Chapter 7

Bias Corrections for a Class of Generalized Linear Models

This chapter seeks to demonstrate the applicability of the projective score methodology in a challenging and useful class of problems. The models to be analyzed have both multivariate nuisance parameters *and* multivariate parameters of interest. The rationale for using the projective scores is that in these models, a conditional approach is possible, but in general intractable. The projective score will be used an an approximation to the conditional score, an approach justified by Corollary 2.1 and Section 4 of Chapter 2. A possible computational difficulty associated with the fact that the number of elements in the Bhattacharyya basis of order t is a polynomial in s^t, is shown *not* to be a problem for the calculation of the U_2 projective score for the class of Generalized Linear Models with canonical link, here under consideration.

7.1 A motivating example

Example 1 *Paired Poisson*

The motivation for the model to be considered comes from a "paired Poisson" experiment. In particular, two independent Poisson variates are observed with means m_1 and m_2, where $m_1 = \exp(\omega)$ and $m_2 = \exp(\omega + \beta)$. Here ω is the nuisance parameter and β is the parameter of interest. β is the log of the ratio of the means, that is $\beta = \log(m_1) - \log(m_2)$. The standard analysis is to condition on a complete and sufficient statistic for the nuisance parameter, here it is $Y_1 + Y_2$. The conditional distribution of Y_2 given that $Y_1 + Y_2 = y_1 + y_2$, is Binomial with parameters $y_1 + y_2$ and $\frac{\exp(\beta)}{1+\exp(\beta)}$. In this example, by conditioning, the problem has been reduced to one in which the conditional distribution is free of ω, the nuisance parameter. The reason for this success is because the conditioning statistic $Y_1 + Y_2$, does not depend on β, so that the problem is a Type I problem in the terminology of Chapter 2. This model has the interesting property that the U_1 score equation is equal to the conditional, U_c score equation.

Example 6 can be re-parameterized into a model in which transformed parameters, β_1 and ω_1 are defined as:

$$\beta_1 = \tfrac{1}{2}\left(\log m_2 - \log m_1\right) \qquad \omega_1 = \tfrac{1}{2}\left(\log m_2 + \log m_1\right),$$

so that

$$\log(m_1) = \omega_1 - \beta_1 \quad \text{and} \quad \log(m_2) = \omega_1 + \beta_1.$$

It follows that the model can be written as:

$$\boldsymbol{Y} \sim Po\left(\exp(Z\beta_1 + \Gamma\omega_1)\right)$$

where

$$Z = \begin{pmatrix} -1 \\ +1 \end{pmatrix} \qquad \Gamma = \begin{pmatrix} +1 \\ +1 \end{pmatrix} \qquad \text{and } Z \perp \Gamma.$$

This parameterization leads to a sensible extension to one of n independent observations: consider n independent Poisson random variables

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim Po(m) = Po\begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}$$

where the mean vector m is assumed to be of the form

$$\log(\boldsymbol{m}) = (\mathbf{Z}, \boldsymbol{\Gamma}) \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\omega} \end{pmatrix} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\Gamma}\boldsymbol{\omega}.$$

Both $\mathbf{Z}$ and $\boldsymbol{\Gamma}$ are taken to be of full rank, and the sum of their ranks is equal to n. Since there are n parameters in the model, the model is saturated. The $\mathbf{Z}$ matrix contains the contrasts of interest, so that using the fact that $\mathbf{Z} \perp \boldsymbol{\Gamma}$, β can be represented as

$$\boldsymbol{\beta} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1} \mathbf{Z}^T \log(\boldsymbol{m}),$$

which explicitly displays its nature as a contrast in the logs of the cell means.

Writing the rows of the $\mathbf{Z}$ and $\boldsymbol{\Gamma}$ matrices, associated with the i^{th} obseravtion as z_i and γ_i, the likelihood for the model is

$$L(\beta, \omega; y) = \prod_{i=1}^{n} \left[\frac{\exp(-\exp(z_i\beta + \gamma_i\omega)) \exp((z_i\beta + \gamma_i\omega)y_i)}{y_i!} \right]$$

and the log-likelihood is

$$l(\beta, \omega; y) = \sum_{i=1}^{n} \left(-\exp(z_i\beta + \gamma_i\omega) + (z_i\beta + \gamma_i\omega)y_i + k(y)\right). \tag{7.1}$$

The log-likelihood in (7.1) can be generalized in the following way: define the matrix $\mathbf{X}$ to be $(\mathbf{Z}, \boldsymbol{\Gamma})$, and the vector θ by $(\beta, \omega)^T$. Denote the i^{th} row of the matrix $\mathbf{X}$ by x_i, and finally denote $x_i\theta$ by η_i. Then the kernel of the log-likelihood in (7.1) can be written as

$$l(\theta; y) = \sum_{i=1}^{n} y_i\eta_i - \sum_{i=1}^{n} b_i(\eta_i), \tag{7.2}$$

where $b_i(\eta_i) = \exp(x_i\theta)$. (7.2) is recognizable as a generalized log-linear model with a canonical link, in which the dispersion parameter ϕ is taken to be known. The model (7.2) will be taken as the starting point for this chapter, the cases where the random component in the model is Poisson, Gamma or Binomial will considered, but a specialization back to the Poisson case will be made in order to perform some numerical calculations. It is important to realize that the models considered are *saturated models*, but in which there is specific interest in a subset of the parameters.

7.2 Projective score calculations

The first and second order projective scores, $\mathbf{U}_1$ and $\mathbf{U}_2$ will be derived. Denoting the j^{th} component of the score function for the parameter of interest β by U_0^j, then on differentiating (7.2) with respect to β^j,

$$U_0^j = \frac{\partial}{\partial \beta^j}\left(\sum_{i=1}^{n} y_i \eta_i - \sum_{i=1}^{n} b_i(\eta_i)\right) = \sum_i y_i z_{ij} - \sum_i b'(\eta_i) z_{ij},$$

where $'$ denotes differentiation with respect to η_i. It follows that

$$\mathbf{U}_0 = \mathbf{Z}^T(\mathbf{y} - b'(\eta))$$

where $b'(\eta)$ is the vector with i^{th} component $b'(\eta_i)$. By analogy, the score for the nuisance parameter $\mathbf{V}_1$ is

$$\mathbf{V}_1 = \boldsymbol{\Gamma}^T(\mathbf{y} - b'(\eta)).$$

Using the recurrence relation given in Theorem 3.1 it can be shown that the second order basis element V_{ij} is the $(i,j)^{th}$ component of the matrix

$$\mathbf{V}_2 = \boldsymbol{\Gamma}^T \left[(\mathbf{y} - b'(\eta))\,(\mathbf{y} - b'(\eta))^T - b''(\eta)\right] \boldsymbol{\Gamma},$$

where $b''(\eta)$ is a diagonal matrix with elements $b''(\eta_i)$. For notational simplicity, the vector $b'(\eta)$ will be written as $\mathbf{m}$, and the diagonal matrix $b''(\eta)$ as $\mathbf{M}$. It is useful to recognize that $\mathbf{m}$ is simply the mean of $\mathbf{Y}$, and that $\mathbf{M}$ is the variance covariance matrix of $\mathbf{Y}$.

In this new notation the vectors U_0 and $\mathbf{V}_1$ are $\mathbf{Z}^T(\mathbf{Y}-\mathbf{m})$ and $\boldsymbol{\Gamma}^T(\mathbf{Y}-\mathbf{m})$ respectively. The matrix $\mathbf{V}_2$ is given by

$$\boldsymbol{\Gamma}^T\left[(\mathbf{y}-\mathbf{m})(\mathbf{y}-\mathbf{m})^T-\mathbf{M}\right]\boldsymbol{\Gamma}.$$

Theorem 7.1 *The first projective score is given by*

$$\mathbf{U}_1=\mathbf{Z}^T\,\boldsymbol{\Xi}(\mathbf{y}-\mathbf{m})$$

where

$$\boldsymbol{\Xi}=\mathbf{Z}\,(\mathbf{Z}^T\,\mathbf{M}^{-1}\,\mathbf{Z})^{-1}\,\mathbf{Z}^T\,\mathbf{M}^{-1}.$$

Proof

Following Chapter 2 (2.16), in order to find the first projective score $U_1=\mathbf{U}_0-A\mathbf{V}_1$, the quantity

$$A=E[\mathbf{U}_0\mathbf{V}_1^T]\,(E[\mathbf{V}_1\mathbf{V}_1^T])^{-1}$$

must be calculated (as usual it is assumed that $E[\mathbf{V}_1\mathbf{V}_1^T]$ is invertible). Now

$$E[\mathbf{U}_0\mathbf{V}_1^T]=E[\mathbf{Z}^T\,(\mathbf{y}-\mathbf{m})\,(\mathbf{y}-\mathbf{m})^T\,\mathbf{\Gamma}]=\mathbf{Z}^T\,E[(\mathbf{y}-\mathbf{m})\,(\mathbf{y}-\mathbf{m})^T]\,\boldsymbol{\Gamma}=\mathbf{Z}^T\,\mathbf{M}\,\boldsymbol{\Gamma}. \tag{7.3}$$

In a similar fashion

$$E[\mathbf{V}_1\mathbf{V}_1^T]=E[\mathbf{\Gamma}^T\,(\mathbf{y}-\mathbf{m})\,(\mathbf{y}-\mathbf{m})^T\,\mathbf{\Gamma}]=\mathbf{\Gamma}^T\,E[(\mathbf{y}-\mathbf{m})\,(\mathbf{y}-\mathbf{m})^T]\,\mathbf{\Gamma}=\mathbf{\Gamma}^T\,\mathbf{M}\,\mathbf{\Gamma} \tag{7.4}$$

