Advanced Econometrics II
TA Session Problems No. 6

Agnieszka Borowska

10.02.2015

Note: this is only a draft of the problems discussed on Tuesday and might contain some typos or more or less imprecise statements. If you find some, please let me know.

Binary Response Models

Recall

$$P_t \equiv \mathbb{P}[y_t = 1|\Omega_t] = \mathbb{E}[y_t|\Omega_t],$$
$$1 - P_t \equiv \mathbb{P}[y_t = 0|\Omega_t] = 1 - \mathbb{E}[y_t|\Omega_t].$$

So, BRN can be thought of as modelling a **conditional expectation**. Since this conditional expectation is also a probability, we need to ensure that it lies in the $0-1$ interval.

For this purpose consider a nonlinear **transformation function** $F(x)$ satisfying the defining properties of a CDF i.e.

$$F(-\infty) = 0,$$
$$F(\infty) = 1,$$
$$f(x) \equiv \frac{dF(x)}{dx} > 0.$$

Then if we use an **index function** to map the vectors X_t and β into a scalar index, we get the required property

$$P_t \equiv \mathbb{E}[y_t \Omega_t] = F(X_t\beta) \in (0,1).$$

BRM: ML Estimation

Chapter	10	11.2
Dependent variable	continuously distributed	binary
Contribution to the likelihood	probability density at y_t	probability that y_t is realized
Likelihood function	joint density	product of Bernoulli trials
Constraint on likelihood	integral of the possible values equal to 1	sum of the possible values equal to 1

y_t - the realized value

Table 1: ML estimation comparison

From Table 1 follows that when the dependent variable can take only **discrete** values, we cannot proceed like in the continuous case and take **joint density** as our likelihood function. Since $y_t = 1$ or $y_t = 0$, with the probability of the former given by $F(X_t\beta)$, now the contribution to the loglikelihood for that observation **if** $y_t = 1$ is $\log F(X_t\beta)$. Similarly, the probability that $y_t = 0$ is $1-F(X_t\beta)$ so the contribution to the loglikelihood for that observation **if** $y_t = 0$ is $\log\Big(1 - F(X_t\beta)\Big)$. This means that the contribution to the likelihood function made by observation t is

$$F(X_t\beta)^{y_t}[1 - F(X_t\beta)]^{1-y_t}.$$

yielding the **contribution to the loglikelihood** function equal to

$$y_t \log F(X_t\beta) + (1 - y_t)\log(1 - F(X_t\beta). \tag{1}$$

Hence the **loglikelihood** function for y is

$$l(y, \beta) = \sum_{t=1}^{n}\Big(y_t \log F(X_t\beta) + (1 - y_t)\log\Big(1 - F(X_t\beta)\Big)\Big). \tag{11.09}$$

BRM: Inference

In subchapter 11.3 it is stated that *it can be shown* that

$$\mathrm{Var}\left(\operatorname{plim} n^{1/2}(\hat{\beta} - \beta_0)\right) = \operatorname{plim}\left(\frac{1}{n}X^T\Upsilon(\beta_0)X\right)^{-1}, \tag{11.15}$$

where $\Upsilon(\beta)$ is a diagonal matrix with the typical element

$$\Upsilon_t(\beta) = \frac{f^2(X_t\beta)}{F(X_t\beta)(1 - F(X_t\beta))}.$$

Below, we will explicitly obtain the **asymptotic covariance matrix** (11.15) for a BRM using general results for ML estimation.

1° First (Ex. 11.7), we will start with deriving of the **asymptotic information matrix** and then compare it with the asymptotic covariance matrix (11.15), to show that the latter is equal to the inverse of the former.

2° Second (Ex. 11.8), we will start with obtaining the **Hessian matrix** corresponding to the loglikelihood function (11.09) and then use the **information matrix equality**[1], to obtain the same result, i.e. that the asymptotic covariance matrix is equal to the RHS of (11.15).

DM, 11.7

1. First, we want to find G_{ti}, the derivative of the contribution to the log-likelihood function made by the t-th observation with respect to β_i.

   The derivative of (1) with respect to β_i is given by

$$G_{ti}(\beta) = \frac{y_t f(X_t\beta)x_{ti}}{F(X_t\beta)} - \frac{(1 - y_t)f(X_t\beta)x_{ti}}{1 - F(X_t\beta)}. \tag{2}$$

2. Next, we need to show that the expectation of G_{ti} is zero when it is evaluated at the true β.

   Notice, that the only source of randomness in (2) is y_t, the expected value of which is equal to $F(X_t\beta)$ (under the DGP characterized by β). Hence by linearity and conditioning on Ω_t, we obtain

$$\begin{aligned}
\mathbb{E}\left(G_{ti}(\beta)\right) &= \frac{\mathbb{E}\left(y_t\right)f(X_t\beta)x_{ti}}{F(X_t\beta)} - \frac{(1 - \mathbb{E}\left(y_t\right))f(X_t\beta)x_{ti}}{1 - F(X_t\beta)} \\
&= \frac{F(X_t\beta)f(X_t\beta)x_{ti}}{F(X_t\beta)} - \frac{(1 - F(X_t\beta))f(X_t\beta)x_{ti}}{1 - F(X_t\beta)} \\
&= \frac{F(X_t\beta)f(X_t\beta)x_{ti}}{F(X_t\beta)} - \frac{(1 - F(X_t\beta))f(X_t\beta)x_{ti}}{1 - F(X_t\beta)} \\
&= f(X_t\beta)x_{ti} - f(X_t\beta)x_{ti} \\
&= 0,
\end{aligned}$$

   which is the required result.

