Econ 203B: Single Equation Models

Solutions for Problem Set 5

Gissele Gajate/Michael Powell
Department of Economics, UCLA
March 3rd, 2006

1 Hayashi Chapter 7, Analytical Exercises

1 (Kullback-Leibler information inequality) Let $f(y|x;\theta)$ be a parametric family of hypothetical conditional density functions, with the true density function given by $f(y|x;\theta_0)$. Suppose $E[\log f(y|x,\theta)]$ exists and is finite for all θ. The Kullback-Leibler information inequality states that if

$$Prob[f(y|x;\theta) \neq f(y|x,\theta_0)] > 0,$$

then

$$E[\log f(y|x,\theta)] < E[\log f(y|x,\theta_0)].$$

Prove this by taking the following steps. In the proof, for simplicity only, assume that $f(y|x;\theta) > 0$ for all (y,x) and θ, so that it is legitimate to take logs of the density $f(y|x,\theta)$.

a. Suppose Prob$[f(y|x;\theta) \neq f(y|x;\theta_0)] > 0$. Let $w = (y,x')'$ and define

$$a(w) \equiv \frac{f(y|x;\theta)}{f(y|x;\theta_0)}$$

Verify that $a(w) \neq 1$ with positive probability, so that $a(w)$ is a nonconstant random variable.

Solution We know that

$$\begin{aligned} 0 &< \text{Prob}[f(y|x;\theta) \neq f(y|x;\theta_0)] \\ &= \text{Prob}\left[\frac{f(y|x;\theta)}{f(y|x;\theta_0)} \neq 1\right] \\ &= \text{Prob}[a(w) \neq 1] \end{aligned}$$

And therefore, $a(w)$ is a nonconstant random variable.

b. The strict version of Jensen's inequality states that if $c(x)$ is a strictly concave function and x is a nonconstant random variable, then $E[c(x)] < c(E[x])$.

Using this fact, show that

$$E[\log a(w)] < \log(E[a(w)])$$

Hint: $\log(x)$ is strictly concave.

Solution Since this part of the question is so seemingly unsubstantive, I will begin by showing the hint:

$$\frac{\partial \log(x)}{\partial x} = \frac{1}{x}; \quad \frac{\partial^2 \log(x)}{\partial x^2} = -\frac{1}{x^2} < 0 \ \forall x \neq 0$$

Therefore, $\log(x)$ is strictly concave (except at 0, where it is not defined.) Since, as established in part (a), $a(w)$ is a nonconstant random variable, we have the result:

$$E[\log a(w)] < \log(E[a(w)])$$

c. Show that $E[a(w)] = 1$. Hint: The conditional mean of $a(w)$ equals 1 because

$$\begin{aligned} E[a(w)|x] &= \int a(w) f(y|x;\theta_0)\,dy \\ &= \int \frac{f(y|x;\theta)}{f(y|x;\theta_0)} f(y|x;\theta_0)\,dy \\ &= \int f(y|x;\theta_0)\,dy \\ &= 1 \end{aligned}$$

The last equality holds because $f(y|x,\theta)$ is a density function for any x and θ. Note well that the conditional expectation is taken with respect to the true conditional density $f(y|x;\theta_0)$.

Solution Since $E[a(w)|x] = 1$, by the law of iterated expectations, we have that

$$E[a(w)] = E[E[a(w)|x]] = E[1] = 1$$

d. Finally, show the desired result.

From part (b), we know that $E[\log a(w)] < \log(E[a(w)])$. From part ($c$), we know that $E[a(w)] = 1$ and therefore, $\log(E[a(w)]) = 0$. This gives us:

$$\begin{aligned} 0 &> E[\log a(w)] \\ &= E\left[\log \frac{f(y|x;\theta)}{f(y|x;\theta_0)}\right] \\ &= E[\log f(y|x;\theta) - \log f(y|x;\theta_0)] \\ &= E[\log f(y|x;\theta)] - E[\log f(y|x;\theta_0)] \end{aligned}$$

Rearranging gives us

$$E[\log f(y|x;\theta)] < E[\log f(y|x;\theta_0)]$$

Which is the desired result.

2 (Information matrix equality) For ML, the score vector and the Hessian for observation (y,x) can be rewritten (with $w \equiv (y,x')'$) as:

$$\begin{aligned} \underset{p\times 1}{s(w;\theta)} &= \frac{\partial \log f(y|\, x;\theta)}{\partial\theta}, \\ \underset{p\times p}{H(w;\theta)} &= \frac{\partial s(w;\theta)}{\partial\theta'} = \frac{\partial^2 \log f(y|\, x;\theta)}{\partial\theta\partial\theta'}. \end{aligned}$$

As usual, for simplicity, assume $f(y|\, x;\theta) > 0$ for all (y,x) and θ, so it is legitimate to take logs of the density function.

a. Assuming that integration (i.e., taking expectations) and differentiation can be interchanged, show that $E[s(w;\theta_0)] = 0$. Hint: Since $f(y|\, x;\theta)$ is a hypothetical density, its integral is unity:

$$\int f(y|\, x;\theta)\, dy = 1.$$

Differentiate both sides of this equation with respect to θ and then change the order of differentiation and integration to obtain the identity

$$\int s(w;\theta)\, f(y|\, x;\theta)\, dy = \underset{(p\times 1)}{0}.$$

Set $\theta = \theta_0$ and obtain $E[s(w;\theta_0)|\, x] = 0$. Finally, by the law of iterated expectations, we have $E[s(w;\theta_0)] = E[E[s(w;\theta_0)|\, x]] = E[0] = 0$.

Solution The hint really solves the problem, except for one non-trivial step, but for formalities, we begin by noting that for any θ, the function $f(y|\, x;\theta)$ is a density function and therefore,

$$\int f(y|\, x;\theta)\, dy = 1$$

Taking derivatives of both sides with respect to θ gives us:

$$0 = \frac{\partial}{\partial\theta} 1 = \frac{\partial}{\partial\theta}\int f(y|\, x;\theta)\, dy = \int \frac{\partial f(y|\, x;\theta)}{\partial\theta} dy \tag{1}$$

Next, note that

$$\frac{\partial \log f(y|\, x;\theta)}{\partial\theta} = \frac{1}{f(y|\, x;\theta)}\frac{\partial f(y|\, x;\theta)}{\partial\theta}$$

Or, rearranging,

$$\begin{aligned} \frac{\partial f(y|\, x;\theta)}{\partial\theta} &= \frac{\partial \log f(y|\, x;\theta)}{\partial\theta} f(y|\, x;\theta) \\ &= s(w;\theta)\, f(y|\, x;\theta) \end{aligned}$$

Substituting this into (1) gives us:

$$0 = \int \underset{(p\times 1)}{s(w;\theta)} f(y|\, x;\theta)\, dy = E_\theta[s(w;\theta)] \tag{2}$$

Where E_θ denotes the expectation operator given that θ is the true value of the parameter.

Substituting $\theta = \theta_0$ gives us the final result:

$$E[s(w;\theta_0)] \equiv E_{\theta_0}[s(w;\theta_0)] = 0$$

b. Show the information matrix equality:

$$-E\left[H\left(w;\theta_0\right)\right]=E\left[s\left(w;\theta_0\right)s\left(w;\theta_0\right)'\right].$$

Hint: Differentiating both sides of the above identity in the previous hint and assuming that the order of differentiation and integration can be interchanged, we obtain

$$\int\frac{\partial}{\partial\theta'}\underbrace{\left[s\left(w;\theta\right)f\left(y|\,x;\theta\right)\right]}_{(p\times 1)}dy=\underset{(p\times p)}{0}.$$

Show that the integrand can be written as

$$\begin{aligned}&\frac{\partial}{\partial\theta'}\left[s\left(w;\theta\right)f\left(y|\,x;\theta\right)\right]\\=\;&H\left(w;\theta\right)f\left(y|\,x;\theta\right)+s\left(w;\theta\right)s\left(w;\theta\right)'f\left(y|\,x;\theta\right).\end{aligned}$$

Solution Rewriting (2):

$$0=E\left[s\left(w;\theta\right)\right]=\int s\left(w;\theta\right)f\left(y|\,x;\theta\right)dy$$

And taking derivatives with respect to θ', we have:

$$\begin{aligned}0&=\frac{\partial}{\partial\theta'}\int s\left(w;\theta\right)f\left(y|\,x;\theta\right)dy\\&=\int\frac{\partial}{\partial\theta'}s\left(w;\theta\right)f\left(y|\,x;\theta\right)dy\\&=\int\left\{\left[\frac{\partial}{\partial\theta'}s\left(w;\theta\right)\right]f\left(y|\,x;\theta\right)+s\left(w;\theta\right)\frac{\partial}{\partial\theta'}f\left(y|\,x;\theta\right)\right\}dy\\&=\int H\left(w;\theta\right)f\left(y|\,x;\theta\right)dy+\int s\left(w;\theta\right)\frac{\partial\log f\left(y|\,x;\theta\right)}{\partial\theta'}f\left(y|\,x;\theta\right)dy\\&=\int H\left(w;\theta\right)f\left(y|\,x;\theta\right)dy+\int s\left(w;\theta\right)s\left(w;\theta\right)'f\left(y|\,x;\theta\right)dy\\&=E_\theta\left[H\left(W;\theta\right)\right]+E_\theta\left[s\left(W;\theta\right)s\left(W;\theta\right)'\right]\end{aligned}$$

