Answer key to problem set # 3

ECON 342

Prof. Shakeeb Khan
T.A. J. Marcelo Ochoa

Spring, 2009

Problem (Hayashi Chapter 7, page 469 #4)**.** *Solution.* What needs to be proved is: If $\mathbb{E}\left(x_t x_t'\right)$ is nonsingular, then $\Phi(x_t'\theta) \neq \Phi(x_t'\theta_0)$ for $\theta \neq \theta_0$. Since the CDF $\Phi(\cdot)$ is strictly increasing, we only need to show that $x_t'\theta \neq x_t'\theta_0$ for $\theta \neq \theta_0$. See the argument on Example 7.8 in the book. ♢

Problem (Hayashi Chapter 7, page 501 #1)**.** *Solution.* **(a)** We have,

$$a(\mathbf{w}) \neq 1 \leftrightarrow f(y|x;\theta) \neq f(y|x;\theta_0)$$

and by hypothesis $\mathbb{P}\left(f(y|x;\theta) \neq f(y|x;\theta_0)\right) > 0$ then,

$$\mathbb{P}\left(a(\mathbf{w}) \neq 1\right) = \mathbb{P}\left(f(y|x;\theta) \neq f(y|x;\theta_0)\right) > 0$$

Thus, $a(\mathbf{w})$ is a nonconstant random variable.

(b) Let $c[a(\mathbf{w})] = \ln a(\mathbf{w})$, then using Jensen's inequality we have,

$$\mathbb{E}\left(\ln a(\mathbf{w})\right) < \ln \mathbb{E}\left(a(\mathbf{w})\right)$$

as we wanted to show.

(c)

$$\begin{aligned}\mathbb{E}\left(a(\mathbf{w})|x\right) &= \int a(\mathbf{w}) f(y|x,\theta_0)dy \\ &= \int \frac{f(y|x,\theta)}{f(y|x,\theta_0)} f(y|x,\theta_0)dy \\ &= \int f(y|x,\theta)dy \\ &=1\end{aligned}$$

then, using the law of iterated expectations we have,

$$\mathbb{E}\left(\mathbb{E}\left(a(\mathbf{w})|x\right)\right) = \mathbb{E}\left(a(\mathbf{w})\right) = 1$$

as we wanted to show.

(d) Using the previous results we have,

$$\mathbb{E}\left(\ln a(\mathbf{w})\right) < \ln \mathbb{E}\left(a(\mathbf{w})\right) = \ln 1 = 0$$

then, by definition:

$$\mathbb{E}\left(\ln a(\mathbf{w})\right) = \mathbb{E}\left(f(y|x,\theta) - f(y|x,\theta_0)\right) < 0$$

then,

$$\mathbb{E}\left(f(y|x,\theta)\right) = \mathbb{E}\left(f(y|x,\theta_0)\right)$$

as we wanted to show.

◊

Problem 1. Here we will use Matlab to use the NLS Probit estimator using the Gauss Newton Method. (See the answer to question 2 on page 500 of Hayashi to see how.)

a) Download the Excel `hitstat.xls` from the class web page: the columns correspond to at bats, home runs, rbis, stolen bases and price of the player. First construct a dummy left hand side variable which is 1 if the price exceeds 15 and 0 otherwise. Do NLS Probit of this dummy on an intercept, homeruns, rbis and stolen bases. Construct a 95% confidence interval for each of the coefficients.

Solution. Let

$$p_i = \mathbb{P}\left(y_i = 1|\mathbf{x}_i\right)$$

i.e., the probability that Y_i is one, and $1 - p_i$ the probability that Y_i is 0, then

$$Y_i \sim \text{ Bernoulli } (p_i)$$

therefore

$$\mathbb{E}\left(Y_i|\mathbf{x}_i\right) = p_i$$

and

$$\mathrm{V}\left(Y_i|\mathbf{x}_i\right) = p_i(1 - p_i).$$

We can represent the probability that $y_i = 1$ as,

$$p_i = \mathbb{P}\left(\mathbf{x}_i'\beta + \varepsilon_i > 0\right) = \mathbb{P}\left(\varepsilon_i > -\mathbf{x}_i\beta\right)$$

where $\mathbf{x}_i = (1, x_i)$ and $\beta = (\beta_1, \beta_2)'$.

