(a) $W_i = \alpha + \beta D_i + u_i \quad i = 1, 2, \ldots, N$

$D_i = \begin{cases} 1 & \text{if } i \text{ is a non-native worker} \\ 0 & \text{if } i \text{ is a native worker.} \end{cases}$

$$\hat{\beta}_{ols} = \frac{\sum_{i=1}^{N}(D_i - \bar{D}) W_i}{\sum_{i=1}^{N}(D_i - \bar{D})^2} \quad \begin{matrix} \rightarrow N^r. \\ \rightarrow D^r. \end{matrix}$$

$\underline{N^r.}$: $\sum_{i=1}^{N} D_i W_i - \bar{D} \sum_{i=1}^{N} W_i = \bar{W}_{non} n_{non} - \bar{D}\bar{W}N$

$(\because \bar{D} = n_{NON}/N) = (\bar{W}_{non} - \bar{W})\, n_{non}$

$\left(\because \bar{W} = \frac{n_{NON}\bar{W}_{NON} - n_{NAT}\bar{W}_{NAT}}{N}\right) \quad = (N\bar{W}_{non} - \bar{W}_{non} n_{non} - \bar{W}_{nat} n_{nat}) \frac{n_{non}}{N}$

$= (\bar{W}_{non} - \bar{W}_{nat}) \frac{n_{non} \cdot n_{nat}}{N}$

$\underline{D^r.}$: $\sum_{i=1}^{N} (D_i - \bar{D}) D_i = \sum_{i=1}^{N} D_i^2 - \bar{D}\sum_{i=1}^{N} D_i$

$= n_{non} - \frac{n_{non}}{N} \cdot n_{non} = \frac{n_{nat} \cdot n_{non}}{N}.$

Hence, $\boxed{\hat{\beta}_{ols} = \bar{W}_{non} - \bar{W}_{nat}}$

and $\hat{\alpha}_{ols} = \bar{W} - \hat{\beta}_{ols} \cdot \bar{D} = \frac{\bar{W}_{nat} n_{nat} + \bar{W}_{non} n_{non}}{N} - (\bar{W}_{non} - \bar{W}_{nat}) \frac{\times n_{non}}{N}$

$\boxed{\hat{\alpha}_{as} = \bar{W}_{nat}}$

Therefore, $\hat{\beta}_{ols}$ provides the difference between average wages of non-native immigrant workers and native resident workers of New Delhi. That is, $\hat{\beta}_{ols}$ provides avg. wages of non-natives relative to the native workers.

$\hat{\alpha}_{ols}$ provides the average wages of native resident workers.

(b) $\hat{\beta}_{ols} > 0$ suggests that on-average non-native immigrants earn a higher wage in our sample relative to the native residents of New Delhi

$\hat{\beta}_{ols} > \hat{\alpha}_{ols}$ provides the inference that wages of non-native workers are on-average greater than twice the wages of native workers in our sample;

Alternatively, $\hat{\beta}_{ols} > \hat{\alpha}_{ols}$ suggests that the wage of an average non-native worker is greater than twice the wage of an average native worker in our sample.

Now, $\hat{\beta}_{OLS} > 0 \Rightarrow \overline{W}_{non} > \overline{W}_{nat}$

and $\hat{\beta}_{OLS} < \hat{\alpha}_{OLS} \Rightarrow \overline{W}_{non} < 2\overline{W}_{nat}$

$\Rightarrow \overline{W}_{non} \in (\overline{W}_{nat}, 2\overline{W}_{nat})$

∴ The above condition provides an inference that the wage of an average non-native worker is higher than the wage of an average native worker in our sample, [however the former's wage is less than twice the wage of averag native worker in our sample].

———x———.

E) SLRM assumption: $E(u_i | D_i) = 0 \ \forall \ i = 1, 2, \ldots, N$ which implies $E(u_i D_i) = 0 \ \forall i$ and $Corr(u_i, D_i) = 0$ for causal inference

- Equivalent statistical property of $\hat{\beta}_{OLS}$: $E(\hat{\beta}_{OLS} | D_i) = \beta$ (i.e. Unbiasedness of a estimator)

- $\hat{\beta}_{OLS}$ is unlikely to provide a causal inference because education levels are not included as a regressor and it may be a case that natives and non-natives are not equally educated in term of professional degree etc.