Where the fourth equality holds because

$$\frac{\partial\log f\left(y|\,x;\theta\right)}{\partial\theta'}=\frac{1}{f\left(y|\,x;\theta\right)}\frac{\partial}{\partial\theta'}f\left(y|\,x;\theta\right)$$

Or

$$\frac{\partial}{\partial\theta'}f\left(y|\,x;\theta\right)=\frac{\partial\log f\left(y|\,x;\theta\right)}{\partial\theta'}f\left(y|\,x;\theta\right)$$

Rearranging the above result gives us:

$$-E_\theta\left[H\left(W;\theta\right)\right]=E_\theta\left[s\left(W;\theta\right)s\left(W;\theta\right)'\right]$$

Evaluating this at θ_0, we have the desired result:

$$-E\left[H\left(W;\theta_0\right)\right]=E\left[s\left(W;\theta_0\right)s\left(W;\theta_0\right)'\right]$$

3 (Trinity for linear regression model) Consider the linear regression model with normal errors, whose conditional density for observation t is

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

Let $\left(\hat{\beta}, \hat{\sigma}^2\right)$ be the unrestricted ML estimate of $\theta = \left(\beta', \sigma^2\right)'$ and let $\left(\tilde{\beta}, \tilde{\sigma}^2\right)$ be the restricted ML estimate subject to the constraint $R\beta = c$ where R is an $r \times K$ matrix of known constants. Assume that $\Theta = \mathbb{R}^K \times \mathbb{R}_{++}$ and that $E\left[x_t x_t'\right]$ is nonsingular. Also, let

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

a. Verify that $\hat{\beta}$ minimizes the sum of squared residuals. So it is the OLS estimator. Verify that $\tilde{\beta}$ minimizes the sum of squared residuals subject to the constraint $R\beta = c$. So it is the restricted least squares estimator.

Solution First, I would like to switch to more preferable notation. Let $\hat{\beta}_{UR}^{ML} = \hat{\beta}$ and $\hat{\beta}_R^{ML} = \tilde{\beta}$. Also, note that Hayashi defines x_t as a column vector of the data for the t'th observation. In class, we define it as a row vector of data for the t'th observation, but this will not change anything. Deriving $\hat{\beta}_{UR}^{ML}$, then, we have:

$$\begin{aligned} \frac{\partial}{\partial\beta}\log L\left(\{y_t\}|\{x_t\};\beta,\sigma^2\right) &= \frac{\partial}{\partial\beta}\sum_{t=1}^n \log f\left(y_t|x_t;\beta,\sigma^2\right) \\ &= \frac{\partial}{\partial\beta}\sum_{t=1}^n\left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma^2\right) - \frac{\left(y_t - x_t'\beta\right)^2}{2\sigma^2}\right) \\ &= -\sum_{t=1}^n \frac{\left(y_t - x_t'\beta\right)x_t}{\sigma^2} = 0 \end{aligned}$$

Rearranging, we get:

$$0 = \sum_{i=1}^n x_t\left(y_t - x_t'\hat{\beta}_{UR}^{ML}\right) = \sum_{i=1}^n x_t y_t - \sum_{i=1}^n x_t x_t'\hat{\beta}_{UR}^{ML}$$

Or, in matrix notation, with $X = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix}$, we have:

$$\begin{aligned} X'X\hat{\beta}_{UR}^{ML} &= X'Y \\ \hat{\beta}_{UR}^{ML} &= \left(X'X\right)^{-1}X'Y \end{aligned}$$

Which is the same as the OLS estimator for β. Proceeding similarly to derive $\hat{\beta}_R^{ML}$, we want to

$$\max_{\beta:R\beta=c}\log L\left(\{y_t\}|\{x_t\};\beta,\sigma^2\right) = \max_{\beta:R\beta=c}\sum_{t=1}^n\left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma^2\right) - \frac{\left(y_t - x_t'\beta\right)^2}{2\sigma^2}\right)$$

The Lagrangian for this constrained optimization problem is:

$$L\left(\beta,\sigma^2,\lambda\right) = \sum_{t=1}^n\left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma^2\right) - \frac{\left(y_t - x_t'\beta\right)^2}{2\sigma^2}\right) + \underset{(1\times p)}{\lambda}\left(\underset{(p\times k)}{R}\underset{(k\times 1)}{\beta} - c\right)$$

Taking first order conditions with respect to β and λ, we get:

$$\begin{aligned}
(\beta) \quad &: \quad -\sum_{t=1}^{n} \frac{\left(y_t - x_t'\hat{\beta}_R^{ML}\right) x_t}{\hat{\sigma}_R^2} + R'\lambda' = 0 \\
(\lambda) \quad &: \quad R\hat{\beta}_R^{ML} = c
\end{aligned}$$

Or

$$\begin{aligned}
\sum_{t=1}^{n} \frac{\left(y_t - x_t'\hat{\beta}_R^{ML}\right) x_t}{\hat{\sigma}_R^2} &= R'\lambda' \\
\frac{1}{\hat{\sigma}_R^2}\left(X'Y - X'X\hat{\beta}_R^{ML}\right) &= R'\lambda'
\end{aligned}$$

And solving for $\hat{\beta}_R^{ML}$, we have:

$$\begin{aligned}
X'Y - X'X\hat{\beta}_R^{ML} &= \hat{\sigma}_R^2 R'\lambda' \\
X'X\hat{\beta}_R^{ML} &= X'Y - \hat{\sigma}_R^2 R'\lambda' \\
\hat{\beta}_R^{ML} &= \left(X'X\right)^{-1} X'Y - \hat{\sigma}_R^2 \left(X'X\right)^{-1} R'\lambda' \\
&= \hat{\beta}_{UR}^{ML} - \hat{\sigma}_R^2 \left(X'X\right)^{-1} R'\lambda' \qquad (1)
\end{aligned}$$

Plugging this into the constraint:

$$R\hat{\beta}_R^{ML} = c$$

We get:

$$R\hat{\beta}_R^{ML} - \hat{\sigma}_R^2 R\left(X'X\right)^{-1} R'\lambda' = c$$

And solving for λ':

$$\begin{aligned}
\hat{\sigma}_R^2 R\left(X'X\right)^{-1} R'\lambda' &= R\hat{\beta}_R^{ML} - c \\
\lambda' &= \frac{1}{\hat{\sigma}_R^2}\left[R\left(X'X\right)^{-1} R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right)
\end{aligned}$$

Finally, substituting this into (1) gives us:

$$\begin{aligned}
\hat{\beta}_R^{ML} &= \hat{\beta}_{UR}^{ML} - \hat{\sigma}_R^2\left(X'X\right)^{-1} R'\frac{1}{\hat{\sigma}_R^2}\left[R\left(X'X\right)^{-1} R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right) \\
&= \hat{\beta}_{UR}^{ML} - \left(X'X\right)^{-1} R'\left[R\left(X'X\right)^{-1} R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right)
\end{aligned}$$

I claim that this is the same solution as we would achieve by estimating the restricted OLS model. To show this, we solve for the following value

$$\begin{aligned}
\underset{\beta:R\beta=c}{\arg\min}\,\left(Y - X\beta\right)'\left(Y - X\beta\right) &= \underset{\beta:R\beta=c}{\arg\min}\left(Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta\right) \\
&= \underset{\beta:R\beta=c}{\arg\min}\left(Y'Y - 2\beta'X'Y + \beta'X'X\beta\right) \\
&= \underset{\beta:R\beta=c}{\arg\min}\frac{1}{2}\left(Y'Y - 2\beta'X'Y + \beta'X'X\beta\right)
\end{aligned}$$

The Lagrangian for this problem is:

$$L(\beta, \lambda) = \frac{1}{2}\left(Y'Y - 2\beta' X'Y + \beta' X'X\beta\right) + \lambda\left(R\beta - c\right)$$

Taking first order conditions, we have:

$$\begin{aligned}(\beta) &: \quad \frac{1}{2}\left(-2X'Y + 2X'X\hat{\beta}_R^{OLS}\right) + R'\lambda' = 0 \\ (\lambda) &: \quad R\beta = c\end{aligned}$$

Solving for $\hat{\beta}_R^{OLS}$, we have:

$$\begin{aligned}X'X\hat{\beta}_R^{OLS} &= X'Y - R'\lambda' \\ \hat{\beta}_R^{OLS} &= \left(X'X\right)^{-1}X'Y - \left(X'X\right)^{-1}R'\lambda' \\ &= \hat{\beta}_{UR}^{OLS} - \left(X'X\right)^{-1}R'\lambda' \qquad (2)\end{aligned}$$

And substituting this into the constraint:

$$R\hat{\beta}_{UR}^{OLS} - R\left(X'X\right)^{-1}R'\lambda' = c$$

Solving for λ', we get:

$$\begin{aligned}R\left(X'X\right)^{-1}R'\lambda' &= R\hat{\beta}_{UR}^{OLS} - c \\ \lambda' &= \left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{OLS} - c\right)\end{aligned}$$

And substituting this into (2)

$$\hat{\beta}_R^{OLS} = \hat{\beta}_{UR}^{OLS} - \left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{OLS} - c\right)$$

And conclude that since $\hat{\beta}_{UR}^{OLS} = \hat{\beta}_{UR}^{ML}$, we have that $\hat{\beta}_R^{OLS} = \hat{\beta}_R^{ML}$.

b. Let $Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\log f\left(y_t|\, x_t; \beta, \sigma^2\right)$. Show that

$$\begin{aligned}Q_n\left(\hat{\theta}\right) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2} - \frac{1}{2}\log\left(\frac{SSR_U}{n}\right), \\ Q_n\left(\tilde{\theta}\right) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2} - \frac{1}{2}\log\left(\frac{SSR_R}{n}\right),\end{aligned}$$

where $SSR_U\left(\equiv \sum\left(y_t - x_t'\hat{\beta}\right)^2\right)$ is the unrestricted sum of squared residuals and $SSR_R\left(\equiv \sum\left(y_t - x_t'\tilde{\beta}\right)^2\right)$ is the restricted sum of squared residuals. Hint: Show that $\hat{\sigma}^2 = \frac{SSR_U}{n}$ and $\tilde{\sigma}^2 = \frac{SSR_R}{n}$.

