Lecture 4 - Frisch-Waugh, Omitted Variables

Alice Wu

UW Madison Econ 695

This Lecture

- Frisch-Waugh Theorem
- Omitted Variable Bias
- Motivation for Instrument Variables: Returns to Education

Motivation: Returns to Schooling

- There is a substantial literature in labor economics that aims to causally estimate the returns to education (1 more year in school, college degree, postgrad degree, online education...)
- Can we estimate the returns to education in an OLS regression (e.g., a Mincer regression)? The estimate is likely to be biased even with a rich set of controls.
    - there is always something unobserved that endogenously influences the education choice (e.g., a person's ability)
    - individuals with higher returns to education are also selected into higher education (selection bias)
- Today we will focus on omitted variable bias, and discuss how it affects the OLS estimates for returns to education. In the next lecture, we will learn how instrument variables can help us address omitted var bias, selection bias, and other endogeneity concerns and obtain causal estimates of the returns to education.

Frisch-Waugh Theorem

- We posit a population regression:

$$y_i = x_i'\beta + u_i \tag{1}$$

$$x_i := \begin{bmatrix} x_{1i} \\ \dots \\ x_{Ki} \end{bmatrix}$$

where $x_{1i} = 1$.

- What is the population coefficient β? Recall

$$\beta = argmin_b E[(y_i - x_i'b)^2] = E[x_i x_i']^{-1} E[x_i y_i]$$

$$\text{if } x_i' = [1, x_{2i}]$$

$$\beta_2 = \frac{cov(y_i, x_{2i})}{var(x_{2i})}$$

Frisch-Waugh Theorem

- What is the coefficient β_k on x_{ki} in multivariate regressions? Turns out we also have a closed-form formula via the Frisch-Waugh Theorem (FW):

$$\beta_k = \frac{cov(y_i, \xi_i)}{var(\xi_i)}$$

where $\xi_i = x_{ki} - E^*[x_{ki}|x_{(-k)i}]$ is the residual over the projection of x_{ki} on *all* other covariates. β_k = coefficient from univariate regression of y_i on x_{ki}, after "partialling out" other $x's$.

Proof of FW

Proof: $x_i' = (x_{1i}, x_{2i} ... x_{ki} ... x_{Ki})$ has K elements.
Let $x_{(\sim k)i}$ be x_i *after removing row* k.
Now write the "auxiliary" regression of x_{ki} on $x_{(\sim k)i}$:

$$x_{ki} = \underbrace{x_{(\sim k)i}'\pi}_{E^*[x_{ki}|x_{(-k)i}]} + \xi_i \tag{2}$$

As usual, the FOC for π require

$$E[x_{(\sim k)i}\xi_i] = 0 \tag{3}$$

Finally, since $y_i = x_i'\beta^* + u_i$ we can write:

$$\begin{aligned} E[\xi_i y_i] &= E[\xi_i(\beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + ... + \beta_K x_{Ki} + u_i)] \\ &= \beta_1 E[\xi_i x_{1i}] + \beta_2 E[\xi_i x_{2i}] + ... + \beta_k E[\xi_i x_{ki}] + ... + \beta_K E[\xi_i x_{Ki}] \\ &\quad + E[\xi_i u_i] \end{aligned}$$

Now notice that from the FOC for π, $E[\xi_i x_{mi}] = 0$ unless $m = k$.

$$E[\xi_i y_i] = \beta_k E[\xi_i x_{ki}] + \underbrace{E[\xi_i u_i]}_{=0} = \beta_k E[\xi_i x_{ki}] \tag{4}$$

$E[\xi_i u_i] = E[(x_{ki} - x_{(\sim k)i}'\pi)u_i] = 0$ because u_i is orthogonal to all the $x's$.
Finally: $E[\xi_i x_{ki}] = E[\xi_i(x_{(\sim k)i}'\pi + \xi_i)] = E[\xi_i^2]$ using the FOC for π

FW in Univariate Case

One extremely useful version of FW: Suppose we have a constant and one other x variable: $x_i' = (1, x_{2i})$. Consider the population regression:

$$y_i = \beta_1 + \beta_2 x_{2i} + u_i$$

in lecture 1 we showed:

$$\begin{aligned} \beta_2 &= E[(x_i - E[x_i])^2]^{-1} E[(x_i - E[x_i])y_i] \\ &= Var[x_i]^{-1} Cov[x_i, y_i] \end{aligned}$$

From FW, we can get β_2 from a '2 step' approach: first regress x_{2i} on the other regressor (i.e., 1), then regress y_i on the residual from the first regression. But what is the auxiliary regression of x_{2i} on a constant? This is:

$$x_{2i} = \pi \times 1 + \xi_i$$

And $\pi = E[x_{i2}]$ is the solution. So in this case, $\xi_i = x_{i2} - E[x_{i2}]$.

