Closed form of hessian of negative binomial GLM with model for dispersion and mean parameter

David Fischer

August 3, 2018

Contents

1 Model description

1.1 Likelihood

The negative binomial likelihood of the GLM is:

$$L(y|\theta_i^m, \theta_i^r) = \frac{\Gamma(r(\theta^r)+y)}{y!\Gamma(r(\theta^r))}(\frac{m(\theta^m)}{r(\theta^r)+m(\theta^m)})^y(\frac{r(\theta^r)}{r(\theta^r)+m(\theta^m)})^{r(\theta^r)} \tag{1}$$

The negative binomial log-likelihood of the GLM is:

$$\begin{aligned} LL(y|\theta_i^m, \theta_i^r) &= \log L(y|\theta_i^m, \theta_i^r) \\ &= \log(\Gamma(r(\theta^r) + y)) - \log(y!\Gamma(r(\theta^r))) + y * (\log(m(\theta^m)) - \log(r(\theta^r) + m(\theta^m))) \\ &+ r(\theta^r) * (\log(r(\theta^r)) - \log(r(\theta^r) + m(\theta^m))) \end{aligned} \tag{2}$$

1.2 Parameter models

The mean (m) and dispersion (r) model of the GLM with design matrices X^m and X^r look as follows:

$$r(\theta^r) = \exp(\langle X^r, \theta_i^r \rangle) \tag{3}$$

$$m(\theta^m) = \exp(\langle X^m, \theta_i^m \rangle) \tag{4}$$

1.3 Derivatives of parameter models

Using the chain rule $\frac{f(m)}{d\theta_i^m} = \frac{f(m)}{m} \frac{m}{d\theta_i^m}$ we can decompose the first derivative. Firstly, the derivative of the mean model with respect to its parameters is:

$$\begin{aligned} \frac{d}{d\theta_i^m} m(\theta^m) &= \frac{d}{d\theta_i^m} \exp(\langle X^m, \theta_i^m \rangle) \\ &= \exp(\langle X^m * \theta^m \rangle) * X_i^m \\ &= m(\theta^m) * X_i^m \end{aligned} \tag{5}$$

Where X_i^m is the column of X^m that corresponds to θ_i^m. Equivalently for the dispersion model:

$$\begin{aligned} \frac{d}{d\theta_i^r} r(\theta^r) &= \frac{d}{d\theta_i^r} \exp(\langle X^r, \theta_i^r \rangle) \\ &= \exp(\langle X^r * \theta^r \rangle) * X_i^r \\ &= r(\theta^r) * X_i^r \end{aligned} \tag{6}$$

2 Miscancellous

2.1 Polygamma function

In the following, we also need the polygamma function ψ to compute the derivative of the gamma function:

$$\frac{d^{(n+1)}}{dy^{(n+1)}} \log(\Gamma(y)) = \psi_n(y) \tag{7}$$

$$\begin{aligned} \frac{d}{d\theta_i^r} \log \Gamma(r(\theta^r)) &= \frac{d}{d\theta_i^r} \Big(r(\theta^r) \Big) * \psi_0(r(\theta^r)) \\ &= r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \end{aligned} \tag{8}$$

$$\begin{aligned} \frac{d^2}{d(\theta_i^r)^2} \log \Gamma(r(\theta^r)) &= \frac{d}{d\theta_i^r} \frac{d}{d\theta_i^r} \log(\Gamma(y)) \\ &= \frac{d}{d\theta_i^r} \Big(r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \Big) \\ &= X_i^r * \frac{d}{d\theta_i^r} \Big(r(\theta^r) \Big) * \psi_0(r(\theta^r)) + X_i^r * r(\theta^r) * \frac{d}{d\theta_i^r} \Big(\psi_0(r(\theta^r)) \Big) \\ &= X_i^r * r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) + X_i^r * r(\theta^r) * r(\theta^r) * X_i^r * \psi_1(r(\theta^r)) \Big) \\ &= r(\theta^r) * (X_i^r)^2 * (\psi_0(r(\theta^r)) + r(\theta^r) * \psi_1(r(\theta^r))) \end{aligned} \tag{9}$$

$\psi_0(y)$ and $\psi_1(y)$ can be directly evaluated with *scipy.special.polygamma*().

