Economics 361
Problem Set #4

Jun Ishii *
Department of Economics
Amherst College

Fall 2022

Question 1: Ordinary Least Squares (OLS) Model

(a) Show that

$$\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X}_N)(Y_i-\bar{Y}_N) = \frac{1}{N}\sum_{i=1}^{N}(X_iY_i) - \left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}Y_i\right)$$

and

$$\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X}_N)^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i^2) - \left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)^2$$

ANS: Note that both $\bar{X}_N$ and $\bar{Y}_N$ are scalars that can be be taken "out" of the summation

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X}_N)(Y_i-\bar{Y}_N) &= \frac{1}{N}\sum_{i=1}^{N}\left(X_iY_i - X_i\bar{Y}_N - Y_i\bar{X}_N + \bar{X}_N\bar{Y}_N\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}X_iY_i - \bar{X}_N\frac{1}{N}\sum_{i=1}^{N}Y_i - \bar{Y}_N\frac{1}{N}\sum_{i=1}^{N}X_i + \frac{1}{N}N\bar{X}_N\bar{Y}_N \\
&= \frac{1}{N}\sum_{i=1}^{N}X_iY_i - \bar{X}_N\bar{Y}_N - \bar{Y}_N\bar{X}_N + \bar{X}_N\bar{Y}_N \quad = \quad \frac{1}{N}\sum_{i=1}^{N}X_iY_i - \bar{X}_N\bar{Y}_N
\end{aligned}$$

The steps for $\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X}_N)^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X}_N)(X_i-\bar{X}_N)$ are similar.

(b) Let $Z \equiv Y - BLP_{MSE}(Y|X)$. Show that $E[Z]=0$ and $\text{Cov}(X,Z)=0$.

*Office: Converse Hall 315 Phone: (413) 542-2901 E-mail: jishii@amherst.edu

ANS: Recall that

$$\begin{aligned} BLP_{MSE}(Y|X) &= \underbrace{(E[Y] - b^* E[X])}_{a^*} + b^* X \\ & \text{where } b^* = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \end{aligned}$$

Therefore

$$\begin{aligned} E[Z] &= E[\ Y - (E[Y] - b^* E[X] + b^* X)\ ] \\ &= E[Y] - \underbrace{E[E[Y]]}_{=E[Y]} - b^* \left(\underbrace{E[E[X]]}_{=E[X]} - E[X] \right) = 0 \\ \text{Cov}(X,Z) &= E[XZ] - \underbrace{E[X]\ E[Z]}_{=0 \text{ as } E[Z]=0} \\ &= E[\ XY - X\ \ (E[Y] - b^* E[X] + b^* X)\ ] \\ &= \underbrace{E[XY] - E[X]E[Y]}_{\text{Cov}(X,Y)} - b^* \underbrace{\left(E[X^2] - (E[X])^2\right)}_{\text{Var}(X)} \\ &= \text{Cov}(X,Y) - \frac{\text{Cov}(X,Y)}{\text{Var}(X)} \text{Var}(X) = 0 \end{aligned}$$

(c) Let $\tilde{Z} \equiv Y - \tilde{a} - \tilde{b}X$. Solve for the $(\tilde{a}, \tilde{b})$ that set (i) the sample mean of $\tilde{Z}$ equal to zero and (ii) the sample covariance between X and $\tilde{Z}$ equal to zero.

ANS: From $\tilde{Z}_N = 0$

$$\begin{aligned} \tilde{Z}_N &= \frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \tilde{a} - \tilde{b} X_i \right) = 0 \\ &= \underbrace{\frac{1}{N} \sum_{i=1}^{N} Y_i}_{\bar{Y}_N} - \tilde{a} - \tilde{b} \underbrace{\frac{1}{N} \sum_{i=1}^{N} X_i}_{\bar{X}_N} = 0 \\ \Rightarrow\ \tilde{a} &= \bar{Y}_N - \tilde{b} \bar{X}_N \end{aligned}$$

