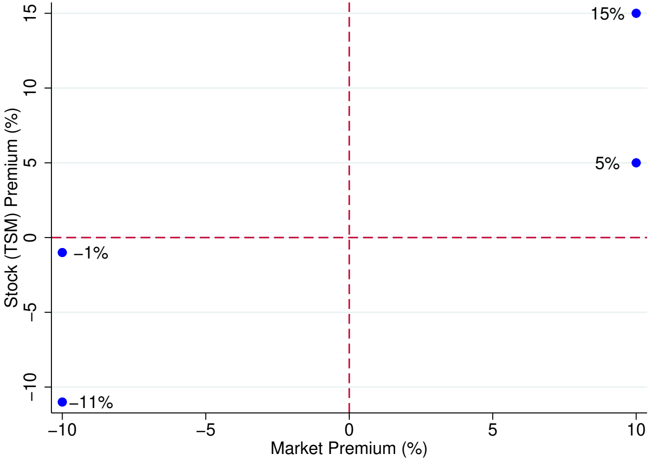
Financial Econometrics

Basic Regression

Tim C.C. Hung 洪志清

February 21^{st}, 2022

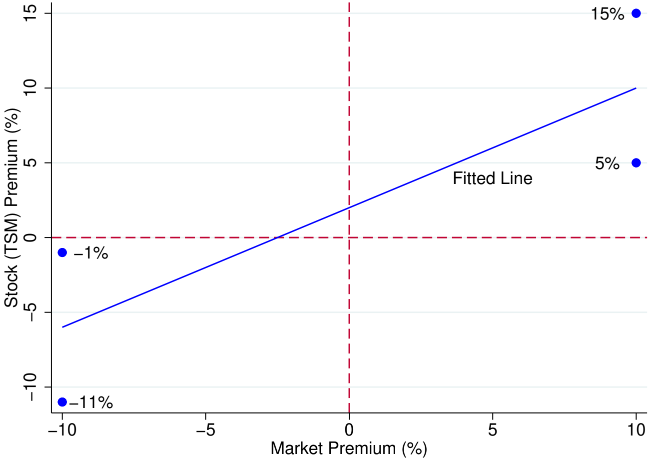
Example: Naive CAPM



- What is the relationship between TSM premium and market premium?

Example: Naive CAPM

國立臺灣大學 National Taiwan University



- A fitted line that describes their relationship.

Conditional Expectation Function

- A simple (univariate) linear regression model (of the population) can be stated as the following:

$$r(x) = E[Y|X = x] = \beta_0 + \beta_1 x$$

- This **partially** describes the data generating process in the population. (We will talk more about the DGP in multivariate regressions.)

- Y = dependent variable
- X = independent variable
- β_0, β_1 = population intercept and population slope
  Things we are interested! What we want to estimate!

Sample Linear Regression Function

- The estimated or sample regression function is:

$$\widehat{r}(X_i) = \widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i$$

- $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the estimated intercept and slope.

- $\widehat{Y}_i$ is the fitted/predicted value.

- We also have the residuals, $\widehat{u}_i$, which are the differences between the true Y and the predicted ones: $\widehat{u}_i = Y_i - \widehat{Y}_i$.

- You can think of the residuals as the prediction errors of our estimates.

Things to Catch Up!

- How to run and read regressions?
- Mechanics: How to estimate the intercept and slope?
- Properties: How good are these estimates?
- Statistic inference: How do we test the estimates and how do we interpret them?

Ordinary Least Squares (OLS)

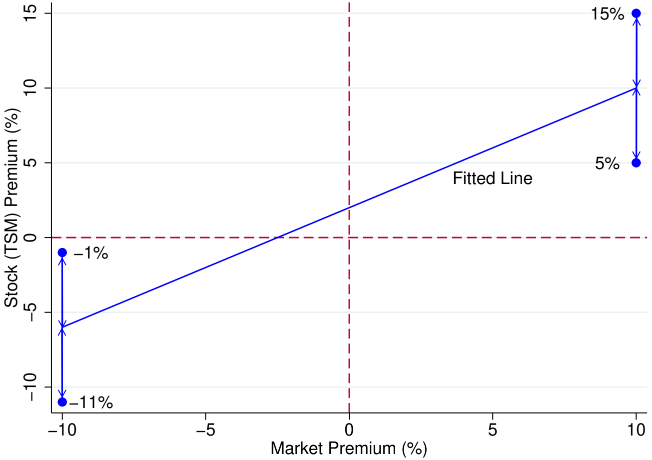
国立臺灣大學 National Taiwan University

- An estimator of the regression line.
- It is derived by minimizing the squared prediction errors of the regression:

$$(\widehat{\beta}_0, \widehat{\beta}_1) = \arg\min_{b_0, b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

- Alternatively, the OLS estimates are the intercept and slope that minimize the sum of the squared residuals.