This last matrix, (7.4), will be invertible provided $\mathbf{M}$ is full rank, which in the Poisson case is equivalent to none of the cell means being equal be zero. Putting (7.3) and (7.4) together yields:

$$A=\mathbf{Z}^T\,\mathbf{M}\,\mathbf{\Gamma}\,(\mathbf{\Gamma}^T\,\mathbf{M}\,\mathbf{\Gamma})^{-1},$$

and

$$\mathbf{U}_1 = \mathbf{Z}^T(\mathbf{y}-\mathbf{m}) - \mathbf{Z}^T\mathbf{M}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T(\mathbf{y}-\mathbf{m}) \quad (7.5)$$

$$= \mathbf{Z}^T(\mathbf{I} - \mathbf{M}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T)(\mathbf{y}-\mathbf{m}). \quad (7.6)$$

Because of the orthogonality of $\mathbf{Z}$ and $\boldsymbol{\Gamma}$, (7.6) can be rewritten without reference to $\boldsymbol{\Gamma}$ by noting that $\mathbf{M}^{\frac{1}{2}}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T\mathbf{M}^{\frac{1}{2}}$ is the projection operator onto the subspace spanned by $\mathbf{M}^{\frac{1}{2}}\boldsymbol{\Gamma}$, and that (7.6) equals

$$\mathbf{Z}^T(\mathbf{I} - \mathbf{M}\mathbf{M}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{-\frac{1}{2}})(\mathbf{y}-\mathbf{m}). \quad (7.7)$$

Since $\mathbf{Z}$ and $\boldsymbol{\Gamma}$ are by definition orthogonal it follows that

$$\mathbf{Z}^T\mathbf{M}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\boldsymbol{\Gamma} = \mathbf{Z}^T\boldsymbol{\Gamma} = 0,$$

and furthermore since $\mathbf{Z} \oplus \boldsymbol{\Gamma} = \mathbb{R}^n$

$$\mathbf{M}^{\frac{1}{2}}\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T\mathbf{M}^{\frac{1}{2}} = \mathbf{I} - \mathbf{M}^{-\frac{1}{2}}\mathbf{Z}(\mathbf{Z}^T\mathbf{M}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{M}^{-\frac{1}{2}}. \quad (7.8)$$

Substituting (7.8) back into (7.6) and using the definition of $\boldsymbol{\Xi}$ we obtain

$$\begin{aligned}\mathbf{U}_1 &= \mathbf{Z}^T(\mathbf{I} - \mathbf{M}\mathbf{M}^{-\frac{1}{2}}(\mathbf{I} - \mathbf{M}^{-\frac{1}{2}}\mathbf{Z}(\mathbf{Z}^T\mathbf{M}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{M}^{-\frac{1}{2}})\mathbf{M}^{-\frac{1}{2}})(\mathbf{y}-\mathbf{m}) \\ &= \mathbf{Z}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{M}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{M}^{-1}(\mathbf{y}-\mathbf{m}) \\ &= \mathbf{Z}^T\boldsymbol{\Xi}(\mathbf{y}-\mathbf{m}). \quad (7.9)\end{aligned}$$

■

The matrix $\boldsymbol{\Xi}$ is simply the projection operator for a projection onto the subspace spanned by the columns of $\mathbf{Z}$, but with a metric defined by $\mathbf{M}$. Recalling that U_0 is equal to $\mathbf{Z}^T(\mathbf{y}-\mathbf{m})$ it is apparent that U_1 is a "weighted" version of U_0 in which the weights are the diagonal elements of the inverse of $\mathbf{M}$, in other words U_1 has achieved the standard inverse variance weighting!

In the following derivations it is often necessary to explicitly consider the elements of certain matrices. These matrices will be either of dimensions $n \times n$, $n^2 \times n$, $n \times n^2$ or $n^2 \times n^2$. If $\mathbf{A}$ is an $n \times n$ matrix then its (i, r) element will be denoted by $\mathbf{A}_i^r$. If the matrix is $n^2 \times n$ then its rows can be indexed by i and j where $1 \leq i \leq n$ and $1 \leq j \leq n$, and its columns can be indexed by r. Similarly an $n^2 \times n^2$ matrix can be indexed by the pairs $((i, j), (r, s))$.

In order to find $\mathbf{U}_2$ it is convenient to work with the matrix $\mathbf{V}_2^*$, the orthogonalized version of $\mathbf{V}_2$. Since it is necessary to regress out all of the elements of $\mathbf{V}_2^*$ from $\mathbf{U}_1$ the vector $\mathrm{vec}(\mathbf{V}_2^*)$ is used, where (vec) standards for the operator that takes the elements of a matrix and places them into a vector.

By using the result that for any three matrices A_1, A_2 and A_3

$$\mathrm{vec}(A_1 A_2 A_3) = \left(A_1 \otimes A_3^T\right)\ \mathrm{vec}(A_2)$$

where $\otimes$ denotes the Kronecker product, it follows that

$$\mathrm{vec}(\mathbf{V}_2) = (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathrm{vec}((\mathbf{y} - \mathbf{m})\,(\mathbf{y} - \mathbf{m})^T - \mathbf{M}).$$

By definition, $\mathrm{vec}(\mathbf{V}_2^*) = \mathrm{vec}(\mathbf{V}_2) - B\,\mathbf{V}_1$ where B is defined so that

$$E[(\mathrm{vec}(\mathbf{V}_2) - B\,\mathbf{V}_1)\,\mathbf{V}_1^T] = 0,$$

and therefore

$$E[\mathrm{vec}(\mathbf{V}_2)\,\mathbf{V}_1^T] = B\,E[\mathbf{V}_1\,\mathbf{V}_1^T].$$

The term $E[\mathbf{V}_1\,\mathbf{V}_1^T]$ is given by (7.4), so the only new calculation needed to find B is to obtain the matrix $E[\mathrm{vec}(\mathbf{V}_2)\,\mathbf{V}_1^T]$. Explicitly writing

$$E[\mathrm{vec}(\mathbf{V}_2)\,\mathbf{V}_1^T] = E[(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathrm{vec}((\mathbf{y} - \mathbf{m})\,(\mathbf{y} - \mathbf{m})^T - \mathbf{M})\,(\mathbf{y} - \mathbf{m})^T\,\mathbf{\Gamma}],$$

then

$$\begin{aligned} E[\mathrm{vec}(\mathbf{V}_2)\,\mathbf{V}_1^T] &= (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)E[\mathrm{vec}((\mathbf{y}-\mathbf{m})\,(\mathbf{y}-\mathbf{m})^T - \mathbf{M})\,(\mathbf{y}-\mathbf{m})^T]\,\mathbf{\Gamma} \\ &= (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\,\mathbf{M}_s\,\mathbf{\Gamma}, \end{aligned} \tag{7.10}$$

where

$$\begin{aligned} \mathbf{M}_s &= E[\mathrm{vec}((\mathbf{y}-\mathbf{m})\,(\mathbf{y}-\mathbf{m})^T - \mathbf{M})\,(\mathbf{y}-\mathbf{m})^T] \\ &= E[\mathrm{vec}((\mathbf{y}-\mathbf{m})\,(\mathbf{y}-\mathbf{m})^T)\,(\mathbf{y}-\mathbf{m})^T]. \end{aligned}$$

The reason for the s as the subscript is to stand for “skewness”, since M_s is composed of third central moments. $\mathbf{M}_s$ is an $n^2 \times n$ dimensional matrix. Using the pair (i, j) to index the rows and r to index the columns it is clear that the only non-zero elements of $\mathbf{M}_s$ are those for which $i = j = r$. The matrix $\mathbf{M}_s$ is very sparse, for the case where $n = 2$ it looks like

$$\begin{pmatrix} \kappa^{1,1,1} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \kappa^{2,2,2} \end{pmatrix}$$

where $\kappa^{i,j,k}$ is the third cumulant of $\mathbf{Y}$. All the mixed cumulants are zero because all the Y_i are mutually independent. $\mathbf{M}_s$ can be also be written as $\Delta\widetilde{\mathbf{M}}$, where Δ is the $n^2 \times n$ matrix with (ij, r) element Δ^r_{ij} equal to zero unless $i = j = r$, in which case $\Delta^r_{ij} = 1$, and the matrix $\widetilde{\mathbf{M}}$ is an $n \times n$ diagonal matrix with the third univariate cumulants down the diagonal.

Putting together (7.10) and (7.4) we have that

$$B = (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\,\mathbf{M}_s\,\mathbf{\Gamma}\,(\mathbf{\Gamma}^T\,\mathbf{M}\,\mathbf{\Gamma})^{-1},$$

and so

$$\mathrm{vec}(\mathbf{V}_2^*) = \mathrm{vec}(\mathbf{V_2}) - (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\, \mathbf{M}_s\, \mathbf{\Gamma}\, (\mathbf{\Gamma}^T\, \mathbf{M}\, \mathbf{\Gamma})^{-1}\, \boldsymbol{\Gamma}^T\, (\mathbf{y} - \mathbf{m}) =$$
$$(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\, \left\{\mathrm{vec}((\mathbf{y} - \mathbf{m})\, (\mathbf{y} - \mathbf{m})^T - \mathbf{M}) - \mathbf{M}_s\, \mathbf{\Gamma}\, (\mathbf{\Gamma}^T\, \mathbf{M}\, \mathbf{\Gamma})^{-1} \mathbf{\Gamma}^T (\mathbf{y} - \mathbf{m})\right\}.$$

Using the relationship given by (7.8) this can be re-expressed as

$$\mathrm{vec}(\mathbf{V}_2^*) = (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\, \left\{\mathrm{vec}((\mathbf{y} - \mathbf{m})\, (\mathbf{y} - \mathbf{m})^T - \mathbf{M}) - \mathbf{M}_s\, \mathbf{M}^{-1}\, (\mathbf{I} - \mathbf{\Xi})\, (\mathbf{y} - \mathbf{m})\right\}. \tag{7.11}$$

