Review Session 1: Homework 3 & OLS Regression

Jiahui Shui
jishui@ucsd.edu

Rady School of Management, UCSD

November 8, 2024

Hints for HW3

- Period: January 1926 until December 2018
- CRSP_vw and CRSP_vwx: with / without dividends
- Don't spend to much time on Question 8 and 9

Conditional Expectations

Consider $\mathbb{E}[X|Y]$.

- If X and Y are independent, then $\mathbb{E}[X|Y] = \mathbb{E}X$
- If X is $\sigma(Y)$-measurable, then $\mathbb{E}[X|Y] = X$. In particular, $\mathbb{E}[f(Y)|Y] = f(Y)$
- $\mathbb{E}[Xf(Y)|Y] = f(Y)\mathbb{E}[X|Y]$ (pulling out known factors)
- $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$. Law of Iterated Expectation

Q: Suppose that X, Y are i.i.d, what is $\mathbb{E}[X|X+Y]$.
Hint: What is $\mathbb{E}[X+Y|X+Y]$.

Regression Review (Small/Finite Sample)

- Vector form:
$$y_t = x_t'\beta + \varepsilon_t$$
- Matrix form:
$$y = X\beta + \varepsilon$$

In finite sample theory, we have 4 assumptions:

- (**H1**) Linear Model: $y = X\beta + \varepsilon$
- (**H2**) Strict Exogeneity: $\mathbb{E}[\varepsilon_i|X] = 0$
  - $\mathbb{E}[y|X] = X\beta$
  - $\mathbb{E}[\varepsilon_i] = 0$
  - $\mathbb{E}[\varepsilon_i x_{j,k}] = 0, \forall j, k$
  - $\mathrm{Cov}(\varepsilon_i, x_{j,k}) = 0, \forall j, k$
- (**H3**) No Multicollinearity: $\mathrm{rank}(X) = K$
- (**H4**) Spherical disturbance: $\mathrm{Var}(\varepsilon|X) = \sigma^2 I_n$
  - Conditional Homoskedasticity: $\mathbb{E}[\varepsilon_i^2|X] = \sigma^2, \forall i$
  - No correlation: $\mathbb{E}[\varepsilon_i \varepsilon_j|X] = 0, \forall i \neq j$

Regression Review (Large Sample)

- (**A1**): Linear model. $y = X\beta + \varepsilon$
- (**A2**): Ergodic stationarity: $\{y_t, x_t\}$ is stationary and ergodic. (If two processes are far enough, say x_k and x_{t+k} as $t \to \infty$, then they can be thought as "independent")
- (**A3**): (Predetermined regressors): $\mathbb{E}[x_{t,i}\varepsilon_t] = 0$, $\forall i, t$. Define $g_t = x_t\varepsilon_t$, then $\mathbb{E}[g_t] = 0$.
- (**A4**): $\mathbb{E}[x_t x_t']$ has full rank.
- (**A5**): $\mathbb{E}[g_t g_t'] < \infty$ and g_t is a martingale difference sequence. Also, $\mathbb{E}[g_t g_t']$ has full rank.

Regression Review (Large Sample)

- Under A1-A4 $\Rightarrow \text{plim}_{T\to\infty} \hat{\beta} = \beta$.
- Additionally, if we have A5, then

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \text{Avar}(\hat{\beta})) \quad (1)$$

where

$$\text{Avar}(\hat{\beta}) = \Sigma_{xx}^{-1} S \Sigma_{xx}^{-1}, \quad \Sigma_{xx} = \mathbb{E}[x_t x_t'], \quad S = \mathbb{E}[g_t g_t'] \quad (2)$$

If x_t is a scalar, then

$$\Sigma_{xx} = \sigma_x^2, \quad S = \mathbb{E}[x_t^2 \varepsilon_t^2], \quad \text{Avar}(\hat{\beta}) = \frac{\mathbb{E}[x_t^2 \varepsilon_t^2]}{\sigma_x^4}$$