FW extension

- In fact, there is a slightly more general version of FW. Suppose we are interested in a subset of regressors, e.g., (x_{1i}, x_{2i}). Then the coefficients (β_1, β_2) can be expressed as the outcome of a two-step process: first consider the population regression of (x_{1i}, x_{2i}) on all the other regressors, then consider the population regression of y_i on the pair of residuals.
- A version of this result: suppose that $x_i' = (1, x_{2i}, x_{3i}, ...x_{Ki})$. Then we can get the coefficients on the non-constant regressors by considering the population regression of y on the set of variables $(x_{2i} - E[x_{2i}], x_{3i} - E[x_{3i}]...)$, residualized over the constant. And they will have:

$$\begin{pmatrix} \beta_2 \\ \beta_3 \\ ... \\ \beta_K \end{pmatrix} = Var[x_{2i}, x_{3i}, ...x_{Ki}]^{-1} Cov[(x_{2i}, x_{3i}, ...x_{Ki})', y_i]$$

- People often express the pop. regression $y_i = x_i'\beta + u_i$ in terms of variances and covariances, but this is a little sloppy unless y_i and *all the elements* of x_i have mean 0. In that case, you can write:

$$\begin{aligned} \beta &= E[x_i x_i']^{-1} E[x_i y_i] \\ &= Var[x_i]^{-1} Cov[x_i, y_i] \end{aligned}$$

FW for OLS

Now let's move from the population regression to the OLS regression. Recall the objective is

$$min_{\ b} \quad \sum_{i=1}^{N}(y_i - x_i'b)^2$$

The FOC is:

$$\begin{aligned}
\sum_{i=1}^{N} x_i(y_i - x_i'\hat{\beta}) &= 0 \qquad \Rightarrow \frac{1}{N}\sum_{i=1}^{N} x_i(y_i - x_i'\hat{\beta}) \\
\Rightarrow \frac{1}{N}\sum_{i=1}^{N} x_i y_i &= \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)\hat{\beta} \\
\Rightarrow \hat{\beta} &= \left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\frac{1}{N}\sum_{i=1}^{N} x_i y_i
\end{aligned}$$

c/w population regression:

$$\beta = E[x_i x_i']^{-1}E[x_i y_i]$$

FW for OLS

So we are "matching moments":

- We replace $E[x_i x_i']$ with $S_{xx} = \frac{1}{N}\sum_{i=1}^{N} x_i x_i'$.
- We replace $E[x_i y_i]$ with $S_{xy} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i$.
- Computer programs compute S_{xx}, S_{xy} and invert S_{xx} very efficiently.

Frisch-Waugh (FW) for OLS: The k^{th} row of $\hat{\beta}$ is:

$$\hat{\beta}_k = E[\hat{\xi}_i^2]^{-1} E[\hat{\xi}_i y_i]$$

where $\hat{\xi}_i$ is the *estimated residual* from an OLS regression of x_{ki} on all the other $x's$:

$$x_{ki} = x'_{(\sim k)i}\hat{\pi} + \hat{\xi}_i$$

Prove FW for OLS:

1. OLS: get $\hat{\beta}$, define $\hat{u}_i = y_i - x_i'\hat{\beta}$. We know $\frac{1}{N}\sum_{i=1}^{N} x_i \hat{u}_i = 0$ (Why?)
2. OLS for auxiliary model: $\hat{\xi}_i = x_{ki} - x'_{(\sim k)i}\hat{\pi}$. We know $\frac{1}{N}\sum_{i=1}^{N} x_{(\sim k)i}\hat{\xi}_i = 0$ (Why?)

- Write $y_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + ... + \hat{\beta}_k x_{ki} + ... + \hat{\beta}_K x_{Ki} + \hat{u}_i$. Now form

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i y_i = \frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i(\hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + ... + \hat{\beta}_k x_{ki} + ... + \hat{\beta}_K x_{Ki} + \hat{u}_i)$$

What terms are equal to 0 from the 2 FOC?

Example: Mincer Equation

- A classic benchmark for estimating wage determination equations is called the Mincer (earnings) equation, proposed by Mincer in his dissertation in 1958 and book in 1974.
- Consider a population regression of log earnings (as a measure of human capital) on years of education S and years of labor market experience $X = age - S - 6$,

$$log(y) = \beta_0 + rS + \beta_1 X + \beta_2 X^2 + u$$

in which r represents the return to education. How would you interpret r using the Frisch-Waugh Theorem?

$$r = \frac{cov(log(y), \widetilde{S})}{var(\widetilde{S})}$$

$$\widetilde{S} = S - E^*[S|1, X, X^2]$$

  - How would you interpret β_1?

