Lecture Notes 1

1. Economic Forecasting

1. Forecasting – the process of estimating or predicting unknown situations

Example – usually economists predict future economic variables

Forecasting applies to a variety of data
(1) time – series data – predicting future data points in time
Examples – unemployment, interest rates, exchange rates etc.
(2) cross-sectional data – predicting within a class of a variable taken at one point in time
Examples – Usually associated with households, individuals, businesses, etc.
How many businesses bankrupted in 2009?
How many people are living in poverty in 2009?
(3) panel data – combines cross-sectional and time series data
Example – examining unemployment over time in several regions

Thus, forecasting is an enormous subject.
Time-series alone can use advanced math and is quite a complex topic
Combines statistics, calculus, and economics
Examples – ARIMA modeling and Vector Auto Regressions

2. Tools for forecasting

Forecasting extensively uses econometrics
Foundation of econometrics is least squares

Also called ordinary least squares, regression, or multiple regression

(1) Method

$$y_i = \beta_1 + \beta_2 x_i + u_i \qquad i = 1, 2, \cdots, n$$

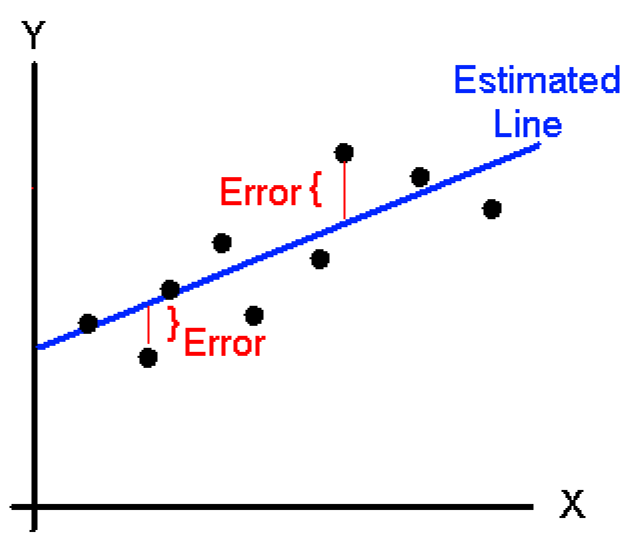
where y_i and x_i are paired observations or variables
β_1 and β_2 are unknown parameters
n is the total number of observations
u_i is the error term associated with observation i
Error is also called stochastic, white noise, or random
Has an expected value of zero

Time series data – the error is almost never random
Instead they call it an innovation
The error term contains information

When we estimate the parameters, they are denoted by hats

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i \qquad i = 1, 2, \cdots, n$$

The plot is below with the data points



We want to minimize the errors

Some errors are positive

Other errors are negative

We cannot add the errors because they may cancel

We square the errors terms to make them all positive

(2) Derivation

Starting with the equation

$$y_i = \beta_1 + \beta_2 x_i + u_i \quad i = 1, 2, \cdots, n$$

Solve for u_t, which yields

$$u_i = y_i - \beta_1 - \beta_2 x_i$$

Square the errors to make them positive

$$u_i^2 = (y_i - \beta_1 - \beta_2 x_i)^2$$

This is only for one data point. We want to minimize the total errors of all the data points.

Sum over all the data points

Define Sum of Squared Errors (SSE)

$$SSE = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2$$

We want to find the minimum, thus we take the first partial derivatives with respect to the betas

$$\frac{\partial SSE}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)^2$$

$$\frac{\partial SSE}{\partial \beta_1} = 2\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)\frac{\partial}{\partial \beta_1}(y_i - \beta_1 - \beta_2 x_i)$$

$$\frac{\partial SSE}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)$$

The second step is the Chain Rule from Calculus
I can put the 2 in front of the summation because each term in the summation has a 2

