Elements of Econometrics. Lecture 15. Maximum Likelihood Estimation. Limited Dependent Variable Models.

FCS, 2022-2023

Maximum likelihood estimation

MLE is widely used in estimating various models, and for some of them it is the principal estimation method.

It provides the estimates of parameters θ which maximise joint probability (or probability density) of the sample available:

$f(y_1, y_2, \ldots, y_n; \theta) \rightarrow max$, or

$f(y_1;\theta) f(y_2;\theta) \ldots f(y_n;\theta) \rightarrow max$ for the case of independent y_i.

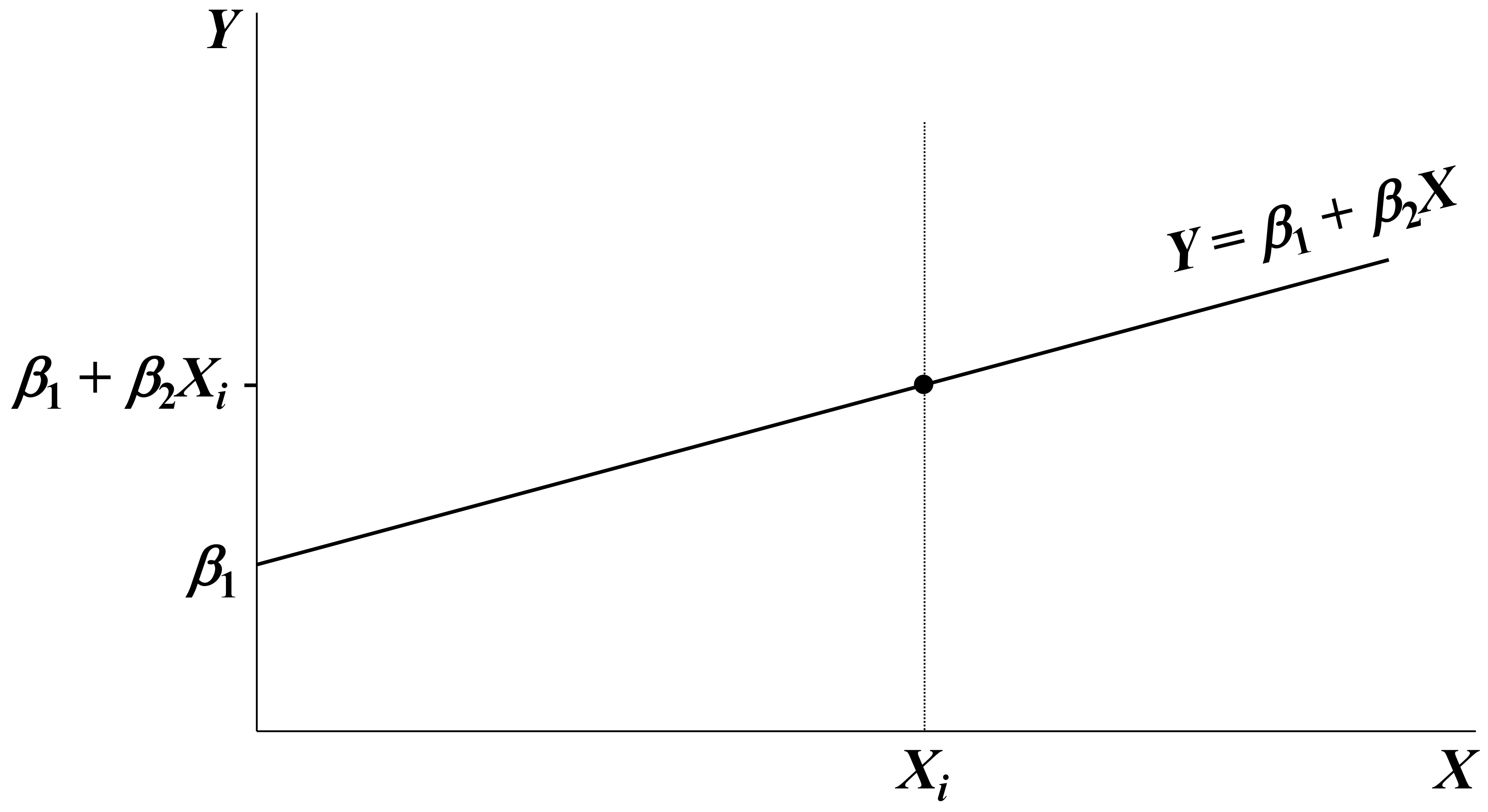
MLE is **usually consistent** and often unbiased.