(d) · If $Corr(D_i, e_i) = 0$ then ~~is~~ there is a ~~greater~~ likelihood that $E(u_i | D_i) = 0$ in eq. (1) and thus it is possible that $\hat{\beta}_{OLS} = \hat{\gamma}_{1,OLS}$.

· If $Corr(D_i, e_i) > 0$ then $\hat{\beta}_{OLS} \neq \hat{\gamma}_{1,OLS}$ is definitely true. Infact, since worker identity and education levels positively impact wages ($\because \hat{\beta}_{OLS} > 0, \hat{\gamma}_{1,OLS} > 0$ and $\hat{\gamma}_{2} > 0$), and non-natives are more educated than natives ($\because Corr(D_i, e_i) > 0$) it is likely that D_i was picking up the impact of education levels as well in eq. (1). These compounded impacts of worker identity and education will be isolated in eq. (2). Hence, we expect

$$\hat{\beta}_{OLS} > \hat{\gamma}_{1,OLS}$$ ~~is~~.

Let $n = 100$; $n_{non} = 50$; $n_{nat} = 50$

(e) $w_i = \gamma_0 + \gamma_1 D_i + \gamma_2 e_i + u_i$

$$\hat{\gamma}_{OLS} = \begin{bmatrix} \hat{\gamma}_{0,OLS} \\ \hat{\gamma}_{1,OLS} \\ \hat{\gamma}_{2,OLS} \end{bmatrix} = (X'X)^{-1} X' \underline{W}$$

where $\underset{100\times3}{X} = \begin{bmatrix} 1 & D_1 & e_1 \\ 1 & D_2 & e_2 \\ 1 & D_3 & e_3 \\ \vdots & \vdots & \vdots \\ 1 & D_{100} & e_{100} \end{bmatrix}_{100\times3}$; $\underset{100\times1}{\underline{W}} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_{100} \end{bmatrix}$

$$\underset{3\times3}{(X'X)} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ D_1 & D_2 & D_3 & \cdots & D_{100} \\ e_1 & e_2 & e_3 & \cdots & e_{100} \end{bmatrix} \begin{bmatrix} 1 & D_1 & e_1 \\ 1 & D_2 & e_2 \\ 1 & D_3 & e_3 \\ \vdots & \vdots & \vdots \\ 1 & D_{100} & e_{100} \end{bmatrix}$$

$\because \left(\begin{array}{l} \sum_i D_i = n_{non} \\ \sum_i D_i^2 = n_{non} \end{array} \right)$

$$= \begin{bmatrix} n & n_{non} & \sum_{i=1}^{N} e_i \\ & n_{non} & \sum_{i=1}^{N} D_i e_i \\ & & \sum_{i=1}^{N} e_i^2 \end{bmatrix} = \begin{bmatrix} 100 & 50 & \sum_{i=1}^{N} e_i = \bar{e} \cdot 100 \\ & 50 & \bar{e}_{non} \cdot 50 \\ & & \sum_{i=1}^{N} e_i^2 \end{bmatrix}_{3\times3}$$

$$\underset{3\times100}{X'} \underset{100\times1}{\underline{W}} = \begin{bmatrix} \sum_{i=1}^{N} w_i \\ \sum_{i=1}^{N} D_i w_i \\ \sum_{i=1}^{N} e_i w_i \end{bmatrix}_{3\times1} = \begin{bmatrix} \bar{W} \cdot 100 \\ \bar{W}_{non} \cdot n_{non}^{=50} \\ \sum_{i=1}^{N} e_i w_i \end{bmatrix}$$

If all native workers are college graduates and non-native workers are not. Then

$$e_i = \begin{cases} 1 \text{ if } i \text{ is a native worker} \\ 0 \text{ if } i \text{ is a non native worker.} \end{cases}$$

we know that $D_i = \begin{cases} 1 \text{ if } i \text{ is a non-native worker} \\ 0 \text{ if } i \text{ is a native worker.} \end{cases}$

Then $(X'X) = \begin{bmatrix} 100 & 50 & 50 \\ 50 & 50 & 0 \\ 50 & 0 & 50 \end{bmatrix}$ is a singular matrix.

$\left[\because \sum_i D_i e_i = 0 \quad \sum_i e_i = n_{nat} = \sum_i e_i^2\right]$

because Col. I − Col. II = Col. III in this matrix.

