ECON 3150/4150, Spring term 2013. Lecture 1

Ragnar Nymoen

University of Oslo

15 January 2013

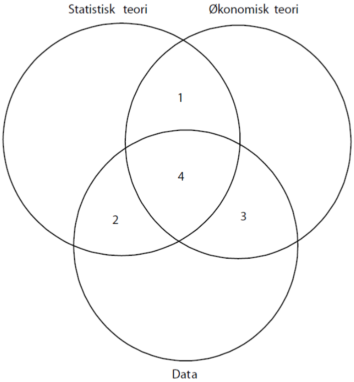
References to Lecture 1 and 2

- Hill, Griffiths and Lim, 4 ed (HGL)
  - Ch 1-1.5;
  - Ch 2.8-2.9,4.3-4.3.1.3
- Bårdsen and Nymoen (BN)
  - Kap 1-2
  - Kap 3
- Reference to ECON 2130 : "Om enkel lineær regresjon I" by Harald Goldstein
- A note about data types in econometrics by Erik Biorn (posted on the web page)

The goal of econometrics

- The Econometric Project: Use real world data and statistical theory to obtain empirical knowledge about relationships that hold outside the given sample.
- Statistical inference is a main concept: Generalization of empirical evidence from the data "to the population"
- Historically, inference theory was "imported" from mathematical statistics, and further developed to increase its relevance for economic data and theories.
- Inference is not only about confirmation of economic theory: In principle equally valuable to reject as to confirm that a theory holds in the population.
- *Econometric models* are the hallmark of econometrics

Econometrics is a combined discipline I



- Several of the intersections are of interest, but
- area 4 represents genuine *econometric models*

Econometric models I

- The main econometric models that we learn in this course are regression models with deterministic and stochastic (i.e. random) explanatory variables
- To begin with, we simplify and look at the model where there is a single explanatory variable which is deterministic (there is no uncertainty about which values the variable takes)
- We dub this Regression Model 1 (RM1)
- Next we explain the statistical theory needed to establish the regression model with stochastic explanatory variables, (RM2)
- Then we expand RM2 in different directions that are important for the relevance of the model for real world economic data.

Econometric models II

- We also explore “the limits of the regression model”, but the full introduction to other econometric model than the regression model is left for more advanced courses in econometrics

Data types I

- Cross section: A data set where the variables vary across n individuals $i = 1, 2, \ldots, n$
- Time series: A data set where the variables vary over T time periods: $t = 1, \ldots, T$
- Panel data: Variation both across individuals and over time.

There are other distinctions between data types as well'

- Micro/macro
- Experimental/non-experimental

Data types II

- In this course we will concentrate on the common ground between cross-section and time series data
- Will use notation like (Y_i, X_i) $i = 1, 2, \ldots, n$ for the most
- But will use (Y_t, X_t) $t = 1, 2, \ldots, T$ when it is relevant.

Econometrics courses in our programmes

- ECON 2130 *Statistikk 1*
- ECON 3145/4150 **Introductory Econometrics**
- ECON 4136 *Applied Statistics and Econometrics*
- ECON 4160 *Econometrics-Modelling and System Estimation*
- ECON 4130 Statistics 2
- ECON 5101/02/03 Advanced coursed in time series, panel data, micro econometrics

Econometric software

- **Stata**
  - An introduction to Stata is given in the Computer classes that start next week
  - Used in ECON 4136, and in advanced courses in micro and panel data
- **OxMetrics–PcGive**
  - Used in ECON 4160 (integrated CC and seminars)
  - ECON 5101, dynamic econometrics
- EViews
- TSP
- MicroFit
- RATS
- *Gretl*
- *R*

Historical and methodological background

The singular contribution to modern econometrics is *The Probability Approach to Econometrics* by Trygve Haavelmo from 1944.

PREFACE

This study is intended as a contribution to econometrics. It represents an attempt to supply a theoretical foundation for the analysis of interrelations between economic variables. It is based upon modern theory of probability and statistical inference. A few words may be said to justify such a study.

