Lecture 2: Linear Regression (Part I)

Jaakko Meriläinen

5304 Econometrics @ Stockholm School of Economics

Warm-Up: Female Leaders and Policy

- Suppose that we want to understand what is the impact of having a female leader on policy
  - Relative to their share in population, women are under-represented in all political positions
  - There is evidence that women and men have different policy preferences
  - Does this also imply that female representation has an impact on policy decisions?

- Important question, but hard to study: female leaders are not allocated randomly!
  - For example, cross-sectional comparisons are difficult to interpret
  - The fact that women are better represented in a particular country or locality may reflect the political preferences of the group that elects them
  - The correlation between policy outcomes and women's participation then may not imply a causal effect of women's participation

Chattopadhyay and Duflo (2004)

- Since the mid-1990s, one third of Village Council (*Gram Panchayat*) head positions in India have been randomly reserved for a woman
- Indian Village Councils are responsible for the provision of many local public goods in rural areas
- Do reservations affect the provision of these public goods?
- Yes—leaders invest more in infrastructure that is directly relevant to the needs of their own genders

Female Representation in Gram Panchayats

TABLE I

FRACTION OF WOMEN AMONG PRADHANS IN RESERVED AND UNRESERVED GP

	Reserved GP (1)	Unreserved GP (2)
West Bengal		
Total Number	54	107
Proportion of Female Pradhans	100%	6.5%
Rajasthan		
Total Number	40	60
Proportion of Female Pradhans	100%	1.7%

Differences in Women and Men's Policy Preferences

TABLE IV

ISSUES RAISED BY WOMEN AND MEN IN THE LAST 6 MONTH

	West Bengal						Rajasthan					
	Women			Men	Average	Difference	Women			Men	Average	Difference
	Reserved	Unreserved	All				Reserved	Unreserved	All			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Other Programs												
Public Works	.84	.84	.84	.85	.84	−.01	.60	.64	.62	.87	.74	−.26
Welfare Programs	.12	.09	.10	.04	.07	.06	.25	.14	.19	.03	.04	.16
Child Care	.00	.02	.01	.01	.01	.00	.04	.09	.07	.01	.02	.06
Health	.03	.04	.04	.02	.03	.02	.06	.08	.07	.04	.03	.03
Credit or Employment	.01	.01	.01	.09	.05	−.08	.06	.06	.05	.04	.09	.01
Total Number of Issues	153	246	399	195			72	88	160	155		
Breakdown of Public Works Issues												
Drinking Water	.30	.31	.31	.17	.24	.13	.63	.48	.54	.43	.49	.09
Road Improvement	.30	.32	.31	.25	.28	.06	.09	.14	.13	.23	.18	−.11
Housing	.10	.11	.11	.05	.08	.05	.02	.04	.03	.04	.04	−.01
Electricity	.11	.07	.08	.10	.09	−.01	.02	.04	.03	.02	.02	.01
Irrigation and Ponds	.02	.04	.04	.20	.12	−.17	.02	.02	.02	.04	.03	−.02
Education	.07	.05	.06	.12	.09	−.06	.02	.07	.05	.13	.09	−.09
Adult Education	.01	.00	.00	.01	.00	.00	0	0	.00	.00	.00	.00
Other	.09	.11	.10	.09	.09	.01	.19	.21	.20	.12	.28	.05
Number of Public Works Issues	128	206	334	166			43	56	99	135		
Public Works												
Chi-square	8.84		71.72				7.48		16.38			
p-value	.64		.00				.68		.09			

Notes: 1. Each cell lists the number of times an issue was mentioned, divided by the total number of issues in each panel. 2. The data for men in West Bengal comes from a subsample of 48 villages. 3. Chi-square values placed across two columns test the hypothesis that issues come from the same distribution in the two columns.

