Page 23

Tuesday, March 01, 2016 8:49 AM

Lecture 14

5.6 The problem of prediction

Based on our sample data, we have the following sample regression:

$$\boxed{\hat{Y}_i = \underset{\hat{\beta}_1}{98.5238} + \underset{\hat{\beta}_2}{0.5929}X_i}$$

UNBIASED ESTIMATOR OF $E(Y|X_i)$

We can use the above regression to "Predict" or "Forecast" the future consumption expenditure Y corresponding to some given level of income X

There are two kinds of predictions which are:

[1] **Mean prediction** We will predict the conditional mean value of Y corresponding to a chosen X (i.e X_0)

[2] **Individual Prediction** We will predict an individual Y value corresponding to (i.e X_0)

Mean Prediction

EX: SUPPOSE WE WANT TO ESTIMATE $\hat{Y}_0$ WHEN $X_0 = 100$. IN OTHER WORDS, $E[Y|X_0=100] = ?$

$\hat{Y}_0 = 98.5238 + 0.5929(100) = 157.8138$

We know that

$$\boxed{\hat{Y}_0 \sim N(E(\hat{Y}_0), \text{var}(\hat{Y}_0))}$$

→ $\hat{Y}_0$ IS NORMALLY DISTRIBUTED WITH MEAN $E(\hat{Y}_0)$ AND VARIANCE $VAR(\hat{Y}_0)$.

where

$$E(\hat{Y}_0) = \beta_1 + \beta_2 X_0$$

and

$$\text{var}(\hat{Y}_0) = \sigma^2\left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]$$

AVAILABLE IN APPENDIX 5A.4 (PAGE 146)

GIVEN $X = X_0$, THE TRUE MEAN PREDICTION $E(Y_0|X_0)$ IS GIVEN BY

$E(Y_0|X_0) = \beta_1 + \beta_2 X_0$ — ①

WHICH CAN BE ESTIMATED FROM

$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$ — ②

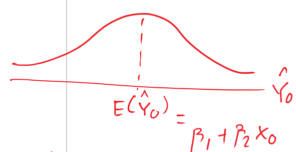
Q: WHY $\hat{Y}_0$ CAN BE USED TO ESTIMATE $E(Y_0|X_0)$?

A: $E(\hat{Y}_0) = E(\hat{\beta}_1) + E(\hat{\beta}_2)X_0$

$E(\hat{Y}_0) = \beta_1 + \beta_2 X_0$

SO $E(\hat{Y}_0) = E(Y_0|X_0) = \beta_1 + \beta_2 X_0$

WE PROVE THAT $\hat{Y}_0$ IS UNBIASED ESTIMATOR OF $E(Y_0|X_0)$!!! #



SAMPLE 1 → $\hat{Y}_0$
SAMPLE 2 → $\hat{Y}_0$
SAMPLE 3 → $\hat{Y}_0$

BY MAKING USE OF $var(a+b) = var(a) + var(b) + 2\cdot cov(a,b)$

& $var(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum x_i^2}$ & $var(\hat{\beta}_1) = \dfrac{\sum X_i^2}{n\sum x_i^2}\cdot\sigma^2$ & $cov(\hat{\beta}_1, \hat{\beta}_2) = -\bar{X}\left(\dfrac{\sigma^2}{\sum x_i^2}\right)$

WE CAN GET $var(\hat{Y}_0) = \sigma^2\left[\dfrac{1}{n} + \dfrac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]$. #

CAN BE ESTIMATED BY 1ST UNBIASED ESTIMATOR $\hat{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n-2}$.

$= \frac{\sum \hat{u}_i^2}{n-2}$

If we replace the unknown σ^2 by the unbiased estimator $\hat{\sigma}^2$ we can get

$$\boxed{t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{se(\hat{Y}_0)}}$$

which is the t distribution with n-2 df.
Therefore, we can derive the confidence interval for the true $E(Y_0|X_0)$ as following:

$$Pr[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\frac{\alpha}{2}} se(\hat{Y}_0) \leq \beta_1 + \beta_2 X_0 \leq \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\frac{\alpha}{2}} se(\hat{Y}_0)] = 1 - \alpha$$