From the sample covariance of X, Z equal to 0

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X}_N)(\tilde{Z}_i - \tilde{Z}_N) &= \frac{1}{N}\sum_{i=1}^{N}(X_i\tilde{Z}_i) - \bar{X}_N\tilde{Z}_N \qquad \text{using result from (a)} \\
&= \frac{1}{N}\sum_{i=1}^{N}\Big(X_i\ (Y_i - \tilde{a} - \tilde{b}X_i)\Big) \qquad \text{as } \tilde{Z}_N = 0 \\
&= \frac{1}{N}\sum_{i=1}^{N}X_iY_i - \underbrace{\Big(\bar{Y}_N - \tilde{b}\bar{X}_N\Big)}_{\tilde{a}}\bar{X}_N - \tilde{b}\frac{1}{N}\sum_{i=1}^{N}X_i^2 \\
&= \left(\frac{1}{N}\sum_{i=1}^{N}X_iY_i - \bar{X}_N\bar{Y}_N\right) - \tilde{b}\left(\frac{1}{N}\sum_{i=1}^{N}X_i^2 - (\bar{X}_N)^2\right) \\
\Rightarrow \ \tilde{b} &= \frac{\frac{1}{N}\sum_{i=1}^{N}X_iY_i - \bar{X}_N\bar{Y}_N}{\frac{1}{N}\sum_{i=1}^{N}X_i^2 - (\bar{X}_N)^2}
\end{aligned}
$$

(d) Use the above results to explain the following:

> The OLS estimator is the moment-based estimator that equates the population moments $E[Z] = 0$ and $\text{Cov}(X, Z) = 0$ to their sample analog.

ANS: Note that $(\tilde{a}, \tilde{b})$ above are simply the bivariate OLS estimators. So the OLS estimators can be derived by setting the sample mean of Z equal to the population mean (0) and the sample covariance of X, Z equal to the population covariance (0). This suggests that the OLS estimator can be rationalized as the moment-based estimator with $E[Z]$ and $\text{Cov}(X, Z)$ as the relevant moments.

(e) Show that $E[XZ] = 0$

ANS: In part (d) we showed that $E[Z] = 0$ and $\text{Cov}(X, Z) = 0$. Note that $\text{Cov}(X, Z) = E[XZ] - E[Z]\ E[Z]$. Therefore, $E[XZ] = 0$

(f) Let $\tilde{Z}$ be defined as in **(c)**. This time, solve for the $(\tilde{a}, \tilde{b})$ that set (i) the sample mean of Z equal to zero and (ii) the sample mean of XZ equal to zero. Note that

$$
\text{Sample Mean of } XZ \ = \ \frac{1}{N}\sum_{i=1}^{N}X_i\tilde{Z}_i
$$

ANS: From **(a)**, we can see that imposing the sample mean of Z equal to zero **and** the sample covariance of X and Z equal to zero implies that the sample mean of XZ must be zero, as the sample covariance equals the sample mean of XZ minus the product of the sample mean of X and the sample mean of Z. Therefore, the answer here must be the same as the answer in **(c)**

(g) Use the above results to explain the following:

> The OLS estimator is the moment-based estimator that equates the population moments $E[Z]=0$ and $E[XZ]=0$ to their sample analog.

ANS: See answer for **(d)**

Moment conditions of the form $E[XZ]=0$ are known as **orthogonality conditions**. We will be exploring estimators based on other orthogonality conditions later in the course.

Above, we defined Z as the difference between Y and the $BLP_{MSE}(Y|X)$ and rationalized the OLS estimator using the orthogonality condition between X and this Z. However, OLS is often rationalized using the orthogonality condition between X and the difference between Y and the $BP_{MSE}(Y|X)$. We will not fully develop this particular rationalization here. But we can do the first two steps ... Note: $BP_{MSE}(Y|X)$ is not necessarily the same as $BLP_{MSE}(Y|X)$

(h) Let $W \equiv Y - E[Y|X]$. Show that $E[W]=0$ and $\mathrm{Cov}(X,W)=0$.