The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. But the bridge itself was never completely built. So far, the common procedure has been

Historical and methodological background

The singular contribution to modern econometrics is *The Probability Approach to Econometrics* by Trygve Haavelmo from 1944.

PREFACE

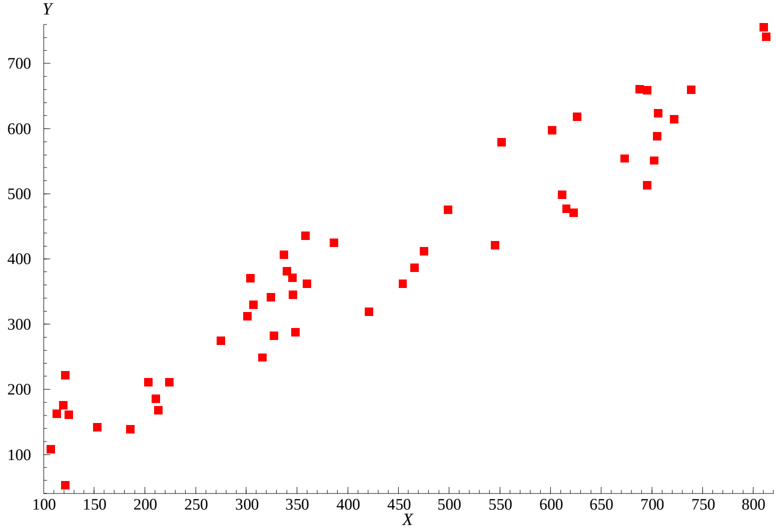
This study is intended as a contribution to econometrics. It represents an attempt to supply a theoretical foundation for the analysis of interrelations between economic variables. It is based upon modern theory of probability and statistical inference. A few words may be said to justify such a study.

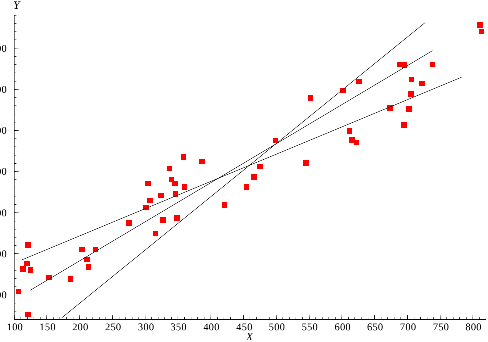
The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. But the bridge itself was never completely built. So far, the common procedure has been

Econometric models and regression

- Econometric models are the conjunction that Haavelmo spoke of.
- As mentioned: In this course, regression models will be central
- But we start with a review of regression as a method of fitting a straight line to the observed data points
- To keep it apart from regression models, we call this “mere” curve fitting, or regression without statistical inference.
- For those who took "Statistikk 1", a very good reference for review is Harald Goldstein's *Om enkel lineær regresjon I*

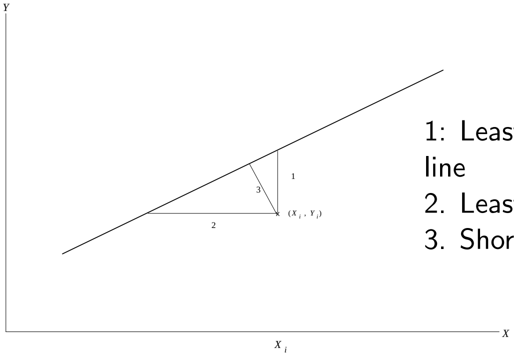
Scatter plot and least squares fit



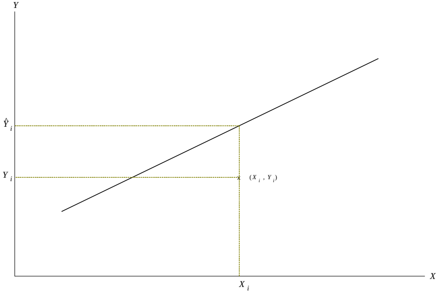


- Which line is best?
- Idea: Minimize sum of squared errors!
- But which errors?

