Multiple Regression and FWL

Gov 2001: Quantitative Social Science Methods I

Week 10, Lecture 19

Spring 2026

For Today

Required Reading

- Blackwell, Chapter 6 (multiple regression sections)
- Angrist & Pischke, §3.1.3 (pp. 34–37)

Today: Adding more variables and the Frisch-Waugh-Lovell theorem.

Roadmap

1. Why multiple regression?
2. The multiple regression model
3. Interpretation of coefficients
4. Frisch-Waugh-Lovell theorem
5. "Controlling for" and partialling out

Part I: Why Multiple Regression?

The Problem with Simple Regression

Simple regression: $Y = \alpha + \beta X + \varepsilon$

Issue: Other variables might affect Y and be correlated with X.

Example: Wages, education, and ability

- Y = log wage
- X_1 = years of education
- X_2 = ability (unobserved)

If ability affects both education and wages, simple regression of wage on education may be misleading.

The Idea of Multiple Regression

Solution: Include multiple explanatory variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

Benefits:

1. "Control for" confounding variables
2. Separate the effects of different variables
3. Better predictions (more information)

Goal: Estimate β_1—the effect of X_1 holding other variables constant.

Part II: The Multiple Regression Model

Multiple Regression Model

Population model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

More compactly:

$$Y_i = \mathbf{X}_i' \boldsymbol{\beta} + \varepsilon_i$$

where:

- $\mathbf{X}_i = (1, X_{1i}, X_{2i}, \ldots, X_{ki})'$ is a $(k + 1) \times 1$ vector
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ is the parameter vector

OLS with Multiple Regressors

Minimize:

$$SSR = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1i} - \cdots - b_k X_{ki})^2$$

Matrix form:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

where $\mathbf{X}$ is the $n \times (k+1)$ design matrix and $\mathbf{Y}$ is the $n \times 1$ outcome vector.

(We'll see more matrix notation in Week 11.)

The Two-Variable Case

Consider:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

OLS estimates:

$$\hat{\beta}_1 = \frac{\widehat{\text{Cov}}(X_1, Y) \cdot \widehat{\text{Var}}(X_2) - \widehat{\text{Cov}}(X_2, Y) \cdot \widehat{\text{Cov}}(X_1, X_2)}{\widehat{\text{Var}}(X_1) \cdot \widehat{\text{Var}}(X_2) - [\widehat{\text{Cov}}(X_1, X_2)]^2}$$

This formula is complex—there's an easier way to think about it.

Part III: Interpretation of Coefficients

Interpreting β_1

Model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

Interpretation of β_1:
The expected change in Y associated with a one-unit increase in X_1, **holding X_2 constant**.

Mathematically:

$$\beta_1 = \frac{\partial \mathbb{E}[Y|X_1, X_2]}{\partial X_1}$$

(When the CEF is linear.)

“Holding Constant” = Ceteris Paribus

Ceteris paribus: “All else equal”

β_1 answers: If we compare two observations with the same X_2 but X_1 differs by one unit, how much do we expect Y to differ?

Example:

$$\widehat{\text{Wage}} = 10 + 2 \cdot \text{Education} + 5 \cdot \text{Experience}$$

$\hat{\beta}_{\text{Education}} = 2$: Among workers with the **same experience**, each additional year of education is associated with \$2 higher wages.

When Simple and Multiple Regression Differ

Simple regression:

$$Y_i = \alpha + \gamma X_{1i} + u_i$$

Multiple regression:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

Question: When is $\hat{\gamma} \neq \hat{\beta}_1$?

Answer: When X_1 and X_2 are correlated!
If $\widehat{\text{Cov}}(X_1, X_2) = 0$, then $\hat{\gamma} = \hat{\beta}_1$.

Part IV: The FWL Theorem

Frisch-Waugh-Lovell (FWL) Theorem

FWL Theorem

Consider the regression:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

The OLS estimate $\hat{\beta}_1$ can be obtained by:

1. Regress X_1 on X_2 (and a constant), get residuals $\tilde{X}_1$
2. Regress Y on $\tilde{X}_1$ (no constant needed)

The slope from step 2 equals $\hat{\beta}_1$.

FWL: Intuition

Step 1: Regress X_1 on X_2

$$\tilde{X}_{1i} = X_{1i} - \hat{\delta}_0 - \hat{\delta}_1 X_{2i}$$

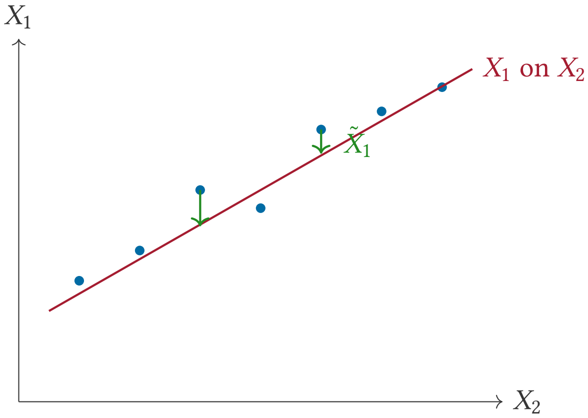
$\tilde{X}_1$ is the part of X_1 that is **not explained by** X_2.