Solution Once again, I will change notation slightly. Let $\hat{\sigma}_{UR}^2 = \hat{\sigma}^2$ and $\hat{\sigma}_R^2 = \tilde{\sigma}^2$, which are understood to be the unrestricted and restricted ML estimators of σ^2, respectively. First, we will proceed to derive $\hat{\sigma}_{UR}^2$ as per the hint. The problem is to find:

$$\left(\hat{\beta}_{UR}^{ML}, \hat{\sigma}_{UR}^2\right) = \arg\max \sum_{t=1}^{n}\left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma^2\right) - \frac{\left(y_t - x_t'\beta\right)^2}{2\sigma^2}\right)$$

Since we have already solved for $\hat{\beta}_{UR}^{ML}$ in part (a), we only have to look at the following FOC:

$$\left(\sigma^2\right) : -\frac{n}{2\hat{\sigma}_{UR}^2} + \sum_{t=1}^{n} \frac{\left(y_t - x_t'\hat{\beta}_{UR}^{ML}\right)^2}{2\left(\hat{\sigma}_{UR}^2\right)^2} = 0$$

Rearranging to solve for $\hat{\sigma}_{UR}^2$, we have:

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

$$\begin{aligned} \hat{\sigma}_{UR}^2 &= \frac{1}{n}\sum_{t=1}^{n} \left(y_t - x_t\hat{\beta}_{UR}^{ML}\right)^2 \\ &= \frac{1}{n} SSR_{UR} \end{aligned}$$

Proceeding similarly to solve for $\hat{\sigma}_R^2$, we have the following Lagrangian:

$$L\left(\beta, \sigma^2, \lambda\right) = \sum_{t=1}^{n} \left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma^2\right) - \frac{\left(y_t - x_t'\beta\right)^2}{2\sigma^2}\right) + \underset{(1\times p)}{\lambda} \left(\underset{(p\times k)}{R}\underset{(k\times 1)}{\beta} - c\right)$$

Which yields the following FOC wrt σ^2:

$$\left(\sigma^2\right) : -\frac{n}{2\hat{\sigma}_R^2} + \sum_{t=1}^{n} \frac{\left(y_t - x_t'\hat{\beta}_R^{ML}\right)^2}{2\left(\hat{\sigma}_R^2\right)^2} = 0$$

Solving for $\hat{\sigma}_R^2$,

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

$$\begin{aligned} \hat{\sigma}_R^2 &= \frac{1}{n}\sum_{t=1}^{n} \left(y_t - x_t'\hat{\beta}_R^{ML}\right)^2 \\ &= \frac{1}{n} SSR_R \end{aligned}$$

Define $\hat{\theta}_{UR}^{ML} = \hat{\theta}$ and $\hat{\theta}_R^{ML} = \tilde{\theta}$ in a similar shift of notation. We then have:

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \left(-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\sigma^2\right) - \frac{\left(y_t - x_t'\beta\right)^2}{2\sigma^2}\right)$$

Substituting in $\hat{\theta}_{UR}^{ML} = \left(\hat{\beta}_{UR}^{ML}, \hat{\sigma}_{UR}^2\right)$, we have:

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

And substituting in $\hat{\theta}_R^{ML} = \left(\hat{\beta}_R^{ML,}\hat{\sigma}_R^2\right)$, we have:

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

c. Verify that $\hat{\Sigma}$ given here, although not the same as $-\frac{1}{n}\sum_{t=1}^{n}H\left(w_t;\hat{\theta}\right)$, is consistent for $-E\left[H\left(w_t;\theta_0\right)\right]$.

Verify that $\tilde{\Sigma}$ given here, although not the same as $-\frac{1}{n}\sum_{t=1}^{n}H\left(w_t;\tilde{\theta}\right)$, is consistent for $-E\left[H\left(w_t;\theta_0\right)\right]$.

Hint: From the discussion in Example 7.8, $\hat{\theta}$ is consistent. As mentioned in Section 7.4, $\tilde{\theta}$ too is consistent under the null. Given the consistency of $\hat{\theta}$ and $\tilde{\theta}$, it should be easy to show that $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ are consistent for σ^2.

Solution In order to show that $p\lim\hat{\Sigma}_{UR}^{ML} = -E\left[H\left(w_t;\theta_0\right)\right]$ and $p\lim\hat{\Sigma}_R^{ML} = -E\left[H\left(w_t;\theta_0\right)\right]$, I will first derive $-E\left[H\left(w_t;\theta_0\right)\right]$. Recall that $H\left(w_t;\theta_0\right) = \left.\frac{\partial^2 m}{\partial\theta\partial\theta'}\right|_{\theta_0}$.

$$\begin{aligned}
m\left(w_t;\theta_0\right) &= -\frac{1}{2}\log(2\pi)-\frac{1}{2}\log\left(\sigma^2\right)-\frac{\left(y_t-x_t'\beta\right)^2}{2\sigma^2}\\
&= -\frac{1}{2}\log(2\pi)-\frac{1}{2}\log\left(\sigma^2\right)-\frac{\left(y_t-x_t'\beta\right)^2}{2\sigma^2}
\end{aligned}$$

Taking first derivatives:

$$\begin{aligned}
\frac{\partial m}{\partial\beta} &= -\frac{\left(y_t-x_t'\beta\right)\left(-x_t\right)}{\sigma^2}\\
&= \frac{1}{\sigma^2}\left(x_t y_t - x_t x_t'\beta\right)
\end{aligned}$$

$$\frac{\partial m}{\partial\sigma^2} = -\frac{1}{2\sigma^2}+\frac{\left(y_t-x_t'\beta\right)^2}{2\left(\sigma^2\right)^2}$$

And putting these together, we have:

$$\frac{\partial m}{\partial\theta} = \frac{\partial m}{\partial\begin{bmatrix}\beta\\ \sigma^2\end{bmatrix}} = \begin{bmatrix}\frac{\partial m}{\partial\beta}\\ \frac{\partial m}{\partial\sigma^2}\end{bmatrix} = \begin{bmatrix}\frac{1}{\sigma^2}\left(x_t y_t - x_t x_t'\beta\right)\\ -\frac{1}{2\sigma^2}+\frac{\left(y_t-x_t'\beta\right)^2}{2\left(\sigma^2\right)^2}\end{bmatrix}$$

Next, taking second derivatives, we have:

$$\begin{aligned}
\frac{\partial^2 m}{\partial\beta\partial\beta'} &= -\frac{1}{\sigma^2}x_t x_t'\\
\frac{\partial^2 m}{\partial\beta\partial\sigma^2} &= -\frac{1}{\left(\sigma^2\right)^2}\left(x_t y_t - x_t x_t'\beta\right)\\
\frac{\partial^2 m}{\partial\sigma^2\partial\sigma^2} &= \frac{1}{2\left(\sigma^2\right)^2}-\frac{1}{\left(\sigma^2\right)^3}\left(y_t-x_t'\beta\right)^2
\end{aligned}$$

Putting these together to form the Hessian matrix, we have:

$$\begin{aligned}\frac{\partial^2 m}{\partial\theta\partial\theta'} &= \frac{\partial}{\partial\theta'}\left(\frac{\partial m}{\partial\theta}\right) = \left[\begin{array}{cc} \frac{\partial}{\partial\beta'}\left(\frac{\partial m}{\partial\theta}\right) & \frac{\partial}{\partial\sigma^2}\left(\frac{\partial m}{\partial\theta}\right) \end{array}\right] \\ &= \left[\begin{array}{cc} -\frac{1}{\sigma^2}x_t x_t' & -\frac{1}{(\sigma^2)^2}\left(x_t y_t - x_t x_t'\beta\right) \\ -\frac{1}{(\sigma^2)^2}\left(x_t y_t - x_t x_t'\beta\right) & \frac{1}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\left(y_t - x_t'\beta\right)^2 \end{array}\right]\end{aligned}$$