Will be used for Binary Choice and Limited Dependent Variable Models.

MLE is generally the most asymptotically efficient estimator when the population model $f(y;\theta)$ is correctly specified.

MLE is sometimes the **minimum variance unbiased estimator.**

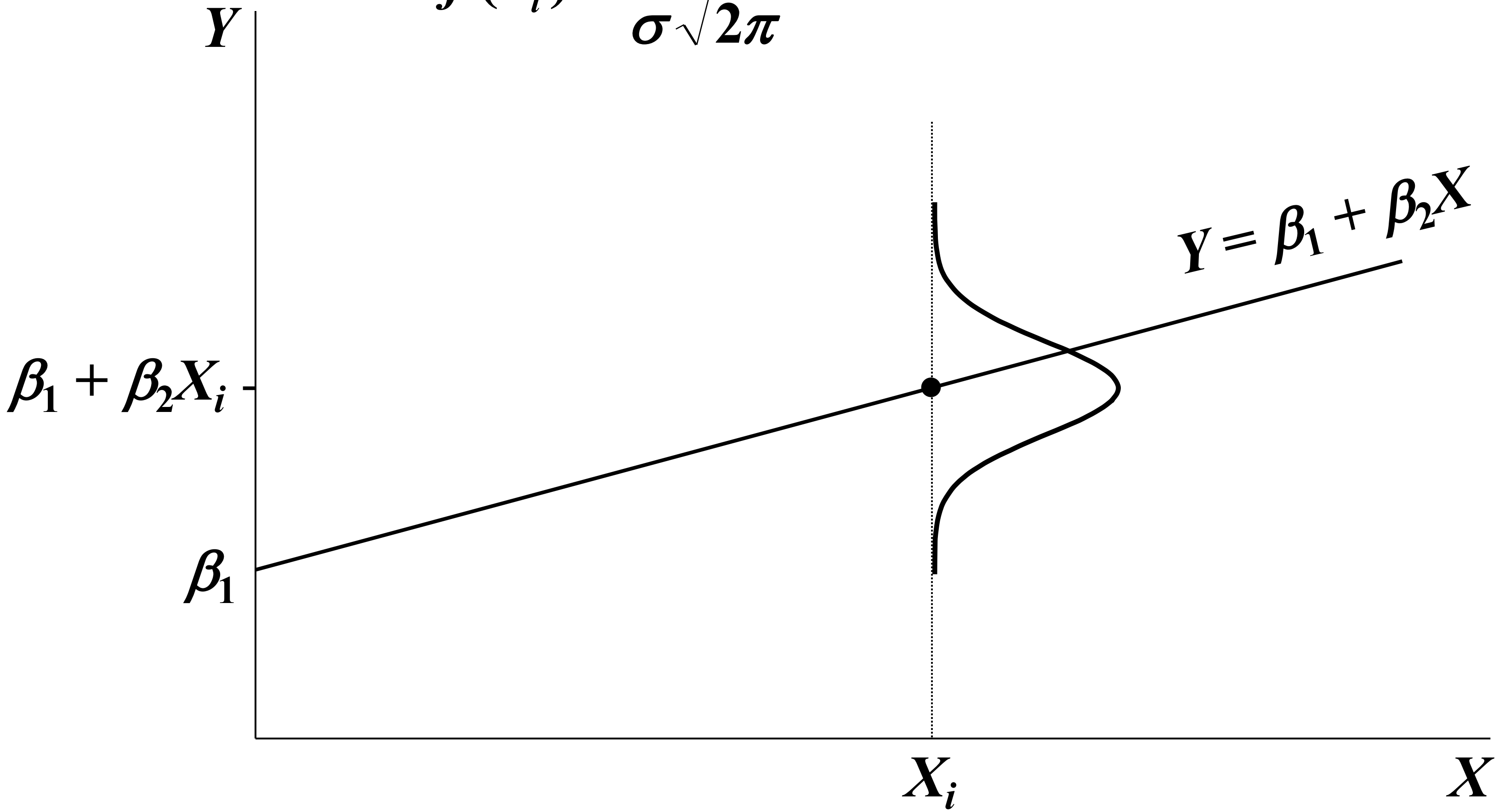
MAXIMUM LIKELIHOOD ESTIMATION OF REGRESSION COEFFICIENTS



We will now apply the maximum likelihood principle to regression analysis, using the simple linear model $Y = \beta_1 + \beta_2 X + u$.

