Empirical Methods

Topic 1c:

The CLRM: Interpretation and Multicollinearity

The Classical Linear Regression Model Interpretation

CLRM Interpretation: Intro I

- Time for a change in focus
- To this point, we've focused mostly on *calculating* the CLRM
- This (short) deck is going to teach you how to *interpret* the CLRM coefficients
    - Critically important: you *must* know what it is you're getting out when you type 'ols y x1 x2 x3' in Stata
        - (Or the equivalent in R)
- I'll also discuss the (more-important-than-you'd-think) topic of Multicollinearity

Interpretation of Multiple Regression Coefficients I

- How then should you interpret one of the OLS coefficient, $\hat{\beta}_k$?
- Well... it turns out there are two ways to answer this question:
  1. At a conceptual level
  2. At a practical level
- Let's consider each in turn...

Conceptual Interpretation of $\hat{\beta}_k$ I

- First the conceptual interpretation
  - Often this is *implicit*, but I want to make it *explicit*
- Recall that OLS estimates the (assumed-linear) population regression function

$$\begin{aligned} y_i &= E(y_i|x_i) + \epsilon_i \\ &= x_i'\beta + \epsilon_i \\ &= \beta_1 + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + \epsilon_i \end{aligned}$$

Conceptual Interpretation of $\hat{\beta}_k$ II

- As emphasized in the last deck, there are two conceptual interpretations for $\hat{\beta}_k$:
  1. A "correlational" interpretation:
     - ★ Where $\hat{\beta}_k$ measures the "(partial) correlation" between y_i and x_{ik}
     - ★ (i.e. the correlation between y_i and x_{ik} controlling for $x_{i,-k}$)
  2. A "causal" interpretation:
     - ★ Where $\hat{\beta}_k$ measures the causal effect on y_i from a one-unit change in x_{ik}
     - ★ But only if the CLRM Assumption 2 (Mean-zero error) is satisfied!
- For the rest of this deck, we will assume that (A2) *is* satisfied
  - Just don't forget it's in the background!

Practical Interpretation of $\hat{\beta}_k$ I

- OK, so we'll assume (A2, Mean-zero error) is satisfied
  - And so $\hat{\beta}_k$ measures a causal effect as a conceptual matter
- But *what* causal effect?
  - As a *practical* matter, how do we interpret $\hat{\beta}_k$?

Practical Interpretation of $\hat{\beta}_k$ II

- We turn to calculus to help us:
- We can caculate the partial derivative of $E(y_i|x_i)$ w.r.t. x_{ik}:

$$\begin{aligned} \frac{\partial E(y_i|x_i)}{\partial x_{ik}} &= \frac{\partial[\beta_1 + \beta_2 x_{i2} + \ldots + \beta_K x_{iK}]}{\partial x_{ik}} \\ &= \beta_k \end{aligned}$$

Practical Interpretation of $\hat{\beta}_k$ III

$$\begin{aligned} \frac{\partial E(y_i|x_i)}{\partial x_{ik}} &= \frac{\partial[\beta_1 + \beta_2 x_{i2} + ... + \beta_K x_{iK}]}{\partial x_{ik}} \\ &= \beta_k \end{aligned}$$

(As long as Assumption 2 holds...)

- β_k measures the change in the expected value of y_i when we change x_{ik} by a small amount (e.g. one unit)...
  - ..."controlling for" (i.e. holding fixed) the other variables in the regression
    - ★ (Recall this is the interpretation of a partial derivative)
- We call β_k the marginal effect of x_{ik} on y_i
  - And $\hat{\beta}_k$ is our estimate of this marginal effect

Practical Interpretation of $\hat{\beta}_k$ IV: Midpoint Summary

- This "controlling for" other covariates interpretation of regression coefficients is *very important*
- You can get very different estimates of economic effects depending on the other covariates for which you control
  - See Stata Example in class
- Why???
  - ____

