Environmental Econometrics

Jérôme Adda

j.adda@ucl.ac.uk

Office # 203

Syllabus

Course Description:

This course is an introductory econometrics course. There will be 2 hours of lectures per week and a class (in the computer lab) each week. No previous knowledge of econometrics is assumed. By the end of the term, you are expected to be at ease with basic econometric techniques such as setting up a model, testing assumptions and have a critical view on econometric results. The computer classes introduce you to real life problems, and will help you to understand the theoretical content of the lectures. You will also learn to use a powerful and widespread econometric software, STATA.

Understanding these techniques will be of great help for your thesis over the summer, and will help you in your future workplace.

For any contact or query, please send me an email or visit my web page at:

`http://www.ucl.ac.uk/~uctpjea/teaching.html`.

My web page contains documents which might prove useful such as notes, previous exams and answers.

Books:

There are a lot of good basic econometric books but the main book to be used for reference is the Wooldridge (J. Wooldridge (2003) Introductory Econometrics, MIT Press.). Other useful books are:

- Gujurati (2001) Basic Econometrics, Mc Graw-Hill. (Introductory text book)
- Wooldridge (2002) Econometric Analysis of Cross Section and Panel Data, MIT Press. (More advanced).
- P. Kennedy, 3rd edition (1993) A Guide to Econometrics, Blackwell. (Easy, no maths).

Course Content

1. Introduction
   What is econometrics? Why is it useful?

2. The linear model and Ordinary Least Squares
   Model specification. Introduction to simple regression and method of ordinary least squares (OLS) estimation.

3. Extension to multiple regression
   Properties of OLS. Omitted variable bias. Measurement errors.

4. Hypothesis Testing
   Goodness of fit, R^2. Hypothesis tests (t and F).

5. Heteroskedasticity and Autocorrelation
   Generalized least squares. Heteroskedasticity: Examples; Causes; Consequences; Tests; Solutions. Autocorrelation: Examples; Causes; Consequences; Tests; Solutions.

6. Simultaneous Equations and Endogeneity
   Simultaneity bias. Identification. Estimation of simultaneous equation models. Measurement errors. Instrumental variables. Two stage least squares.

7. Limited Dependent Variable Models
   Problem Using OLS to estimate models with 0-1 dependent variables. Logit and probit models. Censored dependent variables. Tobit models.

8. Time Series
   AR and MA Processes. Stationarity. Unit roots.

Definition and Examples

Econometrics: statistical tools applied to economic problems.

Examples: using data to:

- Test economic hypotheses.
- Establish a link between two phenomenons.
- Assess the impact and effectiveness of a given policy.
- Provide an evaluation of the impact of future public policies.

Provide a qualitative but also a quantitative answer.

Example 1: Global Warming

- Measuring the extent of global warming.
  - when did it start?
  - How large is the effect?
  - has it increased more in the last 50 years?

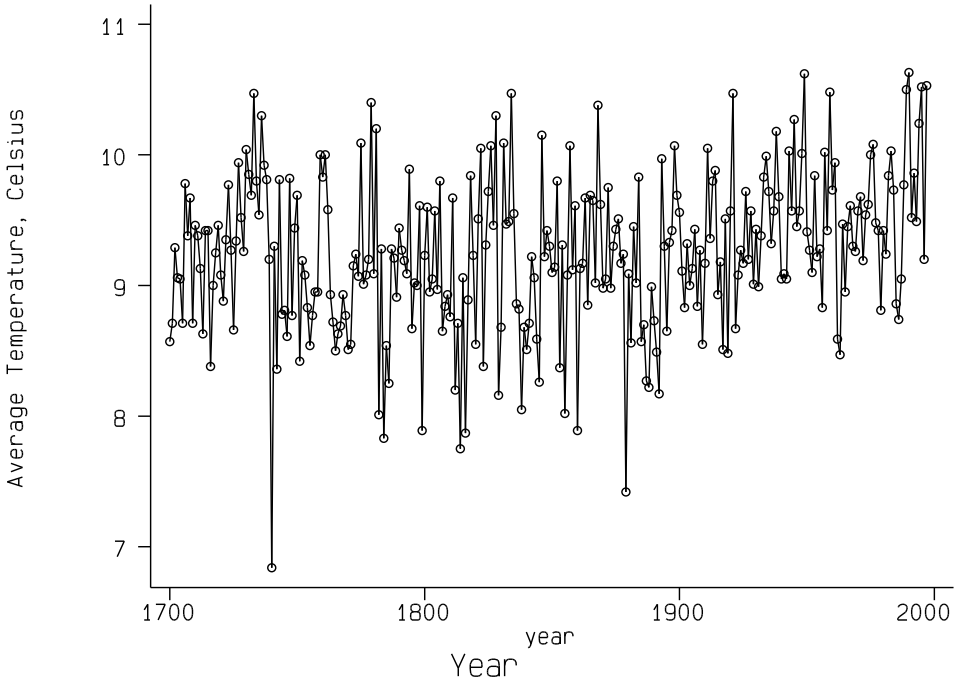
- What are the causes of global warming?
  - Does carbon dioxide *cause* global warming?
  - Are there other determinants?
  - What are their respective importance?

- Average temperature in 50 years if nothing is done?

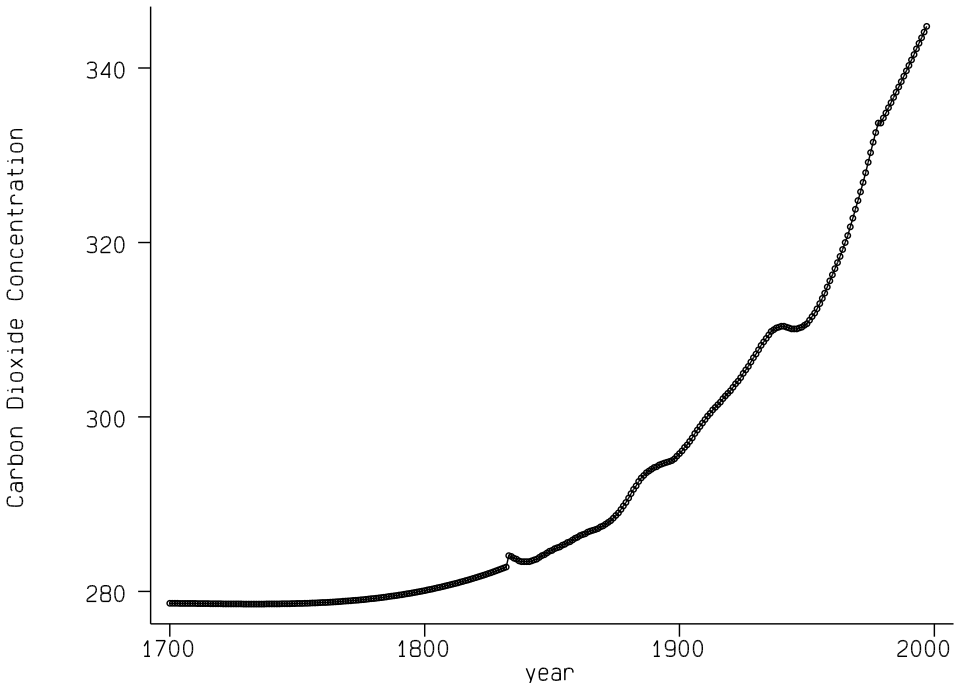
- Average temperature if carbon dioxide concentration is reduced by 10%?