Example

FROM OUR SAMPLE

$$var(\hat{Y}_0) = \hat{\sigma}^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$$

$$= 800.4762 \left[\frac{1}{6} + \frac{(100 - 750)^2}{175,000} \right]$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = \frac{3201.90}{6-2} = 800.4762$$

$$= 2065.991$$

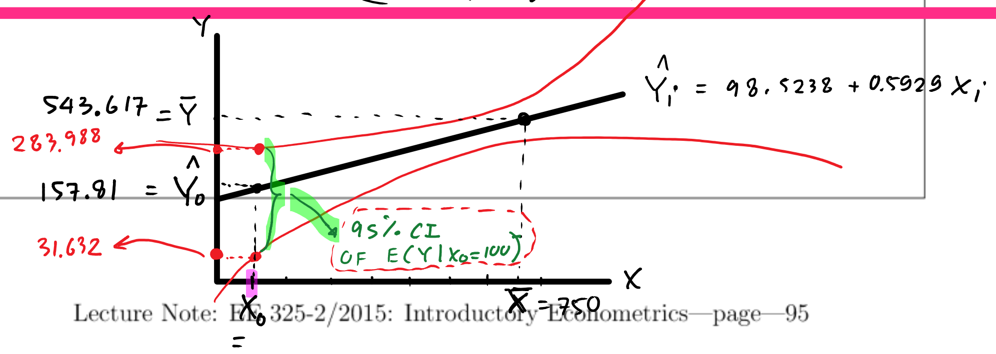
$$se(\hat{Y}_0) = \sqrt{2065.991} = 45.453$$

$$t_{\frac{0.05}{2}, 4} = 2.776$$

THEREFORE, GIVEN $X_0 = 100$, IN REPEATED SAMPLING, 95 OUT OF 100 CASES OF INTERVALS LIKE BELOW "WILL CONTAIN" OR "WILL INCLUDE" THE TRUE MEAN VALUE $\beta_1 + \beta_2 X_0$

$$\underbrace{\hat{\beta}_1 + \hat{\beta}_2 X_0}\ \underbrace{- t_{\frac{0.05}{2}, 4}}\ \underbrace{se(\hat{Y}_0)} \qquad \underbrace{+ t_{\frac{0.05}{2}, 4}}\ \underbrace{se(\hat{Y}_0)}$$

$$\boxed{\begin{array}{c} 157.81 - (2.776)(45.453) \leq E(Y|X_0 = 100) \leq 157.81 + (2.776)(45.453) \\ 31.632 \leq E(Y|X_0 = 100) \leq 283.988 \end{array}}$$



P-I-Y TRY THIS ⇒ FIND 95% CI FOR $E(Y|X_0 = 1200)$.

Individual Prediction

We can predicting an individual Y value, Y_0 corresponding to a given X value (X_0) :

THE BEST LINEAR UNBIASED ESTIMATOR OF Y_0 IS ALSO :

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$$

but the variance in this case is:

$$\mathrm{var}(Y_0 - \hat{Y}_0) = E[Y_0 - \hat{Y}_0]^2 = \sigma^2\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right]$$

We can show that Y_0 follows the normal distribution:

$$Y_0 \sim N(\hat{Y}_0, \mathrm{var}(Y_0 - \hat{Y}_0))$$

Therefore, we can construct the confidence interval for the Y_0 as well.

From our example:

$$t = \frac{Y_0 - \hat{Y}_0}{se(Y_0 - \hat{Y}_0)} \sim t$$

95% CI FOR Y_0 CORRESPONDING TO $X_0 = 100$ IS

$$\hat{Y}_0 - t_{\frac{\alpha}{2}} \cdot se(Y_0 - \hat{Y}_0) \leq Y_0 \mid X_0 = 100 \leq \hat{Y}_0 + t_{\frac{\alpha}{2}} \cdot se(Y_0 - \hat{Y}_0)$$

FIRST, WE COMPUTE $\mathrm{var}(Y_0 - \hat{Y}_0) = 800.4762\left[1 + \frac{1}{6} + \frac{(100 - 750)^2}{175{,}000}\right]$

$$= 2866.467$$

$$se(Y_0 - \hat{Y}_0) = \sqrt{2866.467} = 53.539.$$

THEREFORE, 95% CI FOR INDIVIDUAL Y_0 CORRESPONDING TO $X_0 = 100$ IS :

$$9.1840 \leq Y_0 \mid X_0 = 100 \leq 308.435$$

NOTICE THAT CI FOR INDIVIDUAL Y_0 IS WIDER THAN CI FOR CONDITIONAL MEAN OF Y_0.

