Ordinary Least Squares Regression

Goals for this unit

- More on notation and terminology
- OLS scalar versus matrix derivation

Some Preliminaries

In this class we will be learning to analyze

- Cross Section Data
- Panel Data (Longitudinal Data)
- Time Series Data is ignored in this class (take ECON 408)

Cross Section Data

Information at a point in time for N individuals, firms, or economic units.

id	year	wage	education	experience
1	1996	15	12	10
2	1996	23	14	8
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
99	1996	9	10	2
100	1996	54	16	15

Balanced Panel Data

Information at T points in time for each of N individuals, firms, or economic units.

id	year	wage	education	experience
1	1996	15	12	10
1	1997	17	12	11
1	1998	20	12	12
2	1996	23	14	8
2	1997	20	15	9
2	1998	21	15	10
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
99	1996	9	10	2
99	1997	13	10	3
99	1998	15	10	4
100	1996	54	16	15
100	1997	58	17	16
100	1998	74	18	17

Here $T = 3$ (indexed by years) and N=100 (indexed by id)

Unbalanced Panel Data

Information on up to T points in time for each of N individuals, firms, or economic units.

id	year	wage	education	experience
1	1997	17	12	11
1	1998	25	12	12
2	1996	23	14	8
2	1998	21	15	10
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
99	1996	9	10	2
99	1997	13	10	3
99	1998	15	10	4
100	1996	58	17	16
100	1997	74	18	17

Here $T = 3$ (indexed by years) and N=100 (indexed by id)

Time Series Data (Not Covered)

id	year	wage	education	experience
1	1986	15	12	8
1	1987	17	12	9
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
1	2005	33	13	27
1	2006	35	13	28

Here $T = 20$ (indexed by years) and $N=1$ (indexed by id)

Yardsticks

- Unbiased: $E[b] = \beta$
- Consistent: $E[b] \to \beta$, as N approaches ∞

Yardsticks, cont.

- Efficiency: The estimator uses all information at hand for the **best** estimate of β and the variance/covariance matrix.

"An estimator is asymptotically efficient if it is consistent, asymptotically normally distributed, and has an asymptotic covariance that is not larger than the asymptotic covariance matrix of any other consistent, asymptotically distributed estimator."

OLS

Consider a situation where we want to test a hypothesis about economic behavior or some type of economic phenomena.

- What are the returns to education?
- How do premiums impact choice of Obamacare?

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$$

OLS and linear algebra

$$\begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i1} & x_{i2} & x_{i3} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & x_{N3} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_N \end{bmatrix} \tag{1}$$

Exercise

- Check for conformability
- Show that any row in the above relationship can be written

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i \quad (2)$$

- Why is a column of ones needed in $\mathbf{x}$?

OLS and linear algebra, cont.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i1} & x_{i2} & x_{i3} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & x_{N2} & x_{N3} \end{bmatrix}_{N\times 4} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}_{4\times 1} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_N \end{bmatrix}_{N\times 1}$$

Estimation Problem

- For most empirical research in the Social Sciences, interested in marginal effects or elasticities. The ME are:

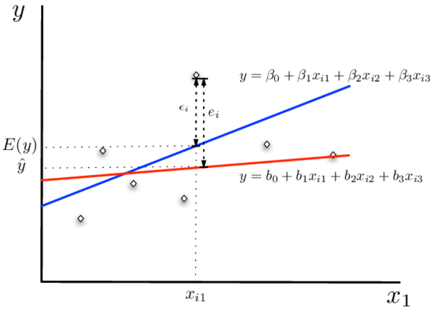
$$\frac{\partial y_i}{\partial x_{ik}} = \beta_k$$

- Don't know β or ϵ.
- However, given an estimate for β ($\mathbf{b}$), we can construct estimates for ϵ_i (e_i) as

$$e_i = y_i - \left(b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3}\right) = y_i - \mathbf{x}_i \mathbf{b}$$

where $\mathbf{x}_i = \begin{bmatrix} 1 & x_{i1} & x_{i2} & x_{i3} \end{bmatrix}$