Which squared error?

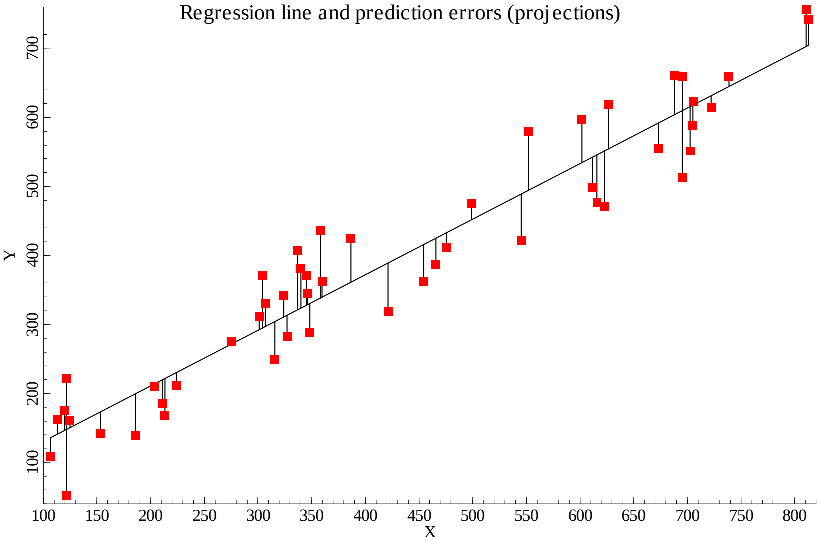


1: Least vertical distance to line
2. Least horizontal
3. Shortest distance to line



- Choose 1 when want to minimize squared errors from predicting Y_i linearly from X_i
- Residual:

$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$, where $\hat{Y}_i$ is predicted value


Regression line and prediction errors (projections)
Y
X
100
200
300
400
500
600
700
100
150
200
250
300
350
400
450
500
550
600
650
700
750
800

Ordinary least squares (OLS) estimates I

- The different lines that we considered placing in the scatter-plot correspond to different values of the parameter β_0 and β_1 in the linear function that connects given numbers X_1,X_2, $\ldots X_n$ with Y_1^{fitted},Y_2^{fitted}, $\ldots$, Y_n^{fitted}:

$$Y_i^{fitted} = \beta_0 - \beta_1 X_i,\ i = 1, 2, \ldots, n$$

- We obtain the best fit $Y_i^{fitted} \equiv \hat{Y}_i$ $(i = 1, 2, \ldots, n)$

$$\hat{Y}_i = \hat{\beta}_0 - \hat{\beta}_1 X_i,\ i = 1, 2, \ldots, n \tag{1}$$

by finding the estimates of β_0 and β_1 that minimizes the sum of squared residuals $\sum_{i=1}^{n} \left(Y_i - Y_i^{fitted}\right)^2$:

$$S(\beta_0,\beta_2) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{2}$$

Ordinary least squares (OLS) estimates II

Consequently $\hat{\beta}_0$ and $\hat{\beta}_1$ are determined by the 1oc's:

$$\overline{Y} - \hat{\beta}_0 - \hat{\beta}_1 \overline{X} = 0 \tag{3}$$

$$\sum_{i=1}^{n} X_i Y_i - \hat{\beta}_0 \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0 \tag{4}$$

where

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{5}$$

is the *sample mean* (*empirical mean*) of X.

- It is expected that you can solve the simultaneous equation system (3)-(4). Work with exercises to Seminar 1!