Specifying a functional form for the CDF of ε_i say $F(\varepsilon_i)$ we have,

$$p_i = 1 - F(-\mathbf{x}_i'\beta)$$

with a PDF symmetric around zero we can re-write this results as,

$$p_i = F(\mathbf{x}_i'\beta)$$

In our case we will assume that,

$$F(\alpha) = \Phi(\alpha)$$

i.e., a standard normal cdf. You can show that $\Phi(\cdot)$ is symmetric and has first-derivative equal to,

$$\Phi'(\alpha) = -\alpha\phi(\alpha)$$

In this problem we will obtain our estimates solving the following optimization problem:

$$\min_{\beta} -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i - \Phi(\mathbf{x}_i'\beta)\Big)^2$$

Let

$$Q(\beta) = \frac{1}{N}\sum_{i=1}^{N} m(\mathbf{w}_t, \beta)$$

be our objective function. To solve this problem we will use the following algorithm:

Initialization Choose a guess β_0 and stopping parameters δ and $\epsilon > 0$

Step 1: Compute,

$$\nabla Q(\beta^m) = -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i - \Phi(\mathbf{x}_i'\beta^m)\Big)\phi(\mathbf{x}_i'\beta^m)\mathbf{x}_i$$

and,

$$H(\beta) = \frac{1}{N}\sum_{i=1}^{N}\phi(\mathbf{x}_i'\beta^m)^2\mathbf{x}_i\mathbf{x}_i'$$

Step 2: Compute,

$$\beta^{m+1} = \beta^m + [H(\beta^m)]^{-1}\nabla Q(\beta^m)$$

Step 3: If

$$\frac{\|\beta^m - \beta^{m+1}\|}{1 + \|\beta^m\|} < \epsilon$$

go to step 4; else go to step 1.

Step 4: If

$$\frac{\|\nabla f(\beta^{m+1})\|}{1 + \|f(\beta^m)\|} < \delta$$

STOP and report success; else STOP and report convergence to a non-optimal point.

	Coefficient	Std.Dev.	95% Conf. Int.
$\hat{\beta}_1$	-10.0277	2.674	[-15.269,-4.786]
$\hat{\beta}_1$	0.0766	0.107	[-0.134,0.287]
$\hat{\beta}_1$	0.0909	0.049	[-0.006,0.188]
$\hat{\beta}_1$	0.2099	0.060	[0.093,0.327]

Table 1: Estimation results of the model $\mathrm{y}_i = \Phi(\mathbf{x}_i'\beta) + \varepsilon_i$with $\varepsilon_i \sim N(0,1)$ using NLS.

To calculate standard errors remember that the asymptotic distribution of this M-estimator is,

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N\Big(\mathbf{0}, \mathbb{E}\left(H(\mathbf{w};\beta)\right)^{-1} \Sigma \mathbb{E}\left(H(\mathbf{w};\beta)\right)^{-1}\Big)$$

Denote,

$$s(\mathbf{w}_i, \beta) = 2\Big(y_i - \Phi(\mathbf{x}_i'\beta)\Big)\phi(\mathbf{x}_i'\beta)\mathbf{x}_i$$

and,

$$H(\mathbf{w}_i, \beta) = \frac{\partial s(\mathbf{w}_i, \beta)}{\partial \beta'} = -2\phi(\mathbf{x}_i'\beta)^2\mathbf{x}_i\mathbf{x}_i' - 2\Big(y_i - \Phi(\mathbf{x}_i'\beta)\Big)\phi(\mathbf{x}_i'\beta)\mathbf{x}_i\mathbf{x}_i'$$

then, an estimate of the asymptotic variance is,

$$\hat{V} = \left[\frac{1}{N}\sum_{1}^{N} H(\mathbf{w}_i;\beta)\right]^{-1} \left[\frac{1}{N}\sum_{1}^{N} s(\mathbf{w}_i,\beta)s(\mathbf{w}_i,\beta)'\right] \left[\frac{1}{N}\sum_{1}^{N} H(\mathbf{w}_i;\beta)\right]^{-1}$$

and we obtain the variance for the coefficients from the diagonal elements of $\hat{V}/T$. The estimation results are in Table 1.

$\Diamond$