Proceeding with the calculations for $\mathbf{U}_2$,writing $\mathbf{U}_2$ as $\mathbf{U}_1 - C\,\mathrm{vec}(\mathbf{V}_2^*)$, where C is defined so that

$$E[(\mathbf{U}_1 - C\,\mathrm{vec}(\mathbf{V}_2^*))\,\mathrm{vec}(\mathbf{V}_2)^T] = 0, \tag{7.12}$$

it is clear that both $E[\mathbf{U}_1\,\mathrm{vec}(\mathbf{V}_2)^T]$ and $E[\mathrm{vec}(\mathbf{V}_2^*)\,\mathrm{vec}(\mathbf{V}_2)^T]$ must be found. Now

$$\begin{aligned} E[\mathbf{U}_1\,\mathrm{vec}(\mathbf{V}_2)^T] &= E[\mathbf{Z}^T\, \mathbf{\Xi}(\mathbf{y} - \mathbf{m})\, (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathrm{vec}((\mathbf{y} - \mathbf{m})\, (\mathbf{y} - \mathbf{m})^T - \mathbf{M})^T] \\ &= \mathbf{Z}^T\, \mathbf{\Xi} E[(\mathbf{y} - \mathbf{m})\,\mathrm{vec}((\mathbf{y} - \mathbf{m})\, (\mathbf{y} - \mathbf{m})^T - \mathbf{M})^T](\mathbf{\Gamma} \otimes \mathbf{\Gamma}) \\ &= \mathbf{Z}^T\, \mathbf{\Xi}\, \mathbf{M}_s^T\, (\mathbf{\Gamma} \otimes \mathbf{\Gamma}). \end{aligned} \tag{7.13}$$

Furthermore

$$E[\mathrm{vec}(\mathbf{V}_2^*)\,\mathrm{vec}(\mathbf{V}_2)^T] =$$
$$E\left[(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\, \left\{\mathrm{vec}((\mathbf{y} - \mathbf{m})\, (\mathbf{y} - \mathbf{m})^T - \mathbf{M}) - \mathbf{M}_s\, \mathbf{M}^{-1}\, (\mathbf{I} - \mathbf{\Xi})\, (\mathbf{y} - \mathbf{m})\right\}\right.$$
$$\left.\mathrm{vec}((\mathbf{y} - \mathbf{m})\, (\mathbf{y} - \mathbf{m})^T - \mathbf{M})^T\, (\mathbf{\Gamma} \otimes \mathbf{\Gamma})\right],$$

which can be rewritten as

$$(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\, \left\{\mathbf{M}_k - \mathbf{M}_s\, \mathbf{M}^{-1}\, (\mathbf{I} - \mathbf{\Xi})\, \mathbf{M}_s^T\right\}\, (\mathbf{\Gamma} \otimes \mathbf{\Gamma}), \tag{7.14}$$

where

$$\mathbf{M}_k = E[\mathrm{vec}((\mathbf{y} - \mathbf{m})\, (\mathbf{y} - \mathbf{m})^T - \mathbf{M})\, (\mathrm{vec}((\mathbf{y} - \mathbf{m})\, (\mathbf{y} - \mathbf{m}^T) - \mathbf{M}))^T].$$

The matrix $\mathbf{M}_k$ contains fourth central moments and products of second central moments of the Y_i. Indexing the rows of $\mathbf{M}_k$ by (i,j) and the columns by (r,s) we have the following characterization in terms of cumulants, for its elements:

$$\{\mathbf{M}_k\}_{ij}^{rs} = \begin{cases} 2\kappa^{i,r}\kappa^{j,s} + \kappa^{i,j,r,s} & \text{if } i=j=r=s \\ \kappa^{i,r}\kappa^{j,s} & \text{if } i=r\neq j=s \\ \kappa^{i,s}\kappa^{j,r} & \text{if } i=s\neq j=r \\ 0 & \text{otherwise} \end{cases}$$

where $\kappa^{i,j,r,s}$ is the fourth cumulant of $\mathbf{Y}$. It follows that $(\mathbf{M}_k)_{ij}^{rs}$ can be written as the sum of two matrices

$$\{\mathbf{M}_k\}_{ij}^{rs} = \left(\mathbf{M}_i^r\mathbf{M}_j^s + \mathbf{M}_i^s\mathbf{M}_j^r\right) + \Delta_{ij}^{\alpha}\overline{\mathbf{M}}_{\alpha}^{\beta}\Delta_{\beta}^{rs} \tag{7.15}$$

where $\overline{\mathbf{M}}$ is an $n\times n$ diagonal matrix with diagonal elements equal to the fourth univariate cumulants. There is a strong similarity of form between (7.15) and (3.10), even though the later was derived under an *i.i.d.* sampling scheme, whereas the present context has only the independence assumption.

To illustrate the structure of M_k, consider the case when $\mathbf{Y}$ is made up of just two observations, then M_k is of the form

$$\begin{pmatrix} \kappa^{1,1,1,1}+2(\kappa^{1,1})^2 & 0 & 0 & 0 \\ 0 & \kappa^{1,1}\kappa^{2,2} & \kappa^{2,2}\kappa^{1,1} & 0 \\ 0 & \kappa^{2,2}\kappa^{1,2} & \kappa^{2,2}\kappa^{1,1} & 0 \\ 0 & 0 & 0 & \kappa^{2,2,2,2}+2(\kappa^{2,2})^2 \end{pmatrix}.$$

Returning to (7.12) the matrix C is

$$C = E[\mathbf{U}_1\, v(\mathbf{V}_2)^T]\ \left(E[v(\mathbf{V}_2^*)\, v(\mathbf{V}_2)^T]\right)^+$$

where $^+$ represents the Moore-Penrose generalized-inverse. Putting together (7.13) and (7.14) gives

$$C = \mathbf{Z}^T\,\mathbf{\Xi}\,\mathbf{M}_s^T\,(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\left((\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)\left\{\mathbf{M}_k - \mathbf{M}_s\,\mathbf{M}^{-1}\,(\mathbf{I}-\mathbf{\Xi})\;\mathbf{M}_s^T\right\}\;(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\right)^+. \tag{7.16}$$

Writing $\mathbf{\Sigma}$ for $\mathbf{M}_k - \mathbf{M}_s\,\mathbf{M}^{-1}\,(\mathbf{I}-\mathbf{\Xi})\;\mathbf{M}_s^T$, we finally have

Theorem 7.2 *The* $\mathbf{U}_2$ *projective score is*

$$\begin{aligned}\mathbf{U}_2 = \mathbf{Z}^T\,\mathbf{\Xi}(\mathbf{y}-\mathbf{m}) - \mathbf{Z}^T\,\mathbf{\Xi}\,\mathbf{M}_s^T\,(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\left((\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)\mathbf{\Sigma}\,(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\right)^+\;(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)\\ \left\{vec((\mathbf{y}-\mathbf{m})\,(\mathbf{y}-\mathbf{m})^T - \mathbf{M}) - \mathbf{M}_s\,\mathbf{M}^{-1}\,(\mathbf{I}-\mathbf{\Xi})\;(\mathbf{y}-\mathbf{m})\right\}.\end{aligned} \tag{7.17}$$

Proof

This result is obtained by substituting in the expressions obtained for $\mathrm{vec}(\mathbf{V}_2^*)$ and C obtained in (7.11) and (7.16) respectively, into the definition of $\mathbf{U}_2 = \mathbf{U}_1 - C\mathrm{vec}(\mathbf{V}_2^*)$. ∎

7.3 Substitution version of the $\mathbf{U}_2$ projective score

Consider what happens to (7.17) when the *m.l.e.* of ω for β fixed is substituted. The *m.l.e.*, $\hat{\mathbf{m}}_\beta = b'(\mathbf{Z}\beta + \boldsymbol{\Gamma}\hat{\omega}_\beta)$ is a solution to the score equation:

$$\mathbf{\Gamma}^T\,(\mathbf{y}-\mathbf{m}) = 0,$$

so that

$$\hat{\mathbf{m}}_\beta = (\mathbf{y} - \mathbf{Z}c) \tag{7.18}$$

for some vector c. This follows because the only vectors that are orthogonal to $\boldsymbol{\Gamma}$ are those in the subspace spanned by the columns of $\mathbf{Z}$ since $\boldsymbol{\Gamma}\oplus\mathbf{Z} = \mathbb{R}^n$. The cumulants in the matrices $\mathbf{M}_k$ and $\mathbf{M}_s$ will be estimated by the corresponding functions of the estimates of the means. Substituting $\hat{\mathbf{m}}_\beta$ back into (7.17) the $\mathbf{U}_2$ function becomes

$$\mathbf{U}_2(\hat{\mathbf{m}}_\beta) =$$
$$\mathbf{Z}^T\,\mathbf{Z}\,c - \mathbf{Z}^T\,\hat{\boldsymbol{\Xi}}\,\hat{\mathbf{M}}_s^T\,(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\left((\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Gamma}^T)\hat{\boldsymbol{\Sigma}}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\right)^+(\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Gamma}^T)\left\{\mathrm{vec}(\mathbf{Z}\,c\,c^T\mathbf{Z}^T - \hat{\mathbf{M}})\right\}.$$

Now

$$\mathrm{vec}(\mathbf{Z}\,c\,c^T\,\mathbf{Z}^T - \hat{\mathbf{M}}) = \mathrm{vec}(\mathbf{Z}\,c\,c^T\,\mathbf{Z}^T) - \mathrm{vec}(\hat{\mathbf{M}})$$

and

$$(\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Gamma}^T)\,\mathrm{vec}(\mathbf{Z}\,c\,c^T\,\mathbf{Z}^T) = \mathrm{vec}(\boldsymbol{\Gamma}^T\,\mathbf{Z}\,c\,c^T\,\mathbf{Z}^T\boldsymbol{\Gamma}) = 0,$$

from the orthogonality of $\mathbf{Z}$ and $\boldsymbol{\Gamma}$. Furthermore

$$\mathrm{vec}(\hat{\mathbf{M}}) = \Delta\hat{\mathbf{M}}\,\mathbf{1},$$

where $\mathbf{1}$ is an n-dimensional vector of ones. Hence, $\mathbf{U}_2(\hat{\mathbf{m}}_\beta)$ is expressible as

$$\mathbf{U}_2(\hat{\mathbf{m}}_\beta) = \mathbf{Z}^T\,\mathbf{Z}\,c + \mathbf{Z}^T\,\hat{\boldsymbol{\Xi}}\,\hat{\widetilde{\mathbf{M}}}\Delta^T\,(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\left((\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Gamma}^T)\hat{\boldsymbol{\Sigma}}\,(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\right)^+\,(\boldsymbol{\Gamma}^T\otimes\boldsymbol{\Gamma}^T)\,\Delta\hat{\mathbf{M}}\mathbf{1}, \tag{7.19}$$

where $\hat{\widetilde{\mathbf{M}}}\Delta^T$ has been written for $\hat{\mathbf{M}}_s^T$ in order to reveal a symmetric structure.

