Econometrics II HW 1

Jacob Aguirre

January 27, 2023

1 Question 1

Problem: 3.5

1. (a) It is easy to see that because $Var(\bar{y}_N = \frac{\sigma^2}{N}$, then $Var[\sqrt{N}(\bar{y}_N - \mu]$ is equal to $N(\frac{\sigma^2}{N}) = \sigma^2$.
2. (b) Recognize that by Central Limit Theorem (CLT), we have that $\sqrt{N}(\bar{y}_N - \mu)$ is approximated by to $Normal(0, \sigma^2)$, and so the $Avar[\sqrt{N}(\bar{y}_N - \mu)] = \sigma^2$.
3. (c) It is easy to see that we can obtain $Avar(\bar{y}_N)$ by dividing $Avar[\sqrt{N}(\bar{y}_N - \mu)]$ by N. Thus, we can see that $Avar(\bar{y}_N) = \frac{\sigma^2}{N}$. Obviously, this coincides with the variance of $\bar{y}_N$.
4. (d) Recognize that the asymptotic std of $\bar{y}_N$ is the square root of its asymptotic variance, written as $\frac{\sigma}{\sqrt{N}}$.
5. (e) We can see that to obtain the asymptotic S.E. of $\bar{y}_N$, we need to write a consistent estimator of σ. We can see that the asymptotic standard error of $\bar{y}_N = \frac{\hat{\sigma}}{\sqrt{N}}$.

2 Question 2

Problem 4.1

1. (a) We can find that by exponentiation, we have the equation,

$$
\begin{aligned}
wage &= \exp(\beta_0 + \beta_1 Married + \beta_2 educ + z\gamma + u), && (1)\\
&= \exp(u)\exp(\beta_0 + \beta_1 Married + \beta_2 educ + z\gamma). && (2)
\end{aligned}
$$

This would give us,

$$\mathbb{E}(wage|x) = \mathbb{E}[\exp(u)|x]\exp(\beta_0 + \beta_1 married + \beta_2 educ + z\gamma),$$

where we can see that the x term *obviously* denotes the explanatory variables. So, consider the case if u and x are independent. This would lead to

$$
\begin{aligned}
\mathbb{E}[\exp(u)|x] = \mathbb{E}[\exp(u)] = \delta_0,\\
\implies \mathbb{E}(wage|x) = \delta_0 \exp(\beta_0 + \beta_1 married + \beta_2 educ + z\gamma).
\end{aligned}
$$

Naturally, we set married =1 if married and =0 if not. Then, we can write that

$$\frac{\delta_0 \exp(\beta_0 + \beta_1 + \beta_2 edduc + z\gamma) - \delta_0 \exp(\beta_0 + \beta_2 educ + z\gamma)}{\delta_0 \exp(\beta_0 + \beta_2 educ + z\gamma)} = \exp(\beta_1) - 1$$

Which leads us to concluding the percentage difference is indeed $100 \cdot [\exp(\beta_1) - 1]$.

2. (b) Obviously since $\theta_1 = 100 \cdot [\exp(\beta_1) - 1] = g(\beta_1)$, we can take the derivative of g with respect to β_1: $\frac{\partial g}{\partial \beta_1} = 100 \cdot \exp(\beta_1)$. So, this gives us the asymptotic standard error of $\hat{\theta_1}$ using the delta method. So, this would give us the absolute value of $\frac{\partial \hat{g}}{\partial \beta_1} \cdot$ s.e.$(\hat{\beta}_1)$:

$$se(\hat{\theta_1}) = 100 \cdot [\exp(\hat{\beta_1}] \cdot se(\hat{\beta_1}).$$

3. (c) We can see that the proportionate change in the expected wage from $educ_0 \to educ_1$ is given by,

$$[\exp(\beta_2 educ_1) - \exp(\beta_2 educ_0)]/\exp(\beta_2 educ_0) = \exp[\beta_2(educ_1 - educ_0)] - 1 = \exp(\beta_2 \Delta educ) - 1.$$

Now, applying this to the argument in part b, take $\hat{\theta}_2 = 100 \cdot [\exp(\beta_2 \Delta educ) - 1]$ and then,

$$se(\hat{\theta_2} = 100 \cdot |\Delta educ| \exp(\hat{\beta_2} \Delta educ) se(\hat{\beta_2}).$$

4. (d) We find that for the estimated model, $\hat{\beta_1} = .199, se(\hat{\beta_1}) = .039, \hat{\beta_2} = .065$, and then finally we have $se(\hat{\beta_2}) = .006$. Therefore, we can find that $\hat{\theta_1} = 22.01$ and $se(\hat{\theta_1}) = 4.76$. For $\hat{\theta_2}$, we take $\Delta educ = 4$. Then, we have $\hat{\theta_2} = 29.7$ and $se(\hat{\theta_2}) = 3.11$

3 Question 3

Problem: 4.2

1. See that for each i we have, $E(u_i|X) = 0$. By independence of i, we see that $E(u_i|X) = E(u_i|x_i)$ because (u_i, x_i) is independent of the explanatory variables for all other observations. Letting U be the $N \times 1$ vector of all errors, we can see this would imply $E(U|X) = 0$. Yet, consider that if $\hat{\beta} = \beta + (X'X)^{-1}X'U$ and so,

$$E(\hat{\beta}|X) = \beta + (X'X)^{-1}X'E(U|X) = \beta + (X'X)^{-1}X' \cdot 0 = \beta$$

