Introductory Econometrics

Simple and Multiple Regression Analysis

Monash Econometrics and Business Statistics

Semester 1, 2018

Recap

Stages of empirical analysis in business and economics:

1. Understanding the problem
2. Formulating an appropriate conceptual model to tackle the problem
3. Collecting appropriate data
4. Looking at data (Descriptive Analytics)
5. Estimating the model, making inference, predictions and policy prescriptions as appropriate (Predictive and Prescriptive Analytics)
6. Evaluating, learning and improving each of the previous steps, and iterating until the problem is solved

Lecture Outline: Wooldridge: Ch 2 - 2.3, Appendix B-1, B-2, B-4

- Explaining y using x with a model
- Simple linear regression (textbook reference Chapter 2, 2-1)
- The OLS estimator (textbook reference, Ch 2, 2-2,2-3)
- Simple linear regression in matrix form
- Geometric interpretation of least squares (not in the textbook)
- Stretching our imagination to multiple regression - Appendix E of the textbook, section E-1
- **Assumption:** I assume that almost everyone has done Part A of the tutorial on matrices

What is a model?

- To study the relationship between random variables, we use a *model*
- A complete model for a random variable is a model of its probability distribution
- For example, a possible model for heights of adult men is that it is normally distributed, and since normal distribution is fully described by its mean and variance, we only have to estimate the mean and variance from a sample of observations on adult men and we have estimated a model example1
- But when studying two or more random variables, modelling the joint distribution is difficult and requires a lot of data.
- So, we model the conditional distribution of y given x.
- In particular we provide a model for $E(y \mid x)$

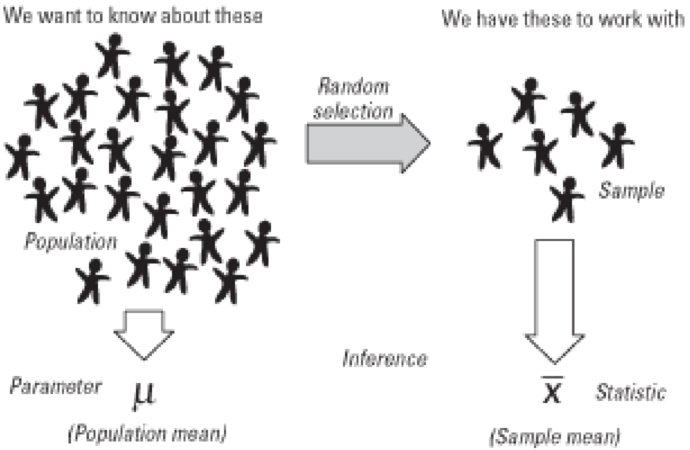
Terminology and notation

- Notation: Unlike first year, we use lower case letters for random variables
- Terminology: There are many different ways people refer to the target variable y that we want to explain using variable x. These include:

y	x
Dependent Variable	Independent Variable
Explained Variable	Explanatory Variable
Response Variable	Control Variable
Predicted Variable	Predictor Variable
Regressand	Regressor

- The expressions
  "Run a regression of y on x", and,
  "Regress y on x"
  both mean "Estimate the model $y = \beta_0 + \beta_1 x + u$ using the ordinary least squares method"

Modelling mean



$$E(y \mid x) = \beta_0 + \beta_1 x$$

(Conditional expectation function)

$$\hat{y} = \widehat{E(y \mid x)} = \hat{\beta}_0 + \hat{\beta}_1 x$$

(Sample regression function)

Laws of Probability

- To understand what a conditional expectation is, we start with laws of probability

1. Probability of any event is a number between 0 and 1. The probabilities of all possible outcomes of a random variable add up to 1
2. If A and B are mutually exclusive events then

$$P(A \text{ or } B) = P(A) + P(B)$$

3. If A and B are two events, then

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

Joint probability density function

- The ultimate goal of this unit is to study the relationship between one variable y with many variables x_1 to x_k. But let's start with only one x, and with the discrete case.

- Suppose y is the number of bathrooms and x is the number of bedrooms in an apartment in Melbourne. Assume y has two possible values, 1 and 2, and x has three possible values 1, 2 and 3. The joint pdf gives us the probabilities of every possible outcome of (x, y).