The aim is now to solve the $\mathbf{U}_2(\hat{\mathbf{m}}_\beta)$ for c, and then use (7.18) to obtain the estimate of $\hat{\mathbf{m}}_\beta$. In order to obtain the estimator of β note that

$$\hat{\mathbf{m}}_\beta = b'(\mathbf{Z}\beta + \boldsymbol{\Gamma}\hat{\omega}_\beta)$$

so that

$$b'^{-1}(\hat{\mathbf{m}}_\beta) = \mathbf{Z}\beta + \boldsymbol{\Gamma}\hat{\omega}_\beta$$

and

$$(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T b'^{-1}(\hat{\mathbf{m}}_\beta) = \beta. \tag{7.20}$$

The critical difficulty in solving (7.19) is that it requires that the generalized inverse of the matrix $\left(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\hat{\mathbf{\Sigma}}\,(\mathbf{\Gamma} \otimes \mathbf{\Gamma})\right)$ be found. This would in general be a formidable task, since the matrix is of dimensions $s^2 \times s^2$, and s will usually be the same order as the sample size n. However, in this model, it is possible to obtain some remarkable simplifications, the principle reason being that the matrix with elements $\left(\mathbf{M}_i^r \mathbf{M}_j^s + \mathbf{M}_i^s \mathbf{M}_j^r\right)$, has a very simple generalized inverse, given by (3.12).

The simplifications will be obtained for the matrix

$$\Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma}) \left((\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{\Sigma}\,(\mathbf{\Gamma} \otimes \mathbf{\Gamma})\right)^+ (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\,\Delta.$$

The matrix $\boldsymbol{\varSigma} = \mathbf{M}_k - \mathbf{M}_s\,\mathbf{M}^{-1}\,(\mathbf{I} - \mathbf{\Xi})\;\mathbf{M}_s^T$ can be decomposed as follows:

$$\begin{aligned}
\boldsymbol{\varSigma}_{ij}^{rs} &= \left(\mathbf{M}_i^r \mathbf{M}_j^s + \mathbf{M}_i^s \mathbf{M}_j^r\right) + \left(\Delta \overline{\mathbf{M}} \Delta^T\right)_{ij}^{rs} - \left(\Delta \widetilde{\mathbf{M}} \mathbf{M}^{-1} (\mathbf{I} - \mathbf{\Xi}) \widetilde{\mathbf{M}} \Delta^T\right)_{ij}^{rs} \\
&= \left(\mathbf{M}_i^r \mathbf{M}_j^s + \mathbf{M}_i^s \mathbf{M}_j^r\right) + \left(\Delta \left(\overline{\mathbf{M}} - \widetilde{\mathbf{M}} \mathbf{M}^{-1} \widetilde{\mathbf{M}}\right) \Delta^T\right)_{ij}^{rs} + \left(\Delta \widetilde{\mathbf{M}} \mathbf{M}^{-1} \mathbf{\Xi} \widetilde{\mathbf{M}} \Delta^T\right)_{ij}^{rs},
\end{aligned} \tag{7.21}$$

which can be rewritten as

$$\boldsymbol{\varSigma} = \mathbf{A} + \Delta \mathbf{B} \Delta^T$$

where

$$\mathbf{A}_{ij}^{rs} = \left(\mathbf{M}_i^r \mathbf{M}_j^s + \mathbf{M}_i^s \mathbf{M}_j^r\right) + \left(\Delta \left(\overline{\mathbf{M}} - \widetilde{\mathbf{M}} \mathbf{M}^{-1} \widetilde{\mathbf{M}}\right) \Delta^T\right)_{ij}^{rs} \quad \text{and} \quad \mathbf{B} = \widetilde{\mathbf{M}} \mathbf{M}^{-1} \mathbf{\Xi} \widetilde{\mathbf{M}}.$$

By substituting in $\mathbf{\Xi}$ explicitly, $\mathbf{B}$ can be re-expressed as

$$\mathbf{B} = \widetilde{\mathbf{M}} \mathbf{M}^{-1} \mathbf{Z}\,(\mathbf{Z}^T\,\mathbf{M}^{-1}\,\mathbf{Z})^{-1}\,\mathbf{Z}^T\,\mathbf{M}^{-1} \widetilde{\mathbf{M}} = \mathbf{C}\mathbf{C}^T$$

where

$$\mathbf{C} = \widetilde{\mathbf{M}} \mathbf{M}^{-1} \mathbf{Z}\,(\mathbf{Z}^T\,\mathbf{M}^{-1}\,\mathbf{Z})^{-\frac{1}{2}}.$$

Lemma 7.1 *Both the matrices* $\mathbf{A}$ *and* $\Delta\mathbf{B}\Delta^T$ *are non-negative definite.*

Proof

The fact that $\Delta\mathbf{B}\Delta^T$ is non-negative definite is immediate from its representation as a matrix multiplied by its transpose. To see that $\mathbf{A}$ is also non-negative definite, consider first the elements of $\mathbf{A}$ for which $i = j = r = s$, then $\mathbf{A}_{ij}^{rs} = \mathbf{A}_{ii}^{ii} = 2\kappa_2^2 + \kappa_4 - \kappa_3\kappa_2^{-1}\kappa_3$ where κ_t is the t^{th} cumulant of Y_i. (Note the change back to *power* notation for the cumulants, which allows a more compact notation because of the fact that all joint cumulants are zero since the random variables Y_i are independent.) It is a simple exercise to show that this quantity must be greater than or equal to 0 (Rao, 1973, see Exercise 2.2, p.143). The remaining elements in $\mathbf{A}$ can be written in terms of a matrix DD^T so that they comprise a non-negative definite matrix. Hence $\mathbf{A}$ is the sum of two non-negative definite matrices and is therefore itself non-negative definite. ■

Since $\mathbf{A}$ is a non-negative definite matrix, it has a unique symmetric square root $\mathbf{A}^{\frac{1}{2}}$ (Arnold, 1981, p.454, Theorem A.6). Therefore, the matrix

$$\begin{aligned}(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{\Sigma}(\mathbf{\Gamma} \otimes \mathbf{\Gamma}) &= (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\left(\mathbf{A}^{\frac{1}{2}}\mathbf{A}^{T\frac{1}{2}} + \Delta\mathbf{C}\mathbf{C}^T\Delta^T\right)(\mathbf{\Gamma} \otimes \mathbf{\Gamma}) \\ &= \mathbf{P}\mathbf{P}^T + \mathbf{Q}\mathbf{Q}^T, \qquad (7.22)\end{aligned}$$

where $\mathbf{P} = (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{A}^{\frac{1}{2}}$ and $\mathbf{Q} = (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\Delta\mathbf{C}$. Equation (7.22) is a particularly useful representation, since now a series of well known generalized inverse results can be applied to help simplify the problem. The next lemma establishes that $\mathbf{Q}$ lies in the column space of $\mathbf{P}$, which allows further simplifications of $(\mathbf{P}\mathbf{P}^T + \mathbf{Q}\mathbf{Q}^T)^+$ to be made.

Lemma 7.2 *Assuming that* $2\mathbf{M}_i^r\mathbf{M}_j^s + \left\{\Delta\left(\overline{\mathbf{M}} - \widetilde{\mathbf{M}}\mathbf{M}^{-1}\ \widetilde{\mathbf{M}}\right)\Delta^T\right\}_{ij}^{rs} \neq 0$ *for* $i = j = r = s$*, then* $\mathbf{Q}$ *lies in the column space of* $\mathbf{P}$*.*

Proof

Since $\mathbf{A}$ is a non-negative definite symmetric matrix it has a decomposition $\mathbf{A} = R\Lambda R^T$ (Rao, 1973, p.39, 1c.3) where R is an orthogonal matrix and Λ is a diagonal matrix of non-negative numbers, so that $\mathbf{A} = R\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}R^T = R\Lambda^{\frac{1}{2}}R^T R\Lambda^{\frac{1}{2}}R^T$ since $R^T R = I$. Hence $\mathbf{P} = (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)R\Lambda^{\frac{1}{2}}R^T$. R is the matrix of eigenvectors of $\mathbf{A}$ and Λ is the diagonal matrix of eigenvalues.

For the particular case where $\mathbf{M}$ is a diagonal matrix, it can be shown that R and Λ are of the following form:

$$\{R\}_{ij}^{rs} = \begin{cases} 1 & \text{if } i = j = r = s \\ \frac{1}{\sqrt{2}} & \text{if } i = r \neq j = s \text{ and } i < j \\ -\frac{1}{\sqrt{2}} & \text{if } i = r \neq j = s \text{ and } i > j \\ \frac{1}{\sqrt{2}} & \text{if } i = s \neq j = r \\ 0 & \text{otherwise} \end{cases}$$

and the diagonal elements of Λ

$$\{\Lambda\}_{ij}^{rs} = \begin{cases} 2\mathbf{M}_i^r\mathbf{M}_j^s + \left\{\Delta\left(\overline{\mathbf{M}} - \widetilde{\mathbf{M}}\mathbf{M}^{-1}\widetilde{\mathbf{M}}\right)\Delta^T\right\}_{ij}^{rs} & \text{if } i = r, j = s \text{ and } i \leq j \\ 0 & \text{otherwise} \end{cases}$$

The above decomposition was established using `Mathematica`. Now consider the matrix

$$\begin{aligned} W &= \mathbf{P}R\Lambda^{-\frac{1}{2}}R^T\Delta\mathbf{C} \\ &= (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{A}^{\frac{1}{2}}R\Lambda^{-\frac{1}{2}}R^T\Delta\mathbf{C} \\ &= (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)R\Lambda^{\frac{1}{2}}R^T R\Lambda^{-\frac{1}{2}}R^T\Delta\mathbf{C} \\ &= (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)R\Lambda^{\frac{1}{2}}\Lambda^{-\frac{1}{2}}R^T\Delta\mathbf{C} \\ &= (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)R\tilde{I}R^T\Delta\mathbf{C}, \end{aligned}$$

where $\Lambda^{-\frac{1}{2}}$ is defined as the diagonal matrix, with elements equal to the reciprocal of the square root of the elements of Λ, but when an element of Λ is equal to 0 the corresponding element in $\Lambda^{-\frac{1}{2}}$ is set equal to 0. The matrix $\widetilde{I}$ is an partial identity matrix, in which the elements that correspond to the 0 diagonal elements of Λ are also 0 in $\widetilde{I}$.