MAXIMUM LIKELIHOOD ESTIMATION OF REGRESSION COEFFICIENTS

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_i - \beta_1 - \beta_2 X_i}{\sigma}\right)^2}$$



We will assume that the disturbance term in the model has a normal distribution.

The mean value of the distribution of Y_i is $\beta_1 + \beta_2 X_i$. Its standard deviation is σ, the standard deviation of the disturbance term.

The density function for the distribution of Y_i is as shown.

MAXIMUM LIKELIHOOD ESTIMATION OF REGRESSION COEFFICIENTS

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_i - \beta_1 - \beta_2 X_i}{\sigma}\right)^2}$$

$$f(Y_1) \times \ldots \times f(Y_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1 - \beta_1 - \beta_2 X_1}{\sigma}\right)^2} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n - \beta_1 - \beta_2 X_n}{\sigma}\right)^2}$$

$$L(\beta_1, \beta_2, \sigma | Y_1, \ldots, Y_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1 - \beta_1 - \beta_2 X_1}{\sigma}\right)^2} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n - \beta_1 - \beta_2 X_n}{\sigma}\right)^2}$$

$$\log L = \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1 - \beta_1 - \beta_2 X_1}{\sigma}\right)^2} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n - \beta_1 - \beta_2 X_n}{\sigma}\right)^2}\right)$$

We will choose β_1, β_2, and σ to maximize the log likelihood, given the data on Y and X.

MAXIMUM LIKELIHOOD ESTIMATION OF REGRESSION COEFFICIENTS

$$\log L = \log\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y_1-\beta_1-\beta_2X_1}{\sigma}\right)^2}\times\ldots\times\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y_n-\beta_1-\beta_2X_n}{\sigma}\right)^2}\right)$$

$$= \log\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y_1-\beta_1-\beta_2X_1}{\sigma}\right)^2}\right)+\ldots+\log\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y_n-\beta_1-\beta_2X_n}{\sigma}\right)^2}\right)$$

$$= n\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{1}{2}\left(\frac{Y_1-\beta_1-\beta_2X_1}{\sigma}\right)^2-\ldots-\frac{1}{2}\left(\frac{Y_n-\beta_1-\beta_2X_n}{\sigma}\right)^2$$

$$= n\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{\sigma^{-2}}{2}Z$$

$$\text{where } Z = \left[(Y_1-\beta_1-\beta_2X_1)^2+\ldots+(Y_n-\beta_1-\beta_2X_n)^2\right]$$

To maximize the log-likelihood, we need to minimize *Z*. Choosing estimators of β_1 and β_2 to minimize *Z* is identical to the OLS procedure. Hence the ML estimators of β_1 and β_2 are identical to the least squares estimators.

MAXIMUM LIKELIHOOD ESTIMATION OF REGRESSION COEFFICIENTS

$$\log L = n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{\sigma^{-2}}{2} Z =$$

$$= n \log\left(\frac{1}{\sigma}\right) + n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{\sigma^{-2}}{2} Z =$$

$$= -n \log \sigma + n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{\sigma^{-2}}{2} Z$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \sigma^{-3} Z = \sigma^{-3}\left(Z - n\sigma^2\right)$$

$$\widehat{\sigma}^2 = \frac{Z}{n} = \frac{\sum e_i^2}{n}$$

The ML estimator of the variance σ^2 is the sum of the squares of the residuals divided by *n*. It is biased for finite samples (to obtain an unbiased estimator, we should divide by *n–k*, where *k* is the number of parameters). Bias disappears as the sample size becomes large.

Maximum Likelihood Estimation of the Logit Model

$$L = \prod_i \Pr(Y = Y_i | X_i, \beta) = \prod_{i:Y_i=1} F(\beta_1 + \beta_2 \cdot X_i) \cdot \prod_{i:Y_i=0} (1 - F(\beta_1 + \beta_2 \cdot X_i))$$

$$\to \max_\beta \; l(\beta) = \log L = \sum_i (\log p(Y = Y_i | X_i, \beta)) =$$

$$= \sum_{i:Y_i=1} \log F(\beta_1 + \beta_2 \cdot X_i) + \sum_{i:Y_i=0} \log(1 - F(\beta_1 + \beta_2 \cdot X_i)) =$$

$$= \sum_i Y_i (\log F(\beta_1 + \beta_2 \cdot X_i)) + \sum_i (1 - Y_i)(\log(1 - F(\beta_1 + \beta_2 \cdot X_i))) \to \max_\beta$$

For logit model $\quad F(\beta_1 + \beta_2 \cdot X_i) = \dfrac{1}{1 + e^{-(\beta_1 + \beta_2 \cdot X_i)}}$

EViews: Quick – Estimate Equation – Equation Specification (type) –

Method: Binary - Logit

LOGIT MODEL: EXAMPLE.

