CHULALONGKORN BUSINESS SCHOOL
FLAGSHIP FOR LIFE

AACSB ACCREDITED · EFMD EQUIS ACCREDITED · EFMD EPAS ACCREDITED

MSF Chula

Financial Econometrics

Lecture VII:
Maximum Likelihood Estimator & Limited Dependent Variable Models

Narapong Srivisal, Ph.D.

Outline

Maximum Likelihood Estimator (MLE) skip :(

Limited Dependent Variable Models

- **Binary Responses** $y \to 1$, $y \to 0$
  - Linear Probability Model
  - Probit Model
  - Logit Model
- **Censored Responses** cannot OLS, it will be biased
  - Tobit Model

Maximum Likelihood Estimator

- The idea of the **Maximum Likelihood Estimator (MLE)** is to find $\hat{\theta}$, an estimator for parameter θ, that makes our observed data set most likely to happen
- Suppose that $X_1, \dots X_n \sim iid\ X$ which has the PDF $f(X;\theta)$ or the PMF $P(X;\theta)$, where θ is an unknown parameter of interest.
- Then, the likelihood of observing $X_i = x$ is the density at x, which is $f(x;\theta)$ or the probability that $X_i = x$, which is $P(x;\theta)$
- If our sample is from an iid distribution, the probability of observing $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ is simply the joint probability

$$P\{X_1 = x_1, \dots, X_n = x_n\} = P\{X_1 = x_1\} \times \cdots \times P\{X_n = x_n\}$$

- **The likelihood of observing our iid dataset** is

$$L(\theta; X) = \prod_{i=1}^{n} f(x_i;\theta)\ or \prod_{i=1}^{n} P(x_i;\theta)$$

- Hence, the Maximum Likelihood Estimator (MLE) is then

$$\hat{\theta}_n^{MLE} = \text{argmax}_\theta \left\{ \prod_{i=1}^{n} f(x_i;\theta) \right\} \;\; or \;\; \text{argmax}_\theta \left\{ \prod_{i=1}^{n} P(x_i;\theta) \right\}$$

- A likelihood function is always positive as it's multiplication of density and is often in the form of power of a density function, which is quite difficult to find the FOC and solve for the solution.
- Since $\ln(.)$ is a strictly monotone (increasing) function, the maximum of $\ln\big(f(X)\big)$ is the same as the maximum of $f(X)$ for any function $f(X) > 0$.
- So, when finding a MLE, people often transform a likelihood function to a log-likelihood function, which is in additive form and easier to take derivative

$$\hat{\theta}_n^{MLE} = \underset{\theta}{\text{argmax}} \left\{ \ln\left(\prod_{i=1}^{n} P(X_i;\theta) \right) \right\} = \underset{\theta}{\text{argmax}} \left(\sum_{i=1}^{n} \ln\big(P(X_i;\theta)\big) \right)$$

Example

- Suppose $X_1, \dots, X_n \sim Bernoulli(p)$.
- The probability of observing $X_i = 1$ is p and the probability of observing $X_i = 0$ is $1 - p$. If our dataset contains m samples taking value 1, then the likelihood is

$$L(p; X) = p^m (1-p)^{n-m}$$

- The Maximum Likelihood Estimator is the solution to $\max_p \{p^m (1-p)^{n-m}\}$
- The FOC is

$$mp^{m-1}(1-p)^{n-m} + p^m[(n-m)(1-p)^{n-m-1}(-1)] = 0$$

$$m(1-p) + p(n-m)(-1) = 0$$

- Or the log-likelihood function is

$$l(p; X) = \ln(p^m (1-p)^{n-m}) = m \ln p + (n-m) \ln(1-p)$$

- The FOC is

$$\frac{m}{p} - \frac{n-m}{1-p} = 0 \;\Rightarrow\; m(1-p) - p(n-m) = 0$$

- The solution is

$$\hat{p}^{MLE} = \frac{m}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Example

- Consider the linear regression model with the assumption that the error term is Normally Distributed:

$$Y_i = \beta_0 + \beta_1 X_i + U_i,$$

$$U_i \sim iid\ N(0, \sigma^2)$$

- Recall that the PDF of U_i is $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{U_i^2}{2\sigma^2}\right\}$
- So, the log-likelihood is

$$l(\beta_0, \beta_1; X, Y) = \sum_{i=1}^{n} \left\{-\ln\sqrt{2\pi} - \frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(Y_i - \beta_0 - \beta_1 X_i)^2\right\}$$

- Therefore, the likelihood estimators for β_0, β_1 are

$$\underset{\beta_0,\beta_1}{\operatorname{argmax}} \sum_{i=1}^{n} \{-(Y_i - \beta_0 - \beta_1 X_i)^2\} = \underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{n} \{(Y_i - \beta_0 - \beta_1 X_i)^2\}$$

- Notice that this is the to minimize sum of squares of the errors, which is the same as the OLS estimator.

Limited Dependent Variable Models

Limited Dependent Variable

- In some situations, we may be interested in estimating models in which the dependent variable or regressand can take values only within a limited range.

  y — 1 invest
  0 not invest
  binary respond

- E.g. **to explain the probability** that an individual has a brokerage account, our regressand is a binary random variable indicating whether an individual has a brokerage account.
- For such case, we will consider a **binary response model**.

  there's a lot of individuals that don't invest

- E.g. to study determinants of individual's level investment in stock market, our sample may contain a number of individuals with zero investment which implicitly means there is a **threshold** for individuals to enter the financial market.

  censored people

- For such case, we will consider a **censored response model**.