Example 1: Global Warming

Average Temperature in Central England (1700-1997)



Atmospheric Concentration of Carbon Dioxide (1700-1997)



Example 2:
Willingness to Pay for new Policy

Data on WTP for better waste service management in Kuala Lumpur. Survey of 500 households.

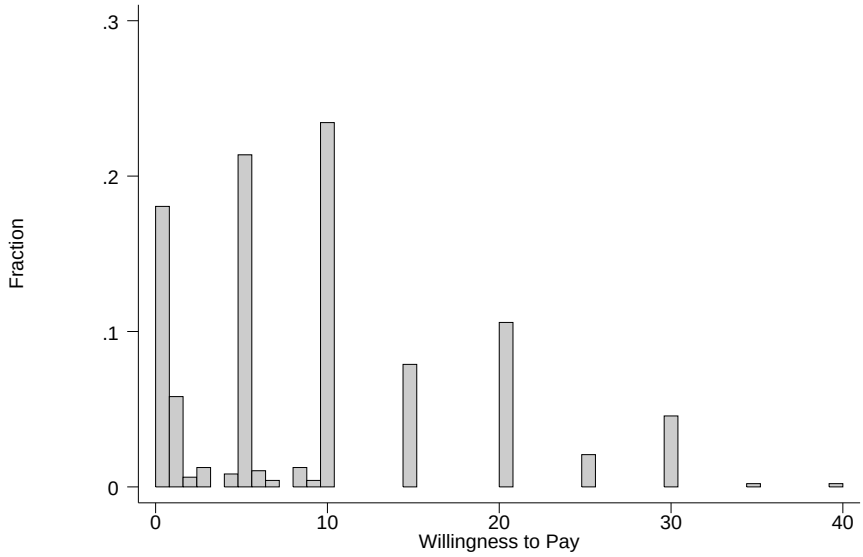
- How is WTP distributed?

- Is WTP influenced by income?

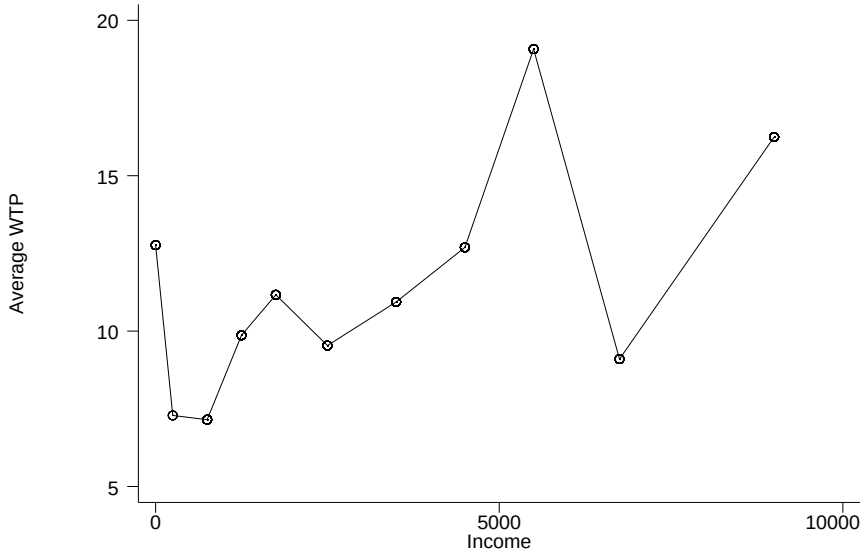
- What is the effect on WTP of a 10% tax cut on income tax?

Example 2: WTP

Distribution of WTP for better Service



WTP and Income



Causality

- We often observe that two variables are correlated.
  - Examples:
    - Individuals with higher education earn more.
    - Parental income is correlated with child's education.
    - Smoking is correlated with peer smoking.
    - Income and health are correlated.

- However this does NOT establish causal relationships.

Causality

- If a variable Y is causally related to X, then changing X will LEAD to a change in Y.
  - For example: Increasing VAT may cause a reduction of demand.
  - Correlation may **not** be due to causal relationship:
    * Part or the whole correlation may be induced by both variables depending on some common factor and does not imply causality.
    * For example: Individuals who smoke may be more likely to be found in similar jobs. Hence, smokers are more likely to be surrounded by smokers, which is usually taken as a sign of peer effects. The question is how much an increase in smoking by peers results in higher smoking.
    * Brighter people have more education AND earn more. The question is how much of the increased in earnings is caused by the increased education.

Causality

- The course in its more advanced phase will deal with the issue of causality and ways that we have of establishing and measuring causal relationships

The Regression Model

- The basic tool in Econometrics is the Regression Model.
- Its simplest form is the two variable linear regression Model:

$$Y_i = \alpha + \beta X_i + u_i$$

Explanation of Terms:

- Y_i: The DEPENDENT Variable. The Dependent Variable is the variable we are modeling.
- X_i: The EXPLANATORY variable. The Explanatory Variable X is the variable of interest whose impact on Y we wish to measure.
- u_i: the error term. The error term reflects *all other factors* determining the dependent variable.
- $i = 1, \ldots, N$: The observation indicator.
- α and β are parameters to be estimated.

Example:

$$\text{Temperature}_i = \alpha + \beta \text{ year}_i + u_i$$

Assumptions

- During most of the lectures we will assume that u and X are NOT correlated.
- This assumption will allow us to interpret the coefficient β as the effect of X on Y.
- Note that
$$\beta = \frac{\partial Y_i}{\partial X_i}$$
which we will call the marginal effect of X on Y.
- This coefficient will be interpreted as the ceteris paribus impact of a change in X on Y.
- **Aim:** To use data to estimate the coefficients α and β.

