm{T}}X^{\mathrm{T}}X\bar{\varphi} + (\frac{n}{2\varphi_1} + \frac{1}{\varphi_1^2}\bar{\varphi}^{\mathrm{T}}X^{\mathrm{T}}X\bar{\varphi})\Big\} \\ &\propto (\frac{1}{\varphi_1})^{p+2}. \end{aligned}$$

At this point we are able to construct the integrand for the directional test:

$$t^{d-1}|J\{\hat{\varphi}(t); s(t)\}|^{-\frac{1}{2}}\exp\left[\ell\{\hat{\varphi}(0); s(t)\} - \ell\{\hat{\varphi}(t); s(t)\}\right] \propto t^{d-1}\hat{\sigma}^2(t)^{\frac{n-p-2}{2}}.$$

This hypothesis is a linear hypothesis, meaning that the canonical parameter can be partitioned into the parameter of interest and a nuisance parameter. To see this, we note that $A\beta = \psi$ is equivalent to $A\beta/\sigma^2 = \psi/\sigma^2$. Consequently, H_ψ holds if and only if

$$\begin{bmatrix} A \; 2\psi \end{bmatrix}\varphi(\theta) = M\varphi(\theta) = 0. \tag{12}$$

The dimension of M is $d \times p + 1$. We define $\tilde{M}$ to be a matrix formed by adding $p + 1 - d$ rows to M in a manner so that all of the rows of $\tilde{M}$ are linearly independent. We then see that $u^{\mathrm{T}}(y)\varphi(\theta) = u^{\mathrm{T}}(y)\tilde{M}^{-1}\tilde{M}\varphi(\theta)$ since $\tilde{M}$ is invertible, and we can redefine our canonical parameter to be $\tilde{M}\varphi(\theta)$. By (12) this new canonical parameter can be partitioned into the parameter of interest and a nuisance parameter. Thus we are in the scenario covered in Davison et al. (2014). No nuisance parameter adjustment term is required in this case.

S.5.3 Making a change of variables

As a reminder

$$\begin{aligned}\hat{\sigma}^2(t) &= \frac{1}{n}\{u_1(t) - u_2^{\mathrm{T}}(t)\hat{\beta}(t)\} \\ &= \frac{1}{n}\{u_1(t) - u_2^{\mathrm{T}}(t)(X^{\mathrm{T}}X)^{-1}u_2(t)\}.\end{aligned}$$

After some algebra the terms involving t disappear from the above expression and we are left with an expression that only involves t^2.

$$\begin{aligned}\hat{\sigma}^2(t) &= \{\hat{\sigma}^2_{\psi} - \frac{t^2}{n}(y^0 - X\hat{\beta}_{\psi})^{\mathrm{T}}X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(y^0 - X\hat{\beta}_{\psi})\} \\ &= (a - t^2 b).\end{aligned}$$

Thus $t_{max} = (\frac{a}{b})^{\frac{1}{2}}$. We find the integral in the numerator to be

$$\int_1^{\sqrt{\frac{a}{b}}} t^{d-1}(a - bt^2)^{\frac{n-p-2}{2}}dt = k\int_1^{\sqrt{\frac{a}{b}}} \big(2\frac{b}{a}t\big)\big(\frac{b}{a}t^2\big)^{\frac{d-2}{2}}\big(1 - \frac{b}{a}t^2\big)^{\frac{n-p-2}{2}}dt.$$

Make the change of variables $x = \frac{b}{a}t^2$:

$$= k\int_{\frac{b}{a}}^{1} x^{\frac{d-2}{2}}(1-x)^{\frac{n-p-2}{2}}dx.$$

Make the change of variables $x = \frac{1}{z}$:

$$\begin{aligned}&= k\int_{\frac{a}{b}}^{1} -\frac{1}{z^2}\big(\frac{1}{z}\big)^{\frac{d-2}{2}}\big(1 - \frac{1}{z}\big)^{\frac{n-p-2}{2}}dz \\ &= k\int_1^{\frac{a}{b}} \big(\frac{1}{z}\big)^{\frac{d+2}{2}}\big(1 - \frac{1}{z}\big)^{\frac{n-p-2}{2}}dz \\ &= k\int_1^{\frac{a}{b}} \big(\frac{1}{z}\big)^{\frac{n+d-p}{2}}(z-1)^{\frac{n-p-2}{2}}dz.\end{aligned}$$