Outcomes in Reserved and Non-Reserved Villages

TABLE V

EFFECT OF WOMEN'S RESERVATION ON PUBLIC GOODS INVESTMENTS

	West Bengal			Rajasthan		
	Mean, Reserved GP	Mean, Unreserved GP	Difference	Mean, Reserved GP	Mean, Unreserved GP	Difference
Dependent Variables	(1)	(2)	(3)	(4)	(5)	(6)
A. *Village Level*						
Number of Drinking Water Facilities	23.83	14.74	9.09	7.31	4.69	2.62
Newly Built or Repaired	(5.00)	(1.44)	(4.02)	(.93)	(.44)	(.95)
Condition of Roads (1 if in good	.41	.23	.18	.90	.98	−.08
condition)	(.05)	(.03)	(.06)	(.05)	(.02)	(.04)
Number of Panchayat Run	.06	.12	−.06			
Education Centers	(.02)	(.03)	(.04)			
Number of Irrigation Facilities	3.01	3.39	−.38	.88	.90	−.02
Newly Built or Repaired	(.79)	(.8)	(1.26)	(.05)	(.04)	(.06)
Other Public Goods (ponds, biogas,	1.66	1.34	.32	.19	.14	.05
sanitation, community buildings)	(.49)	(.23)	(.48)	(.07)	(.06)	(.09)
Test Statistics: Difference Jointly Significant			4.15			2.88
(*p*-value)			(.001)			(.02)
B. *GP Level*						
1 if a New Tubewell Was Built	1.00	.93	.07			
		(.02)	(.03)			
1 if a Metal Road Was Built or Repaired	.67	.48	.19			
	(.06)	(.05)	(.08)			
1 if There Is an Informal Education	.67	.82	−.16			
Center in the GP	(.06)	(.04)	(.07)			
1 if at Least One Irrigation Pump Was Built	.17	.09	.07			
	(.05)	(.03)	(.05)			
Test Statistics: Difference Jointly Significant			4.73			
(*p*-value)			(.001)			

Notes: 1. Standard errors in parentheses. 2. In West Bengal, there are 322 observations in the village level regressions, and 161 in the GP level regressions. There are 100 observations in the Rajasthan regressions. 3. Standard errors are corrected for clustering at the GP level in the village level regressions, using the Moulton (1986) formula, for the West Bengal regressions.

Recap

- In the last lecture, we discussed the goal of obtaining causal estimates
- We focused on the issue of identification using the potential outcomes framework
  - How do we solve the selection problem?
  - How a randomized trial helps in this issue
- But experiments are rare in economics (although less so)
  - Not everything *can* be consciously manipulated in controlled settings
  - Not everything *should* be consciously manipulated in controlled settings
- What do we do then?

Next Steps

- The rest of this course will focus on settings where we do not have randomized allocation but still want causal estimates
  - The **bulk** of applied micro literature
  - Experiments might be the ideal, but they are not the norm

- We will focus on various techniques:
  - Regression analysis
  - Instrumental variables
  - Panel models (especially difference-in-differences)
  - Regression discontinuity designs

- We will now talk about linear regression—a building block for (almost) all that follows

This Lecture

1. Warm-up, recap, and overview
2. Introducing linear regression
3. Deriving OLS estimators
   - Bivariate linear regression
   - Finding $\hat{\beta}$
   - Multivariate linear regression
4. Multivariate linear regression: Assumptions and unbiasedness

Specifying a Model

- Our goal is to study the relationship between some outcome variable y and another variable x
- To do this, we need to confront three issues:
  - How do we allow for other factors to affect y?
  - What is the functional form of the relationship?
  - The identification problem: how can we be sure that the relationship is a *ceteris paribus* relationship between y and x?
- We will be discussing these issues throughout the course

Terminology

- You have all seen linear regression before
- You will hear these variables being referred to in different ways:

y	x
Dependent variable	Independent variable
Explained variable	Explanatory variable
Response variable	Control variable
Predicted variable	Predictor variable
Regressand	Regressor
LHS variable	RHS variable

Let Us Specify a Model!

- Specify the following model:

$$y = \underbrace{\beta_0}_{intercept} + \underbrace{\beta_1}_{slope} x + \underbrace{u}_{error}$$

- u, the error term, represents factors other than x to affect y

- A change in x is assumed to have a linear effect on y when $\triangle u = 0$, and the slope parameter β_1 captures the effect on y

- Whether this is a ceteris paribus (causal) effect will depend on the relationship between x and u **in the population**—we will discuss this a lot during this course

Setting E(u) = 0

- First, we make a simplifying assumption (with no loss of generality): the expected value of u is zero in the population

$$E(u) = 0$$

- We can do this because the presence of β_0 allows us to adjust the intercept, leaving the slope the same

- If, say, $E(u) = \alpha_0$, we can always write out the specification as:

$$y = (\beta_0 + \alpha_0) + \beta_1 x + (u - \alpha_0)$$

- Now the error term has an expectation of zero, but the important part is that the slope, β_1, has not changed

