1

a Let X_3 be a indicator of whether one is in manufacturing sector or not. The linear regression model could be as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1 \cdot X_2 + \beta_3 X_1 \cdot X_3 + \epsilon$$

b If Y is log-wage, then the coefficients change Δ is interpreted as 100Δ percentage change.

$$\begin{aligned}
\log(Y') - \log(Y) &= \Delta \\
\log\left(\frac{Y'}{Y}\right) &= \Delta \\
\log\left(\frac{Y'}{Y} - 1 + 1\right) &= \Delta
\end{aligned}$$

Because $\log\left(\frac{Y'}{Y} - 1 + 1\right) \approx \frac{Y'}{Y} - 1$, this implies

$$\begin{aligned}
\frac{Y'}{Y} - 1 &= \Delta \\
Y' &= 100\,(1+\Delta)\,\%Y
\end{aligned}$$

2

a

$$Y = \beta_1 X_1 + \phi\,(X_2, X_3) + \epsilon$$

b

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

The restriction imposed in the less general model is that $\phi(\cdot)$ is linear in X_2 and X_3.

3

$$\begin{aligned}
Y =& m\,(0,0)\,(1 - \mathbf{1}\{x_1 = x_{10}\})\,(1 - \mathbf{1}\{x_2 = x_{20}\}) \\
&+ m\,(1,0)\,\mathbf{1}\{x_1 = x_{10}\}\,(1 - \mathbf{1}\{x_2 = x_{20}\}) \\
&+ m\,(0,1)\,(1 - \mathbf{1}\{x_1 = x_{10}\})\,\mathbf{1}\{x_2 = x_{20}\} \\
&+ m\,(1,1)\,\mathbf{1}\{x_1 = x_{10}\}\,\mathbf{1}\{x_2 = x_{20}\}\,.
\end{aligned}$$

4

Consider the original model as

$$Y = X\beta + \epsilon.$$

The original OLS estimate $\hat{\beta}$ is

$$\hat{\beta} = (X'X)^{-1}\,X'Y.$$

Now if we multiply the dependent variables by c, $\tilde{Y} = cY$, the new OLS estimate is

$$\begin{aligned}\tilde{\hat{\beta}} &= \left(X'X\right)^{-1} X'\tilde{Y} \\ &= \left(X'X\right)^{-1} X'cY \\ &= c\left(X'X\right)^{-1} X'Y \\ &= c\hat{\beta}.\end{aligned}$$

5

Assume there are K regressors. The auxiliary regression is

$$X_j = \gamma_1 X_1 + \cdots \gamma_{j-1} X_{j-1} + \gamma_{j+1} X_{j+1} + \cdots + \nu.$$

Denote the estimates as $\hat{\gamma}_i$, the residual is

$$\hat{\nu} = X_j - \sum_{i=1, i\neq j}^{K} \hat{\gamma}_i X_i,$$

and jth OLS coefficient by regressing Y is

$$\hat{\beta}_j = \frac{\hat{\nu}'Y}{\hat{\nu}'\hat{\nu}}.$$

If now we multiply X_j by c, as the conclusion of Q4 suggests, the estimates in auxiliary regression will be multiplied by c:

$$\tilde{\hat{\gamma}} = c\hat{\gamma},$$

and

$$\begin{aligned}\tilde{\hat{\nu}} &= \tilde{X}_j - \sum_{i=1, i\neq j}^{K} \tilde{\hat{\gamma}}_i X_i \\ &= cX_j - \sum_{i=1, i\neq j}^{K} c\hat{\gamma}_i X_i \\ &= c\left(X_j - \sum_{i=1, i\neq j}^{K} \hat{\gamma}_i X_i\right) \\ &= c\hat{\nu}.\end{aligned}$$