Taking expected values of each element of $-\frac{\partial^2 m}{\partial\theta\partial\theta'}$, we have:

$$\begin{aligned} E\left[\frac{1}{\sigma^2}x_t x_t'\right] &= \frac{1}{\sigma^2}E\left[x_t x_t'\right] \\ E\left[\frac{1}{(\sigma^2)^2}\left(x_t y_t - x_t x_t'\beta\right)\right] &= E\left[\frac{1}{(\sigma^2)^2}\left(x_t\left(x_t'\beta+\varepsilon\right) - x_t x_t'\beta\right)\right] \\ &= E\left[\frac{\varepsilon}{(\sigma^2)^2}\right] = -\frac{1}{(\sigma^2)^2}E\left[\varepsilon\right] = 0 \\ E\left[-\frac{1}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}\left(y_t - x_t'\beta\right)^2\right] &= -\frac{1}{2(\sigma^2)^2} + \frac{E\left[\left(y_t - x_t'\beta\right)^2\right]}{(\sigma^2)^3} \\ &= -\frac{1}{2(\sigma^2)^2} + \frac{\sigma^2}{(\sigma^2)^3} = \frac{1}{2(\sigma^2)^2}\end{aligned}$$

Putting this together yields:

$$E\left[-H\left(w_t;\theta_0\right)\right] = E\left[-\left.\frac{\partial^2 m}{\partial\theta\partial\theta'}\right|_{\theta_0}\right] = \left[\begin{array}{cc} \frac{1}{\sigma_0^2}E\left[x_t x_t'\right] & 0 \\ 0 & \frac{1}{2\left(\sigma_0^2\right)^2}\end{array}\right]$$

With this out of the way, I will show convergence in probability element by element of $\hat{\Sigma}_{UR}^{ML}$ and $\hat{\Sigma}_{R}^{ML}$ to $E\left[-H\left(w_t;\theta_0\right)\right]$. I claim that this will suffice to show the result. A richer argument could be made using the Cramer-Wold device and noting that convergence in distribution to a constant is equivalent to convergence in probability to a constant.

For $\hat{\Sigma}_{UR}^{ML}$, we have the following results. First, it is apparent that

$$\frac{1}{n}\sum_{t=1}^{n} x_t x_t' \xrightarrow{p} E\left[x_t x_t'\right]$$

Provided that $\{x_t x_t'\}$ are i.i.d. (which is satisfied since $\{x_t\}$ are i.i.d. and thus, $\{g\left(x_t\right)\}$ are i.i.d. for any function g - in particular, $g\left(x_t\right) = x_t x_t'$) and that $E\left[|x_t x_t'|\right] < +\infty$. Next, we need to show that $\hat{\sigma}_{UR}^2 \xrightarrow{p} \sigma_0^2$.

$$\begin{aligned}\hat{\sigma}_{UR}^2 &= \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} = \frac{1}{n}\varepsilon' M_X \varepsilon \\ &= \frac{1}{n}\varepsilon'\left(I - X\left(X'X\right)^{-1}X'\right)\varepsilon \\ &= \frac{1}{n}\varepsilon'\varepsilon - \frac{1}{n}\varepsilon' X\left(X'X\right)^{-1}X'\varepsilon \\ &= \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 - \left(\frac{1}{n}\sum_{t=1}^{n}x_t\varepsilon_t\right)'\left(\frac{1}{n}\sum_{t=1}^{n}x_t x_t'\right)^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}x_t\varepsilon_t\right) \\ &\xrightarrow{p} E\left[\varepsilon_t^2\right] - \underbrace{E\left[x_t\varepsilon_t\right]'}_{=0}\left[E\left[x_t x_t'\right]\right]^{-1}E\left[x_t\varepsilon_t\right]\end{aligned}$$

That is, $\hat{\sigma}_{UR}^2 \xrightarrow{p} E\left[\varepsilon_t^2\right] = \sigma_0^2$ by Slutsky's theorem, the Mann-Wald theorem, and the weak law of large numbers given appropriate moment conditions (which I will assume). Using the Mann Wald theorem again, we have that $\frac{1}{\hat{\sigma}_{UR}^2} \xrightarrow{p} \frac{1}{\sigma_0^2}$ and $\frac{1}{2\left(\hat{\sigma}_{UR}^2\right)^2} \xrightarrow{p} \frac{1}{2(\sigma^2)^2}$. By Slutsky's theorem, then, we have that

$$\frac{1}{\hat{\sigma}_{UR}^2}\frac{1}{n}\sum_{t=1}^{n} x_t x_t' \xrightarrow{p} \frac{1}{\sigma_0^2} E\left[x_t x_t'\right]$$

And therefore, element by element, we have that $\hat{\Sigma}_{UR}^{ML} \xrightarrow{p} E\left[-H\left(w_t;\theta_0\right)\right]$. Proceeding similarly for $\hat{\Sigma}_{R}^{ML}$, we have that $\frac{1}{n}\sum_{t=1}^{n} x_t x_t' \xrightarrow{p} E\left[x_t x_t'\right]$ by the weak law of large numbers provided that $\{x_t x_t'\}$ are i.i.d. (they are) and $E\left[|x_t x_t'|\right] < +\infty$. Next, we have by part (b) that

$$\hat{\sigma}_R^2 = \frac{\hat{\varepsilon}_R' \hat{\varepsilon}_R}{n}$$

But

$$\begin{aligned}
\hat{\varepsilon}_R &= \left(Y - X\hat{\beta}_R^{ML}\right) \\
&= \left(Y - X\hat{\beta}_{UR}^{ML} + X\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right)\right)
\end{aligned}$$

Where the second equality holds by part (a). Proceeding further,

$$\begin{aligned}
\hat{\varepsilon}_R'\hat{\varepsilon}_R &= \left(Y - X\hat{\beta}_{UR}^{ML} + X\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right)\right)' \\
&\quad \cdot\left(Y - X\hat{\beta}_{UR}^{ML} + X\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right)\right) \\
&= \left(Y - X\hat{\beta}_{UR}^{ML}\right)'\left(Y - X\hat{\beta}_{UR}^{ML}\right) \\
&\quad + \left[\begin{array}{c} \left(R\hat{\beta}_{UR}^{ML} - c\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}R\left(X'X\right)^{-1}X' \\ \cdot X\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right) \end{array}\right] \\
&\quad + 2\left(Y - X\hat{\beta}_{UR}^{ML}\right)'\left[X\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right)\right] \\
&= \hat{\varepsilon}_{UR}'\hat{\varepsilon}_{UR} + \left(R\hat{\beta}_{UR}^{ML} - c\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right)
\end{aligned}$$

Where the last equality holds because

$$\begin{aligned}
&2\left(Y - X\hat{\beta}_{UR}^{ML}\right)'\left[X\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right)\right] \\
= \;& 2Y'\underbrace{M_X X}_{=0}\left(X'X\right)^{-1}R'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right) = 0
\end{aligned}$$

This result gives us

$$\hat{\sigma}_R^2 = \hat{\sigma}_{UR}^2 + \frac{1}{n}\left(R\hat{\beta}_{UR}^{ML} - c\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right)$$

Which can easily be used to show that $\hat{\sigma}_R^2 - \hat{\sigma}_{UR}^2 \geq 0$. Next,

$$\begin{aligned}
0 &\leq \hat{\sigma}_R^2 - \hat{\sigma}_{UR}^2 \leq \frac{1}{n}\left(R\hat{\beta}_{UR}^{ML} - c\right)'\left[R\left(X'X\right)^{-1}R'\right]^{-1}\left(R\hat{\beta}_{UR}^{ML} - c\right) \\
&\leq \frac{1}{n}\left|\left|R\hat{\beta}_{UR}^{ML} - c\right|\right|'\left|\left|\left[R\left(X'X\right)^{-1}R'\right]^{-1}\right|\right|\left|\left|R\hat{\beta}_{UR}^{ML} - c\right|\right|
\end{aligned}$$

By the Cauchy-Schwarz inequality. Since $\hat{\beta}_{UR}^{ML} \xrightarrow{p} \beta_0$, under the null, $R\hat{\beta}_{UR}^{ML} - c \xrightarrow{p} 0$. If we assume that $\left\| \left[R\left(X'X\right)^{-1} R' \right]^{-1} \right\| < +\infty$, we have that

$$\hat{\sigma}_R^2 - \hat{\sigma}_{UR}^2 \xrightarrow{p} 0$$

(Note that alternatively, instead of appealing to the result that $R\hat{\beta}_{UR}^{ML} - c \xrightarrow{p} 0$, I could have used the weaker result that $\left\| R\hat{\beta}_{UR}^{ML} - c \right\| < +\infty$ and therefore the term on the right is of the form $\frac{1}{n} \cdot C$ where $C < +\infty$. Thus, it converges in probability to zero.) This gives us:

$$\hat{\sigma}_R^2 \xrightarrow{p} \hat{\sigma}_{UR}^2$$

But since $\hat{\sigma}_{UR}^2 \xrightarrow{p} \sigma_0^2$, we have that $\hat{\sigma}_R^2 \xrightarrow{p} \sigma_0^2$. Exactly as above, we have that by the Mann-Wald theorem, $\frac{1}{\hat{\sigma}_R^2} \xrightarrow{p} \frac{1}{\sigma_0^2}$ and $\frac{1}{2\left(\hat{\sigma}_R^2\right)^2} \xrightarrow{p} \frac{1}{2\left(\sigma_0^2\right)^2}$. By Slutsky's theorem, then, we have that $\frac{1}{\hat{\sigma}_R^2} \frac{1}{n} \sum_{t=1}^{n} x_t x_t' \xrightarrow{p} \frac{1}{\sigma_0^2} E\left[x_t x_t'\right]$. Therefore, by the less-than-rigorous application of the Cramer-Wold device, $\hat{\Sigma}_R^{ML} \xrightarrow{p} E\left[-H\left(w_t; \theta_0\right)\right]$.

