Econometrics I: Solutions of final exam

LU ANG, Hiroki Kato*

August 16, 2020

Contents

1 Question 1

(1)

First we replace β by OLS estimator $\hat{\beta}$ and we can have the following expression:

$$y = X\hat{\beta} + e$$

Note that OLSE is the estimator that minimizes the sum of squared error terms, In our case denote as e. Therefore we obtain the expression as follows:

$$S(\hat{\beta}) = \sum_{t=1}^{n} e_i^2 = e'e = (y - X\hat{\beta})'(y - X\hat{\beta}) = (y - \hat{\beta}'y')(y - X\hat{\beta})$$

$$= y'y - y'X'\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} = y'y - 2y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

*If you have any errors in handouts and materials, please contact me via lvang12@hotmail.com or vge008kh@student.econ.osaka-u.ac.jp. Room 503.

To minimize $S(\hat{\beta})$ with respect to $\hat{\beta}$, we set the first derivative of $S(\hat{\beta})$ equal to zero:

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

Solve the equation we can obtain the OLS estimator as:

$$\hat{\beta} = (X'X)^{-1}X'y \tag{1}$$

(2)

In order to calculate the mean and variance of $\hat{\beta}$ we first need to rewrite (1) as:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + u) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u \qquad (2)\end{aligned}$$

Becasue X is nonstochastic Then we take the expectation and variance for both sides:

$$E(\hat{\beta}) = E(\beta + (X'X)^{-1}X'u) = \beta + (X'X)^{-1}X'E(u) = \beta$$

where $E(u) = 0$ by the Assumption

$$
\begin{aligned}
V(\hat{\beta}) &= E((\hat{\beta}-\beta)(\hat{\beta}-\beta)') = E((X'X)^{-1}X'u((X'X)^{-1}X'u)') \\
&= E((X'X)^{-1}X'uu'X(X'X)^{-1}) \\
&= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 I_T X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

where $E(uu') = V(u) = \sigma^2$ by the Assumption

(3)

The first step is to construct a linear unbiased estimator, $\tilde{\beta}$ Since a linear estimator is a function of dependent variable, y, define $\tilde{\beta} = Cy$ where C is a $k \times T$ matrix. Then, the expectation of $\tilde{\beta}$ is

$$E(\tilde{\beta}) = E(C(X\beta + u)) = CX\beta$$

If $\tilde{\beta}$ is an unbiased estimator, it must hold that

$$CX = I_k$$

Next we take the variance of $\tilde{\beta} = Cy$

$$V(\tilde{\beta}) = E((\tilde{\beta}-\beta)(\tilde{\beta}-\beta)') = E(Cuu'C') = \sigma^2 CC'$$

Defining $C = D + (X'X)^{-1}X'$, $V(\tilde{\beta})$ is rewritten as:.

$$V(\tilde{\beta}) = \sigma^2 CC' = \sigma^2(D + (X'X)^{-1}X')(D + (X'X)^{-1}X')'$$

Moreover, because $\tilde{\beta}$ is unbiased, we have the following:

$$CX = I_k = (D + (X'X)^{-1}X')X = DX + I_k$$

Therefore, we have the following condition:

$$DX = 0$$

Accordingly, $V(\tilde{\beta})$ is rewritten as:

$$\begin{aligned} V(\tilde{\beta}) &= \sigma^2 CC' = \sigma^2(D + (X'X)^{-1}X')(D + (X'X)^{-1}X')' \\ &= \sigma^2(X'X)^{-1} + \sigma^2 DD' = V(\hat{\beta}) + \sigma^2 DD' \end{aligned} \tag{3}$$

Notice that $V(\hat{\beta}) - V(\hat{\beta})$ is a positive definite matrix. That is $V(\hat{\beta}) - V(\hat{\beta}) > 0$. Thus $V(\hat{\beta})$ has the smallest variance among all unbiased estimator

(4)