Step 2: Regress Y on $\tilde{X}_1$

$$\hat{\beta}_1 = \frac{\sum_i \tilde{X}_{1i} Y_i}{\sum_i \tilde{X}_{1i}^2}$$

This uses only the "residual variation" in X_1—the part orthogonal to X_2.

FWL: Visual Intuition



Residuals $\tilde{X}_1$ are the vertical distances—variation in X_1 unexplained by X_2.

FWL: The Formula

$$\hat{\beta}_1 = \frac{\widehat{\text{Cov}}(\tilde{X}_1, Y)}{\widehat{\text{Var}}(\tilde{X}_1)}$$

where $\tilde{X}_1$ = residuals from regressing X_1 on X_2.

Key insight: $\hat{\beta}_1$ uses only the variation in X_1 that is **orthogonal to** X_2.

This is why we call it "controlling for" or "partialling out" X_2.

FWL Example: Wages

Goal: Effect of education on wages, controlling for experience.

Method 1: Multiple regression

$$\widehat{\text{Wage}}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{Educ}_i + \hat{\beta}_2 \text{Exper}_i$$

Method 2: FWL

1. Regress Education on Experience, get residuals $\widetilde{\text{Educ}}$
2. Regress Wage on $\widetilde{\text{Educ}}$

Both methods give the **same** $\hat{\beta}_1$!

Why FWL Matters

Conceptual clarity:

- "Controlling for X_2" means using only the variation in X_1 that is orthogonal to X_2
- Helps understand what multiple regression actually does

Practical uses:

- Visualizing partial effects
- Understanding fixed effects (panel data)
- Computing when matrix inversion is slow

Part V: Partialling Out

"Partialling Out" a Variable

To find the effect of X_1 controlling for X_2:

Full procedure:

1. Regress X_1 on X_2: get residuals $\tilde{X}_1$
2. Regress Y on X_2: get residuals $\tilde{Y}$
3. Regress $\tilde{Y}$ on $\tilde{X}_1$: the slope is $\hat{\beta}_1$

Intuition: Remove the influence of X_2 from both X_1 and Y, then see what relationship remains.

Added Variable Plot

An **added variable plot** (or partial regression plot):

- x-axis: $\tilde{X}_1$ (residuals from X_1 on X_2)
- y-axis: $\tilde{Y}$ (residuals from Y on X_2)

The slope of this scatterplot is $\hat{\beta}_1$.

Useful for:

- Visualizing the partial relationship between X_1 and Y
- Detecting influential observations
- Checking linearity after controlling for covariates

With Many Control Variables

Model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

FWL says $\hat{\beta}_1$ is obtained by:

1. Regress X_1 on $X_2, X_3, \ldots, X_k$, get residuals $\tilde{X}_1$
2. Regress Y on $\tilde{X}_1$

$\hat{\beta}_1$ uses only the variation in X_1 that is **orthogonal to all other regressors**.

Implications of FWL

1. Adding controls can change coefficients dramatically
If X_1 is highly correlated with X_2, there's little "residual variation" in X_1 left over.

2. Multicollinearity
If $X_1 \approx$ linear function of other X's, then $\tilde{X}_1 \approx 0$.
$\Rightarrow \widehat{\mathrm{Var}}(\tilde{X}_1) \approx 0 \Rightarrow$ huge standard errors.

3. "Bad controls" problem
If you control for a post-treatment variable, you might remove the very variation you wanted.

Part VI: Properties of Multiple Regression

Properties Extend to Multiple Regression

All the key properties from simple regression extend:

Residual properties:

- $\sum_i \hat{e}_i = 0$
- $\sum_i X_{ji}\hat{e}_i = 0$ for each regressor j

Unbiasedness:
If $\mathbb{E}[\varepsilon|X_1, \ldots, X_k] = 0$, then $\mathbb{E}[\hat{\beta}_j] = \beta_j$ for all j.

Gauss-Markov:
Under homoskedasticity and no autocorrelation, OLS is BLUE.

R^2 in Multiple Regression

$$R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST}$$

Important: R^2 can only **increase** when you add variables.
(Even if the new variable is pure noise!)

Adjusted R^2:

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)}$$

Penalizes for adding variables; can decrease.

Summary

Key ideas:

1. Multiple regression includes several explanatory variables

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \varepsilon$$

2. β_1 = effect of X_1 holding other variables constant
3. FWL theorem: $\hat{\beta}_1$ uses only the variation in X_1 orthogonal to other X's
4. "Controlling for" = partialling out
5. Properties (unbiasedness, BLUE) extend from simple regression

Looking Ahead

Next lecture: Omitted Variable Bias

- What if we *don't* control for an important variable?
- The OVB formula
- When does omitting a variable bias our estimate?
- "Bad controls" and post-treatment variables

The coefficient β_1 measures the relationship between X_1 and Y, holding other variables constant.

FWL: This uses only the variation in X_1 that is orthogonal to the other regressors.

"Controlling for" means partialling out.