d. Show that the Wald, LM, and LR statistics using $\hat{\Sigma}$ and $\tilde{\Sigma}$ given here in the formulas in Table 7.2 can be written as

$$\begin{aligned}
W &= n \cdot \frac{\left(R\hat{\beta} - c\right)' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta} - c\right)}{SSR_U}, \\
LM &= n \cdot \frac{\left(y - X\tilde{\beta}\right)' P \left(y - X\tilde{\beta}\right)}{SSR_R} \\
LR &= n \cdot \left\{ \log\left(\frac{SSR_R}{n}\right) - \log\left(\frac{SSR_U}{n}\right) \right\},
\end{aligned}$$

where y $(n \times 1)$ and X $(n \times K)$ are the data vector and matrix associated with y_t and x_t, and $P = X\left(X'X\right)^{-1} X'$. Hint: The $A(\theta)$ $(r \times (K+1))$ in Table 7.2 is

$$\begin{pmatrix} \underset{(r \times K)}{R} & \underset{(r \times 1)}{0} \end{pmatrix}.$$

Solution Table 7.2 lists the three tests:

$$\begin{aligned}
W &= na\left(\hat{\theta}\right)' \left[A\left(\hat{\theta}\right) \hat{\Sigma}^{-1} A\left(\hat{\theta}\right)'\right]^{-1} a\left(\hat{\theta}\right) \\
LM &= n \left(\frac{\partial Q_n\left(\tilde{\theta}\right)}{\partial \theta}\right)' \tilde{\Sigma}^{-1} \left(\frac{\partial Q_n\left(\tilde{\theta}\right)}{\partial \theta}\right) \\
LR &= 2n \cdot \left[Q_n\left(\hat{\theta}\right) - Q_n\left(\tilde{\theta}\right)\right]
\end{aligned}$$

In my notation, these are:

$$\begin{aligned}
W &= na\left(\hat{\theta}_{UR}^{ML}\right)' \left[A\left(\hat{\theta}_{UR}^{ML}\right) \hat{\Sigma}_{UR}^{-1} A\left(\hat{\theta}_{UR}^{ML}\right)'\right]^{-1} a\left(\hat{\theta}_{UR}^{ML}\right) \\
LM &= n \left(\frac{\partial Q_n\left(\hat{\theta}_R^{ML}\right)}{\partial \theta}\right)' \hat{\Sigma}_R^{-1} \left(\frac{\partial Q_n\left(\hat{\theta}_R^{ML}\right)}{\partial \theta}\right) \\
LR &= 2n \cdot \left[Q_n\left(\hat{\theta}_{UR}^{ML}\right) - Q_n\left(\hat{\theta}_R^{ML}\right)\right]
\end{aligned}$$

Where $\theta = \begin{bmatrix} \beta & \sigma^2 \end{bmatrix}$ (and therefore, $\hat{\theta}_{UR}^{ML} = \begin{bmatrix} \hat{\beta}_{UR}^{ML} & \hat{\sigma}_{UR}^2 \end{bmatrix}$ and $\hat{\theta}_{R}^{ML} = \begin{bmatrix} \hat{\beta}_{R}^{ML} & \hat{\sigma}_{R}^2 \end{bmatrix}$). For this problem, we have that $H_0 : R\beta = c$ or equivalently, $H_0 : R\beta - c = 0$ or $H_0 : a(\theta) = 0$ where $a(\theta) = R\beta - c$. We can easily derive $A(\theta)$ by:

$$\begin{aligned} A(\theta) &= \frac{\partial}{\partial \theta} a(\theta) = \frac{\partial}{\partial \begin{bmatrix} \beta & \sigma^2 \end{bmatrix}} (R\beta - c) \\ &= \begin{bmatrix} \frac{\partial}{\partial \beta}(R\beta - c) & \frac{\partial}{\partial \sigma^2}(R\beta - c) \end{bmatrix} \\ &= \begin{bmatrix} R & 0 \end{bmatrix} \end{aligned}$$

Since R is not a function of θ, we have that $A\left(\hat{\theta}_{UR}^{ML}\right) = \begin{bmatrix} R & 0 \end{bmatrix}$.

With these preliminaries out of the way, I will proceed to show the result for W. First, it is given that

$$\hat{\Sigma}_{UR}^{ML} = \begin{bmatrix} \frac{1}{\hat{\sigma}_{UR}^2} \frac{1}{n} \sum_{t=1}^n x_t x_t' & 0 \\ 0 & \frac{1}{2\left(\hat{\sigma}_{UR}^2\right)^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{n\hat{\sigma}_{UR}^2} X'X & 0 \\ 0 & \frac{1}{2\left(\hat{\sigma}_{UR}^2\right)^2} \end{bmatrix}$$

Recall the result that if A and B are invertible matrices and $C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$, then C is invertible and $C^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}$. Using this, we have that

$$\hat{\Sigma}_{UR}^{-1} = \begin{bmatrix} n\hat{\sigma}_{UR}^2 (X'X)^{-1} & 0 \\ 0 & 2\left(\hat{\sigma}_{UR}^2\right)^2 \end{bmatrix}$$

Now it just remains to substitute in all the relevant values:

$$\begin{aligned} W &= na\left(\hat{\theta}_{UR}^{ML}\right)' \left[A\left(\hat{\theta}_{UR}^{ML}\right) \hat{\Sigma}_{UR}^{-1} A\left(\hat{\theta}_{UR}^{ML}\right)'\right]^{-1} a\left(\hat{\theta}_{UR}^{ML}\right) \\ &= n\left(R\hat{\beta}_{UR}^{ML} - c\right)' \left[\begin{bmatrix} R & 0 \end{bmatrix} \begin{bmatrix} n\hat{\sigma}_{UR}^2 (X'X)^{-1} & 0 \\ 0 & 2\left(\hat{\sigma}_{UR}^2\right)^2 \end{bmatrix} \begin{bmatrix} R' \\ 0 \end{bmatrix}\right]^{-1} \left(R\hat{\beta}_{UR}^{ML} - c\right) \\ &= n\left(R\hat{\beta}_{UR}^{ML} - c\right)' \left[Rn\hat{\sigma}_{UR}^2 (X'X)^{-1} R'\right]^{-1} \left(R\hat{\beta}_{UR}^{ML} - c\right) \\ &= n\left(R\hat{\beta}_{UR}^{ML} - c\right)' \left[Rn\frac{SSR_{UR}}{n} (X'X)^{-1} R'\right]^{-1} \left(R\hat{\beta}_{UR}^{ML} - c\right) \\ &= \frac{n\left(R\hat{\beta}_{UR}^{ML} - c\right)' \left[R(X'X)^{-1} R'\right]^{-1} \left(R\hat{\beta}_{UR}^{ML} - c\right)}{SSR_{UR}} \end{aligned}$$

This establishes the first result. We now show the result for the LM statistic. Using the same argument as above, recall that:

$$\hat{\Sigma}_{R}^{ML} = \begin{bmatrix} \frac{1}{\hat{\sigma}_{R}^2} \frac{1}{n} \sum_{t=1}^n x_t x_t' & 0 \\ 0 & \frac{1}{2\left(\hat{\sigma}_{R}^2\right)^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{n\hat{\sigma}_{R}^2} X'X & 0 \\ 0 & \frac{1}{2\left(\hat{\sigma}_{R}^2\right)^2} \end{bmatrix}$$

And therefore,

$$\hat{\Sigma}_{R}^{-1} = \begin{bmatrix} n\hat{\sigma}_{R}^2 (X'X)^{-1} & 0 \\ 0 & 2\left(\hat{\sigma}_{R}^2\right)^2 \end{bmatrix}$$

Next, we have that

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

The first derivatives are:

$$\begin{aligned} \frac{\partial}{\partial\beta}Q_n(\theta) &= -\frac{1}{n}\sum_{t=1}^{n}\frac{(y_t - x_t'\beta)\,x_t}{\sigma^2} \\ \frac{\partial}{\partial\sigma^2}Q_n(\theta) &= -\frac{1}{2\sigma^2} + \frac{1}{n}\sum_{t=1}^{n}\frac{(y_t - x_t'\beta)^2}{2\left(\sigma^2\right)^2} \end{aligned}$$

Evaluating these at $\hat{\theta}_R^{ML}$, we have:

$$\begin{aligned} \frac{\partial}{\partial\beta}Q_n\left(\hat{\theta}_R^{ML}\right) &= -\frac{1}{n\hat{\sigma}_R^2}\sum_{t=1}^{n}x_t\left(y_t - x_t'\hat{\beta}_R^{ML}\right) \\ &= -\frac{1}{n\hat{\sigma}_R^2}\left(X'Y - X'X\hat{\beta}_R^{ML}\right) \\ &= -\frac{1}{SSR_R}\left(X'Y - X'X\hat{\beta}_R^{ML}\right) \end{aligned}$$

And

$$\begin{aligned} \frac{\partial}{\partial\sigma^2}Q_n\left(\hat{\theta}_R^{ML}\right) &= -\frac{1}{2\hat{\sigma}_R^2} + \frac{1}{2n\left(\hat{\sigma}_R^2\right)^2}\sum_{t=1}^{n}\left(y_t - x_t'\hat{\beta}_R^{ML}\right)^2 \\ &= -\frac{n}{2SSR_R} + \frac{1}{2n\left(\frac{SSR_R}{n}\right)^2}SSR_R \\ &= -\frac{1}{2}\frac{n}{SSR_R} + \frac{1}{2}\frac{n}{(SSR_R)} \\ &= 0 \end{aligned}$$