Note that due to the chain rule, both homogenous and heterogenous secondary derivatives of the polygamma function can be computed via ψ_1:

$$\begin{aligned}\frac{d^2}{dx_i x_j}\log\Gamma(y(x)) &= \frac{dy}{dx_i}\frac{dy}{dx_j}\frac{d^2}{dy^2}\log\Gamma(y(x)) \\ &= \frac{dy}{dx_i}\frac{dy}{dx_j}\psi_1(y(x))\end{aligned} \tag{10}$$

$$\begin{aligned}\frac{d^2}{d\theta_i^r\theta_j^r}\log\Gamma(r(\theta^r)) &= \frac{d}{d\theta_j^r}\frac{d}{d\theta_i^r}\log(\Gamma(y)) \\ &= \frac{d}{d\theta_j^r}\Big(r(\theta^r) * X_i^r * \psi_0(r(\theta^r))\Big) \\ &= X_i^r * \frac{d}{d\theta_j^r}\Big(r(\theta^r)\Big) * \psi_0(r(\theta^r)) + X_i^r * r(\theta^r) * \frac{d}{d\theta_j^r}\Big(\psi_0(r(\theta^r))\Big) \\ &= X_i^r * r(\theta^r) * X_j^r * \psi_0(r(\theta^r)) + X_i^r * r(\theta^r) * r(\theta^r) * X_j^r * \psi_1(r(\theta^r))\Big) \\ &= r(\theta^r) * X_i^r * X_j^r * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r)))\end{aligned} \tag{11}$$

3 Jacobians

3.1 Jacobian mean model

The Jacobians of the GLM wrt to the m model is:

$$\begin{aligned}\frac{d}{d\theta_i^m}LL(y|\theta^m,\theta^r) &= \frac{d}{d\theta_i^m}\log(\Gamma(r(\theta^r)+y)) - \frac{d}{d\theta_i^m}\log(y!\Gamma(r(\theta^r))) \\ &+ \frac{d}{d\theta_i^m}y * \Big(\log(m(\theta^m)) - \log(r(\theta^r)+m(\theta^m))\Big) \\ &+ \frac{d}{d\theta_i^m}r(\theta^r)\Big(\log(r(\theta^r)) - \log(r(\theta^r)+m(\theta^m))\Big) \\ &= y * \Big(\frac{d}{d\theta_i^m}\log(m(\theta^m)) - \frac{d}{d\theta_i^m}\log(r(\theta^r)+m(\theta^m))\Big) \\ &+ r(\theta^r) * \Big(\frac{d}{d\theta_i^m}\log(r(\theta^r)) - \frac{d}{d\theta_i^m}(\log(r(\theta^r)) + m(\theta^m))\Big) \\ &= y * (\frac{1}{m(\theta^m)}\frac{dm(\theta^m)}{d\theta_i^m} - \frac{1}{r(\theta^r)+m(\theta^m)}\frac{dm(\theta^m)}{d\theta_i^m}) \\ &- \frac{r(\theta^r)}{r(\theta^r)+m(\theta^m)}\frac{dm(\theta^m)}{d\theta_i^m} \\ &= \frac{dm(\theta^m)}{d\theta_i^m}(\frac{y}{m(\theta^m)} - \frac{y+r(\theta^r)}{r(\theta^r)+m(\theta^m)}) \\ &= m(\theta^m) * X_i^m(\frac{y}{m(\theta^m)} - \frac{y+r(\theta^r)}{r(\theta^r)+m(\theta^m)}) \\ &= X_i^m * y - X_i^m * (y + r(\theta^r)) * \frac{m(\theta^m)}{r(\theta^r)+m(\theta^m)}\end{aligned} \tag{12}$$

3.2 Jacobian dispersion model

The Jacobians of the GLM wrt to the r model is:

$$\begin{aligned}
\frac{d}{d\theta_i^r} LL(y|\theta^m, \theta^r) &= \frac{d}{d\theta_i^r} \log(\Gamma(r(\theta^r) + y)) - \frac{d}{d\theta_i^r} \log(y!\Gamma(r(\theta^r))) \\
&+ y * \frac{d}{d\theta_i^r} \Bigg(\log(m(\theta^m)) - \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ \frac{d}{d\theta_i^r} \Bigg(r(\theta^r) * \Bigg(\log(r(\theta^r)) - \log(r(\theta^r) + m(\theta^m)) \Bigg) \Bigg) \\
&= r(\theta^r) * X_i^r * \psi_0(r(\theta^r) + y) + r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \\
&- y * \frac{d}{d\theta_i^r} \Bigg(\log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ \frac{d}{d\theta_i^r} \Bigg(r(\theta^r) * \log(r(\theta^r)) \Bigg) - \frac{d}{d\theta_i^r} \Bigg(r(\theta^r) \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&= r(\theta^r) * X_i^r * \psi_0(r(\theta^r) + y) + r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \\
&- y * \frac{1}{r(\theta^r) + m(\theta^m)} r(\theta^r) * X_i^r \\
&+ \Bigg(r(\theta^r) * X_i^r * \log(r(\theta^r)) + r(\theta^r) * r(\theta^r) * X_i^r * \frac{1}{r(\theta^r)} \Bigg) \\
&- \Bigg(r(\theta^r) * X_i^r * \log(r(\theta^r) + m(\theta^m)) + r(\theta^r) * r(\theta^r) * X_i^r * \frac{1}{r(\theta^r) + m(\theta^m)} \Bigg) \\
&= r(\theta^r) * X_i^r * \psi_0(r(\theta^r) + y) + r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \\
&- \frac{1}{r(\theta^r) + m(\theta^m)} r(\theta^r) * X_i^r * (r(\theta^r) + y) \\
&+ r(\theta^r) * X_i^r * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg)
\end{aligned} \tag{13}$$

4 Block-wise entries of hessian matriy

The hessian can be decomposed into a block-wise hessian of the form $H = [[H^{m,m}, H^{m,r}], [H^{r,m}, H^{r,r}]]$, where $H^{r,m} = H^{m,r}$ as H is symmetric.

4.1 Mean-model block $H^{m,m}$

4.1.1 Mean-model block diagonal elements

The diagonal elements of the mean-model-block of the hessian (using the quotient rule) are:

$$
\begin{aligned}
H^{m,m}_{i,i} &= \frac{d^2}{d\theta^m_i d\theta^m_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^m_i} \frac{d}{d\theta^m_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^m_i} \left(X^m_i * y - X^m_i * (y + r(\theta^r)) * \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \right) \\
&= -X^m_i * (y + r(\theta^r)) \frac{d}{d\theta^m_i} \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \\
&= -X^m_i * (y + r(\theta^r)) * \frac{\frac{d}{d\theta^m_i}\left(m(\theta^m) \right) * (r(\theta^r) + m(\theta^m)) - m(\theta^m) * \frac{d}{d\theta^m_i}\left(r(\theta^r) + m(\theta^m) \right)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * (y + r(\theta^r)) * \frac{m(\theta^m) * X^m_i * (r(\theta^r) + m(\theta^m)) - m(\theta^m) * m(\theta^m) * X^m_i}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * (y + r(\theta^r)) * m(\theta^m) * X^m_i * \frac{r(\theta^r) + m(\theta^m) - m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * (y + r(\theta^r)) * m(\theta^m) * X^m_i * \frac{r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * X^m_i * m(\theta^m) * \frac{y/r(\theta^r) + 1}{(1 + m(\theta^m)/r(\theta^r))^2}
\end{aligned} \tag{14}
$$

4.1.2 Mean-model block off-diagonal elements

The off-diagonal elements of the mean-model-block of the hessian are:

$$
\begin{aligned}
H^{m,m}_{i,j} &= \frac{d^2}{d\theta^m_i d\theta^m_j} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^m_j} \frac{d}{d\theta^m_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^m_j} \left(X^m_i * y - X^m_i * (y + r(\theta^r)) * \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \right) \\
&= -X^m_i * (y + r(\theta^r)) \frac{d}{d\theta^m_j} \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \\
&= -X^m_i * (y + r(\theta^r)) * \frac{\frac{d}{d\theta^m_j}\left(m(\theta^m) \right) * (r(\theta^r) + m(\theta^m)) - m(\theta^m) * \frac{d}{d\theta^m_j}\left(r(\theta^r) + m(\theta^m) \right)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * (y + r(\theta^r)) * \frac{m(\theta^m) * X^m_j * (r(\theta^r) + m(\theta^m)) - m(\theta^m) * m(\theta^m) * X^m_j}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * (y + r(\theta^r)) * m(\theta^m) * X^m_j * \frac{r(\theta^r) + m(\theta^m) - m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * (y + r(\theta^r)) * m(\theta^m) * X^m_j * \frac{r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * X^m_j * m(\theta^m) * \frac{y/r(\theta^r) + 1}{(1 + m(\theta^m)/r(\theta^r))^2}
\end{aligned} \tag{15}
$$