Hence, $\hat{\gamma}_{OLS}$ is unidentified when $e_i = 1$ for all native workers and 0 therwise

Alternatively, one can see that under given e_i and D_i definitions

$$Corr(D_i, e_i) = -1., \text{ i.e.,}$$

D_i and e_i are perfectly negatively correlated. Hence, X is not full rank

$\Rightarrow (X'X)$ is not invertible! $\Rightarrow \hat{\gamma}_{OLS}$ is unidentified (f). $\tilde{e}_i = \begin{cases} 3 & \text{if is a college graduate} \\ 2 & \text{if is not a college graduate} \end{cases}$

i.e,

$$w_i = \gamma_0 + \gamma_1 D_i + \gamma_2 \tilde{e}_i + u_i$$

$$w_i = \gamma_0 + \gamma_1 D_i + \gamma_2 (e_i + 2) + u_i$$

$$w_i = (\gamma_0 + 2\gamma_2) + \gamma_1 D_i + \gamma_2 e_i + u_i$$

$$w_i = \tilde{\gamma}_0 + \gamma_1 D_i + \gamma_2 e_i + u_i$$

Since redefining e_i only leads to an unidentifiable change in the intercept vector $\hat{\gamma}_{OLS}$ will remain exactly the same between parts (c) and (f). Afterall, $\Delta w_i = \gamma_2$ for both model specifications. when $\Delta e_i = 1$

(g) 25 out of 50 non-native workers are college graduates

None out of 50 native workers are college graduates

We know that $Var(\hat{\gamma}_{OLS} | X) = \sigma^2 (X'X)^{-1}$

where σ^2 is variance of u_i

$(X'X)$ vector in part (e) becomes

$$(X'X) = \begin{bmatrix} 100 & 50 & 25 \\ 50 & 50 & 25 \\ 25 & 25 & 25 \end{bmatrix} = 25 \begin{bmatrix} 4 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} = 25Z$$

more $Z = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

Hence, $(X'X)^{-1} = 25^{-1} \cdot Z^{-1}$

$$|Z| = 4(2-1) - 2(2-1) + 1(2-2)$$
$$= 4 - 2 = 2.$$

$$Z^{-1} = \frac{1}{2} \begin{bmatrix} (2-1) & (1-2) & (2-2) \\ (1-2) & (-1+4) & (2-4) \\ (2-2) & (-4+2) & (4+8) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 \\ -1 & +3 & -2 \\ 0 & -2 & +4 \end{bmatrix}$$

$$V(\hat{\gamma}_{os} | X) = \frac{\sigma^2}{50} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 3 & -2 \\ 0 & -2 & 4 \end{bmatrix}$$

ALTERNATIVE SOLUTION PART (f) ①

Consider : $y = \beta_0 + \beta_1 x + u$.

Here; $\boxed{\hat{\beta}_1^{OLD} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) y_i}{\sum_{i=1}^{N} (x_i - \bar{x})^2}}$

$\boxed{\hat{\beta}_0^{OLD} = \bar{y} - \hat{\beta}_1 \bar{x}}$

Now; suppose x is replaced by an alternate regressor; say $\boxed{\tilde{x} = x + A}$; then the model becomes:

$$\boxed{y = \beta_0 + \beta_1 \tilde{x} + u.}$$

where $\tilde{x}_i = x_i + A \quad \forall i$. &

$$\boxed{\bar{\tilde{x}} = \bar{x} + A.}$$

Here; $\hat{\beta}_1^{NEW} = \frac{\sum_{i=1}^{N} (\tilde{x}_i - \bar{\tilde{x}}) y_i}{\sum_{i=1}^{N} (\tilde{x}_i - \bar{\tilde{x}})^2}$

$$= \frac{\sum_{i=1}^{N} (x_i + \cancel{A} - \bar{x} - \cancel{A}) y_i}{\sum_{i=1}^{N} (x_i + \cancel{A} - \bar{x} - \cancel{A})^2}$$

$$\boxed{\hat{\beta}_1^{NEW} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) y_i}{\sum_{i=1}^{N} (x_i - \bar{x})^2} = \hat{\beta}_1^{OLD}}$$

$$\boxed{\hat{\beta}_0^{NEW} = \bar{y} - \hat{\beta}_1^{NEW} (\bar{\tilde{x}}) = \bar{y} - \hat{\beta}_1^{OLD} (\bar{x} + A) = \hat{\beta}_0^{OLD} - \hat{\beta}_1^{OLD} \cdot A}$$