Make the change of variables $t = z - 1$:

$$\begin{aligned}&= k\int_0^{\frac{a}{b}-1} \big(\frac{1}{t+1}\big)^{\frac{n+d-p}{2}}t^{\frac{n-p-2}{2}}dt \\ &= k\int_0^{\frac{a-b}{b}} \big(\frac{1}{t+1}\big)^{(\frac{n-p}{2}+\frac{d}{2})}t^{\frac{n-p}{2}-1}dt.\end{aligned}$$

Make the final change of variables $t = \frac{n-p}{d}x$:

$$
\begin{aligned}
&= k' \int_0^{\frac{(a-b)d}{b(n-p)}} \Big(\frac{1}{\frac{n-p}{d}x+1}\Big)^{(\frac{n-p}{2}+\frac{d}{2})} \Big(\frac{n-p}{d}x\Big)^{\frac{n-p}{2}-1} dx \\
&= k'' \int_0^{\frac{(a-b)d}{b(n-p)}} \frac{\Gamma(\frac{n-p}{2}+\frac{d}{2})}{\Gamma(\frac{n-p}{2})\Gamma(\frac{d}{2})} \frac{n-p}{d-1} \Big(\frac{1}{\frac{n-p}{d}x+1}\Big)^{(\frac{n-p}{2}+\frac{d}{2})} \Big(\frac{n-p}{d}x\Big)^{\frac{n-p}{2}-1} dx \\
&= k'' P_W\left(W > \frac{\frac{b}{d}}{\frac{a-b}{n-p}}\right).
\end{aligned}
$$

where $W \sim F(d, n-p)$. The above sequence of changes of variables is equivalent to making the single change $x = \{(n-p)a - dbt^2\}/(dbt^2)$. Now performing the exact same sequence of changes of variables on the integral in the denominator will result in a similar expression, however the bounds of the integral over the F-distribution will be different. The bounds of the integral will change as follows:

$$
\int_0^{\sqrt{\frac{a}{b}}} \Rightarrow \int_0^1 \Rightarrow \int_1^\infty \Rightarrow \int_0^\infty .
$$

Thus the integral in the denominator of the directional p-value equals

$$
\int_1^{\sqrt{\frac{a}{b}}} t^{q-1}(a-bt^2)^{\frac{n-p-2}{2}} dt = k'' P_W(W > 0) = k''.
$$

We find $n(a-b)$ to be

$$
(y^0)^{\mathrm{T}} y^0 - (y^0)^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} y^0 + 2(X\hat{\beta} - y^0)^{\mathrm{T}} X\hat{\beta}_\psi .
$$

We know that $(y^0 - X\hat{\beta})^{\mathrm{T}} X\hat{\beta} = 0$. Thus

$$
\begin{aligned}
(X\hat{\beta} - y^0)^{\mathrm{T}} X\hat{\beta}_\psi &= \frac{1}{n}(y^0)^{\mathrm{T}}\{I_n - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\}X(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\{A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\}^{-1}(A\hat{\beta} - \psi) \\
&= 0 \\
\implies (a-b) &= (y^0)^{\mathrm{T}}\{I_n - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\}y^0.
\end{aligned}
$$

