Problem Set 4

1. (a) **OLS.1:**$\mathrm{E}(u|x) = 0 \Rightarrow \mathrm{E}(x'u) = 0$, remember that $x = (1\ x_1 \ldots x_K)$ Meaning that u has mean zero and is uncorrelated to all regressors.
OLS.2: rank $\mathrm{E}(x'x) = K$. Meaning that $X'X$ has the full rank and invertible, so there's no perfect collinearity between regressors.
OLS.3: $\mathrm{E}(u^2 x'x) = \sigma^2 \cdot \mathrm{E}(x'x)$ (homoscedasticity). Squared errors are uncorrelated with all x_i, x_i^2 and $x_i x_j$. Meaning, the value of X does not influence the variance of errors, variance is constant for different values of X.
Assumptions needed for

- identification: OLS.1 + OLS.2
- unbiasedness: in general, OLS estimator is biased (see slide 25 and 26)
- consistency: OLS.1 + OLS.2
- asymptotic normality: OLS.1 + OLS.2

Hence, OLS.3 can simply make estimation more efficient (lower variance estimation). But it is hard to set this assumption in practice.

(b) OLS aims to minimize the SSR. In matrix notation the SSR is

$$\begin{aligned} U'U &= (y - X\beta)'(y - X\beta) = y'y - 2X'y\beta + \beta'X'X\beta \\ \frac{\partial SSR}{\partial \beta} &= -2X'y + 2X'X\beta \stackrel{!}{=} 0 \end{aligned}$$

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

<u>Consistency:</u>

$$\begin{aligned} \hat{\beta} &= \beta + (X'X)^{-1}X'U = \beta + \left(\sum x_i' x_i\right)^{-1} \cdot \sum x_i' u_i \\ &= \beta + \left(\frac{1}{N}\sum x_i' x_i\right)^{-1} \cdot \frac{1}{N}\sum x_i' u_i \\ \text{plim}(\hat{\beta}) &= \beta + \text{plim}\left[\left(\frac{1}{N}\sum x_i' x_i\right)^{-1} \cdot \frac{1}{N}\sum x_i' u_i\right] \\ &\stackrel{Slutsky}{=} \beta + \left(\text{plim}\left(\frac{1}{N}\sum x_i' x_i\right)\right)^{-1} \cdot \text{plim}\left(\frac{1}{N}\sum x_i' u_i\right) \end{aligned}$$

Since x_i is iid, each element of matrix $x_i' x_i$ is iid.
$\Rightarrow \left[\text{plim}\left(\frac{1}{N}\sum x_i' x_i\right)\right]^{-1} \underset{WLLN}{\rightarrow} [E(x'x)]^{-1}$ which exists by OLS.2 The same reasoning holds for the second term:
$\text{plim}\left(\frac{1}{N}\sum x_i' u_i\right) \underset{WLLN}{\rightarrow} E(x'u) = 0$ by assumption OLS.1
$\text{plim}(\hat{\beta}) = \beta + (E(x'x))^{-1} \cdot 0 = \beta$

Asymptotic normality: Rearrange the OLS formula from the consistency proof:

$$\begin{aligned}\hat\beta-\beta &= \left(\frac{1}{N}\sum x_i'x_i\right)^{-1}\cdot\left(\frac{1}{N}\sum x_i'u_i\right)\\ \sqrt{N}(\hat\beta-\beta) &= \left(\frac{1}{N}\sum x_i'x_i\right)^{-1}\cdot\left(\frac{1}{\sqrt{N}}\sum x_i'u_i\right)\end{aligned}$$

Asymptotic behavior of two parts:

i. $\text{plim}\left(\frac{1}{N}\sum x_i'x_i\right)\overset{WLLN}{=}\text{E}(x'x)$

ii. Since $x_i'u_i$ is iid and $\text{E}(x_i'u_i)=0$ by OLS.1 we can apply the CLT on it:

$N^{-\frac{1}{2}}\sum x_i'u_i\xrightarrow{d}N(0,B)$, where $B=\text{Var}(x_i'u_i)=\text{E}(u_i^2x_i'x_i)$.

