ECON 31380: Discussion Section 1

Steven Buschbach

University of Chicago

October 9, 2024

Table of Contents

Table of Contents

Introduction

- I am a fourth-year PhD student in the Economics Department. My research focus is in public economics, especially topics around health insurance.
- Each week, we will be either reviewing topics mentioned in lecture, discussing homework, and/or covering an extension of class material. If you have something you want covered, please let me know!
- Feel free to contact me at sbuschbach@uchicago.edu.

Table of Contents

Unbiasedness

- **Definition:**
  An estimator $\hat{\theta}_n$ of a parameter θ is **unbiased** if its expected value equals the true parameter value:
  $$\mathbb{E}[\hat{\theta}_n] = \theta$$
- **Intuition:**
  The estimator is, on average, correct. Unbiasedness means that in repeated samples, the estimator will be centered around the true value of the parameter. This is a **finite sample property**.
- Unbiasedness is a useful property that we generally want our estimators to have, but sometimes we are willing to trade off some bias for a large reduction in variance of the estimator.

Consistency

- **Definition:**
  An estimator $\hat{\theta}_n$ of parameter θ is **consistent** if it converges in probability to the true parameter θ as sample size goes to infinity:

$$\hat{\theta}_n \xrightarrow{p} \theta$$

  Recall the definition of convergence in probability. For any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}[|\hat{\theta}_n - \theta| > \epsilon] = 0$$

- **Intuition:**
  With a large enough sample, the estimator will give values very close to the true parameter. This is an **asymptotic property**.

- Sample means are consistent estimators for population means. What “law” gives us this? What regularity condition(s) do we need?

Comparing Unbiasedness and Consistency

- Unbiasedness does not imply consistency, consistency does not imply unbiasedness. Finite sample and asymptotic properties are different!
- Consider the problem of estimating the mean of $X \sim \mathcal{P}$, i.e., we want an estimator of $\theta = \mathbb{E}[X]$. We collect sample $\{X_i\}_{i=1}^n$. Are the following estimators unbiased? Consistent?

Comparing Unbiasedness and Consistency

- Unbiasedness does not imply consistency, consistency does not imply unbiasedness. Finite sample and asymptotic properties are different!
- Consider the problem of estimating the mean of $X \sim \mathcal{P}$, i.e., we want an estimator of $\theta = \mathbb{E}[X]$. We collect sample $\{X_i\}_{i=1}^n$. Are the following estimators unbiased? Consistent?
  - $\hat{\theta}_n = X_1$

Comparing Unbiasedness and Consistency

- Unbiasedness does not imply consistency, consistency does not imply unbiasedness. Finite sample and asymptotic properties are different!
- Consider the problem of estimating the mean of $X \sim \mathcal{P}$, i.e., we want an estimator of $\theta = \mathbb{E}[X]$. We collect sample $\{X_i\}_{i=1}^n$. Are the following estimators unbiased? Consistent?
  - $\hat{\theta}_n = X_1$
  - $\hat{\theta}_n = \frac{1}{n-2} \sum_{i=1}^n X_i$

Comparing Unbiasedness and Consistency

- Unbiasedness does not imply consistency, consistency does not imply unbiasedness. Finite sample and asymptotic properties are different!
- Consider the problem of estimating the mean of $X \sim \mathcal{P}$, i.e., we want an estimator of $\theta = \mathbb{E}[X]$. We collect sample $\{X_i\}_{i=1}^n$. Are the following estimators unbiased? Consistent?
  - $\hat{\theta}_n = X_1$
  - $\hat{\theta}_n = \frac{1}{n-2} \sum_{i=1}^n X_i$
  - $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i$

Comparing Unbiasedness and Consistency

- Unbiasedness does not imply consistency, consistency does not imply unbiasedness. Finite sample and asymptotic properties are different!
- Consider the problem of estimating the mean of $X \sim \mathcal{P}$, i.e., we want an estimator of $\theta = \mathbb{E}[X]$. We collect sample $\{X_i\}_{i=1}^n$. Are the following estimators unbiased? Consistent?
  - $\hat{\theta}_n = X_1$
  - $\hat{\theta}_n = \frac{1}{n-2}\sum_{i=1}^n X_i$
  - $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$
  - $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n (X_i - 5)$

Comparing Unbiasedness and Consistency

- Unbiasedness does not imply consistency, consistency does not imply unbiasedness. Finite sample and asymptotic properties are different!
- Consider the problem of estimating the mean of $X \sim \mathcal{P}$, i.e., we want an estimator of $\theta = \mathbb{E}[X]$. We collect sample $\{X_i\}_{i=1}^n$. Are the following estimators unbiased? Consistent?
  - $\hat{\theta}_n = X_1$
  - $\hat{\theta}_n = \frac{1}{n-2}\sum_{i=1}^n X_i$
  - $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$
  - $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n (X_i - 5)$
- Another example: the usual 2SLS estimator for IV is consistent, but biased in finite samples.