The Zero Conditional Mean (ZCM) Assumption

- A much more stringent (but crucial) assumption is that the average value of u does not depend on the value of x
- This can be written as:

$$E(u|x) = E(u) = 0$$

- The first equality is the crucial assumption which means that the expected value of u does not vary with x

Returns to Education and the ZCM Assumption

- A classic question in labor economics is on the effect of education on wages
- This could be modeled as

$$wage = \beta_0 + \beta_1 x + u$$

- Here x is years of schooling, and u contains all other factors that affect wage
- For simplicity, assume the only other factor is (unobserved) ability
- $E(u|x) = 0$ means that the average ability in those with, say, 8 years of education was the same as those who had finished university
- But is this plausible...?

How Do We Interpret β_1?

- Take the expectation of y conditional on x and use $E(u|x) = 0$ to get:

$$E(y|x) = E(\beta_0 + \beta_1 x + u|x)$$
$$E(y|x) = \beta_0 + \beta_1 x$$

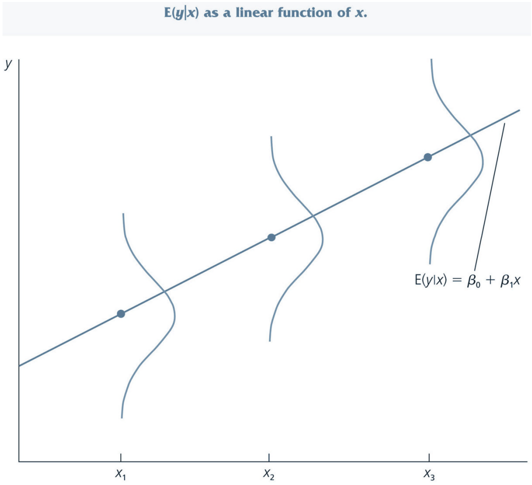
- This is the "population regression function" (PRF), which is a linear function of x

- β_1 expresses the change in expected value of y if x increases one unit

- We can use the PRF to split y up into systematic and unsystematic part:

$$\begin{aligned} y &= E(y|x) + u \\ &= \beta_0 + \beta_1 x + u \end{aligned}$$

Population Regression Function (Wooldridge 2013)



OLS in the Bivariate Case

- We will estimate β_0 and β_1, using a random sample of size n from the population
- So, we write (for each $i = 1, \ldots, n$):

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- The question now is: how do we use the data to obtain estimates of the slope and intercept parameters?
- We will derive an **estimator** for each parameter
- This is "a rule for combining data to produce a numerical value for a population parameter"

OLS

- We will use the assumptions that we have made to motivate OLS:

$$E(u) = 0$$
$$Cov(x, u) = E[(x - E(x))(u - E(u))] = E[xu - E(x)u] = E(xu) - E(x)E(u) = E(xu) = 0$$

- The latter equation is implied by mean independence

- Rewrite u in terms of observable variables x and y and unknown parameters:

$$E(y - \beta_0 - \beta_1 x) = 0$$
$$E[x(y - \beta_0 - \beta_1 x)] = 0$$

Finding $\widehat{\beta}$

- Let us use the **method of moments** approach—replace the population moments by their sample analogues

- Choose $\widehat{\beta}_0$ and $\widehat{\beta}_1$ to solve the sample analogues:

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0$$
$$\frac{1}{n}\sum_{i=1}^{n}x_i\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0$$

Finding $\widehat{\beta}_0$

- From the first equation:

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0$$
$$\Leftrightarrow \overline{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{x}$$
$$\Leftrightarrow \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

- $\widehat{\beta}_0$ is expressed as a function of $\widehat{\beta}_1$, so now we just need to find the latter

Finding $\widehat{\beta}_1$

- Recall that $\frac{1}{n}\sum_{i=1}^{n} x_i \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right) = 0$ and plug in our expression for $\widehat{\beta}_0$

$$\begin{aligned}
&\sum_{i=1}^{n} x_i \left[y_i - \left(\overline{y} - \widehat{\beta}_1 \overline{x}\right) - \widehat{\beta}_1 x_i\right] = 0 \\
\Leftrightarrow &\sum_{i=1}^{n} x_i \left(y_i - \overline{y}\right) = \widehat{\beta}_1 \sum_{i=1}^{n} x_i \left(x_i - \overline{x}\right) \\
\Leftrightarrow &\sum_{i=1}^{n} \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right) = \widehat{\beta}_1 \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2
\end{aligned}$$

