Cornell University
Department of Economics

Econ 620 – Spring 2007
Instructor: Professor Kiefer
TA: Jae Ho Yun

Suggested Solutions for Midterm Exam 2007

1. We have to estimate the following model:

$$\begin{bmatrix} y \\ y^* \end{bmatrix} = \begin{bmatrix} X \\ X^* \end{bmatrix}\beta + \begin{bmatrix} \varepsilon \\ \varepsilon^* \end{bmatrix}$$
$$\text{where } \varepsilon \sim N(0, \sigma^2 I_N), \quad \varepsilon^* \sim N(0, I_{n_0})$$

(a) Is $\widetilde{\beta}$ linear?

$$\begin{aligned} \widetilde{\beta} &= \left(\begin{bmatrix} X \\ X^* \end{bmatrix}' \begin{bmatrix} X \\ X^* \end{bmatrix} \right)^{-1} \begin{bmatrix} X \\ X^* \end{bmatrix}' \begin{bmatrix} y \\ y^* \end{bmatrix} \\ &= (X'X + X^{*\prime}X^*)^{-1}(X'y + X^{*\prime}y^*) \\ &= (X'X + X^{*\prime}X^*)^{-1}X'y + (X'X + X^{*\prime}X^*)^{-1}X^{*\prime}y^* \end{aligned}$$

Since $\widetilde{\beta}$ can be represented by $\widetilde{\beta} = Ay + B$(for some matrix A and B), $\widetilde{\beta}$ is affine in y. Strictly speaking, we can say that this is linear in y, only if $B = 0$.
(grading policy: If you pointed out that $\widetilde{\beta}$ does not involve nonlinear functions of y, then you would be given a full credit even if you said $\widetilde{\beta}$ is linear in y.)

(b) Is $\widetilde{\beta}$ unbiased?

$$\begin{aligned} \widetilde{\beta} &= [X'X + X^{*\prime}X^*]^{-1}[X'y + X^{*\prime}y^*] \\ &= [X'X + X^{*\prime}X^*]^{-1}[X'(X\beta + \varepsilon) + X^{*\prime}(X^*\beta_0 + \varepsilon^*)] \\ &= [X'X + X^{*\prime}X^*]^{-1}[X'X\beta + X'\varepsilon + X^{*\prime}X^*\beta_0 + X^{*\prime}\varepsilon^*] \\ E\widetilde{\beta} &= [X'X + X^{*\prime}X^*]^{-1}[X'X\beta + E(X'\varepsilon) + X^{*\prime}X^*\beta_0 + E(X^{*\prime}\varepsilon^*)] \\ &= [X'X + X^{*\prime}X^*]^{-1}[X'X\beta + X^{*\prime}X^*\beta_0] \end{aligned}$$

Hence, generally $\widetilde{\beta}$ is biased unless β_0 is same as β by sheer luck.

(c) sampling variance of $\widetilde{\beta}$

$$\begin{aligned}
\widetilde{\beta} - E\widetilde{\beta} &= [X'X + X^{*\prime}X^*]^{-1}[X'X\beta + X'\varepsilon + X^{*\prime}X^*\beta_0 + X^{*\prime}\varepsilon^*] \\
&\quad -[X'X + X^{*\prime}X^*]^{-1}[X'X\beta + X^{*\prime}X^*\beta_0] \\
&= [X'X + X^{*\prime}X^*]^{-1}[X'\varepsilon + X^{*\prime}\varepsilon^*] \\
V(\widetilde{\beta}) &= E[(\widetilde{\beta} - E\widetilde{\beta})(\widetilde{\beta} - E\widetilde{\beta})'] \\
&= E[(X'X + X^{*\prime}X^*)^{-1}(X'\varepsilon + X^{*\prime}\varepsilon^*)(X'\varepsilon + X^{*\prime}\varepsilon^*)'(X'X + X^{*\prime}X^*)^{-1}] \\
&= E[(X'X + X^{*\prime}X^*)^{-1}(X'\varepsilon\varepsilon'X + X'\varepsilon\varepsilon^{*\prime}X^{*\prime} + X^{*\prime}\varepsilon^*\varepsilon'X + X^{*\prime}\varepsilon^*\varepsilon'X) \\
&\quad (X'X + X^{*\prime}X^*)^{-1}] \\
&= (X'X + X^{*\prime}X^*)^{-1}(X'E(\varepsilon\varepsilon')X + X'E(\varepsilon\varepsilon^{*\prime})X^{*\prime} + X^{*\prime}E(\varepsilon^*\varepsilon')X + \\
&\quad X^{*\prime}E(\varepsilon^*\varepsilon^{*\prime})X^*)(X'X + X^{*\prime}X^*)^{-1} \\
&= (X'X + X^{*\prime}X^*)^{-1}(\sigma^2 X'X + X^{*\prime}X^*)(X'X + X^{*\prime}X^*)^{-1}
\end{aligned}$$