Notice the moment-generating function of $X \sim N(u, \Sigma)$ is given by:

$$\phi(\theta) \equiv E(exp(\theta' X)) = E(exp(\theta u + \frac{1}{2}\theta' \Sigma \theta)$$

In our case we know that the standard error $u \sim N(0, \sigma^2 I_T)$. i.e.

$$\phi(\theta_u) \equiv E(exp(\theta_u' X)) = E(exp(\frac{1}{2}\theta_u' \theta_u)$$

Next in order to derive a distribution of $\hat{\beta}$, we write the moment-generating function of $\hat{\beta}$ as follows:

$$\phi(\theta) \equiv E(exp(\theta_\beta'\hat{\beta})) = E(exp(\theta_\beta'\beta + \theta_\beta'(X'X)^{-1}X'u)$$
$$= exp(\theta_\beta'\beta)E(exp(\theta_\beta'(X'X)^{-1}X'u)) = exp(\theta_\beta'\beta)\phi_u(\theta_\beta'(X'X)^{-1}X')$$
$$= exp(\theta_\beta'\beta)exp(\frac{\sigma^2}{2}\theta_\beta'(X'X)^{-1}\theta_\beta) = exp(\theta_\beta'\beta + \frac{\sigma^2}{2}\theta_\beta'(X'X)^{-1}\theta_\beta)$$

where $\theta_u = X(X'X)^{-1}\theta_\beta$

This indicate that:

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$$

This expression can be also rewritten as:

$$\sqrt{T}(\beta - \hat{\beta}) \sim N(0, \sigma^2(\frac{1}{T}X'X)^{-1})$$

Here, by weak law of Large Number, We assume that:

$$(\frac{1}{T}X'X)^{-1} \xrightarrow{p} M_{xx}^{-1} < \infty$$

Alternatively we can also apply CLT to derive this distribution, in order to do that, first we can rewrite the equation (2) as:

$$\sqrt{T}(\beta - \hat{\beta}) = (\frac{1}{T}X'X)^{-1}(\frac{1}{\sqrt{T}}X'u)$$

Applying the central limit theorem yields:

$$\frac{1}{\sqrt{T}}X'u \xrightarrow{d} N(0, \sigma^2 M_{xx})$$

Finally by applying Slutsky theorem we obtain:

$$\sqrt{T}(\beta - \hat{\beta}) \sim N(0, \sigma^2 M_{xx}^{-1}M_{xx}M_{xx}^{-1}) = N(0, \sigma^2 M_{xx}^{-1})$$

(5)

In order to prove $E(s^2) = \sigma^2$, we first substitute $\hat{\beta}$ with $(X'X)^{-1}X'y$

$$
\begin{aligned}
y - X\hat{\beta} &= y - X(\beta + (X'X)^{-1}X'u) \\
&= (y - X\beta) + X(X'X)^{-1}X'u \\
&= (I_T - X(X'X)^{-1}X')u \\
&= Mu
\end{aligned}
$$

where $P = X(X'X)^{-1}X'$ is the projection matrix. $M = (I_T - P)$, which maps to vectors of response values to the vector of residual values. Both P and M are idempotent and symmetric. i.e. $P^2 = P$,$P' = P$,$M^2 = M$,$M' = M$

$$
\begin{aligned}
s^2 &= \frac{1}{T-k}(Mu)'Mu \\
&= \frac{1}{T-k}u'MMu \\
&= \frac{1}{T-k}u'Mu
\end{aligned}
$$