Since $\left(\frac{1}{N}\sum x_i'x_i\right)\xrightarrow{p}\text{E}(x'x)^{-1}=A^{-1}$ and $\frac{1}{\sqrt{N}}\sum x_i'u_i\xrightarrow{d}z$, where $z\sim N(0,B)$, we can apply Cramer's Theorem on $\sqrt{N}(\hat\beta-\beta)$:

$$\begin{aligned}\sqrt{N}(\hat\beta-\beta)\xrightarrow{d}\tilde z=A^{-1}z\text{ with }\text{E}(\tilde z) &= A^{-1}\,\text{E}(z)=A^{-1}\,\text{E}(x_i'u_i)=0\\ \text{and }\text{Var}(\tilde z) &= \text{E}(A^{-1}zz'A^{-1})\\ &= A^{-1}\,\text{E}(u_i^2x_i'x_i)A^{-1}\\ &= A^{-1}BA^{-1}\\ \rightarrow\sqrt{N}(\hat\beta-\beta)\xrightarrow{d}N(0,A^{-1}BA^{-1})\end{aligned}$$

2. (a) Using the formula for OV bias from p.10 of the formulary: $\text{plim}\,\hat\beta_i=\beta_i+\gamma\delta_i$ where δ_i measures the relationship between x_i and OV and γ measures the relationship between OV and y. A higher number of semesters is associated with a higher average grade. But a higher number of semesters is associated with a lower ability ($\delta_i<0$) and a lower ability is associated with a higher average grade ($\gamma<0$). Therefore $\hat\beta$ is larger than β and we are faced with an upward bias.

   (b) In order to reduce this bias we should include ability into the model. However, ability is not observed and thus cannot be included. We have to find a proxy for it, e.g. the average grade of participants achieved in the bachelor program or an IQ test score. The proxy should be added to the original regression.

3. (a) To get some intuition, assume $\hat\beta_1=0.05$. If a man is married, his wage increases by 5% (log-linear interpretation, Y in logs, X binary). Now calculate the exact effect θ_1. Denote further married $\coloneqq x_1$, educ $\coloneqq x_2$.

$$\text{E}(wage|x_1,x_2,z)=\text{E}(e^{\log(wage)}|x_1,x_2,z)=\text{E}(\exp(\beta_0+\beta_1x_1+\beta_2x_2+z\gamma+u)|x_1,x_2,z)$$
$$=\exp(\beta_0+\beta_1x_1+\beta_2x_2+z\gamma)\cdot\underbrace{\text{E}(\exp(u)|x_1,x_2,z)}_{\substack{\text{if } x_i,u \text{ independent,}\\ \text{this is constant}=:\delta_0}}=\delta_0\cdot\exp(\beta_0+\beta_1x_1+\beta_2x_2+z\gamma)$$

The exact percentage difference θ_1 is given as:

$$\frac{\text{E}(wage|x_1=1,x_2,z)-\text{E}(wage|x_1=0,x_2,z)}{\text{E}(wage|x_1=0,x_2,z)}\cdot 100=$$
$$=\frac{\delta_0\cdot\exp(\beta_0+\beta_1+\beta_2x_2+z\gamma)-\delta_0\cdot\exp(\beta_0+\beta_2x_2+z\gamma)}{\delta_0\cdot\exp(\beta_0+\beta_2x_2+z\gamma)}\cdot 100=(\exp(\beta_1)-1)\cdot 100$$

Check exact effect for the example above: $(\exp(0.05) - 1) \cdot 100 = 5.13\%$.
For higher effects approximation gets worse: $(\exp(0.2) - 1) \cdot 100 = 22.14\%$ instead of 20%.

