Notes on computing the Fisher Information matrix for MARSS models. Part III Overview of Harvey 1989

EE Holmes, National Marine Fisheries Service & University of Washington

2016-06-16

This is part of a series on computing the Fisher Information for Multivariate Autoregressive State-Space Models. Part I: Background, Part II: Louis 1982, Part III: Harvey 1989, Background, Part IV: Harvey 1989, Implementation. [1]

Part II discussed the approach by Louis 1982 which uses the full-data likelihood and the first derivative of that that is part of the M-step of the EM algorithm. The conclusion of part II was that that approach is doable but computationally expensive because it scales with T^2 at least.

Here I will review the more common approach (Harvey 1989, pages 140-142, section 3.4.5 Information matrix) which uses the prediction error form of the likelihood function to calculate the observed Fisher Information $\mathcal{I}(\hat{\theta}, y)$. A related paper is Cavanaugh and Shumway (1996), which presents an approach for calculating the *expected* Fisher Information.

Harvey 1989 recursion for the expected and observed Fisher Information matrix

Harvey (1989), pages 140-142, shows how to write the Hessian of the log-likelihood function using the prediction error form of the likelihood. The prediction error form is:

$$\log L = \sum_{t=1}^{T} l_t = \sum_{t=1}^{T} p(y_t|y_1^{t-1}) \tag{1}$$

The Hessian of the log-likelihood can then be written as

$$\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} = \sum \frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} \tag{2}$$

and this can be written in terms of derivatives of the innovations v_t and the variance of the innovations F_t. This is shown in Equation 3.4.66 in Harvey (1989). There are a couple differences between the equation below and 3.4.66 in Harvey. First, 3.4.66 has a typo; the $[I - F_t v_t v_t^\top]$ should be within the trace (as below). Second, I have written out the derivative with respect to θ_j that appears in the first trace term.

$$
\begin{aligned}
\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j} = {} & \frac{1}{2} tr \left[\left[F_t^{-1} \frac{\partial F_t}{\partial \theta_j} F_t^{-1} \frac{\partial F_t}{\partial \theta_i} - F_t^{-1} \frac{\partial^2 F_t}{\partial \theta_i \partial \theta_j} \right] \left[I - F_t^{-1} v_t v_t^\top \right] \right] - \\
& \frac{1}{2} tr \left[F_t^{-1} \frac{\partial F_t}{\partial \theta_i} F_t^{-1} \frac{\partial F_t}{\partial \theta_j} F_t^{-1} v_t v_t^\top \right] + \\
& \frac{1}{2} tr \left[F_t^{-1} \frac{\partial F_t}{\partial \theta_i} F_t^{-1} \left[\frac{\partial v_t}{\partial \theta_j} v_t^\top + v_t \frac{\partial v_t^\top}{\partial \theta_j} \right] \right] - \\
& \frac{\partial^2 v_t^\top}{\partial \theta_i \partial \theta_j} F_t^{-1} v_t + \frac{\partial v_t^\top}{\partial \theta_i} F_t^{-1} \frac{\partial F_t}{\partial \theta_j} F_t^{-1} v_t - \frac{\partial v_t^\top}{\partial \theta_i} F_t^{-1} \frac{\partial v_t}{\partial \theta_j}
\end{aligned}
\tag{3}
$$

---

[1] Holmes, E. E. 2016. Notes on computing the Fisher Information matrix for MARSS models. Part III Overview of Harvey 1989.