Key Issues

The Key issues are:

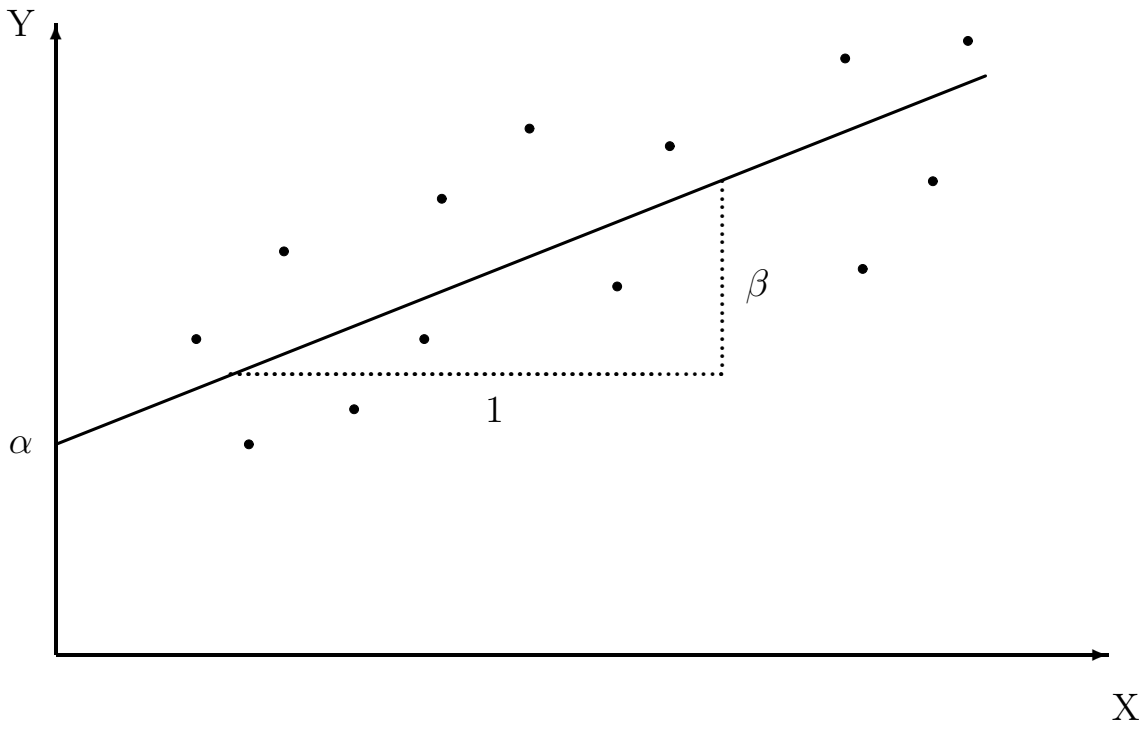
- Estimating the Coefficients of the regression line that fits this data best in the most efficient way possible.
- Making inferences about the model based on these estimates.
- Using the model.

Regression Line

Model :

$$Y_i = \alpha + \beta X_i + u_i$$

Graphical Interpretation:

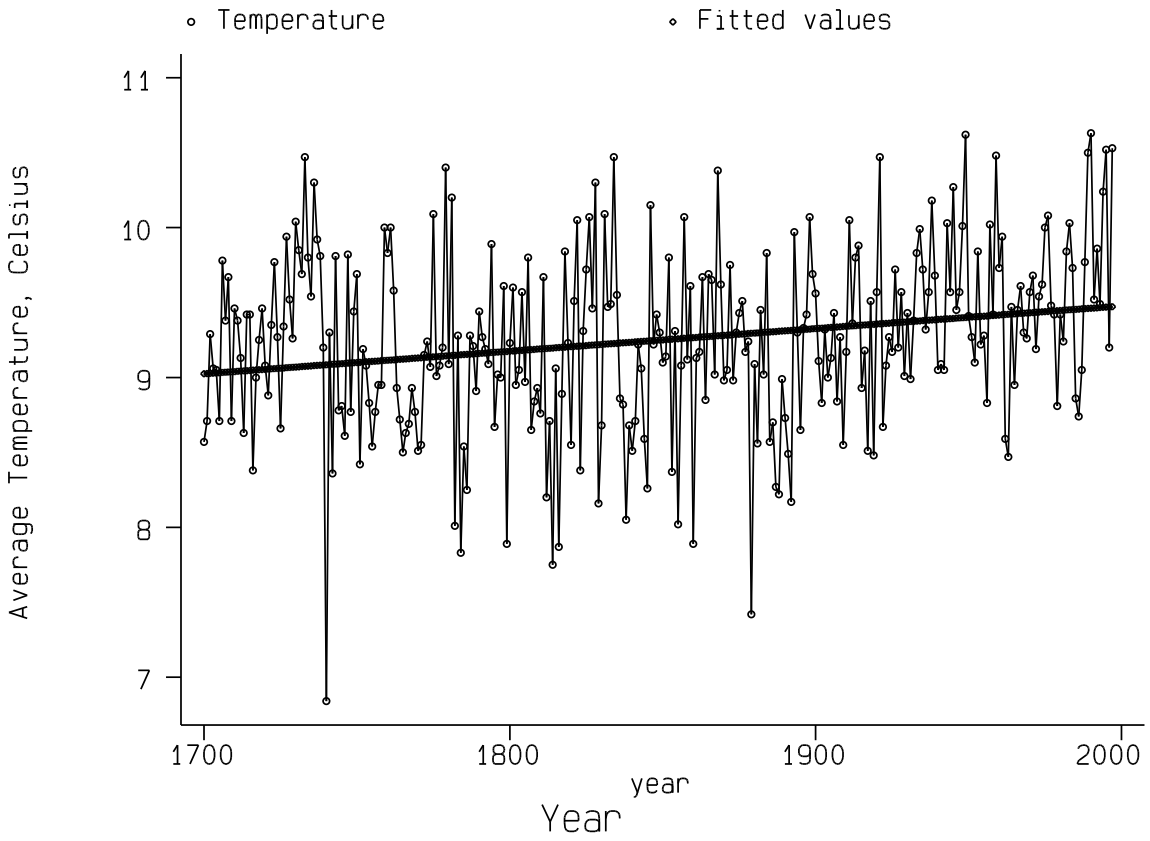


- The distance between any point and the fitted line is the estimated residual.
- This summarizes the impact of other factors on Y.
- As we will see, the chosen best line is fitted using the assumption that these other factors are not correlated with X.

An Example: Global Warming

The Fitted Line

Intercept (β_0): 6.45
Estimated slope (β_1): 0.0015



Model Specifications

- **Linear model:**

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$
$$\frac{\partial Y_i}{\partial X_i} = \beta_1$$

Interpretation: When X goes up by 1 unit, Y goes up by β_1 units.

- **Log-Log model** (constant elasticity model):

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

$$Y_i = e^{\beta_0} X_i^{\beta_1} e^{u_i}$$

$$\begin{aligned} \frac{\partial Y_i}{\partial X_i} &= e^{\beta_0} \beta_1 X_i^{\beta_1 - 1} e^{u_i} \\ \frac{\partial Y_i / Y_i}{\partial X_i / X_i} &= \beta_1 \end{aligned}$$

Interpretation: When X goes up by **1%**, Y goes up by β_1 **%**.