Because $u'Mu$ is a scalar, thus $tr(u'Mu) = u'Mu$. Then we can obtain

$$
\begin{aligned}
E(s^2) &= \frac{1}{n}E(tr(u'Mu)) \\
&= \frac{1}{T-k}E(tr(u'(I_T - X(X'X)^{-1}X')u)) \\
&= \frac{1}{T-k}E(tr((I_T - X(X'X)^{-1}X')uu')) \\
&= \frac{1}{T-k}tr((I_T - X(X'X)^{-1}X')E(uu')) \\
&= \frac{1}{T-k}\sigma^2 tr((I_T - X(X'X)^{-1}X')) \\
&= \frac{1}{T-k}\sigma^2(tr(I_T) - tr(X(X'X)^{-1}X')) \\
&= \frac{1}{T-k}\sigma^2(tr(I_T) - tr((X'X)^{-1}X'X)) \\
&= \frac{1}{T-k}\sigma^2(tr(I_T) - tr(I_k) \\
&= \frac{1}{T-k}\sigma^2(T-k) \\
&= \sigma^2
\end{aligned}
$$

(6)

We apply the Lagrange multiplier to calculate the restricted estimator. In order to minimize $(y - X\beta)'(y - X\beta)$ with the restriction $R\beta = r$. We can write the Loss function as:

$$L = (y - X\tilde{\beta})'(y - X\tilde{\beta}) - 2\tilde{\lambda}'(R\tilde{\beta} - r)$$

Here $\tilde{\beta}$ and $\tilde{\lambda}$ are the estimators that minimize L. Then the F.O.C can be obtained:

$$
\begin{aligned}
\frac{\partial L}{\partial \tilde{\beta}} &= -2X'(y - X\tilde{\beta}) - 2R'\tilde{\lambda} = 0 \\
\frac{\partial L}{\partial \tilde{\lambda}} &= -2(R\tilde{\beta} - r) = 0
\end{aligned}
$$

Solving the equation for $\tilde{\beta}$ we have following expression:

$$\tilde{\beta} = (X'X)^{-1}X'y + (X'X)^{-1}R'\tilde{\lambda} = \hat{\beta} + (X'X)^{-1}R'\tilde{\lambda} \tag{4}$$

Multiply R by both side we have following expression:

$$R\tilde{\beta} = R\hat{\beta} + R(X'X)^{-1}R'\tilde{\lambda}$$

Because we have the restriction $R\tilde{\beta} = r$, we substitute the Left side:

$$r = R\hat{\beta} + R(X'X)^{-1}R'\tilde{\lambda}$$

Thus we can solve $\tilde{\lambda}$ as:

$$\tilde{\lambda} = (R(X'X)^{-1}R')^{-1}(r - R\hat{\beta})$$

Next we substitute $\tilde{\lambda}$ back into equation (4) we can obtain:

$$\tilde{\beta} = \hat{\beta} + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta})$$

(7)

Since $R\hat{\beta} \sim N(R\beta, \sigma^2 R(X'X)^{-1}R')$ and Under the restriction $R\beta = r$ we have:

$$\frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - r)}{\sigma^2} \sim \chi^2(G)$$

where $rank(R) = G \leq k$

Also we know that:

$$\frac{(n-k)s^2}{\sigma^2} = \frac{\hat{u}'\hat{u}}{\sigma^2} = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{\sigma^2} \sim \chi^2(T-k)$$

Therefore, we have the following distribution:

$$\frac{(R\hat{\beta}-r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta}-r)/G}{(y-X\hat{\beta})'(y-X\hat{\beta})/(T-k)} \sim F(G, T-k)$$

Then, using $\tilde{\beta} = \hat{\beta} + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta})$, we can derive that:

$$(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(R\hat{\beta}-r) = \hat{\beta} - \tilde{\beta} \tag{5}$$

Multiply R by both sides in equation (5) we can have the following expression:

$$(R\hat{\beta}-r) = R(\hat{\beta}-\tilde{\beta}) \tag{6}$$

Substitute this expression back into the numerator we can obtain:

$$\begin{aligned}(R\hat{\beta}-r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta}-r) &= (\hat{\beta}-\tilde{\beta})'R'(R(X'X)^{-1}R')^{-1}R(\hat{\beta}-\tilde{\beta}) \\ &= (\hat{\beta}-\tilde{\beta})'X'X(\hat{\beta}-\tilde{\beta})\end{aligned}$$