Example: Naive CAPM



- Minimizing the sum of squares of the arrowed lines.

Derive the OLS estimators

- Let's think about n pairs of sample observations: $(Y_1, X_1), (Y_2, X_2), (Y_3, X_3), ..., (Y_n, X_n)$

- Let $\{b_0, b_1\}$ be possible value for $\{\beta_0, \beta_1\}$.

- Define the least square objective function:

$$S(b_0, b_1) = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

- How do we derive the LS estimators for β_0 and β_1? We want to minimize this function, which is actually a very well-defined calculus problem.
    1. Take partial derivative of S with respect to b_0 and b_1.
    2. Set each of the partial derivatives to 0.
    3. Solve for $\{b_0, b_1\}$.

Take Partial Derivatives

National Taiwan University

$$
\begin{aligned}
S(b_0, b_1) &= \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2 \\
&= \sum_{i=1}^{n} (Y_i^2 + b_0^2 + b_1^2 X_i^2 - 2Y_i b_0 - 2Y_i b_1 X_i + 2b_0 b_1 X_i) \\
\frac{\partial S(b_0, b_1)}{\partial b_0} &= \sum_{i=1}^{n} (-2Y_i + 2b_0 + 2b_1 X_i) \\
&= -2 \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) \\
\frac{\partial S(b_0, b_1)}{\partial b_1} &= \sum_{i=1}^{n} (-2Y_i X_i + 2b_0 X_i + 2b_1 X_i^2) \\
&= -2 \sum_{i=1}^{n} X_i (Y_i - b_0 - b_1 X_i)
\end{aligned}
$$

Solve for the Intercept

$$
\begin{aligned}
\frac{\partial S(b_0, b_1)}{\partial b_0} &= -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) \\
0 &= -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) \\
0 &= \sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i) \\
0 &= \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} b_0 - \sum_{i=1}^{n} b_1 X_i \\
b_0 * n &= (\sum_{i=1}^{n} Y_i) - b_1(\sum_{i=1}^{n} X_i) \\
b_0 &= \bar{Y} - b_1 \bar{X}
\end{aligned}
$$

A Helpful Lemma

- Lemmas are like helper results that are often invoked repeatedly.

Lemma (Deviations from the Mean Sum to 0)

$$
\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X}) &= \sum_{i=1}^{n}(X_i) - n\bar{X} \\
&= \sum_{i=1}^{n}(X_i) - n * \sum_{i=1}^{n}(X_i/n) \\
&= \sum_{i=1}^{n}(X_i) - \sum_{i=1}^{n}(X_i) \\
&= 0
\end{aligned}
$$

Solve for the Slope

国立臺灣大學 National Taiwan University

$$
\begin{aligned}
0 &= -2\sum_{i=1}^{n} X_i(Y_i - b_0 - b_1 X_i) \\
&= \sum_{i=1}^{n} X_i(Y_i - b_0 - b_1 X_i) \\
&= \sum_{i=1}^{n} X_i(Y_i - (\bar{Y} - b_1\bar{X}) - b_1 X_i) \quad \text{(sub in } b_0\text{)} \\
&= \sum_{i=1}^{n} X_i(Y_i - \bar{Y} - b_1(X_i - \bar{X})) \\
&= \sum_{i=1}^{n} X_i(Y_i - \bar{Y}) - b_1\sum_{i=1}^{n} X_i(X_i - \bar{X}) \\
b_1\sum_{i=1}^{n} X_i(X_i - \bar{X}) &= \sum_{i=1}^{n} X_i(Y_i - \bar{Y}) - \bar{X}\sum_{i=1}^{n}(Y_i - \bar{Y}) \quad \text{(add 0)}
\end{aligned}
$$

Solve for the Slope

$$b_1 \sum_{i=1}^{n} X_i(X_i - \bar{X}) = \sum_{i=1}^{n} X_i(Y_i - \bar{Y}) - \bar{X} \sum_{i=1}^{n} (Y_i - \bar{Y})$$

$$= \sum_{i=1}^{n} X_i(Y_i - \bar{Y}) - \sum_{i=1}^{n} \bar{X}(Y_i - \bar{Y})$$

$$b_1 (\sum_{i=1}^{n} X_i(X_i - \bar{X}) - \sum_{i=1}^{n} \bar{X}(X_i - \bar{X})) = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) \text{ (add 0)}$$

$$b_1 \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X}) = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$b_1 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

The OLS Estimators

- We're Done! Here are the OLS estimators:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X}$$

- Can you calculate $\widehat{\beta}_0$ and $\widehat{\beta}_1$ in our CAPM example by hand?