ICEF STUDENTS UoL First Class Honours Degrees, pre-covid year, Depending on their Econometrics Performance

Dependent Variable: FIRST
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 01/04/19 Time: 11:21
Sample (adjusted): 3 241
Included observations: 239 after adjustments
Convergence achieved after 8 iterations
Coefficient covariance computed using observed Hessian

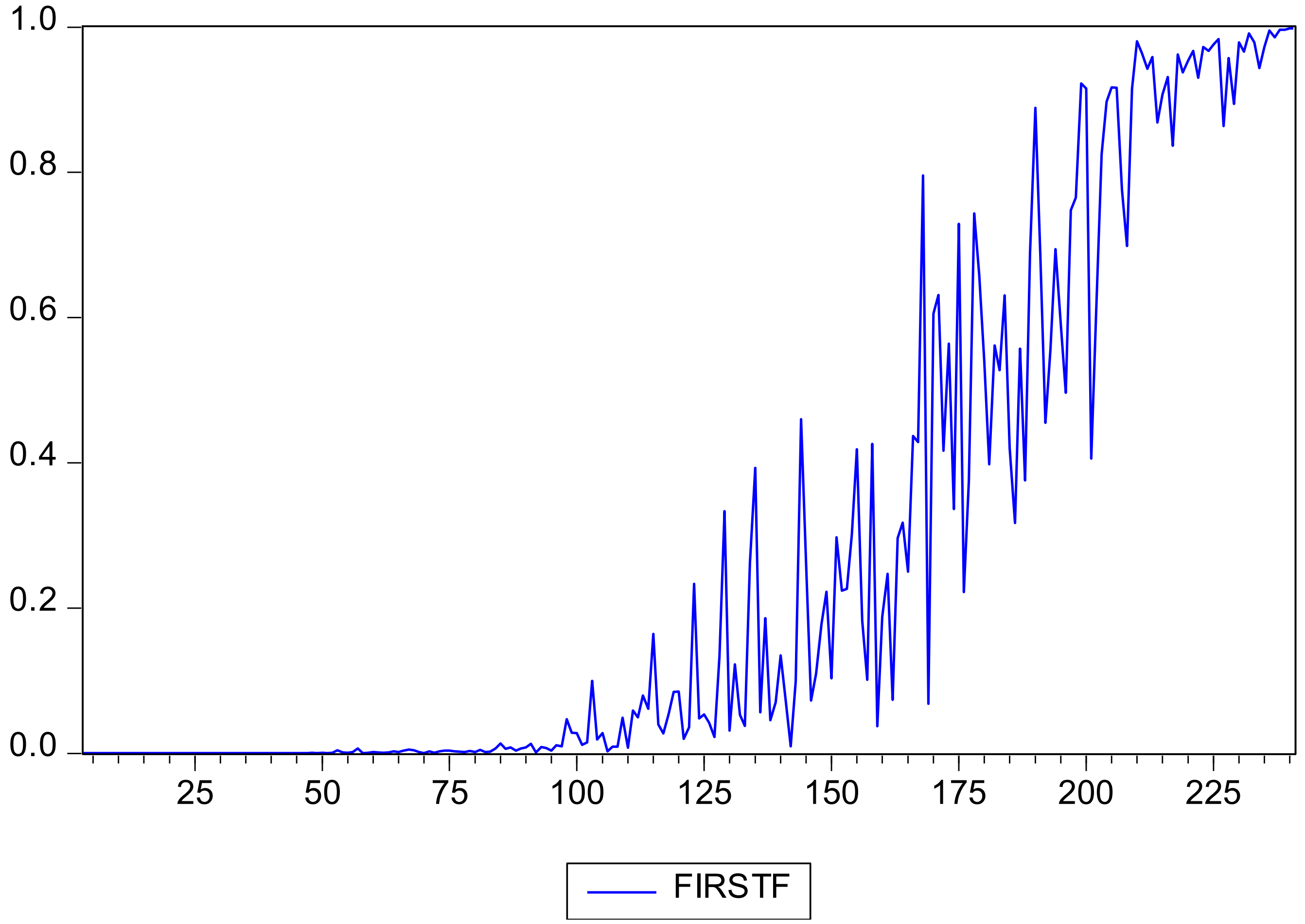
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-15.28950	2.357285	-6.486062	0.0000
SEM1	0.078481	0.024139	3.251175	0.0011
UOL	0.183450	0.034530	5.312829	0.0000

McFadden R-squared	0.620387	Mean dependent var	0.276151
S.D. dependent var	0.448031	S.E. of regression	0.265737
Akaike info criterion	0.472503	Sum squared resid	16.66544
Schwarz criterion	0.516140	Log likelihood	-53.46405
Hannan-Quinn criter.	0.490087	Deviance	106.9281
Restr. deviance	281.6763	Restr. log likelihood	-140.8381
LR statistic	174.7482	Avg. log likelihood	-0.223699
Prob(LR statistic)	0.000000		

LOGIT MODEL: EXAMPLE.