4.1.3 Mean-model block elements

The equation is always the same and does not simplify for the diagonal:

$$H_{i,j}^{m,m} = -X_i^m * X_j^m * m(\theta^m) * \frac{y/r(\theta^r) + 1}{(1 + m(\theta^m)/r(\theta^r))^2} \tag{16}$$

This is also the result from Zwilling et al.

4.2 Dispersion-model block $H^{r,r}$

4.2.1 Dispersion-model block diagonal elements

The diagonal elements of the dispersion-model-block of the hessian (using the quotient rule) are: This should yield the same apart of i, j as the off-diagonal elements in the next section. This is true.

$$
\begin{aligned}
H^{r,r}_{i,i} &= \frac{d^2}{d\theta^r_i d\theta^r_i} LL(y|\theta^m,\theta^r) \\
&= \frac{d}{d\theta^r_i}\frac{d}{d\theta^m_i} LL(y|\theta^m,\theta^r) \\
&= \frac{d}{d\theta^r_i}\Bigg(r(\theta^r) * X^r_i * \psi_0(r(\theta^r)+y) + r(\theta^r) * X^r_i * \psi_0(r(\theta^r)) \\
&- \frac{1}{r(\theta^r)+m(\theta^m)} r(\theta^r) * X^r_i * (r(\theta^r)+y) \\
&+ r(\theta^r) * X^r_i * \bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m)) \bigg) \Bigg) \\
&= \frac{d}{d\theta^r_i}\left(\frac{d}{d\theta^r_i} \log\Gamma(r(\theta^r)+y) \right) + \frac{d}{d\theta^r_i}\left(\frac{d}{d\theta^r_i} \log\Gamma(r(\theta^r)) \right) \\
&- \frac{d}{d\theta^r_i}\left(\frac{1}{r(\theta^r)+m(\theta^m)} r(\theta^r) * X^r_i * (r(\theta^r)+y) \right) \\
&+ \frac{d}{d\theta^r_i}\left(r(\theta^r) * X^r_i * \bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m)) \bigg) \right) \\
&= r(\theta^r) * (X^r_i)^2 * \psi_0(r(\theta^r)+y) + r(\theta^r) * \psi_1(r(\theta^r)+y)) \\
&+ r(\theta^r) * (X^r_i)^2 * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \frac{d}{d\theta^r_i}\left(\frac{r(\theta^r) * X^r_i}{r(\theta^r)+m(\theta^m)} * (r(\theta^r)+y) \right) \\
&+ \frac{d}{d\theta^r_i}\left(r(\theta^r) * X^r_i * \bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m)) \bigg) \right) \\
&= r(\theta^r) * (X^r_i)^2 * \psi_0(r(\theta^r)+y) + r(\theta^r) * \psi_1(r(\theta^r)+y)) \\
&+ r(\theta^r) * (X^r_i)^2 * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \left(\frac{d}{d\theta^r_i}\frac{r(\theta^r) * X^r_i}{r(\theta^r)+m(\theta^m)} * (r(\theta^r)+y) + \frac{r(\theta^r) * X^r_i}{r(\theta^r)+m(\theta^m)} * \frac{d}{d\theta^r_i}(r(\theta^r)+y) \right) \\
&+ X^r_i * \left(\frac{d}{d\theta^r_i}\bigg(r(\theta^r) \bigg) * \bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m)) \right) \\
&+ r(\theta^r) * \frac{d}{d\theta^r_i}\bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m)) \bigg) \bigg) \\
&= r(\theta^r) * (X^r_i)^2 * \psi_0(r(\theta^r)+y) + r(\theta^r) * \psi_1(r(\theta^r)+y)) \\
&+ r(\theta^r) * (X^r_i)^2 * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \left(\frac{r(\theta^r) * X^r_i * X^r_i * (r(\theta^r)+m(\theta^m)) - r(\theta^r) * X^r_i * r(\theta^r) * X^r_i}{(r(\theta^r)+m(\theta^m))^2} * (r(\theta^r)+y) + \frac{r(\theta^r) * X^r_i}{r(\theta^r)+m(\theta^m)} * r(\theta^r) * X^r_i \right) \\
&+ X^r_i * \left(r(\theta^r) * X^r_i * \bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m)) \right) \\
&+ r(\theta^r) * \left(r(\theta^r) * X^r_i \frac{1}{r(\theta^r)} - r(\theta^r) * X^r_i \frac{1}{r(\theta^r)+m(\theta^m)} \right) \bigg)
\end{aligned} \tag{17}
$$