Intuition of the OLS Estimator

National Taiwan University

- The intercept equation tells us that the regression line goes through the point $(\bar{Y}, \bar{X})$: $\bar{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{X}$

- The slope for the regression line can be written as the following:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\text{(Sample) Cov(X,Y)}}{\text{(Sample) Var(X)}}$$

- The higher the covariance between X and Y, the higher the slope.

- What if X_i does not vary? $\rightarrow$ undefined fraction $\rightarrow$ non-identifiable.
- What if Y_i does not vary? Flat line!

Mechanical Properties of the OLS Estimator

National Taiwan University

- The sample mean of the residual is zero: $\frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i = 0$

- The residuals will be uncorrelated with the predictor: $\sum_{i=1}^{n} X_i\widehat{u}_i = 0$

  $\rightarrow \widehat{Cov}(X_i, \widehat{u}_i) = 0$

- The residuals will be uncorrelated with the fitted values: $\sum_{i=1}^{n}\widehat{Y}_i\widehat{u}_i = 0$

  $\rightarrow \widehat{Cov}(\widehat{Y}_i, \widehat{u}_i) = 0$

- You should be able to derive the above results from the PDs.

- We can rewrite $\widehat{\beta}_1 = \sum_{i=1}^{n} W_i Y_i$, where $W_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

  This is why OLS is a linear estimator.

Sampling distribution of the OLS estimator

- Remember, OLS is an estimator -
  it's a machine that we plug samples to get our estimates.

Sample 1: $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ → OLS → $(\widehat{\beta}_0, \widehat{\beta}_1)_1$

Sample 2: $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ → OLS → $(\widehat{\beta}_0, \widehat{\beta}_1)_2$

$\vdots$

Sample $k-1$: $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ → OLS → $(\widehat{\beta}_0, \widehat{\beta}_1)_{k-1}$

Sample k: $\{(Y_1, X_1), \ldots, (Y_n, X_n)\}$ → OLS → $(\widehat{\beta}_0, \widehat{\beta}_1)_k$

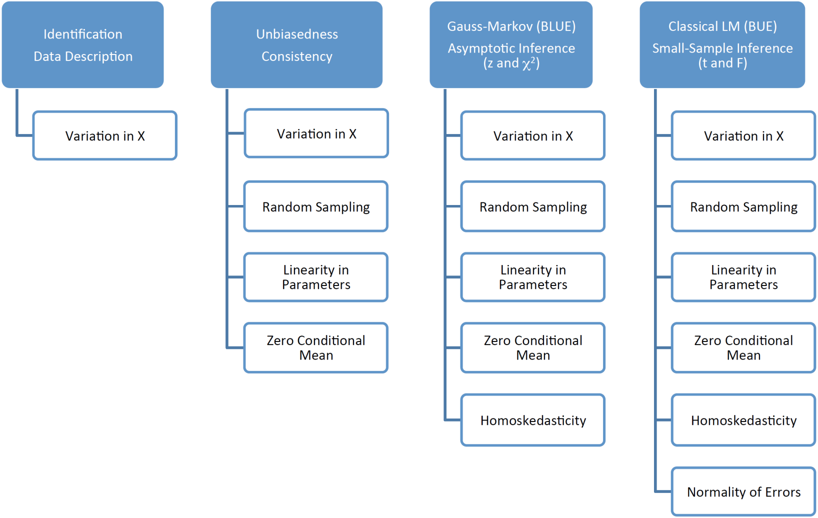
- Just like the sample mean or the sample variance, it has a sampling distribution, with a sampling variance/standard error.

- Now, what is the sampling distribution of the OLS slope? $\widehat{\beta}_1 \sim ?(?, ?)$

Classic Linear Regression Model (CLRM) Assumptions

1. Linearity in Parameters: The population model is linear in its parameters and correctly specified.
2. Random Sampling: The observed data represent a random sample from the population described by the model.
3. Variation in X: There is variation in the explanatory variable.
4. Zero conditional mean: Expected value of the error term is zero conditional on all values of the explanatory variable.
5. Homoskedasticity: The error term has the same variance conditional on all values of the explanatory variable.
6. Normality: The error term is independent of the explanatory variables and normally distributed.

Hierarchy of OLS Assumptions



OLS Assumption 1

- **Assumption I. Linearity in Parameters**:
  The population regression model is linear in its parameters and correctly specified as:
  $Y_i = \beta_0 + \beta_1 X_i + u_i$

- Note that it can be nonlinear in variables
  - OK: $Y_i = \beta_0 + \beta_1 X_i + u_i$ or
    $Y_i = \beta_0 + \beta_1 X_i^2 + u_i$ or
    $Y_i = \beta_0 + \beta_1 log(X_i) + u_i$
  - Not OK: $Y_i = \beta_0 + \beta_1^2 X_i + u_i$ or
    $Y_i = \beta_0 + exp(\beta_1) X_i + u_i$

- β_0, β_1: Population parameters: fixed and unknown.
- u_i: Unobserved random variable with $E[u_i] = 0$
  Usually called the error term or the innovation term.
  Captures all other factors influencing Y_i other than X_i
  VERY IMPORTANT: $u_i \neq \widehat{u_i}$, the error term is not the residual!!

