Estimating standard errors in rank deficient models

We give some technical details on the estimation of standard errors in rank deficient models. Further details can be found in the Appendix to Currie (2013).

We consider the following setup. We have a generalized linear model or GLM with model matrix $\boldsymbol{X}$, $n \times p$, $n > p$, and rank $p - q$, $q > 0$; we denote the regression coefficients by $\boldsymbol{\theta}$, $p \times 1$. Then there exists a matrix $\boldsymbol{H}$

$$\boldsymbol{X}_{\text{Aug}} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{H} \end{bmatrix}$$

where $\boldsymbol{H}$ is $q \times p$ such that $\boldsymbol{X}_{\text{Aug}}$ has full column rank p. In our present application $\boldsymbol{H}$ is a possible constraints matrix. We note in particular that $\boldsymbol{H}$ is not unique. We suppose that $\boldsymbol{H\theta} = \boldsymbol{0}$.

For example, in the APC model, $\boldsymbol{X}$ is $n_a n_y \times (n_a + n_y + n_c)$ where n_a is the number of ages, n_y is the number of years and $n_c = n_a + n_y - 1$ is the number of cohorts; the constraints matrix $\boldsymbol{H}$ is $3 \times (n_a + n_y + n_c)$. See Macdonald et al (2018, chaps. 10-13) for some examples of model matrices.

Define

$$\boldsymbol{\Delta} = \boldsymbol{X}'\hat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{H}'\boldsymbol{H}.$$

Here $\hat{\boldsymbol{W}}$ is a diagonal matrix of weights; in the case of a Poisson model the entries in $\hat{\boldsymbol{W}}$ are the fitted numbers of deaths. Note that, unlike $\boldsymbol{X}'\hat{\boldsymbol{W}}\boldsymbol{X}$ which is singular, $\boldsymbol{\Delta}$ is non-singular. Now let

$$\boldsymbol{\Psi} = \boldsymbol{\Delta}^{-1} - \boldsymbol{\Delta}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{\Delta}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{\Delta}^{-1}.$$

Then

$$\text{Var}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Psi}.$$

This is Corollary 2 in the Appendix to my 2013 paper. We note that $\boldsymbol{\Psi}$ is a singular variance matrix; this occurs because $\boldsymbol{X}$ is not of full rank.

Now let's look at the standard errors of the fitted values. The full covariance matrix, $\boldsymbol{V}$ say, of the fitted values $\boldsymbol{X}\hat{\boldsymbol{\theta}}$ is given by

$$\boldsymbol{V} = \text{Var}(\boldsymbol{X}\hat{\boldsymbol{\theta}}) = \boldsymbol{X}\text{Var}(\hat{\boldsymbol{\theta}})\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{\Psi}\boldsymbol{X}'.$$

The matrix $\boldsymbol{V}$ can be rather large and if we are only interested in its diagonal elements, the variances of the fitted values, then we have the following neat formula

$$\text{diag}\{\boldsymbol{X}\boldsymbol{\Psi}\boldsymbol{X}'\} = [(\boldsymbol{X}\boldsymbol{\Psi}) * \boldsymbol{X}]\boldsymbol{1}_p$$

where $*$ indicates element-by-element multiplication and $\boldsymbol{1}_p$ is a vector of 1s with length p.

It is a simple matter to check numerically that $\boldsymbol{V}$ does not depend on the particular chosen $\boldsymbol{H}$. Jim Howie of the Mathematics Department here at Heriot-Watt has provided the following elegant mathematical proof.

It is convenient to work with row vectors so here $\boldsymbol{x}$ will stand for a row vector rather than the usual column vector. We first prove the following lemma.

Lemma: Let $\boldsymbol{H}$, $q \times p$, be such that

$$\boldsymbol{X}_{\text{Aug}} = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{H} \end{bmatrix}$$

has full column rank p. Let $\breve{\boldsymbol{H}}$, $q \times p$, be an alternative version of $\boldsymbol{H}$ such that

$$\breve{\boldsymbol{X}}_{\text{Aug}} = \begin{bmatrix} \boldsymbol{X} \\ \breve{\boldsymbol{H}} \end{bmatrix}$$

also has full column rank p. Then there exists an invertible matrix $\boldsymbol{Y}$, $p \times p$, such that

$$\begin{aligned} \boldsymbol{XY} &= \boldsymbol{X}, \text{ and} && (1) \\ \boldsymbol{HY} &= \breve{\boldsymbol{H}}. && (2) \end{aligned}$$