$y \downarrow, x \rightarrow$	1	2	3	marginal f_y
1	0.40	0.24	0.04	
2	0.00	0.16	0.16	
marginal f_x				

- The entries show the probabilities of different possible combination of bedrooms and bathrooms (x, y). For example the top left cell shows $P(x = 1 \& y = 1) = 0.40$.
- Check the first law of probability: all probabilities are between 0 and 1 and the probabilities of all possible outcomes sum to 1
- Using the second law of probability, we can use the joint pdf to deduce the pdf of x by itself (called the marginal density of x), and also the marginal density of y

Conditional density

- Using the third law of probability, we can also deduce the conditional distribution of number of bathrooms in an apartment given that it has 1 bedroom.

$$P(y = 1 \mid x = 1) = \frac{P(y = 1 \ \& \ x = 1)}{P(x = 1)} = \frac{0.40}{0.40} = 1.00$$

$$P(y = 2 \mid x = 1) =$$

- Similarly, we can deduce the conditional distribution of y given $x = 2$

$$P(y = 1 \mid x = 2) = \frac{P(y = 1 \ \& \ x = 2)}{P(x = 2)} =$$

$$P(y = 2 \mid x = 2) =$$

- And y given $x = 3$

$$P(y = 1 \mid x = 3) =$$

$$P(y = 2 \mid x = 3) =$$

Conditional expectation function

- Each of these conditional densities has an expected value

$y \mid x = 1$	$f_{y\mid x=1}$
1	1.00
2	0.00

$$\Rightarrow \quad E(y \mid x = 1) = 1 \times 1.00 + 2 \times 0.00 = 1.00$$

$y \mid x = 2$	$f_{y\mid x=2}$
1	0.60
2	0.40

$$\Rightarrow \quad E(y \mid x = 2) = 1 \times 0.60 + 2 \times 0.40 = 1.40$$

$y \mid x = 3$	$f_{y\mid x=3}$
1	0.20
2	0.80

$$\Rightarrow \quad E(y \mid x = 3) = 1 \times 0.20 + 2 \times 0.80 = 1.80$$

- Plot the expected values for different values of x. Do they fit on a straight line? What is the equation of that line?

- When y and x have many possible outcomes or when they are continuous random variables (like birth weight and number of cigarettes during pregnancy, or price of a house and its land size) we cannot enumerate the joint density and perform the same exercise
- Therefore we go after the conditional expectation function directly

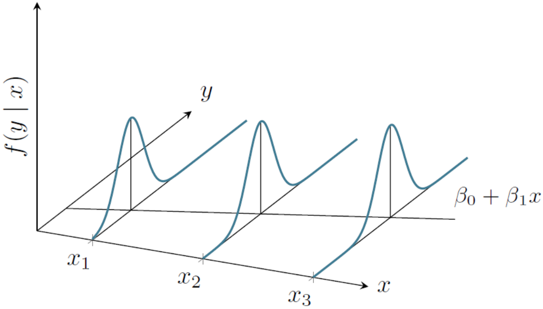
$$E(y \mid x) = \beta_0 + \beta_1 x \tag{PRF}$$

- For example, if y is the price of a house and x is its area, the mean of the price of houses with area x is given by $\beta_0 + \beta_1 x$
- The price of each house can be written as random variations around this central value

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

  where u is a random variable with $E(u \mid x) = 0$, which implies that $E(u) = 0$ also

The simple linear regression model in a picture



The simple linear regression model in equation form

- The following equation specifies the conditional mean of y given x

$$y = \beta_0 + \beta_1 x + u \text{ with } E(u \mid x) = 0$$

- It is an incomplete model because it does not specify the probability distribution of y conditional of x
- If we add the assumptions that $Var(u \mid x) = \sigma^2$ and that the conditional distribution of u given x is normal, then we have a complete model, which is called the "classical linear model" and was shown in the picture on the previous slide.
- In summary: we make the assumption that in the big scheme of things, data are generated by this model, and we want to use observed data to learn the unknowns β_0, β_1 and σ^2 in order to predict y using x.