Visualizing OLS



Note that for every point,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + e_i$$

Or, for all points

$$\mathbf{y} = \mathbf{x}\beta + \epsilon = \mathbf{xb} + \mathbf{e}$$

Choosing $\mathbf{b}$

One way to calculate $\mathbf{b}$ is to minimize the sum squared errors (SSE):

$$\sum_{i=1}^{n}(e_i)^2 = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3}))^2$$

Exercise: show that $\sum_{i=1}^{n}(e_i)^2$ is equivalent to

$$(\mathbf{y} - \mathbf{x}\mathbf{b})'(\mathbf{y} - \mathbf{x}\mathbf{b}) = \mathbf{e}'\mathbf{e}$$

The Minimization Problem

$$\min_{\mathbf{b}} \quad SSE(\mathbf{b})$$

$$\min_{\mathbf{b}} \quad \mathbf{e}'\mathbf{e}$$

The Minimization Problem

$$\min_{\mathbf{b}} \quad SSE(\mathbf{b})$$

$$\min_{\mathbf{b}} \quad \mathbf{e}'\mathbf{e}$$

Simplifying and using the inner product rule:

$$\min_{\mathbf{b}} \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{x}\mathbf{b} + \mathbf{b}'\mathbf{x}'\mathbf{x}\mathbf{b}$$

First order conditions:

$$\frac{\partial S}{\partial \mathbf{b}} = -2\mathbf{x}'\mathbf{y} + \mathbf{2x}'\mathbf{xb} = 0$$

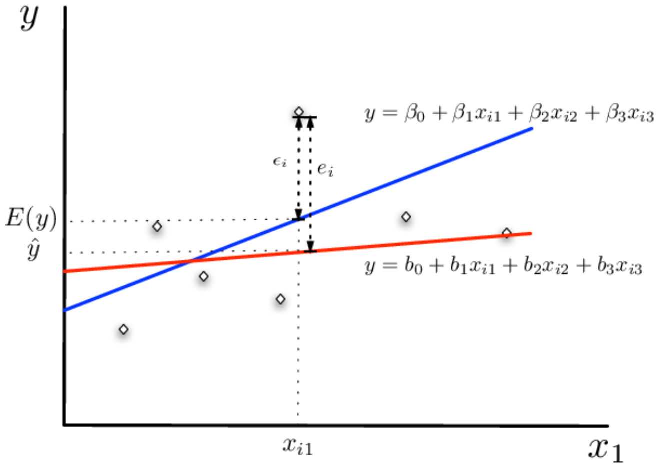
First order conditions:

$$\frac{\partial S}{\partial \mathbf{b}} = -2\mathbf{x}'\mathbf{y} + \mathbf{2x}'\mathbf{x}\mathbf{b} = 0$$

can be simplified to solve for the OLS estimator **b**

$$\mathbf{b} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$$

What have we just solved for?



Implementation in Stata

Some Assumptions we will use for deriving model

Recall the assumptions in the classic multiple regression model:

1. Linear in parameters
2. There is no linear relationships between the columns in the matrix **x**
3. Exogeneity of the independent variables: $E[\epsilon_i | x_{j1}, x_{j2}, \ldots, x_{j3}] = 0$, for all j. This condition ensures that the independent variables for observation i, or for any other observation are not useful for predicting the disturbance terms in the model.
4. Homoskedasticity and no autocorrelation: $E[\epsilon_i | \epsilon_j]$ $\forall j \in n$ and share the same variance σ^2.
5. Disturbance terms are normally distributed, and given condition (3) and (4): $\epsilon \sim N(0, \sigma^2 I)$

Properties: Unbiased

Is $E(\mathbf{b}) = \beta$?