Set the partial derivative to zero, in order to find minimum value

$$\frac{\partial SSE}{\partial \beta_1} = 0$$

Now solve equation for β_1,
It is debatable when you should add hats to the estimators
I added at this step when partial was set to zero

$$\begin{aligned}
\frac{\partial SSE}{\partial \beta_1} = 0 &= -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i\right) \\
&= \sum_{i=1}^{n}\left(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i\right) \\
&= \sum_{i=1}^{n}(y_i) + \sum_{i=1}^{n}\left(-\hat{\beta}_1\right) + \sum_{i=1}^{n}\left(-\hat{\beta}_2 x_i\right) \\
&= \sum_{i=1}^{n}(y_i) - n\hat{\beta}_1 - \hat{\beta}_2\sum_{i=1}^{n}(x_i)
\end{aligned}$$

Summation is a linear operator
We can apply the summation to all terms in parenthesis

- β_1 is constant that is summed n times
- β_2 is constant and multiplied by all x's in summation

Can bring this to the front

$$n\beta_1 = \sum_{i=1}^{n}(y_i) - \hat{\beta}_2 \sum_{i=1}^{n}(x_i)$$

$$\beta_1 = \frac{\sum_{i=1}^{n}(y_i)}{n} - \hat{\beta}_2 \frac{\sum_{i=1}^{n}(x_i)}{n}$$

$$\beta_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$

The last step works because we substitute the average for y and average for x into the equation

Repeating these steps to get the estimator for β_2

$$\frac{\partial SSE}{\partial \beta_2} = \frac{\partial}{\partial \beta_2} \sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)^2$$

$$\frac{\partial SSE}{\partial \beta_2} = 2\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)\frac{\partial}{\partial \beta_2}(y_i - \beta_1 - \beta_2 x_i)$$

$$\frac{\partial SSE}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)(x_i)$$

Similarly, set the partial to zero and solve for β_2,

$$\frac{\partial SSE}{\partial \beta_2} = 0 = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i\right)(x_i)$$

$$= -2\sum_{i=1}^{n}\left(y_i x_i - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2\right)$$

We substitute the estimator for β_1 into the equation,

$\beta_1 = \bar{y} - \hat{\beta}_2 \bar{x}$, which yields

$$0 = -2\sum_{i=1}^{n}\left(y_i x_i - \left(\bar{y} - \hat{\beta}_2\bar{x}\right)x_i - \hat{\beta}_2 x_i^2\right)$$

$$0 = \sum_{i=1}^{n}\left(y_i x_i - \left(\bar{y} - \hat{\beta}_2\bar{x}\right)x_i - \hat{\beta}_2 x_i^2\right)$$

$$0 = \sum_{i=1}^{n}\left(y_i x_i - \bar{y}x_i + \hat{\beta}_2\bar{x}x_i - \hat{\beta}_2 x_i^2\right)$$

$$0 = \sum_{i=1}^{n}(y_i x_i) - \bar{y}\sum_{i=1}^{n}(x_i) - \hat{\beta}_2\sum_{i=1}^{n}\left(x_i^2 - \bar{x}x_i\right)$$

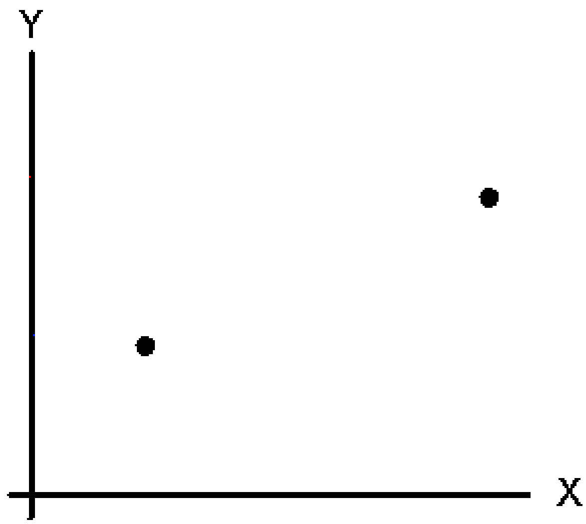
I did not break the last summation apart
This is to solve for the estimator for β_2

$$\hat{\beta}_2\sum_{i=1}^{n}\left(x_i^2 - \bar{x}x_i\right) = \sum_{i=1}^{n}(y_i x_i) - \bar{y}\sum_{i=1}^{n}(x_i)$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(y_i x_i) - \sum_{i=1}^{n}(\bar{y}x_i)}{\sum_{i=1}^{n}\left(x_i^2 - \bar{x}x_i\right)}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(y_i x_i) - \bar{y}\sum_{i=1}^{n}(x_i)}{\sum_{i=1}^{n}\left(x_i^2\right) - \bar{x}\sum_{i=1}^{n}(x_i)}$$

