ECON 672

Week 3: Brief Regression Review

Samuel Rowe 12/15/2022

Overview

- Conditional Expectation Function
- Variance Operator
- Covariance Operator
- Correlation Operator
- Ordinary Least Squares
- Algebraic Properties of OLS
- Goodness of Fit

Overview

- Expected Value of OLS
- Bias of Expected Value of OLS
- Variance of OLS Estimators
- Robust Standard Errors
- Cluster Robust Standard Errors

Conditional Expectation Function

- You have two population variables x and y
  - We want to see how much y varies with a change in x
- Three Questions
  - 1) Is y affected another factor besides x
  - 2) What is the functional form connecting y and x?
  - 3) How do we disentangle causal effects from a correlation between x and y

Conditional Expectation Function

- Assume a population model
  - $y = \beta_0 + \beta_1 x + u$
  - We want to see the causal effects of x (RHS) on y (LHS)
- The population model allows for additional factors to influence y due to the error term u
  - β_0 is the coefficient of the intercept
  - β_1 is the slope parameter and coefficient of interest for causal effects
- We need **data** and **assumptions**

Conditional Expectation Function

- Normalization Assumption
  - $E(u) = 0$
  - Anything leftover gets put in β_0 (Wooldridge, 2009)
- Mean Independence Assumption
  - $E(u \mid x) = E(u) \ \forall \ x$
  - The expected value of the error term is the same across all slices of the population
- Zero Conditional Mean Assumption
  - Put assumptions 1 and 2 together
  - $E(u \mid x) = 0 \ \forall \ x$

Conditional Expectation Function

- Conditional Expectation Function (CEF) implies (Angrist & Pischke, 2009)
  - $E(y \mid x) = \beta_0 + \beta_1 x$
  - For a specific value, we write $E(y \mid X_i = x)$
- CEF is a population model and we rarely have a population in our data set

Variance Operator

- Variance Operator $V(\,.\,)$
    - The variance operator shows the variance of population
- Consider the variance of a random variable W
    - $V(W) = \sigma^2 = E\left[(W - E(W))^2\right]$
- When we have a sample
    - $S^2 = (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

Variance Operator

- Variance Operator Properties
- Variance of a line
  - $V(aX + b) = a^2V(X) + V(b) = a^2V(X)$
- Variance of a constant b
  - $V(b) = 0$
- Variance of the sum of two random variables
  - $V(X + Y) = V(X) + V(Y) + 2\big(E(XY) - E(X)E(Y)\big)$

Covariance Operator

- The Covariance Operator is represented by the last part of the variance of the sum of two random variables
  - $C(X, Y) = E(XY) - E(X)E(Y)$
- It represents the linear dependence between two random variables X and Y
  - If two random variables move in the same direction
    - $C(X, Y) > 0$
  - If two random variables move in the opposite direction
    - $C(X, Y) < 0$
  - If two random variables are independent
    - $C(X, Y) = 0$

Correlation Operator

- Covariance operator measures if two random variables move together
- Correlation operator measures the magnitude of the covariance oftwo random variables move together
  - How much do X and Y move together
  - $$Corr(X, Y) = \frac{C(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{(E[(X - E(X))^2])(E[(Y - E(Y))^2]}}$$
- The correlation coefficient is bounded between -1 and 1
  - The closer to -1 or 1 means a stronger correlation
  - A correlation near zero means no correlation

Ordinary Least Squares (OLS)

- OLS estimates of the parameters of interest from the population model by minimizing the sum of squared residuals (Wooldridge, 2009)
- If we have data on x and y and a population model of $E[y \mid x] = \beta_0 + \beta_1 x$
  - We can plug in x and y into our population model $y_i = \beta_0 + \beta_1 x_i + u_i$
  - We don't observe u_i even though we know it's there

Ordinary Least Squares (OLS)

- We have two assumptions after we plug in x and y (Wooldridge, 2009)
- $E[u] = 0$
  - Our expected value of residual is 0
  - $E[u|x] = 0$
- $C(x, u) = E[xu] = 0$
  - This means the error term is independent of x

Ordinary Least Squares (OLS)