This is simply the mean squared error under $\hat{\beta}$. Now we simplify nb as

$$
\begin{aligned}
&(y^0 - X\hat{\beta}_\psi)^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(y^0 - X\hat{\beta}_\psi) = (y^0)^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y^0 - 2(y^0)^{\mathrm{T}} X\hat{\beta}_\psi + \hat{\beta}_\psi^{\mathrm{T}} X^{\mathrm{T}} X\hat{\beta}_\psi \\
&= (y^0)^{\mathrm{T}}(X\hat{\beta} - X\hat{\beta}_\psi) - (y^0)^{\mathrm{T}} X\hat{\beta}_\psi + \hat{\beta}_\psi^{\mathrm{T}} X^{\mathrm{T}} X\beta_\psi \\
&= \frac{1}{2}(y^0)^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\hat{\lambda} - (y^0)^{\mathrm{T}} X\hat{\beta} + \frac{1}{2}(y^0)^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\hat{\lambda} + \hat{\beta}_\psi^{\mathrm{T}} X^{\mathrm{T}} X\hat{\beta}_\psi \\
&= (y^0)^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\hat{\lambda} - (y^0)^{\mathrm{T}} X\hat{\beta} + \hat{\beta}X^{\mathrm{T}} X\hat{\beta} - \hat{\beta}^{\mathrm{T}} X^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\hat{\lambda} + \frac{1}{4}\hat{\lambda}^{\mathrm{T}} A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\hat{\lambda} \\
&= \frac{1}{4}\hat{\lambda}^{\mathrm{T}} A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\hat{\lambda} \\
&= (A\hat{\beta} - \psi)^{\mathrm{T}} \left\{A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\right\}^{-1}(A\hat{\beta} - \psi).
\end{aligned}
$$

Finally we find that the directional p-value equals the desired quantity:

$$
p(\psi) = \mathrm{P}_W\left\{W \geq \frac{(A\hat{\beta} - \psi)^{\mathrm{T}}\{A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\}^{-1}(A\hat{\beta} - \psi)/d}{(y^0)^{\mathrm{T}}\{I - X(X^{\mathrm{T}}X)X\}y^0/(n-p)}\right\}.
$$

S.6 Multivariate normal mean

S.6.1 Calculating $\exp[\ell\{\hat{\varphi}(0); s(t)\} - \ell\{\hat{\varphi}(t); s(t)\}]$

Let $y_i \sim N_p(\mu, \Lambda^{-1})$ be n observations from a multivariate Gaussian distribution with unknown concentration matrix Λ. Here we wish to test $H_\psi : \mu = \psi$ using directional testing. The parameter of interest is μ, while the nuisance parameter is Λ. The log-likelihood of these observations is given by

$$
\begin{aligned}
\ell(\mu, \Lambda) &= -\frac{n}{2}\log(|\Lambda^{-1}|) - \frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)^{\mathrm{T}}\Lambda(y_i - \mu) \\
&= \frac{n}{2}\log(|\Lambda|) - \frac{1}{2}\sum_{i=1}^{n} y_i^{\mathrm{T}}\Lambda y_i + \frac{1}{2}\mu^{\mathrm{T}}\Lambda n\bar{y} + \frac{1}{2}n\bar{y}^{\mathrm{T}}\Lambda\mu - \frac{n}{2}\mu^{\mathrm{T}}\Lambda\mu.
\end{aligned}
$$

By using the fact that the trace of a product of matrices is invariant under cyclic permutations and the trace of the product of two square matrices is the dot product of the vectorization of these matrices we can rewrite the log-likelihood as

$$
\ell(\mu, \Lambda) = \begin{bmatrix} \Lambda\mu \\ \mathrm{vec}(\Lambda) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} n\bar{y} \\ \mathrm{vec}(-\frac{1}{2}\sum_{i=1}^{n} y_i y_i^{\mathrm{T}}) \end{bmatrix} + \frac{n}{2}\log(|\Lambda|) - \frac{n}{2}\mu^{\mathrm{T}}\Lambda\Lambda^{-1}\Lambda\mu \qquad (13)
$$

We can rewrite the above log-likelihood in terms of the canonical parameter, φ, and sufficient statistic u:

$$\begin{aligned}
\ell\{\varphi; u(y)\} =& \varphi^{\mathrm{T}} u + \frac{n}{2}\log(|\varphi_2|) - \frac{n}{2}\varphi_1^{\mathrm{T}}\varphi_2^{-1}\varphi_1, \\
\varphi =& \begin{bmatrix}\varphi_1\\ \varphi_2\end{bmatrix} = \begin{bmatrix}\Lambda\mu\\ \Lambda\end{bmatrix} = \begin{bmatrix}\lambda\psi\\ \lambda\end{bmatrix}, \\
\theta =& \begin{bmatrix}\psi\\ \lambda\end{bmatrix} = \begin{bmatrix}\mu\\ \Lambda\end{bmatrix}.
\end{aligned}$$

Throughout, we will be treat φ_2 as both a matrix and the vectorization of a matrix depending upon the context it is used in. The constrained MLE under H_ψ is found by simply maximizing $\ell(\mu, \Lambda)$ with respect to Λ while setting $\mu = \psi$. This yields the standard covariance matrix estimate
$\hat{\Lambda}_\psi^{-1} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \psi)(y_i - \psi)^{\mathrm{T}}$. This matrix as well as ψ will be used to find $s(t)$.