- **Log-lin model:**

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

$$\begin{aligned} \frac{\partial Y_i}{\partial X_i} &= \beta_1 e^{\beta_0} e^{\beta_1 X_i} e^{u_i} \\ \frac{\partial Y_i / Y_i}{\partial X_i} &= \beta_1 \end{aligned}$$

Interpretation: When X goes up by 1 **unit**, Y goes up by $100\beta_1$ **%**.

An Example: Global Warming

- Linear Model:

$$T_i = \beta_0 + \beta_1 \text{year}_i + u_i$$

  - T_i: average annual temperature in central England, in Celsius.

OLS Results, linear model:

Variable	Estimates
β_0 (constant)	6.45
β_1 (year)	0.0015

On average, the temperature goes up by 0.0015 degrees each year, so 0.15 each centuries.

- Log-Lin Model:

$$\ln(Temperature_i) = \beta_0 + \beta_1 year_i + u_i$$

OLS Results, Log-Lin Model: :

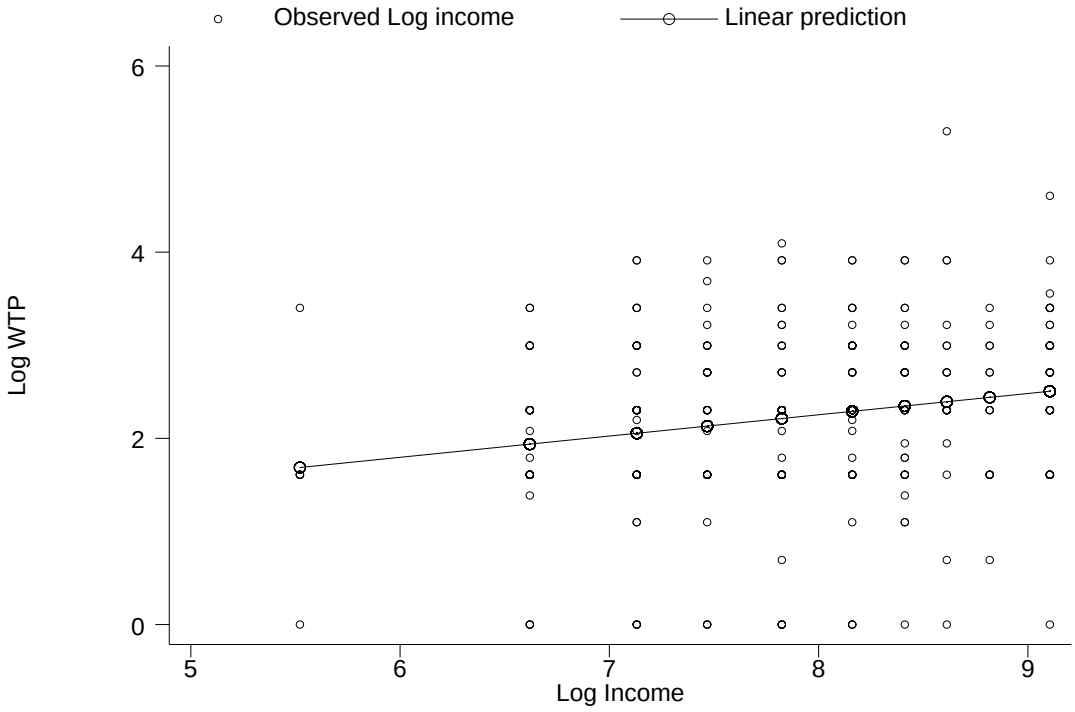
Variable	Estimates
β_0 (constant)	2.17
β_1 (year)	0.00023

The temperature goes up by 0.023% each year, so 2.3% each centuries.

An Example: WTP

Log WTP and Log Income

Intercept: 0.42
slope: 0.23



- A one percent increase in income increases WTP by 0.23%.

- So a 10% tax cut would increase WTP by 2.3%.

More Advanced Models

- In many occasions we will consider more elaborate models where a number of explanatory variables will be included.

- The regression models in this case will take the more general form:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_{ki} + u_i$$

- There are k explanatory variables and a total of $k + 1$ coefficients to estimate (including the intercept).

- Each coefficient represents the ceteris paribus effect of changing one variable.

Data Sources

- Time Series Data:
  - Data on variables observed over time. Typically Macroeconomic measures such as GDP, Inflation, Prices, Exchange Rates, Interest Rates, etc.
  - Used to study and simulate macroeconomic relationships and to test macro hypotheses
- Cross Section Survey Data:
  - Data at a given point in time on individuals, households or firms. Examples are data on expenditures, income, hours of work, household composition, investments, employment etc.
  - Used to study household and firm behaviour when variation over time is not required.
- Panel Data:
  - Data on individual units followed over time.
  - Used to study dynamic aspects of household and firm behaviour and to measure the impact of variables that vary predominantly over time.

Type of variables

- continuous.
  - temperature.
  - age.
  - income.
- categorical/ qualitative
  - ordered
    - answers such that small /medium /large.
    - income coded into categories.
  - non ordered
    - answers such that Yes/No, Blue/Red, Car/Bus/Train.

The linear model we have written accommodate well continuous variables as they have units. From now on, we will assume that the dependent variable is continuous. The course will explain later on how to deal with qualitative depend variables.

Properties of OLS

The Model

- We return to the classical linear regression model to learn formally how best to estimate the unknown parameters. The model is:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- where β_0 and β_1 are the coefficients to be estimated.

Assumptions of the Classical Linear Regression Model

- **Assumption 1:** $E(u_i|X) = 0$
  - The expected value of the error term has mean zero *given any value of the explanatory variable*. Thus observing a high or a low value of X does not imply a high or a low value of u.

    X and u are uncorrelated.
  - This implies that changes in X are not associated with changes in u in any particular direction - Hence the associated changes in Y can be attributed to the impact of X.
  - This assumption allows us to interpret the estimated coefficients as reflecting causal impacts of X on Y.
  - Note that we condition on the whole set of data for X in the sample not on just one .

Assumptions of the Classical Linear Regression Model

- **Assumption 2:** HOMOSKEDASTICITY (Ancient Greek for Equal variance)

$$Var(u_i|X) \equiv E(u_i - E(u_i|X)|X)^2 = E(u_i^2|X) = \sigma^2$$

where σ^2 is a positive and finite constant that does not depend on X.

  - This assumption is not of central importance, at least as far as the interpretation of our estimates as causal is concerned.
  - The assumption will be important when considering hypothesis testing.
  - This assumption can easily be relaxed. We keep it initially because it makes derivations simpler.

Assumptions of the Classical Linear Regression Model

- **Assumption 3:** The error terms are uncorrelated with each other.

$$cov(u_i, u_j|X) = 0 \quad \forall i, j \quad i \neq j$$

  - When the observations are drawn sequentially over time (time series data) we say that there is no serial correlation or no autocorrelation.
  - When the observations are cross sectional (survey data) we say that we have no spatial correlation.
  - This assumption will be discussed and relaxed later in the course.