By construction W lies in the column space of $\mathbf{P}$. Now the matrix $R\widetilde{I}R^T$ has as its (ij, rs) element $\frac{1}{2}\left(I_i^r I_j^s + I_i^s I_j^r\right)$ (this is provided the assumption that $2\mathbf{M}_i^r\mathbf{M}_j^s + \left\{\Delta\left(\overline{\mathbf{M}} - \widetilde{\mathbf{M}}\mathbf{M}^{-1}\widetilde{\mathbf{M}}\right)\Delta^T\right\}_{ij}^{rs} \neq 0$ holds), where I here denotes the $n \times n$ identity matrix, so that $R\widetilde{I}R^T\Delta = \Delta$, and hence

$$W = (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\Delta\mathbf{C} = \mathbf{Q},$$

so that $\mathbf{Q}$ lies in the column space of $\mathbf{P}$ as required. ■

Using Lemma 7.1 and a standard result for the Moore-Penrose inverse of a sum of non-negative definite matrices (Boullion & Odell, 1971, Theorem 6, p.16), it follows that

$$\begin{aligned}\left(\mathbf{P}\mathbf{P}^T + \mathbf{Q}\mathbf{Q}^T\right)^+ &= \mathbf{P}^{+T}\left(I - \mathbf{P}^+\mathbf{Q}\left(I + \mathbf{Q}^T\mathbf{P}^{+T}\mathbf{P}^+\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{P}^{+T}\right)\mathbf{P}^+ \\ &= \left(\mathbf{P}\mathbf{P}^T\right)^+ - \left(\mathbf{P}\mathbf{P}^T\right)^+\mathbf{Q}\left(I + \mathbf{Q}^T\mathbf{P}^{+T}\mathbf{P}^+\mathbf{Q}\right)^{-1}\mathbf{Q}^T\left(\mathbf{P}\mathbf{P}^T\right)^+ \\ &= \mathbf{S}^+ - \mathbf{S}^+\mathbf{Q}\left(I + \mathbf{Q}^T\mathbf{S}^+\mathbf{Q}\right)^{-1}\mathbf{Q}^T\mathbf{S}^+ \qquad (7.23)\end{aligned}$$

where the fact that $\mathbf{P}^{+T}\mathbf{P}^+ = \left(\mathbf{P}\mathbf{P}^T\right)^+$ has been used , and $\mathbf{P}\mathbf{P}^T$ is written as $\mathbf{S}$. This result is used in the proof of

Lemma 7.3

$$\Delta^T(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\left((\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)\mathbf{\Sigma}(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\right)^+(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)\Delta = \mathbf{R} - \mathbf{R}\mathbf{C}\left(I + \mathbf{C}^T\mathbf{R}\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{R} \qquad (7.24)$$

where $\mathbf{R} = \Delta^T(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\mathbf{S}^+(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)\Delta$.

The result is obtained by substituting (7.23) into the LHS of (7.24), which gives

$$\Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma}) \left(\mathbf{S}^+ - \mathbf{S}^+\mathbf{Q} \left(I + \mathbf{Q}^T\mathbf{S}^+\mathbf{Q}\right)^{-1} \mathbf{Q}^T\mathbf{S}^+\right) (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \Delta =$$
$$\Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma})\mathbf{S}^+(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \Delta - \Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma})\mathbf{S}^+\mathbf{Q} \left(I + \mathbf{Q}^T\mathbf{S}^+\mathbf{Q}\right)^{-1} \mathbf{Q}^T\mathbf{S}^+\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \Delta,$$

and then writing $\mathbf{Q} = (\mathbf{\Gamma} \otimes \mathbf{\Gamma})\Delta\mathbf{C}$ one obtains

$$\begin{aligned} &\Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma})\mathbf{S}^+(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \Delta - \\ &\qquad \Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma})\mathbf{S}^+(\mathbf{\Gamma} \otimes \mathbf{\Gamma})\Delta\mathbf{C} \left(I + \mathbf{C}^T\Delta(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{S}^+(\mathbf{\Gamma} \otimes \mathbf{\Gamma})\Delta\mathbf{C}\right)^{-1} \\ &\qquad\qquad \mathbf{C}^T\Delta(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{S}^+(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \Delta \end{aligned} \tag{7.25}$$

Denoting $\Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma})\mathbf{S}^+(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \Delta$ as $\mathbf{R}$, (7.25) simplifies to

$$\mathbf{R} - \mathbf{R}\mathbf{C} \left(I + \mathbf{C}^T\mathbf{R}\mathbf{C}\right)^{-1} \mathbf{C}^T\mathbf{R},$$

which is the required result. ■

This first simplification has not achieved the required reduction in dimensionality of the problem, because the matrix $\mathbf{R}$ still contains the generalized inverse of the matrix $\mathbf{S}$, which is of dimensions $s^2 \times s^2$.

Writing out $\mathbf{R}$ explicitly,

$$\begin{aligned} \mathbf{R} &= \Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma})S^-(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \Delta \\ &= \Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma}) \left(\mathbf{P}\mathbf{P}^T\right)^- (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \Delta \\ &= \Delta^T (\mathbf{\Gamma} \otimes \mathbf{\Gamma}) \left((\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{A}(\mathbf{\Gamma} \otimes \mathbf{\Gamma})\right)^+ (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T) \Delta, \end{aligned}$$

where, as before $\mathbf{A}_{ij}^{rs} = \left(\mathbf{M}_i^r\mathbf{M}_j^s + \mathbf{M}_i^s\mathbf{M}_j^r\right) + \left(\Delta\left(\overline{\mathbf{M}} - \widetilde{\mathbf{M}}\mathbf{M}^{-1}\widetilde{\mathbf{M}}\right)\Delta^T\right)_{ij}^{rs}$. It is apparent that the problem of simplification of $\mathbf{A}$ is similar to that tackled in Lemma 7.1, but now there is a slight technicality in that the matrix $\Delta\left(\overline{\mathbf{M}} - \widetilde{\mathbf{M}}\mathbf{M}^{-1}\widetilde{\mathbf{M}}\right)\Delta^T$ is not obviously non-negative definite. However, it will be shown that for the GLM models under consideration all the diagonal elements of $\overline{\mathbf{M}} - \widetilde{\mathbf{M}}\mathbf{M}^{-1}\widetilde{\mathbf{M}}$ will have the *same* sign. If the sign is positive, then $\mathbf{A}$ will remain unchanged, but if the sign is negative then $\mathbf{A}$ will be written as

$$\mathbf{A}_{ij}^{rs} = \left(\mathbf{M}_i^r\mathbf{M}_j^s + \mathbf{M}_i^s\mathbf{M}_j^r\right) - \left(\Delta\left(\widetilde{\mathbf{M}}\mathbf{M}^{-1}\widetilde{\mathbf{M}} - \overline{\mathbf{M}}\right)\Delta^T\right)_{ij}^{rs}.$$

Let $\mathbf{D}$ be the square root of non-negative definite diagonal matrix with elements equal to $|(\overline{\mathbf{M}} - \widetilde{\mathbf{M}}\mathbf{M}^{-1}\widetilde{\mathbf{M}})_i^i|$. Consequently

$$(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{A}(\mathbf{\Gamma} \otimes \mathbf{\Gamma}) = (\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\left(\mathbf{EE} \pm \Delta\mathbf{DD}\Delta^T\right)(\mathbf{\Gamma} \otimes \mathbf{\Gamma}),$$

where $(\mathbf{EE})_{ij}^{rs} = \left(\mathbf{M}_i^r\mathbf{M}_j^s + \mathbf{M}_i^s\mathbf{M}_j^r\right)$, and the $\pm$ indicates the sign of the diagonal elements of $\overline{\mathbf{M}} - \widetilde{\mathbf{M}}\mathbf{M}^{-1}\widetilde{\mathbf{M}}$. One can show in a similar fashion to Lemma 7.2 that $(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\Delta\mathbf{D}$ lies in the column space of $(\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{E}$, and by an argument similar to Lemma 7.3 that

Lemma 7.4

$$\mathbf{R} = \mathbf{H} \mp \mathbf{HD}(I \pm \mathbf{D}^T\mathbf{HD})^{-1}\mathbf{D}^T\mathbf{H},$$

where $\mathbf{H}$ *is the matrix* $\Delta^T(\mathbf{\Gamma} \otimes \mathbf{\Gamma})\left((\mathbf{\Gamma}^T \otimes \mathbf{\Gamma}^T)\mathbf{EE}(\mathbf{\Gamma} \otimes \mathbf{\Gamma})\right)^+(\mathbf{\Gamma} \otimes \mathbf{\Gamma})\Delta$.

The matrix $\mathbf{H}$ is still an unwieldy object, being of dimensions $s^2 \times s^2$, but it is of a form that allows the fundamental simplification that makes the entire problem computationally tractable.