ICEF STUDENTS UoL First Class Honours Degrees, pre-covid year,

Depending on their Econometrics Performance: Logit Forecast



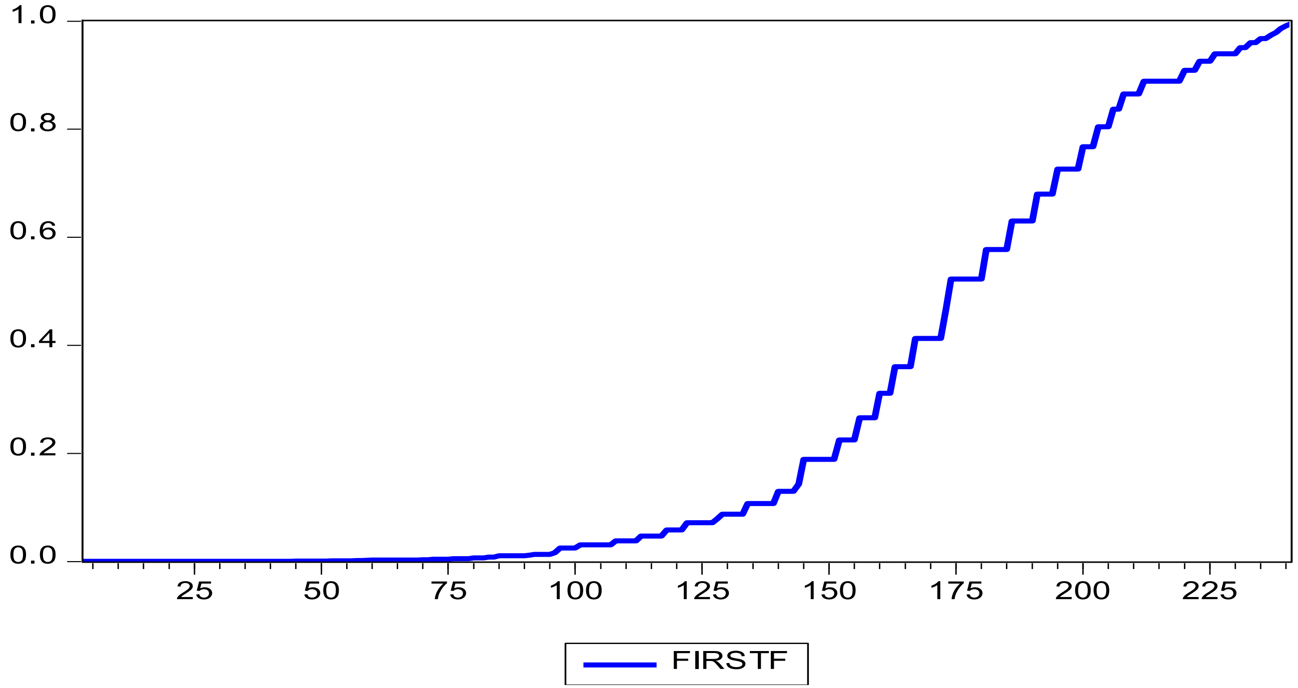
LOGIT MODEL: EXAMPLE.

ICEF STUDENTS UoL First Class Honours Degrees, pre-covid year:

139 obs, Dependent Variable – FIRST, SEM1 proxy omitted

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-14.72458	2.103972	-6.998467	0.0000
UOL	0.221113	0.031821	6.948581	0.0000