Proof: The proof is by construction of $\boldsymbol{Y}$. Let $\boldsymbol{U}$ be a maximal linearly independent subset of the rows of $\boldsymbol{X}$. Now $\boldsymbol{X}$ is $n \times p$ with rank $p - q$ so $\boldsymbol{U}$ is $(p - q) \times p$. Define $\boldsymbol{A}$ as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{H} \end{bmatrix}. \tag{3}$$

Now the rows of $\boldsymbol{H}$ are linearly independent of the rows of $\boldsymbol{X}$ and hence of $\boldsymbol{U}$. Thus $\boldsymbol{A}$, $p \times p$, has rank p and so is non-singular. In the same way, the matrix $\boldsymbol{B}$ where

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{U} \\ \breve{\boldsymbol{H}} \end{bmatrix} \tag{4}$$

is also $p \times p$ and non-singular. We define $\boldsymbol{Y}$ as

$$\boldsymbol{Y} = \boldsymbol{A}^{-1}\boldsymbol{B} \tag{5}$$

and note that $\boldsymbol{Y}$ is non-singular. Evidently, from the definitions of $\boldsymbol{Y}$, $\boldsymbol{A}$ and $\boldsymbol{B}$, we have

$$\begin{aligned} \boldsymbol{AY} &= \boldsymbol{B} \\ \Rightarrow \boldsymbol{UY} &= \boldsymbol{U} && (6) \\ \text{and } \boldsymbol{HY} &= \breve{\boldsymbol{H}}. && (7) \end{aligned}$$

Now let $\boldsymbol{x}$ be a row of $\boldsymbol{X}$ and let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{p-q}$ be the rows $\boldsymbol{U}$. Now $\boldsymbol{x}$ is linearly dependent on the rows of $\boldsymbol{U}$ so

$$\boldsymbol{x} = f_1\boldsymbol{u}_1 + \ldots + f_{p-q}\boldsymbol{u}_{p-q}$$

for some coefficients $f_1, \ldots, f_{p-q}$. Further, $\boldsymbol{u}_i \boldsymbol{Y} = \boldsymbol{u}_i$, for $i = 1, \ldots, p-q$ by (6), so

$$\begin{aligned} \boldsymbol{x}\boldsymbol{X} &= f_1 \boldsymbol{u}_1 \boldsymbol{Y} + \ldots f_{p-q} \boldsymbol{u}_{p-q} \boldsymbol{Y} \\ &= f_1 \boldsymbol{u}_1 + \ldots f_{p-q} \boldsymbol{u}_{p-q} \\ &= \boldsymbol{x}. \end{aligned}$$

Since this holds for all rows of $\boldsymbol{X}$ we have $\boldsymbol{XY} = \boldsymbol{X}$ and together with (7) we have proved (1) and (2). $\square$

We have provided an example of a matrix $\boldsymbol{Y}$ which satisfies (1) and (2). In fact, such a matrix is unique and is given by (5) where $\boldsymbol{A}$ and $\boldsymbol{B}$ are defined in (3) and (4) respectively. For completeness, we demonstrate uniqueness in

Corollary: The matrix $\boldsymbol{Y}$ which satisfies (1) and (2) is unique.

Proof: We use the same notation as the lemma. Let $\boldsymbol{Y}^*$ be any matrix satisfying (1) and (2), ie, $\boldsymbol{XY}^* = \boldsymbol{X}$ and $\boldsymbol{HY}^* = \breve{\boldsymbol{H}}$.

$$\begin{aligned} \boldsymbol{XY}^* &= \boldsymbol{X} \\ \Rightarrow \boldsymbol{UY}^* &= \boldsymbol{U} \text{ since the rows of } \boldsymbol{U} \text{ are a subset of the rows of } \boldsymbol{X} \\ \Rightarrow \begin{bmatrix} \boldsymbol{U} \\ \boldsymbol{H} \end{bmatrix} \boldsymbol{Y}^* &= \begin{bmatrix} \boldsymbol{U} \\ \breve{\boldsymbol{H}} \end{bmatrix} \\ \Rightarrow \boldsymbol{AY}^* &= \boldsymbol{B} \text{ by (3) and (4)} \\ \Rightarrow \boldsymbol{Y}^* &= \boldsymbol{A}^{-1}\boldsymbol{B} \end{aligned}$$

which is the definition of $\boldsymbol{Y}$ in (5). $\square$

We remark $\boldsymbol{A}$ and $\boldsymbol{B}$ are not unique since they are dependent on the choice of $\boldsymbol{U}$; nevertheless the product $\boldsymbol{A}^{-1}\boldsymbol{B}$ is unique.

