ECON220C Discussion Section 1

Review on Causality

Lapo Bini

Roadmap

1. Feedback 220B
2. From Linear Projection to Structural Model
3. DAG Representation
4. Exercise: Conditional Mean Independence
5. Exercise: Probability Limit OLS

ASYMPTOTIC LINEAR REPRESENTATION

Goal for Today

Suppose we have some data $\{y_i, x_i\}_{i=1}^n$, we want to understand what is required to go from a prediction model (reduced form) to a causal model (structural model). We will go through:

1. Linear projection: no assumptions $\Longrightarrow$ mathematical device.
2. Correlation: different interpretation of linear regression: $E[x_i u_i] = 0$.
3. Causal model: $E[u_i | x_i] = 0$ plus causal structure (DAG).

Linear Projection

Suppose we have some data $\{y_i, x_i\}_{i=1}^{n}$ and we estimate $y_i = \delta_0 + x_i\delta_1 + e_i$ by OLS to study the predictive power of x_i. The linear projection gives us:

1. $\hat{\delta}$ consistent for true δ, known as the **minimum mean square error linear predictor** since δ solves the following problem:

$$\min_{\mathrm{d}_0, \mathrm{d}_1 \in \mathbb{R}} E\left[(y_i - \mathbf{d}_0 - x_i\mathbf{d}_1)^2\right]$$

2. The linear projection **always** satisfies $E[x_i e_i] = 0$ and $E[e_i] = 0$.

Linear Projection

The results $E[x_i e_i] = 0$ and $E[e_i] = 0$ are tautological, why?

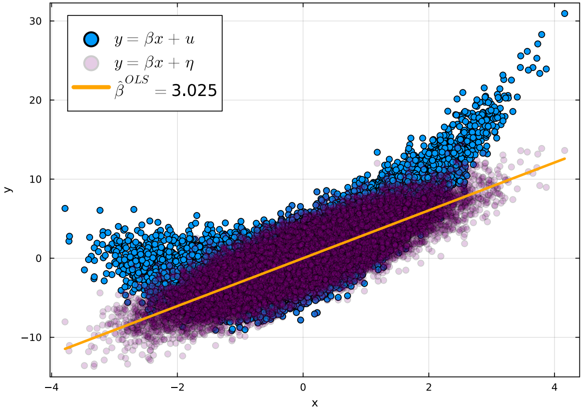
Linear Regression as Correlation

Model: $y_i = x_i^T \beta + u_i$

- If we impose $E[u_i x_i] = 0$, then $\hat{\beta}_{OLS}$ is consistent for the true parameter β which can be interpreted in terms of correlation.
- How is this different from $E[x_i e_i] = 0$ before?
- Remember example from 220B: $y_i = \beta x_i + u_i$, $u_i = x_i^2 + \eta_i$ with true parameter $\beta = 3$, and $x_i \sim \mathcal{N}(0, 1)$, $\eta_i \sim \mathcal{N}(0, 4)$, $x_i \perp \eta_i$.

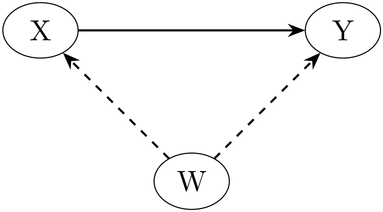
Linear Regression as Correlation

Model: $y_i = \beta x_i + u_i$ $u_i = x_i^2 + \eta_i$ with true parameter $\beta = 3$, and $x_i \sim \mathcal{N}(0,1)$, $\eta_i \sim \mathcal{N}(0,4)$, $x_i \perp \eta_i$.



Causality (I/II)

(i) We need to impose a precise causal structure. Let's consider the Directed Acyclic Graph (DAG) below:



(ii) Linear causal structure:

Causality (II/II)

(iii) Exogeneity condition: $E[u_i|x_i] = 0 \rightarrow$ after controlling for the covariates, the error term does not know anything about the regressors.

Question: in a linear regression model, say $y = \beta_1 x_1 + \beta_2 x_2 + u$, if $E[u_i|\mathbf{x}_i] = 0$ we interpret β_1 as the causal effect of x_1 holding x_2 constant. This is a partial effect: what happens to y when you change x_1 by one unit, keeping x_2 fixed. Is that meaningful if x_1 and x_2 are correlated?

Exercise 1 (Final 2024)

Consider a causal system $Y \leftarrow X\beta + W\gamma + U$ where the variable on the left side is caused by the variables on the right side. Let Z be a potential causal factor of X, W, U and suppose an iid sample $\{Y_i, X_i, W_i, U_i\}_{i=1}^n$ is collected from the above model. Explain whether β and γ can be consistently estimated under each of the following conditions, assuming that there is no perfect multicollinearity:

(i) $E[U|X, W, Z] = W\gamma_w + Z\gamma_z$

(ii) $E[U|X, W, Z] = X\gamma_x + Z\gamma_z$

(iii) $E[U|X, W, Z] = Z\gamma_z$

(iv) $E[U|X, W, Z] = W^2\gamma_w + Z\gamma_z$

(v) $E[U|X, W, Z] = X\gamma_x + Z^2\gamma_z$

Exercise 2 (Midterm 2024)

Suppose X and Y are generated according to the following model:

$$Y = X^3 + U$$

where X and U are independent random variables with mean 0. We have an iid sample $\{Y_i, X_i, U_i\}_{i=1}^{n}$ from the above population model. Now suppose we do not know the above model and we estimate the following linear regression model by OLS and QML:

$$Y + \alpha + X\beta + error$$

where $error \sim iid\, \mathcal{N}(0, \sigma_e^2)$.

Exercise 2 (Midterm 2024)

(i) Find a closed form solution for the probability limit of the linear projection coefficient $\hat{\beta}_{OLS}$.

(ii) Define $e = Y - \alpha - X\beta$, show that $Cov(e, X) = 0$: Is e independent of X?

(iii) Suppose you want to compute the asymptotic variance of $\sqrt{n}(\hat{\beta}_{OLS} - \beta)$. Do you need to use a variance estimator robust to conditional heteroskedasticity?

(iv) The QML estimator is based on the wrong model and the wrong assumption on the error term (normality). What does $\hat{\beta}_{QML}$ converge to (in probability) as $n \to \infty$?