Therefore, we have that:

$$\begin{aligned} \frac{\partial}{\partial\theta}Q_n\left(\hat{\theta}_R^{ML}\right) &= \frac{\partial}{\partial\begin{bmatrix}\beta\\ \sigma^2\end{bmatrix}}Q_n\left(\hat{\theta}_R^{ML}\right) \\ &= \begin{bmatrix} \frac{\partial}{\partial\beta}Q_n\left(\hat{\theta}_R^{ML}\right) \\ \frac{\partial}{\partial\sigma^2}Q_n\left(\hat{\theta}_R^{ML}\right) \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{SSR_R}\left(X'Y - X'X\hat{\beta}_R^{ML}\right) \\ 0 \end{bmatrix} \end{aligned}$$

Plugging in these results gives us:

$$
\begin{aligned}
LM &= n\left(\frac{\partial Q_n\left(\hat{\theta}_R^{ML}\right)}{\partial\theta}\right)'\hat{\Sigma}_R^{-1}\left(\frac{\partial Q_n\left(\hat{\theta}_R^{ML}\right)}{\partial\theta}\right)\\
&= n\cdot\begin{bmatrix} -\frac{\left(X'Y-X'X\hat{\beta}_R^{ML}\right)'}{SSR_R} & 0\end{bmatrix}\begin{bmatrix} n\hat{\sigma}_R^2\left(X'X\right)^{-1} & 0\\ 0 & 2\left(\hat{\sigma}_R^2\right)^2\end{bmatrix}\begin{bmatrix} -\frac{\left(X'Y-X'X\hat{\beta}_R^{ML}\right)}{SSR_R}\\ 0\end{bmatrix}\\
&= n\frac{\left(X'Y-X'X\hat{\beta}_R^{ML}\right)'}{SSR_R}\,n\underbrace{\hat{\sigma}_R^2}_{\frac{SSR_R}{n}}\left(X'X\right)^{-1}\frac{\left(X'Y-X'X\hat{\beta}_R^{ML}\right)}{SSR_R}\\
&= n\frac{\left(X'Y-X'X\hat{\beta}_R^{ML}\right)'\left(X'X\right)^{-1}\left(X'Y-X'X\hat{\beta}_R^{ML}\right)}{SSR_R}\\
&= n\frac{\left(Y-X\hat{\beta}_R^{ML}\right)'X\left(X'X\right)^{-1}X'\left(Y-X\hat{\beta}_R^{ML}\right)}{SSR_R}\\
&= n\frac{\left(Y-X\hat{\beta}_R^{ML}\right)'P_X\left(Y-X\hat{\beta}_R^{ML}\right)}{SSR_R}
\end{aligned}
$$

Finally, for the LR statistic, recall from part (b) that

$$
\begin{aligned}
Q_n\left(\hat{\theta}_{UR}^{ML}\right) &= -\frac{1}{2}\log\left(2\pi\right)-\frac{1}{2}-\frac{1}{2}\log\left(\frac{SSR_{UR}}{n}\right)\\
Q_n\left(\hat{\theta}_{R}^{ML}\right) &= -\frac{1}{2}\log\left(2\pi\right)-\frac{1}{2}-\frac{1}{2}\log\left(\frac{SSR_{R}}{n}\right)
\end{aligned}
$$

Therefore, we have:

$$
\begin{aligned}
LR &= 2n\cdot\left[Q_n\left(\hat{\theta}_{UR}^{ML}\right)-Q_n\left(\hat{\theta}_{R}^{ML}\right)\right]\\
&= 2n\cdot\left\{-\frac{1}{2}\log\left(2\pi\right)-\frac{1}{2}-\frac{1}{2}\log\left(\frac{SSR_{UR}}{n}\right)-\left[-\frac{1}{2}\log\left(2\pi\right)-\frac{1}{2}-\frac{1}{2}\log\left(\frac{SSR_{R}}{n}\right)\right]\right\}\\
&= n\left\{\log\left(\frac{SSR_R}{n}\right)-\log\left(\frac{SSR_{UR}}{n}\right)\right\}
\end{aligned}
$$

e. Show that the three statistics can also be written as

$$
\begin{aligned}
W &= n\cdot\frac{SSR_R-SSR_U}{SSR_U},\\
LM &= n\cdot\frac{SSR_R-SSR_U}{SSR_R},\\
LR &= n\cdot\log\left(\frac{SSR_R}{SSR_U}\right).
\end{aligned}
$$

Hint: As we have shown in an analytical exercise of Chapter 1,

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

Solution Appealing to the second equality in the hint, we have:

$$SSR_R - SSR_{UR} = \left(R\hat{\beta}_{UR}^{ML} - c\right)' \left[R\left(X'X\right)^{-1} R'\right]^{-1} \left(R\hat{\beta}_{UR}^{ML} - c\right) = \frac{SSR_{UR}}{n} W$$

Solving for W, we have:

$$W = n\frac{SSR_R - SSR_{UR}}{SSR_{UR}}$$

Which establishes the first result. For the second result, we use the third equality of the hint:

$$SSR_R - SSR_{UR} = \left(Y - X\hat{\beta}_R^{ML}\right) P_X \left(Y - X\hat{\beta}_R^{ML}\right) = \frac{SSR_R}{n} LM$$

Solving for LM, we have:

$$LM = n\frac{SSR_R - SSR_{UR}}{SSR_R}$$

This establishes the second result. Finally, we have:

$$\begin{aligned} LR &= n\left\{\log\left(\frac{SSR_R}{n}\right) - \log\left(\frac{SSR_{UR}}{n}\right)\right\} \\ &= n\left\{\log\left(\frac{\frac{SSR_R}{n}}{\frac{SSR_{UR}}{n}}\right)\right\} = n\left\{\log\left(\frac{SSR_R}{SSR_{UR}}\right)\right\} \end{aligned}$$

Which is the third result.

f. Show that $W \geq LR \geq LM$. (These inequalities do not always hold in nonlinear regression models.)

Solution First, note that $W \geq LR \geq LM$ if and only if $\frac{W}{n} \geq \frac{LR}{n} \geq \frac{LM}{n}$, so I will show this instead.

$$\begin{aligned} \frac{W}{n} &= \frac{SSR_R - SSR_{UR}}{SSR_{UR}} = \frac{SSR_R}{SSR_{UR}} - 1 \\ \frac{LR}{n} &= \log\left(\frac{SSR_R}{SSR_{UR}}\right) \\ \frac{LM}{n} &= \frac{SSR_R - SSR_{UR}}{SSR_R} = 1 - \frac{SSR_{UR}}{SSR_R} = 1 - \frac{1}{\frac{SSR_R}{SSR_{UR}}} \end{aligned}$$

We have then that

$$\begin{aligned} \frac{W}{n} - \frac{LR}{n} &= \frac{SSR_R}{SSR_{UR}} - \left(1 + \log\frac{SSR_R}{SSR_{UR}}\right) \\ &= \frac{SSR_R}{SSR_{UR}} - \log\frac{SSR_R}{SSR_{UR}} - 1 \equiv f\left(\frac{SSR_R}{SSR_{UR}}\right) \end{aligned}$$

Characterizing this function further, we have that

$$f(1) = 1 - \log(1) - 1 = 1 - 0 - 1 = 0$$

And

$$\frac{df}{d\left(\frac{SSR_R}{SSR_{UR}}\right)} = 1 - \frac{1}{\frac{SSR_R}{SSR_{UR}}} \geq 0$$

Whenever

$$\begin{aligned} 1 &\geq \frac{1}{\frac{SSR_R}{SSR_{UR}}} \\ \frac{SSR_R}{SSR_{UR}} &\geq 1 \end{aligned}$$

Which always holds. Therefore, we have that $\frac{W}{n} - \frac{LR}{n} \geq 0$ or $\frac{W}{n} \geq \frac{LR}{n}$

Next, we have

$$\begin{aligned} \frac{LR}{n} - \frac{LM}{n} &= \log\left(\frac{SSR_R}{SSR_{UR}}\right) - \left(1 - \frac{1}{\frac{SSR_R}{SSR_{UR}}}\right) \\ &= \log\left(\frac{SSR_R}{SSR_{UR}}\right) + \frac{1}{\frac{SSR_R}{SSR_{UR}}} - 1 \equiv g\left(\frac{SSR_R}{SSR_{UR}}\right) \end{aligned}$$

Proceeding similarly, we have:

$$g(1) = \log(1) + \frac{1}{1} - 1 = 0 + 1 - 1 = 0$$

And

$$\frac{dg}{d\left(\frac{SSR_R}{SSR_{UR}}\right)} = \frac{1}{\frac{SSR_R}{SSR_{UR}}} - \frac{1}{\left(\frac{SSR_R}{SSR_{UR}}\right)^2} \geq 0$$

Whenever

$$\begin{aligned} \frac{1}{\frac{SSR_R}{SSR_{UR}}} &\geq \frac{1}{\left(\frac{SSR_R}{SSR_{UR}}\right)^2} \\ \frac{SSR_R}{SSR_{UR}} &\geq 1 \end{aligned}$$

Which always holds. Therefore, we have that $\frac{LR}{n} - \frac{LM}{n} \geq 0$ or $\frac{LR}{n} \geq \frac{LM}{n}$. Putting this all together, we get the result that $W \geq LR \geq LM$.