Moreover the numerator is represented as follows:

$$\begin{aligned}(y-X\tilde{\beta})'(y-X\tilde{\beta}) &= (y-X\hat{\beta}-X(\tilde{\beta}-\hat{\beta}))'(y-X\hat{\beta}-X(\tilde{\beta}-\hat{\beta})) \\ &= (y-X\hat{\beta})'(y-X\hat{\beta}) + (\tilde{\beta}-\hat{\beta})'X'X(\tilde{\beta}-\hat{\beta}) \\ &\quad - (y-X\hat{\beta})'X(\tilde{\beta}-\hat{\beta}) - (\tilde{\beta}-\hat{\beta})'X'(y-X\hat{\beta}) \\ &= (y-X\hat{\beta})'(y-X\hat{\beta}) + (\hat{\beta}-\tilde{\beta})'X'X(\hat{\beta}-\tilde{\beta}) \end{aligned} \tag{7}$$

where $X'(y-X\hat{\beta}) = X'\hat{u} = 0$

Summarizing, we have following representation:

$$\begin{aligned}(R\hat{\beta}-r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta}-r) &= (\hat{\beta}-\tilde{\beta})'X'X(\hat{\beta}-\tilde{\beta}) \\ &= (y-X\tilde{\beta})'(y-X\tilde{\beta}) - (y-X\hat{\beta})'(y-X\hat{\beta}) \\ &= \tilde{u}'\tilde{u} - \hat{u}'\hat{u}\end{aligned}$$

Therefore we have:

$$\frac{(R\hat{\beta}-r)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta}-r)/G}{(y-X\hat{\beta})'(y-X\hat{\beta})/(T-k)} = \frac{(\tilde{u}'\tilde{u}-\hat{u}'\hat{u})/G}{\hat{u}'\hat{u}/(T-k)} \sim F(G, T-k)$$

(8)

The Coefficient of Determination R^2 for questions (1) and (6) are as follows:

$$R_1^2 = 1 - \frac{\hat{u}'\hat{u}}{y'(I_T - \frac{1}{T}ii')y}$$

$$R_6^2 = 1 - \frac{\tilde{u}'\tilde{u}}{y'(I_T - \frac{1}{T}ii')y}$$

Substitute above into $\frac{(R_1^2 - R_6^2)/G}{(1-R_1^2)/(T-k)}$ we can obtain:

$$\frac{(\frac{\tilde{u}'\tilde{u}}{y'(I_T-\frac{1}{T}ii')y} - \frac{\hat{u}'\hat{u}}{y'(I_T-\frac{1}{T}ii')y})/G}{(\frac{\tilde{u}'\tilde{u}}{y'(I_T-\frac{1}{T}ii')y})/(T-k)} = \frac{(\tilde{u}'\tilde{u}-\hat{u}'\hat{u})/G}{\hat{u}'\hat{u}/(T-k)}$$

Which is exactly the same expression as (7), thus we can obtain:

$$\frac{(R_1^2 - R_6^2)/G}{(1-R_1^2)/(T-k)} \sim F(G, T-k)$$

2 Question 2

(9)

The likelihood function is defined as $L(P;X) = \prod_{i=1}^{n} f(X_i;P)$ in our case:

$$L(P;X) = P^{\sum_{i=1}^{n} X_i}(1-P)^{\sum_{i=1}^{n}(1-X_i)} \tag{8}$$

The maximum likelihood estimate of P is the P such that:

$$\max_P L(P;X) \iff \max_P logL(P;X)$$

Therefore we can take log for both sides and the log likelihood function can be written as:

$$logL(P;X) = \sum_{i=1}^{n} X_i logP + \sum_{i=1}^{n}(1-X_i)log(1-P) \tag{9}$$

Obtain the first order condition:

$$\frac{\partial logL(P;X)}{\partial P} = \frac{\sum_{i=1}^{n} X_i}{P} - \frac{\sum_{i=1}^{n}(1-X_i)}{1-P} = 0$$