Finding $s(t)$ amounts to first finding a vector s_ψ which when added to the observed value of the sufficient statistic has the constrained MLE, $\hat{\varphi}_\psi$, as its MLE. Once this is found $s(t)$ is given by $s(t) = (1-t)s_\psi$. The partial derivatives of $\ell(\varphi, s)$ with respect to φ are

$$\begin{aligned}
&\frac{\partial \ell}{\partial \varphi_1} = (u_1 + s_1) - n\varphi_2^{-1}\varphi_1 \\
\Longrightarrow \qquad & s_1 = n\psi - u_1, \\
&\frac{\partial \ell}{\partial \varphi_2} = (u_2 + s_2) + \frac{n}{2}\varphi_2^{-1} + \frac{n}{2}\varphi_2^{-1}\varphi_1\varphi_1^{T}\varphi_2^{-1} \\
\Longrightarrow \qquad & s_2 = -u_2 - \frac{n}{2}\hat{\Lambda}_\psi^{-1} - \frac{n}{2}\psi\psi^{\mathrm{T}}.
\end{aligned}$$

In much the same way that s_ψ was found we find the maximum likelihood estimate for φ as we vary t in $s(t)$. The partial derivatives of $\ell\{\varphi, s(t)\}$ with respect to φ are

$$\begin{aligned}
\frac{\partial \ell}{\partial \varphi_1} &= \{u_1 + s_1(t)\} - n\varphi_2^{-1}\varphi_1, \\
\Longrightarrow \qquad \hat{\varphi}_2^{-1}(t)\hat{\varphi}_1(t) &= \frac{1}{n}\{u_1 + s_1(t)\} \\
&= \frac{1}{n}\{n\psi + t(u_1 - n\psi)\} \\
&= \psi + t(\bar{y} - \psi).
\end{aligned}$$

Notice that the above MLE agrees with what one might reasonably expect since $\varphi_2^{-1}\varphi_1$ is the mean vector. We use the above formula to solve for $\hat{\varphi}_2^{-1}(t)$:

$$\begin{aligned}
\frac{\partial \ell}{\partial \varphi_2} =& \{u_2 + s_2(t)\} + \frac{n}{2}\varphi_2^{-1} + \frac{n}{2}\varphi_2^{-1}\varphi_1\varphi_1^{\mathrm{T}}\varphi_2^{-1}, \\
\implies \qquad \hat{\varphi}_2^{-1}(t) =& \frac{2}{n}[-u_2 - s_2(t) - \frac{n}{2}\{\hat{\varphi}_2^{-1}(t)\hat{\varphi}_1(t)\}\{\hat{\varphi}_2^{-1}(t)\hat{\varphi}_1(t)\}^{\mathrm{T}}] \\
=& \frac{2}{n}\{-u_2 + (1-t)(u_2 + \frac{n}{2}\hat{\Lambda}_\psi^{-1} + \frac{n}{2}\psi\psi^{\mathrm{T}})\} - \frac{1}{n^2}\{u_1 + s_1(t)\}\{u_1 + s_1(t)\}^{\mathrm{T}} \\
=& (1-t)(\hat{\Lambda}_\psi^{-1} + \psi\psi^{\mathrm{T}}) - \frac{2}{n}tu_2 - \{\psi + t(\bar{y} - \psi)\}\{\psi + t(\bar{y} - \psi)\}^{\mathrm{T}} \\
=& (1-t)\hat{\Lambda}_\psi^{-1} + t(\psi\psi^{\mathrm{T}} - \bar{y}\psi^{\mathrm{T}} - \psi\bar{y}^{\mathrm{T}}) - \frac{2}{n}tu_2 - t^2(\bar{y} - \psi)(\bar{y} - \psi)^{\mathrm{T}} \\
=& \hat{\Lambda}_\psi^{-1} - t^2(\bar{y} - \psi)(\bar{y} - \psi)^{\mathrm{T}}.
\end{aligned}$$

