The Cox–Reid modified score: a comment

Manuel Arellano

April 10, 2005

1. The Cox-Reid modified likelihood

Let $\ell_i(\beta, \eta_i)$ be an individual log-likelihood conditioned on z_i, and let $d_{\beta i}(\beta, \eta_i)$, $d_{\eta i}(\beta, \eta_i)$, $d_{\eta\eta i}(\beta, \eta_i)$ and $d_{\beta\eta i}(\beta, \eta_i)$ be first and second partial derivatives. The first argument is a vector common parameter β and η_i is a scalar individual effect. Let $\ell_i(\beta, \widehat{\eta}_i(\beta))$ be the concentrated log likelihood, so that $d_{\beta i}(\beta, \widehat{\eta}_i(\beta))$ is the concentrated score.

Suppose there exists an information orthogonal reparameterization of the fixed effects, meaning that there is a function $\eta_i = \eta(\beta, \lambda_i)$ such that the reparameterized log likelihood $\ell_i^*(\beta, \lambda_i) = \ell_i(\beta, \eta(\beta, \lambda_i))$ satisfies at true values:

$$E\left(\frac{\partial^2 \ell_i^*(\beta_0, \lambda_{0i})}{\partial\beta\partial\lambda_i}\right) = 0$$

where the expectation is conditioned on fixed effects and covariates.

The modified concentrated log likelihood function of Cox and Reid (1987) can be written as $L_M(\beta) = \sum_i \ell_{Mi}(\beta)$ with

$$\ell_{Mi}(\beta) = \ell_i^*\left(\beta, \widehat{\lambda}_i(\beta)\right) - \frac{1}{2}\log\left[-d_{\lambda\lambda i}^*\left(\beta, \widehat{\lambda}_i(\beta)\right)\right], \tag{1.1}$$

where $\widehat{\lambda}_i(\beta)$ is the MLE of λ_i for given β, and $d_{\lambda\lambda i}^*(\beta, \lambda_i) = \partial^2\ell_i^*/\partial\lambda_i^2$.

An orthogonal reparameterization may not exist, but if it does it is not unique in general, and different reparameterizations may lead to different modified likelihoods, so that a Cox–Reid estimator may not exist or there may be many of them.

Finding an orthogonal reparameterization Since we have

$$\begin{aligned}\frac{\partial \ell_i^*}{\partial \lambda_i} &= \frac{\partial \ell_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \lambda_i} \\ \frac{\partial^2 \ell_i^*}{\partial \beta \partial \lambda_i} &= \left(\frac{\partial^2 \ell_i}{\partial \beta \partial \eta_i} + \frac{\partial^2 \ell_i}{\partial \eta_i^2}\frac{\partial \eta_i}{\partial \beta}\right)\frac{\partial \eta_i}{\partial \lambda_i} + \frac{\partial^2 \eta_i}{\partial \beta \partial \lambda_i}\frac{\partial \ell_i}{\partial \eta_i},\end{aligned}$$

evaluating at true values, taking expectations and equating to zero, it turns out that an orthogonal function $\eta(\beta, \lambda_i)$ must satisfy

$$E\left(\frac{\partial^2 \ell_i\left(\beta_0, \eta(\beta_0, \lambda_{0i})\right)}{\partial \beta \partial \eta_i}\right) + \frac{\partial \eta_i\left(\beta_0, \lambda_{0i}\right)}{\partial \beta} E\left(\frac{\partial^2 \ell_i\left(\beta_0, \eta(\beta_0, \lambda_{0i})\right)}{\partial \eta_i^2}\right) = 0.$$

Thus,

$$\frac{\partial \eta_i\left(\beta_0, \lambda_{0i}\right)}{\partial \beta} = -\frac{E\left[d_{\beta\eta i}\left(\beta_0, \eta_{0i}\right)\right]}{E\left[d_{\eta\eta i}\left(\beta_0, \eta_{0i}\right)\right]} = -q_i\left(\beta_0, \eta_{0i}\right), \text{ say.} \tag{1.2}$$