Solving the equation we can obtain the MLE estimators as:

$$\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{10}$$

(10)

Recall that Bernoulli distribution has the mean $E(X) = P$ and the variance:

$$E(X) = Pr(X=1) \times 1 + Pr(X=0) \times 0 = P$$

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= Pr(X=1) \times 1^2 + Pr(X=0) \times 0^2 - P^2 \\ &= P(1-P) \end{aligned}$$

Then the variance of MLE estimators are calculated as:

$$E(\hat{P}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = P \tag{11}$$

$$V(\hat{P}) = V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{P(1-P)}{n} \tag{12}$$

(11)

First, Fisher's information matrix(In our case, it is just a scalar since we only have one estimator) $I(P)$ is given as follows:

$$\begin{aligned} I(P) = V(\frac{\partial logL(P;X)}{\partial P} = -E(\frac{\partial^2 logL(P;X)}{\partial P^2}) &= -E(-\frac{\sum_{i=1}^{n} X_i}{P^2} - \frac{n-\sum_{i=1}^{n} X_i}{(1-P)^2}) \\ &= \frac{\sum_{i=1}^{n} E(X_i)}{P^2} + \frac{n-\sum_{i=1}^{n} E(X_i)}{(1-P)^2} \\ &= (\frac{n}{P} + \frac{n}{1-P}) \\ &= \frac{n}{P(1-P)} \end{aligned} \tag{13}$$

Cramer-Rao lower bound which is given as:

$$I(P)^{-1} = \frac{P(1-P)}{n}$$

Next we are going to $\hat{P}$ has the smallest variance. Suppose that an unbiased estimator of P as $s(X)$, i.e. $E(s(X)) = P$
The expectation of $s(X)$:

$$E(s(X)) = \int s(x)L(P;x)dx$$

Differentiating the above with respect to P

$$\begin{aligned}\frac{\partial E(s(X))}{\partial P} &= \int s(x)\frac{\partial L(P;x)}{\partial P}dx = \int s(x)\frac{\partial logL(P;x)}{\partial P}L(P;x)dx \\ &= Cov(s(X), \frac{\partial logL(P;X)}{\partial P})\end{aligned}$$

In our case, $s(X)$ and P are just scalars, thus:

$$\begin{aligned}(\frac{\partial E(s(X))}{\partial P})^2 &= (Cov(s(X), \frac{\partial L(P;x)}{\partial P}))^2 = \rho^2 V(s(X))V(\frac{\partial logL(P;X)}{\partial P}) \\ &\geq V(s(X))V(\frac{\partial logL(P;X)}{\partial P})\end{aligned}$$

where ρ is the correlation coefficient between $s(X)$ and $\frac{\partial logL(P;X)}{\partial P}$ and $|\rho| \leq 1$

Therefore, we have the following inequality:

$$V(s(X)) \geq \frac{(\frac{\partial E(s(X))}{\partial P})^2}{V(\frac{\partial logL(P;X)}{\partial P})}$$

Since $s(X)$ is an unbiased estimator of P. i.e. $E(s(X)) = P$

Therefore, we obtain:

$$V(s(X)) \geq \frac{1}{V(\frac{\partial logL(P;X)}{\partial P})} = (I(P))^{-1} \tag{14}$$

In our case

$$(I(P))^{-1} = V(\hat{P}) = \frac{P(1-P)}{n}$$

Thus we have proved $\hat{P}$ has the smallest variance among all unbiased estimator

(12)

In order to prove $\hat{P}$ is a consistent estimator of P we need to prove:

$$\lim_{n\to\infty} P(|\hat{P}-P|<\epsilon)=1$$

for any positive ϵ

Recall that Chebyshev's Inequality states as:

$$P(g(X)\geq k)\leq \frac{E(g(X))}{k}$$

for $g(X)\geq 0$

In our case let us set $g(X)=(\hat{P}-P)^2,\ e^2=k,\ E(g(X))=V(\hat{P})=\dfrac{P(1-P)}{n}$
if $n\longrightarrow\infty$,

$$P((\hat{P}-P)^2\geq k)=P(|\hat{P}-P|\geq\epsilon)\leq\frac{P(1-P)}{n\epsilon^2}\to 0$$