The OLS estimator

- Let's leave the theory universe and go back to the data world
- We have a random sample of n observations on two variables x and y in two columns of a spreadsheet
- Let's denote them by $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- We want to see if these two columns of numbers are related to each other
- In the simple case that there is only one x, we look at the scatter plot, which is very informative visual tools and give us a good idea of the correlation between y and x (bodyfat and weight)
- Unfortunately though, in some business and economic applications the signal is too weak to be detected by data visualisation alone (asset return and size)

The OLS estimator

- How can we determine a straight line that fits our data best?
- We find $\widehat{\beta}_0$ and $\widehat{\beta}_1$ that minimise the **sum of squared residuals**

$$SSR(b_0, b_1) = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2$$

- My favourite visual explanation of this is in http://youtu.be/jEEJNz0RK4Q
- The first order conditions are:

$$\frac{\partial SSR}{\partial b_0}|_{\widehat{\beta}_0, \widehat{\beta}_1} = -2\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0$$
$$\frac{\partial SSR}{\partial b_1}|_{\widehat{\beta}_0, \widehat{\beta}_1} = -2\sum_{i=1}^{n} x_i(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0$$

- Two equations and two unknowns, we can solve to get $\widehat{\beta}_0$ and $\widehat{\beta}_1$

- We obtain the set of equations

$$\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n} (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{n} x_i \hat{u}_i = \sum_{i=1}^{n} x_i (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0$$

- And after some algebra (see page 28 of the textbook, 5th ed., or page 26, 6th ed.)

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

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$$

- Several important things to note:

  1. $$\widehat{\beta_1} = \frac{\widehat{Cov(x,y)}}{\widehat{Var(x)}} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2}$$

     Those of you who are doing finance, now realise where the name "*beta*" of a stock has come from!

  2. From the formula for $\widehat{\beta_0}$ we see that $\bar{y} = \widehat{\beta_0} + \widehat{\beta_1}\bar{x}$, which shows that $(\bar{x}, \bar{y})$ lies on the regression line, i.e. regression prediction is most accurate for the sample average.

  3. Both $\widehat{\beta_0}$ and $\widehat{\beta_1}$ are functions of sample observations only. They are estimators, i.e. formulae that can be computed from the sample. If we collect a different sample from the same population, we get different estimates. So these estimators are random variables.

Estimators we have seen so far

Population parameter	its estimator
population mean $\mu_y = E(y)$	sample average $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
population variance $\sigma_y^2 = E(y - \mu_y)^2$	sample variance $s_y^2 = \hat{\sigma}_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
population st. dev. $\sigma_y = \sqrt{\sigma_y^2}$	sample st. dev. $\hat{\sigma}_y = \sqrt{\hat{\sigma}_y^2}$
population covariance $\sigma_{xy} = E(x - \mu_x)(y - \mu_y)$	sample covariance $\hat{\sigma}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
population correlation coeff. $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$	sample corr. coeff. $\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$
slope and intercept of the PRF β_1 and β_0 in $E(y \mid x) = \beta_0 + \beta_1 x$	their OLS estimators $\widehat{\beta_1} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2}$ $\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$

Simple linear regression in matrix form

- For each of the n observations (x_1, y_1),(x_2, y_2),. . .,(x_n, y_n), our model implies

$$y_i = \beta_0 + \beta_1 x_i + u_i \ , \ \ i = 1, \ldots, n$$

- We stack all of these on top of each other

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \beta_0 + \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \beta_1 + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

- We define the $n \times 1$ vectors $\mathbf{y}$ (dependent variable vector) and $\mathbf{u}$ (error vector):

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \text{ and } \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

and the $n \times 2$ matrix of regressors

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

and the 2×1 parameter vector

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

which allow us to write the model simply as

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

- In multiple regression where we have k explanatory variables plus the intercept

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times(k+1)}{\mathbf{X}} \underset{(k+1)\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\mathbf{u}}$$

where

$$\underset{n\times(k+1)}{\mathbf{X}} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

and

$$\underset{(k+1)\times 1}{\boldsymbol{\beta}} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