The Fisher Information matrix is the negative of the expected value (over all possible data) of 2:

$$I(\theta) = -E\left[\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right] \tag{4}$$

Thus for the Fisher Information matrix, we take the expectation (over all possible data) of the sum (over t) of Equation 3 (3.4.66 in Harvey 1989). On pages 141-142, Harvey shows that the expected value of Equation 3 can be simplified and the i,j element of the Fisher Information matrix can be written as (Equation 3.4.69 in Harvey 1989):

$$I_{ij}(\theta) = \frac{1}{2}\sum_t\left[tr\left[F_t^{-1}\frac{\partial F_t}{\partial \theta_i}F_t^{-1}\frac{\partial F_t}{\partial \theta_j}\right]\right] + E\left[\sum_t\left(\frac{\partial v_t}{\partial \theta_i}\right)^\top F_t^{-1}\frac{\partial v_t}{\partial \theta_j}\right] \tag{5}$$

Equation 5 (3.4.69 in Harvey 1989) is the Fisher Information and is evaluated at the true parameter values θ. We do not know θ and instead we estimate the Fisher Information using our estimates of θ. The two estimates of $I(\theta)$ that are used are called the expected and observed Fisher Information matrices. The expected Fisher Information is

$$I(\hat{\theta}) = -E\left[\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right]|_{\theta=\hat{\theta}} = -E\left[\sum\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}\right]|_{\theta=\hat{\theta}} \tag{6}$$

and the observed Fisher Information is

$$\mathcal{I}(\hat{\theta}, y) = -\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}|_{\theta=\hat{\theta}} = -\sum\frac{\partial^2 l_t}{\partial \theta_i \partial \theta_j}|_{\theta=\hat{\theta}} \tag{7}$$

The $|_{\theta=\hat{\theta}}$ means 'evaluated at'. l_t is a function of θ. We take the derivative of that function and then evaluate that derivative at $\theta = \hat{\theta}$. The expectation (which is an integral) is over that possible values of the data y which are generated from the model with θ.

The observed Fisher Information drops the expectation and the expected Fisher Information does not. The expectation is taken over all possible data, and we have only one observed data set. On first blush, it may seem that it is impossible to compute the expectation and that we must always use the observed Fisher Information. However, for some models, one can write down the expectations analytically. One could simulate from the MLEs to get the expectations—this is the idea behind bootstrapping. In a bootstrapping approach one uses the MLE to generate data. This is an approximation since what we would like is to simulate data from the true parameters. The mean and variance of data generated from the MLEs versus data generated the true parameters often have nice asymptotic properties.

However it is common to use the observed Fisher Information matrix. This is what one is using when one uses the Hessian of the log-likelihood function evaluated at the MLEs. To get an analytical equation for the observed Fisher Information matrix, we use Equation 3 for l_t and take the sum to get the Hessian of the log-likelihood function (2). This is the same Hessian that you can get numerically. In R, you can use the fdHess function in the nmle package or the optim function.

Partially observed, partially expected Fisher Information matrix

Equation 5 (Equation 3.4.69 in Harvey) is a simplification the expected value of the sum of equation 3. The simplification occurs because a number of terms in equation 3 drop out or cancel out when you take the expectation (see bottom of page 141 in Harvey 1989). The only terms that remain are those shown in equation 5. Harvey (1989) does not say how to compute the expectation in equation 5 (which is his 3.4.69). Cavanaugh and Shumway (1996) do not say how to compute it either and suggest that it is infeasible (page 1

in paragraph after their equation 1). Instead they say that you can drop the expectation in equation 5 and get the observed Fisher Information:

$$\mathcal{I}_{ij}(\theta) = \frac{1}{2}\sum_t \left[tr\left[F_t^{-1}\frac{\partial F_t}{\partial \theta_i}F_t^{-1}\frac{\partial F_t}{\partial \theta_j}\right]\right] + \sum_t \left(\frac{\partial v_t}{\partial \theta_i}\right)^\top F_t^{-1}\frac{\partial v_t}{\partial \theta_j} \tag{8}$$

This however is halfway between the expected Fisher Information matrix and the observation Fisher Information matrix because equation 5 is what you get after doing the expectation and dropping some of the terms in equation 3. If you compare what you get from equation 8 and what you get from a numerical estimate of the Hessian of the log-likelihood function at the MLE, you will see that they are different. The variance of the former is less than the variance of the latter. This is what you expect since the former has had the expectation applied to some terms in equation 3 (Harvey's 3.4.66).

This does not mean that equation 8 should not be used, but rather that if you compare it to the output from a numerically computed Hessian, they will not be the same.

