Regression [in the population]

$$\left[\begin{array}{l} \text{given linear model } Y_i = \alpha + \beta X_i + \epsilon_i \\ \alpha, \beta \text{ s.t. } E(\epsilon_i) = 0 \quad E(\epsilon_i X_i) = 0 \end{array}\right]$$

① $\alpha, \beta = \underset{a,b}{\text{argmin}}\, E\{(Y_i - [a + bX_i])^2\}$

① $\alpha + \beta X_i$ is the best linear predictor of Y_i

$$\frac{\partial MSE}{\partial a} = -2E[\{Y_i - a - bX_i\}] = 0$$

$$\Leftrightarrow E(\epsilon_i) = 0$$

$$\frac{\partial MSE}{\partial b} = -2E[\{Y_i - a - bX_i\}X_i] = 0$$

$$\Leftrightarrow E(X_i \epsilon_i) = 0$$

② if CEF is linear, then regression is the CEF

(ii) if the CEF is linear, then $\alpha + \beta X_i$ is the CEF

$$E[Y_i | X_i] = \alpha + \beta X_i$$

$$\varepsilon_i = Y_i - E[Y_i | X_i]$$

$$E[\varepsilon_i | X_i] = 0$$

$$E[\varepsilon_i] = E[E[\varepsilon_i | X_i]] = 0$$

$$E[X_i \varepsilon_i] = E[X_i E[\varepsilon_i | X_i]] = 0$$

③ it's the best linear approx to CEF

$$\rightarrow (\alpha, \beta) = \underset{a,b}{\text{argmin}}\, E\{[E[Y_i | X_i] - (a + bX_i)]^2\}$$

(iii) even if the CEF is not linear, $\alpha + \beta X_i$ is the best linear approx. to it

OLS [In the data]

$$(\hat{\alpha}_{OLS}, \hat{\beta}_{OLS}) = \underset{a,b}{\arg\min} \sum_i (y_i - a - b x_{1i})^2$$

$$\hat{\beta}_{OLS} = \frac{S_{xy}}{S_x^2}$$

$$E[\hat{\beta}_{OLS}] = \beta \;?$$

is $\hat{\beta}_{OLS}$ an unbiased estimator of β?

$$\hat{\beta}_{OLS} = \frac{S_{xy}}{S_x^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y}) \cancel{\frac{1}{n-1}}}{\sum (x_i - \bar{x})^2 \cancel{\frac{1}{n-1}}}$$

$$\sum (x_i - \bar{x}) y_i - \bar{y} \cancel{\sum (x_i - \bar{x})} \rightarrow 0$$

$$\sum [(x_i - \bar{x}) y_i - (x_i - \bar{x}) \bar{y}]$$

$$\underline{\sum (x_i - \bar{x}) y_i} - \bar{y} \underbrace{\sum (x_i - \bar{x})}_{[\sum x_i - n\bar{x}]}$$

$$= \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x}) x_i}$$

$$= \frac{\sum (x_i - \bar{x}) [\alpha + \beta x_i + \varepsilon_i]}{\sum (x_i - \bar{x}) x_i}$$

$$= \alpha \underbrace{\frac{\sum (x_i - \bar{x})}{\sum (x_i - \bar{x}) x_i}}_{0} + \beta \underbrace{\frac{\sum (x_i - \bar{x}) x_i}{\sum (x_i - \bar{x}) x_i}}_{1} + \frac{\sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x}) x_i}$$

$$\hat{\beta}_{OLS} = \beta + \frac{\sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x}) x_i}$$

$$E[\hat{\beta}_{OLS}] = \beta$$

$$E[\hat{\beta}_{OLS}] = E\left[\beta + \frac{\sum(X_i - \bar{X})\varepsilon_i}{\sum(X_i - \bar{X})X_i}\right]$$

$\boxed{\text{ass: } X_i \text{ is fixed}}$

$$\boxed{\sum X_i\varepsilon_i - \bar{X}\sum\varepsilon_i} \neq 0$$

in a given sample, this does not need to be true!

$$= \beta + E\left[\frac{\sum(X_i - \bar{X})\varepsilon_i}{\sum(X_i - \bar{X})X_i}\right]$$

$$= \beta + \frac{\sum(X_i - \bar{X})}{\sum(X_i - \bar{X})X_i} E[\varepsilon_i] = 0$$

$$E[\hat{\beta}_{OLS}] = \beta + \underbrace{E\left[\frac{\sum(X_i - \bar{X})\varepsilon_i}{\sum(X_i - \bar{X})X_i}\right]}_{=}$$

$$E\left[E\left[\frac{\sum(X_i - \bar{X})\varepsilon_i}{\sum(X_i - \bar{X})X_i} \,\middle|\, X_i\right]\right]$$

$$E\left[\frac{\sum(X_i - \bar{X})E[\varepsilon_i | X_i]}{\sum(X_i - \bar{X})X_i}\right]$$

$$\Leftarrow$$

$$E(\varepsilon_i) = 0 \not\Rightarrow E(\varepsilon_i | X) = 0$$

$$E(\varepsilon_i X_i) = 0 \Leftarrow E(\varepsilon_i | X_i) = 0$$

$$E(\underbrace{Y_i - \alpha - \beta X_i}_{= E[Y_i | X_i]} \mid X_i) = 0$$

$\boxed{\text{ass: CEF is linear}}$

Regression anatomy

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

$$X_{1i} = \underbrace{\delta_0 + \delta_{12} X_{2i}}_{\substack{= \hat{X}_{1i} \\ \text{(fitted values)}}} + \underbrace{\tilde{x}_{1i}}_{\text{residuals}}$$

$$E[\hat{X}_{1i} \tilde{x}_{1i}] = 0$$

since $E[\tilde{x}_{1i}] = 0$
and $E[X_{2i} \tilde{x}_{1i}] = 0$
{regression residual uncorrelated with the regressors that made them}

$$\beta_1 = \frac{cov(Y_i, \tilde{x}_{1i})}{V(\tilde{x}_{1i})}$$

$$cov(y_i, \tilde{x}_{1i}) = E[(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i)\tilde{x}_{1i}]$$

$$= E[\beta_0 \tilde{x}_{1i}] = \beta_0 E[\tilde{x}_{1i}] = 0$$

$$+ \beta_1 E[X_{1i} \tilde{x}_{1i}] \rightarrow \beta_1 \left[E[(\hat{X}_{1i} + \tilde{x}_{1i})(\tilde{x}_{1i})] \right] = \beta_1 V(\tilde{x}^2_{1i})$$

$$\underbrace{E(\hat{X}_{1i} \tilde{x}_{1i})}_{= 0} + \underbrace{E[\tilde{x}^2_{1i}]}_{= V(\tilde{x}^2_{1i})}$$

$$\hat{X}_{1i} = \delta_0 + \delta_1 X_1$$

$$+ \beta_2 \underbrace{E[X_{2i} \tilde{x}_{1i}]}_{} \rightarrow = 0$$

$$+ \underbrace{E[\varepsilon_i \tilde{x}_1]}_{} = 0 \rightarrow E[\varepsilon_i (X_1 - \delta_0 - \delta_1 X_2)]$$

$$V(X) = E[X^2] - E[X]^2$$

$$cov(Y_i, \tilde{x}_{1i}) = \beta_1 var(\tilde{x}_{1i})$$

In the pset: $\frac{cov(\tilde{y}_i, \tilde{x}_{1i})}{V(\tilde{x}_{1i})} = \beta_1$

$$Y_i = \alpha_1 + \alpha_{12} X_{2i} + \tilde{y}_i$$