ANS: Perhaps easiest way to show this is to use the Law of Iterated Expectations

$$\begin{aligned}
E[W] &= E[Y - E[Y|X]] = E[Y] - E_X[E[Y|X]] = E[Y] - E[Y] = 0 \\
\mathrm{Cov}(X,W) &= E[XW] - E[X]E[W] = E[XW] = E_X[E[XW|X]] \\
&= E_X[\ E[XY - XE[Y|X] \mid X\ ]\ ] = E_X[\ XE[Y|X] - XE[Y|X]\ ] = 0
\end{aligned}$$

(i) Show that $E[XW]=0$

> **ANS**: Showed it while deriving $\mathrm{Cov}(X,W)=0$

Violation of the orthogonality condition $E[XW]=0$ is often dubbed an "**endogeneity** problem." We will explore this issue later in the course.

Question 2: A Production Example of OLS in Action

Suppose that firms in an industry produced output (Y) using two types of input, capital (K) and labor (L). The relationship between the amount of inputs used and the amount of output produced is given by the following **production function**:

$$Y = AK^{\beta}L^{\gamma} \tag{1}$$

This particular function is known as the **Cobb-Douglas** function, named after Charles Cobb and Paul Douglas, two former Amherst College professors.[1] The Cobb-Douglas function is one of the most widely used and known functions in economics.

The function consists of two key parameters $\{\beta, \gamma\}$ that are of economic interests. In this question, we show how one might use the ordinary least squares (OLS) model to estimate those unknown parameters using only a sample of $\{Y, K, L\}$. In a sense, we are re-tracing the steps of Paul Douglas.

Consider the following linear equation involving Y, K, L

$$ln(Y) = ln(A) + \alpha_1 \; ln(K) + \alpha_2 \; ln(L) \tag{2}$$

where $ln(\cdot)$ refers to the natural log transformation and $\{\alpha_1, \alpha_2\}$ are parameters

(a) What is the relationship between $\{\alpha_1, \alpha_2\}$ and $\{\beta, \gamma\}$?

ANS: $\alpha_1 = \beta$ and $\alpha_2 = \gamma$. Just take the natural log of $Y = AK^{\beta}L^{\gamma}$ and see that $ln(Y) = lnA + \beta \; ln(K) + \gamma \; ln(L)$

Suppose you are given the following *random* sample of size $N > 3$ from a population (Y, K, L) determined by the above Cobb-Douglas production function

$$\{ \; (Y_1 = y_1, A_1 = a_1, K_1 = k_1, L_1 = l_1) \cdots (Y_N = y_N, A_N = a_N, K_N = k_N, L_N = l_N) \; \}$$

(b) Explain how you *may* be able to **solve** for $\{\alpha_1, \alpha_2\}$. What must be true of the random sample in order for you to be able to solve for $\{\alpha_1, \alpha_2\}$?

ANS: With a sample of $N > 2$, you essentially have more linear equations (N) than unknowns (2 – α_1, α_2)

$$\begin{aligned} ln(Y_1) &= ln(A_1) + \alpha_1 \; ln(K_1) + \alpha_2 \; \ln(L_1) \\ ln(Y_2) &= ln(A_2) + \alpha_1 \; ln(K_2) + \alpha_2 \; \ln(L_2) \\ &\vdots \\ ln(Y_N) &= ln(A_N) + \alpha_1 \; ln(K_N) + \alpha_2 \; \ln(L_N) \end{aligned}$$

Therefore, as long as 2 of the equations are linearly independent, you can solve for $\{\alpha_1, \alpha_2\}$ by solving the system of equation. There is no estimation involved. The *exact* values of $\{\alpha_1, \alpha_2\}$ can

[1] For the history, see "The Cobb-Douglas Production Function Once Again: Its History, Its Testing, and Some New Empirical Values," by P.H. Douglas, *Journal of Political Economy*, October 1976, pp.903-915.

be solved. This requirement of at least 2 linearly independent equations is analogous to the full rank assumption for the OLS model ...

Suppose that you did **not** observe $\{A_i\}_{i=1}^N$. Each A_i is now an unobserved "productivity shock" and A effectively a random variable.