- Remember that $\mathbf{y}, \mathbf{X}$ are observable, $\boldsymbol{\beta}$ and $\mathbf{u}$ are unknown and unobservable

- Also remember that $\mathbf{Xb}$ for any $(k+1)\times 1$ vector $\mathbf{b}$ is an $n \times 1$ vector that is a linear combination of columns of $\mathbf{X}$

The power of abstraction

- In the real world we have goals such as:
  - We want to determine the added value of education to wage after controlling for experience and IQ based on a random sample of 1000 observations
  - We want to establish if the return on a firm's share price is related to its value measured by its book to market ratio and firm's size based on 24 yearly observations on 500 stocks
  - We want to predict the value of house in the city of Monash given characteristics such as land area, number of bedrooms, number of bathrooms, proximity to public transport, etc., based on data on the last 100 houses sold in Monash
  - We want to forecast hourly electricity load based on all information available at the time that forecast is made, based on hourly electricity load in the last 3 years
- In all cases we want to get a good estimate of $\boldsymbol{\beta}$ in the model $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$ based on our sample of observations
- We have now turned these very diverse questions into a single mathematical problem

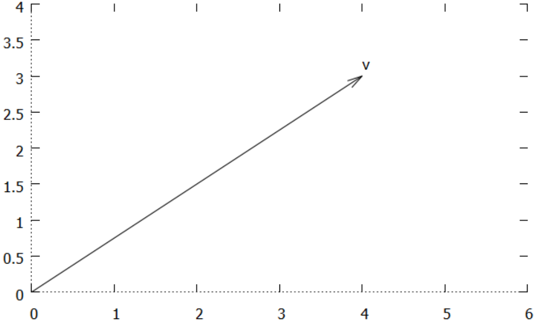
The OLS solution

- OLS finds the vector in the column space of **X** which is closest to **y**. Let's see what this means
- To visualise we are limited to 3 dimensions at most
- Assume we want to explain house prices with number of bedrooms
- We have data on 3 houses, with 4, 1 and 1 bedrooms which sold for 10, 4 and 6 $\times \$100,000$
- Want to find a good $\widehat{\boldsymbol{\beta}}$

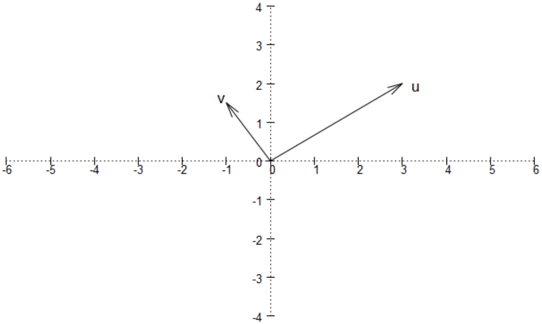
$$\begin{bmatrix}10\\4\\6\end{bmatrix} = \begin{bmatrix}1\\1\\1\end{bmatrix}\hat{\beta}_0 + \begin{bmatrix}4\\1\\1\end{bmatrix}\hat{\beta}_1 + \hat{\mathbf{u}} = \begin{bmatrix}1 & 4\\1 & 1\\1 & 1\end{bmatrix}\widehat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$$

Vectors and vector spaces

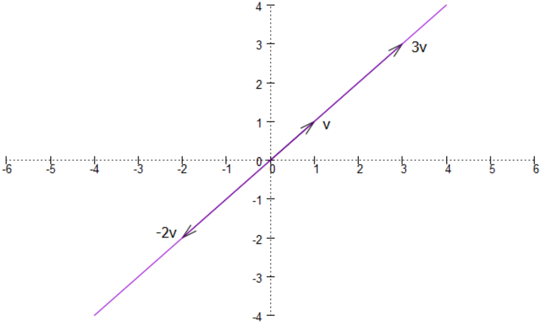
- An n-dimensional vector is an arrow from the origin to the point in $\Re^n$ with coordinates given by its elements
- Example: $\mathbf{v} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$