- The last line follows from the properties of the sum operator:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i(y_i - \bar{y}) - \bar{x}\sum_{i=1}^{n}(y_i - \bar{y}) = \sum_{i=1}^{n} x_i(y_i - \bar{y}) - \bar{x}(n\bar{y} - n\bar{y}) = \sum_{i=1}^{n} x_i(y_i - \bar{y})$$

Finding $\widehat{\beta}_1$

- Hence, if

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 > 0,$$

the estimated slope parameter $\widehat{\beta}_1$ equals

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

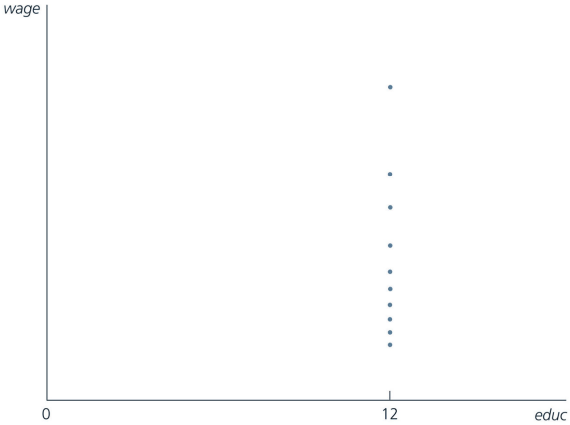
Finding $\widehat{\beta}_1$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{Cov(x,y)}{Var(x)}$$

- So $\widehat{\beta}_1$ is the sample covariance between x and y divided by the sample variance of x
- The sample covariance of x and y determines the sign
- If x and y are positively (negatively) correlated then the estimate for β_1 will be positive (negative)
- $\sum_{i=1}^n (x_i - \overline{x})^2 > 0$ is only assumption needed to **calculate** the OLS estimator
- In practice this imposes is that there needs to be some variation in x

No Variation in x (Wooldridge 2013)

A scatterplot of wage against education when $educ_i = 12$ for all i.



Why Is This Estimator Called Ordinary Least Squares?

- $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$ is the **predicted value** for individual i
- The **residual** for i is thus:

$$\begin{aligned}\widehat{u}_i &= y_i - \widehat{y}_i \\ &= y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\end{aligned}$$

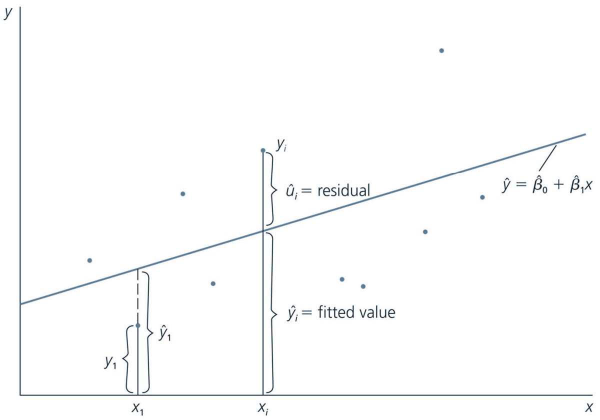
- $\widehat{\beta}_0$ and $\widehat{\beta}_1$ minimize the sum of squared residuals/errors (SSE):

$$\sum_{i=1}^{n} \widehat{u}_i^2 = \sum_{i=1}^{n} (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2$$

- Solving the minimization exercise would lead to the same solution we derived earlier!

Fitted Values and Residuals (Wooldridge 2013)

Fitted values and residuals.



Example: Mincerian Earnings Function

- Suppose we want to understand the relationship between (log) earnings and educational attainment
- We want to run the following regression:

$$\ln(\text{wage}) = \beta_0 + \beta_1 \text{education} + u$$

- This formulation has the attractive feature that the **rate of return to education** $\left(\frac{\partial \ln \text{wage}}{\partial \text{education}}\right)$ is captured by parameter β_1:

$$\frac{\partial \ln \text{wage}}{\partial \text{education}} = \beta_1$$

Example: Mincerian Earnings Function

- Wooldridge (2013) shows the regression output:

$$\widehat{ln(wage)} = 0.584 + 0.083.educ$$

- This implies that predicted ln(wage) will $\uparrow$ by 0.083 with an extra year of education
- This approximates a 8.3% increase in wages