(c) How does the above affect your ability to **solve** for $\{\alpha_1, \alpha_2\}$ using the random sample?

ANS: You can no longer solve the system of equations as you always have more unknowns than equations. With $ln(A)$ random, this means that the value of $ln(A)$ can differ for each observation – so each $ln(A_i)$ is essentially an unknown!

$$\begin{aligned} ln(Y_1) &= ln(A_1) + \alpha_1\ ln(K_1) + \alpha_2\ \ln(L_1) \\ ln(Y_2) &= ln(A_2) + \alpha_1\ ln(K_2) + \alpha_2\ \ln(L_2) \\ &\vdots \\ ln(Y_N) &= ln(A_N) + \alpha_1\ ln(K_N) + \alpha_2\ \ln(L_N) \end{aligned}$$

So there are now N+2 unknowns: $\{\alpha_1, \alpha_2\}$ plus $\{ln(A_1), \ldots, ln(A_N)\}$. But there are still only N equations at most. This system of equations is *underidentified* and each of the parameters cannot be uniquely solved.

Suppose you know that $E[ln(A) \mid ln(K), ln(L)] = \mu$, where μ is some parameter; the conditional mean of $ln(A)$ given $(ln(K), ln(L))$ does not vary with $(ln(K), ln(L))$.

(d) What is the $E[ln(Y) \mid ln(K), ln(L)]$ in terms of $\{\mu, \alpha_1, \alpha_2, ln(K), ln(L)\}$?

ANS:

$$\begin{aligned} E[ln(Y)|ln(K), ln(L)] &= E[ln(A) + \alpha_1\ ln(K) + \alpha_2\ ln(L)|ln(K), ln(L)] \\ &= E[ln(A)|ln(K), ln(L)] + \alpha_1\ ln(K) + \alpha_2\ ln(L) \\ &= \mu + \alpha_1\ ln(K) + \alpha_2\ ln(L) \end{aligned}$$

(e) Using your answers in **(a)-(d)**, explain how you may properly apply the ordinary least squares (OLS) model to your sample and obtain estimates of $\{\mu, \alpha_1, \alpha_2\}$.

ANS: We know from earlier discussion that OLS can provide us with a moment-based estimator of the parameters of the $\text{BLP}_{MSE}(Y|X)$. The answer in **(d)** indicates that the Linearity Condition is satisfied, implying that the $\text{BLP}_{MSE}(Y|X)$ is also the $\text{BP}_{MSE}(Y|X)$. This means that $\{\mu, \alpha_1, \alpha_2\}$ also correspond to the parameters of the $\text{BLP}_{MSE}(Y|X)$. Thus, OLS for this problem can be thought as providing us with a moment-based estimator of $\{\mu, \alpha_1, \alpha_2\}$. Moreover, given that we have a random sample, the Spherical Errors assumption is also satisfied. As long as the sample is such that we satisfy the full rank assumption, the OLS estimator should be the minimum variance linear unbiased estimator of $\{\mu, \alpha_1, \alpha_2\}$, according to the Gauss-Markov theorem.

$$E[b^{ols}|ln(K), ln(L)] = E\left[\begin{array}{c|c} b_0^{ols} & \\ b_1^{ols} & ln(K), ln(L) \\ b_2^{ols} & \end{array} \right] = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

Define the following matrices concerning your $N = 10$ sized random sample

$$L = \begin{pmatrix} ln(y_1) \\ \vdots \\ ln(y_{10}) \end{pmatrix} \qquad H = \begin{pmatrix} 1 & ln(k_1) & ln(l_1) \\ \vdots & \vdots & \vdots \\ 1 & ln(k_{10}) & ln(l_{10}) \end{pmatrix}$$

L is a (10×1) column vector and H is a (10×3) matrix.