The Projection Matrix, The Residual Maker, and Partitioned Regression

Mathematical Interpretation of Multiple Regression I

- You've probably all seen or heard this much before
- I'm going to go one step further...
  - ...and show you how you can use Partitioned Matrices to provide another intuition for multiple regression coefficients
  - I call this a "Mathematical Interpretation of Multiple Regression"
- I do this for two reasons:
  1. It will deepen your intuition for the interpretation of $\hat{\beta}_k$
  2. The math involved is used in many settings in econometrics
     - ★ Especially Instrumental Variables

The Projection Matrix I

- We start by introducing the *Projection Matrix, P_X*
- Recall the definition of the fitted value for y:

$$\begin{aligned} \hat{y} &= X\hat{\beta} \\ &= X(X'X)^{-1}X'y \\ &= \underbrace{X(X'X)^{-1}X'}_{P_X}y \\ &= P_X y \end{aligned}$$

where $P_X = X(X'X)^{-1}X'$ is an $N \times N$ "idempotent" matrix

- (About which more in a moment...)

The Projection Matrix II

$$\boxed{P_X = X(X'X)^{-1}X'}$$

- P_X plays a special role in econometrics:
  - It gives the fitted value of <whatever> when you regress <whatever> on X. Thus
    - $P_X y$ is the fitted value of y when you regress it on X
    - $P_X z$ is the fitted value of z when you regress it on X

The Residual Maker I

- Another matrix we particularly need is P_X's "sibling"
- Recall the definition of the residual, e:

$$\begin{aligned} e &= y - X\hat{\beta} \\ &= y - X(X'X)^{-1}X'y \\ &= (I_N - X(X'X)^{-1}X')y \\ &= \underbrace{(I_N - X(X'X)^{-1}X')}_{M_X} y \\ &= M_X y \end{aligned}$$

where $M_X = I_N - X(X'X)^{-1}X' = I_N - P_X$ is (also) an $N \times N$ "idempotent" matrix

The Residual Maker II

- M_X also plays a special role in econometrics:
  - It gives the residual of <whatever> when you regress <whatever> on X. Thus
    - ★ $M_X y$ is the residual of y when you regress it on X
    - ★ $M_X z$ is the residual of z when you regress it on X
- I call M_X the "residual maker" (for X)
  - (M_X is also called the "annihilator matrix"...
    - ★ ...but I find that a bit too aggressive!)

Properties of P_X and M_X

- P_X and M_X have special properties:
  1. They are square and symmetric
     - ★ $P_X = P_X'$, $M_X = M_X'$
  2. They are idempotent
     - ★ How many of you have (not) heard of idempotent matrices?
     - ★ $\equiv P_X P_X = P_X$, $M_X M_X = M_X$
- We will use these properties momentarily...

Mathematical Interpretation of Multiple Regression II

- In order to provide intuition for how to interpret $\hat{\beta}_k$ when there are other regressors, we need to briefly discuss partitioned matrices
- Suppose we care most about one particular element of β, β_k, in the regression of y on X
  - (This is very common in econometrics - where you really care about one/a few variables)
- We therefore "partition" X into X_k and X_{-k}
  - where X_k is $N \times 1$ and X_{-k} is $N \times (K-1)$
- We partition β similarly into β_k and β_{-k}
  - where β_k is 1×1 and β_{-k}, is $(K-1) \times 1$

Mathematical Interpretation of Multiple Regression III

- Then we can write the CLRM as

$$\begin{aligned} y &= X\beta + \epsilon \\ &= X_k\beta_k + X_{-k}\beta_{-k} + \epsilon \end{aligned}$$

where

  - $X = [X_k \ X_{-k}]$ is (still) an $N \times K$ matrix
  - $\beta = [\beta_k \ \beta'_{-k}]'$ is (still) a $K \times 1$ vector*