In Part IV, I show Harvey's recursion for computing the first derivatives of v_t and F_t needed in equations 3 and 5. I extend this recursion to get the second derivative also. Once we have all these, we can use equation 7 with equation 3 to compute the observed Fisher Information matrix and use equation 5 to compute the "observed/expected" Fisher Information.

Writing Equation 3 in vec form

We can compute the Hessian of the log-likelihood by using a for loop of i from 1 to p with an inner for loop for j from i to p. The Hessian is symmetric so the inner loop only needs to go from i to p. However, we can also write the Hessian for time step t in a single line without any for loops using the Jacobian matrices for our derivatives. With the t subscripts of F and v dropped:

$$\begin{aligned} &\frac{1}{2}J_F^\top(F^{-1}\otimes F^{-1})J_F - J_F^\top(F^{-1}vv^\top F^{-1}\otimes F^{-1})J_F - \frac{1}{2}(I_p\otimes[F^{-1}-F^{-1}v_tv_t^\top F^{-1}])\mathcal{J}_F+ \\ &\qquad \frac{1}{2}J_F^\top[3F^{-1}v\otimes F^{-1}+F^{-1}\otimes F^{-1}v]J_v - \mathcal{J}_\sqsubseteq^\top(I_p\otimes F^{-1}v) - J_v^\top F^{-1}J_v \end{aligned} \tag{9}$$

This may or may not be faster but it is more concise. Go to Part IV. to see how to compute these Jacobians using Harvey's recursion.

Derivation of the observed Fisher Information matrix (equation 9)

Note, I am going to drop the t subscript on F and v because things are going to get cluttered; v_1 will refer to the 1st element of the $n\times 1$ column vector v and F_{11} is the (1,1) element of the matrix F. There has to be a loop to go through all the F_t and v_t for $t = 1$ to T.

Terms 1 and 2 of equation 3

The first term of equation 3 is

$$\begin{aligned} &\frac{1}{2}tr\left[\left[F^{-1}\frac{\partial F}{\partial \theta_j}F^{-1}\frac{\partial F}{\partial \theta_i}-\frac{1}{2}F^{-1}\frac{\partial^2 F}{\partial \theta_i\partial \theta_j}\right][I-F^{-1}vv^\top]\right]= \\ &\frac{1}{2}tr\left[F^{-1}\frac{\partial F}{\partial \theta_j}F^{-1}\frac{\partial F}{\partial \theta_i}[I-F_t^{-1}vv^\top]\right]-\frac{1}{2}tr\left[F^{-1}\frac{\partial^2 F}{\partial \theta_i\partial \theta_j}[I-F^{-1}vv^\top]\right]= \\ &\frac{1}{2}tr\left[F^{-1}\frac{\partial F}{\partial \theta_j}F^{-1}\frac{\partial F}{\partial \theta_i}\right]-\frac{1}{2}tr\left[F^{-1}\frac{\partial F}{\partial \theta_j}F^{-1}\frac{\partial F}{\partial \theta_i}F^{-1}vv^\top\right]-\frac{1}{2}tr\left[F^{-1}\frac{\partial^2 F}{\partial \theta_i\partial \theta_j}[I-F^{-1}vv^\top]\right] \end{aligned} \tag{10}$$

The second term of equation 3 is

$$-\frac{1}{2}tr\left[F_t^{-1}\frac{\partial F_t}{\partial \theta_i}F_t^{-1}\frac{\partial F_t}{\partial \theta_j}F_t^{-1}v_tv_t^\top\right] \tag{11}$$

All the matrices within the traces above are symmetric. The trace of products of symmetric matrices is permutation invariant. That means that if A, B, C, and D are symmetric matrices, $tr(ABCD) = tr(ACBD) = tr(ACDB)$, etc. Thus the second term can be rearranged to match the middle term in the first term. Terms $1 + 2$ of equation 3 can thus be written as

$$\frac{1}{2}tr\left[F_t^{-1}\frac{\partial F_t}{\partial \theta_j}F_t^{-1}\frac{\partial F_t}{\partial \theta_i}\right] - tr\left[F_t^{-1}\frac{\partial F_t}{\partial \theta_j}F_t^{-1}\frac{\partial F_t}{\partial \theta_i}F_t^{-1}v_tv_t^\top\right] - \frac{1}{2}tr\left[F_t^{-1}\frac{\partial^2 F_t}{\partial \theta_i\partial \theta_j}\left[I - F_t^{-1}v_tv_t^\top\right]\right] \tag{12}$$