- We plug in u to the population model (Wooldridge, 2009) and use first-order conditions
  - 1) $E[u] = E[y - \beta_0 - \beta_1 x] = 0$
  - 2) $C(x, u) = E[ux] = E\left[x(y - \beta_0 - \beta_1 x)\right] = 0$
- We will use these assumptions and sample averages to get $\hat{\beta}_0$ and $\hat{\beta}_1$
  - $\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$
  - $\frac{1}{n}\sum_{i=1}^{n} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$

Intercept Coefficient (OLS)

- We will use sample data since we do not have access to population data assumption 1
  - $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \frac{1}{n}\sum_{i=1}^{n}(y_i) - \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_0) - \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_1 x_i) = \ldots$
  - $\ldots = \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x} = 0$
    - Where $\bar{y}$ and $\bar{x}$ are sample averages
- The intercept coefficient
  - $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Slope Coefficient (OLS)

- We plug sample averages into assumption 2 to get our slope coefficient $\hat{\beta}_1$
- We'll drop $\dfrac{1}{n}$ since it doesn't affect the solution
- $\dfrac{1}{n}\sum_{i=1}^{n} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = \sum_{i=1}^{n} x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$
  - Plug in $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
  - $\sum_{i=1}^{n} x_i\left(y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i\right) = 0$

Slope Coefficient (OLS)

- $\sum_{i=1}^{n} x_i \left(y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i\right) = 0$
- $\sum_{i=1}^{n} x_i (y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n} x_i (x_i - \bar{x})$
- Given $\sum_{i=1}^{n} x_i (x_i - \bar{x}) = \sum_{i=1}^{n} (x_i - \bar{x})^2$ and $\sum_{i=1}^{n} x_i (y_i - \bar{y}) = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
- Assuming $\sum_{i=1}^{n} (x_i - \bar{x}) > 0$

Slope Coefficient (OLS)

- Our slope coefficient is
  - $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$
- This will pop up a lot for our estimators
  - $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \dfrac{Sample\ Covariance}{Sample\ Variance}$
- The Takeway
  - Variation in x helps us understand and identify its impact on y

Residuals

- We are able to get fitted values and residuals, $\hat{u}_i$, from our sample model
  - $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- $\hat{u}_i$ is a sample term and not a population one, since u is not observed
  - $\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$
- Sum of Squared Residuals
  - $\sum_{i=1}^{n} \hat{u}_i^{\,2} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$
  - OLS gets its name from the first order condition that minimizes sum of squared residuals

Algebraic Properties of OLS

- OLS residuals always sum to zero
  - $\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$
- Covariance of x and u always sums to zero and y-hat is a linear function of x
  - $\sum_{i=1}^{n} x_i \hat{u}_i = \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$ and $\sum_{i=1}^{n} \hat{y}_i \hat{u}_i = 0$
- The point $(\bar{x}, \bar{y})$
  - $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

Goodness of Fit

- For each observation, $y_i = \hat{y}_i + \hat{u}_i$
- Total Sum of Squares (SST)
    - $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$
- Residual Sum of Squares (SSR)
    - $SSR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{u}_i^2$
- Explained Sum of Squares (SSE)
    - $SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

Goodness of Fit

- With algebraic properties
  - $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}\left[\hat{u}_i - (\hat{y}_i - \bar{y})\right]^2$
  - $SST = SSE + SSR$
- Assuming the Total Sum of Squares is greater than 0
  - $R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$
  - How much variation in x explains the variation in y is calculated by R^2
  - R^2 is helpful for functional form, but **it won't tell us anything about bias of our causal estimate**

Expected Value of OLS

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimators for population model β_0 and β_1
  - How are $\hat{\beta}_0$ and $\hat{\beta}_1$ distributed?
  - $\hat{\beta}_0$ and $\hat{\beta}_1$ will differ across different sampling distributions
- $\hat{\beta}$ is unbiased under a set of assumption so that
  - $E[\hat{\beta}] = \beta$

Assumptions and Expected Value of OLS

- 1) Linear in Parameters Assumption
  - $y = \beta_0 + \beta_1 x + u$
- 2) Random Sampling Assumption
  - We have a random sample of size n with a set of numbers
    - $\{(x_i, y_i) : i = 1,2,...,n\}$
  - Where i is a random sampling draw from the population so that
    - $y_i = \beta_0 + \beta_1 x_i + u_i, \ i = 1,2,...,n$