$y \sim$ Beunoulli

$E[y] = 1 \cdot p + 0 \cdot (1-p) = p$

$E[y|x] = E[\beta x + u|x]$ assume $\to 0$

don't have to do anything

$\Pr\{y=1|x\} = \beta x + E[u|x]$

$\Rightarrow$ OLS unbiased & consistent

$\Downarrow$ LINEAR PROBABILITY MODEL (LPM)

$\Pr\{y=1|x\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$

when interpret $\beta \Rightarrow \beta_1$: $x_1 \uparrow 1$ unit $\to$ $y \uparrow \beta_1$ unit

$\Pr\{y=1|x\} \uparrow$ unit

want to study prob not y itself

* drawback : $\hat{\beta} x \Rightarrow$ may not in $[0, 1]$

no restriction for y to be 0 or 1

Binary Response Model

- Consider a simple model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + U_i$$

- If Y is a binary random variable, then $E[Y] = P\{Y = 1\}$.
- So, if we make an assumption that $E[U|X] = 0$, then

$$P\{Y = 1|X\} = E[Y|X] = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$$

- and we can interpret the predicted value

$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_k X_{ki}$$

as the predicted probability of Y_i conditional on the regressors.

- This is called the **Linear Probability Model (LPM)**, which can be estimated by the OLS
- A big problem with this model is that the predicted value $\widehat{Y}_i$ can be less than zero or greater than one, especially when we use out of sample data.

$$Pr\{Y=1 \mid X\} = G(\beta X)$$

return value in $(0,1)$ ← give value $(0,1)$

CDF (cumulative distribution function)

model:

① CDF of standard normal ≡ PROBIT model

$\Phi(\beta X)$

② CDF of logistic distribution ≡ LOGIT model

- A better way to model binary response dependent variable is to find **a function $G(.)$ which has range $[0, 1]$** and assume

no error term in this

$$\boldsymbol{P\{Y_i = 1|X_i\} = G(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) = G(X_i\beta)}$$

- This is the same idea as having a **latent variable Y^***, which is not observed and determined by X and unobserved factor U such that Y would take value 1 if $Y^* > 0$

one way to write model formally

- More precisely, let

$$Y_i^* = X_i\beta + U_i \text{ and}$$
$$Y_i = I\{Y_i^* > 0\}$$

- Note that if $G(.)$ is a cdf of U which is a symmetric distribution,

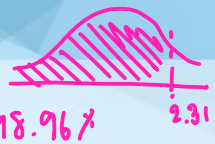
$$P\{Y_i = 1|X_i\} = P\{Y_i^* > 0|X_i\} = P\{X_i\beta + U_i > 0|X_i\}$$
$$= P\{U_i > -X_i\beta|X_i\} = 1 - G(-X_i\beta) = G(X_i\beta)$$

cdf function of $X_i\beta$

- Since this is a non-linear model, we can apply the MLE technique to estimate the parameter β's.

e.g. $X\hat{\beta} = 2.31$

$pr\{y=1|x\} = 98.96\%$



Probit Model

- **Probit model assume that $U_i \sim iid\ N(0,1)$**; so, the model is

$$Y_i^* = X_i\beta + U_i; Y_i = 1\ if\ Y_i^* > 0$$

CDF standard normal ⇒ use z-table to find prob

$$\boldsymbol{P\{Y_i = 1|X_i\} = \Phi(X_i\beta)}$$

- Since $P\{Y_i = 1|X_i\} = \Phi(X_i\beta)$ and $P\{Y_i = 0|X_i\} = 1 - \Phi(X_i\beta)$, then the likelihood of observation i is

$$L(\beta; X_i, Y_i) = P(Y_i|X_i;\beta) = \Phi(X_i\beta)^{Y_i}[1 - \Phi(X_i\beta)]^{1-Y_i}$$

- Or the log-likelihood is

$$l(\beta; X_i, Y_i) = Y_i \ln \Phi(X_i\beta) + (1 - Y_i)\ln\big(1 - \Phi(X_i\beta)\big)$$

- Therefore, the Probit estimator, which is based on the MLE framework, is

$$\hat{\beta}^{Probit} = \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{n} \{Y_i \ln \Phi(X_i\beta) + (1 - Y_i)\ln\big(1 - \Phi(X_i\beta)\big)\}$$

- There is no analytical closed-form expression for $\hat{\beta}^{Probit}$, but we have to calculate the value of Probit estimates numerically.

Problem set 2 : ข้อ 5

Logit Model

$X\hat{\beta} \Rightarrow 2.31 \Rightarrow Pr\{y=1\} = \frac{e^{2.31}}{1+e^{2.31}} = 90.97\%$

- Logit model assume that the function $G(.)$ takes form of the Logistic function:

$$\boldsymbol{P\{Y=1|X\} = \Lambda(X\beta) = \frac{\exp(X\beta)}{1+\exp(X\beta)}}$$

- To estimate the model, we can apply the MLE technique similar to the Probit estimator:

$$\hat{\beta}^{Logit} = \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{n} \{Y_i \ln[\Lambda(X_i\beta)] + (1-Y_i)\ln\big(1-\Lambda(X_i\beta)\big)\}$$

Prob of IPO $\longleftarrow$ ESG

x

$y = \begin{cases} 1 & \text{IPO} \\ 0 & \text{not} \end{cases}$

logit y small medium

model: $Pr\{y=1 \mid x\} = \underset{\text{logistic function}}{\Lambda}(\beta_0 + \beta_1 S + \beta_2 M)$