2. (b) Recognize that for $\hat{\beta}$,

$$Var(\hat{\beta}|X) = Var[(X'X)^{-1}X'U|X] = (X'X)^{-1}X'Var(U|X)X(X'X)^{-1}.$$

Since we know htat $E(U|X) = 0, Var(U|X) = E(UU'|X)$. So, we can see that for the diagonal terms,

$$E(u_i^2|X) = E(u_i^2|x_i) = Var(u_i|x_i) = \sigma^2,$$

where the least equality is equal to the homoskedasticity assumption. Then, for each of the covariance terms we need to show that $E(u_i u_h|X) = 0, \forall i \neq h, i, h = 1, ..., N$. Recognizing that $E(u_i u_h|X) = E(u_i u_h|x_i x_h)$ and $E(u_i|x_i, u_h, x_h) = E(u_i|x_i) = 0$. Then, $E(u_i u_h|x_i, u_h, x_h) = E(u_i|x_i, u_h, x_h)u_h = 0$. It follows by iterated expecations that conditioning yields a zero conditional mean, represented by $E(u_i u_h|x_i, x_h) = 0$. Thus, the proof is completed.

4 Question 4

Problem: 4.4

1. More proofs.
2. Recognize that for each $i, \hat{u}_i = y_i - x_i\hat{\beta} = u_i - x_i(\hat{\beta} - \beta)$ and so,
$$\hat{u}_i^2 = u_i^2 - 2u_i x_i(\hat{\beta} - \beta) + [x_i(\hat{\beta} - \beta)]^2. \quad (3)$$
$$\implies N^{-1}\sum_{i=1}^{N}\hat{u}_i^2 x_i' x_i = N^{-1}\sum_{i=1}^{N} u_i^2 x_i' x_i - 2N^{-1}\sum_{i=1}^{N}[u_i x_i(\hat{\beta} - \beta)]x_i' x_i + N^{-1}\sum_{i=1}^{N}[x_i(\hat{\beta} - \beta)]^2 x_i' x_i. \quad (4)$$
After algebra, we can rewrite our second term as the sum of K terms as,
$$N^{-1}\sum_{i=1}^{N}[u_i x_{ij}(\hat{\beta}_j - \beta_j)]x_i' x_i = (\hat{\beta}_j - \beta_j)N^{-1}\sum_{i=1}^{N}(u_i x_{ij})x_i' x_i = o_p(1)_p(1), \quad (5)$$
in which $\hat{\beta}_j - \beta_j = o_p(1)$ and $N^{-1}\sum_{i=1}^{N}(u_i x_{ij})x_i' x_i = O_p(1)$ whenever we have that $E[|u_i x_{ij} x_{ih} x_{ik}|] < \infty$ for all j, h, k. Following through with the third term, this can be writtern as sum of K^2 terms,
$$(\hat{\beta}_j - \beta_j)(\hat{\beta}_h - \beta_h)N^{-1}\sum_{i=1}^{N}(x_{ij} x_{ih})x_i' x_i = o_p(1) \cdot o_p(1) \cdot O_p(1) = o_p(1), \quad (6)$$
in which $N^{-1}\sum_{i=1}^{N}(x_{ij} x_{ih})x_i' x_i = O_p(1)$ whenever $E[|x_{ij} x_{ih} x_{ik} x_{im}|] < \infty, \forall j, h, k, m$. Thus, we've shown,
$$N^{-1}\sum_{i=1}^{N}\hat{u}_i^2 x_i' x_i = N^{-1}\sum_{i=1}^{N} u_i^2 x_i' x_i + o_p(1), \quad (7)$$
thus we are done.

5 Question 5

Problem: 4.9

1. Consider that we can simply subtract $\log(y_{-1})$ from both sides yielding,
$$\Delta\log(y) = \beta_0 + x\beta + (\alpha_1 - 1)\log(y_{-1}) + u. \quad (8)$$
Clearly, we can recognize that the intercept and slope estimates are the same for x.Thus, we have that the coef for $\log(y_{-1})$ becomes $\alpha_1 - 1$.

6 Question 6

Problem: Show that for a regression model, if a regressor x_j is measured with error, then it will be endogenous.

Proof: If a regressor x_j is measured with error, it is called an endogenous regressor. This is because the measurement error in x_j is correlated with the error term in the regression model, causing a bias in the estimated coefficient for x_j. To demonstrate this, consider a simple linear regression model:

$y_i = \beta_0 + \beta_1 x_{i,j} + u_i$

where y_i is the dependent variable, $x_{i,j}$ is the endogenous regressor measured with error, β_0 and β_1 are the coefficients, and u_i is the error term. The measurement error in $x_{i,j}$ is represented by the term ϵ_i. Therefore, the true value of $x_{i,j}$ is given by $x^*_{i,j} = x_{i,j} + \epsilon_i$.

Substituting this into the regression equation gives:

$y_i = \beta_0 + \beta_1(x_{i,j} + \epsilon_i) + u_i$

This implies that the error term u_i and the measurement error ϵ_i are correlated, as they both affect the value of y_i. Therefore, the estimated coefficient for $x_{i,j}$, $\hat{\beta}_1$, will be biased.