Properties: Unbiased

Is $E(\mathbf{b}) = \beta$?

$$\begin{aligned} E(\mathbf{b}) &= E[(\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}'\mathbf{y}] \\ &= E[(\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}'(\mathbf{x}\beta + \epsilon)] \\ &= E[(\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}'\mathbf{x}\beta + (\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}'\epsilon] \\ &= \beta + E[(\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}'\epsilon] \end{aligned} \tag{3}$$

Need $E[\mathbf{x}'\epsilon] = 0$. Two equivalent ways to think about this:

1. Errors uncorrelated with independent variables.
2. Independent variables have no useful information for predicting errors.

Properties: For unbiased $\mathbf{b}$ need $E[\mathbf{x}'\epsilon] = 0$

Implications:

If errors correlated with independent variables we have endogeneity and the results are **biased**.

Exercise: investigate dimensions of $E[\mathbf{x}'\epsilon] = 0$

Properties: Variance of the estimator

- The variance of any random variable (scalar) is defined as

$$\begin{aligned} Var(a) &= E[(a - E(a))^2] \\ &= E[(a - \mu_a)^2] \end{aligned}$$

- We are interested in the variance/covariance relationship between elements of our estimated vector **b** in order to perform statistical inference on our parameter estimates. The definition of a variance/covariance **matrix** for a random vector **b** is

$$\begin{aligned} Var(\mathbf{b}) =& E[(\mathbf{b} - \mathbf{E}[\mathbf{b}])(\mathbf{b} - \mathbf{E}[\mathbf{b}])'] \\ =& E[(\mathbf{b} - \beta)(\mathbf{b} - \beta)'] \end{aligned}$$

Exercise: what are dimensions?

Properties: Variance of the estimator, cont.

$$
\begin{aligned}
Var(b) &= E[(\mathbf{b}-\beta)(\mathbf{b}-\beta)'] \\
&= E[((\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}'\mathbf{y}-\beta)((\mathbf{x}'\mathbf{x})^{-\mathbf{1}}\mathbf{x}'\mathbf{y}-\beta)']
\end{aligned}
$$

substitute $\mathbf{y}=\mathbf{x}\beta+\epsilon$, check for conformability and simplify.

Properties: Variance of the estimator, cont.

$$\begin{aligned} Var(\mathbf{b}) &= (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'E[\epsilon\epsilon']\mathbf{x}((\mathbf{x}'\mathbf{x})^{-1})' \\ &= (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'E[\epsilon\epsilon']\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1} \end{aligned}$$

To simplify further, we invoke the assumption above ($\epsilon \sim N(0, \sigma^2 I)$), so that $E[\epsilon\epsilon'] = \sigma^2 I$.

Exercise: simplify $Var(\mathbf{b})$ further.

Properties: Variance of the estimator, cont.

Finally, we have

$$Var(b) = \sigma^2 (\mathbf{x}'\mathbf{x})^{-\mathbf{1}}{}_{k \times k}$$

I refer to this as the **Variance/Covariance Matrix of the Parameters**.

How to recover σ^2?

- σ^2 is the variance of the error. Given assumption (3), errors are homoskedastic- shared by everyone in the population.
- σ^2 is unobserved, but estimate it by using estimated errors for each observation and recover sample average

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{N-K} = \frac{(\mathbf{y}-\mathbf{xb})'(\mathbf{y}-\mathbf{xb})}{N-K}$$

Estimated Variance/Covariance Matrix of Parameters

$$var(\mathbf{b}) = s^2(\mathbf{x}'\mathbf{x})^{-\mathbf{1}}$$

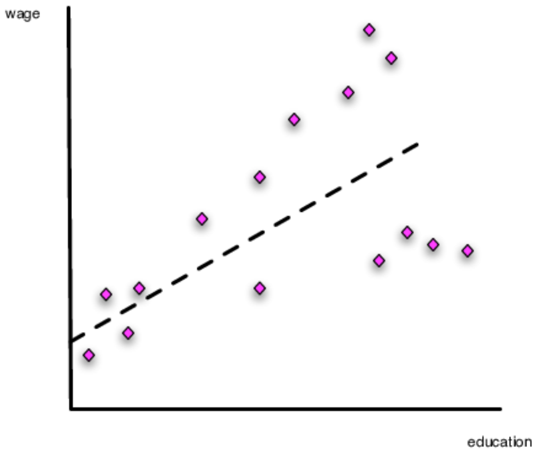
Use diagonal elements of this matrix for the variance of **each** parameter estimated in the model:

$$Var[\mathbf{b}] = s^2(\mathbf{x}'\mathbf{x})^{-\mathbf{1}} = s^2 \begin{bmatrix} 0.0665 & -0.0042 & -0.0035 & -0.0014 \\ -0.0042 & 0.5655 & 0.0591 & -0.0197 \\ -0.0035 & 0.0591 & 0.0205 & -0.0046 \\ -0.0014 & -0.0197 & -0.0046 & 0.0015 \end{bmatrix}$$

So, the standard error from our regression for b_0 (assumed to be the first parameter) is $\sqrt{s^2 \times .0665}$

Implement in Stata

Heteroskedasticity



Testing for Heterskedasticity

- The symptoms of Heteroskedasticity are embodied in the relationship between the linear predictor **xb** and the estimated error.
- Idea: see if the variances of the estimated model errors (**e**) are a function of the independent variables (**x**)

Testing for Heterskedasticity: What stata does (estat hettest)

- **Step 1**: Recover model residuals ($\mathbf{e} = \mathbf{y} - \mathbf{xb}$) and predictions ($\widehat{\mathbf{y}} = \mathbf{xb}$).
- **Step 2**: Using the residuals, calculate

$$r = \frac{diagonal(\mathbf{ee}')}{\frac{1}{N}(\mathbf{y} - \mathbf{xb})'(\mathbf{y} - \mathbf{xb})}$$

- **Step 3**: Run the regression $\mathbf{r} = \delta_{\mathbf{0}} + \delta_{\mathbf{1}}\widehat{\mathbf{y}} + \mathbf{v}$, and recover the Model Sum of Squares (MSS): $(\widehat{\mathbf{r}} - \overline{\widehat{\mathbf{r}}})'(\widehat{\mathbf{r}} - \overline{\widehat{\mathbf{r}}})$
- **Step 4**: Using the $\frac{MSS}{2}$ (distributed $\chi^2(1)$), perform the following hypothesis test:
  - H_0 : No Heteroskedasticity: $\sigma_i^2 = \sigma^2 \quad \forall i \in N$
  - H_1 : Heteroskedasticity: $\sigma_i^2 \neq \sigma^2 \quad \forall i \in N$

The Fix

-Remember back to our proof

$$Var(\mathbf{b}) = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'E[\epsilon\epsilon']\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}$$

-OLS assumes we can estimate $Var(\epsilon)$ as

$$var(e) = s^2 \times I_{N\times N} = \begin{bmatrix} s^2 & 0 & 0 & \dots & 0 \\ 0 & s^2 & 0 & \dots & 0 \\ 0 & 0 & s^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & s^2 \end{bmatrix}$$

The Fix, cont.

For robust standard errors, use this

$$var(e) = \hat{\mathbf{V}} = \begin{bmatrix} s_1^2 & 0 & 0 & \dots & 0 \\ 0 & s_2^2 & 0 & \dots & 0 \\ 0 & 0 & s_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & s_N^2 \end{bmatrix}$$

where

$$s_i = e_i = y_i - \mathbf{x}_i \mathbf{b} = y_i - \hat{y}_i$$

Robust Standard Errors

The robust standard error variance covariance matrix is then:

$$Var(\mathbf{b}) = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\hat{\mathbf{V}}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}$$

Notice we use information in the estimated error to fix the Heteroskedasticity problem.

Things to reflect on:

1. Should we always use robust standard errors?
2. We are able to leverage estimated errors in this way because **b** is unbiased in the presence of heteroskedasticity. Therefore, our estimates for ϵ_i are unbiased.