Regression in Observational Studies

Julian Gerez

Columbia University

POLS-GU4722: Statistical Theory and Causal Inference

Spring 2021

Agenda

Simple Least Squares Estimation

We observe n i.i.d samples of $\{Y_i, X_i\}_{i=1}^{n}$, where Y_i is the outcome variable and X_i is the predictor variable, and $\mathbb{E}(\epsilon_i) = 0$.

$$Y_i \; = \; \alpha + \beta X_i + \epsilon_i$$

The residuals are $\hat{\epsilon} = Y_i - \hat{\alpha} - \hat{\beta} X_i$, and we want to minimze the sum of squared residuals:

$$\underset{(\hat{\alpha}, \hat{\beta})}{\text{argmin}} \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2$$

How do we solve? Take the partial derivatives of $\hat{\alpha}$ and $\hat{\beta}$ and set these equal to zero. Easy to see that these will be a minimum.

$$\frac{\partial SSR}{\partial \hat{\alpha}} = \sum_{i=1}^{n} -2(Y_i - \hat{\alpha} - \hat{\beta} X_i)$$

$$\frac{\partial SSR}{\partial \hat{\beta}} = \sum_{i=1}^{n} -2X_i(Y_i - \hat{\alpha} - \hat{\beta} X_i)$$

Simple Least Squares Estimation: Constant

Recall $\frac{\partial SSR}{\partial \hat{\alpha}} = \sum_{i=1}^{n} -2(Y_i - \hat{\alpha} - \hat{\beta}X_i)$.

$$
\begin{aligned}
0 &= \sum_{i=1}^{n} -2(Y_i - \hat{\alpha} - \hat{\beta}X_i) \\
0 &= \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}X_i) \\
0 &= \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{\alpha} - \sum_{i=1}^{n} \hat{\beta}X_i \\
0 &= \sum_{i=1}^{n} Y_i - n\hat{\alpha} - \sum_{i=1}^{n} \hat{\beta}X_i \\
\hat{\alpha} &= \frac{\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \hat{\beta}X_i}{n} \\
\hat{\alpha} &= \overline{Y} - \hat{\beta}\overline{X}
\end{aligned}
$$

Simple Least Squares Estimation: Coefficient

Recall $\frac{\partial SSR}{\partial \hat{\beta}} = \sum_{i=1}^{n} -2X_i(Y_i - \hat{\alpha} - \hat{\beta}X_i)$ and that $\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$.

$$0 = \sum_{i=1}^{n} -2X_i(Y_i - \hat{\alpha} - \hat{\beta}X_i)$$

$$0 = \sum_{i=1}^{n}(X_iY_i - \hat{\alpha}X_i - \hat{\beta}X_i^2)$$

$$0 = \sum_{i=1}^{n}(X_iY_i - X_i\overline{Y} + \hat{\beta}X_i\overline{X} - \hat{\beta}X_i^2)$$

$$0 = \sum_{i=1}^{n}(X_iY_i - X_i\overline{Y}) + \hat{\beta}\sum_{i=1}^{n}(X_i^2 - X_i\overline{X})$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_iY_i - X_i\overline{Y})}{\sum_{i=1}^{n}(X_i^2 - X_i\overline{X})}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \xrightarrow{p} \frac{Cov(X_i, Y_i)}{\text{Var}(X_i)} = \rho_{XY}\sqrt{\frac{\text{Var}(Y_i)}{\text{Var}(X_i)}}$$

Unbiasedness of Least Squares Estimator

First, we want to show that $\mathbb{E}(\hat{\beta}) = \beta$.

$$
\begin{aligned}
\hat{\beta} &= \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\
&= \frac{\sum_{i=1}^{n}\Big[(\alpha + \beta X_i + \epsilon_i) - (\alpha + \beta\overline{X} + \overline{\epsilon})\Big](X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\
&= \frac{\sum_{i=1}^{n}\Big[\beta(X_i - \overline{X}) + \epsilon_i - \overline{\epsilon}\Big](X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\
&= \frac{\beta\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{i=1}^{n}(X_i - \overline{X})\epsilon_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\
\hat{\beta} - \beta &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})\epsilon_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2}
\end{aligned}
$$

