

# Econ 203B Econometrics

Raphael Lam

## 1 Problem Set 5 Solutions<sup>1</sup>

Hayashi Chp. 7 Question 3:

$$\log f(y_t|x_t; \beta, \sigma^2) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{(y_t - x_t'\beta)^2}{2\sigma^2}$$

Let  $(\hat{\beta}, \hat{\sigma}^2)$  be the unrestricted MLE and  $(\tilde{\beta}, \tilde{\sigma}^2)$  be the restricted MLE with restriction  $R\beta = c$ . Also let

$$\hat{\Sigma} = \begin{bmatrix} \frac{1}{\hat{\sigma}^2} \frac{1}{n} \sum_{t=1}^n x_t x_t' & 0 \\ 0 & \frac{1}{2(\hat{\sigma}^2)^2} \end{bmatrix}, \tilde{\Sigma} = \begin{bmatrix} \frac{1}{\tilde{\sigma}^2} \frac{1}{n} \sum_{t=1}^n x_t x_t' & 0 \\ 0 & \frac{1}{2(\tilde{\sigma}^2)^2} \end{bmatrix}$$

Solution:

First we must derive the MLE of  $\beta$ . The joint log-likelihood is just the sum of the log distributions.

$$\begin{aligned} \mathcal{L}(\beta, y_t, x_t) &= \sum \log f(y_t|x_t) \\ &= \sum_{t=1}^n -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{(y_t - x_t'\beta)^2}{2\sigma^2} \end{aligned}$$

The FOC for a maximum are

$$\begin{aligned} \frac{\partial \mathcal{L}(\beta, y_t, x_t)}{\partial \beta} &= \sum_{t=1}^n x_t \frac{(y_t - x_t'\hat{\beta})}{\hat{\sigma}^2} = 0 \\ \frac{\partial \mathcal{L}(\beta, y_t, x_t)}{\partial \sigma^2} &= -\frac{1}{2} \frac{n}{\hat{\sigma}^2} + \sum_{t=1}^n \frac{2(y_t - x_t'\hat{\beta})^2}{2\hat{\sigma}^2} = 0 \end{aligned}$$

Solving for  $\beta$  and  $\sigma^2$  :

---

<sup>1</sup>This solution is prepared using in parts the solutions of A. Hollenhorst, D. Miyakawa, and A. Mezza.

$$\hat{\beta}_{MLE} = \left( \sum_{t=1}^n x_t x_t' \right)^{-1} \sum_{t=1}^n x_t y_t \quad (1)$$

$$\hat{\sigma}_{MLE}^2 = \frac{1}{n} \sum_{t=1}^n \left( y_t - x_t' \hat{\beta} \right)^2 \quad (2)$$

**Remark 1** In the Hayashi exercise, the  $x_t$  denotes a column vector of dimension  $1 \times k$ , instead of a row vector used in the lecture notes or previous exercise.

Note that the MLE estimator is the same as OLS estimator on minimizing the sum of the residuals squared.

$$\hat{\beta}_{OLS} \in \arg \min_{\beta} \sum (y_t - x_t' \beta)^2 \quad (3)$$

Taking the first order condition on above and derive the OLS estimator, we can see that  $\hat{\beta}_{MLE} = \hat{\beta}_{OLS}$ .

Now we solve for the MLE in the restricted model where  $R\beta = c$ . The constrained MLE problem can be set up as a lagrangian as follows.

$$\tilde{\mathcal{L}}(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \sum_{t=1}^n \frac{(y_t - x_t' \beta)^2}{2\sigma^2} + \lambda' (R\beta - c)$$

Substitute our estimate on  $\tilde{\sigma}^2 = \frac{1}{n} \sum (y_t - x_t' \beta)^2$  to the log likelihood function:

$$\begin{aligned} \tilde{\mathcal{L}}(\beta) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \left( \frac{1}{n} \sum (y_t - x_t' \beta)^2 \right) - \sum_{t=1}^n \frac{(y_t - x_t' \beta)^2}{2 \frac{1}{n} \sum (y_t - x_t' \beta)^2} + \lambda' (R\beta - c) \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \left( \frac{1}{n} \sum (y_t - x_t' \beta)^2 \right) - \frac{2}{n} + \lambda' (R\beta - c) \end{aligned}$$

As the function  $\frac{n}{2} \log \left( \frac{1}{n} \sum (y_t - x_t' \beta)^2 \right)$  is monotonic transformation of the sum of residuals squared, maximization w.r.t.  $\beta$  on the above function is equivalent as minimizing the sum of residuals squared subject to the constraint. (Notice the negative sign before the term - maximizing  $-f(a)$  w.r.t.  $a$  is the same as minimizing  $f(a)$  w.r.t.  $a$ ).