McFadden R-squared	0.576171	Mean dependent var	0.276151



Maximum Likelihood Estimation of the Probit Model

$$L = \prod_i Pr(Y = Y_i | X_i, \beta) = \prod_{i:Y_i=1} F(\beta_1 + \beta_2 \cdot X_i) \cdot \prod_{i:Y_i=0} (1 - F(\beta_1 + \beta_2 \cdot X_i)) \to \max_\beta$$

$$l(\beta) = \log L = \sum_i (\log p(Y = Y_i | X_i, \beta)) =$$
$$= \sum_{i:Y_i=1} \log F(\beta_1 + \beta_2 \cdot X_i) + \sum_{i:Y_i=0} \log(1 - F(\beta_1 + \beta_2 \cdot X_i)) =$$
$$= \sum_i Y_i (\log F(\beta_1 + \beta_2 \cdot X_i)) + \sum_i (1 - Y_i)(\log(1 - F(\beta_1 + \beta_2 \cdot X_i))) \to \max_\beta$$

For probit model $F(\beta_1 + \beta_2 \cdot X_i)$ –

cumulative function of standardized normal distribution

The goodness of fit in maximum likelihood estimation

1\. *"Pseudo-R^2"* (or *McFadden* R^2) = $1-\dfrac{\log L}{\log L_0}$

where $\log L_0$ is the natural logarithm of the value the likelihood function would take with only the intercept in the regression.

$\log L < 0$, since $0 < L < 1$.

The values of Pseudo-R^2 range from 0 to 1; the closer this coefficient is to 1, the better the fit.

The goodness of fit in maximum likelihood estimation

***2. The likelihood ratio*:**

$$LR = 2\log\left(\frac{L}{L_0}\right) = 2(\log L - \log L_0)$$

The likelihood ratio is used to test the following hypothesis:
H_0: the coefficients of all explanatory variables are equal to zero
H_1: the coefficient of at least one explanatory variable is not equal to zero.
Under the null hypothesis, the statistic LR has a χ^2-distribution with (k-1) degrees of freedom, where k is the number of parameters estimated, and, accordingly, (k-1) is the number of explanatory variables.

3. The significance of individual coefficients is tested via z-statistics, whose distribution approaches the standard normal in large samples.

The Likelihood Ratio Test for Variables Exclusion Restrictions

Maximum likelihood estimation (MLE), provides with a log-likelihood, L

As in F test, we estimate the restricted and unrestricted models, then form $LR = 2(L_{ur} - L_r) \sim \chi^2_q$ where q is the number of restrictions (excluded explanatory variables).

Goodness-of-fit measures for Logit and Probit models

– Percent correctly predicted (*G(z)=F(z)*)

$$\tilde{y}_i = \begin{cases} 1 & \text{if } G(\mathbf{x}_i\hat{\beta}) \geq .5 \\ 0 & \text{otherwise} \end{cases}$$

Individual i's outcome is predicted as one if the probability for this event is larger than .5, then percentage of correctly predicted y = 1 and y = 0 is counted

- Pseudo R-squared

$$\tilde{R}^2 = 1 - \log L_{ur} / \log L_0$$

Compare maximized log-likelihood of the model with that of a model that only contains a constant (and no explanatory variables)

- Correlation based measures

$$Corr(y_i, \tilde{y}_i), \;\; Corr(y_i, G(\mathbf{x}_i\hat{\beta}))$$

Look at correlation (or squared correlation) between predictions or predicted prob. and true values

Probit or Logit?

- Both the probit and logit are nonlinear and require maximum likelihood estimation
- The functions F, pseudo-R^2, LR statistics and p-values are usually close to each other
- No real reason to prefer one over the other
- Traditionally logit was wider used because the logistic function leads to a more easily computed model
- Now, probit is easy to compute with standard packages, and is used widely

TOBIT ANALYSIS: CENSORED SAMPLE

Let $Y^* = \beta_1 + \beta_2 X + u$
Y is observed, and

$$Y = Y^* \quad \text{if} \quad Y^* \geq 0$$
$$Y = 0 \quad \text{if} \quad Y^* < 0$$