Comparing Unbiasedness and Consistency

- Unbiasedness does not imply consistency, consistency does not imply unbiasedness. Finite sample and asymptotic properties are different!
- Consider the problem of estimating the mean of $X \sim \mathcal{P}$, i.e., we want an estimator of $\theta = \mathbb{E}[X]$. We collect sample $\{X_i\}_{i=1}^n$. Are the following estimators unbiased? Consistent?
    - $\hat{\theta}_n = X_1$
    - $\hat{\theta}_n = \frac{1}{n-2} \sum_{i=1}^n X_i$
    - $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i$
    - $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n (X_i - 5)$
- Another example: the usual 2SLS estimator for IV is consistent, but biased in finite samples.
- Another example: the OLS estimator is unbiased and consistent iff $\mathbb{E}[u|x] = 0$ (among other regularity assumptions).

Asymptotic Unbiasedness

- **Definition:**
  An estimator $\hat{\theta}_n$ of θ is **asymptotically unbiased** if its bias disappears as the sample size grows:
  $$\lim_{n \to \infty} \mathbb{E}[\hat{\theta}_n] = \theta$$
- **Intuition:**
  For large n, the estimator behaves as if it were unbiased. That is, there may be bias in finite samples, but that bias tends to zero as n increases.
- If the variance of $\hat{\theta}_n$ tends to zero, then asymptotic unbiasedness implies consistency. (Try proving this with Chebyshev's inequality!).

Table of Contents

Overview of Frisch-Waugh-Lovell Theorem

FWL states that in a regression model with outcome y and multiple regressors X_1 and X_2 (each of which can be a matrix containing more than one covariate), we can obtain the coefficient on X_1 by:
(1) regressing X_1 on X_2,
(2) regressing y on X_2, and then
(3) running a regression with the residuals of the two previous regressions.

This first two steps is often referred to as **residualizing** or **partialing out** the effect of X_2.

This is an often-used result in econometrics for many reasons. Among other things, it can be used design estimators with better computational efficiency (e.g., in fixed effects models or more generally in panel data) by taking advantage of dimensionality reduction.

Frisch-Waugh-Lovell: Formal Statement

Our regression model is

$$y = X_1\beta_1 + X_2\beta_2 + u$$

We are interested in using data $\{(y_i, X_{1i}, X_{2i})\}_{i=1}^{n}$ to estimate β_1.

Frisch-Waugh-Lovell: Formal Statement

Our regression model is

$$y = X_1\beta_1 + X_2\beta_2 + u$$

We are interested in using data $\{(y_i, X_{1i}, X_{2i})\}_{i=1}^n$ to estimate β_1.

Let $\tilde{\beta}_1$ be the estimate of β_1 using OLS on the full model.
Let $\hat{\beta}_1$ be the estimate of β_1 from the following modified regression procedure:

1. Regress y on X_2, obtain residuals $M_{X_2}y$
2. Regress X_1 on X_2, obtain residuals $M_{X_2}X_1$
3. Regress $M_{X_2}y$ on $M_{X_2}X_1$ to get coefficient $\hat{\beta}_1$.

(Recall, M_X is the **residual matrix** or **annihilator matrix** that projects onto the orthogonal complement of X, i.e., $M_X = I - X(X'X)^{-1}X'$).

Frisch-Waugh-Lovell: Formal Statement

Our regression model is

$$y = X_1\beta_1 + X_2\beta_2 + u$$

We are interested in using data $\{(y_i, X_{1i}, X_{2i})\}_{i=1}^n$ to estimate β_1.

Let $\tilde{\beta}_1$ be the estimate of β_1 using OLS on the full model.
Let $\hat{\beta}_1$ be the estimate of β_1 from the following modified regression procedure:

1. Regress y on X_2, obtain residuals $M_{X_2}y$
2. Regress X_1 on X_2, obtain residuals $M_{X_2}X_1$
3. Regress $M_{X_2}y$ on $M_{X_2}X_1$ to get coefficient $\hat{\beta}_1$.

(Recall, M_X is the **residual matrix** or **annihilator matrix** that projects onto the orthogonal complement of X, i.e., $M_X = I - X(X'X)^{-1}X'$).