Assumptions and Expected Value of OLS

- 3) Sample Variation in Explanatory Variables Assumption
    - The sample outcomes on x, $\{x, \ i = 1,2,...,n\}$ are not the same value
    - If there is no variation in x then this assumption fails
- 4) Zero Conditional Mean Assumption
    - $E[u \mid x] = 0$
    - The most crucial assumption, but the one most likely to fail
    - We can still calculate $\hat{\beta}$ if this assumption fails

Assumptions and Expected Value of OLS

- When all four assumptions hold
  - $E[\hat{\beta}] = \beta$
- If any of these assumptions fail, then $\hat{\beta}$ is no longer unbiased
- Stata Example

Omitted Variable Bias in OLS

- The 4th assumption is least likely to hold due to omitted variable bias (or confounders)
- Suppose a model: $Y = \alpha + \beta_1 X + \beta_2 U + \varepsilon$
  - U is unobserved
  - We estimate $Y = \alpha + \beta_1 X + \eta$
  - Where $\eta = \beta_2 U + \varepsilon$

Omitted Variable Bias in OLS

- We we estimate our model without unobserved U
- $\hat{\beta}_1 = \dfrac{C(Y,X)}{V(X)} = \beta_1 + \beta_2 \dfrac{C(U,X)}{V(X)} = \beta_1 + \dfrac{C(Y,U)}{V(U)} \dfrac{C(U,X)}{V(X)}$

Direction of Bias	*C(X,U)>0*	*C(X,U)<0*
C(Y,U)>0	Upward Bias	Downward Bias
C(Y,U)<0	Downward Bias	Upward Bias

Variance of the OLS Estimator

- The four assumptions of OLS estimator say nothing about the variance of the estimator
    - We need a measure of dispersion in the sampling distribution of the estimator
    - We need an estimator of the population variance
- We will add a fifth assumption to our OLS estimator
    - $V(u \mid x) = \sigma^2$

Variance of the OLS Estimator

- $V(u \mid x) = \sigma^2$
  - The variance of the population error term is constant (Homoskedasticity)
  - This assumption says that the variance of the population error term u is constant across any value of the explanatory variable
- With assumptions 4 and 5 in terms of conditional means and variance of y for population models
  - $E[y \mid x] = \beta_0 + \beta_1 x$
  - $V(y \mid x) = \sigma^2$

Variance of the OLS Esimator

- Sample Variance Under the 1st through 5th assumptions
  - $V(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \dfrac{\sigma^2}{SST_x}$
  - $V(\hat{\beta}_0) = \dfrac{\sigma^2 \frac{1}{n} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$
- Takeaways
  - As variation in x (SST_x) increases, the variance in $\hat{\beta}_1$ and $\hat{\beta}_0$ decreases
  - As σ^2 increases, the variance in $\hat{\beta}_1$ and $\hat{\beta}_0$ increases

Variance of the OLS Estimator

- Under assumptions 1-5
  - $E[\hat{\sigma}^2] = \sigma^2$
- Where the unbiased estimator of the population variance is
  - $\hat{\sigma}^2 = \dfrac{1}{(n-2)} \displaystyle\sum_{i=1}^{n} \hat{u}^2 = \dfrac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{(n-2)} = \dfrac{SSR}{(n-2)}$
  - Where $(n-2)$ is a degrees of freedom adjustment for two first order conditions
    - $\displaystyle\sum_{i=1}^{n} \hat{u}_i = 0$ and $\displaystyle\sum_{i=1}^{n} x_i \hat{u}_i = 0$

Variance of the OLS Estimator

- The estimator for the standard error for β is $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$
- $$se(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{SST_x}} = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$
- Stata or R will estimate the standard error

Robust Standard Errors

- What is the likelihood that the fifth assumption of homoskedasticity holds?
  - It is possible to assume that residuals are never homoskedastic
- When residuals are heteroskedastic (or not constant over x)
  - The standard errors of the estimator are biased
- Eicker, Huber, and White created a solution for a valid estimator of the variance of $\hat{\beta}$ called robust standard errors
  - Easy to implement in Stata (use the option robust after reg)
  - $V(\hat{\beta}) = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \hat{u}_i{}^2}{SST_x^2}$

Cluster Robust Standard Errors

- In addition to heteroskedastic residuals, we need to be concerned if the residuals are correlated within groups
- For example, residuals were correlated within schools for the Tennessee STAR experiment
    - We have a situation were we need to cluster the standard errors at the school level
- Easy to implement in Stata with the cluster(*group*) option