Your trusty research assistant, Sam, has done the following calculations for you:

$$(H'H) = \begin{pmatrix} 10.00000 & 51.68789 & 31.29016 \\ 51.68789 & 269.3835 & 159.9216 \\ 31.29016 & 159.9216 & 113.8035 \end{pmatrix} \qquad (H'L) = \begin{pmatrix} 33.84597 \\ 174.65849 \\ 117.20837 \end{pmatrix}$$

$$(H'H)^{-1} = \begin{pmatrix} 15.87792 & -2.74415 & -0.50941 \\ -2.74415 & 0.496662 & 0.056572 \\ -0.50941 & 0.056572 & 0.069352 \end{pmatrix}$$

(f) Use the above calculations provided by Sam to calculate the OLS estimate of $\{\mu, \alpha_1, \alpha_2\}$

ANS:

$$b^{ols} = (H'H)^{-1}(H'L) = \begin{pmatrix} \hat{\mu}_\lambda \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} -1.59387 \\ 0.498338 \\ 0.767866 \end{pmatrix}$$

There may be some rounding differences.

An important concept in economics concerning production is **economies of scale.** If output doubles when we double the amount of capital and labor, our production function exhibits **constant returns to scale**. If output increases by less than double, it exhibits **decreasing returns to scale**. And if output increases by more than double, it exhibits **increasing returns to scale.**

(g) Based on your answer in **(f)**, do you think the production function that generated your random sample exhibits constant, decreasing, or increasing returns to scale?

ANS: Suppose we increase both K and L by some factor, δ. From Equation (1)

$$Y_{\text{new}} = A(\delta K)^\beta(\delta L)^\gamma = \underbrace{AK^\beta L^\gamma}_{Y} (\delta)^{\beta+\gamma}$$

Therefore, Y also increases by a factor of δ if $\beta + \gamma = 1$ (constant returns). If $\beta + \gamma < 1$ then Y increases by a factor less than δ (decreasing returns). If $\beta + \gamma > 1$ then Y increases by a factor more than δ (increasing returns).

Our OLS estimate of β is 0.3124 and of γ 0.6647. So our OLS estimate of $\beta + \gamma$ is $0.498338 + 0.767866 = 1.266204$. Technically, this implies that our OLS estimates indicate some increasing returns to scale. But note that we have not factored in the precision with which we have estimated

β and γ. The formal answer to this question requires us to conduct a proper **hypothesis test** that factors the precision of our estimates.

ASIDE: The data used to answer **(f)** were computer simulations from a Cobb-Douglas production function where $\beta = 0.4, \gamma = 0.6$. So the estimates are not that bad – especially for a sample of only 10 observations.

You are told that the $\text{Var}(ln(A) \mid ln(K), ln(L)) = \frac{1}{12}$.

(h) What is the $\text{Var}(ln(Y) \mid ln(K), ln(L))$?

ANS:

$$\begin{aligned}
\text{Var}(ln(Y) \mid ln(K), ln(L)) &= \text{Var}(ln(A) + \beta ln(K) + \gamma ln(L) \mid ln(K), ln(L)) \\
&= \text{Var}(ln(A) \mid ln(K), ln(L)) \quad = \quad \frac{1}{12}
\end{aligned}$$

Note: This means "$\sigma^2 = \frac{1}{12}$"

(i) Based on the matrices provided above and your answer to part **(h)**, what is the variance of your OLS estimate for α_1? What is the covariance between your OLS estimate for α_1 and your OLS estimate for α_2?

ANS: Recall that the variance-covariance matrix for the OLS estimators is $\sigma^2(X'X)^{-1}$. As shown above, $\sigma^2 = \frac{1}{12}$ and $(X'X)^{-1}$ here is the same as $(H'H)^{-1}$.

$$\text{Var}\begin{pmatrix} \hat{\mu}_{ols} \\ \hat{\beta}_{ols} \\ \hat{\gamma}_{ols} \end{pmatrix} = \frac{1}{12}\begin{pmatrix} 15.87792 & -2.74415 & -0.50941 \\ -2.74415 & 0.496662 & 0.056572 \\ -0.50941 & 0.056572 & 0.069352 \end{pmatrix}$$

The variance of $\hat{\alpha_1}$ (which is $\hat{\beta}_{ols}$) is $\frac{0.496662}{12} = 0.041389$. The covariance of $\hat{\alpha_1}$ and $\hat{\alpha_2}$ (which is $\hat{\gamma}_{ols}$) is $\frac{0.056572}{12} = 0.004714$.