Multivariate Regression

- So far, we have only looked at regressions with one explanatory variable
- In practice, we would often want to have multiple such variables on the right-hand-side
- Why?
  - To make the zero conditional mean assumption more plausible
  - To relax the functional form by which x relates to y
  - To reflect theoretical considerations where multiple variables together determine y
- It is straightforward to go from the bivariate to the multivariate case in regression analysis

OLS on Multivariate Models

- Consider the following model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- OLS chooses $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$ to minimize the sum of squared residuals:

$$argmin_{\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1} - \ldots - \widehat{\beta}_k x_{ik} \right)^2$$

- x_{ij}: observation i for independent variable j

OLS on Multivariate Models

- This gives us the following $k+1$ first order conditions

$$\sum_{i=1}^{n}\left(y_i-\widehat{\beta}_0-\widehat{\beta}_1 x_{i1}-\ldots-\widehat{\beta}_k x_{ik}\right)=0$$
$$\sum_{i=1}^{n}x_{i1}\left(y_i-\widehat{\beta}_0-\widehat{\beta}_1 x_{i1}-\ldots-\widehat{\beta}_k x_{ik}\right)=0$$
$$\sum_{i=1}^{n}x_{i2}\left(y_i-\widehat{\beta}_0-\widehat{\beta}_1 x_{i1}-\ldots-\widehat{\beta}_k x_{ik}\right)=0$$
$$\vdots$$
$$\sum_{i=1}^{n}x_{ik}\left(y_i-\widehat{\beta}_0-\widehat{\beta}_1 x_{i1}-\ldots-\widehat{\beta}_k x_{ik}\right)=0$$

- $k+1$ linear equations in $k+1$ unknowns

Multivariate Linear Regression in Matrix Form

- Consider the regression $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u,\ \ i = 1, 2, ..., n$
- Here y_i is the dependent variable for observation i and x_{ij}, $j = 1, 2, ..., k$ are the independent variables
- For each i, define a row vector $\mathbf{x}_i = (1, x_{i1}, x_{i2}, ..., x_{ik})$
- $\beta = (\beta_0, \beta_1, ..., \beta_k)$ is a column vector of parameters

The Model

- Stack up the n vectors of independent variables (x_i)as follows:

$$\mathbf{X} \equiv \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ . \\ . \\ . \\ \mathbf{x_n} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & . & . & . & x_{1k} \\ 1 & x_{21} & x_{22} & . & . & . & x_{2k} \\ . & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ 1 & x_{n1} & x_{n2} & . & . & . & x_{nk} \end{bmatrix}$$

- Here $\mathbf{x_1}$ is the vector of all independent variables for individual 1

The Model

- Let $\mathbf{y}$ be a row vector of observations of y
- Let u be a row vector of observations of unobservable errors
- This then gives us the linear regression model in matrix notation:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$

- This is **just** a change in notation and xactly equivalent to $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u,\ i = 1, 2, ..., n$

Deriving the OLS Estimators

- The OLS estimator minimizes the sum of squared residuals

- This gives the FOC:

$$\sum_i^n \mathbf{x_i}'(\mathbf{y_i} - \mathbf{x_i}\widehat{\beta}) = \mathbf{0}$$

which can be written as:

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\widehat{\beta}) = \mathbf{0}$$

or

$$(\mathbf{X}'\mathbf{X})\widehat{\beta} = \mathbf{X}'\mathbf{y}$$

and finally pre-multiplying both sides by $(\mathbf{X}'\mathbf{X})^{-1}$ we obtain

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Deriving the OLS Estimators

- Note that the FOC

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\widehat{\beta}) = \mathbf{0}$$

  is exactly the same as writing

$$\mathbf{X}'\mathbf{u} = \mathbf{0}$$

  where **u** is the vector of residuals

Interpreting the OLS Regression Equation

- The **predicted value** for unit i is:

$$\widehat{y}_i = \widehat{\beta}_{0i} + \widehat{\beta}_1 x_{1i} + \widehat{\beta}_2 x_{2i} + \ldots + \widehat{\beta}_k x_{ki}$$

- We can then write down the OLS regression line (the sample regression function, SRF) as:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \ldots + \widehat{\beta}_k x_k$$

- We can also express this in terms of changes:

$$\Delta\widehat{y} = \widehat{\beta}_1 \Delta x_1 + \widehat{\beta}_2 \Delta x_2 + \ldots + \widehat{\beta}_k \Delta x_k$$