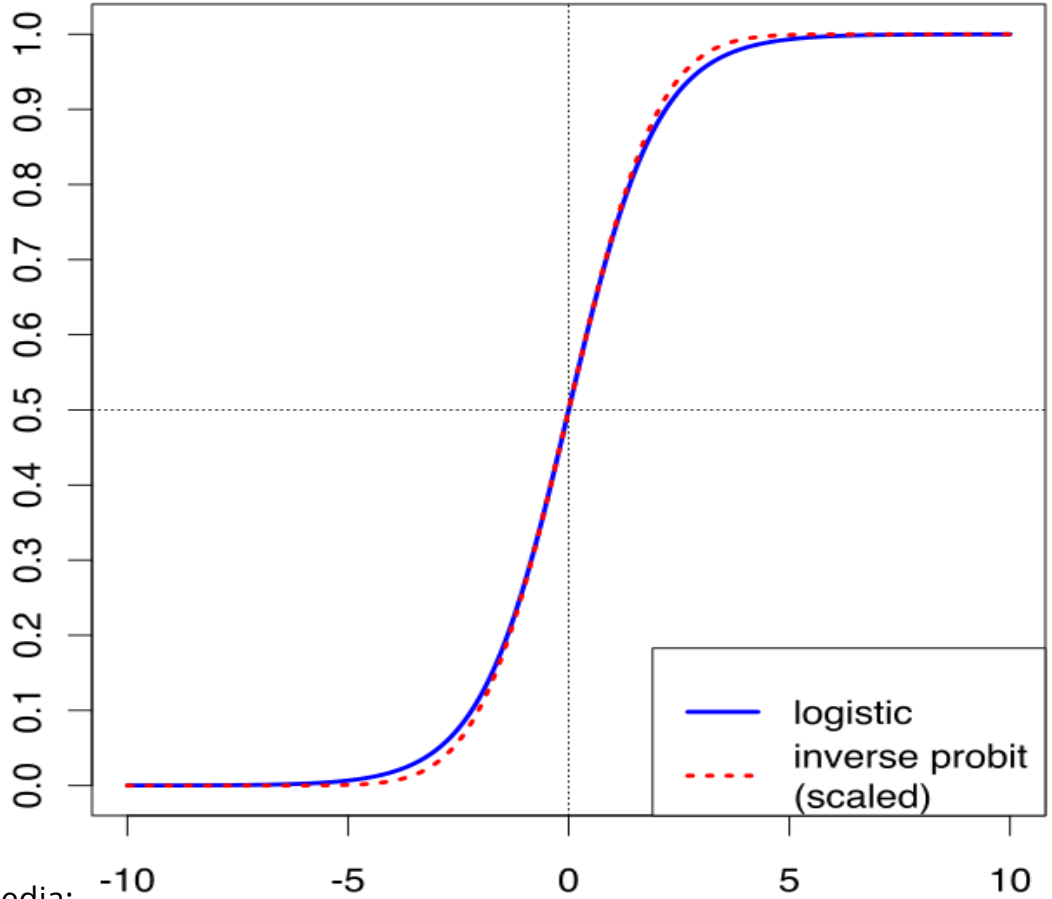
the one that doesn't have dummy is base case

small $\Rightarrow \dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$

medium $\Rightarrow \dfrac{e^{\beta_0+\beta_2}}{1+e^{\beta_0+\beta_2}}$

large $\Rightarrow \dfrac{e^{\beta_0}}{1+e^{\beta_0}}$

$\Phi(.)$ vs $\Lambda(.)$



Picture from wikipedia:
http://upload.wikimedia.org/wikipedia/commons/thumb/f/f1/Logistic-sigmoid-vs-scaled-probit.svg/300px-Logistic-sigmoid-vs-scaled-probit.svg.png

Marginal Effect and Interpretation of β's

- Consider the model:

$$P\{Y = 1|X\} = G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \ln X_3)$$

- The marginal *ceteris paribus* effect of X_1 on $P\{Y = 1|X\}$ is

$$\frac{\partial P\{Y = 1|X\}}{\partial X_1} = g(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \ln X_3)\beta_1$$

- Since $G(.)$ is a CDF, then $g(.)$ is its corresponding PDF:

  - **Probit:** $\boldsymbol{g(X\beta) = \frac{1}{\sqrt{2\pi}} exp\{-(X\beta)^2\}}$
  - **Logit:** $\boldsymbol{g(X\beta) = \frac{exp(X\beta)}{(1+exp(X\beta))^2}}$

formula of PDF

- PDF is greater than zero. So, the marginal effect of X_1 has the same direction as β_1
- Notice the effect is not just β_1 but also weighted by the density or likelihood of the observed point. The higher the likelihood, the higher the effect.

- The marginal effect of X_2 on $P\{Y = 1|X\}$ is

$$\frac{\partial P\{Y = 1|X\}}{\partial X_2} = g(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \ln X_3)\beta_2$$

- Hence, the effect of X_1 relative to the effect of X_2 is

$$\frac{g(\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 \ln X_3)\beta_1}{g(\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 \ln X_3)\beta_2} = \frac{\beta_1}{\beta_2}$$

- The marginal effect of X_3 on $P\{Y = 1|X\}$ is

$$\frac{\partial P\{Y = 1|X\}}{\partial X_3} = g(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \ln X_3)\frac{\beta_3}{X_3}$$

$$\partial P\{Y = 1|X\} = g(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \ln X_3)\beta_3 \frac{\partial X_3}{X_3}$$

- So, the marginal effect of increasing X_3 by 1% is

$$g(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \ln X_3)\beta_3/100$$

want to estimate probit by MLE

Estimating Probit Model

y x_1 x_2 x_3 x_4 x_5 x_6

```
. probit married age weight educ motheduc fatheduc lwage

Iteration 0:   log likelihood = -1314.8424
Iteration 1:   log likelihood = -1217.7696
Iteration 2:   log likelihood = -1217.2746
Iteration 3:   log likelihood = -1217.2746

Probit regression                               Number of obs     =      2,220
                                                LR chi2(6)        =     195.14
                                                Prob > chi2       =     0.0000
Log likelihood = -1217.2746                     Pseudo R2         =     0.0742

------------------------------------------------------------------------------
     married |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0749375   .0106561     7.03   0.000     .0540518    .0958231
      weight |   2.13e-07   1.94e-07     1.10   0.273    -1.68e-07    5.94e-07
        educ |  -.0678883    .013566    -5.00   0.000    -.0944771   -.0412994
    motheduc |  -.0226065   .0127978    -1.77   0.077    -.0476897    .0024766
    fatheduc |  -.0057566   .0107569    -0.54   0.593    -.0268396    .0153265
       lwage |   .5663235   .0744527     7.61   0.000     .4203989    .7122482
       _cons |  -3.868142    .455581    -8.49   0.000    -4.761064    -2.97522
------------------------------------------------------------------------------
```