Taking the first order condition on the restricted case and we can see that  $\tilde{\sigma}^2$  is similar to before

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (y_t - x_t' \tilde{\beta})^2$$

As above we see that these FOC, and in fact the original maximization problem is equivalent to minimizing the sum of squared error subject to the constraint. Thus we again will have that  $\tilde{\beta}$  is equivalent to the restricted OLS estimator.

b)

Using the above result, and define

$$Q_n(\theta) = \frac{1}{n} \sum \log f(y_t | x_t; \beta, \sigma^2) \quad (4)$$

where  $\theta$  is the parameter vector containing  $\beta$  and  $\sigma^2$ .

$$\begin{aligned} Q_n(\hat{\theta}) &= \frac{1}{n} \sum_{t=1}^n \left[ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\hat{\sigma}^2) - \frac{(y_t - x_t' \hat{\beta})^2}{2\hat{\sigma}^2} \right] \\ &= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\hat{\sigma}^2) - \frac{1}{n} \sum_{t=1}^n \frac{(y_t - x_t' \hat{\beta})^2}{2\hat{\sigma}^2} \\ &= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \left( \frac{1}{n} \sum_{t=1}^n (y_t - x_t' \hat{\beta})^2 \right) - \frac{1}{2} \\ &= -\frac{1}{2} \log(2\pi) - \frac{1}{2} - \frac{1}{2} \log \left( \frac{SSR_U}{n} \right) \end{aligned}$$

Using the result from part a), we can see that

$$Q_n(\tilde{\theta}) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} - \frac{1}{2} \log \left( \frac{SSR_R}{n} \right)$$

c)

First we derive the Hessian matrix as follows:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \beta} &= x_t \frac{(y_t - x_t' \beta)}{\sigma^2} \\ \frac{\partial \mathcal{L}}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} + \frac{(y_t - x_t' \beta)^2}{2(\sigma^2)^2} \end{aligned}$$

So the score function  $S(w_t; \theta) = \left[ x_t \frac{(y_t - x_t' \beta)}{\sigma^2} - \frac{1}{2\sigma^2} + \frac{(y_t - x_t' \beta)^2}{2(\sigma^2)^2} \right]$

Taking second derivatives for the Hessian matrix.

$$H(w_t; \theta) = \begin{bmatrix} -\frac{1}{\sigma^2} x_t x_t' & -\frac{1}{(\sigma^2)^2} x_t (y_t - x_t' \beta)' \\ -\frac{1}{(\sigma^2)^2} x_t' (y_t - x_t' \beta) & \frac{1}{2(\sigma^2)^2} - \frac{(y_t - x_t' \beta)^2}{4(\sigma^2)^3} \end{bmatrix}$$

Taking the expectation, using the LIE, and notice that we are interested in the negative expectation evaluated at the true parameters:

$$-E_0 [H(w_t; \theta_0)] = \begin{bmatrix} \frac{1}{\sigma_0^2} E [x_t x_t'] & 0 \\ 0' & \frac{1}{2(\sigma_0^2)^2} \end{bmatrix}$$

To show that  $\widehat{\Sigma} \xrightarrow{p} -E [H(w_t; \theta_0)]$ , we have to show that  $\widehat{\Sigma}_{11} = \frac{1}{\widehat{\sigma}^2} \frac{1}{n} \sum [x_t x_t']$  (with dimension of  $k \times k$ )  $\rightarrow_p \frac{1}{\sigma_0^2} E [x_t x_t']$ , and the second term  $\widehat{\Sigma}_{22} = \frac{1}{2(\widehat{\sigma}^2)^2} \rightarrow_p \frac{1}{2(\sigma_0^2)^2}$

The proof, by now, should be familiar to you. If not, take a look on previous assignments. The idea is that using the LLN (appropriate version), we can show that  $\widehat{\sigma}^2$  is a consistent estimator for  $\sigma^2$ , which implies any continuous function of  $\widehat{\sigma}^2$  is consistent, specifically  $(\widehat{\sigma}^2)^2 \xrightarrow{p} (\sigma^2)^2$ . Slutsky Theorem is useful in this context. We also have to assume certain conditions on the expectation of  $E(x_t x_t')$  to be finite. Combining these, we can see that

$$\widehat{\Sigma} \xrightarrow{p} \begin{bmatrix} \frac{1}{\widehat{\sigma}^2} E [x_t x_t'] & 0 \\ 0' & \frac{1}{2(\widehat{\sigma}^2)^2} \end{bmatrix} = -E_0 [H(w_t; \theta_0)]$$

Likewise, we have to show that  $\widetilde{\Sigma} \xrightarrow{p} -E_0 H(w_t; \theta_0)$  by relying on the proof that

$$1) \quad \widetilde{\Sigma}_{11} = \frac{1}{\widetilde{\sigma}^2} \frac{1}{n} \sum [x_t x_t'] \rightarrow_p \frac{1}{\sigma_0^2} E [x_t x_t'] \quad (5)$$

$$2) \quad \widetilde{\Sigma}_{22} = \frac{1}{2(\widetilde{\sigma}^2)^2} \rightarrow_p \frac{1}{2(\sigma_0^2)^2} \quad (6)$$

$$3) \quad \widetilde{\varepsilon}' \widetilde{\varepsilon} = \widehat{\varepsilon}' \widehat{\varepsilon} + (R\widehat{\beta} - c)' [R(X'X)^{-1}R']^{-1} (R\widehat{\beta} - c) \quad (7)$$

where the third condition can be shown using the notes that I gave you earlier on the restricted least square. From above and previous

homework assignments, we have verified that  $\hat{\varepsilon}'\hat{\varepsilon} \rightarrow^p \sigma^2$ . But here we are verifying that  $\tilde{\varepsilon}'\tilde{\varepsilon} \rightarrow^p \sigma^2$  as well. The idea is simple if we expand out the second term in the third condition. that  $R\beta = c$

$$\begin{aligned} R\hat{\beta} - c &= R\beta - c - R(X'X)^{-1}X'\varepsilon \\ &= -R(X'X)^{-1}X'\varepsilon = -R\left(\frac{1}{n}\sum x_t x_t'\right)\left(\frac{1}{n}\sum x_t \varepsilon_t\right) \quad (8) \end{aligned}$$