We can write the first trace of equation 12 as a vector product using the relation $tr(A^\top B) = vec(A)^\top vec(B)$. Note that the matrices in the traces in equation 12 are symmetric. If A is symmetric, $A^\top = A$ and $tr(AB) = vec(A)^\top vec(B)$.

$$tr\left[F^{-1}\frac{\partial F}{\partial \theta_j}F^{-1}\frac{\partial F}{\partial \theta_i}\right] = vec\left(F^{-1}\frac{\partial F}{\partial \theta_j}F^{-1}\right)^\top vec\left(\frac{\partial F}{\partial \theta_i}\right) =$$
$$\left((F^{-1}\otimes F^{-1})vec\left(\frac{\partial F}{\partial \theta_j}\right)\right)^\top vec\left(\frac{\partial F}{\partial \theta_i}\right) = \tag{13}$$
$$vec\left(\frac{\partial F}{\partial \theta_j}\right)^\top (F^{-1}\otimes F^{-1})vec\left(\frac{\partial F}{\partial \theta_i}\right)$$

That is for the i,j element. This matrix is symmetric so it is also the j,i element. The derivative of $vec(F)$ with respect to θ (as opposed to the j-th element of θ) is the Jacobian matrix of $vec(F)$.

$$J_F = \begin{bmatrix}\frac{\partial vec(F)}{\theta_1} & \frac{\partial vec(F)}{\theta_2} & \cdots & \frac{\partial vec(F)}{\theta_p}\end{bmatrix} = \begin{bmatrix}\frac{\partial F_{11}}{\theta_1} & \frac{\partial F_{11}}{\theta_2} & \cdots & \frac{\partial F_{11}}{\theta_p} \\ \frac{\partial F_{21}}{\theta_1} & \frac{\partial F_{21}}{\theta_2} & \cdots & \frac{\partial F_{21}}{\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{nn}}{\theta_1} & \frac{\partial F_{nn}}{\theta_2} & \cdots & \frac{\partial F_{nn}}{\theta_p}\end{bmatrix} \tag{14}$$

The full matrix for the first part of equation 12 is then

$$\frac{1}{2}J_F^\top(F^{-1}\otimes F^{-1})J_F \tag{15}$$

The middle trace of equation 12 is similar to the first and we end up with:

$$vec\left(\frac{\partial F}{\partial \theta_j}\right)^\top (F^{-1}\otimes F^{-1})vec\left(\frac{\partial F}{\partial \theta_i}F^{-1}vv^\top\right) =$$
$$vec\left(\frac{\partial F}{\partial \theta_j}\right)^\top (F^{-1}\otimes F^{-1})(vv^\top F^{-1}\otimes I_n)vec\left(\frac{\partial F}{\partial \theta_i}\right) = \tag{16}$$
$$vec\left(\frac{\partial F}{\partial \theta_j}\right)^\top (F^{-1}vv^\top F^{-1}\otimes F^{-1})vec\left(\frac{\partial F}{\partial \theta_i}\right)$$

We can write this in terms of the Jacobian of $vec(F)$:

$$J_F^\top(F^{-1}vv^\top F^{-1}\otimes F^{-1})J_F \tag{17}$$

4

The third part of equation 12 involves the second derivatives $\partial^2 F / \partial\theta_i \partial\theta_j$.