Theorem 7.3 *Let the matrix* $\mathbf{H}$ *be defined as before, then*

$$\mathbf{H} = \tfrac{1}{2}\left[\mathbf{\Gamma}\left(\mathbf{\Gamma}^T\mathbf{M}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}^T\right] * \left[\mathbf{\Gamma}\left(\mathbf{\Gamma}^T\mathbf{M}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}^T\right], \tag{7.26}$$

where $*$ *denotes the Hadamard product.*

The value of the theorem, lies in the fact that the required generalized inverses are of matrices with dimension $s \times s$, rather than $s^2 \times s^2$. Furthermore, the orthogonality of $\boldsymbol{\Gamma}$ and $\mathbf{Z}$ means that the identity

$$\mathbf{\Gamma}\left(\mathbf{\Gamma}^T\mathbf{M}\mathbf{\Gamma}\right)^{-1}\mathbf{\Gamma}^T = \mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{Z}\left(\mathbf{Z}^T\mathbf{M}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{M}^{-1},$$

can be used to evaluate this expression in terms of $\mathbf{Z}$ rather than $\boldsymbol{\Gamma}$. Whether one chooses to use the expression involving $\mathbf{Z}$ or the one involving $\boldsymbol{\Gamma}$ would depend on the relative dimensions of r and s, the number of parameters of interest and the number of nuisance parameters respectively.

<u>Proof</u>

By definition $\mathbf{H}$ is equal to

$$\Delta^T\left(\mathbf{\Gamma}\otimes\mathbf{\Gamma}\right)\left(\left(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T\right)\left(\mathbf{E}\mathbf{E}\right)\left(\mathbf{\Gamma}\otimes\mathbf{\Gamma}\right)\right)^{-}\left(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T\right)\Delta,$$

and note that the element $\left(\left(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T\right)\left(\mathbf{E}\mathbf{E}\right)\left(\mathbf{\Gamma}\otimes\mathbf{\Gamma}\right)\right)^{rs}_{ij}$ can be written as

$$\begin{aligned}\boldsymbol{\Gamma}^{T}{}^{\alpha}_{i}\boldsymbol{\Gamma}^{T}{}^{\beta}_{j}\left(\mathbf{M}^{u}_{\alpha}\mathbf{M}^{t}_{\beta}+\mathbf{M}^{t}_{\alpha}\mathbf{M}^{u}_{\beta}\right)\boldsymbol{\Gamma}^{r}_{u}\boldsymbol{\Gamma}^{s}_{t} &= \boldsymbol{\Gamma}^{T}{}^{\alpha}_{i}\mathbf{M}^{u}_{\alpha}\boldsymbol{\Gamma}^{r}_{u}\boldsymbol{\Gamma}^{T}{}^{\beta}_{j}\mathbf{M}^{t}_{\beta}\boldsymbol{\Gamma}^{s}_{t}+\boldsymbol{\Gamma}^{T}{}^{\alpha}_{i}\mathbf{M}^{t}_{\alpha}\boldsymbol{\Gamma}^{r}_{t}\,\boldsymbol{\Gamma}^{T}{}^{\beta}_{j}\mathbf{M}^{u}_{\beta}\boldsymbol{\Gamma}^{s}_{u}\\ &= \boldsymbol{\Omega}^{r}_{i}\boldsymbol{\Omega}^{s}_{j}+\boldsymbol{\Omega}^{s}_{i}\boldsymbol{\Omega}^{r}_{j}\end{aligned}$$

where $\boldsymbol{\Omega}$ is the matrix $\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma}$. The fact that the Moore-Penrose generalized inverse of $\boldsymbol{\Omega}^{r}_{i}\boldsymbol{\Omega}^{s}_{j}+\boldsymbol{\Omega}^{s}_{i}\boldsymbol{\Omega}^{r}_{j}$ is $\frac{1}{4}\left(\mathbf{F}^{r}_{i}\mathbf{F}^{s}_{j}+\mathbf{F}^{s}_{i}\mathbf{F}^{r}_{j}\right)$ (McCullagh, 1987, p.20 Qu.1.12) provides a fundamental simplification. Here, the matrix $\mathbf{F}$ is the inverse of $\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma}$. Consequently, the (i,j) element of $\Delta^T(\mathbf{\Gamma}\otimes\mathbf{\Gamma})S^{-}(\mathbf{\Gamma}^T\otimes\mathbf{\Gamma}^T)\Delta$ can be written as

$$\tfrac{1}{4}\left(\Delta^{rs}_{i}\boldsymbol{\Gamma}^{\alpha}_{r}\boldsymbol{\Gamma}^{\beta}_{s}\left[\mathbf{F}^{t}_{\alpha}\mathbf{F}^{u}_{\beta}+\mathbf{F}^{u}_{\alpha}\mathbf{F}^{t}_{\beta}\right]\boldsymbol{\Gamma}^{T}{}^{\gamma}_{u}\boldsymbol{\Gamma}^{T}{}^{\delta}_{t}\Delta^{j}_{\gamma\delta}\right),$$

or on adding and subtracting $\mathbf{F}^{t}_{\alpha}\mathbf{F}^{u}_{\beta}$

$$\begin{aligned}&\tfrac{1}{4}\left(\Delta^{rs}_{i}\boldsymbol{\Gamma}^{\alpha}_{r}\boldsymbol{\Gamma}^{\beta}_{s}\left[2\,\mathbf{F}^{t}_{\alpha}\mathbf{F}^{u}_{\beta}+\left(\mathbf{F}^{u}_{\alpha}\mathbf{F}^{t}_{\beta}-\mathbf{F}^{t}_{\alpha}\mathbf{F}^{u}_{\beta}\right)\right]\boldsymbol{\Gamma}^{T}{}^{\gamma}_{u}\boldsymbol{\Gamma}^{T}{}^{\delta}_{t}\Delta^{j}_{\gamma\delta}\right)=\\ &\tfrac{1}{2}\left(\Delta^{rs}_{i}\boldsymbol{\Gamma}^{\alpha}_{r}\boldsymbol{\Gamma}^{\beta}_{s}\left[\mathbf{F}^{t}_{\alpha}\mathbf{F}^{u}_{\beta}\right]\boldsymbol{\Gamma}^{T}{}^{\gamma}_{u}\boldsymbol{\Gamma}^{T}{}^{\delta}_{t}\Delta^{j}_{\gamma\delta}\right)+\tfrac{1}{4}\left(\Delta^{rs}_{i}\boldsymbol{\Gamma}^{\alpha}_{r}\boldsymbol{\Gamma}^{\beta}_{s}\left[\mathbf{F}^{u}_{\alpha}\mathbf{F}^{t}_{\beta}-\mathbf{F}^{t}_{\alpha}\mathbf{F}^{u}_{\beta}\right]\boldsymbol{\Gamma}^{T}{}^{\gamma}_{u}\boldsymbol{\Gamma}^{T}{}^{\delta}_{t}\Delta^{j}_{\gamma\delta}\right)\end{aligned}\tag{27}$$

Using the structure of Δ, $\frac{1}{4}\left(\Delta_i^{rs}\boldsymbol{\Gamma}_r^{\alpha}\boldsymbol{\Gamma}_s^{\beta}\left[\mathbf{F}_\alpha^u\mathbf{F}_\beta^t - \mathbf{F}_\alpha^t\mathbf{F}_\beta^u\right]\boldsymbol{\Gamma}^{T\gamma}_{\ u}\boldsymbol{\Gamma}^{T\delta}_{\ t}\Delta_{\gamma\delta}^j\right)$ can be re-expressed as

$$\frac{1}{4}\left(\boldsymbol{\Gamma}_i^{\alpha}\boldsymbol{\Gamma}_i^{\beta}\left[\mathbf{F}_\alpha^u\mathbf{F}_\beta^t - \mathbf{F}_\alpha^t\mathbf{F}_\beta^u\right]\boldsymbol{\Gamma}^{Tj}_{\ u}\boldsymbol{\Gamma}^{Tj}_{\ t}\right).$$

Now consider the elements of the matrix $\mathbf{J} = \mathbf{F}_\alpha^u\mathbf{F}_\beta^t - \mathbf{F}_\alpha^t\mathbf{F}_\beta^u$, for which we have the following characterization:

$$\{\mathbf{J}\}_{\alpha\beta}^{ut} = \begin{cases} 0 & \text{if } u = t \\ 0 & \text{if } \alpha = \beta \\ \{\mathbf{J}\}_{\alpha\beta}^{ut} = -\{\mathbf{J}\}_{\beta\alpha}^{ut} & \text{if } u \neq t \\ \{\mathbf{J}\}_{\alpha\beta}^{ut} = -\{\mathbf{J}\}_{\alpha\beta}^{tu} & \text{if } \alpha \neq \beta \end{cases} \tag{7.28}$$

By definition,

$$\boldsymbol{\Gamma}_i^{\alpha}\boldsymbol{\Gamma}_i^{\beta}\mathbf{J}_{\alpha\beta}^{ut} = \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\boldsymbol{\Gamma}_i^{\alpha}\boldsymbol{\Gamma}_i^{\beta}\mathbf{J}_{\alpha\beta}^{ut} = \sum_{\alpha<\beta}\boldsymbol{\Gamma}_i^{\alpha}\boldsymbol{\Gamma}_i^{\beta}\mathbf{J}_{\alpha\beta}^{ut} + \sum_{\alpha<\beta}\boldsymbol{\Gamma}_i^{\alpha}\boldsymbol{\Gamma}_i^{\beta}\mathbf{J}_{\beta\alpha}^{ut} = 0$$

from the definitions of the elements of $\mathbf{J}$ given in (7.28). Consequently the part of (7.27)

$$\frac{1}{4}\left(\Delta_i^{rs}\boldsymbol{\Gamma}_r^{\alpha}\boldsymbol{\Gamma}_s^{\beta}\left[\mathbf{F}_\alpha^u\mathbf{F}_\beta^t - \mathbf{F}_\alpha^t\mathbf{F}_\beta^u\right]\boldsymbol{\Gamma}^{T\gamma}_{\ u}\boldsymbol{\Gamma}^{T\delta}_{\ t}\Delta_{\gamma\delta}^j\right)$$