Then,

$$\begin{aligned}\tilde{\hat{\beta}}_j &= \frac{\tilde{\hat{\nu}}'Y}{\tilde{\hat{\nu}}'\tilde{\hat{\nu}}} \\ &= \frac{c\hat{\nu}'Y}{c\hat{\nu}'c\hat{\nu}} \\ &= \frac{1}{c}\hat{\beta}_j.\end{aligned}$$

(Without using auxiliary regression:)

$$\tilde{X} = X \underbrace{\begin{bmatrix} 1 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & c & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \end{bmatrix}}_{I_j}.$$

I_j is a $K \times K$ matrix, with diagonal being 1s except the jth diagonal entry equals to c, and all the other entries being 0. Obviously, $I_j = I_j'$. The entries of I_j^{-1} are all the same with I_j except the jth diagonal entry equals to $\frac{1}{c}$.

$$\begin{aligned} \tilde{\hat{\beta}} &= \left(\tilde{X}'\tilde{X}\right)^{-1} \tilde{X}'Y \\ &= \left(I_j'X'XI_j\right)^{-1} I_j'X'Y \\ &= I_j^{-1} \left(X'X\right)^{-1} I_j'^{-1} I_j'X'Y \\ &= I_j^{-1} \left(X'X\right)^{-1} X'Y \\ &= I_j^{-1}\hat{\beta} \\ &= \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \frac{1}{c}\hat{\beta}_j \\ \vdots \\ \hat{\beta}_K \end{bmatrix}. \end{aligned}$$

6

Assume the original model is

$$Y = X\beta + \epsilon,$$

where the first column vector of X is $\mathbf{1}$. The OLS estimator is

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y.$$

If we add a constant c to Y, then

$$\begin{aligned} \tilde{\hat{\beta}} &= \left(X'X\right)^{-1} X' \left(Y + \mathbf{1}c\right) \\ &= \left(X'X\right)^{-1} X'Y + \left(X'X\right)^{-1} X'\mathbf{1}c \\ &= \hat{\beta} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} c. \end{aligned}$$

Notice

$$\left(X'X\right)^{-1}X'\mathbf{1}=\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix},$$

because $\mathbf{1}$ is the first column vector of X.

If we now add a constant c to X_j, then

$$\tilde{X}=X\underbrace{\begin{bmatrix}1&\cdots&c&\cdots&0\\\vdots&\vdots&\vdots&\vdots&\vdots\\\vdots&\vdots&1&\vdots&\vdots\\\vdots&\vdots&\vdots&\vdots&\vdots\\0&\cdots&\cdots&\cdots&1\end{bmatrix}}_{C_j}.$$

Note that C_j has full rank and is invertible.

$$C_j^{-1}=\begin{bmatrix}1&\cdots&-c&\cdots&0\\\vdots&\vdots&\vdots&\vdots&\vdots\\\vdots&\vdots&1&\vdots&\vdots\\\vdots&\vdots&\vdots&\vdots&\vdots\\0&\cdots&\cdots&\cdots&1\end{bmatrix}.$$

The OLS estimator using $\tilde{X}$ is

$$\begin{aligned}\tilde{\hat{\beta}}&=\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'Y\\&=\left(C_j'X'XC_j\right)^{-1}C_j'X'Y\\&=C_j^{-1}\left(X'X\right)^{-1}C_j'^{-1}C_j'X'Y\\&=C_j^{-1}\left(X'X\right)^{-1}X'Y\\&=C_j^{-1}\left(X'X\right)^{-1}\hat{\beta}\\&=\begin{bmatrix}\hat{\beta}_1-c\hat{\beta}_j\\\vdots\\\hat{\beta}_j\\\vdots\\\hat{\beta}_K\end{bmatrix}.\end{aligned}$$

The intuition behind it is that adding c to X_j creates a vector $\tilde{X}$ that is a linear combination of constant variable and the original X_j that are linearly independent. Remember that OLS regression is projecting Y onto the space spanned by X_i column vectors. $\tilde{X}$ takes away a part of explanation of constant vector which is exactly $c\hat{\beta}_j$.