And it is well known that Covariance matrix of OLS estimator, $V(\widehat{\beta}) = \sigma^2(X'X)^{-1}$.

(d) From b), we have the following.

$$\begin{aligned}
\widetilde{\beta} &= [X'X + X^{*\prime}X^*]^{-1}[X'X\beta + X'\varepsilon + X^{*\prime}X^*\beta_0 + X^{*\prime}\varepsilon^*] \\
&= [\frac{X'X}{N} + \frac{X^{*\prime}X^*}{N}]^{-1}[\frac{X'X}{N}\beta + \frac{X'\varepsilon}{N} + \frac{X^{*\prime}X^*}{N}\beta_0 + \frac{X^{*\prime}\varepsilon^*}{N}] \\
\text{Note that } p\lim\frac{X'X}{N} &= Q, p\lim\frac{X^{*\prime}X^*}{N} = 0 (\text{since } n_0 \text{ is fixed}), \\
p\lim\frac{X'\varepsilon}{N} &= 0, p\lim\frac{X^{*\prime}\varepsilon^*}{N} = 0.
\end{aligned}$$

$$\begin{aligned}
p\lim\widetilde{\beta} &= [p\lim\frac{X'X}{N} + p\lim\frac{X^{*\prime}X^*}{N}]^{-1}[p\lim\frac{X'X}{N}\beta + p\lim\frac{X'\varepsilon}{N} \\
&\quad + p\lim\frac{X^{*\prime}X^*}{N}\beta_0 + p\lim\frac{X^{*\prime}\varepsilon^*}{N}] \\
&= Q^{-1}Q\cdot\beta = \beta
\end{aligned}$$

Therefore, $\widetilde{\beta}$ is consistent.

(e) The Asymptotic Distribution of $\widetilde{\beta}$

First, consider $\sqrt{N}(\widetilde{\beta} - \widehat{\beta})$

$$\begin{aligned}
\sqrt{N}\widetilde{\beta} - \sqrt{N}\widehat{\beta} &= [\frac{X'X}{N} + \frac{X^{*\prime}X^*}{N}]^{-1}[\frac{X'X}{\sqrt{N}}\beta + \frac{X'\varepsilon}{\sqrt{N}} + \frac{X^{*\prime}X^*}{\sqrt{N}}\beta_0 + \frac{X^{*\prime}\varepsilon^*}{\sqrt{N}}] \\
&\quad -[\frac{X'X}{N}]^{-1}[\frac{X'X}{\sqrt{N}}\beta + \frac{X'\varepsilon}{\sqrt{N}}]
\end{aligned}$$

Note that $p\lim \frac{X^{*\prime}X^*}{\sqrt{N}} = 0$ and $p\lim \frac{X^{*\prime}\varepsilon^*}{\sqrt{N}} = 0$ since n_0 is fixed even if n goes to infinity.

$$p\lim \sqrt{N}(\widetilde{\beta} - \widehat{\beta}) = Q^{-1}[p\lim \frac{X^{*\prime}X^*}{\sqrt{N}}\beta_0 + p\lim \frac{X^{*\prime}\varepsilon^*}{\sqrt{N}}] = 0$$

Since $p\lim \sqrt{N}(\widetilde{\beta} - \widehat{\beta}) = \sqrt{N}((\widetilde{\beta} - \beta) - (\widehat{\beta} - \beta)) = 0$, $\sqrt{N}(\widetilde{\beta} - \beta)$ has the same limiting distribution as $\sqrt{N}(\widehat{\beta} - \beta)$. We know that $\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$. Hence, $\sqrt{N}(\widetilde{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$.