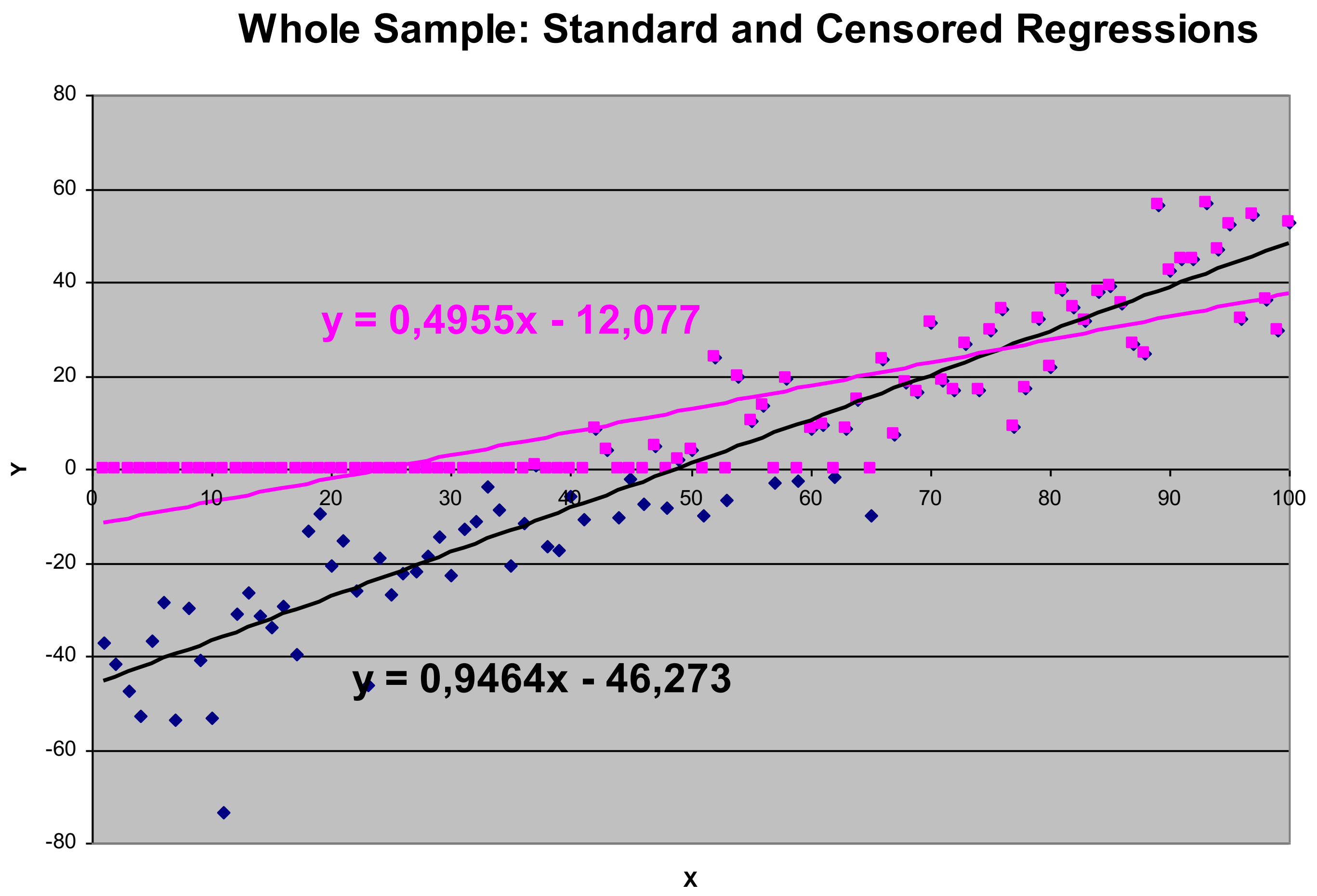
Monte Carlo experiment: X=1;2;…;100; Y*= -50+X+10*nrnd
Regression of *Y* on *X*:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-11.36440	1.852136	-6.135836	0.0000
X	0.493684	0.031841	15.50454	0.0000

R^2=0.71.