The main result on the invariance of $\boldsymbol{X\Psi X'}$ now follows easily.

Proposition: The variance matrix $\boldsymbol{V} = \boldsymbol{X\Psi X'}$ does not depend on the choice of $\boldsymbol{H}$.

Proof: We use the $\breve{}$ accent to denote the alternative definitions of $\boldsymbol{H}$ and $\boldsymbol{\Delta}$. We use (1) and (2) repeatedly to obtain the following identities.

$$\begin{aligned} \breve{\boldsymbol{\Delta}} &= \boldsymbol{X}'\hat{\boldsymbol{W}}\boldsymbol{X} + \breve{\boldsymbol{H}}'\breve{\boldsymbol{H}} \\ &= \boldsymbol{Y}'\boldsymbol{X}'\hat{\boldsymbol{W}}\boldsymbol{XY} + \boldsymbol{Y}'\boldsymbol{H}'\boldsymbol{HY}' \\ &= \boldsymbol{Y}'\boldsymbol{\Delta Y}, \end{aligned} \tag{8}$$

$$\begin{aligned} \breve{\boldsymbol{H}}\breve{\boldsymbol{\Delta}}^{-1}\breve{\boldsymbol{H}}' &= \boldsymbol{HY}(\boldsymbol{Y}'\boldsymbol{\Delta Y})^{-1}\boldsymbol{Y}'\boldsymbol{H}' \text{ by (8)} \\ &= \boldsymbol{H\Delta}^{-1}\boldsymbol{H}' \text{ since } \boldsymbol{Y} \text{ is invertible.} \end{aligned} \tag{9}$$

Further,

$$
\begin{aligned}
\breve{\boldsymbol{\Psi}} &= \breve{\boldsymbol{\Delta}}^{-1} - \breve{\boldsymbol{\Delta}}^{-1}\breve{\boldsymbol{H}}'(\breve{\boldsymbol{H}}\breve{\boldsymbol{\Delta}}^{-1}\breve{\boldsymbol{H}}')^{-1}\breve{\boldsymbol{H}}\breve{\boldsymbol{\Delta}}^{-1} \\
&= (\boldsymbol{Y}'\boldsymbol{\Delta}\boldsymbol{Y})^{-1} - (\boldsymbol{Y}'\boldsymbol{\Delta}\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{\Delta}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{Y}(\boldsymbol{Y}'\boldsymbol{\Delta}\boldsymbol{Y})^{-1} \text{ by (8) and (9)} \\
&= \boldsymbol{Y}^{-1}\boldsymbol{\Delta}^{-1}\boldsymbol{Y}'^{-1} - \boldsymbol{Y}^{-1}\boldsymbol{\Delta}^{-1}\boldsymbol{H}'(\boldsymbol{H}\boldsymbol{\Delta}^{-1}\boldsymbol{H}')^{-1}\boldsymbol{H}\boldsymbol{\Delta}^{-1}\boldsymbol{Y}'^{-1} \text{ since } \boldsymbol{Y} \text{ is invertible} \\
&= \boldsymbol{Y}^{-1}\boldsymbol{\Psi}\boldsymbol{Y}'^{-1}, \qquad (10)
\end{aligned}
$$

and finally

$$
\begin{aligned}
\boldsymbol{X}\breve{\boldsymbol{\Psi}}\boldsymbol{X}' &= \boldsymbol{X}\boldsymbol{Y}^{-1}\boldsymbol{\Psi}\boldsymbol{Y}'^{-1}\boldsymbol{X}' \text{ by (10)} \\
&= \boldsymbol{X}\boldsymbol{\Psi}\boldsymbol{X}'
\end{aligned}
$$

as required, since $\boldsymbol{X}\boldsymbol{Y} = \boldsymbol{X} \Rightarrow \boldsymbol{X}\boldsymbol{Y}^{-1} = \boldsymbol{X}$. □

References

Currie, I.D. (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, **13**, 69-93.

Macdonald, A.S., Richards, S.J. and Currie, I.D. (2018). *Modelling Mortality with Actuarial Applications*, Cambridge University Press: Cambridge.