2 Extra Problems

1. Provide primitive conditions for consistency and asymptotic normality of the NLS and WNLS estimator of β_0 in the logit model:

$$\Pr\left(Y_i = 1 | X_i\right) = \frac{\exp\left(X_i \beta_0\right)}{1 + \exp\left(X_i \beta_0\right)}$$

and derive their asymptotic distributions.

Solution The standard assumptions ensuring consistency and asymptotic normality of the NLS estimator are, divided into three categories:

Model Assumption 1 $Y_i = g\left(X_i, \beta_0\right) + \varepsilon_i$, $i = 1, \ldots, n$ where $g\left(X_i, \beta\right)$ is a known function, X_i is an L dimensional random vector, and β_0 is a vector of unknown parameters belonging to a set $B \subset \mathbb{R}^K$

Model Assumption 2 $\{(X_i, \varepsilon_i)\}_{i=1}^{n}$ are i.i.d.

Model Assumption 3 $E\left[\varepsilon_i | X_i\right] = 0$, $i = 1, \ldots, n$.

Model Assumption 4 $g\left(X_i, \beta_0\right) \neq g\left(X_i, \beta\right)$ for all $\beta \neq \beta_0$.

Consistency Assumption 1 $B \subset \mathbb{R}^K$ is compact. (i.e. closed and bounded.)

Consistency Assumption 2 $g\left(X_i, \beta\right)$ is continuous in β for any X_i.

Consistency Assumption 3 $\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - g\left(X_i, \beta\right)\right)^2 \xrightarrow{p} E\left[\left(Y_i - g\left(X_i, \beta\right)\right)^2\right]$ uniformly in β.

Normality Assumption 1 $\beta_0 \in int\left(B\right)$.

Normality Assumption 2 $g\left(X_i, \beta\right)$ twice continuously differentiable in β for any X_i.

Normality Assumption 3 The matrix

$$\Omega_0 = E\left[\left.\frac{\partial g\left(X_i, \beta\right)}{\partial \beta}\right|_{\beta_0} \left.\frac{\partial g\left(X_i, \beta\right)}{\partial \beta'}\right|_{\beta_0} \left(Y_i - g\left(X_i, \beta_0\right)\right)^2\right]$$

Is finite and positive definite.

Normality Assumption 4 For some neighborhood $\mathcal{N}$ of β_0,

$$E\left[\sup_{\beta \in \mathcal{N}} ||H\left(X_i, Y_i, \beta\right)||\right] < +\infty$$

Normality Assumption 5 $H_0 = E\left[H\left(X_i, Y_i, \beta_0\right)\right]$ is nonsingular.

Before proceeding to confirm these assumptions and discern which are primitive (not implied by the logit model), I will derive the general asymptotic distribution for a NLS estimator and then apply it to the special case of the logit model. First, note that if

$$\begin{aligned}
\hat{\beta}_n^{NLS} &= \arg\min_{\beta} \sum_{i=1}^{n}\left(Y_i - g\left(X_i, \beta\right)\right)^2 \\
&= \arg\min_{\beta} \frac{1}{n}\sum_{i=1}^{n}\underbrace{\frac{1}{2}\left(Y_i - g\left(X_i, \beta\right)\right)^2}_{m\left(X_i, Y_i, \beta\right)}
\end{aligned}$$

Where $g(X_i,\beta) = E[Y_i|X_i]$. For this particular model, we have:

$$g(X_i,\beta) = E[Y_i|X_i] = \Pr(Y_i = 1|X_i) = \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)}$$

Then we have that $\frac{1}{n}\sum_{i=1}^{n} s\left(X_i, Y_i, \hat{\beta}_n^{NLS}\right) = \frac{\partial}{\partial\beta}\frac{1}{n}\sum_{i=1}^{n} m\left(X_i, Y_i, \hat{\beta}_n^{NLS}\right) = 0$. By the mean value theorem, if we expand this expression around β_0, which is possible since $g(X_i,\beta)$ is assumed to be twice continuously differentiable, we have:

$$0 = \frac{1}{n}\sum_{i=1}^{n} s\left(X_i, Y_i, \hat{\beta}_n^{NLS}\right) = \frac{1}{n}\sum_{i=1}^{n} s(X_i, Y_i, \beta_0) + \left(\frac{1}{n}\sum_{i=1}^{n} H\left(X_i, Y_i, \tilde{\beta}\right)\right)\left(\hat{\beta}_n^{NLS} - \beta_0\right)$$

Where $\tilde{\beta} \in \left|\hat{\beta}_n^{NLS}, \beta_0\right|$. Solving for $\left(\hat{\beta}_n^{NLS} - \beta_0\right)$ and premultiplying by $\sqrt{n}$, we get:

$$\sqrt{n}\left(\hat{\beta}_n^{NLS} - \beta_0\right) = \left(\frac{1}{n}\sum_{i=1}^{n} -H\left(X_i, Y_i, \tilde{\beta}\right)\right)^{-1}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} s(X_i, Y_i, \beta_0)\right)$$

Assuming for the moment that all the necessary assumptions hold, we then have that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} s(X_i, Y_i, \beta_0)\right) \xrightarrow{d} N\left(0, E\left[s(X_i, Y_i, \beta_0)\, s(X_i, Y_i, \beta_0)'\right]\right)$$

By the central limit theorem and

$$\left(\frac{1}{n}\sum_{i=1}^{n} -H\left(X_i, Y_i, \tilde{\beta}\right)\right)^{-1} \xrightarrow{p} \left[E\left[-H(X_i, Y_i, \beta_0)\right]\right]^{-1}$$

By the uniform law of large numbers and the Mann-Wald theorem. Then, by Slutsky's theorem, we have:

$$\sqrt{n}\left(\hat{\beta}_n^{NLS} - \beta_0\right) \xrightarrow{d} N\left(0, AV\left(\hat{\beta}_n^{NLS}\right)\right)$$

Where

$$\begin{aligned}
AV\left(\hat{\beta}_n^{NLS}\right) &= \left[\underbrace{E\left[H(X_i, Y_i, \beta_0)\right]}_{\equiv H_0}\right]^{-1}\left[\underbrace{E\left[s(X_i, Y_i, \beta_0)\, s(X_i, Y_i, \beta_0)\right]}_{\equiv \Omega_0}\right]\left[\underbrace{E\left[H(X_i, Y_i, \beta_0)\right]}_{\equiv H_0}\right]^{-1} \\
&= H_0^{-1}\Omega_0 H_0^{-1}
\end{aligned}$$

It remains only to solve for these values (here, $g(X_i,\beta) = \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)}$)

$$\begin{aligned}
s(X_i, Y_i, \beta_0) &= \left.\frac{\partial}{\partial\beta} m(X_i, Y_i, \beta)\right|_{\beta_0} = \left.\frac{\partial}{\partial\beta}\frac{1}{2}\left(Y_i - g(X_i,\beta)\right)^2\right|_{\beta_0} = \left.-\left(Y_i - g(X_i,\beta)\right)\frac{\partial g}{\partial\beta}\right|_{\beta_0} \\
&= \left.\frac{\partial}{\partial\beta}\left(Y_i - \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)}\right)^2\right|_{\beta_0} \\
&= -2\left(Y_i - \frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)}\right)\left(\frac{(1+\exp(X_i\beta_0))\exp(X_i\beta_0)X_i' - \left[\exp(X_i\beta_0)\right]^2 X_i'}{(1+\exp(X_i\beta_0))^2}\right) \\
&= -2\left(Y_i - \Lambda(X_i\beta_0)\right)\left(\Lambda(X_i\beta_0)X_i' - \left[\Lambda(X_i\beta_0)\right]^2 X_i'\right)
\end{aligned}$$

Where $\Lambda(X_i\beta) \equiv \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)}$. And,

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

Which, for sanity's sake, I will not derive explicitly. Next, deriving with the weighted nonlinear least squares estimator, we begin by noting that for the model

$$Y_i = g(X_i, \beta) + \varepsilon_i$$

We have that

$$\begin{aligned} \varepsilon_i &= Y_i - g(X_i, \beta) \\ &= Y_i - E[Y_i | X_i] \end{aligned}$$

And recognizing that $\varepsilon_i = \begin{cases} 1 - E[Y_i | X_i] & \text{if } Y_i = 1 \\ -E[Y_i | X_i] & \text{if } Y_i = 0 \end{cases} = \begin{cases} 1 - \frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)} & \text{if } Y_i = 1 \\ -\frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)} & \text{if } Y_i = 0 \end{cases}$

And

$$\Pr(\varepsilon_i = 1 - E[Y_i | X_i] | X_i) = \Pr(Y_i = 1 | X_i) = E[Y_i | X_i] = \frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)}$$

And, of course

$$\Pr(\varepsilon_i = -E[Y_i | X_i] | X_i) = \Pr(Y_i = 0 | X_i) = 1 - E[Y_i | X_i] = \frac{1}{1+\exp(X_i\beta_0)}$$

Therefore, we have since $\varepsilon_i | X_i \sim Bernoulli\left(\frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)}\right)$

$$Var(\varepsilon_i | X_i) = \frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)}\left(1 - \frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)}\right)$$