Assume  $E(x_t x_t')^{-1}$  is finite and using the LLN such that  $(\frac{1}{n}\sum x_t \varepsilon_t) \rightarrow^p E(x_t \varepsilon_t) = 0$ ,

$$R\hat{\beta} - c \rightarrow^p 0$$

Using the Slutsky Theorem and continuity theorem that  $R(\frac{1}{n}\sum(x_t x_t')^{-1}R' \rightarrow^p R[E(x_t x_t')^{-1}]R'$ , we can see that

$$\tilde{\varepsilon}'\tilde{\varepsilon} = \frac{1}{n}\sum \tilde{\varepsilon}_t^2 \rightarrow^p \sigma^2$$

After the detour, we go back to see that the first two conditions:

$$\begin{aligned} \tilde{\Sigma}_{11} &= \frac{1}{\tilde{\sigma}^2} \frac{1}{n} \sum [x_t x_t'] \rightarrow^p \frac{1}{\sigma_0^2} E[x_t x_t'] \\ \tilde{\Sigma}_{22} &= \frac{1}{2(\tilde{\sigma}^2)^2} \rightarrow^p \frac{1}{2(\sigma_0^2)^2} \end{aligned}$$

So,  $\tilde{\Sigma} \rightarrow^p -E[H(w_t; \theta_0)]$ .

d)

In Hayashi's notation we have.

$$\begin{aligned} a(\theta) &= R\beta - c \\ A(\theta) &= \frac{\partial a(\theta)}{\partial \theta'} \Big|_{\theta_0} = [R \quad 0] \end{aligned}$$

Also, from our results in part a

$$\frac{\partial Q_n(\tilde{\theta})}{\partial \theta} = \begin{bmatrix} x_t \frac{(y_t - x_t' \tilde{\beta})}{\tilde{\sigma}^2} \\ -\frac{1}{2\tilde{\sigma}^2} + \frac{(y_t - x_t' \tilde{\beta})^2}{2(\tilde{\sigma}^2)^2} \end{bmatrix} = \frac{1}{SSR_R} \begin{bmatrix} X'(y - X\tilde{\beta}) \\ 0 \end{bmatrix}$$

Substituting this with the expressions for  $Q_n(\tilde{\theta})$  and  $Q_n(\hat{\theta})$  into the test statistics gives.

$$\begin{aligned}
W &= n(R\hat{\beta} - c)' \left\{ [R \ 0] \hat{\Sigma}^{-1} \begin{bmatrix} R \\ 0 \end{bmatrix} \right\}^{-1} (R\hat{\beta} - c) \\
&= n(R\hat{\beta} - c)' \left\{ [R \ 0] \begin{bmatrix} \hat{\sigma}^2 n(X'X)^{-1} & 0 \\ 0 & 2\hat{\sigma}^4 \end{bmatrix}^{-1} \begin{bmatrix} R \\ 0 \end{bmatrix} \right\}^{-1} (R\hat{\beta} - c) \\
&= n(R\hat{\beta} - c)' \left\{ [R \ 0] [SSR_U] \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & -2\hat{\sigma}^2/n \end{bmatrix} \begin{bmatrix} R \\ 0 \end{bmatrix} \right\}^{-1} (R\hat{\beta} - c) \\
&= n \frac{(R\hat{\beta} - c)' [R(X'X)^{-1}R']^{-1} (R\hat{\beta} - c)}{SSR_U}
\end{aligned}$$

$$LM : n \left( \frac{\partial Q_n(\tilde{\theta})}{\partial \theta} \right)' \tilde{\Sigma}^{-1} \left( \frac{\partial Q_n(\tilde{\theta})}{\partial \theta} \right)$$

$$\begin{aligned}
\text{where } Q_n(\tilde{\theta}) &= \frac{1}{n} \sum \log f(y_t | x_t; \theta) \\
&= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 - \frac{1}{n} \sum \frac{(y_t - x_t' \beta)^2}{2\sigma^2}
\end{aligned}$$

