Cornell University
Department of Economics

Econ 620 – Spring 2008
TA: Jae Ho Yun

Suggested Solution for PS #5

1. (Measurement Error, IV)

   (a) This is a measurement error problem.

   $$\begin{aligned} y_i &= x_i'\beta + \delta t_i^* + \varepsilon_i \\ t_i^* &= \alpha t_i + \upsilon_i \end{aligned}$$

   The problem is that ε_i and υ_i might be correlated. If this is the case, we cannot obtain consistent estimators by OLS, which is called endogeneity problem.

   (b) Get $\widehat{t_i^*}$ by regressing t_i^* on z. Then, get $\widehat{\delta}$ from estimating the model, $y_i = x_i'\beta + \delta\widehat{t_i^*} + \varepsilon_i$.

   (c) We can use the Wu-Hasman test. Under the null hypothesis that there is no endogeneity, $\sqrt{N}(\widehat{\delta}_{OLS} - \widehat{\delta}_{2SLS})$ follows normal distribution asymptotically.
   This is equivalent to testing a significance of $\widetilde{\delta}$ in the following eqatation.

   $$y_i = x_i'\beta + \delta t_i^* + \widetilde{\delta} t_i^* + \varepsilon_i$$

   The test-statistics is t-test for $\widetilde{\delta} = 0$.

2. (Probit Model)

   (a) We can obtain NLS estimator by solving the following problem:

   $$\underset{\beta}{Arg\min} \sum_{i=1}^{N} (d_i - \Phi(x_i\beta))^2.$$

   The corresponding First Order Condition (or Normal equations) is as follows:

   $$-2\sum_{i=1}^{N} (d_i - \Phi(x_i\beta))\phi(x_i\beta)x_i = 0,$$

   where $\phi(x_i\beta)$ is a normal pdf.
   The resulting NLS estimator is biased due to its nonlinear feature, but consistent as discussed in the class. It is not the most efficient estimator, since MLE can achieve Cramer-Rao bound.

(b) Conditional variance of d_i is

$$\Phi(x_i\beta)\left[1-\Phi(x_i\beta)\right],$$

which indicates conditional heteroskedasticity.

(c) Yes, we can improve the efficiency, since we take into account its conditional variance strucutre. Think of GLS! First, estimate $\widetilde{\beta}$ as shown in part (a). Then, solve the following problem:

$$Arg\min_{\beta}\sum_{i=1}^{N}(d_i-\Phi(x_i\beta))^2\frac{1}{\Phi(x_i\widetilde{\beta})\left(1-\Phi(x_i\widetilde{\beta})\right)}.$$

(d) For MLE, construct the loglikelihood function as follows:

$$l(\beta)=\sum_{i=1}^{N}\left[d_i\log\Phi(x_i\beta)+(1-d_i)\log(1-\Phi(x_i\beta)\right].$$

By FOC (score function), we can get the ML estimator:

$$\sum_{i=1}^{N}\left[d_i\frac{\phi(x_i\beta)}{\Phi(x_i\beta)}-(1-d_i)\frac{\phi(x_i\beta)}{1-\Phi(x_i\beta)}\right]x_i=0.$$

3. (2SLS)

(a) First, get $\widehat{Y}_2$ as follows:

$$\widehat{Y}_2=X(X'X)^{-1}X'Y_2=P_XY_2.$$

Then, we can obtain 2SLS estimator by

$$\widehat{\delta}=(Z'P_XZ)^{-1}Z'P_Xy_1.$$

(b) See the following equations showing that $\widehat{\delta}$ is same as $\widehat{\delta}^*$:

$$\begin{aligned}\widehat{\delta}^* &= (Z'P_XZ)^{-1}Z'P_X\widehat{y}_1\\ &= (Z'P_XZ)^{-1}Z'P_XP_Xy_1\\ &= (Z'P_XZ)^{-1}Z'P_Xy_1\\ &= \widehat{\delta}.\end{aligned}$$

4. (IV Estimation)

$$\begin{aligned} y &= X_1\beta+Y\gamma+\epsilon\\ &\ \text{where } X_1 \text{ is } k_1\times N \text{ and } Y \text{ is } G\times N\\ &\ \text{and } X \text{ is } k\times N.\end{aligned}$$

(a) In order to identify δ, k should be at least, $k_1 + G$.

(b) OLS estimator:

$$\begin{aligned}
\widehat{\delta} &= (Z'Z)^{-1}Z'y \\
&= (Z'Z)^{-1}Z'(Z\delta + \epsilon) \\
(\widehat{\delta} - \delta) &= (Z'Z)^{-1}Z'\epsilon
\end{aligned}$$

Applying the asymptotic theory, we will have the following.

$$\sqrt{N}(\widehat{\delta} - \delta) \xrightarrow{d} N\left(0, \sigma^2 p\lim\left(\frac{Z'Z}{N}\right)^{-1}\right)$$

Or you can just say that the (approximated) asymptotic variance of $\widehat{\delta}$ is $\sigma^2(Z'Z)^{-1}$.