$$tr\left[F^{-1}\frac{\partial^2 F}{\partial\theta_i\partial\theta_j}[I - F^{-1}vv^\top]\right] = tr\left[[I - F^{-1}vv^\top]F^{-1}\frac{\partial^2 F}{\partial\theta_i\partial\theta_j}\right] =$$

$$vec\left(F^{-1} - F^{-1}vv^\top F^{-1}\right)^\top vec\left(\frac{\partial^2 F}{\partial\theta_i\partial\theta_j}\right) = \tag{18}$$

$$vec\left(F^{-1} - F^{-1}vv^\top F^{-1}\right)^\top \frac{\partial vec(\partial F/\partial\theta_i)}{\partial\theta_j}$$

Again this is the i,j term. The term on the bottom line on the right is the (θ_i, θ_j) term of the Jacobian of the vec of the Jacobian of F:

$$\mathcal{J}_F = \begin{bmatrix} \frac{\partial vec(J_F)}{\partial\theta_1} & \frac{\partial vec(J_F)}{\partial\theta_2} & \cdots & \frac{\partial vec(J_F)}{\partial\theta_p} \end{bmatrix} = \begin{bmatrix} \frac{\partial F_{11}}{\theta_1\theta_1} & \frac{\partial F_{11}}{\theta_1\theta_2} & \cdots & \frac{\partial F_{11}}{\theta_1\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{nn}}{\theta_1\theta_1} & \frac{\partial F_{nn}}{\theta_1\theta_2} & \cdots & \frac{\partial F_{nn}}{\theta_1\theta_p} \\ \frac{\partial F_{11}}{\theta_2\theta_1} & \frac{\partial F_{11}}{\theta_2\theta_2} & \cdots & \frac{\partial F_{11}}{\theta_2\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{nn}}{\theta_2\theta_1} & \frac{\partial F_{nn}}{\theta_2\theta_2} & \cdots & \frac{\partial F_{nn}}{\theta_2\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{11}}{\theta_p\theta_1} & \frac{\partial F_{11}}{\theta_p\theta_2} & \cdots & \frac{\partial F_{11}}{\theta_p\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{nn}}{\theta_p\theta_1} & \frac{\partial F_{nn}}{\theta_p\theta_2} & \cdots & \frac{\partial F_{nn}}{\theta_p\theta_p} \end{bmatrix} \tag{19}$$

The full matrix for the second part of term $1 + 2$ in Equation 3 is then

$$(I_p \otimes [F^{-1} - F^{-1}vv^\top F^{-1}])\mathcal{J}_F \tag{20}$$

The subscript on the I indicates the size of the identity matrix. In this case, it is a $p \times p$ matrix.

Term 3 of equation 3

With the t subscripts dropped, the 3rd term of equation 3 is

$$\frac{1}{2}tr\left[F^{-1}\frac{\partial F}{\partial\theta_i}F^{-1}\left(\frac{\partial v}{\partial\theta_j}v^\top + v\frac{\partial v^\top}{\partial\theta_j}\right)\right] \tag{21}$$

Using the same procedure as for the above terms, we can write this in terms of vecs. If b and a are $1 \times n$ column vectors,

$$vec(ab^\top) = (b \otimes I_n)vec(a) = (b \otimes I_n)a = (I_n \otimes a)vec(b) = (I_n \otimes a)b \tag{22}$$

Thus,

$$vec\left(\frac{\partial v}{\partial\theta_j}v^\top\right) = (v \otimes I_n)\frac{\partial v}{\partial\theta_j}$$

$$vec\left(v(\partial v/\partial\theta_j)^\top\right) = (I_n \otimes v)\frac{\partial v}{\partial\theta_j} \tag{23}$$

and

$$vec\left(\frac{\partial v}{\partial \theta_j}v^\top + v(\partial v/\partial \theta_j)^\top\right) = (v \otimes I_n + I_n \otimes v)\frac{\partial v}{\partial \theta_j} \tag{24}$$

