# Lecture 1: Review and Simulation Methods

Chris Conlon

February 1, 2021

NYU Stern

# Introduction

Consider a linear regression:

$$Y_i = X_i'\beta + \varepsilon_i \qquad \text{with} \qquad \mathbb{E}[\varepsilon_i | X_i] = 0$$

We've discussed the least squares estimator:

$$\widehat{\beta}_{ols} = \arg\min_{\beta} \sum_{i=1}^{N} (Y_i - X_i'\beta)^2$$

$$\widehat{\beta}_{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

# Regression "Fit"

How "well" does a regression perform?

- $R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$: fraction of variance explained by $X_i$ (and the fraction explained by $\varepsilon_i$).
- Alternative: mean squared error (MSE) $\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$.
  - This is of course what least-squares is actually minimizing!
- Alternative: root mean squared error (RMSE) $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$.
  - The average distance from a point to the line of best fit.
- Alternative: mean absolute deviation (MAD) $\frac{1}{N}\sum_{i=1}^{N}(|y_i - \hat{y}_i|)$.
  - The average residual.
- Alternative: median absolute error (MAE) $\text{median}(|y_i - \hat{y}_i|)$.
  - The median residual (insensitive to outliers).
- If you read enough econometrics papers, you will see enough of these.

# Thanks!