- The **length** of a vector is the square root of sum of squares of its coordinates
$$length(\mathbf{v}) = (\mathbf{v}'\mathbf{v})^{1/2}$$
- Example: vector **v** on the previous slide
- For **u** and **v** of the same dimension
$$\mathbf{u}'\mathbf{v} = 0 \Leftrightarrow \mathbf{u} \text{ and } \mathbf{v} \text{ are perpendicular (orthogonal) to each other}$$
- Example: $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1.5 \end{bmatrix}$



- For any constant c, the vector $c\mathbf{v}$ is on the line that passes through $\mathbf{v}$
- This line is called the "space spanned by $\mathbf{v}$"
- Example: $\begin{bmatrix} 3 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ -2 \end{bmatrix}$ are in the space spanned by $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$



- Consider a matrix **X** that has two columns that are not a multiple of each other

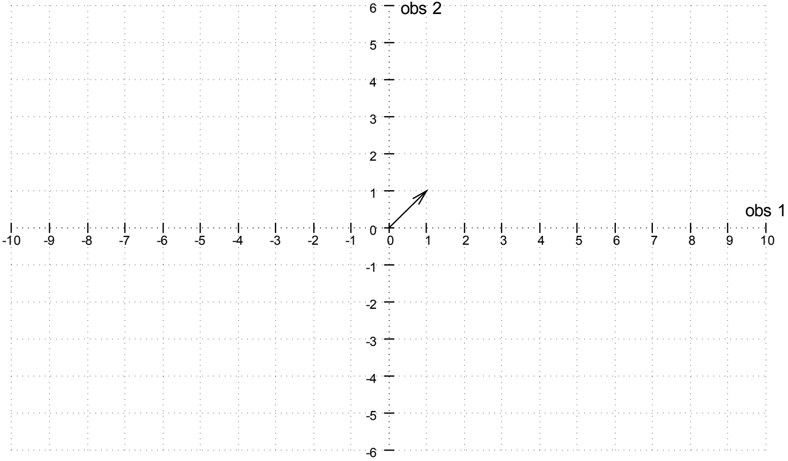
- Geometrically, these vectors form a two dimensional plane that contains all linear combinations of these two vectors, i.e. set of all **Xb** for all **b**. This plane is called the column space of **X**

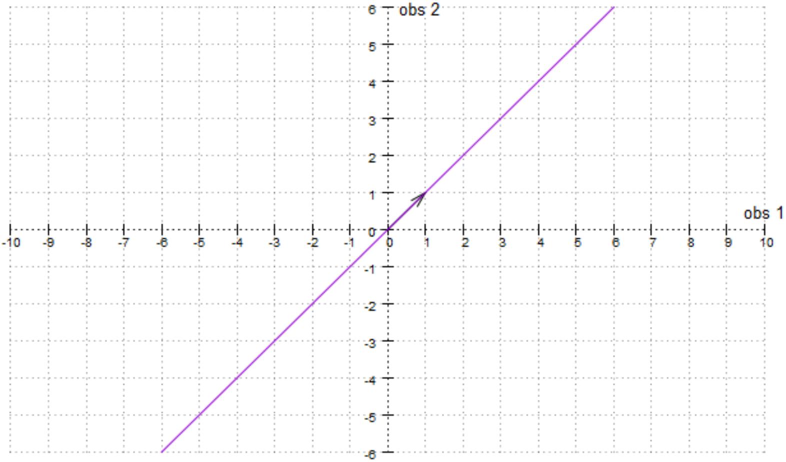
- If the number of rows of **X** is also two, then the column space of **X** is the entire $\Re^2$, that is any two dimensional **y** can be written as a linear combination of columns of **X**