(f) Write $\widetilde{\beta}$ as $A\widehat{\beta} + B\beta_0 + C\varepsilon^*$

Since $\widehat{\beta} = \beta + (X'X)^{-1}X'\varepsilon$, we have $\beta = \widehat{\beta} - (X'X)^{-1}X'\varepsilon$.Plug this into the $\widetilde{\beta}$ formula from b).

$$\begin{aligned}
\widetilde{\beta} &= [X'X + X^{*\prime}X^*]^{-1}[X'X\beta + X'\varepsilon + X^{*\prime}X^*\beta_0 + X^{*\prime}\varepsilon^*] \\
&= [X'X + X^{*\prime}X^*]^{-1}[X'X(\widehat{\beta} - (X'X)^{-1}X'\varepsilon) + X'\varepsilon + X^{*\prime}X^*\beta_0 + X^{*\prime}\varepsilon^*] \\
&= [X'X + X^{*\prime}X^*]^{-1}[X'X\widehat{\beta} - X'\varepsilon + X'\varepsilon + X^{*\prime}X^*\beta_0 + X^{*\prime}\varepsilon^*] \\
&= [X'X + X^{*\prime}X^*]^{-1}X'X\widehat{\beta} + [X'X + X^{*\prime}X^*]^{-1}X^{*\prime}X^*\beta_0 + [X'X + X^{*\prime}X^*]^{-1}X^{*\prime}\varepsilon^* \\
& \text{Therefore,} \\
A &= [X'X + X^{*\prime}X^*]^{-1}X'X, \\
B &= [X'X + X^{*\prime}X^*]^{-1}X^{*\prime}X^* \\
C &= [X'X + X^{*\prime}X^*]^{-1}X^{*\prime}
\end{aligned}$$

2. $Ey = \beta_0 + \beta_1 x_1 + \beta_2 x_2$
$\gamma = \beta_1^2 + \beta_2$

(a) An estimator $\widehat{\gamma}$ for γ

First, $\widehat{\beta}_0, \widehat{\beta}_1$ and $\widehat{\beta}_2$ can be obtained by MLE. Then, we can have a consistent estimator, $\widehat{\gamma} = \widehat{\beta}_1^2 + \widehat{\beta}_2$ by applying Slutsky theorem.

(b) Asymptotic distribution of $\widehat{\gamma}$

First, define $\theta_0 = (\beta_{00}, \beta_{10}, \beta_{20})$(true parameters) and $\widehat{\theta} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)$.
We know that $\sqrt{N}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, i(\theta_0)^{-1})$,
where $i(\theta_0) = \frac{1}{\sigma^2}Q$, $Q = p\lim(\frac{X'X}{N})$ and X is a $N \times 3$ matrix whose columns consist of observations for 1, x_1 and x_2 respectively.
For asymptotic distribution of $\widehat{\gamma}$, we can apply delta method.
Define $R(\beta_0, \beta_1, \beta_2) = \beta_1^2 + \beta_2$. Then, we have:

$$\sqrt{N}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} N(0, \frac{\partial R}{\partial \theta'}(\theta_0) i(\theta_0)^{-1} \frac{\partial R}{\partial \theta}(\theta_0))$$

where $\frac{\partial R}{\partial \theta'}(\theta_0) = (0, 2\beta_{10}, 1)$.