As a check on our work we see that $\hat{\varphi}(0)$ provides the constrained MLE while $\hat{\varphi}(1)$ gives the unconstrained MLE. We see that $\hat{\varphi}_2^{-1}(t)$ is symmetric and thus $\hat{\varphi}_2(t)$ is symmetric meaning that any transpositions of these terms may be ignored in future calculations. Throughout we will use $\hat{\mu}(t)$ to represent $\hat{\varphi}_2^{-1}(t)\hat{\varphi}_1(t)$. Remembering that any terms not involving t can be dropped from our calculations we find $\ell\{\hat{\varphi}(t); s(t)\}$ to be

$$\begin{aligned}
\ell\{\hat{\varphi}(t); s(t)\} =& \hat{\varphi}^{\mathrm{T}}(t)\{u + s(t)\} + \frac{n}{2}\log(|\hat{\varphi}_2(t)|) - \frac{n}{2}\hat{\varphi}_1^{\mathrm{T}}(t)\hat{\varphi}_2^{-1}(t)\hat{\varphi}_1(t) \\
=& n\hat{\varphi}_1^{\mathrm{T}}(t)\hat{\mu}(t) + \mathrm{Tr}\big[\hat{\varphi}_2^{\mathrm{T}}(t)\{u_2 + s_2(t)\}\big] - \frac{n}{2}\hat{\mu}^{\mathrm{T}}(t)\hat{\varphi}_2(t)\hat{\mu}(t) - \frac{n}{2}\log(|\hat{\varphi}_2^{-1}(t)|) \\
=& \mathrm{Tr}\big[\hat{\varphi}_2(t)\big\{\frac{n}{2}\hat{\mu}(t)\hat{\mu}^{\mathrm{T}}(t) - \frac{n}{2}\Lambda_\psi^{-1} - \frac{n}{2}\psi\psi^{\mathrm{T}} + t(u_2 + \frac{n}{2}\Lambda_\psi^{-1} + \frac{n}{2}\psi\psi^{\mathrm{T}})\big\}\big] \\
&- \frac{n}{2}\log(|\hat{\varphi}_2^{-1}(t)|) \\
=& \mathrm{Tr}\big[\hat{\varphi}_2(t)\big\{-\frac{n}{2}\Lambda_\psi^{-1} + \frac{n}{2}t^2(\bar{y} - \psi)(\bar{y} - \psi)^{\mathrm{T}}\big\}\big] - \frac{n}{2}\log(|\hat{\varphi}_2^{-1}(t)|) \\
=& \mathrm{Tr}(-\frac{n}{2}I_p) - \frac{n}{2}\log(|\hat{\varphi}_2^{-1}(t)|) \\
\equiv& -\frac{n}{2}\log(|\hat{\varphi}_2^{-1}(t)|).
\end{aligned}$$

Similarly we find $\ell\{\hat{\varphi}(0); s(t)\}$ as

$$
\begin{aligned}
\ell\{\hat{\varphi}(0); s(t)\} =& \hat{\varphi}^{\mathrm{T}}(0)\{u + s(t)\} \\
=& n\hat{\mu}^{\mathrm{T}}(0)\hat{\varphi}_2(0)\hat{\mu}(t) + \mathrm{Tr}[\hat{\varphi}_2(0)\{u_2 + s_2(t)\}] \\
=& \mathrm{Tr}\big[\hat{\varphi}_2(0)\big\{n\hat{\mu}(t)\psi^{\mathrm{T}} + u_2 + s_2(t)\big\}\big] \\
=& \mathrm{Tr}\big[\hat{\varphi}_2(0)\big\{n\hat{\mu}(t)\psi^{\mathrm{T}} - \frac{n}{2}\Lambda_{\psi}^{-1} - \frac{n}{2}\psi\psi^{\mathrm{T}} + t(u_2 + \frac{n}{2}\Lambda_{\psi}^{-1} + \frac{n}{2}\psi\psi^{\mathrm{T}})\big\}\big] \\
\equiv& \mathrm{Tr}\big[\hat{\varphi}_2(0)\big\{tn(\bar{y} - \psi)\psi^{\mathrm{T}} + t(u_2 + \frac{n}{2}\Lambda_{\psi}^{-1} + \frac{n}{2}\psi\psi^{\mathrm{T}})\big\}\big] \\
=& \mathrm{Tr}\big[\hat{\varphi}_2(0)\big\{tn(\bar{y} - \psi)\psi^{\mathrm{T}} + t(n\psi\psi^{\mathrm{T}} - \frac{n}{2}\bar{y}\psi^{\mathrm{T}} - \frac{n}{2}\psi\bar{y}^{\mathrm{T}})\big\}\big] \\
=& 0.
\end{aligned}
$$