(b) Using the Delta method,

$$\begin{aligned}
\theta_1 &= g(\beta_1) = (\exp(\beta_1) - 1) \cdot 100 \\
\frac{\partial g}{\partial \beta_1} &= \exp(\beta_1) \cdot 100 \\
\Rightarrow \text{Avar}(\hat{\theta}_1) &= \exp(\beta_1) \cdot 100 \cdot \text{Avar}(\hat{\beta}_1) \cdot \exp(\beta_1) \cdot 100 \\
&= 10000 \cdot \exp(2\beta_1) \cdot \text{Avar}(\hat{\beta}_1) \\
\Rightarrow \text{ASE}(\hat{\theta}_1) &= 100 \cdot \exp(\beta_1) \cdot \text{ASE}(\hat{\beta}_1) \\
\Rightarrow \widehat{\text{ASE}}(\hat{\theta}_1) &= 100 \cdot \exp(\hat{\beta}_1) \cdot \widehat{\text{ASE}}(\hat{\beta}_1)
\end{aligned}$$

(c)
$$\begin{aligned}
\theta_2 &= \frac{\delta_0 \exp(\beta_0 + \beta_1 x_1 + \beta_2 (x_2 + \Delta x_2) + z\gamma) - \delta_0 \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + z\gamma)}{\delta_0 \cdot \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + z\gamma)} \cdot 100 \\
&= (\exp(\beta_2 \Delta x_2) - 1) \cdot 100
\end{aligned}$$

Again, for high values of β_2 or Δx_2 the approximation gets imprecise. Now, using the Delta method for $\theta_2 = g(\beta_2)$,

$$\begin{aligned}
\frac{\partial g(\beta_2)}{\partial \beta_2} &= 100 \cdot \Delta x_2 \cdot \exp(\beta_2 \Delta x_2) \\
\Rightarrow \text{Avar}(\hat{\theta}_2) &= 10000 \cdot \Delta x_2^2 \cdot \exp(2\beta_2 \Delta x_2) \cdot \text{Avar}(\hat{\beta}_2) \\
\Rightarrow \widehat{\text{ASE}}(\hat{\theta}_2) &= 100 \cdot \Delta x_2 \cdot \exp(\hat{\beta}_2 \Delta x_2) \cdot \widehat{\text{ASE}}(\hat{\beta}_2)
\end{aligned}$$

4. We are looking for the $\text{plim}(\frac{1}{N}\sum_{i=1}^N \hat{u}_i^2 \boldsymbol{x}_i' \boldsymbol{x}_i)$. Note that

$$\begin{aligned}
\hat{u}_i &= y_i - \boldsymbol{x}_i \hat{\boldsymbol{\beta}} = \boldsymbol{x}_i \boldsymbol{\beta} + u_i - \boldsymbol{x}_i \hat{\boldsymbol{\beta}} = \underbrace{\boldsymbol{x}_i}_{1\times N} \underbrace{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}_{N\times 1} + u_i \\
\hat{u}_i^2 &= u_i^2 + 2u_i \boldsymbol{x}_i (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \boldsymbol{x}_i' \boldsymbol{x}_i (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})
\end{aligned}$$

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^N \hat{u}_i^2 \boldsymbol{x}_i' \boldsymbol{x}_i &= \frac{1}{N}\sum_{i=1}^N u_i^2 \boldsymbol{x}_i' \boldsymbol{x}_i + \frac{1}{N}\sum_{i=1}^N 2u_i \boldsymbol{x}_i \underbrace{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}_{o_p(1)} \boldsymbol{x}_i' \boldsymbol{x}_i \\
&+ \frac{1}{N}\sum_{i=1}^N \underbrace{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'}_{o_p(1)} \boldsymbol{x}_i' \boldsymbol{x}_i \underbrace{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}_{o_p(1)} \boldsymbol{x}_i' \boldsymbol{x}_i \\
\text{plim}\left(\frac{1}{N}\sum_{i=1}^N \hat{u}_i^2 \boldsymbol{x}_i' \boldsymbol{x}_i\right) &= \text{E}(u^2 \boldsymbol{x}' \boldsymbol{x}) + o_p(1)
\end{aligned}$$

5. Possible sources of endogeneity of Age:

i) Classical OVB: Age (strongly) correlates with experience which in turn has clearly an effect on wage.

ii) Measurement error: Could be random, or participants in the survey do not like to communicate their true age (e.g. they claim to be younger).

iii) Sample selection error: If the study is not fully random, the responds may depend heavily on who is asked (e.g. different districts of a city, different states or countries).