The first derivatives

$$\begin{aligned}
\frac{\partial Q_n(\theta)}{\partial \theta} \Big|_{\tilde{\theta}} &= \begin{bmatrix} \frac{1}{n\tilde{\sigma}^2} \sum x_t (y_t - x_t \tilde{\beta}) \\ -\frac{1}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4} \frac{1}{n} \sum (y_t - x_t \tilde{\beta})^2 \end{bmatrix} \\
&= \frac{1}{SSR_R} \begin{bmatrix} X(Y - X'\tilde{\beta}) \\ -\frac{n}{2} + \frac{1}{2\tilde{\sigma}^2} (Y - X'\tilde{\beta})'(Y - X\tilde{\beta}) \end{bmatrix} = \frac{1}{SSR_R} \begin{bmatrix} X(Y - X'\tilde{\beta}) \\ 0 \end{bmatrix} \\
\tilde{\Sigma}^{-1} &= SSR_R \begin{bmatrix} (X'X)^{-1} & \\ & -\frac{2\tilde{\sigma}^2}{n} \end{bmatrix}
\end{aligned}$$

Therefore, the LM statistics based on the estimation on restricted case:

$$\begin{aligned}
LM &= n \frac{1}{SSR_R} \begin{bmatrix} X(Y - X'\tilde{\beta}) \\ 0 \end{bmatrix}' \left( SSR_R \begin{bmatrix} (X'X)^{-1} & \\ & -\frac{2\tilde{\sigma}^2}{n} \end{bmatrix} \right) \frac{1}{SSR_R} \begin{bmatrix} X(Y - X'\tilde{\beta}) \\ 0 \end{bmatrix} \\
&= n \frac{(y - X\tilde{\beta})' X(X'X)^{-1} X' (y - X\tilde{\beta})}{SSR_R}
\end{aligned}$$

The last case on LR test statistics:

$$LR = 2n \left[ Q_n(\hat{\theta}) - Q_n(\tilde{\theta}) \right]$$

$$\text{where } Q_n(\hat{\theta}) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} - \frac{1}{2} \log \left( \frac{SSR_U}{n} \right)$$

$$Q_n(\tilde{\theta}) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} - \frac{1}{2} \log \left( \frac{SSR_R}{n} \right)$$

It is easy to see that

$$LR = n \left\{ \log \left( \frac{SSR_R}{n} \right) - \log \left( \frac{SSR_U}{n} \right) \right\}$$

e)

To write these statistics in terms of the squared residuals by noting that

$$\begin{aligned} SSR_R - SSR_U &= \tilde{\varepsilon}'\tilde{\varepsilon} - \hat{\varepsilon}'\hat{\varepsilon} \\ &= (R\hat{\beta} - c)' [R(X'X)^{-1}R']^{-1} (R\hat{\beta} - c) \\ &= (y - X\tilde{\beta})'X(X'X)^{-1}X'(y - X\tilde{\beta}) \end{aligned}$$

From the above identities the statistics trivially may be written as follows.

$$W = n \frac{SSR_R - SSR_U}{SSR_U}$$

$$LM = n \frac{SSR_R - SSR_U}{SSR_R}$$

$$LR = n \left[ \log \left( \frac{SSR_R}{n} \right) - \log \left( \frac{SSR_U}{n} \right) \right] = n \log \left( \frac{SSR_R}{SSR_U} \right)$$

f)

Show that  $W \geq LR \geq LM$  by noting that  $SSR_U \leq SSR_R$ . Check the notes that I posted on the webpage on non-linear hypotheses testing, it is easy then to see that  $W \geq LR$ . Now, employing the mean value theorem we have for some  $\xi \in \left( \frac{SSR_U}{SSR_R}, 1 \right) < 1$ .

$$\begin{aligned} LR &= n \log \left( \frac{SSR_R}{SSR_U} \right) = n \left( 1 - \frac{1}{\xi} \frac{SSR_U}{SSR_R} \right) \\ &= n \frac{SSR_R - \frac{1}{\xi} SSR_U}{SSR_R} \\ &\leq n \frac{SSR_R - SSR_U}{SSR_R} = LM \end{aligned}$$

Thus in summary we have

$$W \geq LM \geq LR$$

**Problem 1**

Provide the primitive conditions for consistency and asymptotic normality of the NLS estimator of  $\beta_0$  in the logit model:

$$\Pr(y_i = 1|X_i) = \frac{\exp(X_i\beta_0)}{1 + \exp(X_i\beta_0)} \quad (9)$$

and derive the asymptotic properties.

Answer:

Define the conditional expectation:

$$E(y_i|X_i) = \Pr(y_i = 1|X_i) = \frac{\exp(X_i\beta_0)}{1 + \exp(X_i\beta_0)} \equiv \Lambda(X_i\beta_0) \quad (10)$$

Thus the NLS estimator here is

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \left( Y_i - \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)} \right)^2$$

Let  $\varepsilon_i = y_i - \Lambda(X_i\beta_0)$

**Condition 2** Assume the followings

- 1)  $y_i = \Lambda(X_i\beta_0) + \varepsilon_i$  for all  $i = 1 \dots n$  (satisfied by construction)
- 2)  $\{(X_i, \varepsilon_i)\}$  follows i.i.d
- 3)  $E(\varepsilon_i|X_i) = 0$
- 4)  $\Lambda(X_i\beta_0) \neq \Lambda(X_i\beta_0) \forall \beta \neq \beta_0$  (Identification)
- 5)  $\mathcal{B}$  is a compact set of  $\mathcal{R}^K$
- 6)  $\Lambda(X_i\beta_0)$  is continuous in  $\beta$  for any  $X_i$  and is twice differentiable (satisfied in the logistic distribution)
- 7) Uniform convergence:  $\frac{1}{n} \sum (y_i - \Lambda(X_i\beta_0))^2 \rightarrow E [y_i - \Lambda(X_i\beta_0)]^2$  converges in probability uniformly in  $\beta$ .
- 8)  $\beta$  is in the interior of the parameter set  $\mathcal{B}$

$\hat{\beta}_{NLS}$  exists under above assumptions and has a probability convergence to the true parameter  $\beta_0$ .