Example: Mincer Equation

- Why log earnings?

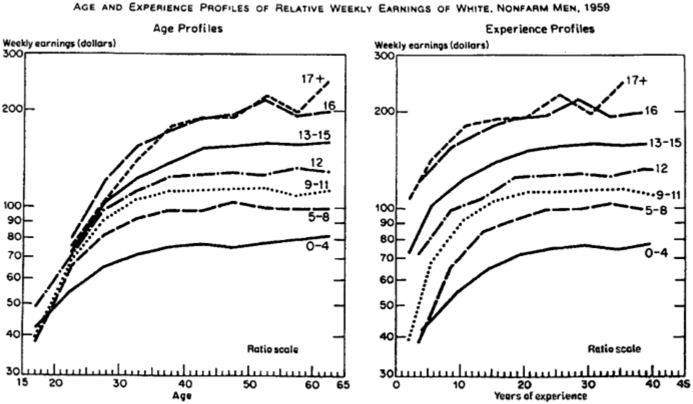
  $$y \propto e^{rS} \times e^{\beta_1 X + \beta_2 X^2}$$

  Mincer (1958) points out education should have a **multiplicative** effect on human capital.
- $\beta_2 < 0$ captures the slower earnings growth among older workers (concave age-earning profiles). But more recent works suggest Mincer equation understates wage growth for younger and predicts a spurious decline in wages among the older (instead, one can estimate the regression separately by age group; see Lemieux 2003 for details.)

- Mincer equation is derived from a formal model of investment in human capital: workers of identical abilities choose the level of schooling that maximizes the net present value of their lifetime earnings. It is simple to estimate, and it fits the data remarkably well.

Example: Mincer Equation

Figure 1

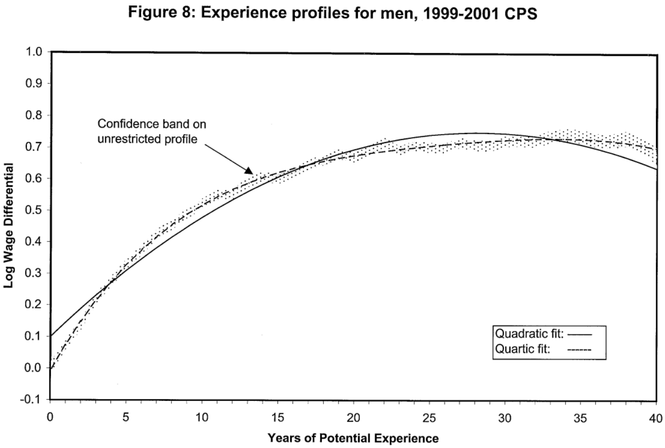


NOTE: Figures on curves indicate years of schooling completed.
SOURCE: 1/1,000 sample of U.S. Census, 1960.

Source: Jacob Mincer, *Schooling, Experience, and Earnings*

Example: Mincer Equation

Figure: Figure 8 of Lemieux (2003)



FW & Omitted Variable Formula

- FW: if you add a regressor x_{ki}, the coefficient is "as if" you added only the part of that regressor that is *unexplained by the other regressors:*

$$x_{ki} - x'_{(\sim k)i}\pi$$

- What about the opposite direction? What happens if you forget a regressor?

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

  Suppose we don't include x_{3i}?
  - auxiliary regression: $x_{3i} = \pi_1 x_{1i} + \pi_2 x_{2i} + \xi_i$, where $E[\xi_i x_{1i}] = E[\xi_i x_{2i}] = 0$
  - Plug the auxiliary regression into the population regression:

$$\begin{aligned} y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 \left(\pi_1 x_{1i} + \pi_2 x_{2i} + \xi_i\right) + u_i \\ &= \left(\beta_1 + \beta_3 \pi_1\right) x_{1i} + \left(\beta_2 + \beta_3 \pi_2\right) + \left(\beta_3 \xi_i + u_i\right) \end{aligned}$$

    importantly, the new residual $\beta_3 \xi_i + u_i$ is orthogonal to (x_{1i}, x_{2i}). What does that imply?