- **Assumption 4:** The variance of X must be non-zero.

$$Var(X_i) > 0$$

  - This is a crucial requirement. It states the obvious: To identify an impact of X on Y it must be that we observe situations with different values of X. In the absence of such variability there is no information about the impact of X on Y.

- **Assumption 5:** The number of observations N is larger than the number of parameters to be estimated.

Fitting a regression model to the Data

- Consider having a sample of N observations drawn randomly from a population. The object of the exercise is to estimate the unknown coefficients β_0 and β_1 from this data.
- To fit a model to the data we need a method that satisfies some basic criteria. The method is referred to as an **estimator**. The numbers produced by the method are referred to as **estimates**; i.e. we need our estimates to have some desirable properties.
- We will focus on two properties for our estimator:
  - Unbiasedness
  - Efficiency [We will leave this for the next lecture]

Unbiasedness

- We want our estimator to be unbiased.
- To understand the concept first note that there actually exist **true** values of the coefficients which of course we do not know. These reflect the true underlying relationship between Y and X. We want to use a technique to estimate these true coefficients. Our results will only be approximations to reality.
- An unbiased estimator is such that the average of the estimates, across an infinite set of different samples of the same size N, is equal to the true value.
- Mathematically this means that

$$E(\hat{\beta}_0) = \beta_0 \qquad \text{and} \qquad E(\hat{\beta}_1) = \beta_1$$

where the $\hat{}$ denotes an estimated quantity.

An Example

True Model:

$$Y_i = 1 + 2X_i + u_i$$

Thus $\beta_0 = 1$ and $\beta_1 = 2$.

	$\hat{\beta}_0$	$\hat{\beta}_1$
Sample 1	1.2185099	1.5841877
Sample 2	.82502003	2.5563998
Sample 3	1.3752522	1.3256603
Sample 4	.92163564	2.1068873
Sample 5	1.0566855	2.1198698
Sample 6	1.048275	1.8185249
Sample 7	.91407965	1.6573014
Sample 8	.78850225	2.9571939
Sample 9	.65818798	2.2935987
Sample 10	1.0852489	2.3455551
Average across samples	.9891397	2.0765179
Average across 500 samples	.98993739	2.0049863

Each sample has 14 observations in all cases (N=14).

Ordinary Least Squares (OLS)

- The Main method we will focus on is OLS, also referred to as Least squares.
- This method chooses the line so that sum of squared residuals (squared vertical distances of the data points from the fitted line) are minimized.
- We will show that this method yields an estimator that has very desirable properties. In particular the estimator is unbiased and efficient (see next lecture)
- Mathematically this is a very well defined problem:

$$\min_{\beta_0,\beta_1} \frac{1}{N}\sum_{i=1}^{N} u_i^2 = \min_{\beta_0,\beta_1} \frac{1}{N}\sum_{i=1}^{N}(Y_i - \beta_0 - \beta_1 X_i)^2$$

First Order Conditions

$$\frac{\partial L}{\partial \beta_0} = -\frac{2}{N}\sum_{i=1}^{N}(Y_i - \beta_0 - \beta_1 X_i) = 0$$

$$\frac{\partial L}{\partial \beta_1} = -\frac{2}{N}\sum_{i=1}^{N}(Y_i - \beta_0 - \beta_1 X_i)X_i = 0$$

This is a set of two simultaneous equations for β_0 and β_1. The estimator is obtained by solving for β_0 and β_1 in terms of means and cross products of the data.

The Estimator

- Solving for β_0 we get

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

  where the bar denotes sample average

- Solving for β_1 we get

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

- Thus the estimator of the slope coefficient can be seen to be the the ratio of the covariance of X and Y to the variance of X.

- We also observe from the first expression that the regression line will always pass through the mean of the data.

- Define the fitted values as

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

  These are also referred to as predicted values.

- The residual is defined as

$$\hat{u}_i = Y_i - \hat{Y}_i$$

Deriving Properties

- First note that within a sample

$$\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u}$$

- Hence

$$Y_i - \bar{Y} = \beta_1 (X_i - \bar{X}) + (u_i - \bar{u})$$

- Substitute this in the expression for β_1 to obtain

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} \left[\beta_1 (X_i - \bar{X})^2 + (X_i - \bar{X})(u_i - \bar{u})\right]}{\sum_{i=1}^{N} (X_i - \bar{X})^2}$$

- Hence, this leads to:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N} (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{N} (X_i - \bar{X})^2}$$

The second part of this expression is called the sample or estimation error. If the estimator is unbiased then this error will have expected value zero.

Deriving Properties, cont.

$$
\begin{aligned}
E(\hat{\beta}_1|X) &= \beta_1 + E\left[\frac{\sum_{i=1}^{N}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{N}(X_i - \bar{X})^2}|X\right] \\
&= \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})E[(u_i - \bar{u})|X]}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \\
&= \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X}) \times 0}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \qquad \text{(using Assumption 1)} \\
&= \beta_1
\end{aligned}
$$

Goodness of Fit

- We measure how well the model fits the data using the R^2.
- This is the ratio of the explained sum of squares to the total sum of squares
  - Define the Total sum of Squares as: $TSS = \sum_{i=1}^{N}(Y_i - \bar{Y})^2$
  - Define the Explained Sum of Squares as: $ESS = \sum_{i=1}^{N}[\hat{\beta}_1(X_i - \bar{X})]^2$
  - Define the Residual Sum of Squares as: $RSS = \sum_{i=1}^{N}\hat{u}_i^2$
- Then we define

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

- The is a measure of how much of the variance of Y is explained by the regressor X.
- The computed R^2 following an OLS regression is always between 0 and 1.
- A low R^2 is not necessarily an indication that the model is wrong - just that the included X has low explanatory power.
- The key to whether the results are interpretable as causal impacts is whether the explanatory variable is uncorrelated with the error term.

An Example

- We investigate the determinants of log willingness to pay as a function of log income:

$$\ln \mathrm{WTP}_i = \beta_0 + \beta_1 \ln \mathrm{income}_i + u_i$$

Variable	Coef.
log income	0.22
constant	0.42

Model sum of squares	11.7
Residual sum of squares	273.7
Total sum of squares	285.4
number of observations	352
R^2	0.041

Precision and Standard Errors

- We have shown that the OLS estimator (under our assumptions) is unbiased.
- But how sensitive are our results to random changes to our sample? The variance of the estimator is a measure of this.
- Consider first the slope coefficient. As we showed this can be decomposed into two parts: The true value and the estimation error:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