For the identification condition, it can be seen that

$$\begin{aligned} & \text{suppose } \Lambda(X_i\beta_0) = \Lambda(X_i\beta) \text{ and } \beta_0 \neq \beta \\ & \frac{\exp(X_i\beta_0)}{1 + \exp(X_i\beta_0)} = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)} \\ & X_i\beta_0 - \log(1 + \exp(X_i\beta_0)) = X_i\beta - \log(1 + \exp(X_i\beta)) \end{aligned}$$

If  $X_i$  is FCR, then the above equality implies that  $\beta_0 = \beta$ , which contradicts that  $\beta_0 \neq \beta$ . The identification under logit case is satisfied.

Let  $m(X_i, y_i; \beta) = \frac{1}{2}(y_i - \Lambda(X_i\beta))^2$ , then we can write down the score function and the hessian in the following:

$$\begin{aligned} S(X_i, y_i; \beta) &= \frac{\partial m(X_i, y_i; \beta)}{\partial \beta} = -(y_i - \Lambda(X_i\beta)) \Lambda'(X_i\beta) X_i' \\ H(X_i, y_i; \beta) &= \frac{\partial^2 m(X_i, y_i; \beta)}{\partial \beta \partial \beta'} = \Lambda'(X_i\beta)^2 X_i' X_i - (y_i - \Lambda(X_i\beta)) \Lambda''(X_i\beta) X_i' X_i \end{aligned}$$

$$\begin{aligned} S(Y_i, X_i, \beta) &= -\frac{\partial \Lambda(X_i\beta)}{\partial \beta} (Y_i - \Lambda(X_i\beta)) \\ \Omega &= s(Y_i, X_i, \beta_0)' s(Y_i, X_i, \beta_0) = \left. \frac{\partial \Lambda(X_i\beta)}{\partial \beta} \right|_{\beta_0} \left. \frac{\partial \Lambda(X_i\beta)}{\partial \beta'} \right|_{\beta_0} (Y_i - \Lambda(X_i\beta_0))^2 \\ H(Y_i, X_i, \beta) &= \frac{\partial s(Y_i, X_i, \beta)}{\partial \beta'} \end{aligned}$$

Further assume that

**Condition 3** 9)  $\Omega_0 = E_0 \{S(X_i, y_i; \beta_0)S(X_i, y_i; \beta_0)'\}$  is finite and positive definite

10)  $E\{H(X_i, y_i; \beta_0)\}$  is non-singular. For the neighbourhood around  $\beta_0$ , the Hessian evaluated at  $\beta$  satisfies the finite property such that  $E\{\sup \|H(X_i, y_i; \beta)\|\} < \infty$

We can write

$$\begin{aligned} H_0 &= E \{ \Lambda(X_i\beta_0)^2 (1 - \Lambda(X_i\beta_0))^2 X_i' X_i \} - E \{ E[\varepsilon_i g(X_i) | X_i] \} \\ &= E \{ \Lambda(X_i\beta_0)^2 (1 - \Lambda(X_i\beta_0))^2 X_i' X_i \} \end{aligned}$$

The asymptotic properties of the  $\hat{\beta}$  can be shown that (see lecture notes for derivation using the mean-value theorem)

$$\sqrt{n}(\hat{\beta}_{NLS} - \beta_0) \rightarrow^d N(0, H_0^{-1}\Omega_0H_0^{-1}) \quad (12)$$

where  $H_0$  is the Hessian and  $\Omega_0$  is the square of the score functions, both evaluated and integrated at the true parameter  $\beta_0$ .

**Remark 4** For a detailed proof of the asymptotic distribution on the logit case, please refer to the Binary choice handouts that I posted. There derives the scores and hessian matrices and use the mean-value theorem to show the asymptotic properties.

## Problem 2

a)

Given that  $X \sim N(0, \sigma^2)$ , we are going to find out the conditional expectation and variance. Assume  $u \sim N(0, 1)$ , then  $X = \sigma u \sim N(0, \sigma^2)$