Note that $\eta_{0i} = \eta(\beta_0, \lambda_{0i})$. In addition,

$$\frac{\partial^2 \eta_i}{\partial \beta \partial \lambda_i} = -\frac{\partial q_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \lambda_i}$$

or

$$\frac{\partial}{\partial \beta} \ln\left|\frac{\partial \eta_i}{\partial \lambda_i}\right| = -\frac{\partial q_i}{\partial \eta_i} = -q_{\eta i}\left(\beta, \eta_i\right), \text{ say.} \tag{1.3}$$

2. Cox–Reid in terms of the original parameterization

Firstly, note that because of the invariance of MLE $\widehat{\eta}_i(\beta) = \eta(\beta, \widehat{\lambda}_i(\beta))$ and

$$\ell_i^*\left(\beta, \widehat{\lambda}_i(\beta)\right) = \ell_i\left(\beta, \widehat{\eta}_i(\beta)\right). \tag{2.1}$$

Next, the term $d^*_{\lambda\lambda i}\left(\beta, \widehat{\lambda}_i(\beta)\right)$ can be calculated as the product of the Fisher information in the (β, η_i) parameterization and the square of the Jacobian of the transformation from (β, η_i) to (β, λ_i). That is, since the second derivatives of ℓ_i^* and ℓ_i are related by the expression

$$\frac{\partial^2 \ell_i^*}{\partial \lambda_i^2} = \frac{\partial^2 \ell_i}{\partial \eta_i^2}\left(\frac{\partial \eta_i}{\partial \lambda_i}\right)^2 + \frac{\partial \ell_i}{\partial \eta_i}\left(\frac{\partial^2 \eta_i}{\partial \lambda_i^2}\right),$$

and $\partial \ell_i / \partial \eta_i$ vanishes at $\widehat{\eta}_i(\beta)$, we have

$$d^*_{\lambda\lambda i}\left(\beta, \widehat{\lambda}_i(\beta)\right) = d_{\eta\eta i}\left(\beta, \widehat{\eta}_i(\beta)\right)\left(\frac{\partial \eta_i}{\partial \lambda_i}\Big|_{\lambda_i = \widehat{\lambda}_i(\beta)}\right)^2. \tag{2.2}$$

Thus, the Cox–Reid log likelihood can be written as

$$\ell_{Mi}(\beta) = \ell_i(\beta, \widehat{\eta}_i(\beta)) - \frac{1}{2}\log\left[-d_{\eta\eta i}(\beta, \widehat{\eta}_i(\beta))\right] + \log\left(\frac{\partial \lambda_i}{\partial \eta_i}\mid_{\eta_i=\widehat{\eta}_i(\beta)}\right). \tag{2.3}$$

The Cox-Reid score Differentiating (2.3):

$$\frac{\partial \ell_{Mi}(\beta)}{\partial \beta} = d_{\beta i}(\beta, \widehat{\eta}_i(\beta)) - \frac{1}{2}\frac{\partial}{\partial \beta}\log\left[-d_{\eta\eta i}(\beta, \widehat{\eta}_i(\beta))\right] - \frac{\partial}{\partial \beta}\log\left(\frac{\partial \eta_i}{\partial \lambda_i}\mid_{\eta_i=\widehat{\eta}_i(\beta)}\right).$$

To look at the form of the last term let us introduce the notation

$$c_i(\beta, \eta_i) = -\log\frac{\partial \eta_i}{\partial \lambda_i},$$

regarded as $c_i(\beta, \eta_i) = c_i(\beta, \eta(\beta, \lambda_i))$. In view of (1.3):

$$-\frac{\partial}{\partial \beta}\log\frac{\partial \eta_i}{\partial \lambda_i} = \frac{\partial c_i}{\partial \beta} + \frac{\partial c_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta} = q_{\eta i}(\beta, \eta_i).$$