OLS Assumption 2 and 3

National Taiwan University

- **Assumption II. Random Sampling**:
  The observed data: (Y_i, X_i) for $i = 1, 2, ..., n$ represent an *i.i.d.* random sample of size n following the population model.

- **Assumption III. Variation in X**:
  The observed data: X_i for $i = 1, 2, ..., n$ are not all the same value.

- Why do we need this? $\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

OLS Assumption 4

- **Assumption IV. Zero Conditional Mean Error**:
  The expected value of the error term is zero conditional on any value of the explanatory variable: $E[u_i|X_i = x] = 0$ for all x.

- $E[u_i|X] = 0$ implies a slightly weaker condition $Cov(X, u) = 0$
- Given random sampling, $E[u|X] = 0$ also implies $E[u_i|X_i] = 0$ for all i.

OLS Unbiasedness

- With CLRM assumptions 1-4, $E[\widehat{\beta}_1] = \beta_1$!

- Let's prove it!

- Again, with $W_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$, we can show that $\sum_{i=1}^n W_i X_i = 1$

$$
\begin{aligned}
\sum_{i=1}^n W_i X_i &= \sum_{i=1}^n \frac{X_i(X_i - \bar{X})}{\sum_{j=1}^n (X_j - \bar{X})^2} \\
&= \frac{1}{\sum_{j=1}^n (X_j - \bar{X})^2} \sum_{i=1}^n X_i(X_i - \bar{X}) \\
&= \frac{1}{\sum_{j=1}^n (X_j - \bar{X})^2} [\sum_{i=1}^n X_i(X_i - \bar{X}) - \sum_{i=1}^n \bar{X}(X_i - \bar{X})] \\
&= \frac{1}{\sum_{j=1}^n (X_j - \bar{X})^2} \sum_{i=1}^n (X_i - \bar{X})^2 = 1
\end{aligned}
$$

OLS Unbiasedness

$$
\begin{aligned}
\widehat{\beta}_1 &= \sum_{i=1}^{n} W_i Y_i \\
&= \sum_{i=1}^{n} W_i(\beta_0 + \beta_1 X_i + u_i) \\
&= \beta_0 (\sum_{i=1}^{n} W_i) + \beta_1 (\sum_{i=1}^{n} W_i X_i) + \sum_{i=1}^{n} W_i u_i \\
&= (0) + \beta_1 * (1) + \sum_{i=1}^{n} W_i u_i \\
&= \beta_1 + \sum_{i=1}^{n} W_i u_i
\end{aligned}
$$

OLS Unbiasedness

$$
\begin{aligned}
E[\widehat{\beta}_1 - \beta_1 | X] &= E[\sum_{i=1}^{n} W_i u_i | X] \\
&= \sum_{i=1}^{n} E[W_i u_i | X] \\
&= \sum_{i=1}^{n} W_i E[u_i | X] \\
&= \sum_{i=1}^{n} W_i 0 \\
&= 0
\end{aligned}
$$

OLS Consistency

- Recall that $\widehat{\beta}_1 = \beta_1 + \sum\limits_{i=1}^{n} W_i u_i$

- Under iid sampling, we have:

$$\sum_{i=1}^{n} W_i u_i = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \xrightarrow{p} \frac{Cov(X_i, u_i)}{Var(X_i)}$$

- Under A4 (zero conditional mean error), we have the slightly weaker property $Cov(X_i, u_i) = 0$ as long as $Var(X_i) > 0$. We then have $\widehat{\beta}_1 \xrightarrow{p} \beta_1$

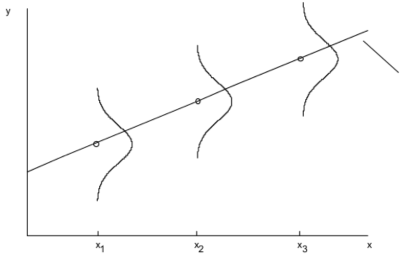
- Filling the blank: $\widehat{\beta}_1 \sim ?(\beta_1, ?)$

OLS Assumption 5

- To derive $Var(\widehat{\beta}_0)$ and $Var(\widehat{\beta}_1)$, let's make the following assumption.

- **Assumption V. Homoskedasticity**:
  The conditional variance of the error term is constant and does not vary as a function of the explanatory variable: $Var(u|X) = \sigma_u^2$

- All errors have an identical error variance, $\sigma_{u_i}^2 = \sigma_u^2$ for all i.