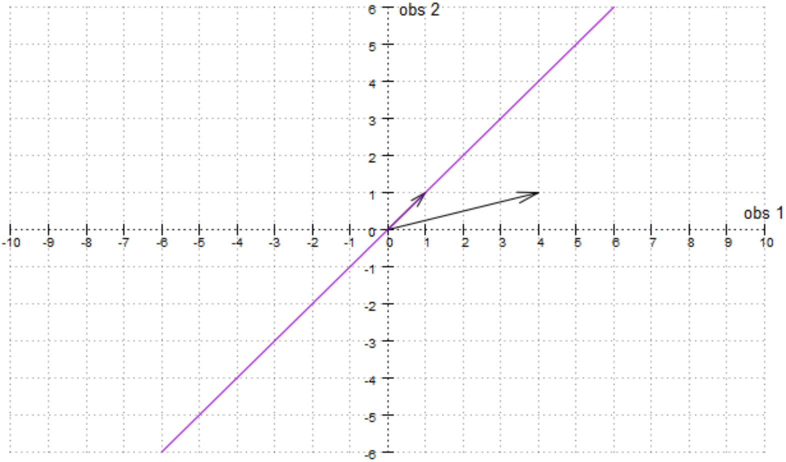
- Consider the house price example, but only with two houses

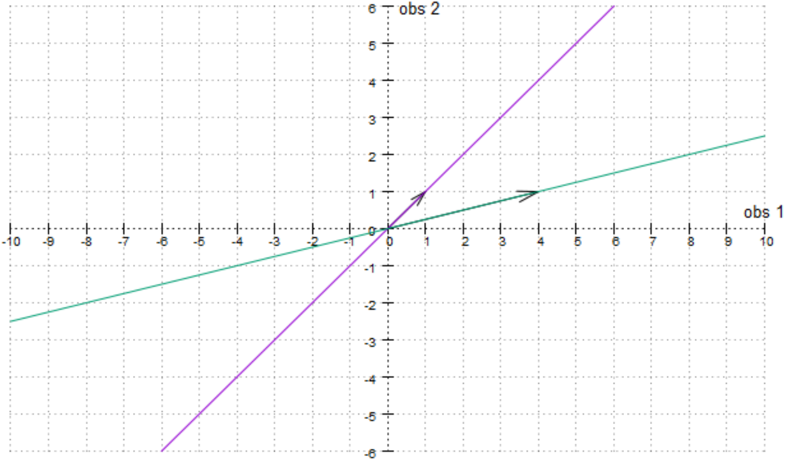
$$\begin{bmatrix} 10 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \hat{\beta}_0 + \begin{bmatrix} 4 \\ 1 \end{bmatrix} \hat{\beta}_1 + \hat{\mathbf{u}} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix} \widehat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$$

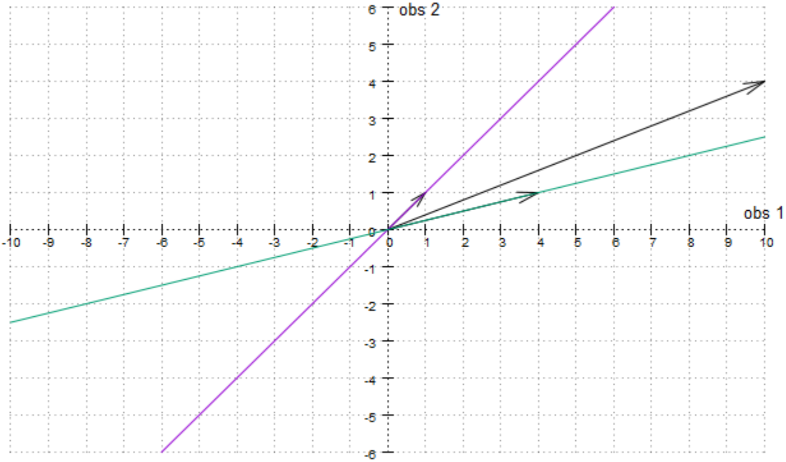
- The price vector can be perfectly explained with a linear combination of columns of **X**, i.e. with zero $\hat{\mathbf{u}}$