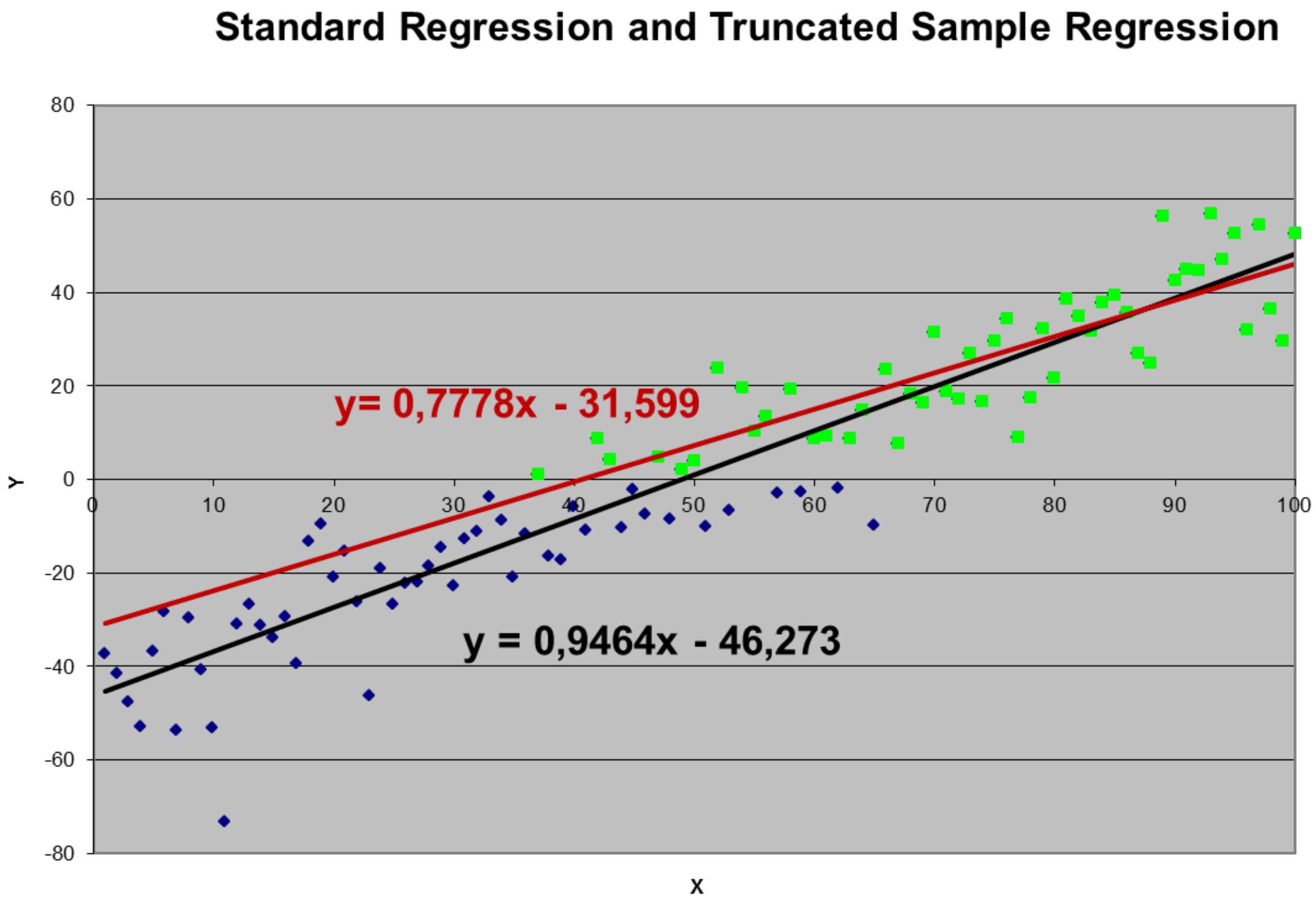
If *Y* is regressed on *X* instead of *Y, then *X* and the disturbance term are related (the smaller is *X*, the greater is the disturbance term), and the OLS estimators are biased.**

TOBIT ANALYSIS: CENSORED REGRESSION EXAMPLE



X=1;2;…;100; Y*=-50+X+10*nrnd

TOBIT ANALYSIS: TRUNCATED SAMPLE EXAMPLE



X=1;2;…;100;   Y*=-50+X+10*nrnd.

Truncated Sample Regression: $\beta_1 + \beta_2 X + u \geq 0$ **→ only observations with** $u \geq -\beta_1 - \beta_2 X$ **are kept in the sample, hence in the remaining sample** X **and** u **related, and the OLS estimators are biased.**

Maximum Likelihood Estimation of the Tobit Model

$$l(\beta) = \log L = \sum_{i:Y_L<Y_i<Y_U} (\log(f((Y_i - \beta_1 - \beta_2 \cdot X_i)/\sigma)) +$$

$$+ \sum_{i:Y_i=Y_L} (\log(F((Y_L - \beta_1 - \beta_2 \cdot X_i)/\sigma)) + \sum_{i:Y_i=Y_U} \log(1 - F((Y_U - \beta_1 - \beta_2 \cdot X_i)/\sigma)) \to \max$$

Here f is probability density function for standardised normal distribution; F – distribution function.

EViews: Quick – Estimate Equation – Equation Specification (type) –

Method: Censored (tobit). Enter censoring points of dependent variable.

Tobit Estimation: Example

Dependent Variable: Y1 Method: ML - Censored Normal (TOBIT)
Included observations: 100 Left censoring (value) at zero
Convergence achieved after 6 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C	-47.75152	5.000606	-9.549148	0.0000
X	0.962746	0.068717	14.01023	0.0000

R-squared 0.856426

Tobit provides consistent estimates; homoscedasticity and normal distribution of the disturbance term are needed for this.