CHAPTER 6: Extensions of The Two-Variable Linear Regression Model

6.1 Functional Form of regression Models

[1] The log-linear model

[2] Semilog models

[3] Reciprocal models

[4] The logarithmic reciprocal model

The Log-linear Model

CONSIDER $Y_i = \beta_1 X_i^{\beta_2} e^{u_i}$ WHERE $e = 2.718$

TAKING ln BOTH SIDES GIVES:

$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i$

$\boxed{\ln Y_i = \alpha_1 + \beta_2 \ln X_i + u_i}$ DOUBLE-LOG MODEL

$\ln Y_i = \alpha_1 + \beta_2 \ln X_i$

$\frac{d \ln Y_i}{dX} = \beta_2 \frac{d(\ln X_i)}{dX}$

$\frac{1}{Y}\frac{dY}{dX} = \beta_2 \frac{1}{X}\frac{\cancel{dX}}{\cancel{dX}}$

$\boxed{\beta_2 = \frac{X}{Y}\frac{dY}{dX} = \frac{\frac{dY}{Y} \times 100}{\frac{dX}{X} \times 100} = \frac{\% \Delta Y}{\% \Delta X} = \text{ELASTICITY OF } Y}$

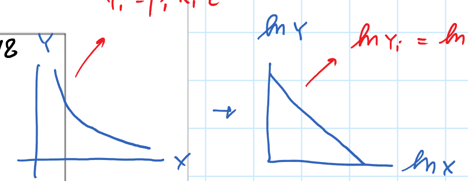
TELLS US ABOUT "PERCENTAGE CHANGE"

$\boxed{\text{SLOPE}: \frac{dY}{dX} = \beta_2 \cdot \frac{Y}{X}}$

TELLS US ABOUT "ABSOLUTE CHANGE"

$Y_i = \beta_1 X_i^{\beta_2} e^{u_i}$ $\qquad$ $\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i$

EX:



LINEAR TRANSFORMATION

The Semilog Models

LOG-LINEAR MODEL (OR LOG-LIN MODEL)

LINEAR-LOG MODEL (OR LIN-LOG MODEL)

LOG-LIN MODEL: $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$. [ECONOMETRIC MODEL]

$\ln Y_i = \beta_1 + \beta_2 X_i$ [MATHEMATICAL MODEL]

$$\frac{d \ln Y}{dX} = \beta_2 \cancel{\frac{dX}{dX}}$$

$$\frac{1}{Y}\frac{dY}{dX} = \beta_2$$

$$\beta_2 = \frac{\frac{dY}{Y}}{dX} = \frac{\frac{\Delta Y}{Y}}{\Delta X}$$

$\frac{\Delta Y}{Y}$ → RELATIVE CHANGE IN Y

ΔX → ABSOLUTE CHANGE IN X

CALLED "SEMI-ELASTICITY OF Y W.R.T. X"

SUPPOSE WE HAVE THIS QUESTION: IF X CHANGES BY 1 UNIT, Y WOULD CHANGE BY ? %.

FROM $\beta_2 = \frac{\frac{dY}{Y}}{dX}$

$$\frac{dY}{Y} = \beta_2 \cdot dX$$

$$100 \times \frac{dY}{Y} = 100 \cdot \beta_2 \cdot dX$$

BECOMES % CHANGE IN Y

PRACTICAL USE

X = YEARS IN SCHOOL

Y = WAGE OR SALARY

$\ln Y = \beta_1 + \beta_2 X_i + u_i$

EX: $\hat{\beta}_2 = 0.094$

$\% \Delta Y = 100 \times 0.094 \cdot 1$ ($= \Delta X$)

$= 9.4$

IF X INCREASES BY 1 UNIT, Y WOULD INCREASE BY 9.4%, ALL ELSE EQUAL.

LIN-LOG MODEL: $Y_i = \beta_1 + \beta_2 \ln X_i + u_i$.