- We also showed that $E(\hat{\beta}_1|X) = \beta_1$
- The definition of the variance is

$$Var(\hat{\beta}_1|X) = E[(\hat{\beta}_1 - \beta_1)^2|X]$$

- Now note that

$$\begin{aligned} E[(\hat{\beta}_1 - \beta_1)^2|X] &= E\left[\left[\frac{\sum_{i=1}^{N}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\right]^2 |X\right] \\ &= \frac{1}{\left[\sum_{i=1}^{N}(X_i - \bar{X})^2\right]^2} E\left[\left[\sum_{i=1}^{N}(X_i - \bar{X})(u_i - \bar{u})\right]^2 |X\right] \\ &= \frac{1}{\left[\sum_{i=1}^{N}(X_i - \bar{X})^2\right]^2} \\ &\quad \cdot \left[\sum_{j=1}^{N}\sum_{i=1}^{N}(X_i - \bar{X})(X_j - \bar{X})E[(u_i - \bar{u})(u_j - \bar{u})|X]\right] \end{aligned}$$

- From Assumption 2

$$Var(u_i|X) = E[(u_i - \bar{u})^2|X] = \sigma^2 \qquad \text{(homoskedasticity)}$$

- From Assumption 3

$$E[(u_i - \bar{u})(u_j - \bar{u})|X] = 0 \qquad \text{(no autocorrelation)}$$

- Hence

$$\begin{aligned} E[(\hat{\beta}_1 - \beta_1)^2|X] &= \frac{1}{\left[\sum_{i=1}^{N}(X_i - \bar{X})^2\right]^2} \sum_{i=1}^{N}(X_i - \bar{X})^2\sigma^2 \\ &= \frac{\sigma^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{1}{N}\frac{\sigma^2}{Var(X)} \end{aligned}$$

- Properties of the variance
  - The Variance reflects the precision of the estimation or the sensitivity of our estimates to different samples.
  - The higher the variance - the lower the precision.
  - The variance increases with the variance of the error term(noise)
  - The variance decreases with the variance of X.
  - The variance decreases with the sample size.
  - The standard error is the square root of the variance:

$$s.e(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)}$$

An Example

- We investigate the determinants of log willingness to pay as a function of log income:

$$\ln \mathrm{WTP}_i = \beta_0 + \beta_1 \ln \mathrm{income}_i + u_i$$

Variable	Coef.	Std. Err.
log income	0.22	0.06
constant	0.42	0.47
number of observations	352	
R^2	0.041	

Efficiency

- An estimator is efficient if within the set of assumptions that we make, it provides the most precise estimates in the sense that the variance is the lowest possible in the class of estimators we are considering.
- How do we choose between the OLS estimator and any other unbiased estimator.
- Our criterion is efficiency.
- Among all the unbiased estimator, which one has the smallest variance?

The Gauss Markov theorem

- Given Assumptions 1-5 the Ordinary Least Squares Estimator is a Best Linear Unbiased Estimator (BLUE)

- This means that the OLS estimator is the most efficient (least variance) estimator in the class of linear unbiased estimators.

Multiple Regression Model

The Multiple Regression Model

- The Multiple regression model takes the form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + u_i$$

- There are k regressors (explanatory Variables) and a constant. Hence there will be k+1 parameters to estimate.

- **Assumption M.1:**
  We will keep the basic least squares assumption - We will assume that the *error term is mean independent of all regressors* (loosely speaking - all Xs are uncorrelated with the error term, i.e.

$$E(u_i|X_1, X_2, \ldots, X_k) = E(u_i|X) = 0$$

Interpretation of the coefficients

- Since the error term is mean independent of the Xs, varying the X's does not have an impact on the error term.
- Thus under Assumption M.1 the coefficients in the regression model have the following simple interpretation:

$$\beta_j = \frac{\partial Y_i}{\partial X_{ij}}$$

- Thus each coefficient measures the impact of the corresponding X on Y keeping all other factors (Xs and u) constant. A ceteris paribus effect.

Dummy Variables

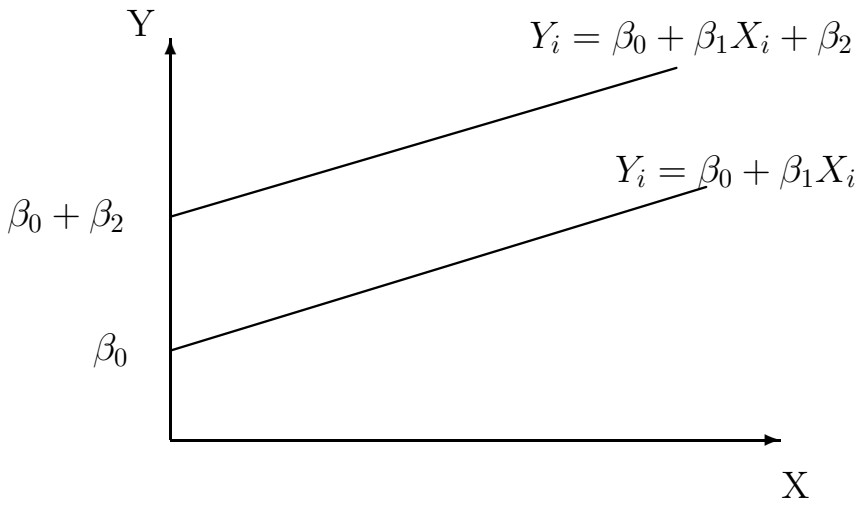
- Some of the explanatory variables are not necessarily continuous variables. Y may also be determined by qualitative factors which are not measured in any units:
  - sex, nationality or race.
  - type of education (vocational, general).
  - type of housing (flat, large house or small house).
- These characteristics are coded into **dummy variables**. These variables take only two values, 0 or 1:

$$\begin{cases} D_i = 0 & \text{if individual is male} \\ D_i = 1 & \text{if individual is female} \end{cases}$$

Dummy Variables: Intercept Specific Relationship

- The dummy variable can be used to build a model with an intercept that vary across groups coded by the dummy variable:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i$$



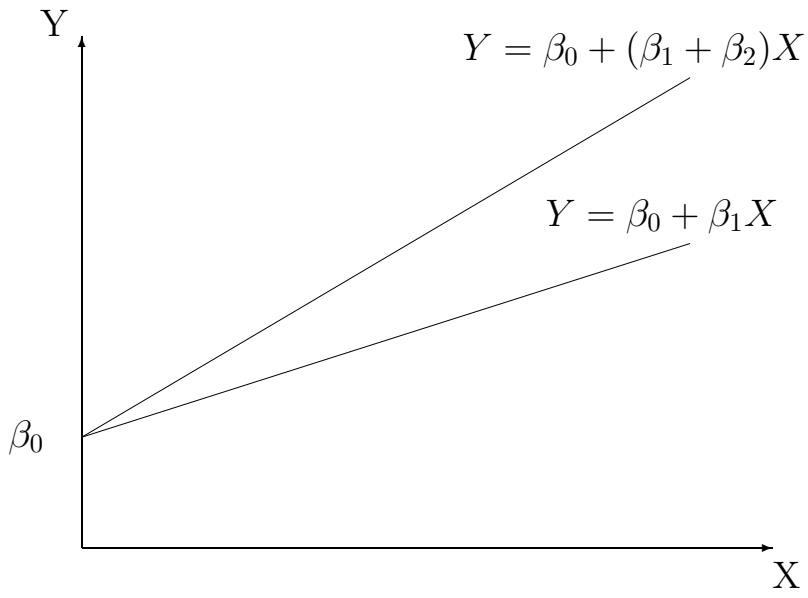
- Interpretation: The observations for which $D_i = 1$ have on average a Y_i which is β_2 units higher.
- Example: WTP, income and sex