The density function can be stated as:

$$f(X|b > X > a) = \frac{f(x)}{\Pr(b > X > a)} = \frac{\frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right)}{\Phi\left(\frac{b}{\sigma}\right) - \Phi\left(\frac{a}{\sigma}\right)}$$

$$\begin{aligned} E(X|a < X < b) &= E(\sigma u|a < \sigma u < b) \\ &= \sigma E\left(u \mid \frac{a}{\sigma} < u < \frac{b}{\sigma}\right) \\ &= \sigma \frac{\int_{a/\sigma}^{b/\sigma} u \phi(u) du}{\Phi(b/\sigma) - \Phi(a/\sigma)} = \frac{\sigma}{\Phi(b/\sigma) - \Phi(a/\sigma)} \left[ \frac{1}{\sqrt{2\pi}} \int_{a/\sigma}^{b/\sigma} u \exp(-u^2/2) du \right] \\ &= \frac{\sigma}{\Phi(b/\sigma) - \Phi(a/\sigma)} \left[ -\phi\left(\frac{b}{\sigma}\right) + \phi\left(\frac{a}{\sigma}\right) \right] \\ &= \frac{\sigma \left[ \phi\left(\frac{a}{\sigma}\right) - \phi\left(\frac{b}{\sigma}\right) \right]}{\Phi(b/\sigma) - \Phi(a/\sigma)} \equiv \sigma \lambda\left(\frac{a}{\sigma}, \frac{b}{\sigma}\right) \end{aligned}$$

To find out the conditional variance, we first find out the conditional expectation for the square of the random variable.

$$\begin{aligned} E(X^2|a < X < b) &= E(\sigma^2 u^2|a < \sigma u < b) \\ &= \int_{a/\sigma}^{b/\sigma} (\sigma u)^2 \left[ \frac{\phi(u)}{\Phi(b/\sigma) - \Phi(a/\sigma)} \right] du \\ &= \frac{\sigma^2}{\Phi(b/\sigma) - \Phi(a/\sigma)} \left[ \frac{a}{\sigma} \phi\left(\frac{a}{\sigma}\right) - \frac{b}{\sigma} \phi\left(\frac{b}{\sigma}\right) + \Phi(b/\sigma) - \Phi(a/\sigma) \right] \end{aligned}$$

and therefore the conditional variance

$$\begin{aligned}
\text{Var}(X|b > X > a) &= E(X^2|b > X > a) - [E(X|b > X > a)]^2 \\
&= \sigma^2 \left[ \frac{\frac{a}{\sigma}\phi\left(\frac{a}{\sigma}\right) - \frac{b}{\sigma}\phi\left(\frac{b}{\sigma}\right)}{\Phi\left(\frac{b}{\sigma}\right) - \Phi\left(\frac{a}{\sigma}\right)} + 1 \right] - \left( \sigma \frac{(\phi\left(\frac{a}{\sigma}\right) - \phi\left(\frac{b}{\sigma}\right))}{\Phi\left(\frac{b}{\sigma}\right) - \Phi\left(\frac{a}{\sigma}\right)} \right)^2 \\
&= \sigma^2 \left[ \frac{\frac{a}{\sigma}\phi\left(\frac{a}{\sigma}\right) - \frac{b}{\sigma}\phi\left(\frac{b}{\sigma}\right)}{\Phi\left(\frac{b}{\sigma}\right) - \Phi\left(\frac{a}{\sigma}\right)} - \left( \frac{(\phi\left(\frac{a}{\sigma}\right) - \phi\left(\frac{b}{\sigma}\right))}{\Phi\left(\frac{b}{\sigma}\right) - \Phi\left(\frac{a}{\sigma}\right)} \right)^2 + 1 \right]
\end{aligned}$$

b)

Consider the censored regression

$$\begin{aligned}
Y_i &= \begin{cases} Y_i^* & \text{for } b > Y_i > a \\ 0 & \text{otherwise} \end{cases} \\
Y_i^* &= X_i\beta_0 + \varepsilon_i \\
\varepsilon_i|X_i &\sim N(0, \sigma^2)
\end{aligned}$$

The distribution of the latent variable  $Y_i^*|X_i$  is as in the CNLR. Since  $(Y_i^* - X_i\beta_0) \sim N(0, \sigma^2)$

$$f(Y_i^*|X_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(Y_i^* - X_i\beta_0)^2}{2\sigma^2}\right) = \frac{1}{\sigma} \phi\left(\frac{Y_i^* - X_i\beta_0}{\sigma}\right)$$

The conditional distribution of the observed  $y_i$  given  $x_i$  has the density:

$$\begin{aligned}
f(Y_i|X_i) &= \begin{cases} f(Y_i^*|X_i) & \text{for } b > Y_i > a \\ 1 - \Pr(b > Y_i^* > a|X_i) & \text{otherwise} \end{cases} \\
f(Y_i|X_i) &= \begin{cases} \frac{1}{\sigma} \phi\left(\frac{Y_i - X_i\beta_0}{\sigma}\right) & \text{for } b > Y_i > a \\ 1 - \Phi\left(\frac{b - X_i\beta_0}{\sigma}\right) + \Phi\left(\frac{a - X_i\beta_0}{\sigma}\right) & \text{for } Y_i = 0 \end{cases}
\end{aligned}$$