6. (a) For example, an omitted variable like *family income* is correlated with *PC* (and *colGPA*), which causes omitted variable bias.

   (b) In general it is difficult to state, since only partial correlation is considered, but the partial correlation goes into the sum in the same direction ($\gamma > 0$, $\delta_3 > 0$). Thus an upward asymptotic bias is likely.

   (c) Possible candidates depending on the availability can be *family income*, *house prices*, *ZIP codes* etc.

7. (a) Because each x_j has finite second moments (by assumption), $\mathrm{Var}(x\beta) < \infty$. Further note, $\mathrm{Var}(u) < \infty$ and $\mathrm{Cov}(x\beta, u) = 0$, since each x_j is uncorrelated with u. Therefore,

$$\mathrm{Var}(y) = \mathrm{Var}(x\beta + u) \overset{\mathrm{Cov}(x\beta,u)=0}{=} \mathrm{Var}(x\beta) + \mathrm{Var}(u) = \mathrm{Var}(x\beta) + \sigma_u^2$$

   Hence, $\sigma_y^2 = \mathrm{Var}(x\beta) + \sigma_u^2$.

   (b)
$$\mathrm{Var}(y_i) = \mathrm{Var}(x_i\beta + u_i) \overset{ind}{=} \mathrm{Var}(x_i\beta) + \mathrm{Var}(u_i) = \mathrm{Var}(x_i\beta) + \sigma^2$$

   The statement of the exercise only makes sense, if we consider x_i deterministic (not a random variable), which we do not in this course. Then $\mathrm{Var}(x_i\beta)$ would be zero. But we view x_i as a random draw along with y_i.

   (c)
$$\begin{aligned}
R^2 &= 1 - \frac{SSR}{TSS} = 1 - \frac{\sum_{i=1}^N \hat{u}_i^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \\
&= 1 - \frac{\frac{1}{N}\sum_{i=1}^N \hat{u}_i^2}{\frac{1}{N}\sum_{i=1}^N (y_i - \bar{y})^2} \underset{\mathrm{WLLN}}{\overset{\mathrm{p}}{\rightarrow}} 1 - \frac{\mathrm{E}(u^2)}{\mathrm{E}((y - \mu_y)^2)} \\
&\overset{\mathrm{E}(u)=0}{=} 1 - \frac{\mathrm{Var}(u)}{\mathrm{Var}(y)} = 1 - \frac{\sigma_u^2}{\sigma_y^2} = \rho^2
\end{aligned}$$

   Hence, $R^2 \overset{p}{\rightarrow} \rho^2$, or in other notation, $\mathrm{plim}(R^2) = \rho^2)$.

   (d) The derivation in part (c) assumed nothing about $\mathrm{Var}(u|x)$. The population R^2 (denoted ρ^2) depends only on the unconditional variances of u and y. Regardless of the nature of heteroskedasticity ($\mathrm{Var}(u|x)$), the regular R^2 consistently estimates the population R^2, ρ^2.

8. (a) The central limit theorem describes the behavior of sums of random variables. The asymptotic normality of $\bar{x} = \frac{1}{N}\sum x_i = \frac{1}{\sqrt{N}}N^{-1/2}\sum x_i$ or $\bar{x}^2$ does not imply that x_i is normally distributed. The distribution of x_i or u_i is unknown and does not change with the sample size.

   (b) The statement is simply wrong. Instead it can be argued that $\hat{u}_i^2$ is a consistent estimator for u_i^2 (see problem 5). But $u_i^2 = \mathrm{E}(u_i^2|x_i) + v_i$, where $\mathrm{E}(v_i|x_i) = 0$. Fortunately, we do not need a consistent estimator for $\mathrm{E}(u_i^2|x_i)$ for inference. We only need one for $\mathrm{E}(u^2x'x)$, which can be found (see problem 5).