Gauss-Markov Assumption

- Together, Assumptions 1-5 imply:

$$E[Y|X] = \beta_0 + \beta_1 X$$

$$Var[Y|X] = \sigma_u^2$$

- Violation: $Var[u|X = x_1] \neq Var[u|X = x_2]$ is called heteroskedasticity.

- Assumption 1-5 are collectively known as the Gauss-Markov assumptions.

Derive Sampling Variance

$$
\begin{aligned}
Var(\widehat{\beta_1}|X) &= Var(\beta_1 + \sum_{i=1}^{n} W_i u_i|X) \\
&= Var(\sum_{i=1}^{n} W_i u_i|X) \\
&= \sum_{i=1}^{n} W_i^2 Var(u_i|X) \qquad \text{(A2: i.i.d.)} \\
&= \sum_{i=1}^{n} W_i^2 \sigma_u^2 \qquad \text{(A5: homoskedasticity)} \\
&= \sigma_u^2 \sum_{i=1}^{n} (\frac{X_i - \bar{X}}{\sum_{j=1}^{n}(X_j - \bar{X})^2})^2 \\
&= \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}
\end{aligned}
$$

National Taiwan University

$$
\begin{aligned}
Var(\widehat{\beta}_0|X) &= Var(\bar{Y} - \widehat{\beta}_1 \bar{X}|X) \\
&= Var(\bar{Y}|X) + \bar{X}^2 Var(\widehat{\beta}_1|X) - 2\bar{X}Cov(\bar{Y}, \widehat{\beta}_1|X) \\
&= \frac{1}{n^2} Var(\sum_{i=1}^{n} Y_i|X) + \frac{\bar{X}^2 \sigma_u^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} - {\color{red}0} \\
&= \frac{\sigma_u^2}{n} + \frac{\bar{X}^2 \sigma_u^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}
\end{aligned}
$$

- Why do we have 0 for $Cov(\bar{Y}, \widehat{\beta}_1|X)$?

$$
\begin{aligned}
Cov(\bar{Y}, \widehat{\beta}_1|X) &= Cov\{\frac{1}{n}\sum Y_i, \frac{\sum(X_j - \bar{X})Y_j}{\sum(X_i - \bar{X})^2}|X\} \\
&= \frac{1}{n\sum(X_i - \bar{X})^2} Cov\{\sum Y_i, \sum(X_j - \bar{X})Y_j|X\} \\
&= \frac{1}{n\sum(X_i - \bar{X})^2} {\color{red}\sum(X_j - \bar{X})} \sum Cov\{Y_i, Y_j|X\} = {\color{red}0}
\end{aligned}
$$

Variance of OLS Estimators

From Gauss-Markov Theorem:

- $Var(\widehat{\beta}_1|X) = \frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$
- $Var(\widehat{\beta}_0|X) = \sigma_u^2 \{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\}$

where $Var(u|X) = \sigma_u^2$ (the error term variance)

- What drives $Var(\widehat{\beta}_1|X)$?
  - $\sigma_u^2 \nearrow \Rightarrow Var(Y_i|X_i) \nearrow \Rightarrow Var(\widehat{\beta_1}|X) \nearrow$
  - $n \nearrow \Rightarrow \sum_{i=1}^n (X_i - \bar{X})^2 \nearrow \Rightarrow Var(\widehat{\beta_1}|X) \searrow$

Estimating Variance of OLS Estimators

National Taiwan University

- How do we estimate $Var(u|X) = \sigma_u^2$?
- One could try to infer from the residual. $\widehat{u}_i = Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_i$
- Reminder:The error terms u_i are NOT the same as the residuals $\widehat{u}_i$.
- Intuitively, the scatter of the residuals around the fitted regression line should reflect the unseen scatter of the true population regression line.
- We can *first* measure the mean squared deviation of the residual:
$$MSD(\widehat{u}) \equiv \frac{1}{n}\sum_{i=1}^{n}(\widehat{u}_i - \bar{\widehat{u}}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i^{\,2}$$

Estimating Variance of OLS Estimators

National Taiwan University

- We will later on (in the multivariate regression session) see that under the univariate regression specification,

$$E[MSD(\widehat{u})] = \frac{n-2}{n}\sigma_u^2$$

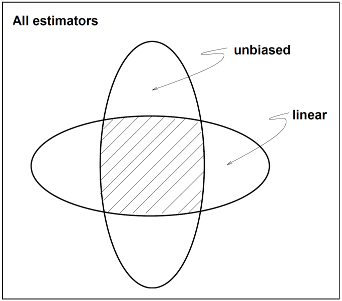
- The 2 in the numerator $n-2$ is the *degree of freedom* in our univariate regression model.