7

$$\log(cy) = \log(c) + \log(y).$$

The coefficient of the constant term x_1 will be shifted by $\log(c)$. Because there is a constant term in the regressors, adding a constant $log(c)$ to the dependent varaible will shift the coeffiicient of constant regressor by $log(c)$ exactly, as we have shown in (6).

8

$$\log(cx_1) = \log(c) + \log(x_1).$$

As we have shown in (6), the coefficient of constant element will be shifted by $-\log(c)\hat{\beta}_1$ while all the other coefficients stay the same.

9

Notice that

$$rank(A) = rank(A'A).$$

"⇒" If A is full rank, then $A'A$ is full rank. Because $A'A$ is square matrix, it implies that it is invertible.

"⇐" If $A'A$ is invertible, then it has full rank. Consequently, A has full rank.

10

(a) The objective function is

$$E\left[(Y - b_0 - X'b)^2\right].$$

The first order derivate with respect to b_0 and b is

$$\begin{aligned} \text{with respect to } b_0 &: E\left[-2\left(Y - b_0 - X'b\right)\right] = 0 \\ \text{with respect to } b &: E\left[2X\left(Y - b_0 - X'b\right)\right] = \mathbf{0} \end{aligned}$$

The first FOC gives

$$\beta_0 = E[Y] - E[X]'\beta.$$

The second FOC gives

$$\begin{aligned} & E[XY] - E[X]\beta_0 - E[XX']\beta = 0 \\ \Leftrightarrow & E[XY] - E[X]E[Y] - E[X]E[X]'\beta - E[XX']\beta = 0 \\ \Leftrightarrow & E[XY] - E[X]E[Y] = \left(E[XX'] - E[X]E[X]'\right)\beta \\ & \beta = Var(X)^{-1} Cov(X,Y). \end{aligned}$$

Notice that $Var(X) = E[XX'] - E[X]E[X]'$ and $Cov(X,Y) = E[XY] - E[X]E[Y]$.

(b)

$$\begin{aligned} E(U) &= E(Y - \beta_0 - X'\beta) \\ &= E\big(Y - E[Y] + E[X]'\beta - X'\beta\big) \\ &= E[Y] - E[Y] + E[X]'\beta - E[X]'\beta \\ &= 0. \end{aligned}$$

$$\begin{aligned} Cov(U, X) &= Cov(Y - \beta_0 - X'\beta, X) \\ &= Cov(Y, X) - Cov(X'\beta, X) \\ &= Cov(Y, X) - Cov(\beta_1 X_1 + \beta_2 X_2 + \cdots \beta_K X_K, X) \\ &= Cov(Y, X) - \beta_1 Cov(X_1, X) - \beta_2 Cov(X_2, X) - \cdots \beta_K Cov(X_K, X) \\ &= Cov(Y, X) - [Cov(X_1, X), Cov(X_2, X), \cdots, Cov(X_K, X)]\beta \\ &= Cov(Y, X) - Var(X)\beta \\ &= Cov(Y, X) - Var(X) Var(X)^{-1} Cov(X, Y) \\ &= 0. \end{aligned}$$

11

(a) Recall the figure illustrated in the lecture where x_2 and x_1 lie on the same line.

(b)

$$\begin{aligned} X'X &= \begin{bmatrix} x_1' \\ a x_1' \end{bmatrix} \begin{bmatrix} x_1 & a x_1 \end{bmatrix} \\ &= \begin{bmatrix} x_1' x_1 & a x_1' x_1 \\ a x_1' x_1 & a^2 x_1' x_1 \end{bmatrix} \\ &= \begin{bmatrix} b_1' \\ a b_1' \end{bmatrix}. \end{aligned}$$

where $b_1' = \begin{bmatrix} x_1' x_1 & a x_1' x_1 \end{bmatrix}$. There is only one linearly independent vector, and therefore the rank is 1.

12

The two coefficients are the same. (Recall the regression of using Y residuals and X auxiliary residuals in the lecture.)