$$
\begin{aligned}
H_{i,i}^{r,r} &= r(\theta^r) * (X_i^r)^2 * \psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y)) \\
&+ r(\theta^r) * (X_i^r)^2 * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \left(\frac{r(\theta^r) * X_i^r * X_i^r * m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} * (r(\theta^r) + y) + \frac{r(\theta^r) * X_i^r}{r(\theta^r) + m(\theta^m)} * r(\theta^r) * X_i^r \right) \\
&+ X_i^r * \Bigg(r(\theta^r) * X_i^r * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ r(\theta^r) * X_i^r * \left(1 - \frac{r(\theta^r)}{r(\theta^r) + m(\theta^m)} \right) \Bigg) \\
&= r(\theta^r) * (X_i^r)^2 * \psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y)) \\
&+ r(\theta^r) * (X_i^r)^2 * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \frac{r(\theta^r) * (X_i^r)^2 * m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} * (r(\theta^r) + y) - \frac{r(\theta^r) * X_i^r}{r(\theta^r) + m(\theta^m)} * r(\theta^r) * X_i^r \\
&+ r(\theta^r) * (X_i^r)^2 * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ r(\theta^r) * (X_i^r)^2 * \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \\
&= r(\theta^r) * (X_i^r)^2 * \psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y)) \\
&+ r(\theta^r) * (X_i^r)^2 * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \frac{r(\theta^r) * (X_i^r)^2 * m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} * (r(\theta^r) + y) \\
&+ r(\theta^r) * (X_i^r)^2 * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ r(\theta^r) * (X_i^r)^2 * \frac{m(\theta^m) - r(\theta^r)}{r(\theta^r) + m(\theta^m)} \\
&= r(\theta^r) * (X_i^r)^2 * \Bigg(\psi_0(r(\theta^r) + y) + \psi_0(r(\theta^r)) - \frac{m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} * (r(\theta^r) + y) \\
&+ \log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) + \frac{m(\theta^m) - r(\theta^r)}{r(\theta^r) + m(\theta^m)} \Bigg) \\
&+ r(\theta^r) * \psi_1(r(\theta^r) + y)) + r(\theta^r) * \psi_1(r(\theta^r)))
\end{aligned} \tag{18}
$$

4.2.2 Dispersion-model block off-diagonal elements

The off-diagonal elements of the dispersion-model-block of the hessian are:

$$
\begin{aligned}
H^{r,r}_{i,j} &= \frac{d^2}{d\theta^r_i d\theta^r_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^r_j} \frac{d}{d\theta^m_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^r_j} \Bigg(r(\theta^r) * X^r_i * \psi_0(r(\theta^r) + y) + r(\theta^r) * X^r_i * \psi_0(r(\theta^r)) \\
&- \frac{1}{r(\theta^r) + m(\theta^m)} r(\theta^r) * X^r_i * (r(\theta^r) + y) \\
&+ r(\theta^r) * X^r_i * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \Bigg) \\
&= \frac{d}{d\theta^r_j} \Bigg(\frac{d}{d\theta^r_i} \log\Gamma(r(\theta^r) + y) \Bigg) + \frac{d}{d\theta^r_j} \Bigg(\frac{d}{d\theta^r_i} \log\Gamma(r(\theta^r)) \Bigg) \\
&- \frac{d}{d\theta^r_j} \Bigg(\frac{1}{r(\theta^r) + m(\theta^m)} r(\theta^r) * X^r_i * (r(\theta^r) + y) \Bigg) \\
&+ \frac{d}{d\theta^r_j} \Bigg(r(\theta^r) * X^r_i * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \Bigg) \\
&= r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y)) \\
&+ r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \frac{d}{d\theta^r_j} \Bigg(\frac{r(\theta^r) * X^r_i}{r(\theta^r) + m(\theta^m)} * (r(\theta^r) + y) \Bigg) \\
&+ \frac{d}{d\theta^r_j} \Bigg(r(\theta^r) * X^r_i * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \Bigg) \\
&= r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y)) \\
&+ r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \Bigg(\frac{d}{d\theta^r_j} \frac{r(\theta^r) * X^r_i}{r(\theta^r) + m(\theta^m)} * (r(\theta^r) + y) + \frac{r(\theta^r) * X^r_i}{r(\theta^r) + m(\theta^m)} * \frac{d}{d\theta^r_j} (r(\theta^r) + y) \Bigg) \\
&+ X^r_i * \Bigg(\frac{d}{d\theta^r_j} \Bigg(r(\theta^r) \Bigg) * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ r(\theta^r) * \frac{d}{d\theta^r_j} \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \Bigg) \\
&= r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y)) \\
&+ r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \Bigg(\frac{r(\theta^r) * X^r_j * X^r_i * (r(\theta^r) + m(\theta^m)) - r(\theta^r) * X^r_i * r(\theta^r) * X^r_j}{(r(\theta^r) + m(\theta^m))^2} * (r(\theta^r) + y) + \frac{r(\theta^r) * X^r_i}{r(\theta^r) + m(\theta^m)} * r(\theta^r) * X^r_j \Bigg) \\
&+ X^r_i * \Bigg(r(\theta^r) * X^r_j * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ r(\theta^r) * \Bigg(r(\theta^r) * X^r_j \frac{1}{r(\theta^r)} - r(\theta^r) * X^r_j \frac{1}{r(\theta^r) + m(\theta^m)} \Bigg) \Bigg)
\end{aligned} \tag{19}
$$

$$
\begin{aligned}
H^{r,r}_{i,j} &= r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y)) \\
&+ r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \left(\frac{r(\theta^r) * X^r_j * X^r_i * m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} * (r(\theta^r) + y) + \frac{r(\theta^r) * X^r_i}{r(\theta^r) + m(\theta^m)} * r(\theta^r) * X^r_j \right) \\
&+ X^r_i * \Bigg(r(\theta^r) * X^r_j * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ r(\theta^r) * X^r_j * \left(1 - \frac{r(\theta^r)}{r(\theta^r) + m(\theta^m)} \right) \Bigg) \\
&= r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y)) \\
&+ r(\theta^r) * X^r_i * X^r_j * \psi_0(r(\theta^r) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&- \Bigg(r(\theta^r) * X^r_j * X^r_i * \frac{m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} * (r(\theta^r) + y) - r(\theta^r) * X^r_i * X^r_j * \frac{r(\theta^r)}{r(\theta^r) + m(\theta^m)} \\
&+ r(\theta^r) * X^r_i * X^r_j \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ r(\theta^r) * X^r_i * X^r_j \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \Bigg) \\
&= r(\theta^r) * X^r_i * X^r_j * \Bigg(\psi_0(r(\theta^r) + y) + \psi_0(r(\theta^r) - \frac{m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} * (r(\theta^r) + y) \\
&- \frac{r(\theta^r)}{r(\theta^r) + m(\theta^m)} + \log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) + \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \Bigg) \\
&+ r(\theta^r) * \psi_1(r(\theta^r) + y)) + r(\theta^r) * \psi_1(r(\theta^r))) \\
&= r(\theta^r) * X^r_i * X^r_j * \Bigg(\psi_0(r(\theta^r) + y) + \psi_0(r(\theta^r) - \frac{m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} * (r(\theta^r) + y) \\
&+ \frac{m(\theta^m) - r(\theta^r)}{r(\theta^r) + m(\theta^m)} + \log(r(\theta^r)) + 1 - \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&+ r(\theta^r) * \psi_1(r(\theta^r) + y)) + r(\theta^r) * \psi_1(r(\theta^r)))
\end{aligned}
\tag{20}
$$

4.3 Offdiagonal-model block $H^{r,m}$

Two-alternative derivations which should and do yield the same.