- As a consequence, an unbiased estimator for the error variance is:

$$\widehat{\sigma}_u^2 = s^2 = \frac{n}{n-2}MSD(\widehat{u}) = \frac{n}{n-2}\frac{1}{n}\sum_{i=1}^{n}\widehat{u_i}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\widehat{u_i}^2$$

OLS is BLUE :(

- Now, under CLRM assumptions 1-5, $\widehat{\beta}_1 \sim ?(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \bar{X})^2})$
- **Gauss-Markov Theorem**:
  Given CLRM assumptions 1-5, OLS is BLUE
  - B est: Lowest variance in class
  - L inear: Among linear estimators
  - U nbiased: Among linear unbiased estimators
  - E stimator



OLS Distribution

National Taiwan University

- According to Central Limit Theorem:
  The sums and means of random variables tend to be *normally* distributed in large samples.

- Therefore, in large sample, $\frac{\widehat{\beta_1}-\beta_1}{SE(\widehat{\beta_1})} \sim N(0,1)$

- Now we know $\widehat{\beta}_1 \sim ?(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2})$ and $\widehat{\sigma}_u^2 = s^2 = \frac{1}{n-2}\sum_{i=1}^{n}\widehat{u_i}^2$,
  can you calculate $SE(\widehat{\beta_1})$ in our CAPM example by hand?
  Do we have a large sample here?

OLS Assumption 6

- What if we don't have large sample?
- **Assumption VI. Normality**:
  The population error term is independent of the explanatory variable, $u \perp X$, and is normally distributed with mean zero and variance σ_u^2
  $u \sim N(0, \sigma_u^2)$ also implies that $Y|X \sim N(\beta_0 + \beta_1 X, \sigma_u^2)$
- *Note: This also implies homoskedasticity and zero conditional mean.*

- The CLRM 1-6 assumptions imply that OLS is BUE
  (minimum variance among all linear and non-linear unbiased estimators)

- Non-normality of the errors is a serious concern in small samples. We can partially check this assumption by looking at the residuals.

- *Something nice to know:*
  Hansen, B. (2022). A Modern Gauss-Markov Theorem. *Econometrica, forthcoming.* finds that A1-A5 $\Rightarrow$ OLS is BUE!

Sampling distribution of OLS slope

National Taiwan University

- If we have Y_i given X_i is distributed $N(\beta_0 + \beta_1 X_i, \sigma_u^2)$, then we have the following at any sample size:

$$\frac{\widehat{\beta}_1 - \beta_1}{SE(\widehat{\beta}_1)} \sim N(0,1)$$

- Furthermore, if we replace the true standard error with the estimated standard error, then we get the following:

$$\frac{\widehat{\beta}_1 - \beta_1}{\widehat{SE}(\widehat{\beta}_1)} \sim t_{n-2}$$

- The standardized coefficient follows a t distribution $n-2$ degrees of freedom. We take off an extra degree of freedom because we had to estimate one more parameter than just the sample mean.

Quick Summary

National Taiwan University

- Under Assumptions 1-5 and in large samples, we know that

$$\widehat{\beta}_1 \sim N(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2})$$

- Under Assumptions 1-6 and in any sample, we know that

$$\frac{\widehat{\beta}_1 - \beta_1}{SE(\widehat{\beta}_1)} \sim t_{n-2}$$

Quick Summary

- Identification / Data Description
  - Variation in X
- Unbiasedness / Consistency
  - Variation in X
  - Random Sampling
  - Linearity in Parameters
  - Zero Conditional Mean
- Gauss-Markov (BLUE) / Asymptotic Inference (z and χ^2)
  - Variation in X
  - Random Sampling
  - Linearity in Parameters
  - Zero Conditional Mean
  - Homoskedasticity
- Classical LM (BUE) / Small-Sample Inference (t and F)
  - Variation in X
  - Random Sampling
  - Linearity in Parameters
  - Zero Conditional Mean
  - Homoskedasticity
  - Normality of Errors

Quick Discussion

National Taiwan University

- **Regression as parametric modeling**:
  Let's summarize the parametric view we have taken thus far.
- Gauss-Markov assumptions:
  - (A1) linearity, (A2) i.i.d. sample, (A3) variation in X, (A4) zero conditional mean error, (A5) homoskedasticity.
  - basically, assume the model is right.
- $\Rightarrow$ OLS is BLUE,
  plus (A6) normal errors and we get small sample SEs and BUE.
- What is the basic approach here?
  - A1 defines a linear model for the outcome:
    $Y_i = \beta_0 + \beta_1 X_i + u_i$
  - A2 and A4 let us write the CEF as function of Xi alone.
    $E[Y_i|X_i] = \mu_i = \beta_0 + \beta_1 X_i$
  - A3 guarantees that the βs are identifiable.
  - A5-A6 define a probabilistic model for the conditional distribution:
    $Y_i|X_i \sim N(\mu_i, \sigma_u^2)$

Quick Discussion

National Taiwan University

- **Agnostic views on regression**
- These assumptions assume we know a lot about how Y_i is generated.
- Justifications for using OLS (like BLUE/BUE) often invoke these assumptions which are unlikely to hold exactly.
- Alternative: take an agnostic view on regression.
  - Use OLS without believing these assumptions.
  - Lean on two things: A2 i.i.d. sample, asymptotics (large-sample properties)
- Lose the distributional assumptions and focus on approximating the best linear predictor.
- If the true CEF happens to be linear, the best linear predictor is it.