13

(a)

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1} X'Y \\
&= (X'X)^{-1} X' \left(X\hat{\beta}_1 + \hat{\alpha} Z + \hat{\epsilon} \right) \\
&= (X'X)^{-1} X'X\hat{\beta}_1 + (X'X)^{-1} X'\hat{\alpha} Z + (X'X)^{-1} X'\hat{\epsilon} \\
&= \hat{\beta}_1 + \hat{\alpha} (X'X)^{-1} X'Z + (X'X)^{-1} X'\hat{\epsilon} \\
&= \hat{\beta}_1 + \hat{\alpha}\hat{\pi}.
\end{aligned}
$$

Note that $(X'X)^{-1} X'Z = \hat{\pi}$ as shown in the auxiliary regression, and $\hat{\epsilon}$ is orthogonal with X.

(b)

$$
\begin{aligned}
\hat{y}_i &= x_i'\hat{\beta}_1 + \hat{\alpha} z_i \\
&= x_i'\hat{\beta}_1 + \hat{\alpha} \left(x_i'\hat{\pi} + \hat{\pi}_0 + \hat{\nu}_{zi} \right) \\
&= x_i'\hat{\beta}_1 + \hat{\alpha} x_i'\hat{\pi} + \hat{\alpha}\hat{\pi}_0 + \hat{\alpha}\hat{\nu}_{zi} \\
&= x_i' \left(\hat{\beta}_1 + \hat{\alpha}\hat{\pi} \right) + \hat{\alpha}\hat{\nu}_{zi} + \hat{\alpha}\hat{\pi}_0 \\
&= x_i'\hat{\beta} + \hat{\alpha}\hat{\nu}_{zi} + \hat{\alpha}\hat{\pi}_0.
\end{aligned}
$$

(c)

$$
\begin{aligned}
\bar{\hat{y}} &= \frac{1}{N} \sum_{i=1}^{N} \hat{y}_i \\
&= \frac{1}{N} \sum_{i=1}^{N} \left(x_i'\hat{\beta} + \hat{\alpha}\hat{\nu}_{zi} + \hat{\alpha}\hat{\pi}_0 \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} x_i'\hat{\beta} + \frac{1}{N} \sum_{i=1}^{N} \hat{\alpha}\hat{\nu}_{zi} + \frac{1}{N} \sum_{i=1}^{N} \hat{\alpha}\hat{\pi}_0 \\
&= \frac{1}{N} \sum_{i=1}^{N} \hat{\tilde{y}}_i + \hat{\alpha}\hat{\pi}_0 \\
&= \bar{\hat{\tilde{y}}} + \hat{\alpha}\hat{\pi}_0.
\end{aligned}
$$

Notice that $\sum_{i=1}^{N} \hat{\nu}_{zi} = 0$ because $\hat{\nu}_{zi}$ is orthogonal with constant vector, $\mathbf{1}\hat{\nu}_{zi} = 0$.

$$
\begin{aligned}
\hat{y}_i - \bar{\hat{y}} &= x_i'\hat{\beta} + \hat{\alpha}\hat{\nu}_{zi} + \hat{\alpha}\hat{\pi}_0 - \bar{\hat{\tilde{y}}} - \hat{\alpha}\hat{\pi}_0 \\
&= \hat{\tilde{y}}_i - \bar{\hat{\tilde{y}}} + \hat{\alpha}\hat{\nu}_{zi}.
\end{aligned}
$$