either logit or probit give same result

ไม่ต่างกันมาก

Estimating Logit Model

```
. logit married age weight educ motheduc fatheduc lwage

Iteration 0:   log likelihood = -1314.8424
Iteration 1:   log likelihood =  -1217.966
Iteration 2:   log likelihood = -1215.3117
Iteration 3:   log likelihood = -1215.3101
Iteration 4:   log likelihood = -1215.3101

Logistic regression                               Number of obs     =      2,220
                                                  LR chi2(6)        =     199.06
                                                  Prob > chi2       =     0.0000
Log likelihood = -1215.3101                       Pseudo R2         =     0.0757

------------------------------------------------------------------------------
     married |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .1329159   .0187288     7.10   0.000     .0962082    .1696236
      weight |   3.50e-07   3.31e-07     1.06   0.290    -2.99e-07    9.99e-07
        educ |  -.1162162   .0232513    -5.00   0.000    -.1617879   -.0706445
    motheduc |  -.0426609   .0219099    -1.95   0.052    -.0856036    .0002818
    fatheduc |   -.008433    .018032    -0.47   0.640    -.0437751    .0269091
       lwage |   .9677625    .127301     7.60   0.000     .7182572    1.217268
       _cons |  -6.735347   .7900325    -8.53   0.000    -8.283782   -5.186911
------------------------------------------------------------------------------
```

ignore this

likelyhood

F-test : H_0 : all $\beta = 0$

H_1 : some $\beta \neq 0$

not simple R^2 like in linear regression

[0,1]

higher : better fit

$\hat{\beta}$ $se(\hat{\beta})$ $\frac{\hat{\beta}}{se(\hat{\beta})}$

Marginal effect

CDF whether logit or probit

$$Pr\{y=1|x\} = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \ln x_3)$$

What is the effect of x_1 on chance that $y=1$?

x_2

diff with respect to x_1

chain rule ⇒ diff นอก, diff ใน

$$\frac{\partial}{\partial x_1} Pr\{y=1|x\} = \underbrace{PDF(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \ln x_3)}_{} \times \beta_1$$

↳ value of PDF will be diff on diff people

always ⊕

↳ ⊕ $x_1 \uparrow \rightarrow y \uparrow$

$x_1 \downarrow \rightarrow y \downarrow$

↳ ⊖ $x_1 \uparrow \rightarrow y \downarrow$

$$\frac{\partial}{\partial x_2} Pr\{y=1|x\} = PDF(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \ln x_3) \times \beta_2$$

(exactly the same for same person ⇒ can compare β_1 & β_2 if we want to compare effect of x_1 verous x_2

Normal vs Logistic distribution

similar mean, shape (0, bell)

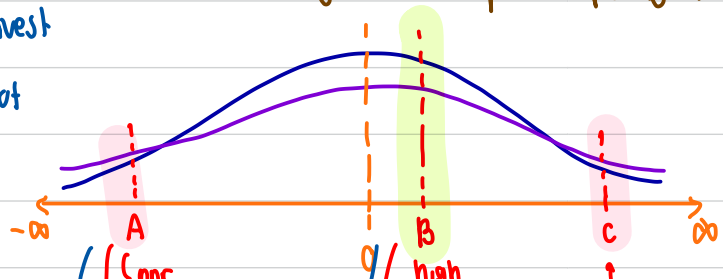
diff in tail → logistic has thicker tail than normal

e.g. $y = \begin{cases} 1 & \text{pass} \\ 0 & \text{fail} \end{cases}$ x : how hard study

if increase x by 1 hour → prob of passing ↑ ?%

effect of x on prob $y=1$ = $\underline{pdf(x\beta)\cdot\beta}$

$y \begin{cases} 1 & \text{invest} \\ 0 & \text{not} \end{cases}$



A: PDF small, read doesn't help; very low income, not gonna invest anyway

B: high effect is higher for people around mean; average student read a little bit → pass; add income → invest

C: gonna pass anyway, x↑ doesn't help; already invest

* margina effect at tail is higher for logistic compared to normal

Estimating marginal effects at the sample mean

```
. mfx, predict(p)
```

```
Marginal effects after logit
      y  = Pr(married) (predict, p)
         =  .74206166
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |   .0254409     .00352    7.24   0.000    .018551    .03233   27.9198
  weight |   6.71e-08     .00000    1.06   0.290   -5.7e-08   1.9e-07    341196
    educ |  -.0222445     .00442   -5.03   0.000   -.030913  -.013576   13.6275
motheduc |  -.0081656     .00419   -1.95   0.051   -.016377   .000046   10.6081
fatheduc |  -.0016141     .00345   -0.47   0.640   -.008379   .005151   10.0788
   lwage |   .1852357     .02416    7.67   0.000    .137891    .23258   6.28542
------------------------------------------------------------------------------
```

marginal effect

value of x to estimate marginal effect

pdf • β

- Note: $-\frac{0.0254409}{0.0222445} = -\frac{0.1329159}{0.1162162} = \frac{\hat{\beta}^{age}}{\hat{\beta}^{edu}}$

Estimating marginal effects

for an average person with 21 years of education and of age 30 years old

```
. mfx, predict(p) at(age=30, edu=21)

warning: no value assigned in at() for variables weight motheduc fatheduc lwage;
   means used for weight motheduc fatheduc lwage

Marginal effects after logit
      y  = Pr(married) (predict, p)
         =  .61689609
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |   .0314127     .00434    7.24   0.000   .022909   .039916      30
  weight |   8.28e-08     .00000    1.05   0.292  -7.1e-08  2.4e-07   341196
    educ |   -.027466     .00656   -4.19   0.000  -.040319 -.014613      21
motheduc |  -.0100823     .00512   -1.97   0.049   -.02011 -.000055 10.6081
fatheduc |   -.001993     .00424   -0.47   0.639  -.010313  .006327 10.0788
   lwage |   .2287164     .03322    6.88   0.000   .163604  .293829 6.28542
------------------------------------------------------------------------------
```