A trick and a simplified derivation I

The trick is to note that

$$\beta_0 + \beta_1 X_i \equiv \alpha + \beta_1 (X_i - \overline{X}) \tag{6}$$

when the intercept parameter α is defined as

$$\alpha \equiv \beta_0 + \beta_1 \overline{X} \tag{7}$$

This means that the best prediction $\hat{Y}_i$ given X_i can be written as

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \equiv \hat{\alpha} + \hat{\beta}_1 (X_i - \overline{X}) \text{ where } \hat{\alpha} \equiv \hat{\beta}_0 + \hat{\beta}_1 \overline{X} \tag{8}$$

and we therefore choose the α and β_1 that minimize

$$S(\alpha, \beta_1) = \sum_{i=1}^{n} \left[Y_i - \alpha - \beta_1 (X_i - \bar{X}) \right]^2 \tag{9}$$

A trick and a simplified derivation II

Calculate the two partial derivatives ("kjerneregelen" for each element in the sums):

$$\frac{\partial S(\alpha,\beta_1)}{\partial\alpha} = 2\sum_{i=1}^{n}\left[Y_i - \alpha - \beta_1(X_i - \bar{X})\right]\cdot(-1)$$
$$\frac{\partial S(\alpha,\beta_1)}{\partial\beta_1} = 2\sum_{i=1}^{n}\left[Y_i - \alpha - \beta_1(X_i - \bar{X})\right]\cdot -(X_i - \bar{X})$$

and choose $\hat{\alpha}$ and $\hat{\beta}_1$ as the solutions of

$$2\sum_{i=1}^{n}\left[Y_i - \hat{\alpha} - \hat{\beta}_1(X_i - \bar{X})\right]\cdot(-1) = 0 \tag{10}$$
$$2\sum_{i=1}^{n}\left[Y_i - \hat{\alpha} - \hat{\beta}_1(X_i - \bar{X})\right]\cdot -(X_i - \bar{X}) = 0 \tag{11}$$

A trick and a simplified derivation III

$$\hat{\alpha} - \bar{Y} = 0 \tag{12}$$

$$\sum_{i=1}^{n}(X_i - \bar{X})Y_i - \hat{\beta}_1 \sum_{i=1}^{n}(X_i - \bar{X})^2 = 0 \tag{13}$$

where

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{14}$$

the empirical mean of Y.

DIY exercise 1: Show that (10) gives (12), and (11) gives (13) and that the solutions of (12) and (13) are

$$\hat{\alpha} = \bar{Y}, \tag{15}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \tag{16}$$

A trick and a simplified derivation IV

Note that for (16) to make sense, we need to assume

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 > 0 \text{ (i.e., } X \text{ is a variable, not a constant)}$$

A generalization of this will be important later, and is then called "absence of perfect multicollinearity".
To obtain $\hat{\beta}_0$ we simply use

$$\hat{\beta}_0 = \hat{\alpha} - \hat{\beta}_1 \overline{X} \tag{17}$$

Residuals and total sum of squares I

Definition of OLS residuals:

$$\hat{\varepsilon}_i = Y_i - \widehat{Y}_i,\ i = 1, 2, ..., n \tag{18}$$

Using this definition in the 1oc's (10) and (13) gives

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \tag{19}$$

$$\sum_{i=1}^{n} \hat{\varepsilon}_i (X_i - \overline{X}) = 0. \tag{20}$$

Residuals and total sum of squares II

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \implies \bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \quad (21)$$

$$\sum_{i=1}^{n} \hat{\varepsilon}_i (X_i - \overline{X}) = 0 \implies \hat{\sigma}_{\varepsilon X} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\varepsilon}_i - \bar{\varepsilon})(X_i - \overline{X}) = 0 \quad (22)$$

where $\hat{\sigma}_{\varepsilon X}$ denotes the (empirical) covariance between the residuals and the explanatory variable.

- These properties always hold when we include the intercept (β_0 or α) in the model
- They generalize to the case of multiple regression as we shall later
- (22) is an *orthogonality condition*. It says that the OLS residuals are uncorrelated with the explanatory variable.