From here we can find the first piece of the conditional density:

$$
\exp[\ell\{\hat{\varphi}(0); s(t)\} - \ell\{\hat{\varphi}(t); s(t)\}] \equiv |\hat{\varphi}_2^{-1}(t)|^{\frac{n}{2}}.
$$

S.6.2 Finding $|J_{\varphi\varphi}\{\hat{\varphi}(t); s(t)\}|$ and the nuisance parameter adjustment

The likelihood in this scenario is the same as that in Example 5.3 of Fraser et. al. (2014). The canonical parameterization is also unchanged and so we can borrow the result that

$$
|J_{\varphi\varphi}\{\hat{\varphi}(t); s(t)\}| = |\hat{\varphi}_2^{-1}(t)|^{p+2}.
$$

The only term involving t in the nuisance parameter adjustment is $|\ell_{\lambda\lambda}\{\hat{\varphi}(0); s(t)\}|$ which in turn only involves t through $\partial^2/\partial\lambda^2\{s^{\mathrm{T}}(t)\varphi\}$. Now $\Lambda\mu$ is linear in $\lambda = \Lambda$ and of course so is Λ. As a result, all second order derivatives of φ with respect to λ disappear, meaning that the nuisance parameter adjustment is constant.

S.6.3 Making a change of variables

Let's call $A = \hat{\Lambda}_{\psi}^{-1}$, $v = (\bar{y} - \psi)$ and $B = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})^{\mathrm{T}}$. By definition $|\Lambda^{-1}(t)| = |A - t^2 vv^{\mathrm{T}}|$. Using the matrix determinant lemma we see that

$$
|\Lambda^{-1}(t)| = \det(A)(1 - t^2 v^{\mathrm{T}} A^{-1} v)
$$

By the constraint that $\hat{\Lambda}^{-1}(t)$ must be a valid covariance matrix, t_{max} is the largest value of t such that $|\hat{\Lambda}^{-1}(t)|$ is positive definite. $\hat{\Lambda}^{-1}(t)$ stops being positive definite as soon as one of its eigenvalues becomes 0. Thus, t_{max} is the solution to $|\hat{\Lambda}^{-1}(t)| = 0$:

$$
\implies t_{max} = (v^{\mathrm{T}} A^{-1} v)^{-1/2}.
$$

Next we derive a formula that will be useful later on. It is easily seen that $A = B + vv^{\mathrm{T}}$. Using the Sherman-Morrison formula on $(B + vv^{\mathrm{T}})^{-1}$ we see that

$$
\begin{aligned}
v^{\mathrm{T}} A^{-1} v &= v^{\mathrm{T}} \left(B^{-1} - \frac{B^{-1} vv^{\mathrm{T}} B^{-1}}{1 + v^{\mathrm{T}} B^{-1} v} \right) v \\
&= \frac{v^{\mathrm{T}} B^{-1} v}{1 + v^{\mathrm{T}} B^{-1} v}.
\end{aligned}
$$

Therefore, $(v^{\mathrm{T}} A^{-1} v)/(1 - v^{\mathrm{T}} A^{-1} v) = v^{\mathrm{T}} B^{-1} v$.