Interpreting the OLS Regression Equation

$$\Delta\widehat{y} = \widehat{\beta}_1\Delta x_1 + \widehat{\beta}_2\Delta x_2 + \ldots + \widehat{\beta}_k\Delta x_k$$

- Interpretation of $\widehat{\beta}_1$: change in $\widehat{y}$ due to a one-unit increase in x_1, keeping all other independent variables fixed
- This is what is meant by the statement "controlling for $x_2, \ldots, x_k$" which you will hear a lot
- $\widehat{\beta}_0$ is the predicted value for y when $x_1 = x_2 = \ldots = x_k = 0$.

Example

- Suppose we have estimated the following relationship between wages, educational attainment, experience, and tenure

$$\widehat{\ln wage} = 0.284 + 0.092 education + 0.0041 experience + 0.022 tenure$$

- Holding *experience* and *tenure* fixed, an extra year of education is predicted to increase wages by approximately 9.2%

- Thought experiment: two people with same *experience* and *tenure* but one year difference in education $\Rightarrow$ we would expect them to have 9.2% higher wage

OLS Properties That Are Always True

- OLS has some properties that, by construction, hold <u>in any sample</u> of data

1. $\sum_{i=1}^{n} \widehat{u}_i = 0$, so $\overline{y} = \overline{\widehat{y}}$
2. $\sum_{i=1}^{n} x_{ij}\widehat{u}_i = 0$ for all j
3. $(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k, \overline{y})$ is always on the OLS regression line

Multivariate Linear Regression: Gauss-Markov Assumptions

1. Linear in parameters

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$

2. No perfect collinearity

The matrix **X** has rank (k+1)

3. Zero conditional mean

$$E(\mathbf{u}|\mathbf{X}) = 0$$

4. Homoskedasticity and no serial correlation (we will return to this when discussing *inference* in Lecture 5)

$$Var(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I_n}$$

Perfect Collinearity

- We have already discussed assumptions (1) and (3)
- What does the second assumption mean?
  - No exact **linear** relationships between variables
  - None of the variables is a constant (intercept included separately)
  - When might this occur?

Examples of Perfect Collinearity

- Perfect collinearity happens when one variable is a **linear** transformation of the other
  - For instance, you will run into problems if you include income in dollars and in cents as separate variables
  - Non-linear relationships are fine (e.g., income and income-squared—but not ln(income) and ln(income-squared)!)

- Perfect collinearity can happen if you include the wrong dummy variables
  - For instance, you include both a dummy for men and a dummy for women
  - The solution is simple: in any set of dummies that exhaust the set of possibilities, you need to exclude one from the regression

- The third possibility is that there is no variation in a given variable (can always be expressed as a multiple of the intercept)

Unbiasedness of OLS

- An estimator is unbiased if its expected value equals the population value

$$E(\widehat{\beta_j}) = \beta_j$$

- In the bivariate case, this means that

$$E(\widehat{\beta_0}) = \beta_0 \quad \text{and} \quad E(\widehat{\beta_1}) = \beta_1$$

- OLS in unbiased given the assumptions we have made!
- Under the Gauss-Markov assumptions, the OLS estimator is BLUE (best linear unbiased estimators)

Unbiasedness of OLS (A Quick Proof in the Multivariate Case)

- Note first that

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

which can be written as

$$\begin{aligned}\widehat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\end{aligned}$$

- Taking expectations conditional on X yields

$$\begin{aligned}\mathbf{E}(\hat{\beta}|\mathbf{X}) &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}(\mathbf{u}|\mathbf{X}) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{0}) \\ &= \beta\end{aligned}$$

Recap of This Lecture

1. Warm-up, recap, and overview
2. Introducing linear regression
3. Deriving OLS estimators
   - Bivariate linear regression
   - Finding $\hat{\beta}$
   - Multivariate linear regression
4. Multivariate linear regression: Assumptions and unbiasedness

(Recommended) Readings

- Chapters 2 and 3 in Wooldridge's book
- (Alternatively) sections 2.11-2.16 in Cunningham's book (more concise, if you have seen this material earlier)
- For the matrix form of the multivariate linear regression model, see the relevant appendix in Wooldridge's book
- If you liked the example in the beginning of the class, you can read the full paper by Chattopadhyay and Duflo (2004)