Solution to Exercise 11.18

⋆11.18 Show that the expectation of the Hessian of the loglikelihood function (11.41), evaluated at the parameter vector $\boldsymbol{\theta}$, is equal to the negative of the $k\times k$ matrix

$$\boldsymbol{I}(\boldsymbol{\theta}) \equiv \sum_{t=1}^{n}\sum_{j=0}^{J}\frac{1}{\Pi_{tj}(\boldsymbol{\theta})}\,\boldsymbol{T}_{tj}^{\top}(\boldsymbol{\theta})\boldsymbol{T}_{tj}(\boldsymbol{\theta}), \tag{11.91}$$

where $\boldsymbol{T}_{tj}(\boldsymbol{\theta})$ is the $1\times k$ vector of partial derivatives of $\Pi_{tj}(\boldsymbol{\theta})$ with respect to the components of $\boldsymbol{\theta}$. Demonstrate that (11.91) can also be computed using the outer product of the gradient definition of the information matrix.

Use the above result to show that the matrix of sums of squares and cross-products of the regressors of the DCAR, regression (11.42), evaluated at $\boldsymbol{\theta}$, is $\boldsymbol{I}(\boldsymbol{\theta})$. Show further that $1/s^2$ times the estimated OLS covariance matrix from (11.42) is an asymptotically valid estimate of the covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$ if the artificial variables are evaluated at $\hat{\boldsymbol{\theta}}$.

The contribution to the loglikelihood function (11.41) by observation t is

$$\sum_{j=0}^{J} d_{tj}\log\Pi_{tj}(\boldsymbol{\theta}).$$

The column vector of derivatives of this contribution with respect to $\boldsymbol{\theta}$ is

$$\sum_{j=0}^{J} d_{tj}\,\frac{1}{\Pi_{tj}(\boldsymbol{\theta})}\,\boldsymbol{T}_{tj}^{\top}(\boldsymbol{\theta}). \tag{S11.17}$$

If we then differentiate (S11.17) with respect to $\boldsymbol{\theta}$, we obtain

$$\sum_{j=0}^{J} d_{tj}\,\frac{-1}{\Pi_{tj}^{2}(\boldsymbol{\theta})}\,\boldsymbol{T}_{tj}^{\top}(\boldsymbol{\theta})\boldsymbol{T}_{tj}(\boldsymbol{\theta}) + \sum_{j=0}^{J} d_{tj}\,\frac{1}{\Pi_{tj}(\boldsymbol{\theta})}\,\boldsymbol{T}_{tj}'(\boldsymbol{\theta}), \tag{S11.18}$$

where $\boldsymbol{T}_{tj}'(\boldsymbol{\theta})$ denotes the $k\times k$ matrix of derivatives of $\boldsymbol{T}_{tj}(\boldsymbol{\theta})$ with respect to the vector $\boldsymbol{\theta}$.

The information matrix is minus the expectation of expression (S11.18), summed over all n. In order to take the expectation, we simply replace d_{tj}, which is the only thing that depends on the dependent variables, by its expectation, which is $\Pi_{tj}(\boldsymbol{\theta})$. The result is

$$\sum_{j=0}^{J}\frac{-1}{\Pi_{tj}(\boldsymbol{\theta})}\,\boldsymbol{T}_{tj}^{\top}(\boldsymbol{\theta})\boldsymbol{T}_{tj}(\boldsymbol{\theta}) + \sum_{j=0}^{J}\boldsymbol{T}_{tj}'(\boldsymbol{\theta}). \tag{S11.19}$$

As we saw when discussing (11.43), the fact that the probabilities sum to unity implies that the vector of derivatives with respect to any parameter must sum to 0, which implies that the vectors of second derivatives must also sum to 0. Therefore, the second term in (S11.19) is equal to 0. Changing the sign of the first term and summing over all n yields expression (11.91) for the information matrix, which is what we were required to show.

Of course, we can also obtain (11.91) by using the definition (10.31) of the information matrix in terms of the contributions to the gradient of the loglikelihood function. A typical contribution is (S11.17). The product of (S11.17) with itself transposed is

$$\sum_{j=0}^{J} d_{tj}\,\frac{1}{\Pi_{tj}^2(\boldsymbol{\theta})}\,\boldsymbol{T}_{tj}^\top(\boldsymbol{\theta})\boldsymbol{T}_{tj}(\boldsymbol{\theta}). \tag{S11.20}$$

The t^{th} contribution to the information matrix is the expectation of expression (S11.20). We can obtain this expectation by replacing d_{tj} by $\Pi_{tj}(\boldsymbol{\theta})$. The result is

$$\boldsymbol{I}_t(\boldsymbol{\theta}) = \sum_{j=0}^{J} \frac{1}{\Pi_{tj}(\boldsymbol{\theta})}\,\boldsymbol{T}_{tj}^\top(\boldsymbol{\theta})\boldsymbol{T}_{tj}(\boldsymbol{\theta}).$$

Summing this over all t yields expression (11.91) for the information matrix, as required.

The regressors of the DCAR, regression (11.42), are $\Pi_{tj}^{-1/2}(\boldsymbol{\theta})\boldsymbol{T}_{tj}(\boldsymbol{\theta})$. The product of this vector transposed times itself is

$$\frac{1}{\Pi_{tj}(\boldsymbol{\theta})}\,\boldsymbol{T}_{tj}^\top(\boldsymbol{\theta})\boldsymbol{T}_{tj}(\boldsymbol{\theta}).$$

Summing over $t = 1, \ldots, n$ and $j = 0, \ldots, J$ yields expression (11.91). Thus the matrix of sums of squares and cross products of the artificial regression is the information matrix, as we were required to show.

It is clear from the preceding result that, if we evaluate the inverse of the information matrix at $\hat{\boldsymbol{\theta}}$, the vector of ML estimates, we obtain $\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}})$, which is an asymptotically valid estimate of the covariance matrix of the MLE. The OLS covariance matrix from (11.42) evaluated at $\hat{\boldsymbol{\theta}}$ is equal to s^2 times $\boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}})$. Thus, if we divide this matrix by $1/s^2$, we obtain an asymptotically valid estimate of the covariance matrix of $\hat{\boldsymbol{\theta}}$.