Mathematical Interpretation of Multiple Regression IV

- What does $\hat{\beta}$ look like now?
- Well, it is still the case that $\hat{\beta} = (X'X)^{-1}X'y$, but now we can write this as:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_k \\ \hat{\beta}_{-k} \end{bmatrix} = \begin{bmatrix} X_k'X_k & X_k'X_{-k} \\ X_{-k}'X_k & X_{-k}'X_{-k} \end{bmatrix}^{-1} \begin{bmatrix} X_k'y \\ X_{-k}'y \end{bmatrix}$$

$$= \text{Trust me on this:}$$

$$= \begin{bmatrix} (X_k'M_{-k}X_k)^{-1}X_k'M_{-k}y \\ (X_{-k}'M_kX_{-k})^{-1}X_{-k}'M_ky \end{bmatrix} *$$

where

- $M_{-k} = I_N - X_{-k}(X_{-k}'X_{-k})^{-1}X_{-k}'$ is the "residual maker" for X_{-k}

Mathematical Interpretation of Multiple Regression V

- Let's focus on the coefficient we care about, $\hat{\beta}_k$

$$\boxed{\hat{\beta}_k = (X_k' M_{-k} X_k)^{-1} X_k' M_{-k} y}$$

- There is a *ton* of intuition in the formula above
    - [Believe it or not]
- But to unlock it we have to do a few... more... things.

Mathematical Interpretation of Multiple Regression VI

$$\boxed{\hat{\beta}_k = (X_k' M_{-k} X_k)^{-1} X_k' M_{-k} y}$$

- Define:

$$X_k^* = M_{-k} X_k$$

- How should we interpret X_k^*?*
  - ______________________________
- What is its dimension?*
  - ______________________________

Mathematical Interpretation of Multiple Regression VII

$$\boxed{\hat{\beta}_k = (X_k' M_{-k} X_k)^{-1} X_k' M_{-k} y}$$

- Now let's manipulate the equation above using these definitions

$$\begin{aligned}
\hat{\beta}_k &= (X_k' M_{-k} X_k)^{-1} X_k' M_{-k} y \\
&= (X_k' M_{-k} M_{-k} X_k)^{-1} X_k' M_{-k} y \quad \text{because } M_{-k} \text{ is idempotent} \\
&= (X_k' M_{-k}{}' M_{-k} X_k)^{-1} X_k' M_{-k}{}' y \quad \text{because } M_{-k} \text{ is symmetric} \\
&= (X_k^{*\prime} X_k^*)^{-1} X_k^{*\prime} y^* \quad \text{because } X_k^* = M_{-k} X_k \text{ and } X_k^{*\prime} = X_k' M_{-k}'
\end{aligned}$$

$$\boxed{\hat{\beta}_k = (X_k^{*\prime} X_k^*)^{-1} X_k^{*\prime} y}$$

Mathematical Interpretation of Multiple Regression VIII*

$$\boxed{\hat{\beta}_k = (X_k^{*\prime} X_k^*)^{-1} X_k^{*\prime} y}$$

- !!!
- The k^{th} coefficient in a multiple OLS regression is *equivalent* to the coefficient in a simple OLS regression of y on the residual from a regression of X_k on all the other regressions ($X_k^* \equiv M_{-k} X_k$).
  - See Stata Example in class
- Thus every time you run a multiple regression and look at the coefficient $\hat{\beta}_k$, you can think of having done a two-step process:
  1. Having run a regression of X_{ik} on all the other vars in the reg, $X_{i,-k}$
     - Yielding the residual e_{i,k_on_-k}
  2. Having run a simple regression of y_i on e_{i,k_on_-k}

Mathematical Interpretation of Multiple Regression IX

$$\boxed{\hat{\beta}_k = (X_k^{*\prime} X_k^*)^{-1} X_k^{*\prime} y}$$

- This is the sense in which including other covariates, X_{-k}, in a regression "controls for" these covariates
  - The only variation left in X_k to identify β_k is the variation left *after* running a regression of X_k on X_{-k}
  - We will see a stark example of this momentarily when we get to multicollinearity

The Constant Term

Latex Color = "LightGrey"

The Constant Term I

- Everything to this point has (implicitly) focused on the $K - 1$ "slope" coefficients in the regression
    - i.e. the coefficients attached to x_{ik} that aren't the constant term
        - ★ (Recall we set the constant term, $x_{i1} = 1 \; \forall i$)
- Let me highlight two features of this constant term:
    1. The impact "controlling for it" has on all the other covariates*
        - ★ We won't cover that this year - essentially it "mean-deviates" all the variables
    2. Its interpretation in a regression