4.3.1 Offdiagonal-model block derived from mean-model jacobian

The off-diagonal elements of the off-block-diagonal hessian (mean-dispersion-model-block) are:

$$
\begin{aligned}
H^{r,m}_{i,j} &= \frac{d^2}{d\theta^m_i d\theta^r_j} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^r_j} \frac{d}{d\theta^m_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^r_j} \left(X^m_i * y - X^m_i * (y + r(\theta^r)) * \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \right) \\
&= -X^m_i * \frac{d}{d\theta^r_j} \left((y + r(\theta^r)) \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \right) \\
&= -X^m_i * \left(r(\theta^r) * X^r_j * \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} - (y + r(\theta^r)) * r(\theta^r) * X^r_j * \frac{m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} \right) \\
&= -m(\theta^m) * X^m_i * r(\theta^r) * X^r_j * \left(\frac{1}{r(\theta^r) + m(\theta^m)} - \frac{y + r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2} \right) \\
&= -m(\theta^m) * X^m_i * r(\theta^r) * X^r_j * \frac{r(\theta^r) + m(\theta^m) - y - r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -m(\theta^m) * X^m_i * r(\theta^r) * X^r_j * \frac{m(\theta^m) - y}{(r(\theta^r) + m(\theta^m))^2} \\
&= m(\theta^m) * X^m_i * r(\theta^r) * X^r_j * \frac{y - m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} \\
&= X^m_i * X^r_j * m(\theta^m) * \frac{y - m(\theta^m)}{(1 + m(\theta^m)/r(\theta^r))^2}
\end{aligned}
\tag{21}
$$

4.3.2 Offdiagonal-model block derived from dispersion-model jacobian

Validation of the off-diagonal elements of the off-block-diagonal hessian (mean-dispersion-model-block) via the Jacobian of the dispersion model:

$$
\begin{aligned}
H^{r,m}_{i,j} &= \frac{d^2}{d\theta^m_i d\theta^r_j} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^m_j}\frac{d}{d\theta^r_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^m_j}\Bigg(r(\theta^r) * X^r_i * \psi_0(r(\theta^r)+y) + r(\theta^r) * X^r_i * \psi_0(r(\theta^r)) \\
&- \frac{1}{r(\theta^r)+m(\theta^m)} r(\theta^r) * X^r_i * (r(\theta^r)+y) \\
&+ r(\theta^r) * X^r_i * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m))\Bigg)\Bigg) \\
&= -\frac{d}{d\theta^m_j}\left(\frac{1}{r(\theta^r)+m(\theta^m)} r(\theta^r) * X^r_i * (r(\theta^r)+y)\right) \\
&+ \frac{d}{d\theta^m_j}\Bigg(r(\theta^r) * X^r_i * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m))\Bigg)\Bigg) \\
&= -r(\theta^r) * X^r_i * (r(\theta^r)+y) * \frac{d}{d\theta^m_j}\frac{1}{r(\theta^r)+m(\theta^m)} \\
&- r(\theta^r) * X^r_i * \frac{d}{d\theta^m_j}\Bigg(\log(r(\theta^r)+m(\theta^m))\Bigg) \\
&= r(\theta^r) * X^r_i * (r(\theta^r)+y) * m(\theta^m) * X^m_j \frac{1}{(r(\theta^r)+m(\theta^m))^2} \\
&- r(\theta^r) * X^r_i * m(\theta^m) * X^m_j \frac{1}{r(\theta^r)+m(\theta^m)} \\
&= r(\theta^r) * X^r_i * m(\theta^m) * X^m_j \left(\frac{r(\theta^r)+y}{(r(\theta^r)+m(\theta^m))^2} - \frac{1}{r(\theta^r)+m(\theta^m)}\right) \\
&= r(\theta^r) * X^r_i * m(\theta^m) * X^m_j \frac{r(\theta^r)+y-r(\theta^r)-m(\theta^m)}{(r(\theta^r)+m(\theta^m))^2} \\
&= r(\theta^r) * X^r_i * m(\theta^m) * X^m_j \frac{y-m(\theta^m)}{(r(\theta^r)+m(\theta^m))^2} \\
&= X^m_i * X^r_j * m(\theta^m) * \frac{y-m(\theta^m)}{(1+m(\theta^m)/r(\theta^r))^2}
\end{aligned} \tag{22}
$$

4.3.3 Offdiagonal-model block elements

$$
H^{r,m}_{i,j} = X^m_i * X^r_j * m(\theta^m) * \frac{y-m(\theta^m)}{(1+m(\theta^m)/r(\theta^r))^2} \tag{23}
$$

This is the result from Zwilling et al. $*-1$.