Following the previous part to calculate the first expected mean and variance,

$$\begin{aligned}
E(Y_i|X_i) &= \int y f(y|X) dy = \int_a^b y \frac{1}{\sigma} \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy = \int_a^b \left(\frac{y - X\beta_0}{\sigma} + \frac{X\beta_0}{\sigma}\right) \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy \\
&= \int_a^b \left(\frac{y - X\beta_0}{\sigma}\right) \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy + \frac{X\beta_0}{\sigma} \int_a^b \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy \\
&= \left[-\sigma \phi\left(\frac{y - X\beta_0}{\sigma}\right)\right]_a^b + X\beta_0 \left[\Phi\left(\frac{b - X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a - X_i\beta_0}{\sigma}\right)\right] \\
&= \sigma \left[\phi\left(\frac{a - X\beta_0}{\sigma}\right) - \phi\left(\frac{b - X_i\beta_0}{\sigma}\right)\right] + X\beta_0 \left[\Phi\left(\frac{b - X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a - X_i\beta_0}{\sigma}\right)\right]
\end{aligned}$$

For the second moment use the same trick as when computing the first moment.

$$\begin{aligned}
E(Y_i^2|X_i) &= \int y^2 f(y|X) dy \\
&= \int_a^b y^2 \frac{1}{\sigma} \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy \\
&= \int_a^b [(y - X\beta_0)^2 + 2yX\beta_0 - (X\beta_0)^2] \frac{1}{\sigma} \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy \\
&= \int_a^b (y - X\beta_0)^2 \frac{1}{\sigma} \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy + 2X\beta_0 \int_a^b y \frac{1}{\sigma} \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy \\
&\quad - (X\beta_0)^2 \int_a^b \frac{1}{\sigma} \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy \\
&= \int_a^b (y - X\beta_0)^2 \frac{1}{\sigma} \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy + 2X\beta_0 E(Y_i|X_i) \\
&\quad - (X\beta_0)^2 \left[ \Phi\left(\frac{b - X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a - X_i\beta_0}{\sigma}\right) \right]
\end{aligned}$$

Using the integration by parts (refer to your calculus knowledge)

$$\begin{aligned}
&\int_a^b (y - X\beta_0)^2 \frac{1}{\sigma} \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy \\
&= -\sigma \left[ (y - X\beta_0) \phi\left(\frac{y - X\beta_0}{\sigma}\right) \right]_a^b + \sigma^2 \int_a^b \frac{1}{\sigma} \phi\left(\frac{y - X\beta_0}{\sigma}\right) dy \\
&= \sigma^2 \left[ \frac{(a - X\beta_0)}{\sigma} \phi\left(\frac{a - X\beta_0}{\sigma}\right) - \frac{(b - X\beta_0)}{\sigma} \phi\left(\frac{b - X\beta_0}{\sigma}\right) \right] \\
&\quad + \sigma^2 \left[ \Phi\left(\frac{y - X\beta_0}{\sigma}\right) \right]_a^b \\
&= \sigma^2 \left[ \frac{(a - X\beta_0)}{\sigma} \phi\left(\frac{a - X\beta_0}{\sigma}\right) - \frac{(b - X\beta_0)}{\sigma} \phi\left(\frac{b - X\beta_0}{\sigma}\right) + \right. \\
&\quad \left. \Phi\left(\frac{b - X\beta_0}{\sigma}\right) - \Phi\left(\frac{a - X\beta_0}{\sigma}\right) \right]
\end{aligned}$$

$$\begin{aligned}
E(Y_i^2|X_i) &= 2X\beta_0 E(Y_i|X_i) - (X\beta_0)^2 \left[ \Phi\left(\frac{b - X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a - X_i\beta_0}{\sigma}\right) \right] \\
&\quad + \sigma^2 \left[ \frac{(a - X\beta_0)}{\sigma} \phi\left(\frac{a - X\beta_0}{\sigma}\right) - \frac{(b - X\beta_0)}{\sigma} \phi\left(\frac{b - X\beta_0}{\sigma}\right) \right. \\
&\quad \left. + \Phi\left(\frac{b - X\beta_0}{\sigma}\right) - \Phi\left(\frac{a - X\beta_0}{\sigma}\right) \right]
\end{aligned}$$

In fact, the question (on a second look) asks us to condition also on  $b > Y_i^* > a$  but this just amounts to multiplying the pdf conditional on  $X_i$  by the inverse of the probability that  $b > Y_i^* > a$ . So we need to multiply the above moments by that same inverse probability. That is

$$\begin{aligned}
E(Y_i|X_i, b > Y_i^* > a) &= \frac{E(Y_i|X_i)}{\Phi\left(\frac{b-X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a-X_i\beta_0}{\sigma}\right)} \\
&= \frac{\sigma \left[ \phi\left(\frac{a-X\beta_0}{\sigma}\right) - \phi\left(\frac{b-X_i\beta_0}{\sigma}\right) \right]}{\left[ \Phi\left(\frac{b-X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a-X_i\beta_0}{\sigma}\right) \right]} + X\beta_0 \\
E(Y_i^2|X_i, b > Y_i^* > a) &= \frac{E(Y_i^2|X_i)}{\Phi\left(\frac{b-X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a-X_i\beta_0}{\sigma}\right)} \\
&= \frac{\sigma^2 \left[ \frac{(a-X\beta_0)}{\sigma} \phi\left(\frac{a-X\beta_0}{\sigma}\right) - \frac{(b-X\beta_0)}{\sigma} \phi\left(\frac{b-X_i\beta_0}{\sigma}\right) \right]}{\left[ \Phi\left(\frac{b-X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a-X_i\beta_0}{\sigma}\right) \right]} \\
Var(Y_i|X_i, b > Y_i^* > a) &= \frac{\left\{ \sigma \left[ \phi\left(\frac{a-X\beta_0}{\sigma}\right) - \phi\left(\frac{b-X_i\beta_0}{\sigma}\right) \right] \right\}^2}{\left[ \Phi\left(\frac{b-X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a-X_i\beta_0}{\sigma}\right) \right]^2} \\
&\quad - \frac{\left[ \frac{(a-X\beta_0)}{\sigma} \phi\left(\frac{a-X\beta_0}{\sigma}\right) - \frac{(b-X\beta_0)}{\sigma} \phi\left(\frac{b-X_i\beta_0}{\sigma}\right) \right]}{\left[ \Phi\left(\frac{b-X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a-X_i\beta_0}{\sigma}\right) \right]}
\end{aligned}$$