[1] Recall: it says that minus the expectation of the asymptotic Hessian is equal to the information matrix.

3. Then we will use the fact that the asymptotic information matrix is equal to

$$\mathcal{I}(\beta)_= \lim_{n\to\infty} \sum_{t=1}^{n} \mathbb{E}\,(G_{ti}G_{tj}),$$

to obtain its typical.

Recall, that the **asymptotic information matrix** is a plim of the information matrix defined as

$$I(\beta) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\,\left(G_t^T(\beta)G_t(\beta)\right).$$

To find its typical element, let us start with calculating

$$G_{ti}(\beta)G_{tj}(\beta) = \frac{y_t^2 f^2(X_t\beta)x_{ti}x_{tj}}{F^2(X_t\beta)} + \frac{(1-y_t)^2 f^2(X_t\beta)x_{ti}x_{tj}}{(1-F(X_t\beta))^2} - \frac{y_t(1-y_t)f^2(X_t\beta)x_{ti}x_{tj}}{F(X_t\beta)(1-F(X_t\beta))}.$$

The expected value of the above expression is

$$\begin{aligned}\mathbb{E}\,(G_{ti}(\beta)G_{tj}(\beta)) &= \frac{f^2(X_t\beta)x_{ti}x_{tj}}{F(X_t\beta)} + \frac{f^2(X_t\beta)x_{ti}x_{tj}}{1-F(X_t\beta)}\\ &= \frac{f^2(X_t\beta)x_{ti}x_{tj}}{F(X_t\beta)(1-F(X_t\beta))},\end{aligned}$$

since

$$\begin{aligned}\mathbb{E}\,\left(y_t^2\right) &= F(X_t\beta),\\ \mathbb{E}\,\left((1-y_t)^2\right) &= 1-F(X_t\beta),\\ \mathbb{E}\,(y_t(1-y_t)) &= 0,\end{aligned}$$

where the last equality arises from y_t being either 0 or 1. Therefore, we arrive at

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\,(G_{ti}(\beta)G_{tj}(\beta)) = \frac{1}{n}\sum_{t=1}^{n}\frac{f^2(X_t\beta)x_{ti}x_{tj}}{F(X_t\beta)(1-F(X_t\beta))}, \tag{3}$$

which in the plim yields the typical element of the asymptotic information matrix $\mathcal{I}(\beta)$.

4. Finally, we need to show that the **asymptotic covariance matrix** (11.15) is equal to the inverse of this asymptotic information matrix.

   In the matrix notation, $\mathcal{I}(\beta)$ can be expressed as

$$\begin{aligned}\mathcal{I}(\beta) &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\,\left(G_t^T(\beta)G_t(\beta)\right)\\ &= \underbrace{\frac{1}{n}\sum_{t=1}^{n} X^T\Upsilon(\beta)X\,.}_{(**)}\end{aligned}$$

   As $(**)$ is indeed the inverse of (11.15), the proof has been completed.

DM, 11.8

1. First, we need to calculate the **Hessian matrix** corresponding to the loglikelihood function (11.09).

   Recall that the ij-th element of the Hessian matrix $H_{ij}(\beta)$ corresponding to the loglikelihood function (11.09) is the derivative with respect to β_j of the typical element of the gradient $g_i(\beta)$ given by

$$g_i(\beta) = \frac{\partial l(\beta)}{\partial \beta_i} = \sum_{t=1}^{n} \frac{(y_t - F(X_t\beta))f(X_t\beta)X_{ti}}{F(X_t\beta)(1 - F(X_t\beta))}, \qquad i = 1, \dots, k.$$

   Taking derivative of each element in the above sum results, for each t, in the ij-th element of the Hessian matrix being a fraction with the denominator equal to

$$F^2(X_t\beta)(1 - F(X_t\beta))^2, \tag{4}$$

   and the numerator given by

$$\begin{aligned} &- f^2(X_t\beta)x_{ti}x_{tj}F(X_t\beta)(1 - F(X_t\beta)) && (5)\\ &+ (y_t - F(X_t\beta))f'(X_t\beta)x_{ti}x_{tj}F(X_t\beta)(1 - F(X_t\beta)) && (6)\\ &- (y_t - F(X_t\beta))f^2(X_t\beta)x_{ti}x_{tj}(1 - F(X_t\beta)) && (7)\\ &+ (y_t - F(X_t\beta))f^2(X_t\beta)x_{ti}x_{tj}F(X_t\beta). && (8) \end{aligned}$$

2. Then, we will use the fact that minus the expectation of the asymptotic Hessian is equal to the asymptotic information matrix to obtain the same result for the latter that you obtained in the previous exercise.

   Conditional on X_t, (4) and (5) are nonstochastic, while (6)–(8) have expectation 0, since $\mathbb{E}\,(y_t - F(X_t\beta)) = 0$. Therefore,

$$\mathbb{E}\left(n^{-1}H_{ij}(\beta)\right) = -\frac{1}{n}\sum_{t=1}^{n} \frac{f^2(X_t\beta)x_{ti}x_{tj}}{F(X_t\beta)(1 - F(X_t\beta))},$$

   which is the opposite of (3), the ij-th element of the information matrix. Since the asymptotic Hessian is equal to minus the asymptotic information matrix, we have shown once again that the asymptotic covariance matrix is given by the RHS of (11.15), which is the desired result.