When A is symmetric, $tr(AB) = vec(A)^\top vec(B)$. Thus term 3 of equation 3 can be written as

$$\frac{1}{2}tr\left[F^{-1}\frac{\partial F}{\partial \theta_i}F^{-1}\left(\frac{\partial v}{\partial \theta_j}v^\top + v\frac{\partial v^\top}{\partial \theta_j}\right)\right] = vec\left(\frac{\partial F}{\partial \theta_i}\right)^\top (F^{-1} \otimes F^{-1})(v \otimes I_n + I_n \otimes v)\frac{\partial v}{\partial \theta_j}$$

$$vec\left(\frac{\partial F}{\partial \theta_i}\right)^\top (F^{-1}v \otimes F^{-1} + F^{-1} \otimes F^{-1}v)\frac{\partial v}{\partial \theta_j} \tag{25}$$

This is the i,j term of the Fisher Information matrix from term 3 in equation 3. To get all terms, we use the Jacobian of vec(F) as above and the Jacobian of v:

$$\frac{1}{2}J_F^\top(F^{-1} \otimes F^{-1})(v \otimes I_n + I_n \otimes v)J_v = \frac{1}{2}J_F^\top[F^{-1}v \otimes F^{-1} + F^{-1} \otimes F^{-1}v]J_v \tag{26}$$

where J_F is defined in equation 14 and J_v is

$$J_v = \begin{bmatrix} \frac{\partial v_1}{\theta_1} & \frac{\partial v_1}{\theta_2} & \cdots & \frac{\partial v_1}{\theta_p} \\ \frac{\partial v_2}{\theta_1} & \frac{\partial v_2}{\theta_2} & \cdots & \frac{\partial v_2}{\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial v_n}{\theta_1} & \frac{\partial v_n}{\theta_2} & \cdots & \frac{\partial v_n}{\theta_p} \end{bmatrix} \tag{27}$$

Term 4 of equation 3

The 4th term of equation 3 is

$$-\frac{\partial^2 v^\top}{\partial \theta_i \partial \theta_j}F^{-1}v \tag{28}$$

This is for the i,j term of the Fisher Information matrix. An equation for all terms can be written as a junction of the the Jacobian of $vec(J_v)$:

$$\mathcal{J}_v = \begin{bmatrix} \frac{\partial vec(J_v)}{\partial \theta_1} & \frac{\partial vec(J_v)}{\partial \theta_2} & \cdots & \frac{\partial vec(J_v)}{\partial \theta_p} \end{bmatrix} = \begin{bmatrix} \frac{\partial v_1}{\theta_1\theta_1} & \frac{\partial v_1}{\theta_1\theta_2} & \cdots & \frac{\partial v_1}{\theta_1\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial v_n}{\theta_1\theta_1} & \frac{\partial v_n}{\theta_1\theta_2} & \cdots & \frac{\partial v_n}{\theta_1\theta_p} \\ \frac{\partial v_1}{\theta_2\theta_1} & \frac{\partial v_1}{\theta_2\theta_2} & \cdots & \frac{\partial v_1}{\theta_2\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial v_n}{\theta_2\theta_1} & \frac{\partial v_n}{\theta_2\theta_2} & \cdots & \frac{\partial v_n}{\theta_2\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial v_1}{\theta_p\theta_1} & \frac{\partial v_1}{\theta_p\theta_2} & \cdots & \frac{\partial v_1}{\theta_p\theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial v_n}{\theta_p\theta_1} & \frac{\partial v_n}{\theta_p\theta_2} & \cdots & \frac{\partial v_n}{\theta_p\theta_p} \end{bmatrix} \tag{29}$$

The right of equation 28, $F^{-1}v$ is a $n \times 1$ matrix. We need to write this as the $np \times p$ matrix:

$$
\begin{bmatrix}
F^{-1}v & 0_{n\times 1} & \ldots & 0_{n\times 1} \\
0_{n\times 1} & F^{-1}v & \ldots & 0_{n\times 1} \\
\vdots & \vdots & \vdots & \vdots \\
0_{n\times 1} & 0_{n\times 1} & \ldots & F^{-1}v
\end{bmatrix} = I_p \otimes F^{-1}v
\tag{30}
$$