## Estimation Methods

There are three major ways of estimating the parameters  $\beta_0$  and  $\sigma_0$ .

### 1) MLE

Define the indicator variable

$$D_i = \begin{cases} 1 & \text{for } b > Y_i > a \\ 0 & \text{otherwise} \end{cases}$$

The MLE is then found by maximizing the expression below over  $(\beta, \sigma^2)$ .

$$\begin{aligned}
L(\beta, \sigma^2) &= \prod_{i=1}^n \Pr(Y_i = 0)^{1-D_i} [f(Y_i^*|X_i, b > Y_i^* > a) \Pr(b > Y_i^* > a)]^{D_i} \\
&= \prod_{i=1}^n \left[ 1 - \Phi\left(\frac{b-X_i\beta_0}{\sigma}\right) + \Phi\left(\frac{a-X_i\beta_0}{\sigma}\right) \right]^{1-D_i} \\
&\quad \cdot \left[ \frac{1}{\sigma} \phi\left(\frac{Y_i - X_i\beta_0}{\sigma}\right) \left( \Phi\left(\frac{b-X_i\beta_0}{\sigma}\right) + \Phi\left(\frac{a-X_i\beta_0}{\sigma}\right) \right) \right]^{D_i}
\end{aligned}$$

Substituting the derivations you obtain from above and take derivatives w.r.t  $\beta$  and obtain the MLE estimator.

## NLS

NLS solves the following maximization.

$$\min_{\beta} \sum_{i=1}^n (Y_i - E(Y_i|X_i))^2$$

$$\hat{\beta}_{NLS} \in \arg \min \sum \left\{ y_i - [X_i\beta + \sigma\lambda \left( \frac{a - X_i\beta}{\sigma}, \frac{b - X_i\beta}{\sigma} \right)] \right\}^2 \tilde{I}\{a < Y_i^* < b\}$$

We solved for  $E(Y_i|X_i)$  above, here we just substitute that expression.

**Remark 5** Here as the function form of  $\text{Var}(\varepsilon_i)$  varies across observations and we have conditional heteroscedasticity. We may have to use WNLS to adjust for the above with appropriate weight on the conditional variance.

$$\hat{\beta}_{WNLS} \in \arg \min \sum \left\{ \frac{\left\{ y_i - [X_i\beta + \sigma\lambda \left( \frac{a - X_i\beta}{\sigma}, \frac{b - X_i\beta}{\sigma} \right)] \right\}^2}{[\text{Var}(\varepsilon_i|X_i, \frac{a - X_i\beta}{\sigma} < u_i^* < \frac{b - X_i\beta}{\sigma})]} \tilde{I}\{a < Y_i^* < b\} \right\}$$

## Heckman Two-Step Estimation Procedures

In step 1 use  $D_i$  as defined above to estimate the probit model and obtain an estimate of  $\frac{\beta_0}{\sigma_0}$ , call it  $\hat{\gamma}$ .

In step 2 use the estimate from step 1 to allow us form a regression model from the conditional expectation function as follows.

$$E(Y_i|X_i, b > Y_i^* > a) = \sigma \frac{\phi\left(a - X_i\hat{\gamma}\frac{a - X_i\beta_0}{\sigma}\right) - \phi\left(\frac{b - X_i\beta_0}{\sigma}\right)}{\Phi\left(\frac{b - X_i\beta_0}{\sigma}\right) - \Phi\left(\frac{a - X_i\beta_0}{\sigma}\right)} + X_i\beta_0$$

Now using  $\hat{\gamma}$  from step 1 write

$$E(Y_i|X_i, b > Y_i^* > a) = \sigma \frac{\phi(a - X_i\hat{\gamma}) - \phi(b - X_i\hat{\gamma})}{\Phi(b - X_i\hat{\gamma}) - \Phi(a - X_i\hat{\gamma})} + X_i\beta_0$$

and regress  $Y_i$  on  $X_i$  and  $\frac{\phi(a - X_i\hat{\gamma}) - \phi(b - X_i\hat{\gamma})}{\Phi(b - X_i\hat{\gamma}) - \Phi(a - X_i\hat{\gamma})}$  to form estimates of  $\beta_0, \sigma$ .

**Remark 6** Please refer to the lecture note or Greene for 2step Heckman estimation procedures.