(c) IV estimator:

$$\begin{aligned}
\widehat{\delta}_{IV} &= (Z'\overline{M}Z)^{-1}Z'\overline{M}y \\
\text{where } \overline{M} &= X(X'X)^{-1}X
\end{aligned}$$

Asymptotic Variance of IV estimator:

$$\begin{aligned}
\widehat{\delta}_{IV} &= (Z'\overline{M}Z)^{-1}Z'\overline{M}y \\
&= (Z'\overline{M}Z)^{-1}Z'\overline{M}(Z\delta + \epsilon) \\
\widehat{\delta}_{IV} - \delta &= (Z'\overline{M}Z)^{-1}Z'\overline{M}\epsilon
\end{aligned}$$

Using the similar technique, we have:

$$\sqrt{N}(\widehat{\delta}_{IV} - \delta) \xrightarrow{d} N\left(0, \sigma^2 p\lim\left(\frac{Z'\overline{M}Z}{N}\right)^{-1}\right)$$

Or you can just say that the (approximated) asymptotic variance of $\widehat{\delta}_{IV}$ is $\sigma^2(Z'\overline{M}Z)^{-1}$.

(d) Asymptotic Variance of $\widehat{\delta} - \widehat{\delta}_{IV}$:

$$\begin{aligned}
\widehat{\delta} - \widehat{\delta}_{IV} &= \{\delta + (Z'Z)^{-1}Z'\epsilon\} - \{\delta + (Z'\overline{M}Z)^{-1}Z'\overline{M}\epsilon\} \\
&= (Z'Z)^{-1}Z'\epsilon - (Z'\overline{M}Z)^{-1}Z'\overline{M}\epsilon
\end{aligned}$$

Note that $E(\widehat{\delta} - \widehat{\delta}_{IV}) = 0$.

$$
\begin{aligned}
Var(\widehat{\delta}-\widehat{\delta}_{IV}) &= E[(\widehat{\delta}-\widehat{\delta}_{IV})(\widehat{\delta}-\widehat{\delta}_{IV})'] \\
&= E[\{(Z'Z)^{-1}Z'\epsilon-(Z'\overline{M}Z)^{-1}Z'\overline{M}\epsilon\}\{(Z'Z)^{-1}Z'\epsilon-(Z'\overline{M}Z)^{-1}Z'\overline{M}\epsilon\}'] \\
&= E[(Z'Z)^{-1}Z'\epsilon\epsilon'Z(Z'Z)^{-1}-(Z'\overline{M}Z)^{-1}Z'\overline{M}\epsilon\epsilon'Z(Z'Z)^{-1} \\
&\quad -(Z'Z)^{-1}Z'\epsilon\epsilon'\overline{M}Z(Z'\overline{M}Z)^{-1}+(Z'\overline{M}Z)^{-1}Z'\overline{M}\epsilon\epsilon\overline{M}Z(Z'\overline{M}Z)^{-1} \\
&= \sigma^2[(Z'Z)^{-1}-(Z'Z)^{-1}-(Z'Z)^{-1}+(Z'\overline{M}Z)^{-1}] \\
&= \sigma^2[(Z'\overline{M}Z)^{-1}-(Z'Z)^{-1}]
\end{aligned}
$$

(e) The rank of the covariance matrix:
The covariance matrix is $\sigma^2[(Z'\overline{M}Z)^{-1}-(Z'Z)^{-1}]$. The rank of this matrix is same as that of $(Z'Z)-(Z'\overline{M}Z)$, since pre- or postmultiplication of some matrix (in our case, $[(Z'\overline{M}Z)^{-1}-(Z'Z)^{-1}]$) by a nonsingular matrix does not change its rank. Premultiply the covariance matrix by $(Z'\overline{M}Z)$ and postmultiply it by $(Z'Z)$.Then, we are end up with $(Z'Z)-(Z'\overline{M}Z)$.

$$
\begin{aligned}
(Z'Z)-(Z'\overline{M}Z) &= Z'(I-\overline{M})Z \\
&= \begin{pmatrix} X_1' \\ Y' \end{pmatrix}(I-\overline{M})\begin{pmatrix} X_1 & Y \end{pmatrix} \\
&= \begin{pmatrix} X_1'\left(I-\overline{M}\right)X_1 & X_1'\left(I-\overline{M}\right)Y \\ Y'\left(I-\overline{M}\right)X_1 & Y'\left(I-\overline{M}\right)Y \end{pmatrix} \\
\text{Note that } \left(I-\overline{M}\right)X_1 &= 0 \\
&= \begin{pmatrix} 0 & 0 \\ 0 & Y'\left(I-\overline{M}\right)Y \end{pmatrix}
\end{aligned}
$$

Since $Y'\left(I-\overline{M}\right)Y$ has a full rank, the rank of the covariance matrix is G.