Exogeneity implies:

$$
\mathbb{E}(\hat{\beta}) - \beta = \mathbb{E}(\hat{\beta} - \beta) = \mathbb{E}_X[\mathbb{E}(\hat{\beta} - \beta)|\mathbf{X}] = 0
$$

Unbiasedness of Least Squares Estimator

Knowing that $\mathbb{E}(\hat{\beta}) = \beta$ makes the proof that $\mathbb{E}(\hat{\alpha}) = \alpha$ easier.

$$\begin{aligned}
\hat{\alpha} &= \overline{Y} - \hat{\beta}\overline{X} \\
&= \alpha + \beta\overline{X} + \overline{\epsilon} - \hat{\beta}\overline{X} \\
\hat{\alpha} - \alpha &= \overline{\epsilon} - (\hat{\beta} - \beta)\overline{X}
\end{aligned}$$

Using exogeneity and $\mathbb{E}(\hat{\beta}) = \beta$:

$$\mathbb{E}(\hat{\alpha}) - \alpha = \mathbb{E}[\overline{\epsilon} - (\hat{\beta} - \beta)\overline{X}] = 0$$

Derivation of Variance for Simple OLS

We want to show that $\text{Var}(\hat{\beta}|\mathbf{X}) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2}$.

$$\begin{aligned}
\text{Var}(\hat{\beta}|\mathbf{X}) &= \text{Var}\left[\frac{\sum_{i=1}^n (Y_i - \overline{Y})(X_i - \overline{X})}{\sum_{i=1}^n (X_i - \overline{X})^2}\Big|\mathbf{X}\right] \\
&= \text{Var}\left[\frac{\sum_{i=1}^n (X_i - \overline{X})\epsilon_i}{\sum_{i=1}^n (X_i - \overline{X})^2}\Big|\mathbf{X}\right] \\
&= \frac{\sum_{i=1}^n (X_i - \overline{X})^2 \text{Var}(\epsilon_i|\mathbf{X})}{\sum_{i=1}^n \left[(X_i - \overline{X})^2\right]^2} \\
&= \frac{\sum_{i=1}^n (X_i - \overline{X})^2 \sigma^2}{\sum_{i=1}^n \left[(X_i - \overline{X})^2\right]^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2}
\end{aligned}$$

Estimating The Sampling Variance

Remember that $\text{Var}(\hat{\beta}|\mathbf{X}) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$, but note that we cannot observe σ^2. In practice we use:

$$\hat{V} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{\epsilon}_i^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Conditionally unbiased: $\mathbb{E}(\hat{\sigma}^2|\mathbf{X}) = \sigma^2$ implies

$$\mathbb{E}(\hat{V}|\mathbf{X}) \;=\; \text{Var}(\hat{\beta}|\mathbf{X})$$

Unconditionally unbiased: $\text{Var}[\mathbb{E}(\hat{\beta}|\mathbf{X})] = 0$ implies

$$\mathbb{E}(\hat{V}) \;=\; \mathbb{E}_X[\mathbb{E}(\hat{V}|\mathbf{X})] = \mathbb{E}[\text{Var}(\hat{\beta}|\mathbf{X})] = \text{Var}(\hat{\beta})$$

Consistency

Recall:

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^{n}(X_i - \overline{X})\epsilon_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Since we assume i.i.d:

$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})\epsilon_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \xrightarrow{p} \frac{\text{Cov}(X_i, \epsilon_i)}{\text{Var}(X_i)}$$