That is, for any ϵ,

$$\lim_{n\to\infty} P(|\hat{P}-P|<\epsilon)=1$$

Thus we have proved $\hat{P}$ is a consistent estimator of P

(13)

In order to prove the asymptotic distribution. let us first focus on the FOC of our likelihood function:

$$\frac{\partial logL(P;X)}{\partial P}=\sum_{i=1}^{n}\frac{\partial logf(X_i;P)}{\partial P}=0$$

Applying Central Limit Theorem as follows:

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial logf(X_i;P)}{\partial P}-E(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial logf(X_i;P)}{\partial P})}{\sqrt{V(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial logf(X_i;P)}{\partial P})}}=\frac{\frac{1}{n}\frac{\partial logL(X_i;P)}{\partial P}-E(\frac{1}{n}\frac{\partial logL(X_i;P)}{\partial P})}{\sqrt{V(\frac{1}{n}\frac{\partial logL(X_i;P)}{\partial P})}}$$

in our case:

$$E(\frac{1}{n}\frac{\partial logL(X_i;P)}{\partial P}) = 0$$

and

$$V(\frac{1}{n}\frac{\partial logL(X_i;P)}{\partial P}) = \frac{1}{n^2}I(\theta)$$

Thus, the asymptotic distribution of $\frac{1}{n}\frac{\partial logL(X_i;P)}{\partial P}$ is given by:

$$\sqrt{n}\left(\frac{1}{n}\frac{\partial logL(X_i;P)}{\partial P} - E(\frac{1}{n}\frac{\partial logL(X_i;P)}{\partial P})\right) = \frac{1}{\sqrt{n}}\frac{\partial logL(X_i;P)}{\partial P} \longrightarrow N(0,\Sigma)$$

where, according to equation (13):

$$\Sigma = V(\frac{1}{\sqrt{n}}\frac{\partial logL(X_i;P)}{\partial P}) = \frac{1}{n}I(P) = \frac{1}{P(1-P)}$$

That is,

$$\frac{1}{\sqrt{n}}\frac{\partial logL(X_i;P)}{\partial P} \longrightarrow N(0,\Sigma)$$

Now, replacing P by $\tilde{P}$, consider the asymptotic distribution of

$$\frac{1}{\sqrt{n}}\frac{\partial logL(\tilde{P};X)}{\partial P}$$

which is expanded around $\tilde{P} = P$ as follows:

$$0 = \frac{1}{\sqrt{n}}\frac{\partial logL(X_i;P)}{\partial P} \approx \frac{1}{\sqrt{n}}\frac{\partial logL(P;X)}{\partial P} + \frac{1}{\sqrt{n}}\frac{\partial^2 logL(P;X)}{\partial P^2}(\tilde{P}-P)$$

Therefore,

$$-\frac{1}{\sqrt{n}}\frac{\partial^2 logL(P;X)}{\partial P^2}(\tilde{P}-P) \approx \frac{1}{\sqrt{n}}\frac{\partial logL(X_i;P)}{\partial P} \longrightarrow N(0,\Sigma) \tag{15}$$

Then the expression can be rewritten as:

$$\sqrt{n}(\tilde{P}-P) \approx (-\frac{1}{\sqrt{n}}\frac{\partial^2 logL(P;X)}{\partial P^2})^{-1}(\frac{1}{\sqrt{n}}\frac{\partial logL(\tilde{P};X)}{\partial P}) \tag{16}$$