Residuals and total sum of squares III

- $\hat{\sigma}_{\varepsilon X} = 0$ occurs because we have defined the OLS residuals in such a way that they measure "what is left unexplained in Y" when we have extract all the explanatory power of X

Total Sum of Squares and Residual Sum of Squares I

We define the Total Sum of Squares for Y as

$$TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{23}$$

We can guess that TSS can be split in Explained Sum of Squares

$$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \overline{\hat{Y}})^2 \tag{24}$$

and Residual Sum of Squares

$$RSS = \sum_{i=1}^{n}(\hat{\varepsilon}_i - \bar{\varepsilon})^2 \tag{25}$$

Total Sum of Squares and Residual Sum of Squares II

so that

$$TSS = ESS + RSS \tag{26}$$

To show this important decomposition, start with

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left[\underbrace{(Y_i - \hat{Y}_i)}_{\hat{\varepsilon}_i} + (\hat{Y}_i - \overline{\hat{Y}})\right]^2$$

where we have used that $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n}(\hat{\varepsilon}_i + \hat{Y}_i) = \overline{\hat{Y}}$ because of (19).
Completing the square gives

Total Sum of Squares and Residual Sum of Squares III

$$\underbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}_{TSS} = RSS + 2\sum_{i=1}^{n}\hat{\varepsilon}_i(\hat{Y}_i - \overline{\hat{Y}}) + ESS$$

Expand the middle term:

$$\begin{aligned}\sum_{i=1}^{n}\hat{\varepsilon}_i(\hat{Y}_i - \overline{\hat{Y}}) &= \sum_{i=1}^{n}\hat{\varepsilon}_i(\hat{\alpha} + \hat{\beta}_1(X_i - \overline{X}) - \overline{\hat{Y}}) \\ &= \hat{\alpha}\underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i}_{(19)} + \hat{\beta}_1\underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i(X_i - \overline{X})}_{(20)} - \overline{\hat{Y}}\underbrace{\sum_{i=1}^{n}\hat{\varepsilon}_i}_{(19)}\end{aligned}$$

Total Sum of Squares and Residual Sum of Squares IV

Therefore

$$\sum_{i=1}^{n} \hat{\varepsilon}_i (\hat{Y}_i - \overline{\hat{Y}}) = 0$$

- The residuals are uncorrelated with the predictions $\hat{Y}_i$.
- Could it be different?
- Hence we have the desired result:

$$TSS = ESS + RSS \tag{27}$$

The coefficient of determination I

To summarize the "goodness of fit" in the form of a single number, *the coefficient of determination*, almost everywhere denoted R^2, is used:

$$R^2 = \frac{ESS}{TSS} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \tag{28}$$
$$= 1 - \text{rate of unexplained } Y \text{ variation}$$

- If $\hat{\beta}_1 = 0$,
  - $RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i - \bar{\varepsilon})^2 = \sum_{i=1}^{n}(Y_i - \hat{\alpha})^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = TSS$.
  - and $R^2 = 0$
- If $RSS = 0$,a perfect fit, then $R^2 = 1$

The coefficient of determination II

- Hence we have the property

$$0 \leq R^2 \leq 1 \tag{29}$$

These results depend on defining the regression function as

$$\hat{Y}_i = \hat{\beta}_0 - \hat{\beta}_1 X_i,$$

as in (1).
If we instead use

$$\hat{Y}_i^{no-i} = \hat{\beta}_1^{no-i} X_i$$

which forces the "regression line" trough the origin:

- the corresponding residuals do not sum to zero,
- the decomposition of TSS breaks down.
- R^2 (as defined above) can be negative!
- Work with Exercises to Seminar 1 to learn more!

Regression and correlation I

We define the *empirical correlation coefficient* between X and Y as

$$r_{X,Y} \equiv \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{X})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X \hat{\sigma}_Y}, \quad (30)$$

- $\hat{\sigma}_{X,Y}$ denotes the empirical covariance between Y and X
- $\hat{\sigma}_X$ and $\hat{\sigma}_Y$ denote the two *empirical standard deviations*
- They are square roots of the empirical variances, e.g., $\hat{\sigma}_X = \sqrt{\hat{\sigma}_X^2} = \sqrt{1/n\sum_{i=1}^{n}(X_i - \overline{X})^2}$

Regression and correlation II

$\hat{\sigma}_{X,Y}$ can be written in three equivalent ways:

$$\hat{\sigma}_{X,Y} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})Y_i$$
$$= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})X_i$$