Also, using hats to denote terms evaluated at $\widehat{\eta}_i(\beta)$, we have

$$\begin{aligned}
-\frac{\partial}{\partial \beta}\log\left(\frac{\partial \eta_i}{\partial \lambda_i}\mid_{\eta_i=\widehat{\eta}_i(\beta)}\right) &= \frac{\partial \widehat{c}_i}{\partial \beta} + \frac{\partial \widehat{c}_i}{\partial \eta_i}\frac{\partial \widehat{\eta}_i(\beta)}{\partial \beta} = q_{\eta i}(\beta, \widehat{\eta}_i(\beta)) + \frac{\partial \widehat{c}_i}{\partial \eta_i}\left(\frac{\partial \widehat{\eta}_i(\beta)}{\partial \beta} - \widehat{\frac{\partial \eta_i}{\partial \beta}}\right) \\
&= q_{\eta i}(\beta, \widehat{\eta}_i(\beta)) + \frac{\partial \widehat{c}_i}{\partial \eta_i}\left(-\frac{d_{\beta\eta i}(\beta, \widehat{\eta}_i(\beta))}{d_{\eta\eta i}(\beta, \widehat{\eta}_i(\beta))} + q_i(\beta, \widehat{\eta}_i(\beta))\right)
\end{aligned}$$

Therefore, the Cox–Reid score is given by

$$\begin{aligned}
\frac{\partial \ell_{Mi}(\beta)}{\partial \beta} &= d_{\beta i}(\beta, \widehat{\eta}_i(\beta)) - \frac{1}{2}\frac{\partial}{\partial \beta}\log\left[-d_{\eta\eta i}(\beta, \widehat{\eta}_i(\beta))\right] \\
&\quad + q_{\eta i}(\beta, \widehat{\eta}_i(\beta)) - \frac{\partial \widehat{c}_i}{\partial \eta_i}\left(\frac{d_{\beta\eta i}(\beta, \widehat{\eta}_i(\beta))}{d_{\eta\eta i}(\beta, \widehat{\eta}_i(\beta))} - q_i(\beta, \widehat{\eta}_i(\beta))\right)
\end{aligned} \tag{2.4}$$

Note that

$$-\frac{\partial}{\partial \lambda_i}\log\frac{\partial \eta_i}{\partial \lambda_i} = \frac{\partial c_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \lambda_i}$$

so that

$$\frac{\partial c_i}{\partial \eta_i} = -\frac{1}{\left(\frac{\partial \eta_i}{\partial \lambda_i}\right)}\frac{\partial}{\partial \lambda_i}\log\frac{\partial \eta_i}{\partial \lambda_i} = -\left(\frac{\partial^2 \eta_i}{\partial \lambda_i}\right) / \left(\frac{\partial \eta_i}{\partial \lambda_i}\right)^2.$$

The term $\partial \widehat{c}_i / \partial \eta_i$ is transformation specific, and it will lead to different estimates for different transformations. However, the last term is irrelevant for the purpose of bias reduction. The modified scores considered by Arellano (2003), Carro (2003), and Woutersen (2002) supress this term and are given by

$$d_{Mi}(\beta) = d_{\beta i}(\beta, \widehat{\eta}_i(\beta)) - \frac{1}{2}\frac{\partial}{\partial \beta}\log\left[-d_{\eta\eta i}(\beta, \widehat{\eta}_i(\beta))\right] + q_{\eta i}(\beta, \widehat{\eta}_i(\beta)). \tag{2.5}$$

References

Arellano, M. (2003): "Discrete Choices with Panel Data", *Investigaciones Económicas*, 27, 423-458.

Carro, J. M. (2003): "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects", CEMFI Working Paper No. 0304.

Cox, D. R. and N. Reid (1987): "Parameter Orthogonality and Approximate Conditional Inference" (with discussion), *Journal of the Royal Statistical Society*, Series B, 49, 1-39.

Woutersen T. (2002): "Robustness Against Incidental Parameters", unpublished.