Quick Discussion

- **Unbiasedness**
- One of the results most people know is that OLS is unbiased, but unbiased for what?
- It is unbiased for the CEF under the assumption that the model is correctly specified.
- However, this could be a quite poor approximation to the true CEF if there is a great deal of non-linearity.
- We will often use OLS as a means to approximate the CEF, but don't forget that it is just an approximation!
- We will come back and revisit some of the assumptions when we get to the multivariate regression session.

Inference: Test Statistic

- Under the null of H_0: $\beta_1 = c$, we can use the following test statistic:

$$\frac{\widehat{\beta}_1 - c}{SE(\widehat{\beta}_1)}$$

- Under the null hypothesis:
  - large samples: $T \sim N(0,1)$
  - any size sample with normal errors: $T \sim t_{n-2}$
  - conservative to use t_{n-2} anyways since t_{n-2} is approximately normal in large samples.

- Under the null, we know the distribution of T and can use that to formulate a rejection region and calculate p-values.

- Choose a level of the test, α, and find rejection regions that correspond to that value under the null distribution:

$$Pr(-t_{\alpha/2,n-2} < T < t_{\alpha/2,n-2}) = 1 - \alpha/2$$

Inference: p-Value and CI

- The interpretation of the p-value is the same:
  the probability of seeing a test statistic at least this extreme if the null hypothesis were true.

- Mathematically:

$$Pr(|\frac{\widehat{\beta}_1 - c}{SE(\widehat{\beta}_1)}| \geq |T_{obs}|)$$

- If the p-value is less than α we would reject the null at the $1-\alpha$ confidence level.

- Confidence Intervals: $\widehat{\beta}_1 \pm t_{\alpha/2,n-2} * SE(\widehat{\beta}_1)$

Goodness of Fit (R^2)

- How do we judge how well a line fits the data?

- One way is to find out how much better we do at predicting Y once we include X into the regression model.

- Prediction errors without X: best prediction is the mean, so our squared errors, or the total sum of squares ($\mathrm{SS}_{\mathrm{tot}}$) would be:

$$SS_{tot} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

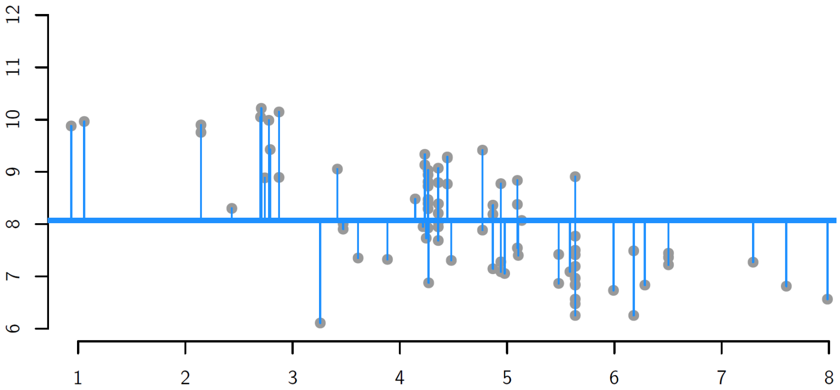
- Once we have estimated our model, we have new prediction errors, which are just the sum of the squared residuals or $\mathrm{SS}_{\mathrm{res}}$:

$$SS_{res} = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$$

Goodness of Fit (R^2)

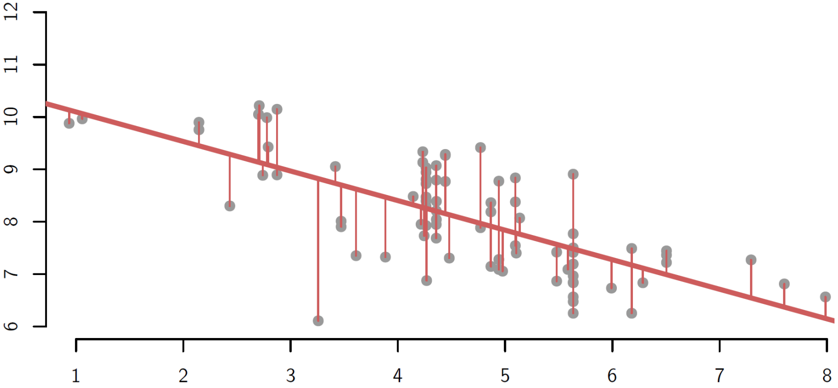
National Taiwan University

- Total Sum of Squares



Goodness of Fit (R^2)

- Residual Sum of Squares



R-Square

- By definition, the residuals have to be smaller than the deviations from the mean, so we might ask the following: how much lower is the SS_{res} compared to the SS_{tot}?