Remark: In lecture 3, Professor Valkanov wrote "intuitively" $\text{Avar}(\hat{\beta}) = \sigma_\varepsilon^2/\sigma_x^2$. This requires additional assumption that $\mathbb{E}[x_t^2 \varepsilon_t^2] = \mathbb{E}[x_t^2]\mathbb{E}[\varepsilon_t^2]$

Econometric issues in return predictability

Consider the following system

$$\begin{aligned} r_{t+1} &= \alpha + \beta dp_t + \varepsilon_{t+1} \\ dp_{t+1} &= \mu + \phi dp_t + u_{t+1} \end{aligned} \tag{3}$$

- In this system, dp_t is highly persistent ($\phi \approx 1$), $\beta > 0$.
- Now we suppose that $\mathbb{E}[dp_{t-1}\varepsilon_t] = 0$ by construction.
- $\hat{\phi}$ tends to be downward biased. This is standard issue in OLS.

$$\mathbb{E}[\hat{\phi}] = \phi - \frac{1+3\phi}{T} + O(1/T^2) \tag{4}$$

- And $\hat{\beta}$ is upward biased, which means that we reject the null of no predictability too often.

$$\mathbb{E}[\hat{\beta} - \beta] = \frac{\sigma_{\varepsilon u}}{\sigma_u^2}\mathbb{E}[\hat{\phi} - \phi] = -\frac{\sigma_{\varepsilon u}}{\sigma_u^2}\frac{1+3\phi}{T} \tag{5}$$

- Q: Why $\sigma_{\varepsilon u} < 0$? A positive *dp* shock usually has no news about dividends, so it means a negative *p* shock and a negative *r* shock.

Regression Review Question 1

1. Consider the following regression equation

$$y_t = \alpha + \beta x_t + \varepsilon_t \tag{6}$$

Assume that $\mathbb{E}[\varepsilon_t | x_t] = 0$ for all t.

Regression Review Question 1

(a) Prove that $\mathbb{E}[\varepsilon_t | x_t] = 0$ implies $\mathbb{E}[\varepsilon_t x_t] = 0$

Solution: By Law of Iterated Expection, we have

$$\mathbb{E}[\varepsilon_t x_t] = \mathbb{E}[\mathbb{E}[\varepsilon_t x_t | x_t]] = \mathbb{E}[x_t \mathbb{E}[\varepsilon_t | x_t]] = 0 \tag{7}$$

Regression Review Question 1

(b) Find the OLS estimator $\hat{\alpha}$ and $\hat{\beta}$

Solution: Define

$$Q(\alpha, \beta) = \sum_{t=1}^{T} (y_t - \alpha - \beta x_t)^2 \tag{8}$$

The FOCs are

$$\begin{aligned} \frac{\partial Q}{\partial \alpha} &= -2 \sum_{t=1}^{T} (y_t - \alpha - \beta x_t) = 0 \\ \frac{\partial Q}{\partial \beta} &= -2 \sum_{t=1}^{T} (y_t - \alpha - \beta x_t) x_t = 0 \end{aligned} \tag{9}$$

The first equation can be simplified as $\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$. Substitute this into the second equation, we get

$$\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t - T\bar{x}\bar{y}}{\sum_{t=1}^{T} x_t^2 - T\bar{x}^2} \tag{10}$$

Regression Review Question 1

(c) Let $\hat{y}_t := \hat{\alpha} + \hat{\beta} x_t$. Define $e_t = y_t - \hat{y}_t$. Show that

$$\sum_{t=1}^{T} e_t = 0$$

Solution: We have

$$\sum_{t=1}^{T} e_t = \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta} x_t) \tag{11}$$

Note that this is just the first FOC.