That is, we have heteroskedasticity. Our weighted nonlinear least squares estimator then becomes:

$$\begin{aligned} \hat{\beta}_n^{WNLS} &= \arg\min_{\beta} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\frac{(Y_i - g(X_i, \beta))^2}{Var(Y_i - g(X_i, \beta))} \\ &= \arg\min_{\beta} \frac{1}{n}\sum_{i=1}^{n} \underbrace{\frac{1}{2}\frac{\left(Y_i - \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)}\right)^2}{\frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)}\left(1 - \frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)}\right)}}_{m(X_i, Y_i, \beta)} \end{aligned}$$

The only difference between this estimator and the NLS estimator is that $m(X_i, Y_i, \beta)$ is a different function. By the argument given above, the asymptotic distribution is given by:

$$\sqrt{n}\left(\hat{\beta}_n^{WNLS} - \beta_0\right) \xrightarrow{d} N\left(0, AV\left(\hat{\beta}_n^{WNLS}\right)\right)$$

Where

$$\begin{aligned} AV\left(\hat{\beta}_n^{WNLS}\right) &= \left[\underbrace{E[H(X_i, Y_i, \beta_0)]}_{\equiv H_0}\right]^{-1}\left[\underbrace{E\left[s(X_i, Y_i, \beta_0)\, s(X_i, Y_i, \beta_0)'\right]}_{\equiv \Omega_0}\right]\left[\underbrace{E[H(X_i, Y_i, \beta_0)]}_{\equiv H_0}\right]^{-1} \\ &= H_0^{-1}\Omega_0 H_0^{-1} \end{aligned}$$

Clearly, since the m function for NLS is not the same as that for $WNLS$, we have that $AV\left(\hat{\beta}_n^{NLS}\right) \neq AV\left(\hat{\beta}_n^{WNLS}\right)$. I will not derive the exact asymptotic distribution, however.

Next, I hope to derive which of the twelve assumptions are implied by the model and which are "primitive" assumptions that we cannot "check." It turns out that precisely the same assumptions required for consistency and asymptotic normality of the NLS estimator ensure consistency and asymptotic normality of the WNLS estimator. Therefore, I will only look at the NLS case.

First we have that, clearly, Model Assumptions 1, 2, and 3 are satisfied if we assume that $\{(X_i, \varepsilon_i)\}$ are i.i.d.. (which must then be a primitive assumption.) Checking Model Assumption 4, we have that if $\beta \neq \beta_0$, then

$$\begin{aligned} g(X_i,\beta) - g(X_i,\beta_0) &= \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} - \frac{\exp(X_i\beta_0)}{1+\exp(X_i\beta_0)} \\ &= \frac{\exp(X_i\beta)\left[1+\exp(X_i\beta_0)\right] - \exp(X_i\beta_0)\left[1+\exp(X_i\beta)\right]}{\left[1+\exp(X_i\beta)\right]\left[1+\exp(X_i\beta_0)\right]} \\ &= \frac{\exp(X_i\beta) - \exp(X_i\beta_0)}{\left[1+\exp(X_i\beta)\right]\left[1+\exp(X_i\beta_0)\right]} \end{aligned}$$

Which is equal to zero if and only if $\exp(X_i\beta) = \exp(X_i\beta_0)$ for all X_i, which occurs if and only if $X_i\beta = X_i\beta_0$ for all X_i, which occurs if and only if $\beta = \beta_0$. Therefore, assumption 4 is satisfied.

Consistency Assumption 1 does not hold in general, but we will assume it here: Assume that $\exists N$ large such that $B \subset [-N, N]^K$, and assume that B is closed. (This is a primitive assumption.)

Consistency Assumption 2 holds, because we know that $\exp(X_i\beta)$ is continuous in β for all X_i, as is $1+\exp(X_i\beta)$ for all X_i and is strictly positive for all X_i. There is a theorem in real analysis that states that the ratio of continuous functions is also continuous. Therefore,

$$g(X_i,\beta) = \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)}$$

Is continuous in β for all X_i.

Consistency Assumption 3 is not easily verifiable, but we will assume it holds. (Primitive assumption) The last part of Consistency Assumption 3 holds by Model Assumption 4, however.

Under the above assumptions, we have that $\hat{\beta}_n^{NLS} \xrightarrow{p} \beta_0$ and $\hat{\beta}_n^{WNLS} \xrightarrow{p} \beta_0$ and are therefore consistent for β_0.

Next, moving on to asymptotic normality, we have that Normality Assumption 1 holds trivially, since if $\beta_0 \in B \backslash int(B)$, then we can define a $B' \supset B$ such that $\beta_0 \in int(B')$. (As no additional restrictions have been placed on our B.)

Normality Assumption 2 holds since both $\exp(X_i\beta)$ and $1+\exp(X_i\beta)$ are both C^2 functions and therefore any ratio of C^2 functions is also C^2.

In order to check Normality Assumption 3, we must first compute Ω_0 :

$$\begin{aligned}
\Omega_0 &= E\left[s(X_i,Y_i,\beta_0)\cdot s(X_i,Y_i,\beta_0)'\right] \\
&= E\left[(Y_i-\Lambda(X_i\beta_0))^2\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)'\right] \\
&= E\left[\varepsilon_i^2\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)'\right] \\
&= E\left[E\left[\varepsilon_i^2\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)'\middle| X_i\right]\right] \\
&= E\left[\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)'E\left[\varepsilon_i^2\middle| X_i\right]\right] \\
&= E\left[\sigma_i^2\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)'\right]
\end{aligned}$$

(Note that if we had conditional homoskedasticity, then we would have had that $\sigma_i^2=\sigma^2$ for all i in which case, we have

$$\Omega_0=\sigma^2E\left[\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)\left(\Lambda(X_i\beta_0)X_i'-(\Lambda(X_i\beta_0))^2X_i'\right)'\right]$$

This matrix is positive definite, since it is a variance-covariance matrix of non-degenerate random variables. We still need to assume that it is finite. That is $\Omega_0<+\infty$ is a primitive assumption.

Normality assumption 4 will be taken as primitive (since it is certainly not easy to check.)

Finally, for normality assumption 5, we have:

$$\begin{aligned}
H_0 &= E\left[H(X_i,Y_i,\beta_0)\right]=E\left[\frac{\partial g(X_i,\beta)}{\partial\beta}\frac{\partial g(X_i,\beta)}{\partial\beta'}-\frac{\partial^2 g(X_i,\beta)}{\partial\beta\partial\beta'}(Y_i-g(X_i,\beta))\bigg|_{\beta_0}\right] \\
&= E\left[\frac{\partial g(X_i,\beta)}{\partial\beta}\frac{\partial g(X_i,\beta)}{\partial\beta'}\bigg|_{\beta_0}\right]-E\left[\frac{\partial^2 g(X_i,\beta)}{\partial\beta\partial\beta'}(Y_i-g(X_i,\beta))\bigg|_{\beta_0}\right] \\
&= E\left[\frac{\partial g(X_i,\beta)}{\partial\beta}\frac{\partial g(X_i,\beta)}{\partial\beta'}\bigg|_{\beta_0}\right]
\end{aligned}$$

Since

$$\begin{aligned}
E\left[\frac{\partial^2 g(X_i,\beta)}{\partial\beta\partial\beta'}(Y_i-g(X_i,\beta))\bigg|\beta_0\right] &= E\left[E\left[\frac{\partial^2 g(X_i,\beta)}{\partial\beta\partial\beta'}\varepsilon_i\bigg|_{\beta_0} X_i\right]\right] \\
&= E\left[E\left[\varepsilon_i| X_i\right]\frac{\partial^2 g(X_i,\beta)}{\partial\beta\partial\beta'}\bigg|_{\beta_0}\right]=0
\end{aligned}$$

We will need to assume then that $H_0=E\left[\frac{\partial g(X_i,\beta)}{\partial\beta}\frac{\partial g(X_i,\beta)}{\partial\beta'}\Big|_{\beta_0}\right]$ is nonsingular. (And thus, this is a primitive assumption.

2. This problem uses the data set MROZ.ASC. the data set is described in the file READMROZ.DOC. It contains 753 observations on women in 1975. Estimate an LPM, Logit, and Probit model of married women's labor force participation as a function of the variables KL6, K618, WA, WE, UN, CIT, which are provided in the data set, a variable say PRIN (property income) generated as FAMINC-(WHRS·WW), and an additional variable, say LWW constructed as follows: Restricting your sample to workers (i.e. the first 428 observations) take the natural logarithm of the wife's wage rate variable WW. Call this new variable LW. Then, for the entire sample construct the square of the wife's experience variable, say EX2. Next, using only the 428 observations from the working sample, estimate OLS, a typical wage determination equation in which LW is regressed on a constant, WA, WE, CIT, EX, and EX2. Then use the parameter estimates of this model for the 325 non-working women to generate predicted or fitted log-wages, say FLW. The variable LWW is just LW for the working sample and FLW for the non-working sample.

a. For each model provide both the estimates of the unknown coefficients of interest and their standard errors. How do they compare across the different models after you make the appropriate rescalings?

b. Estimate the marginal effect of the variable LWW (evaluated at the sample average of the regressors) and provide an estimate of its standard error. Make sure you justify the computation of the latter.

Solution See the attached pages.