chance of getting marriage ↑

whose see effect

for this specific person

Censored Response Model

- The regressand can take on continuous values over some range but is **constrained or censored to be a particular value for a substantial part of the population.**
- Mathematically, for censored from below a particular value c, we can write this model using a latent variable Y^* as follows:

$$Y_i^* = X_i\beta + U_i, \text{ and}$$
$$Y_i = \max(c, Y_i^*)$$

- So, $Y = c$ if $Y^* = X\beta + U \leq c \Rightarrow U \leq c - X\beta$
- Or if censored from above:

$$Y_i^* = X_i\beta + U_i, \text{ and}$$
$$Y_i = \min(c, Y_i^*)$$

- So, $Y = c$ if $Y^* = X\beta + U \geq c \Rightarrow U \geq c - X\beta$

Tobit Model

- This model was originally proposed by Tobin (1958) to analyze households' expenditures on durable goods like a car.
- Tobin noticed that quite a large group of population spent zero dollars on durable goods, because they didn't have adequate income, not that they didn't value the goods.
- **In this Tobit Model, there is a strong assumption that the error term is iid normally distributed and homoskedastic**:

income

$$Y_i^* = X_i\beta + U_i, \text{where } \boldsymbol{U_i|X_i \sim iid\ N(0, \sigma^2)}$$

willingness to spend

$$Y_i = \max(c, Y_i^*)$$

actual spending

real spending

if Y_i^* is not high enough

- Again, to estimate the model, we can apply the MLE technique
- The likelihood of observing $Y_i = c$ is

$$P\{Y_i = c|X_i\} = P\{U_i < c - X_i\beta|X_i\} = \Phi\left(\frac{c - X_i\beta}{\sigma}\right)$$

- For $Y_i > c$, $Y_i^* = Y_i \Rightarrow U_i = Y_i - X_i\beta$. Since U_i is normally distributed, then the likelihood is the normal density:

$$L(\beta, \sigma; X_i, Y_i) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left\{-\frac{1}{2}\left(\frac{Y_i - X_i\beta}{\sigma}\right)^2\right\} = \frac{1}{\sigma}\phi\left(\frac{Y_i - X_i\beta}{\sigma}\right)$$

- Therefore, the log likelihood of the sample is

$$\sum_{i=1}^{n}\left[I\{Y_i = c\}\ln\Phi\left(\frac{c - X_i\beta}{\sigma}\right) + I\{Y_i > c\}\ln\left(\frac{1}{\sigma}\phi\left(\frac{Y_i - X_i\beta}{\sigma}\right)\right)\right]$$

- Again, there is no closed form solution; an estimated value of MLE can be computed using numerical method.
- Note: the Tobit estimator is based on MLE and the strong assumption that $U|X \sim N(0, \sigma^2)$. **MLE is consistent if the normality assumption is roughly true.** But if $U|X$ is very different from normal, then Tobit will not give a good estimator for β

gov wants to increase level of inv in stock mkt
(want to see effect by give 10k Baht income
how much on average inv in stock increase

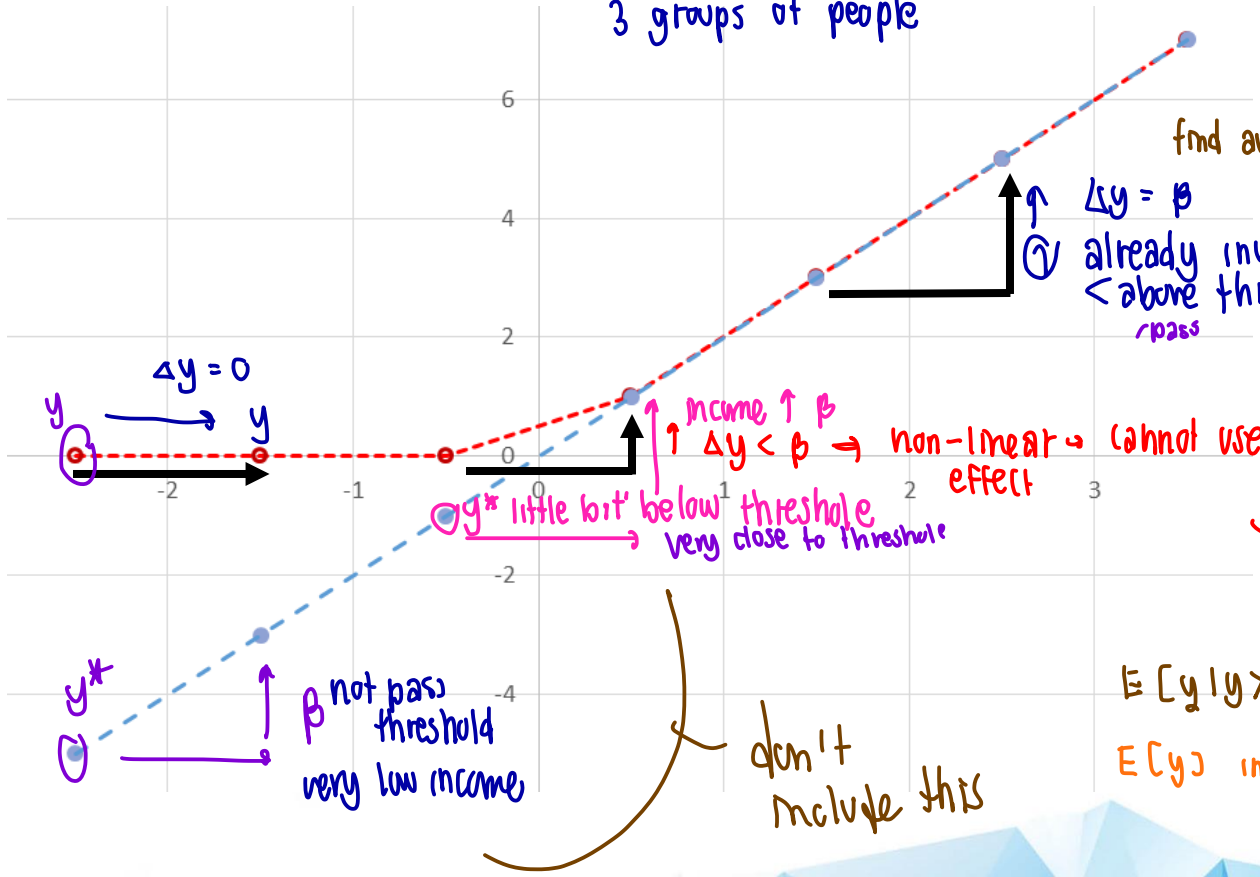
Interpretation of β in Tobit Model

willingness to invest

actual level investment in stock mkt

- Recall that in this model $X\beta + U = Y^*$ but not Y. Thus, **β is not simply the effect of X on Y but on Y^*.**