obs 2
obs 1
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10

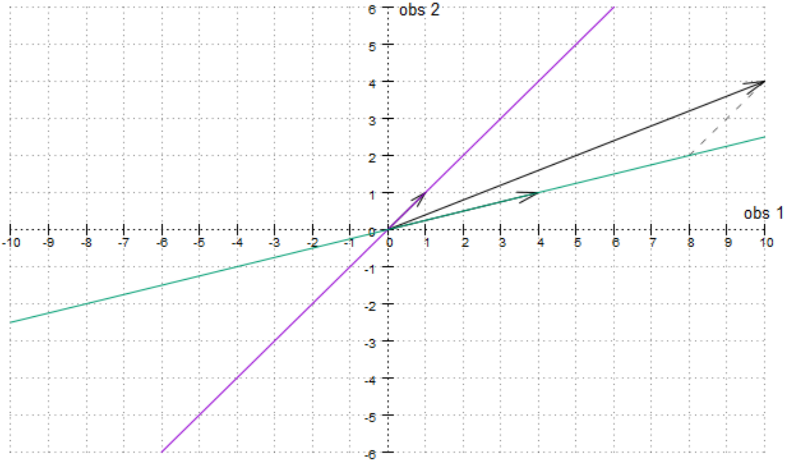

obs 2
obs 1
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10

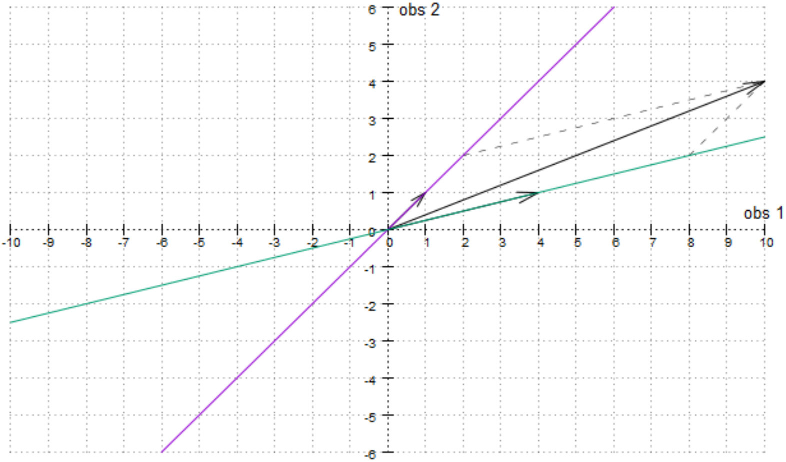

obs 2
obs 1
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10


obs 2
obs 1
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10



obs 2
obs 1
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10

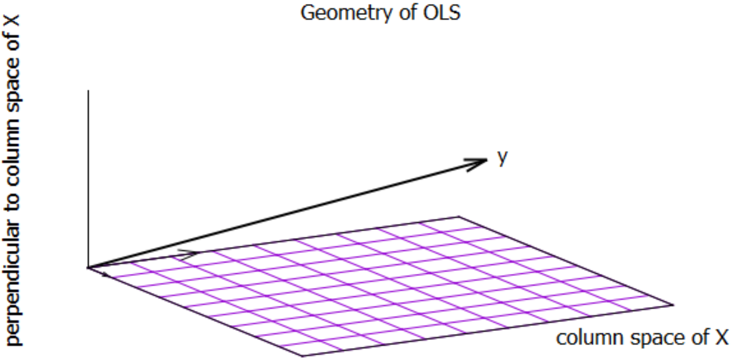

obs 2
obs 1
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10

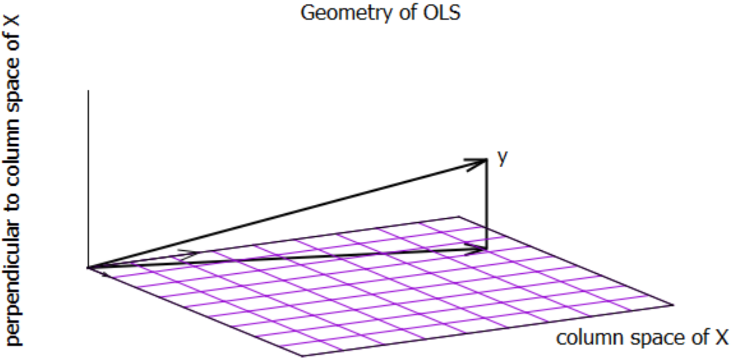

obs 2
obs 1
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7
8
9
10

- This shows that with only two observations, we would suggest

$$\widehat{price} = 2 + 2bedrooms$$

- While this fits the first two houses perfectly, when we add the third house (1 bedroom sold for 6 hundred thousands) we make an error of 2 hundred thousand dollars
- With 3 observations, the 3 dimensional price vector no longer lies in the space of linear combinations of columns of $\mathbf{X}$.
- The closest point in the column space of $\mathbf{X}$ to $\mathbf{y}$ is ...