$Y_i = \beta_1 + \beta_2 \ln X_i$

$$\frac{dY}{dX} = \beta_2 \frac{d(\ln X)}{dX}$$

$$= \beta_2 \frac{1}{X} \cdot \cancel{\frac{dX}{dX}}$$

$$\frac{dY}{dX} = \beta_2 \cdot \frac{1}{X}$$

dY → ABSOLUTE CHANGE IN Y

$$\beta_2 = \frac{dY}{\frac{dX}{X}} \quad \begin{array}{l} \rightarrow \text{ABSOLUTE CHANGE IN } Y \\ \rightarrow \text{RELATIVE CHANGE IN } X \end{array} \quad \left(\frac{\Delta X}{X} = \frac{x_2 - x_1}{x_1} \right)$$

HOW CAN WE UTILIZE THIS MODEL?

SUPPOSE Q: IF X CHANGES BY 1%, Y WOULD CHANGE BY ? UNIT(S)

$$\beta_2 = \frac{dY}{\frac{dX}{X}}$$

$$\Delta Y = \beta_2 \cdot \frac{\Delta X}{X}$$

$$100 \cdot \Delta Y = \beta_2 \cdot \underbrace{\frac{\Delta X}{X} \times 100}_{\text{BECOMES } \% \Delta X}$$

$$\frac{\Delta Y}{\frac{\Delta X}{X} \times 100} = \frac{\beta_2}{100} \quad \text{OR} \quad 0.01 \cdot \beta_2$$

PRATICAL USE: EX: $n = 52$ FAMILIES

UNIT OF CURRENCY: RUPEES

$$\text{FOODEXP}_i = \underbrace{-1283.912}_{\beta_1} + \underbrace{257.27}_{\beta_2} \ln(\text{TOTALEXP}_i).$$

IF TOTAL EXPENDITURE ↑ BY 1%, ON AVERAGE, FOOD EXPENDITURES WILL GO UP BY $0.01 \cdot (257.27) = 2.57$ RUPEES.

FROM $\Delta Y = \beta_2 \cdot \frac{\Delta X}{X}$, WE CAN FIND:

$$\frac{\Delta Y}{\Delta X} = \frac{\beta_2}{X}. \quad [\text{SLOPE}]$$

$$\underbrace{\frac{\% \Delta Y}{\% \Delta X} = \frac{\beta}{Y}}_{\text{D-I-Y: VERIFY THIS}}. \quad [\text{ELASTICITY}]$$

$\frac{X}{Y} \cdot \frac{dY}{dX}$

The Reciprocal Models

CONSIDER $Y_i = \beta_1 + \beta_2 \left[\frac{1}{X_i}\right] + u_i$

FROM $Y_i = \beta_1 + \beta_2 \cdot \left[\frac{1}{X_i}\right]$ → MATHEMATICAL MODEL

IF X_i INCREASES INDEFINITELY, THE TERM $\beta_2 \cdot \frac{1}{X_i}$ WILL APPROACH TO ZERO AND Y_i WILL CONVERGE TO β_1.

ASYMTOTE VALUE OR LIMIT VALUE

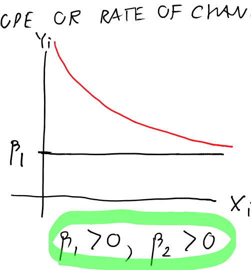
$$\frac{dY}{dX} = \beta_2 \frac{d[X_i^{-1}]}{dX}$$

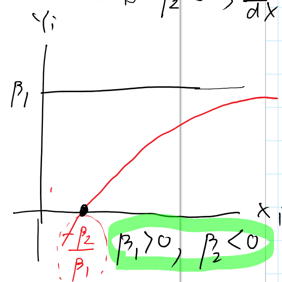
$$\frac{dY}{dX} = -\beta_2 X^{-2} = -\beta_2 \frac{1}{X^2}.$$

IF $\beta_2 > 0$, $\frac{dY}{dX}$ (-ve)

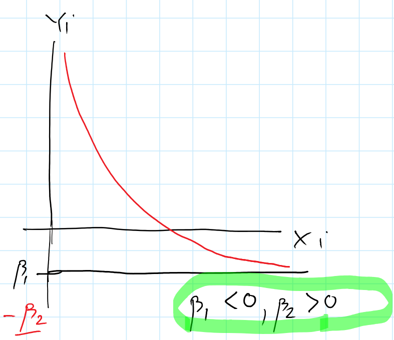
IF $\beta_2 < 0$, $\frac{dY}{dX}$ (+ve)

