1] Intro

$y_t = \underbrace{g(w_t)}_{E[y_t | w_t]} + \varepsilon_t$, where $E[\varepsilon_t | w_t] = 0$

- conditional expectation function

Three topics:

1] $g(w_t) \approx \underset{=f(w_t)}{X_t}'\beta$

2] Estimate β

3] Make inference on $X_t'\beta$.

OLS in population

First, we will begin with some basic properties of the conditional expectation function.

- The CEF is the best predictor of y_t given w_t. That is, $g(w) = \underset{\tilde{g} \in \mathbb{R}}{\operatorname{argmin}} E[(y_t - \tilde{g})^2 | w]$

- To see this, take FOCs:

$(\tilde{g}): 2E[(y-g)|w] = 0$ $g(w) = E[y|w]$ satisfies this.

What is $\underset{\tilde{g} \in G}{\operatorname{argmin}} E[(y - \tilde{g}(w))^2]$?

set of measurable functions

- It turns out $g(w) = E[y|w] = \underset{\tilde{g} \in G}{\operatorname{argmin}} E[(y - \tilde{g}(w))^2]$
- Follows by the law of iterated expectations.

since $E[(y-\tilde{g}(w))^2] = E[E[(y-g(w))^2|w]]$
$\geq E[\min_{\hat{g}\in\mathbb{R}} E[(y-\hat{g})^2|w]]$

Basic properties of OLS:

$\beta = \underset{b\in\mathbb{R}^k}{\operatorname{argmin}}\ Q(b)$, where $Q(b) = E[(y_t - x_t'b)^2]$

- in population
- $x_t'\beta$ is the best linear predictor of y_t with respect to mean squared error.

In finite samples,

$\hat{\beta} = \underset{b\in\mathbb{R}^k}{\operatorname{argmin}}\ Q_n(b)$, where $Q_n(b) = \frac{1}{n}\sum_{t=1}^{n}(y_t - x_t'b)^2$
$= E_n[(y_t - x_t'b)^2]$

- remark: E_n is a short-hand for $\frac{1}{n}\sum_{i=1}^{n}$
- in finite samples, $x_t'\hat{\beta}$ is the best linear predictor of y_t.

<u>Explicit solutions:</u>

The FOCs for $\min_{b\in\mathbb{R}^k} Q(b)$ are given by:

(b): $E[\underbrace{(y_t - x_t'\beta)}_{\text{scalar}} x_t] = 0$ chain rule $\Rightarrow E[x_t y_t - x_t x_t'\beta] = 0$

$\Rightarrow E[x_t x_t']\beta = E[x_t y_t]$ assume $E[x_t x_t']$ is nonsingular

$\Rightarrow \underset{k\times 1}{\beta} = \{E[\underset{k\times 1}{x_t}\underset{1\times k}{x_t'}]\}^{-1} E[\underset{k\times 1}{x_t}\underset{1\times 1}{y_t}]$

$\Rightarrow \beta = [Var(X_t)]^{-1} Cov(X_t, y_t)$

FOCs for $\min_{b \in \mathbb{R}^k} Q_n(b)$ are:

(b): $E_n[X_t(y_t - X_t'\hat{\beta})] = 0$ assuming $E_n[X_t X_t']$ is nonsingular

$\Leftrightarrow \hat{\beta} = [E_n[X_t X_t']]^{-1} E_n[X_t y_t]$

$= \left[\frac{1}{n}\sum_{t=1}^{n} X_t X_t'\right]^{-1} \frac{1}{n}\sum_{t=1}^{n} X_t y_t$

$= (X'X)^{-1}X'Y$, where $X = \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix}$, $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

OLS is the best linear approximation to $g(w) = E[y|w]$. (BLA property)

- ie $\beta = \underset{b \in \mathbb{R}^k}{\arg\min}\, E[\underbrace{(g(w_t) - X_t'b)^2}_{\text{specification error}}]$ given $X_t = f(w_t)$
- that is, β minimizes the expected specification error.
- To show this, note that $y_t = E[y_t|w_t] + \varepsilon_t$, $E[\varepsilon_t|w_t] = 0$

$Q(b) = E[(y_t - X_t'b)^2] = E[(E[y_t|w_t] - X_t'\beta + \varepsilon_t)^2]$

$= E[(E[y_t|w_t] - X_t'b)^2] + E[\varepsilon_t^2]$

$+ 2\underbrace{E[(E[y_t|w_t] - X_t'b)\varepsilon_t]}_{=0 \text{ by law of iterated expectations}}$

$= E[(E[y_t|w_t] - X_t'b)^2] + E[\varepsilon_t^2]$

$= E[(g(w_t) - X_t'b)^2] + E[\varepsilon_t^2]$

Thus, $\beta = \underset{b \in \mathbb{R}^k}{\operatorname{argmin}}\, Q(b) = \underset{b \in \mathbb{R}^k}{\operatorname{argmin}} \left\{ E[(g(w_t) - x_t'b)^2] + E[\varepsilon_t^2] \right\}$

$$= \underset{b \in \mathbb{R}^k}{\operatorname{argmin}}\, E[(g(w_t) - x_t'b)^2]$$

Remark:

$$E[(E[y_t|w_t] - x_t'b)\varepsilon_t] = E\left[E[(E[y_t|w_t] - x_t'b)\varepsilon_t | w_t]\right]$$

$$= E\left[(E[y_t|w_t] - x_t'b)\underbrace{E[\varepsilon_t|w_t]}_{=0}\right] = 0.$$

Buy Judd's Numerical Methods in Economics.