FWL states that $\tilde{\beta}_1 = \hat{\beta}_1$.

FWL: Proof

Simplest proof: Take full OLS model:

$$y = X_1\tilde{\beta}_1 + X_2\tilde{\beta}_2 + \tilde{u}$$

FWL: Proof

Simplest proof: Take full OLS model:

$$y = X_1\tilde{\beta}_1 + X_2\tilde{\beta}_2 + \tilde{u}$$

Multiply both sides by annihilator matrix of X_2:

$$M_{X_2}y = M_{X_2}X_1\tilde{\beta}_1 + M_{X_2}X_2\tilde{\beta}_2 + M_{X_2}\tilde{u}$$

FWL: Proof

Simplest proof: Take full OLS model:

$$y = X_1\tilde{\beta}_1 + X_2\tilde{\beta}_2 + \tilde{u}$$

Multiply both sides by annihilator matrix of X_2:

$$M_{X_2}y = M_{X_2}X_1\tilde{\beta}_1 + M_{X_2}X_2\tilde{\beta}_2 + M_{X_2}\tilde{u}$$

Note that because M_{X_2} projects onto a space orthogonal to X_2, we have $M_{X_2}X_2 = 0$. Also note that one of the basic properties of OLS gives us $\tilde{u}$ orthogonal to X_2, so $M_{X_2}\tilde{u} = \tilde{u}$.

FWL: Proof

Simplest proof: Take full OLS model:

$$y = X_1\tilde{\beta}_1 + X_2\tilde{\beta}_2 + \tilde{u}$$

Multiply both sides by annihilator matrix of X_2:

$$M_{X_2}y = M_{X_2}X_1\tilde{\beta}_1 + M_{X_2}X_2\tilde{\beta}_2 + M_{X_2}\tilde{u}$$

Note that because M_{X_2} projects onto a space orthogonal to X_2, we have $M_{X_2}X_2 = 0$. Also note that one of the basic properties of OLS gives us $\tilde{u}$ orthogonal to X_2, so $M_{X_2}\tilde{u} = \tilde{u}$. Thus, we can write the full OLS model equivalently as

$$M_{X_2}y = M_{X_2}X_1\tilde{\beta}_1 + \tilde{u}$$

which is how we obtained our $\hat{\beta}_1$ in the FWL procedure! Thus, $\tilde{\beta} = \hat{\beta}$.

FWL: Comments

- Basic intuition: you are removing the impact of X_2 on y, removing the impact of X_2 on X_1, and then using the results to isolate the partial effect of X_1 on y.
- In addition to $\tilde{\beta} = \hat{\beta}$, we the same residuals in both regressions: $\tilde{u} = \hat{u}$. Standard inference will be the same in each case, but be careful with DoF adjustments.
- In the bivariate case where we residualize out an intercept, this is equivalent to demeaning the outcome and predictor before regressing the two.

Table of Contents

What is Identification?

- **Definition:**
Identification refers to the ability to learn the true value of a parameter from the data. Intuitively, it answers: if we had infinite data, could we calculate the value of this parameter uniquely?

- **Point Identification:**
A parameter θ is **point identified** if there is exactly one value of θ that is consistent with the observed data. Slightly more formally, consider a parametric model with observed data $X \sim P_X$ that depends on unknown parameter $\theta \in \Theta$. In this case, we say that θ is point identified if

$$P_X(X|\theta_1) = P_X(X|\theta_2) \implies \theta_1 = \theta_2, \quad \forall \theta_1, \theta_2 \in \Theta$$

- You show identification by working with population parameters, not sample parameters!

Why Identification Matters

- Identification is a subtle topic, which we will not cover in detail in this course.
- The canonical reference here is from Arthur Lewbel in the JEL (2019): "The Identification Zoo: Meanings of Identification in Econometrics"
- Without identification, parameter estimates are meaningless because the data cannot uniquely reveal the true parameter values.
- **Point Identification vs. Set Identification:**
  - Point identification means a unique parameter value is found.
  - Set identification (i.e., partial identification) means parameters are constrained to a range.

Example: Non-Identification

Simultaneous Equations
In a system of supply and demand equations:

$$Q = \alpha_0 + \alpha_1 P + u_1 \quad \text{(Supply)}$$

$$Q = \beta_0 + \beta_1 P + u_2 \quad \text{(Demand)}$$

Intuitively, the observed equilibrium could be explained by multiple different values of (α_1, β_1). So these supply and demand parameters are **not identified** without additional information, such as an instrumental variable for one or the other.

Solve the simulatenous equation as check it out for yourself!