Geometry of OLS
perpendicular to column space of X
y
column space of X


Geometry of OLS
perpendicular to column space of X
y
column space of X

- Hence, the shortest $\hat{\mathbf{u}}$ is the one that is perpendicular to column of $\mathbf{X}$, i.e.

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

- Since $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$, we get

$$\begin{aligned} \mathbf{X}'(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) &= 0 \\ \Rightarrow \mathbf{X}'\mathbf{y} &= \mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} \\ \Rightarrow & \boxed{\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}} \end{aligned}$$

- For $\mathbf{X}'\mathbf{X}$ to be invertible, columns of $\mathbf{X}$ must be *linearly independent*
- The top box is more important than the formula for the OLS estimator in the second box. Geometry of OLS is summarised in the first box: The OLS residuals are orthogonal to columns of $\mathbf{X}$

- The vector of OLS predicted values is the orthogonal projection of $\mathbf{y}$ in the column space of $\mathbf{X}$

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

- By definition

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{u}}$$

- Since $\mathbf{y}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{u}}$ form a right-angled triangle, we know (remember the length of a vector)

$$\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\mathbf{u}}'\hat{\mathbf{u}}$$

i.e.,

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} \hat{y}_i^2 + \sum_{i=1}^{n} \hat{u}_i^2 \tag{L2}$$

- Since $(1 \quad 1 \quad \cdots \quad 1)\hat{\mathbf{u}} = 0$ we also have

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i \Rightarrow \bar{y} = \bar{\hat{y}}$$

- Subtracting $n\bar{y}^2$ from both sides of (L2) and rearranging, we get

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} \hat{u}_i^2$$

  or

$$\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}$$

- This leads to the definition of the coefficient of determination R^2, which is a measure of goodness of fit

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

Summary

- Given a sample of n observations, OLS finds the orthogonal projection of the dependent variable vector in the column space of explanatory variables
- The residual vector is orthogonal to each of the explanatory variable vectors, including a column of ones for the intercept

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

- This leads to the formula for the OLS estimator in multiple regression

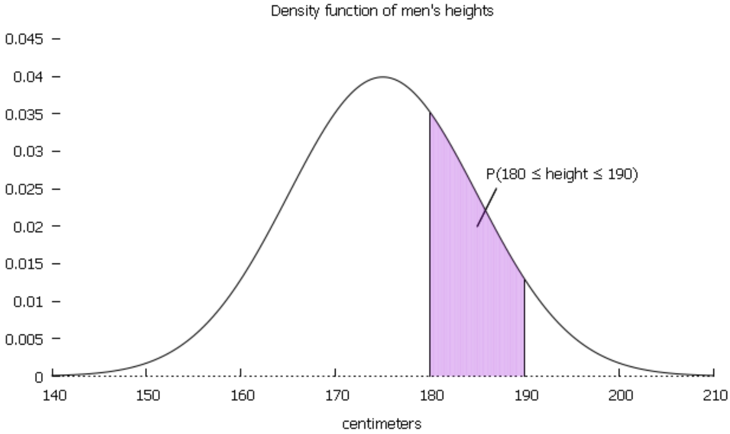
$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- It also leads to

$$\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}$$

- Note the difference between the population parameter $\boldsymbol{\beta}$ and its OLS estimator $\widehat{\boldsymbol{\beta}}$. $\boldsymbol{\beta}$ is constant and does not change, $\widehat{\boldsymbol{\beta}}$ is a function of sample and its value changes for different samples. Why are these good estimators? Next time we explore the statistical properties of $\widehat{\boldsymbol{\beta}}$

- The probability that the height of a randomly selected man lies in a certain interval is the area under the pdf over that interval



Back to what is a model