The Constant Term II*

$$\boxed{y^* \;=\; M_{-k}y \qquad X_k^* \;=\; M_{-k}X_k}$$

- Recall the formulas above for our dependent and explanatory variables when we "control" for X_{-k}
- Let me tell you what this means if $X_{-k} = \iota$,
  - Where ι is mathematically how we write a constant term on a computer
  - (Namely: as an $N \times 1$ vector of ones)

$$\iota \;=\; \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

The Constant Term III*

- It turns out that when $M_{-k} = M_\iota$,

$$\boxed{\begin{array}{ccccc} y^* & = & M_{-k}y & = & y - \bar{y} \\ X^* & = & M_{-k}X_k & = & X - \bar{X} \end{array}}$$

- In essence, including a constant term "mean-deviates" all of the other variables in the econometric model
  - This is just a good thing to know

The Constant Term IV

$$\boxed{\hat{\beta}_1 \quad = \quad \bar{y} - \bar{X}'_{-1}\hat{\beta}_{-1}}$$

- What then is the interpretation of the constant term, $\hat{\beta}_1$?
- Some of you may remember what its interpretation is:
  - The expected value of y_i when (all) x_i are zero
  - Often this isn't very meaningful!
    - (Revisit Stata example in class)

The Constant Term V

- Even if it's not very meaningful, *it is almost always correct* to include a constant term in every regression
  - As if not, you are *forcing* the sample regression function to go through the origin
    - This will generally bias *all* of your other $\hat{\beta}s$ if wrong*

Multicollinearity

Latex Color = "LightGrey"

Multicollinearity I

- There is one last condition that I didn't yet mention for calculation of the OLS estimator
- It's easiest to see for the slope coefficient in simple regression

$$\hat{\beta}_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

- What must be true about the values for $\hat{\beta}_2$ to be calculable?
  - ____________________________________

Multicollinearity II

- The matrix analog of this condition, for...

$$\boxed{(X'X)^{-1}X'y}$$

  - ...is that $X'X$ must be *invertible*
    - ★ i.e. we must be able to take the *inverse* of $X'X$
  - When $X'X$ *isn't* invertible, we say that the model suffers from "perfect multicollinearity"

Multicollinearity III

- Perfect multicollinearity is rare and usually means you made a mistake
  - It happens when there is a perfect linear combination among the explanatory variables
- When this happens, statistical packages like Stata simply drop one of the collinear variables...
  - ...(which one doesn't matter)...
  - ...and runs the regression minus the dropped variable
    - ★ See Stata Example in class

Multicollinearity IV

- It can happen, however, that you get near-perfect multicollinearity
  - This is actually *more* problematic as Stata will still run your regression
  - So you won't necessarily know there is a problem
- Does anyone know what usually happens to the results?*
  - ____
  - See Stata Example in class
- What is the intuition for this?*
  - ____

Multicollinearity V

- To see this last point, recall our formula for $\hat{\beta}_k$

$$\boxed{\begin{aligned}\hat{\beta}_k &= (X_k' M_{-k} X_k)^{-1} X_k' M_{-k} y \\ &= (X_k^{*\prime} X_k^*)^{-1} X_k^{*\prime} y\end{aligned}}$$

  where

$$\begin{aligned} y^* &= M_{-k} y \\ X_k^* &= M_{-k} X_k \end{aligned}$$

  - When there is very high multicollinearity between two variables, k and l, X_k^* doesn't have much variation "left over" after controlling for X_l
    - ★ (And vice versa!)
    - ★ See Stata Example in class
  - As such, $V(\hat{\beta}_k)$ is very high
    - ★ (As is $V(\hat{\beta}_l)$)

Closing thoughts

- OK, that's it for the *estimation* part of the CLRM
- We covered:
  - OLS calculation
  - CLRM assumptions
  - CLRM interpretation
- What's next?
  - *Hypothesis testing* in the CLRM

Table of Contents