Exogeneity implies that $\text{Cov}(X_i, \epsilon_i) = 0$. So if $\text{Var}(X_i) > 0$:

$$\hat{\beta} \xrightarrow{p} \beta$$

Model-Based Asymptotic Inference

Asymptotic distribution and inference $\sqrt{n}(\hat{\beta} - \beta)$:

$$\left(\underbrace{\sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^{n} [X_i - \mathbb{E}(X_i)]\epsilon_i}_{\xrightarrow{d} \mathcal{N}[0,\ \sigma^2 \text{Var}(X_i)]} + \underbrace{\sqrt{n} \cdot (\mathbb{E}[(X_i) - \overline{X}] \frac{1}{n} \sum_{i=1}^{n} \epsilon_i}_{\xrightarrow{p} 0} \right) \times$$

$$\underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \right)^{-1}}_{\xrightarrow{p} \text{Var}(X_i)^{-1}} \xrightarrow{d} \mathcal{N}\left(0,\ \frac{\sigma^2}{\text{Var}(X_i)}\right)$$

Therefore, using a consistent estimator of the standard error:

$$\frac{\hat{\beta} - \beta}{\text{s.e.}} \xrightarrow{d} \mathcal{N}(0, 1)$$

We can calculate confidence intervals as follows:

- $(1 - \alpha) \times 100$ % CI: $[\widehat{\beta} - z_{1-\alpha/2} \cdot \text{s.e.},\ \ \widehat{\beta} + z_{1-\alpha/2} \cdot \text{s.e.}]$
- This will be asymptotically equivalent to the CI based on t-distribution

OLS Derivation with Vector Calculus

For OLS, we try and minimize the sum of squared residuals: $||Y - \mathbf{X}\hat{\beta}||^2$. Let's start by rewriting the SSR.

$$
\begin{aligned}
||Y - \mathbf{X}\hat{\beta}||^2 &= (Y - \mathbf{X}\hat{\beta})^\top (Y - \mathbf{X}\hat{\beta}) \\
&= Y^\top Y - \hat{\beta}^\top \mathbf{X}^\top Y - Y^\top \mathbf{X}\hat{\beta} + \hat{\beta}^\top \mathbf{X}^\top \mathbf{X}\hat{\beta} \\
&= Y^\top Y - 2\hat{\beta}^\top \mathbf{X}^\top Y + \hat{\beta}^\top \mathbf{X}^\top \mathbf{X}\hat{\beta}
\end{aligned}
$$

Now, we find the first order condition $\frac{\partial \mathrm{SSR}}{\partial \hat{\beta}} = 0$:

$$
\begin{aligned}
\frac{\partial \mathrm{SSR}}{\partial \hat{\beta}} &= -2\mathbf{X}^\top Y + 2\mathbf{X}^\top \mathbf{X}\hat{\beta} = 0 \\
(\mathbf{X}^\top \mathbf{X})\hat{\beta} &= \mathbf{X}^\top Y
\end{aligned}
$$

OLS Derivation with Vector Calculus

$(\mathbf{X}^\top \mathbf{X})\hat{\beta} = \mathbf{X}^\top Y$ is known as the normal equation. Now we can solve for $\hat{\beta}$:

$$\begin{aligned}(\mathbf{X}^\top \mathbf{X})^{-1}(\mathbf{X}^\top \mathbf{X})\hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top Y \\ \hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top Y\end{aligned}$$

Is this a minimum? Check second order condition: $\frac{\partial^2 \text{SSR}}{\partial \hat{\beta} \partial \hat{\beta}^\top} \geq 0$.

$$\frac{\partial \text{SSR}}{\partial \hat{\beta}} = -2\mathbf{X}^\top Y + 2\mathbf{X}^\top \mathbf{X}\hat{\beta} \Longrightarrow \frac{\partial^2 \text{SSR}}{\partial \hat{\beta} \partial \hat{\beta}^\top} = 2\mathbf{X}^\top \mathbf{X}$$

How do we know if a matrix is "positive"? Notion of definiteness. Two conditions for a square matrix $\mathbf{A}$ to be positive semi-definite: 1) $\mathbf{A}$ is symmetric and 2) $c^\top \mathbf{A} c \geq 0$ for any column vector c.

$\mathbf{X}^\top \mathbf{X}$ is symmetric and $c^\top \mathbf{X}^\top \mathbf{X} c = ||\mathbf{X}c||^2 \geq 0 \Longrightarrow \mathbf{X}^\top \mathbf{X}$ is positive semi-definite (therefore, it is a minimum).