- We quantify this question with the coefficient of determination or R^2.

$$R^2 = \frac{SS_{tot} - SS_{res}}{SS_{tot}} = 1 - \frac{SS_{res}}{SS_{tot}}$$

- This is the fraction of the total prediction error eliminated by providing information on X.

- Alternatively, this is the fraction of the variation in Y is explained by X.

- $R^2 = 0$: no relationship.
- $R^2 = 1$: perfect linearity (**Démon de Laplace**)

R-Square

- Some more interesting derivations.
  - Define $SS_{tot} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$ (Total Sum of Squares)
  - Define $SS_{res} = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$ (Residual Sum of Squares)
  - Define $SS_{exp} = \sum_{i=1}^{n} (\widehat{Y}_i - \bar{Y})^2$ (Explained Sum of Squares)
- Quick takeaway:

$$SS_{tot} = SS_{res} + SS_{exp}$$

and

$$R^2 = \frac{SS_{tot} - SS_{res}}{SS_{tot}} = \frac{SS_{exp}}{SS_{tot}}$$

R-Square

$$
\begin{aligned}
SS_{tot} &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \\
&= \sum_{i=1}^{n}[(Y_i - \widehat{Y}_i) + (\widehat{Y}_i - \bar{Y})]^2 \\
&= SS_{res} + SS_{exp} + 2\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)(\widehat{Y}_i - \bar{Y})
\end{aligned}
$$

$$
\begin{aligned}
\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)(\widehat{Y}_i - \bar{Y}) &= \sum_{i=1}^{n}\widehat{u}_i(\widehat{\beta}_0 + \widehat{\beta}_1 X_i - \bar{Y}) \\
&= (\widehat{\beta}_0 - \bar{Y})\sum_{i=1}^{n}\widehat{u}_i + \widehat{\beta}_1\sum_{i=1}^{n}\widehat{u}_i X_i \\
&= \underbrace{(0)}_{\sum \widehat{u_i} = 0} + \underbrace{(0)}_{\sum X_i \widehat{u_i} = 0} = 0
\end{aligned}
$$

R-Square v.s. Correlation

National Taiwan University

- Lastly, $R^2 = \sigma^2_{XY} = \sigma^2_{\widehat{Y}Y}$

- Why? First: $\bar{\widehat{Y}} \equiv \frac{1}{n}\sum \widehat{Y}_i = \frac{1}{n}\sum \widehat{\beta}_0 + \frac{1}{n}\sum \widehat{\beta}_1 X_i = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{X} = \bar{Y}$

- So, we can rewrite
$$SS_{exp} = \sum_{i=1}^{n}(\widehat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n}(\widehat{Y}_i - \bar{\widehat{Y}})^2 = (n-1)Var(\widehat{Y})$$

- $R^2 = \frac{SS_{exp}}{SS_{tot}} = \frac{(n-1)Var(\widehat{Y})}{(n-1)Var(Y)} = \frac{Var(\widehat{\beta_0} + \widehat{\beta_1}X)}{Var(Y)} = \widehat{\beta}_1{}^2 \frac{Var(X)}{Var(Y)}$
$= (\frac{Cov(X,Y)}{Var(X)})^2 \frac{Var(X)}{Var(Y)} = \frac{Cov(X,Y)^2}{Var(X)Var(Y)} = (\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}})^2 = \sigma^2_{XY}$

- This result is very useful when we are dealing with non-linear regressions (e.g. *Probit* or *Logit*) that do not have R^2.
$\rightarrow$ We can instead calculate $\sigma^2_{\widehat{Y}Y}$. This statistic is called Pseudo-R^2.

Multivariate Regression

National Taiwan University

- There are a few things we have not covered in univariate regression.
  - The general form of OLS.
  - Prove of BLUE and BUE of OLS under Gauss-Markov.
  - Why $\frac{1}{n-2}\sum_{i=1}^{n}\widehat{u_i}^2$ is an unbiased estimator for σ_u^2
  - How to deal with heteroskedasticity when we do not have A5.
  - Why sample mean is also a least square estimator.
  - Relationship between multiple independent (explanatory) variables.
  - Hypothesis testing of multiple coefficient estimates (and the entire model).
- We will first review for some matrix properties and show you the general multivariate regression derivations.