is equal to zero, so that

$$\begin{aligned} R &= \frac{1}{4}\left(\Delta_i^{rs}\boldsymbol{\Gamma}_r^{\alpha}\boldsymbol{\Gamma}_s^{\beta}\left[\mathbf{F}_\alpha^t\mathbf{F}_\beta^u + \mathbf{F}_\alpha^u\mathbf{F}_\beta^t\right]\boldsymbol{\Gamma}^{T\gamma}_{\ u}\boldsymbol{\Gamma}^{T\delta}_{\ t}\Delta_{\gamma\delta}^j\right) \\ &= \frac{1}{2}\left(\Delta_i^{rs}\boldsymbol{\Gamma}_r^{\alpha}\boldsymbol{\Gamma}_s^{\beta}\left[\mathbf{F}_\alpha^t\mathbf{F}_\beta^u\right]\boldsymbol{\Gamma}^{T\gamma}_{\ u}\boldsymbol{\Gamma}^{T\delta}_{\ t}\Delta_{\gamma\delta}^j\right) \\ &= \frac{1}{2}\Delta^T(\mathbf{\Gamma}\otimes\mathbf{\Gamma})\left(\left(\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma}\right)^{-1}\otimes\left(\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma}\right)^{-1}\right)(\mathbf{\Gamma^T}\otimes\mathbf{\Gamma^T})\Delta \qquad (7.29) \\ &= \frac{1}{2}\Delta^T\left[\mathbf{\Gamma}\left(\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma}\right)^{-1}\mathbf{\Gamma}^T\right]\otimes\left[\mathbf{\Gamma}\left(\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma}\right)^{-1}\mathbf{\Gamma}^T\right]\Delta \qquad (7.30) \\ &= \frac{1}{2}\left[\mathbf{\Gamma}\left(\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma}\right)^{-1}\mathbf{\Gamma}^T\right]*\left[\mathbf{\Gamma}\left(\boldsymbol{\Gamma}^T\mathbf{M}\boldsymbol{\Gamma}\right)^{-1}\mathbf{\Gamma}^T\right], \qquad (7.31) \end{aligned}$$

which is the required result. ■

To summarize the results, the calculation of the simplified form of the $\mathbf{U}_2$ projective score involves three steps: the first of these is to calculate $\mathbf{H}$ as defined in Theorem 7.3. The most intensive computation involved here is to invert an $r \times r$ matrix, or $s \times s$, depending on which of r or s is smaller.

The second step is to calculate $\mathbf{R}$ as defined in Lemma 7.4. The most intensive computation here involves inverting the $n \times n$ matrix $I + \mathbf{D}^T\mathbf{H}\mathbf{D}$. The final step is to calculate

$$\mathbf{R} - \mathbf{R}\mathbf{C}\left(I + \mathbf{C}^T\mathbf{R}\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{R},$$

where $\mathbf{C}$ involves finding the inverse square root of an $r \times r$ matrix.

The conclusion is that this is a computationally feasible agenda, the most intensive operation being the inversion of an $n \times n$ matrix. Furthermore, it may well even be possible to make further simplifications based in the structure of the matrix $I + \mathbf{D}^T\mathbf{H}\mathbf{D}$.

7.4 Calculation of the cumulants

In this section the relevant cumulants are presented for the Normal, Poisson, Gamma and Binomial models.

Table 7.1 displays the first four cumulants, together with the function $\tau(\kappa_2, \kappa_3, \kappa_4)$ defined as:

$$\tau = \kappa_4 - \kappa_3\kappa_2^{-1}\kappa_3.$$

The relevance of τ is that it is required for the calculation of the matrix $\mathbf{D}$ used in Lemma 7.4.

The parameterization used here for the Gamma distribution is

$$f(y; \lambda, r) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} \exp(-\lambda y),$$

where r, the dispersion parameter is assumed known.

Table 7.1: GLM cumulants, and the τ function.

MODEL	Parameters	κ_1	κ_2	κ_3	κ_4	τ
NORMAL	μ, σ^2	μ	σ^2	0	0	0
POISSON	λ	λ	λ	λ	λ	0
GAMMA	λ, r	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$	$\frac{2\,r}{\lambda^3}$	$\frac{6\,r}{\lambda^4}$	$\frac{2r}{\lambda^4}$
BINOMIAL	p	$n\,p$	$n\,p\,q$	$n\,p\,q\,(q-p)$	$n\,p\,q\,(1-6\,p\,q)$	$-2\,np^2q^2$

Several comments are in order here:

- The fact that $\kappa_3 = 0$ in the normal model means that the scores $\mathbf{U}_1$ and $\mathbf{U}_2$ are identical. The reason is because the skewness matrix $\mathbf{M}_s$ is equal to zero, and the correction to the $\mathbf{U}_1$ estimating equation is a linear multiple of $\mathbf{M}_s$, (see Theorem (7.2)).

- For the Poisson model $\tau = 0$. The consequence is that the correction term derived in Lemma 7.4 is not needed, or in the terminology of the Lemma, $\mathbf{R} = \mathbf{H}$. From a computational viewpoint, this means there will be considerably less work in calculating the $\mathbf{U}_2$ score in this Poisson model, and it is in fact the Poisson model that the computational algorithms were first implemented for. From another perspective, it also means that the corrections terms to the $\mathbf{U}_1$ score may be small, because the Poisson model is very like the normal model in that $\tau = 0$, so that the observed difference between the using $\mathbf{U}_1$ and $\mathbf{U}_2$ may seem slight.

- The gamma model has the correction term $\tau = \frac{2r}{\lambda^4}$, which is strictly positive since both λ and r are greater than 0.

- The binomial model is very interesting. The correction term $\tau = -2np^2q^2$ is clearly negative, and may be zero if the probability is on the edge of the parameter space. Furthermore, when $n = 1$, *i.e.* Logistic regression, the term

$$2\mathbf{M}_i^r\mathbf{M}_j^s + \left\{\Delta\overline{\mathbf{M}} - \widetilde{\mathbf{M}}\mathbf{M}^{-1}\widetilde{\mathbf{M}}\Delta^T\right\}_{ij}^{rs} = 0,$$

  for $i = j = r = s$. Consequently the *proof* of Lemma 7.2 is no longer valid. However, empirical work has shown the statement of the Theorem to still hold, but in this Bernoulli case Lemma 7.4 can at times break down, because the matrix

$$(I - \mathbf{D}^T\mathbf{H}\mathbf{D})$$

  is under some circumstances singular. Recall that Lemma 7.4 requires the inversion of this matrix. It is worth re-emphasizing that this problem can only occur if $n = 1$.

- Large sample results can be approached in the Gamma and Binomial problems by letting r and n get large respectively. The rationale is that the Gamma random variable can be considered as the *sum* of r independent exponential random variables with rate λ and the Binomial can be thought of as the *sum* of n independent Bernoulli random variables. In the gamma case, the correction term τ is $O(r)$, and in the Binomial case τ is $O(n)$. But the main term

$$\mathbf{M}_i^r\mathbf{M}_j^s + \mathbf{M}_i^s\mathbf{M}_j^r$$

  is $O(r^2)$ and $O(n^2)$ respectively. Consequently, for large sample sizes, the correction term described in Lemma 7.4 will be negligible, and this may mean that from a practical perspective the correction term $\mathbf{HD}(I \pm \mathbf{D}^T\mathbf{HD})^{-1}\mathbf{D}^T\mathbf{H}$ in Lemma 7.4 is unnecessary.

7.5 Calculations in the Poisson model

To test the practicality of using the projective score methodology the Poisson model was selected. In this section some of the computational issues associated with calculating the projective score $\mathbf{U}_2$, solving it to find estimates of the parameters of interest, and calculating appropriate standard errors for these parameters will be addressed.

For the purposes of finding its root, the substitution version of the $\mathbf{U}_2$ score equation can be considered as a function of c, where c is a vector in $\mathbb{R}^r$. Recall that the *m.l.e.* for $\mathbf{m}$ when β is fixed is (7.18)

$$\hat{\mathbf{m}}_\beta = (\mathbf{y} - \mathbf{Z}c).$$

To solve $\mathbf{U}_2(c) = 0$ for c, an iterative algorithm is needed. In this thesis two methods, a Newton-Raphson and a Fisher Scoring type approach were used. The Newton-Raphson method uses an updating method as follows:

$$c_{n+1} = c_n - [\nabla U_2(c_n)]^{-1}\, U_2(c_n).$$

The derivative matrix is calculated using numerical differentiation. The complexity of $\mathbf{U}_2$ as a function of c makes this numerical differentiation time intensive, however the convergence to a root is typically achieved by 4 iterations.

The second method, the Fisher type scoring, uses the fact that $\mathbf{U}_1 \approx \mathbf{U}_2$. Now, $\mathbf{U}_1 = \mathbf{Z}^T\,\mathbf{\Xi}(\mathbf{y} - \mathbf{m})$, and when $\mathbf{m}$ is substituted by its estimate $(\mathbf{y} - \mathbf{Z}c)$

considerable simplifications occur so that

$$\mathbf{U}_1(c) = \mathbf{Z}^T\mathbf{Z}c,$$

the derivative of which, with respect to c is simply $\mathbf{Z}^T\mathbf{Z}$. The fact that $\mathbf{Z}^T\mathbf{Z}$ is a constant is the reason for referring to the method as Fisher scoring. Clearly it is much quicker to find $\mathbf{Z}^T\mathbf{Z}$ than $[\nabla U_2(c_n)]$, but there is a slight trade-off in the number of iterations needed until convergence.

All iterative methods need a starting value c_0, and in this problem, a very natural starting value is to use the root of the $\mathbf{U}_1(c)$ score equation, which is simply $c = 0$. This seemed to work well in all the data sets tried.

Denoting the solution to $\mathbf{U}_2(c) = 0$ as $\hat{c}$, a transformation of $\hat{c}$ is needed in order to obtain the estimate of the parameter of interest β. For the Poisson problem the function $b(\eta)$ in (7.2) is equal to $\exp(\eta)$, so that the appropriate formula in (7.20) is

$$\tilde{\beta} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T \log(\mathbf{Y} - \mathbf{Z}\hat{c}),$$

where $\tilde{\beta}$ is the desired solution to the $\mathbf{U}_2$.