3 groups of people



⇓

use Tobit model

E[y|y>0]

E[y] includes all groups

- In general, we care more about the effect on Y rather than on the latent variable.
- For simplicity, suppose that $c = 0$. Recall, $E[Y|X] = E\big[E[Y|X, Y^*]\big]$
- So, we can write

$$E[Y|X] = \boldsymbol{E[Y|X, Y^* > c]} \cdot P\{Y^* > c\} + \underbrace{E[Y|X, Y^* \leq c]}_{Y=c=0} \cdot P\{Y^* \leq c\}$$

$$= \boldsymbol{E[X\beta + U|X, (X\beta + U > c)]} \cdot \Phi\left(\frac{X\beta - c}{\sigma}\right) + c \cdot \Phi\left(\frac{c - X\beta}{\sigma}\right)$$

$$= \left\{\boldsymbol{X\beta + \sigma E\left[\frac{U}{\sigma}\middle|X, \left(\frac{U}{\sigma} > -\frac{X\beta}{\sigma}\right)\right]}\right\} \cdot \Phi\left(\frac{X\beta}{\sigma}\right)$$

$$= \left\{\boldsymbol{X\beta + \sigma \frac{\phi\left(-\frac{X\beta}{\sigma}\right)}{1 - \Phi\left(-\frac{X\beta}{\sigma}\right)}}\right\} \cdot \Phi\left(\frac{X\beta}{\sigma}\right) = \left\{\boldsymbol{X\beta + \sigma \frac{\phi\left(\frac{X\beta}{\sigma}\right)}{\Phi\left(\frac{X\beta}{\sigma}\right)}}\right\} \cdot \Phi\left(\frac{X\beta}{\sigma}\right)$$

- In sum, we have

$$E[Y|X, Y > 0] = X\beta + \sigma \frac{\phi\left(\frac{X\beta}{\sigma}\right)}{\Phi\left(\frac{X\beta}{\sigma}\right)}$$

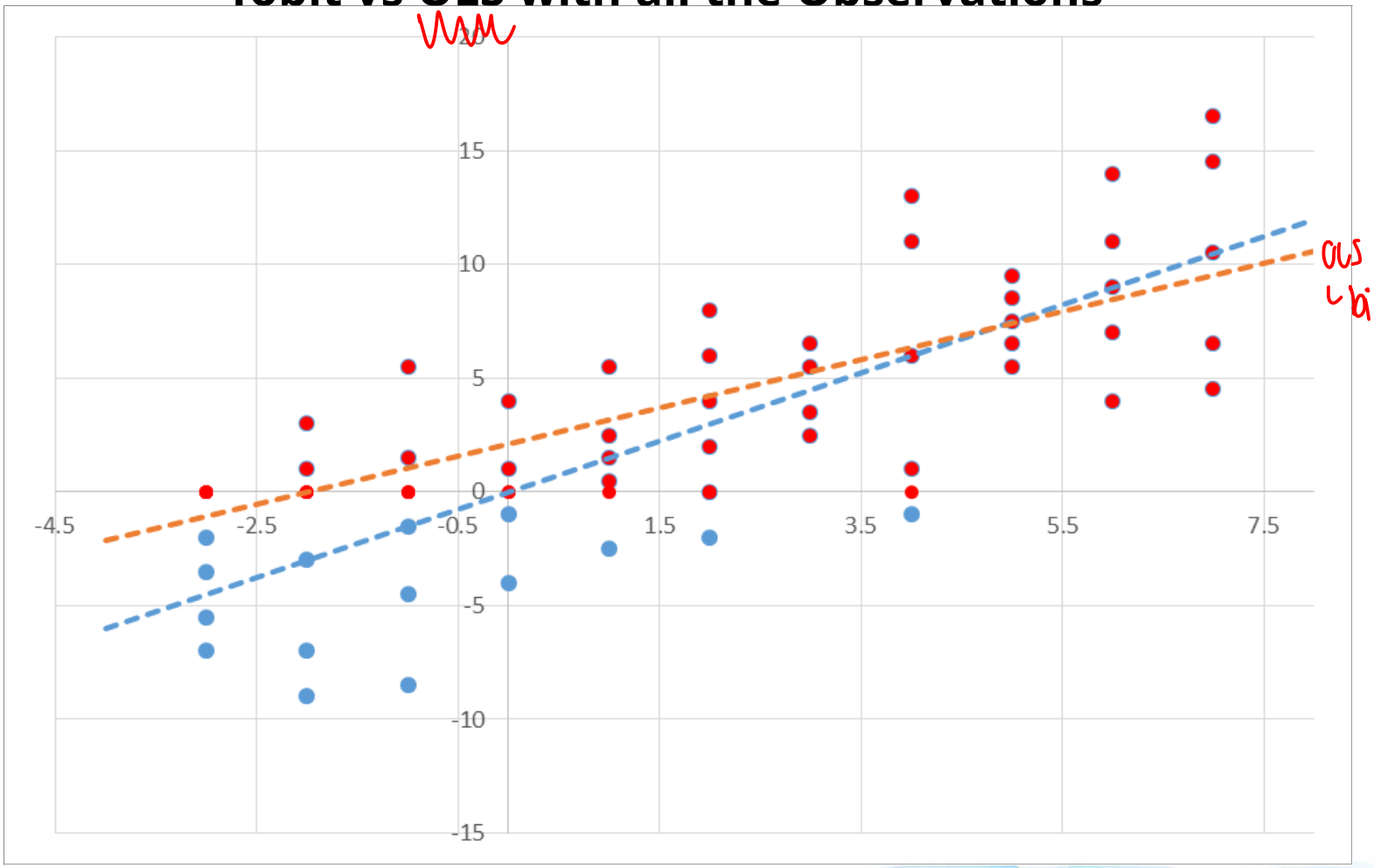
$$E[Y|X] = \left\{ X\beta + \sigma \frac{\phi\left(\frac{X\beta}{\sigma}\right)}{\Phi\left(\frac{X\beta}{\sigma}\right)} \right\} \cdot \Phi\left(\frac{X\beta}{\sigma}\right)$$