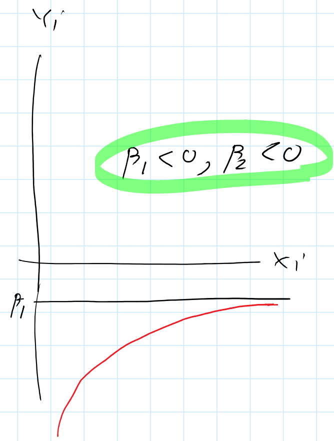
SLOPE OR RATE OF CHANGE





WHEN $Y = 0$, $X = -\frac{\beta_2}{\beta_1}$





APPLICATION

CHILD MORTALITY

$N = 60$ COUNTRIES

$CM = f(PGDP)$

$\beta_1 > 0$ $\beta_2 > 0$

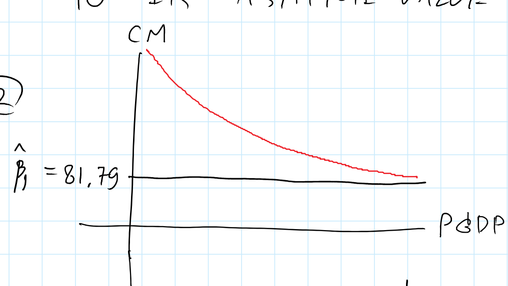
$$\widehat{CM} = 81.7943 + 27{,}237.17 \left[\frac{1}{PGDP}\right]$$

CM = # OF DEATHS OF CHILDREN UNDER 5 YEARS PER 1000 LIVE BIRTHS

PGDP = PER CAPITA GDP (CONSTANT PRICE)

FINDINGS

1. IF PGDP INCREASES INDEFINITELY, CM WILL APPROACH / CONVERGE TO ITS ASYMTOTE VALUE OF ABOUT 82 DEATHS / 1000 LIVE BIRTHS.

2. 

$\frac{\Delta CM}{\Delta PGDP} < 0$

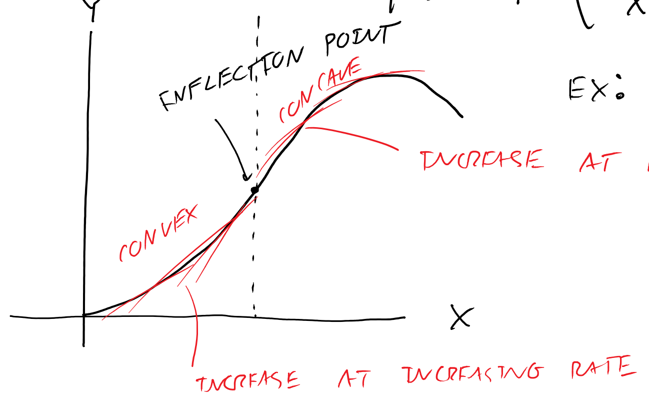
AS $\hat{\beta}_2$ IS POSITIVE, SO $\frac{\Delta CM}{\Delta PGDP}$ IS NEGATIVE.

$$\text{SLOPE} = \frac{dY}{dX} = -\beta_2 \left(\frac{1}{X^2}\right)$$

$$\text{ELASTICITY} = \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{-\beta_2}{X^2} \cdot \frac{X}{Y} = \frac{-\beta_2}{X \cdot Y}$$

The Logarithmic Reciprocal Models

CONSIDER $\ln Y_i = \beta_1 - \beta_2 \left(\frac{1}{X_i}\right) + u_i$



Ex: $Q = F(L, \bar{K})$ SHORT RUN PRODUCTION FUNCTION

$$\frac{d \ln Y}{dX} = -\beta_2 \left(-\frac{1}{X^2}\right) = \beta_2 \cdot \frac{1}{X^2}$$

$$\frac{1}{Y}\frac{dY}{dX} = \beta_2 \cdot \frac{1}{X^2}$$

SO $\boxed{\frac{dY}{dX} = \beta_2 \frac{Y}{X^2}}$

$\boxed{\text{ELASTICITY} = \beta_2 \cdot \frac{1}{X}}$ → VERIFY THIS!