FW & Omitted Variable Formula

- $(\beta_1^o, \beta_2^o) := (\beta_1 + \beta_3\pi_1, \beta_2 + \beta_3\pi_2)$ satisfy FOC for population regression of y_i on (x_{1i}, x_{2i})!
- Conclusion: If

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

  and we don't include x_{3i}, the coefficient on x_{2i} is:

$$\beta_2^o = \beta_2 + \beta_3\pi_2$$

  where π_2 is the coefficient on x_{2i} from the regression of the omitted variable x_{3i} on the remaining $x's$:

$$x_{3i} = \pi_1 x_{1i} + \pi_2 x_{2i} + \xi_i$$

- Intuition: you forgot x_{3i} so the house elf is doing the best he can to predict y given what he has to work with. The best he can do is use the other $x's$ to predict x_{3i}.

OLS version of the Omitted Variable Formula

OLS (sample) version

- Aux. model for the omitted variable: $x_{ji} = \hat{\pi}_1 x_{1i} + \hat{\pi}_2 x_{2i} + \hat{\xi}_i$
- Then:

$$\begin{aligned} y_i &= \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 \left(\hat{\pi}_1 x_{1i} + \hat{\pi}_2 x_{2i} + \hat{\xi}_i \right) + \hat{u}_i \\ &= \underbrace{\left(\hat{\beta}_1 + \hat{\beta}_3 \hat{\pi}_1 \right)} x_{1i} + \underbrace{\left(\hat{\beta}_2 + \hat{\beta}_3 \hat{\pi}_2 \right)} x_{2i} + \underbrace{\left(\hat{\beta}_3 \hat{\xi}_i + \hat{u}_i \right)} \\ &= \hat{\beta}_1^0 x_{1i} + \hat{\beta}_2^0 x_{2i} + \hat{\eta}_i \end{aligned}$$

- Notice $\frac{1}{N} \sum_{i=1}^N (x_{1i}, x_{2i})' \hat{\eta}_i = \frac{1}{N} \sum_{i=1}^N (x_{1i}, x_{2i})' (\hat{\beta}_3 \hat{\xi}_i + \hat{u}_i) = (0, 0)'$. So $(\hat{\beta}_1^0, \hat{\beta}_2^0)$ satisfy FOC for OLS regression of y_i on (x_{1i}, x_{2i})

Returns to Education

- The real importance of the OVF is that we can often think about how the omission of a variable affects the estimated coefficient of variables we include, even if we can't estimate the auxiliary regression.
- Economists have long been interested in the returns to education. One may posit a population regression as follows:

$$log(wage)_i = \beta_1 + \beta_2 Educ_i + \beta_3 Ability_i + u_i$$
$$E[u_i] = E[u_i Educ_i] = E[u_i Ability_i] = 0$$

where the outcome is log wage of a worker (at age 30), Education could be measured by years of schooling or dummies for degree levels, and Ability represents a measure of ability that affects a person's labor market outcome.

  - The coefficient of interest is β_2.
  - Ability is rarely observable, but it is likely to be correlated with education: higher-ability individuals are more likely to go to college/pursue an advanced degree.

Returns to Education

- Consider an auxiliary regression (that we may not estimate):

$$Ability_i = \pi_1 + \pi_2 Educ_i + \xi_i,\ E[\xi_i] = E[\xi_i Educ_i] = 0$$

  - How would you interpret π_2? Hint: $cov(Ability_i, Educ_i)$
- Suppose we have to omit Ability, what would be return to education?

$$\beta_2^o = \beta_2 + \beta_3\pi_2 \neq \beta_2 \text{ unless } \beta_3\pi_2 = 0$$

  This holds both for the population regression and for the OLS estimates.
- Many people (especially those with high education) think that $\beta_3 > 0$ and $\pi_2 > 0$, which leads them to believe that a model that does not control for ability **"overstates"** the effect of education.
- Rethink the Mincer regression: estimating an OLS $log(y) = b_0 + rS + b_1X + b_2X^2 + u$ cannot give us a causal estimate for the return to education. However, the underlying model remains a useful conceptual framework: workers choose education that maximizes the NPV of lifetime earnings $\Rightarrow y = f(s) \times h(x)$, the product of a function of education s and a function of experience x.

Causal Returns to Education

- So how can we get causal estimates of the returns to an additional year of schooling?
- In an ideal experiment, we will randomize how many years a person stays in school. It cannot get any IRB approval.
- Instead, researchers rely on natural experiments that change $Educ_i$ exogenously (unrelated to a person's underlying ability):
  - distance to college
  - lottery for scholarship...
  - changes to compulsory schooling laws
- What do these variables (z's) have in common?

$$cov(z_i, Educ_i) \neq 0$$
$$cov(z_i, Ability_i) = cov(z_i, u_i) = 0$$

We will discuss instrument variables formally in the next lecture.