(f) The asymptotic covariance between $\widehat{\delta}$ and $\widehat{\delta}-\widehat{\delta}_{IV}$:

$$
\begin{aligned}
Cov(\widehat{\delta},\widehat{\delta}-\widehat{\delta}_{IV}) &= E[(\widehat{\delta}-\delta)(\widehat{\delta}-\widehat{\delta}_{IV})'] \\
&= E[\{(Z'Z)^{-1}Z'\epsilon\}\{(Z'Z)^{-1}Z'\epsilon-(Z'\overline{M}Z)^{-1}Z'\overline{M}\epsilon\}'] \\
&= E[(Z'Z)^{-1}Z'\epsilon\epsilon'Z(Z'Z)^{-1}-(Z'Z)^{-1}Z'\epsilon\epsilon'\overline{M}Z(Z'\overline{M}Z)^{-1}] \\
&= \sigma^2[(Z'Z)^{-1}-(Z'Z)^{-1}] \\
&= 0.
\end{aligned}
$$

5. (Order Condition in SEM)

Note that for identification of equation 1, we should have $K \geq K_1+G_1-1$.

(a) $f(x)=1+x$: Here, we have that $K=2$,(Strictly, the rank of X is two) since $f(x)$ is a linear combination of explanatory variables in equation 1. However, $K_1 + G_1 - 1 = 3$. Therefore, this is not identified.

(b) $f(x) = 1 + x + x^2$: Now, $K = 3$. This is just-identified case.

(c) $f(x) = 1 + \exp(x)$: Now, $K = 3$. This is also just-identified case.

6. (K-variable Regression)

(a) Matrix X, β will look like the following.

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Using the usual way of getting the variances of LS estimator, we have;

$$\begin{aligned} Var(\widehat{\beta}_{OLS}) &= (X'X)^{-1} \\ X'X &= \begin{pmatrix} n_{odd} & 0 \\ 0 & n_{even} \end{pmatrix} \\ \text{where n} &= n_{odd} + n_{even}. \end{aligned}$$

Note that, if n is even, $n_{odd} = n_{even}$ and if n is odd, $n_{odd} = n_{even} + 1$. Therefore,

$$Var(\widehat{\beta}_{OLS}) = \begin{pmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2} \end{pmatrix}$$

(b) We are trying to solve the following equations.

$$\begin{aligned} (n_{odd})^{-1} \sum_{odd} (y_i - x_i\beta) &= 0 \\ (n_{even})^{-1} \sum_{even} (y_i - x_i\beta) &= 0 \end{aligned}$$

Then, because of the strange(?) feature of X matrix, we have;

$$\begin{aligned} (n_{odd})^{-1} \sum_{odd} (y_i - \beta_1) &= 0 \\ (n_{even})^{-1} \sum_{even} (y_i - \beta_2) &= 0 \end{aligned}$$

Therefore,

$$\begin{aligned}\beta_1^* &= \frac{\sum_{odd} y_i}{n_{odd}} \\ \beta_2^* &= \frac{\sum_{even} y_i}{n_{even}}\end{aligned}$$

Next, let's figure out the covariance matrix of the above estimator.

$$\beta^* = \begin{pmatrix} \frac{\sum_{odd} y_i}{n_{odd}} \\ \frac{\sum_{even} y_i}{n_{even}} \end{pmatrix} = \begin{pmatrix} \frac{\sum_{odd}(\beta_1+\varepsilon_i)}{n_{odd}} \\ \frac{\sum_{even}(\beta_2+\varepsilon_i)}{n_{even}} \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \frac{\sum_{odd} \varepsilon_i}{n_{odd}} \\ \frac{\sum_{even} \varepsilon_i}{n_{even}} \end{pmatrix}$$

Hence,

$$\begin{aligned} Var(\beta^*) &= E\left[\begin{pmatrix} \frac{\sum_{odd} \varepsilon_i}{n_{odd}} \\ \frac{\sum_{even} \varepsilon_i}{n_{even}} \end{pmatrix} \begin{pmatrix} \frac{\sum_{odd} \varepsilon_i}{n_{odd}} & \frac{\sum_{even} \varepsilon_i}{n_{even}} \end{pmatrix}\right] \\ &= \begin{pmatrix} \frac{1}{n_{odd}} & 0 \\ 0 & \frac{1}{n_{even}} \end{pmatrix}. \end{aligned}$$

The variance of the OLS estimator is $(X'X)^{-1}$. This is same as the above.