not linear

diff with respect to X

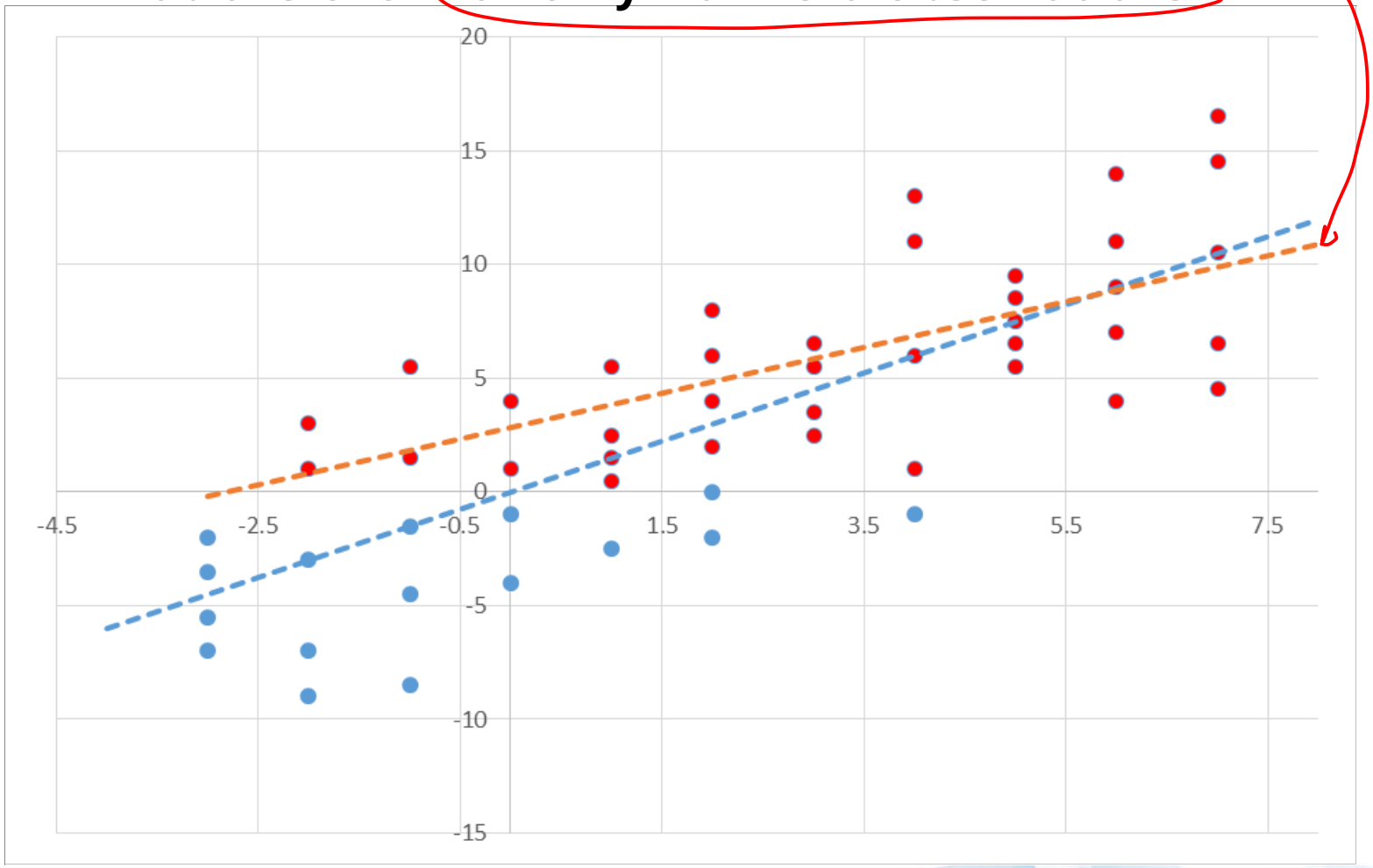
- Therefore, if we apply OLS estimator with only the samples with $Y > 0$, we will get a biased and inconsistent estimate for β, because we omit the variable $\sigma \frac{\phi\left(\frac{X\beta}{\sigma}\right)}{\Phi\left(\frac{X\beta}{\sigma}\right)}$ which is correlated with the regressor X.
- Similarly, if we apply OLS estimator with all the observed samples of Y, we will get a biased and inconsistent estimate as well, since the effect is not in a linear form.