Let $e_i \equiv ln(y_i) - (\hat{\mu} + \hat{\alpha}_1 ln(k_i) + \hat{\alpha}_2 ln(l_i))$ where $\{\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2\}$ are the OLS estimates for $\{\mu, \alpha_1, \alpha_2\}$ you obtained in **(f)**. So e_i is the OLS residual for the ith observation.

(j) Show that the sample average of the OLS residuals equals zero: $\frac{1}{10}\sum_{i=1}^{10} e_i = 0$

HINT: $\sum_i e_i = \sum_i ln(y_i) - (\sum_i \hat{\mu} + \hat{\alpha}_1 \sum_i ln(k_i) + \hat{\alpha}_2 \sum_i ln(l_i))$

ANS: Note that the first column (as well as the first row) of the matrix $(H'H)$ provides you with the sum of the variables across the observations. The first element is the sum of the "ones" and hence 10 (number of observations), the second element the sum of the $ln(k_i)$, and the third element the sum of the $ln(l_i)$. So $\sum_i ln(k_i) = 51.686789$ and $\sum_i ln(l_i) = 31.29016$. The sum of $ln(y_i)$ is the first element of vector $(H'L)$, 33.84597. Therefore

$$\begin{aligned}
\sum_i e_i &= \sum_i ln(y_i) - \left(\sum_i \hat{\mu} + \hat{\alpha}_1 \sum_i ln(k_i) + \hat{\alpha}_2 \sum_i ln(l_i) \right) \\
&= 33.84597 - (\hat{\mu} \times 10 + \hat{\alpha_1} \times 51.68789 + \hat{\alpha_2} \times 31.29016) \\
&= 33.84597 - (-1.59387 \times 10 + 0.498338 \times 51.68780 + 0.767866 \times 31.29016) \quad = \quad 0
\end{aligned}$$

As $\sum_i e_i = 0$, $\frac{1}{10} \sum_{i=1}^{10} e_i = 0$.

Note: may not get exactly zero due to rounding errors (but should be close)

(k) Briefly explain why the result in **(j)** is not a fluke, not a chance happening for the particular realized sample you were given.

ANS: See page 5 of the "Regression Algebra" Handout

$$\frac{1}{N} \sum_i e_i = \frac{1}{N} \sum_i (Y_i - a - X_i b) = \bar{Y} - a - \bar{X}'b = \bar{Y} - \underbrace{(\bar{Y} - \bar{X}'b)}_{=a} - \bar{X}'b = 0$$

This is an important result that we will be re-visiting later.

Question 3: Deviating from the Average

Consider a random experiment characterized by two random variables (X, Y)

- You are told that $E[Y|X] = \alpha + \beta X$ and $\text{Var}[Y|X] = \sigma^2$
- You know the value of σ^2 but not (α, β)

Consider two transformed random variables $W \equiv X - E[X]$ and $Z \equiv Y - E[Y]$

(a) Briefly explain why $E[Y|W] = \alpha + \beta X$ and $\text{Var}[Y|W] = \sigma^2$; i.e. why is conditioning on W the same as conditioning on X?

ANS: Fixing the value of W also fixes the value of X. Additionally, W has the same information as X. Therefore, conditioning on W is the same as conditioning on X (and vice versa)

(b) Solve for $E[Z|W]$ and $\text{Var}[Z|W]$.