The directional p-value is given by

$$p(\psi) = \frac{\int_1^{(v^{\mathrm{T}} A^{-1} v)^{-1/2}} t^{p-1} (1 - t^2 v^{\mathrm{T}} A^{-1} v)^{\frac{n-p-2}{2}} dt}{\int_0^{(v^{\mathrm{T}} A^{-1} v)^{-1/2}} t^{p-1} (1 - t^2 v^{\mathrm{T}} A^{-1} v)^{\frac{n-p-2}{2}} dt}.$$

For simplicity let $C = v^{\mathrm{T}} A^{-1} v$. The changes of variables made here are essentially identical to that in the normal linear regression example in Section 3. Make the change of variables $x = Ct^2$ in the integral in the numerator:

$$\int_1^{C^{-1/2}} t^{p-1} (1 - t^2 C)^{\frac{n-p-2}{2}} dt = k \int_C^1 x^{\frac{p-2}{2}} (1 - x)^{\frac{n-p-2}{2}} dx.$$

Make the change of variables to $x = 1/z$:

$$\begin{aligned}
&= k \int_{C^{-1}}^1 -\frac{1}{z^2} (\frac{1}{z})^{\frac{p-2}{2}} (1 - \frac{1}{z})^{\frac{n-p-2}{2}} dz \\
&= k \int_1^{C^{-1}} (\frac{1}{z})^{\frac{p+2}{2}} (1 - \frac{1}{z})^{\frac{n-p-2}{2}} dz \\
&= k \int_1^{C^{-1}} (\frac{1}{z})^{\frac{n}{2}} (z - 1)^{\frac{n-p-2}{2}} dz.
\end{aligned}$$

Make the change of variables $t = z - 1$:

$$\begin{aligned}
&= k \int_0^{C^{-1}-1} (\frac{1}{t+1})^{\frac{n}{2}} t^{\frac{n-p-2}{2}} dt \\
&= k \int_0^{\frac{1-C}{C}} (\frac{1}{t+1})^{(\frac{n-p}{2} + \frac{p}{2})} t^{\frac{n-p}{2} - 1} dt.
\end{aligned}$$

Make the final change of variables $t = \frac{n-p}{p} x$:

$$\begin{aligned}
&= k' \int_0^{\frac{(1-C)p}{C(n-p)}} \left(\frac{1}{\frac{n-p}{p} x + 1} \right)^{(\frac{n-p}{2} + \frac{p}{2})} (\frac{n-p}{p} x)^{\frac{n-p}{2} - 1} dx \\
&= k'' \int_0^{\frac{(1-C)p}{C(n-p)}} \frac{\Gamma(\frac{n-p}{2} + \frac{p}{2})}{\Gamma(\frac{n-p}{2}) \Gamma(\frac{p}{2})} \frac{n-p}{p-1} (\frac{1}{\frac{n-p}{p} x + 1})^{(\frac{n-p}{2} + \frac{p}{2})} (\frac{n-p}{p} x)^{\frac{n-p}{2} - 1} dx \\
&= k'' P_W \left(W > \frac{n-p}{p} \frac{C}{1-C} \right) \\
&= k'' P_W \left(W > \frac{n-p}{p} v^{\mathrm{T}} B^{-1} v \right),
\end{aligned}$$

where $W \sim F(p, n-p)$. Now performing the exact same sequence of changes of variables on the integral in the denominator will result in a similar expression. The bounds of the integral will change as follows:

$$\int_0^{C^{-1/2}} \Rightarrow \int_0^1 \Rightarrow \int_1^\infty \Rightarrow \int_0^\infty .$$

Thus the integral in the denominator of the directional p-value will equal $k'' P_W(W > 0) = k''$. Hotelling's T^2 statistic is given by $T^2 = (n-1)v^{\mathrm{T}} B^{-1} v$ and since $\frac{n-p}{p(n-1)} T^2 \sim F_{p,n-p}$, the directional test is identical to the p-value obtained from Hotelling's T^2 test.

BIBLIOGRAPHY

Davison, A., Fraser, D.A.S., Reid, N., Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families *Journal of the American Statistical Association* **109**, 302–314.

Fraser, D.A.S. & Massam, H. (1985). Conical tests: observed levels of significance and confidence regions. *Statistiche Hefte* **26**, 1–17.

Fraser, D.A.S., Reid, N., Sartori, N. (2016). Accurate directional inference for vector parameters. *Biometrika* **103**, 625–639.

Rencher, A.C. & Schaalje, G.B. (2008). *Linear Models in Statistics.* 2nd ed. Wiley, New Jersey.

Skovgaard, I.M. (1988). Saddlepoint expansions for directional test probabilities. *Journal of the Royal Statistical Society, Series B*, **50**, 3–32.