The regression coefficient can therefore be re-expressed as

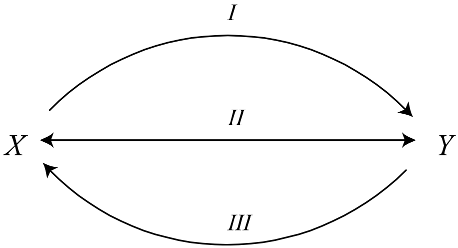
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2}$$
$$= \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}\frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X\hat{\sigma}_Y} = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}r_{X,Y} \tag{31}$$

This shows that

Regression and correlation III

- $r_{X,Y} \neq 0$ is necessary for $\hat{\beta}_1 \neq 0$. Correlation is necessary for finding regression relationships
- Still, $\hat{\beta}_1 \neq r_{XY}$ (in general) and regression analysis is different from correlation analysis.

Regression and causality I



Three possible theoretical causal relationships between X and Y.

- Our regression is causal if I is true, and II (joint causality) and III are not true
- $r_{XY} \neq 0$ in all three cases
- Can also be that a third variable (Z) causes both Y and X (spurious correlation)

Causal interpretation of regression analysis I

- Regression analysis can refute a causal relationship, since correlation is necessary for causality
- But cannot confirm or discover a causal relationship by regression analysis alone
- However, regression analysis is usually done with reference to a conceptual framework (a theory) that points out one direction of causality as more likely or relevant than others.
- For example, when we choose X as the regressor and Y as the regressand, that choice is usually done with reference to a theory that independent changes in X cause a response in Y. If that theory is supported by other evidence, that strengthens the relevance (belief in) a causal interpretation of our regression (combined discipline!)

Causal interpretation of regression analysis II

- Causality represents the sought after "holy grail" in many econometric studies, and recently big advances has been made in studies that utilize "natural experiments", often with the use of large micro data set.
- In time series, there are also possibilities of investigating causality further. The central concept is autonomy with respect to structural breaks. We will look at simple example

- Consider a sample with time series data for X_t and Y_t $(t = 1, 2, \dots T)$
- Time series data has a natural ordering of observations, from past to present, and this is helpful in causality analysis
- Start with recording the two possible regression coefficients:
  - Regress Y on X

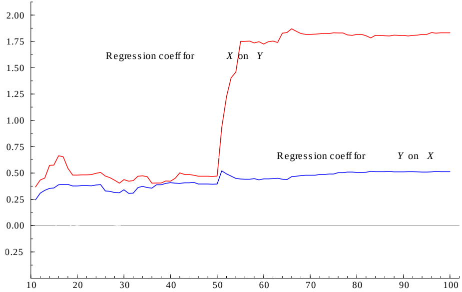
$$\hat{\beta}_1 = \frac{\sum_{t=1}^{T}(X_t - \bar{X})Y_t}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

  - Regress X on Y:

$$\hat{\beta}_1' = \frac{\sum_{t=1}^{T}(X_t - \bar{X})Y_t}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

- Therefore:
$$\hat{\beta}_1\hat{\beta}_1' = r_{XY}^2 \tag{32}$$
- Let us now "sharpen" the notion of causality by demanding that a causal relationship in characterized by parameters that are *invariant* to structural breaks in the sample $(t = 1, 2, \ldots T)$.
- A candidate for a structural break is the correlation coefficient r_{XY}. (32) says that if there is a break in r_{XY} then at least one of $\hat{\beta}_1$and $\hat{\beta}_1'$ must change at the same point in time.
- However if either $\hat{\beta}_1$ or $\hat{\beta}_1'$s are constant despite the break in r_{XY}, we see that the invariance property is validated, and that we have evidence of a one-way causal relationship.

An example of invariance testing I



- The graphs show recursive estimates of $\hat{\beta}_1$ and $\hat{\beta}'_1$
- There is a structural break in period 50 (a higher σ_X that reduces r_{XY})
- $\hat{\beta}'_1$ is not invariant to the break, but $\hat{\beta}_1$ is invariant, supporting that $X \to Y$