ANS:

$$\begin{aligned}
E[Z|W] &= E[Z|X] = E[\ Y - E[Y]\ |X] = E[Y|X] - E[Y] \\
&= (\alpha + \beta X) - \underbrace{(\alpha + \beta E[X])}_{E[Y]\ =\ E_X[\ E[Y|X]\ ]} = \beta(X - E[X]) = \beta W \\
\text{Var}[Z|W] &= \text{Var}[Z|X] = \text{Var}[\ Y - \underbrace{E[Y]}_{\text{constant}}\ |X] = \text{Var}[Y|X] = \sigma^2
\end{aligned}$$

Now, suppose you are given a size N *random* sample from (X, Y)

$$\{\ (X_1 = x_1, Y_1 = y_1) \cdots (X_N = x_N, Y_N = y_N)\ \}$$

Considered a transformed version of this random sample:

$$\{\ (\tilde{X}_1 = \tilde{x}_1, \tilde{Y}_1 = \tilde{y}_1) \cdots (\tilde{X}_N = \tilde{x}_N, \tilde{Y}_N = \tilde{y}_N)\ \}$$

where $\tilde{X}_i = X_i - \bar{X}_N$ and $\tilde{Y}_i = Y_i - \bar{Y}_N$. In other words, variables in each sample observation are subtracted by their sample mean.

(c) Briefly explain why the transformed sample is random if the untransformed sample is random.

ANS: If each observation in the untransformed sample is independently distributed, then each observation in the transformed sample must also be independently distributed as subtracting the same constant for each observation cannot generate dependence. If each observation in the untransformed sample is distributed identically, then each observation in the transformed sample must also be distributed identically as subtracting the same constant for each observation cannot lead to a different distribution (see change of variables). Note: the constant subtracted can differ **across** samples but remains the same **within** the sample. And it is the latter that matters in terms of whether observations within a sample are **independently** and **identically** distributed.

Let (a^*, b^*) be the values of (a, b) that minimizes $\sum_{i=1}^{N}(Y_i - a - bX_i)^2$ and $(\tilde{a}, \tilde{b})$ be the values of (a, b) that minimizes $\sum_{i=1}^{N}(\tilde{Y}_i - a - b\tilde{X}_i)^2$

(d) Solve for $(\tilde{a}, \tilde{b})$ in terms of (a^*, b^*).

ANS: Note that in the first you are essentially applying OLS to the original sample and in the second to the transformed sample. Additionally, it is fairly easy to show that the sample of $\tilde{X}$ is zero and the sample mean of $\tilde{Y}$ is also similarly zero. So ...

$$\begin{aligned}
\tilde{a} &= \underbrace{\frac{1}{N}\sum_{i=1}^{N}\tilde{Y}_i}_{=0} - \tilde{b}\underbrace{\frac{1}{N}\sum_{i=1}^{N}\tilde{X}_i}_{=0} = 0 \\
\tilde{b} &= \frac{\frac{1}{N}\sum_{i=1}^{N}(\tilde{X}_i\tilde{Y}_i) - \overbrace{\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{X}_i\right)}^{=0}\overbrace{\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{Y}_i\right)}^{=0}}{\frac{1}{N}\sum_{i=1}^{N}(\tilde{X}_i^2) - \underbrace{\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{X}_i\right)^2}_{=0}} \\
&= \frac{\frac{1}{N}\sum_{i=1}^{N}(\tilde{X}_i\tilde{Y}_i)}{\frac{1}{N}\sum_{i=1}^{N}(\tilde{X}_i^2)} = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X}_N)(Y_i - \bar{Y}_N)}{\frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X}_N)^2} = b^*
\end{aligned}$$

(e) Use the analogy (moment) principle to provide intuition for the result obtained in (**d**).

ANS: Note that the transformed sample is essentially a random sample of (W, Z). Instead of subtracting the (population) mean, we are subtracting out the sample analog (sample mean). As such $(\tilde{a}, \tilde{b})$ are from a method of moment estimation of $BLP_{MSE}(Z|W)$. Given that $BP_{MSE}(Z|W)$ is linear in W, $(\tilde{a}, \tilde{b})$ are also from a method of moment estimation of $BP_{MSE}(Z|W)$. From **(a)**, $BP_{MSE}(Z|W) = E[Z|W] = \beta W$. β here is the same β as in $BP_{MSE}(Y|X) = E[Y|X] = \alpha + \beta X$. Therefore, we should not be surprised to find that $\tilde{a} = 0$ and $\tilde{b} = b^*$.