Therefore, asymptotic variance of $\sqrt{N}(\widehat{\gamma} - \gamma_0)$ is

$$\sigma^2 \begin{pmatrix} 0 & 2\beta_{10} & 1 \end{pmatrix} Q^{-1} \begin{pmatrix} 0 \\ 2\beta_{10} \\ 1 \end{pmatrix}$$

(c) Test the hypothesis that $\gamma = 0$

Use $LR = 2(l(\widehat{\theta}) - l(\overline{\theta}_0)), LM(Score) = \frac{1}{N} s(\overline{\theta}_0)' i(\overline{\theta}_0) s(\overline{\theta}_0)$ or $Wald = N(\widehat{\theta} - \theta_0)' i(\widehat{\theta})^{-1} (\widehat{\theta} - \theta_0)$ with approximating distribution $\chi^2(1)$. Now, $\widehat{\theta}$ is unrestriced ML estimator and $\overline{\theta}_0$ is restricted ML estimator.

(d) Compare Wald test and Score(LM) test

First, consider the Wald statistics. It can be derived from the limiting distribution of $\sqrt{N}(\widehat{\gamma} - \gamma_0)$ in b) by replacing θ_0 by $\widehat{\theta}$.

$$\sqrt{N}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} N(0, \frac{\partial R}{\partial \theta'}(\theta_0) i(\theta_0)^{-1} \frac{\partial R}{\partial \theta}(\theta_0)) \approx N(0, \frac{\partial R}{\partial \theta'}(\widehat{\theta}) i(\widehat{\theta})^{-1} \frac{\partial R}{\partial \theta}(\widehat{\theta}))$$

where $\frac{\partial R}{\partial \theta'}(\widehat{\theta}) = (0, 2\widehat{\beta}_1, 1)$ and $i(\widehat{\theta})^{-1} = \widehat{\sigma^2}(\frac{X'X}{N})^{-1}$.

Therefore, the Wald statistics is:

$$W = N\widehat{\gamma}^2 [\frac{\partial R}{\partial \theta'}(\widehat{\theta}) i(\widehat{\theta})^{-1} \frac{\partial R}{\partial \theta}(\widehat{\theta})]^{-1} \sim \chi^2(1)$$

Next, consider the Score statistics. For this test, we have to estimate the model under the restricstion. It will be that $y = \beta_0 + \beta_1 x_1 - \beta_1^2 x_2 + \varepsilon$ since $\beta_2 = -\beta_1^2$ under the null hypothesis. Note that this is a nonlinear regression model. So, it is more difficult to estimate this kind of models than ordinary linear models. For the score statistics, plug the restricted estimators(call them $\overline{\theta}$) into score function and construct the score statistics. The resulting statistics is as follows.

$$Score = N(\frac{1}{N} X' \overline{e})' \frac{1}{\overline{\sigma}^2} [(\frac{1}{N} X'X)^{-1}](\frac{1}{N} X' \overline{e}) \sim \chi^2(1)$$

where $\overline{e}$ is a residual vector from restricted model and $\overline{\sigma}^2$ is a corresponding variance estimator.(Note that $\frac{1}{\overline{\sigma}^2} X' \overline{e}$ is a score function)

Therefore, in our case, Wald test is more convenient to implement than the score test.

(e) Compare the tests when $\beta_1 = \beta_2 = 0$.

In this case, Score test is easy to implement. From the Score statistics from d), $\overline{e}$ now becomes a vector of demeaned y under the null hypothesis.

$$Score = N[\frac{1}{N} X'(y - \overline{y} \cdot 1)]' \frac{1}{\overline{\sigma}^2} [(\frac{1}{N} X'X)^{-1}][\frac{1}{N} X'(y - \overline{y} \cdot 1)] \sim \chi^2(2)$$

However, for the Wald test, we have to consider two restrictions. Suppose that $R(\widehat{\theta}) = \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix}$ $\frac{\partial R}{\partial \theta'}(\widehat{\theta}) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $i(\widehat{\theta})^{-1} = \widehat{\sigma^2}(\frac{X'X}{N})^{-1}$. All the estimators came from the unrestricted model. The Wald statistics is as follows.

$$W = N \cdot R(\widehat{\theta})' \left[\frac{\partial R}{\partial \theta'}(\widehat{\theta}) i(\widehat{\theta})^{-1} \frac{\partial R}{\partial \theta}(\widehat{\theta}) \right]^{-1} R(\widehat{\theta}) \sim \chi^2(2)$$

Hence, in this case, score test is easy to implement.