Note that, Using the law of large number

$$
\begin{aligned}
-\frac{1}{\sqrt{n}}\frac{\partial^2 logL(P;X)}{\partial P^2} \longrightarrow & \lim_{n\to\infty}\frac{1}{n}\left(-E(\frac{\partial^2 logL(P;X)}{\partial P^2})\right) \\
&= \lim_{n\to\infty}\frac{1}{n}\left(V(\frac{\partial logL(P;X)}{\partial P})\right) \\
&= \lim_{n\to\infty}\frac{1}{n}I(P) = \frac{1}{P(1-P)} = \Sigma \qquad (17)
\end{aligned}
$$

Combining the result of (15),(16)(17), and applying slutsky's theorem we can obtain:

$$
\begin{aligned}
\sqrt{n}(\tilde{P}-P) \approx (-\frac{1}{\sqrt{n}}\frac{\partial^2 logL(P;X)}{\partial P^2})^{-1}(\frac{1}{\sqrt{n}}\frac{\partial logL(\tilde{P};X)}{\partial P}) \longrightarrow & N(0,\Sigma^{-1}\Sigma\Sigma^{-1}) \\
&= N(0,\Sigma^{-1}) \\
&= N(0,P(1-P))
\end{aligned}
$$

(14)

The Wald test states:

$$h(\hat{\theta})(R_\theta(I(\theta))^{-1}R'_\theta)^{-1}h(\hat{\theta})' \to \chi^2(G)$$

Furthermore, as $n \longrightarrow \infty$ we have $R_{\hat{\theta}} \to R_\theta$ and $I(\hat{\theta}) \to I(\hat{\theta})$

$$h(\hat{\theta})(R_{\hat{\theta}}(I(\hat{\theta}))^{-1}R'_{\hat{\theta}})^{-1}h(\hat{\theta})' \to \chi^2(G)$$

where $h(\theta) = 0$ is the null hypothesis and $R_\theta = \dfrac{\partial logL(\theta)}{\partial\theta}$

In our case $h(P) = P - 0.5 = 0$, $R_P = 1$, $I(P)^{-1} = \frac{P(1-P)}{n}$, $G = 1$

Then our test statistic:

$$h(\hat{P})(R_{\hat{P}}(I(\hat{P}))^{-1}R'_{\hat{P}})^{-1}h(\hat{P})' = \frac{n}{\hat{P}(1-\hat{P})}(\hat{P}-0.5)^2 \sim \chi^2(1)$$

where $\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is our MLE obtain in question (9)

Compare the test statistic, if it is greater than the critical value $\chi^2(1)$ we should reject the null hypothesis, otherwise we can not reject the null hypothesis

(15)

Likelihood Ratio Test states:

$$LR = -2(logL(\tilde{\theta}) - logL(\hat{\theta})) \longrightarrow \chi^2(G)$$

In our case under the null hypothesis: h(P)=P-0.5=0

$$h(\tilde{P}) = 0$$

is always satisfied. i.e. $\tilde{P} = 0.5$

the test statistic is as follows:

$$\begin{aligned} -2(logL(\tilde{P}) - logL(\hat{P})) &= -2[\sum_{i=1}^{n} X_i log\tilde{P} + \sum_{i=1}^{n}(1-X_i)log(1-\tilde{P}) \\ &\quad - \sum_{i=1}^{n} X_i log\hat{P} - \sum_{i=1}^{n}(1-X_i)log(1-\tilde{P})] \\ &= -2[\sum_{i=1}^{n} X_i log\tilde{P}/\hat{P} + \sum_{i=1}^{n}(1-X_i)log(1-\tilde{P})/(1-\hat{P})] \end{aligned}$$

substitute $\tilde{P} = 0.5$ and $\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$ into LR we can obtain our test statistic. Comparing the test statistic, if it is greater than the critical value $\chi^2(1)$ we should reject the null hypothesis, otherwise we can not reject the null hypothesis

3 Question 3

(16)

The OLSE is now given by $\hat{\beta} = (X'X)^{-1}X'y$. Substituting the original regression equation into y yields

$$\hat{\beta} = \beta + (X'X)^{-1}X'u \tag{18}$$

Taking expectation on both sides gives

$$E(\hat{\beta}) = \beta + E[(X'X)^{-1}X'u] = \beta + E[(X'X)^{-1}X'E(u|X)]. \tag{19}$$