Thus the full matrix for the i,j terms in the Fisher Information matrix from term 4 of equation 3 is

$$
-\mathcal{J}_{\sqsubseteq}^{\top}(I_p \otimes F^{-1}v)
\tag{31}
$$

Term 5 of equation 3 Term 5 is

$$
\frac{\partial v^{\top}}{\partial \theta_i} F^{-1} \frac{\partial F}{\partial \theta_j} F^{-1} v
\tag{32}
$$

This is a scalar and thus its vec is equal to itself. We can rewrite equation 32 using the following relation:

$$
vec(a^{\top}ABCc) = (c^{\top}\otimes a^{\top})vec(ABC) = a^{\top}(c^{\top}\otimes I_n)vec(ABC) = c^{\top}(a^{\top}\otimes I_n)(C^{\top}\otimes A)vec(B) = c^{\top}(a^{\top}C^{\top}\otimes A)vec(B)
\tag{33}
$$

Thus equation 32 is

$$
\frac{\partial v^{\top}}{\partial \theta_i} F^{-1} \frac{\partial F}{\partial \theta_j} F^{-1} v = \frac{\partial v^{\top}}{\partial \theta_i}(v^{\top}\otimes I_n)(F^{-1}\otimes F^{-1})vec\left(\frac{\partial F}{\partial \theta_j}\right)
\tag{34}
$$

This is for the i,j term of the Fisher Information matrix. For the full matrix, we use the Jacobian of v (equation 27) and the Jacobian of vec(F) (equation 14):

$$
J_v^{\top}(v^{\top}\otimes I_n)(F^{-1}\otimes F^{-1})J_F = J_v^{\top}(v^{\top}F^{-1}\otimes F^{-1})J_F
\tag{35}
$$

Term 6 of equation 3

Term 6 is

$$
-\frac{\partial v^{\top}}{\partial \theta_i} F^{-1} \frac{\partial v}{\partial \theta_j}
\tag{36}
$$

This is for the i,j term of the Fisher Information matrix and we can write it immediately as the full matrix in terms of the Jacobian of v:

$$
\frac{\partial v^{\top}}{\partial \theta_i} F^{-1} \frac{\partial F}{\partial \theta_j} F^{-1} v = J_v^{\top} F^{-1} J_v
\tag{37}
$$

Putting all the terms together

Putting all the terms together, we have the full observed Fisher Information matrix:

$$
\frac{1}{2}J_F^{\top}(F^{-1}\otimes F^{-1})J_F - J_F^{\top}(F^{-1}vv^{\top}F^{-1}\otimes F^{-1})J_F - \frac{1}{2}(I_p\otimes[F^{-1}-F^{-1}v_tv_t^{\top}F^{-1}])\mathcal{J}_F +
$$
$$
\frac{1}{2}J_F^{\top}[F^{-1}v\otimes F^{-1}+F^{-1}\otimes F^{-1}v]J_v - \mathcal{J}_{\sqsubseteq}^{\top}(I_p\otimes F^{-1}v) + J_v^{\top}(v^{\top}F^{-1}\otimes F^{-1})J_F - J_v^{\top}F^{-1}J_v
\tag{38}
$$

7

We can simplify this a little by noting that all terms are symmetric matrices and the transpose or a symmetric matrix is equal to itself.

$$J_v^\top (v^\top F^{-1} \otimes F^{-1}) J_F = J_F^\top (F^{-1} v \otimes F^{-1}) J_v \tag{39}$$

Thus the full observed Fisher Information matrix is

$$\frac{1}{2} J_F^\top (F^{-1} \otimes F^{-1}) J_F - J_F^\top (F^{-1} v v^\top F^{-1} \otimes F^{-1}) J_F - \frac{1}{2} (I_p \otimes [F^{-1} - F^{-1} v_t v_t^\top F^{-1}]) \mathcal{J}_F +$$
$$\frac{1}{2} J_F^\top [3 F^{-1} v \otimes F^{-1} + F^{-1} \otimes F^{-1} v] J_v - \mathcal{J}_{\sqsubseteq}^\top (I_p \otimes F^{-1} v) - J_v^\top F^{-1} J_v \tag{40}$$