Variable	Coefficient	st. err
log income	0.22	0.06
sex (1=Male)	0.01	0.09
constant	0.42	0.47

Dummy Variables: Slope Specific Relationship

- The dummy variable can also be interacted with a continuous variable, to get a slope specific to each group:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i D_i + u_i$$



- Interpretation: For observations with $D_i = 0$, a one unit increase in X_i leads to an increase of β_1 units in Y_i. For those with $D_i = 1$, Y_i increases by $\beta_1 + \beta_2$ units.
- Example: WTP, income and sex

Variable	Coefficient	st. err
log income	0.23	0.06
sex (1=Male)*log income	0.003	0.01
constant	0.42	0.47

Least Squares in the Multiple Regression Model

- We maintain the same set of assumptions as in the one variable regression model.
- We modify assumption 1 to assumption M1 to take into account the existence of many regressors.
- The OLS estimator is chosen to minimise the residual sum of squares exactly as before.
- Thus β_0, $\beta_1, \ldots, \beta_k$ are chosen to minimise

$$S = \sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_k X_{ik})^2$$

- Differentiating S with respect to each coefficient in turn we obtain a set of $k+1$ equations constituting the first order conditions for minimising the residual sum of squares S. These equations are called the Normal Equations.

A solution for two regressors

- With two regressors this represents a two equation system with two unknowns, i.e. β_1 and β_2.
- The solution for β_1 is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_{i2} - \bar{X}_2)X_{i2}\sum_{i=1}^{N}(X_{i2} - \bar{X}_2)X_{i1} - \sum_{i=1}^{N}(Y_i - \bar{Y})X_{i2}\sum_{i=1}^{N}(Y_i - \bar{Y})X_{i1}}{\sum_{i=1}^{N}(X_{i2} - \bar{X}_2 X_{i2})\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)X_{i1} - \sum_{i=1}^{N}(X_{i2} - \bar{X}_2)X_{i1}\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)X_{i2}}$$

- This formula can also be written as

$$\hat{\beta}_1 = \frac{cov(Y, X_1)Var(X_2) - cov(X_1, X_2)cov(Y, X_2)}{Var(X_1)Var(X_2) - cov(X_1, X_2)^2}$$

  Similarly we can derive the formula for the other coefficient (β_2)

- Note that the formula for $\hat{\beta}_1$ is now different from the formula we had in the two variable regression model. This now takes into account the presence of the other regressor(s).
- The extent to which the two formulae differ depends on the covariance of X_1 and X_2.
- When this covariance is zero we are back to the formula for the one variable regression model.

The Gauss Markov Theorem

- The Gauss Markov Theorem is valid for the multiple regression model. We need however to modify assumption A.4.
- Define the covariance matrix of the regressors X to be

$$cov(X) = \begin{bmatrix} Var(X_1) & cov(X_1, X_2) & \dots & cov(X_1, X_k) \\ cov(X_1, X_2) & Var(X_2) & \dots & cov(X_2, X_k) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_1, X_k) & cov(X_2, X_k) & \dots & Var(X_k) \end{bmatrix}$$

- **Assumption M.4:** We assume that cov(X) **positive definite** and hence can be inverted.
- **Theorem:** Under Assumptions M.1 A.2 and A3 and M.4 the Ordinary Least Squares Estimator (OLS) is Best in the class of Linear Unbiased estimators (BLUE).
- As before this means that OLS provides estimates that are least sensitive to changes in the data - given the stated assumptions.

Goodness of Fit

- The R^2 is non decreasing in the number of explanatory variables.
- To compare two different model, one would like to adjust for the number of explanatory variables: adjusted R^2:

$$\bar{R}^2 = 1 - \frac{\sum_i \hat{u}_i^2/(N-k)}{\sum_i y_i^2/(N-1)}$$

- The adjusted and non adjusted R^2 are related:

$$\bar{R}^2 = 1 - (1 - R^2)\frac{N-1}{N-k}$$

- Note that to compare two different R^2 the dependent variable must be the same:

$$\ln Y_i = \beta_0 + \beta_1 X_i + u_i$$
$$Y_i = \alpha_0 + \alpha_1 X_i + u_i$$

  cannot be compared as the Total Sum of Squares are different.

An Example

- We investigate the determinants of log willingness to pay.
- We include as explanatory variables:
  - log income,
  - education coded as low, medium and high,
  - age of the head of household, in years.
  - household size.

Variable	Coef.	Std Err.	t-stat
log income	0.14	0.07	2.2
medium education	0.47	0.16	2.9
high education	0.58	0.18	3.1
age	0.0012	0.004	0.3
household size	0.008	0.02	0.4
constant	0.53	0.55	0.96
number of observations		352	
R^2		0.0697	
adjusted R^2		0.0562	

interpretation:

- When income goes up by 1%, WTP goes up by 0.14%.
- low education is the reference group (we have omitted this dummy variable). Medium educated individuals have a WTP 47% higher than the low educated ones and high educated 58% more.

Omitted Variable Bias

- Suppose the true regression relationship has the form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$$

- Instead we decide to estimate:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \nu_i$$

- We will show that *in general* this omission will lead to a **biased** estimate of the effect of β_1.

- Suppose we use OLS on the second equation. As we know we will obtain:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)\nu_i}{\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2}$$

- The question is : What is the expected value of the last expression on the right hand side. For an unbiased estimator this will be zero. Here we will show that it is not zero.

Omitted Variable Bias

- First note that according to the true model we have that

$$\nu_i = \beta_2 X_{i2} + u_i$$

- We can substitute this into the expression for the OLS estimator to obtain

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)\beta_2 X_{i2} + \sum_{i=1}^{N}(X_{i1} - \bar{X}_1)u_i}{\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2}$$

- Now we can take expectations of this expression.

$$E[\hat{\beta}_1|X] = \beta_1 + \frac{\sum_{i=1}^{N} E[(X_{i1} - \bar{X}_1)\beta_2 X_{i2}|X] + \sum_{i=1}^{N} E[(X_{i1} - \bar{X}_1)u_i|X]}{\sum_{i=1}^{N}(X_{i1} - \bar{X}_1)^2}$$

  The last expression is zero under the assumption that u is mean independent of X [Assumption M.1].