What if we had the graph below



Which line is the best fit? We could use algebra. This is not stochastic, since there is only two points

2. Desirable Properties of Estimators

No single estimator dominates all other estimators in terms of these properties

$$y_i = \beta_1 + \beta_2 x_i + u_i \qquad i = 1, 2, \cdots, n$$

Estimators of β_1 and β_2 are random.
Depend on the sample of x's and y's.
They have a probability distribution

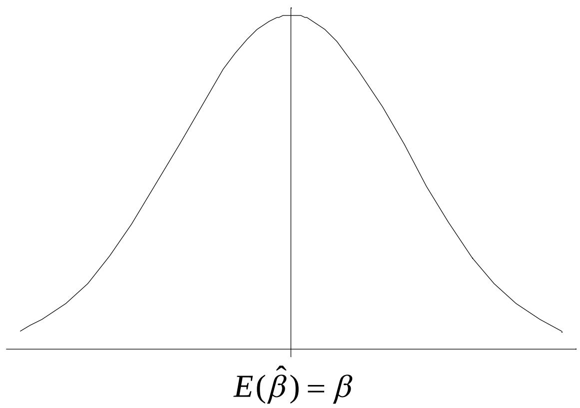
1. Unbiased

An estimator is unbias when the distribution of the estimator has the true value of the parameter as its mean value:

(7) $$E(\hat{\beta}) = \beta$$

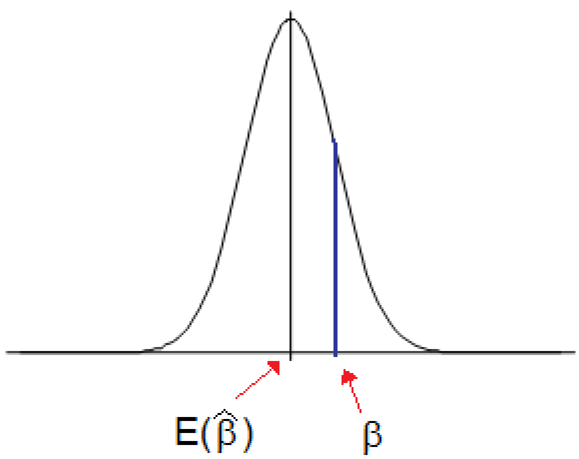
where E is the expectation operator
expected value is the mean

β is the true parameter value
$\hat{\beta}$ is the estimate of the parameter.



An estimator is bias, if the mean of the distribution of the estimator does not equal the true value. Mathematically, this is written as:

(8) $E(\hat{\beta}) \neq \beta$.



The amount of bias is given by:

(9) $\text{bias}(\hat{\beta}) = E(\hat{\beta}) - \beta$

The desirability of an unbiased estimator is that the mean of the estimator's distribution is centered around the true parameter value.

2. Efficiency – refers to variance

Variance is how spread out the data is
$\hat{\beta}$ is a more efficient estimator than $\tilde{\beta}$, if the sample variance of $\hat{\beta}$ is smaller than the variance of $\tilde{\beta}$.
This is difficult to determine the most efficient estimator.
The biased estimator has a smaller variance

3. Minimum Mean Squared Error – accounts for both bias and efficiency

MSE is defined as:

(10) $$MSE = \left[bias(\hat{\beta})\right]^2 + var(\hat{\beta})$$

Bias is could be positive or negative.
That is why it is squared.
If the estimator is unbiased, then the MSE = $var(\hat{\beta})$.

Example – choose the better estimator

	Estimator 1	Estimator 2
Bias	0	2
Variance	15	5

Remember – we are forecasting. Who cares if the estimator is biased.

4. Consistency – large sample properties

As the sample size approaches infinity, the estimator collapse to the true parameter value
Also called asymptotic