Regression Review Question 1

(d) Show that $\hat{\beta}$ is unbiased, i.e. $\mathbb{E}[\hat{\beta}|X] = \beta$, where $X = (x_1, \cdots, x_T)'$

Solution: Note that

$$\mathbb{E}[\hat{\beta}|X] = \mathbb{E}\left[\frac{\sum_{t=1}^{T} x_t y_t - T\bar{x}\bar{y}}{\sum_{t=1}^{T} x_t^2 - T\bar{x}^2}\middle|X\right] = \frac{\sum_{t=1}^{T} x_t \mathbb{E}[y_t|X] - T\bar{x}\mathbb{E}[\bar{y}|X]}{\sum_{t=1}^{T} - T\bar{x}^2} \tag{12}$$

and

$$\mathbb{E}[y_t|X] = \mathbb{E}[\alpha + \beta x_t + \varepsilon_t|X] = \alpha + \beta x_t \tag{13}$$

Then

$$\mathbb{E}[\bar{y}|X] = \frac{1}{T}\mathbb{E}[\sum_{t=1}^{T} y_t|X] = \alpha + \beta\bar{x} \tag{14}$$

Substituting those two equations into equation (12) yields

$$\mathbb{E}[\hat{\beta}|X] = \beta \tag{15}$$

Notes: Unbiased property requires $\mathbb{E}[\varepsilon_t|X] = 0$.

Regression Review Question 1

(e) Now, consider the following regression equation

$$y_t = \beta x_t + \varepsilon_t \tag{16}$$

Find the OLS estimator $\hat{\beta}$. Calculate $\sum_{t=1}^{T} e_t$ again.

Solution: Now

$$Q(\beta) = \sum_{t=1}^{T} (y_t - \beta x_t)^2 \tag{17}$$

The FOC gives us

$$\frac{\partial Q(\beta)}{\beta} = -2 \sum_{t=1}^{T} (y_t - \beta x_t) x_t = 0 \Rightarrow \hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} \tag{18}$$

Now $\sum_{t=1}^{T} e_t$ is not necessarily to be 0.

$$\sum_{t=1}^{T} e_t = \sum_{t=1}^{T} y_t - \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} \sum_{t=1}^{T} x_t \tag{19}$$

Regression Review Question 2

2. Consider the following AR(1) process

$$x_t = \rho x_{t-1} + \varepsilon_t \tag{20}$$

where $|\rho| < 1$, $\{\varepsilon_t\}_{t=0}^{\infty}$ is white noise with variance σ^2. Suppose that $\mathbb{E}[x_{t-1}\varepsilon_t] = 0$.
(a) Is the OLS estimator $\hat{\rho}$ unbiased? Is $\hat{\rho}$ consistent?

Solution: No, $\hat{\rho}$ is biased. Since unbiased estimation requires that $\mathbb{E}[\varepsilon_t|X] = 0$, meaning ε_t is uncorrelated with all x_t, both past and future. Obviously here ε_t is correlated to future x_t.
But $\hat{\rho}$ is consistent. Consistency only requires predetermined explanatory variables, i.e. $\mathbb{E}[x_{t-1}\varepsilon_t] = 0$.

Regression Review Question 2

(b) Find $\mathbb{E}[x_t]$ and $\mathrm{Var}(x_t)$

Solution: There are some tricks here. If we impose x_t is stationary here, then $\mathbb{E}[x_t] = \mathbb{E}[x_{t-1}], \mathrm{Var}(x_t) = \mathrm{Var}(x_{t-1})$. Hence

$$\mathbb{E}[x_t] = \rho\mathbb{E}[x_{t-1}] \Rightarrow \mathbb{E}[x_t] = 0 \tag{21}$$

$$\mathrm{Var}(x_t) = \rho^2\mathrm{Var}(x_{t-1}) + \sigma^2 \Rightarrow \mathrm{Var}(x_t) = \frac{\sigma^2}{1-\rho^2} \tag{22}$$

Regression Review Question 3

3. Consider the following regression

$$\begin{aligned} y_t &= \alpha + \beta y_{t-1} + u_t \\ u_t &= v_t + \theta v_{t-1} \end{aligned} \tag{23}$$

where v_t is i.i.d with mean 0 and variance σ^2. Is the OLS estimator $\hat{\beta}$ unbiased? Is it consistent?

Solution: Consider

$$\mathrm{Cov}(y_{t-1}, u_t) = \mathrm{Cov}(v_{t-1} + \theta v_{t-2}, v_t + \theta v_{t-1}) = \theta\sigma^2$$

Hence it is biased and not consitent.