(d) R^2_{large} of the larger model is

$$\begin{aligned}\frac{\sum_i^N \left(\hat{y}_i - \bar{\hat{y}}_i\right)^2}{\sum_i^N \left(y_i - \bar{y}_i\right)} &= \frac{\sum_i^N \left(\hat{\tilde{y}}_i - \bar{\hat{\tilde{y}}} + \hat{\alpha}\hat{\nu}_{zi}\right)^2}{\sum_i^N \left(y_i - \bar{y}_i\right)} \\ &= \frac{\sum_i^N \left[\left(\hat{\tilde{y}}_i - \bar{\hat{\tilde{y}}}\right)^2 + \left(\hat{\alpha}\hat{\nu}_{zi}\right)^2 + 2\left(\hat{\tilde{y}}_i - \bar{\hat{\tilde{y}}}\right)\left(\hat{\alpha}\hat{\nu}_{zi}\right)\right]}{\sum_i^N \left(y_i - \bar{y}_i\right)} \\ &= \frac{\sum_i^N \left[\left(\hat{\tilde{y}}_i - \bar{\hat{\tilde{y}}}\right)^2 + \left(\hat{\alpha}\hat{\nu}_{zi}\right)^2\right]}{\sum_i^N \left(y_i - \bar{y}_i\right)}.\end{aligned}$$

Notice that $\sum_{i=1}^N \hat{\tilde{y}}_i \hat{\nu}_{zi} = \sum_{i=1}^N x_i'\hat{\beta}\hat{\nu}_{zi} = \mathbf{1}\hat{\nu}_z' X\hat{\beta} = 0$ where $\hat{\nu}_z$ is the $N \times 1$ residual vector that is orthogonal with X. $\sum_{i=1}^N \bar{\hat{\tilde{y}}}\hat{\nu}_{zi} = \bar{\hat{\tilde{y}}}\sum_{i=1}^N \hat{\nu}_{zi} = 0$ as $\hat{\nu}_z$ is orthogonal with constant vector.

R^2_{small} of the smaller model is

$$\frac{\sum_i^N \left(\tilde{\hat{y}}_i - \bar{\tilde{\hat{y}}}_i\right)^2}{\sum_i^N \left(y_i - \bar{y}_i\right)} = \frac{\sum_i^N \left(\hat{\tilde{y}}_i - \bar{\hat{\tilde{y}}}_i\right)^2}{\sum_i^N \left(y_i - \bar{y}_i\right)}.$$

Because the denominators are the same and R^2_{large} has strictly larger numerator if $\hat{\alpha} \neq 0$ (additional regressor is linearly indepedent), R^2_{large} is strictly larger.

(e) Note that the total variance in y_i can be decomposed into two parts when there is a constant term for any OLS esimator $\hat{y}_i$:

$$\begin{aligned}\sum_{i=1}^N \left(y_i - \bar{y}\right)^2 &= \sum_{i=1}^N \left(y_i - \hat{y}_i + \hat{y}_i - \bar{\hat{y}} + \bar{\hat{y}} - \bar{y}\right)^2 \\ &= \sum_{i=1}^N \left(y_i - \hat{y}_i + \hat{y}_i - \bar{\hat{y}}\right)^2 \\ &= \sum_{i=1}^N \underbrace{\left(y_i - \hat{y}_i\right)^2}_{(*)} + \underbrace{\left(\hat{y}_i - \bar{\hat{y}}\right)^2}_{(**)}.\end{aligned}$$

This is because

$$\begin{aligned}&(1)\ \bar{\hat{y}} = \bar{y}, \\ &(2)\ \sum_{i=1}^N \left(y_i - \hat{y}_i\right)\left(\hat{y}_i - \bar{\hat{y}}_i\right) = \sum_{i=1}^N \left(y_i - \hat{y}_i\right) x_i'\beta = 0,\end{aligned}$$

Remark: these may not hold if there is no constant term in the regression model!
Note that the (*) part is the objective function that we are trying to minimize, and the

(**) is the numerator of R^2. The minimization problem for larger model is

$$\min_{\beta,\alpha} \sum_{i=1}^{N} \left(y_i - x_i'\beta - \alpha z_i\right)^2 .$$

The minimization problem for smaller model is as adding a constraint $\alpha = 0$ to the original problem. As we know, unconstrained optimization problem gives a better solution. Therefore, the (**) part is in general weakly larger for the larger model, which gives a weakly larger R^2.