Note that the second equality comes from the law of iterated expectation. Since X is correlated with u, $E(u|X) \neq 0$. Thus, the second term of equation (**??**) no longer vanishes. The OLSE is biased estimator.
Let X_t be $k \times 1$ vector such that $X = (X_1', \cdots, X_T')$ We reformulate (18) as follows;

$$\hat{\beta} = \beta + \left(\frac{1}{T}\sum_t X_t X_t'\right)^{-1}\left(\frac{1}{T}\sum_t X_t u_t\right) \tag{20}$$

Assume $E(X_t X_t') = M_{xx}$.

By the weak law of large numbers (WLLN) and Slutzky's theorem, we have

$$\operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}\sum_t X_t X_t'\right)^{-1} = M_{xx}^{-1} \tag{21}$$

Since X is correlated with u, $E(X_t u_t) = M_{xu} \neq 0$. By the WLLN, we have

$$\operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}\sum_t X_t u_t\right) = M_{xu}. \tag{22}$$

where γ is $k \times 1$ vector. Taking probability limit on both sides of (20) yields $\operatorname{plim}_{T\to\infty}\hat{\beta} = \beta + M_{xx}M_{xu}$. Thus, the OLSE is inconsistent.

(17)

Let Z_t be $k \times 1$ vector such that $Z = (Z_1', \cdots, Z_T')$. We assume that:

Assumption 1. Z_t is uncorrelated with u_t,i.e.$Cov(Z_t, u_t) = E(Z_t u_t) = 0$;
Assumption 2. Z_t is correlated with X_t,i.e.$E(Z_t X_t') = M_{zx}$.

We reformulate the original regression equation as follows;

$$\frac{Z'y}{T} = \frac{Z'X\beta}{T} + \frac{Z'u}{T}. \tag{23}$$

Taking probability limit on both sides yields

$$\operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t y_t\right) = \operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\right)\beta + \operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t u_t\right). \tag{24}$$

By the assumption 1 and 2, and the WLLN, we have

$$\operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\right)^{-1} \operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}\sum_{t} Z_t y_t\right) = \beta. \tag{25}$$

This implies that the consistent estimator of β is

$$\beta_{iv} = (Z'X)^{-1}Z'y, \tag{26}$$

which is called **instrumental variable estimator**.

(18)

Substituting (**??**) into the regression equation yields

$$\beta_{iv} = \beta + \left(\sum_{t} Z_t X_T'\right)^{-1}\sum_{t} Z_t u_t. \tag{27}$$

We reformulate this equation as follows;

$$\sqrt{T}(\beta_{iv} - \beta) = \left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t u_t\right). \tag{28}$$

Using the WLLN and Slutzky's theorem, we have

$$\operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\right)^{-1} = M_{zx}^{-1}. \tag{29}$$

This comes from the assumption 2. We derive the asymptotic distribution of $T^{-1/2}\sum_t Z_t u_t$. Define $\bar{Z}_t = Z_t u_t$. The assumption 1 leads to $E(\bar{Z}_t) = 0$, and $\text{Var}(\bar{Z}_t) = E(u_t^2 Z_t Z_t') = \sigma^2 M_{zz}$. That is, $\lim_{T\to\infty}((1/T)\sum_t \text{Var}(\bar{Z}_t)) = \sigma^2 M_{zz}$. Applying the general version of central limit theorem yields

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{Z}_t \xrightarrow{d} N(0, \sigma^2 M_{zz}). \tag{30}$$

Using (**??**) and (**??**), we obtain

$$\sqrt{T}(\beta_{iv} - \beta) \xrightarrow{d} N(0, \sigma^2 M_{zx}^{-1} M_{zz} (M_{zx}')^{-1}). \tag{31}$$