Tobit vs OLS with all the Observations



Tobit vs OLS with Only Non-zero Observations

biased



Estimating Tobit Model

estimator

specify threshold value

```
. tobit y wage, ll(0)
```

y x

ul(□)
upper level (censor from above)

lower level of zero
(censor from below)

```
Tobit regression                                  Number of obs     =        500
                                                  LR chi2(1)        =     139.72
                                                  Prob > chi2       =     0.0000

Log likelihood = -2435.7398                       Pseudo R2         =     0.0279
```

goodness of fit

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
wage	1170.073	101.3535	11.54	0.000	970.9405	1369.205
_cons	-3084.496	300.2466	-10.27	0.000	-3674.399	-2494.592
/sigma	1080.463	49.87637			982.4698	1178.457

to do inference

```
          227  left-censored observations at y <= 0
          273     uncensored observations
            0 right-censored observations
```

E.g. of upper level

cannot specifically see this

tobit

Portfolio size = $\beta_0 + \beta_1$ Income

censored at some threshold

- below at zero
- above

e.g. very rich investors don't want to disclose their port size

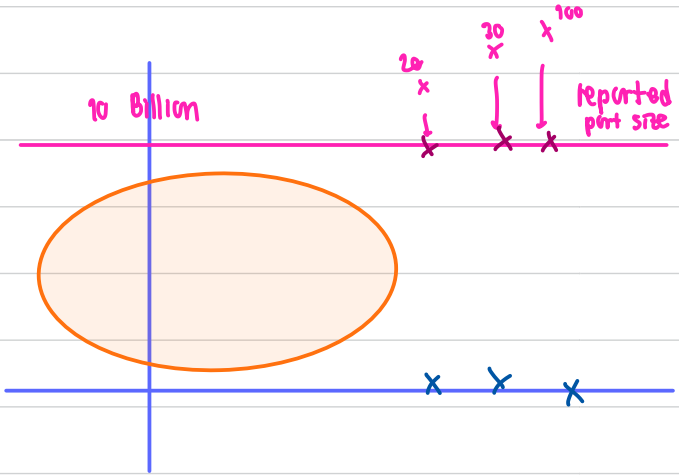
↳ researcher makes choice

① 0

② more than 10 Billion

③ less than n ——— n

censor from above



reported as port size

reported as 10 Billion bcs we didn't ask exactly

↓ assume they have 10 bn consor

Estimating Marginal Effects based on Tobit Model

expected value

e(a,b) → effect on expected value of y → E[y | y ∈ (a,b)]
for y btw a & b

```
. mfx, predict(e(0,.))
Marginal effects after tobit
      y  = E(y|y>0) (predict, e(0,.))
         = 910.27961
```

→ E[y | y > 0]

ที่แสดง y ที่ change for whole sample

for whose x = 2.74554
if increase wage 1 unit → port investment ↑ 456.3374

whose effect

```
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|   [    95% C.I.    ]      X
---------+--------------------------------------------------------------------
    wage |  456.3374      38.109   11.97   0.000    381.646   531.029   2.74554
------------------------------------------------------------------------------
```

$ per person

tell profile to calculate marginal effect

probability [y > 0]

chance of people invest in stock that improve

```
. mfx, predict(p(0,.))
Marginal effects after tobit
      y  = Pr(y>0) (predict, p(0,.))
         = .54714642
```

+∞

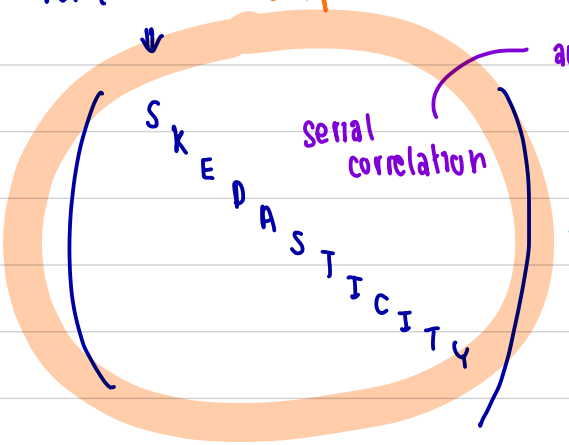
→ ที่ change

```
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|   [    95% C.I.    ]      X
---------+--------------------------------------------------------------------
    wage |  .4290086      .03767   11.39   0.000    .355185   .502832   2.74554
------------------------------------------------------------------------------
```

chance that will invest increase by 42.9%

Review <10 Nov 2023>

$Var(u|x) \Rightarrow se(\hat{\beta})$



autocorrelation

use FGLS กับ pattern ที่เขียน reg, robust ไม่ได้
if we know functional form of it

same $\hat{\beta}$ diff $se(\hat{\beta})$

Homoskedasticity w/o serial correlation : reg y x

Hetero " ——— " w/o " ——— " : reg y x , robust
white robust

↑ volatility : variance/sd of the error term

↓ conditional heteroskedasticity

ARCH, GARCH → variance of error terms
not about serial correlation

heteroskedasticity

help capture asymmetric effect of shock on volatility
& negative shock is more reacted than positive shock

GJR T E

GARCH

to let model capture very long lasting shock → adding σ^2_{t-1}

diff model diff pattern

don't have to restrict coefficients

MLE diff $\hat{\beta}$ from OLS

Volatility :

① Mean equation

② Volatility model

What is your y & X ?

$y = \beta X + u$

mut

return

meaning of error term or shock ?

e.g. Idiosyncratic shock

If not control for mut → u will be total risk

model

Select # of lag → use AIC, BIC

GARCH (1,1) to estimate volatility in return

Homo / Hetero w serial correlation || for time series
Newey y x

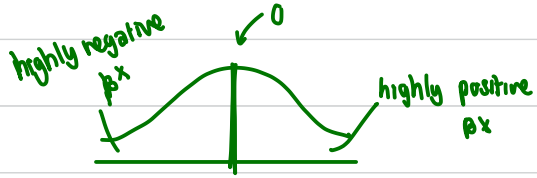
Serial correlation? but have time series data
↳ we can test serial correlation by use ~~Durbin-Watson~~ too old & limitation
Breusch-Godfrey test

Limited Dependent Variable Models:

I) Binary Response model : want to study probability

$y \begin{cases} 1 \\ 0 \end{cases}$

regress y on another thing



, linear prob model

① LPM : prob that $y=1$ is βx → β
② Probit : " ——— " is $\Phi(\beta x)$
③ Logit : " ——— " is $\Lambda(\beta x) = \frac{e^{\beta x}}{1+e^{\beta x}}$

plug in βx to get prob

estimated effect is diff for diff people

$pdf(\beta x) \cdot \beta$

effect is more pronounced for average individual

When x increase, what is effect on y?

II) Censored
⇓
not use OLS (not consistence)
but use Tobit, effect of x on y is not just β ← very complicated