Since the $\mathbf{U}_2$ score equation is originally solved for c, the derivative of $\mathbf{U}_2$ with respect to c, is incorrect as a measure of information on the β scale. In order to obtain the correct information, the Jacobian of the transformation from c to β must be included.

That is

$$\frac{d\mathbf{U}_2}{d\beta} = \frac{d\mathbf{U}_2}{dc}\frac{dc}{d\beta}.$$

It can be shown that the derivative $\frac{d\beta}{dc}$ is equal to $(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\hat{\mathbf{M}}^{-1}\mathbf{Z}$, the inverse of this quantity being used in the calculation of the information on the β scale.

To illustrate how the projective score functions mimic the conditional score function, a simple case will be presented in which there exists a tractable conditional distribution. The data is fabricated for illustrative purposes, and consists of a table of Poisson counts as follows:

$$\mathbf{Y} = \begin{pmatrix} Y_1 & Y_3 \\ Y_2 & Y_4 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 1 & 3 \end{pmatrix} \quad \text{and } E[\mathbf{Y}] = \begin{pmatrix} m_1 & m_3 \\ m_2 & m_4 \end{pmatrix}.$$

The contrast of interest is the column effect, that is $\mathbf{Z}^T = (1, 1, -1, -1)$ so the distribution of Y_1, conditional on all contrasts orthogonal to $\mathbf{Z}$ can be shown to be of the form

$$P(Y_1 = k \,|\, \text{Data contrasts orthogonal to } \mathbf{Z}) \propto \binom{4}{k} \left(\frac{m_1 m_2}{m_3 m_4}\right)^k.$$

Table 7.2 presents the estimates of the parameter β, and the estimated observed information in the U_1, U_2 and conditional score equations. The standard errors of the estimates are also presented, these are calculated as the inverse of the square root of the information.

The root of the U_2 score equation has managed to bridge about 70% of the gap between the roots of the U_1 and U_c score equations. It certainly seems fair to say that it has lived up to its rôle of spanning the divide between unconditional and conditional inference. One interesting feature is the *increase* in the information between U_1 and U_c. The theory developed in Section 5 of Chapter 2 indicates that the information in the score equations should be *decreasing* from U_1 down to U_c. The informations displayed in Table 7.2 are not contradictory to this statement, since they are estimated observed informations, but it is somewhat surprising to see the order reversed. It is also worth remarking that the estimates *shrink* toward the conditional estimate, a phenomena that is repeated in the next set of examples in this

Table 7.2: Estimate and information comparison for the sample data.

	U_1	U_2	U_c
Estimate	0.549	0.488	0.464
Information	6.000	6.940	7.667
s.e.	0.408	0.380	0.361

chapter. We note without proof that exactly the same qualitative characteristics are displayed in the Bernoulli matched pairs model (Example 1). Here the unconditional estimate shrinks by a factor of about one half toward the conditional estimate of the parameter of interest, and the estimated observed information increases between the unconditional and conditional estimate.

Figure 7.1 displays the three score functions, the larger graph is a magnification of the smaller graph about the roots of score equations. It can be seen that, as expected, the substitution version of the U_2 score function is a better approximation to the conditional score function than the U_1 score function.

Two examples will now be presented that allow a comparison of the roots of the $\mathbf{U}_1$ and $\mathbf{U}_2$ estimating equations. The idea is not to fully analyse the data, but to illustrate the differences between the roots of the $\mathbf{U}_1$ and $\mathbf{U}_2$ score equations. The differences between the estimates can be interpreted as either the correction for curvature in the model, or as a movement toward conditionality. The examples

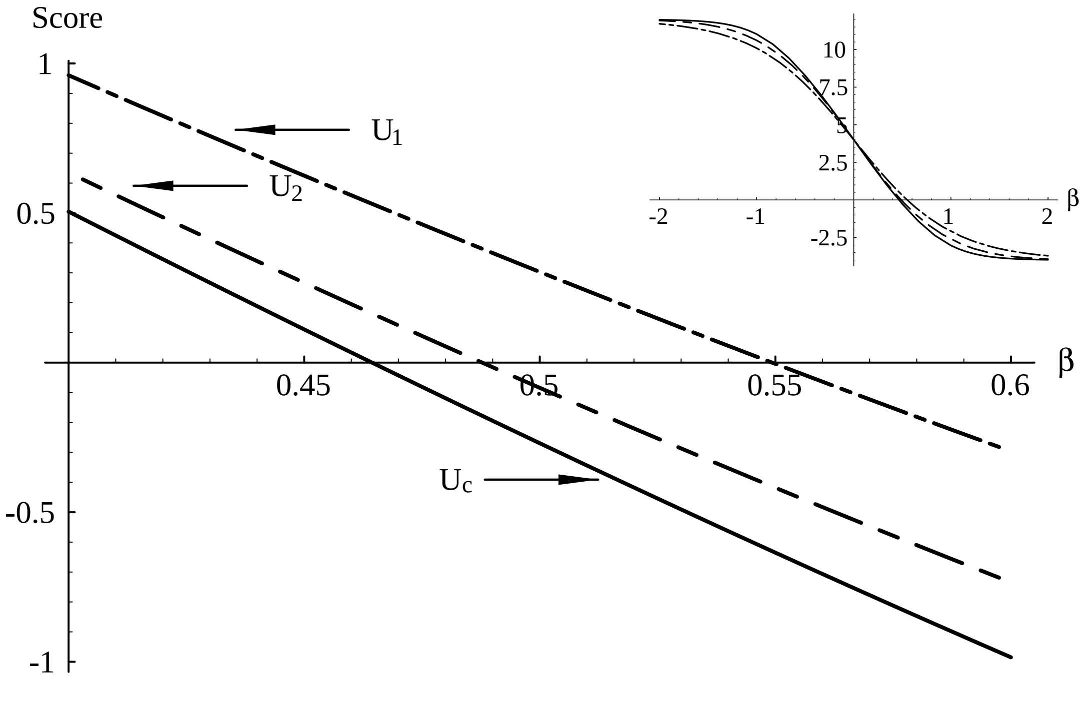

Figure 7.1: Comparison of the U_1, U_2 and U_c projective scores.

Table 7.3: Death, months of exposure, and risk, by age of patient and type of heart valve.

		Types of Heart Valve	
Age		Aortic	Mitral
< 55	Deaths	4	1
	Exposure	1259	2082
	Risk	0.0032	0.0005
55+	Deaths	7	9
	Exposure	1417	1647
	Risk	0.0049	0.0055

Source: Laird and Olivier (1981)

are both taken from Agresti (1990). The first concerns data from a heart valve replacement study. Since there is an *exposure* variable available the goal is to study the *risk* of death. The exposure enters the model through an *offset*, but causes no additional difficulty because it does not affect any of the score equations. The data is in Table 7.3.

Table 7.4 presents the parameter estimates for the main effects. The estimates under Model I, are simply those obtained by fitting a standard main effects model to the data, and are included as a point of interest only. The interest here, is

Table 7.4: Parameter estimates for the heart valve data.

Model I		Model II			
(A,V)	*m.l.e.*	$\hat{U}_1$	$\hat{U}_2$	$\hat{U}_1$	$\hat{U}_2$
$-2\,\hat{\beta}_A$	1.221	1.436	1.366	-	-
	(0.514)	(0.614)	(0.593)	-	-
$-2\,\hat{\beta}_V$	-0.330	-	-	-0.894	-0.871
	(0.438)	-	-	(0.614)	(0.601)

to compare the difference between the roots of the U_1 and U_2 estimating equations, which are in the columns under Model II. In order to obtain the main effect estimate for age, $\hat{\beta}_A$, the $\mathbf{Z}$ matrix would be set equal to $(1, 1, -1, -1)$ and to obtain $\hat{\beta}_V$ the $\mathbf{Z}$ matrix would be $(1, -1, 1, -1)$. Note that these were not estimated simultaneously. In fact, in this model, if the main effects are estimated simultaneously then the U_1 estimates are *identical* to the U_2 estimates, which is a consequence of the fact that if

$$\mathbf{Z} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ -1 & -1 \end{pmatrix},$$

then $U_1 = U_c$!

There is not a noticible difference between the U_1 and U_2 estimates, which

Table 7.5: Death penalty data.

		Death Penalty	
Defendant's Race	Victim's Race	Yes	No
White	White	19	132
	Black	0	9
Black	White	11	52
	Black	6	97

Source: Radelet (1981)

may be explained by the fact that in these Poisson models $\tau = 0$, or that the original score function U_1 is "almost" conditional in the first place, so that there is not far to go to obtain true conditionality. Note though, the consistent shrinking of the estimates and decreasing of the standard errors.

The next set of data, Table 7.5 tabulates the death penalty verdict by defendant's race and victim's race, (see Agresti (1990, section 5.2.2) for a full description of the data). Here, V stands for Victim's Race, D stands for Defendant's Race and P stands for the Death Penalty.

Table 7.6 presents the parameter estimates. The parameter estimates in the Model I column are for the model (VP,DV), which represents the independence

Table 7.6: Parameter estimates for the death penalty data.

Model I		Model II			
(VP,DV)	*m.l.e.*	$\hat{U}_1$	$\hat{U}_2$	$\hat{U}_1$	$\hat{U}_2$
$\hat{\beta}_{DV}$	0.828	0.793	0.772	-	-
	(0.095)	(0.195)	(0.191)	-	-
$\hat{\beta}_{VP}$	0.264	-	-	0.277	0.260
	(0.116)	-	-	(0.195)	(0.191)

of the death penalty verdict and the defendants race given the victim's race. The estimates based on the projective score equations are under the columns for Model II.

The parameters of interest here are $\hat{\beta}_{DV}$ and $\hat{\beta}_{VP}$. Again, note the shrinking of the parameter estimates between the U_1 and U_2 score equations, and the *decrease* in the standard errors of the estimates. The fact that there is a zero cell count in Table 7.5 is an issue. Here, the common practice of adding one half to each cell count was used to move the data away from the edge of the sample space. It would be worth studying the sensitivity of the estimates to this sort of *smoothing* procedure, since conditional approaches to inference are especialy relevant to data sets with small cell counts, in which there will frequently be zero cells.