- This expression can be written more compactly as:

$$E[\hat{\beta}_1|X] = \beta_1 + \beta_2 \frac{cov(X_1, X_2)}{Var(X_1)}$$

Omitted Variable Bias

$$E[\hat{\beta}_1|X] = \beta_1 + \beta_2 \frac{cov(X_1, X_2)}{Var(X_1)}$$

- The bias will be zero in two cases:
  - When the coefficient β_2 is zero. In this case the regressor X_2 obviously does not belong to the regression.
  - When the covariance between the two regressors X_1 and X_2 is zero.
- Thus in general omitting regressors which have an impact on Y (β_2 non-zero) will bias the OLS estimator of the coefficients on the included regressors unless the omitted regressors are uncorrelated with the included ones.

Example

- Determinants of (log) WTP: Suppose true model is:

$$\ln WTP_i = \beta_0 + \beta_1 \ln \text{income}_i + \beta_2 \text{ education}_i + u_i$$

- BUT, you omit education in regression:

$$\ln WTP_i = \alpha_0 + \alpha_1 \ln \text{income}_i + v_i$$

Variable	Coefficient	s.err
log income	0.23	0.06
constant	0.42	0.48
Extended model		
log income	0.19	0.06
education	0.18	0.12
constant	0.59	0.48

- Correlation between Education and income: 0.39.

Summary of Results

- Omitting a regressor which has an impact on the dependent variable and is correlated with the included regressors leads to "omitted variable bias"

- Including a regressor which has no impact on the dependent variable and is correlated with the included regressors leads to a reduction in the efficiency of estimation of the variables included in the regression.

Measurement Error

- Data is often measured with error.
  - reporting errors.
  - coding errors.
- The measurement error can affect either the dependent variable or the explanatory variables. The effect is dramatically different.

Measurement Error on Dependent Variable

- Y_i is measured with error. We assume that the measurement error is **additive and not correlated with** X_i.
- We observe $\check{Y}_i = Y_i + \nu_i$. We regress $\check{Y}_i$ on X_i:

$$\begin{aligned} \check{Y}_i &= \beta_0 + \beta_1 X_i + u_i \\ Y_i &= \beta_0 + \beta_1 X_i + u_i - \nu_i \\ &= \beta_0 + \beta_1 X_i + w_i \end{aligned}$$

- The assumptions we have made for OLS to be unbiased and BLUE are *not* violated. **OLS estimator is unbiased**.
- The variance of the slope coefficient is:

$$\begin{aligned} Var(\hat{\beta}_1) &= \frac{1}{N}\frac{Var(w_i)}{Var(X_i)} \\ &= \frac{1}{N}\frac{Var(u_i - \nu_i)}{Var(X_i)} \\ &= \frac{1}{N}\frac{Var(u_i) + Var(\nu_i)}{Var(X_i)} \\ &\geq \frac{1}{N}\frac{Var(u_i)}{Var(X_i)} \end{aligned}$$

- The variance of the estimator is larger with measurement error on Y_i.

Measurement Error on Explanatory Variables

- X_i is measured with errors. We assume that the error is **additive and not correlated with** X_i.
- We observe $\check{X}_i = X_i + \nu_i$ instead. The regression we perform is Y_i on $\check{X}_i$. The estimator of β_1 is expressed as:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{N}(\check{X}_i - \bar{\check{X}})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(\check{X}_i - \bar{\check{X}})^2} \\
&= \frac{\sum_{i=1}^{N}(X_i + \nu_i - \bar{X})(\beta_0 + \beta_1 X_i + u_i - \bar{Y})}{\sum_{i=1}^{N}(X_i + \nu_i - \bar{X})^2} \\
&= \frac{\sum_{i=1}^{N}\beta_1(X_i - \bar{X})^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2 + \nu_i^2 - 2\nu_i(X_i - \bar{X})}
\end{aligned}$$

$$E(\hat{\beta}_1) = \frac{\beta_1 Var(X_i)}{Var(X_i) + Var(\nu_i)} \leq \beta_1$$

- Measurement error on X_i leads to a **biased OLS estimate**, biased towards zero. This is also called **attenuation bias**.

Example

- True model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad \text{with} \qquad \beta_0 = 1 \;\; \beta_1 = 1$$

- X_i is measured with error. We observe $\tilde{X}_i = X_i + \nu_i$.
- Regression results:

	$\text{Var}(\nu_i)/\text{Var}(X_i)$			
	0	0.2	0.4	0.6
β_0	1	1.08	1.28	1.53
β_1	2	1.91	1.7	1.45

Hypotheses Testing

Hypothesis Testing

- We may wish to test prior hypotheses about the coefficients we estimate.
- We can use the estimates to test whether the data rejects our hypothesis.
- An example might be that we wish to test whether an elasticity is equal to one.
- We may wish to test the hypothesis that X has no impact on the dependent variable Y.
- We may wish to construct a confidence interval for our coefficients.

Hypothesis

- A hypothesis takes the form of a statement of the true value for a coefficient or for an expression involving the coefficient.
  - The hypothesis to be tested is called the null hypothesis.
  - The hypothesis which it is tested against is called the alternative hypothesis.
  - Rejecting the null hypothesis *does not imply accepting the alternative*.
  - We will now consider testing the simple hypothesis that the slope coefficient is equal to some fix value.

Setting up the hypothesis

- Consider the simple regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- We wish to test the hypothesis that $\beta_1 = b$ where b is some known value (for example zero) against the hypothesis that β_1 is not equal to b. We write this as follows:

$$H_0 : \qquad \beta_1 = b$$

$$H_a : \qquad \beta_1 \neq b$$

Distribution of the OLS slope coefficient

- To test the hypothesis we need to know the way that our estimator is distributed.
- We start with the simple case where we assume that the error term in the regression model is a normal random variable with mean zero and variance σ^2. This is written as:

$$u_i \sim \mathcal{N}(0, \sigma^2)$$

- Now recall that the OLS estimator can be written as:

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{N} w_i u_i$$

with

$$w_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{N} (X_i - \bar{X})^2}$$

- Thus the OLS estimator is equal to a constant (β_1) plus a weighted sum of normal random variables,
- Weighted sums of normal random variables are also normal, so the OLS coefficient is a Normal random variable.

Distribution of the OLS slope coefficient

- What is the mean and what is the variance of this random variable?
  - Since OLS is unbiased the mean is β_1.
  - We have derived the variance and shown it to be:

$$Var(\hat{\beta}_1) = \frac{1}{N}\frac{\sigma^2}{Var(X)}$$

  - This means that:

$$z = \frac{\hat{\beta}_1 - b}{\sqrt{Var(\hat{\beta}_1)}} \sim \mathcal{N}(0,1)$$

  - The difficulty with using this result is that we do not know the variance of the OLS estimator because we do not know σ^2, which needs to be estimated.

Distribution of the OLS slope coefficient

- An unbiased estimator of the variance of the residuals is the residual sum of squares divided by the number of observations minus the number of estimated parameters. This quantity (N-2) in our case is called the degrees of freedom. Thus

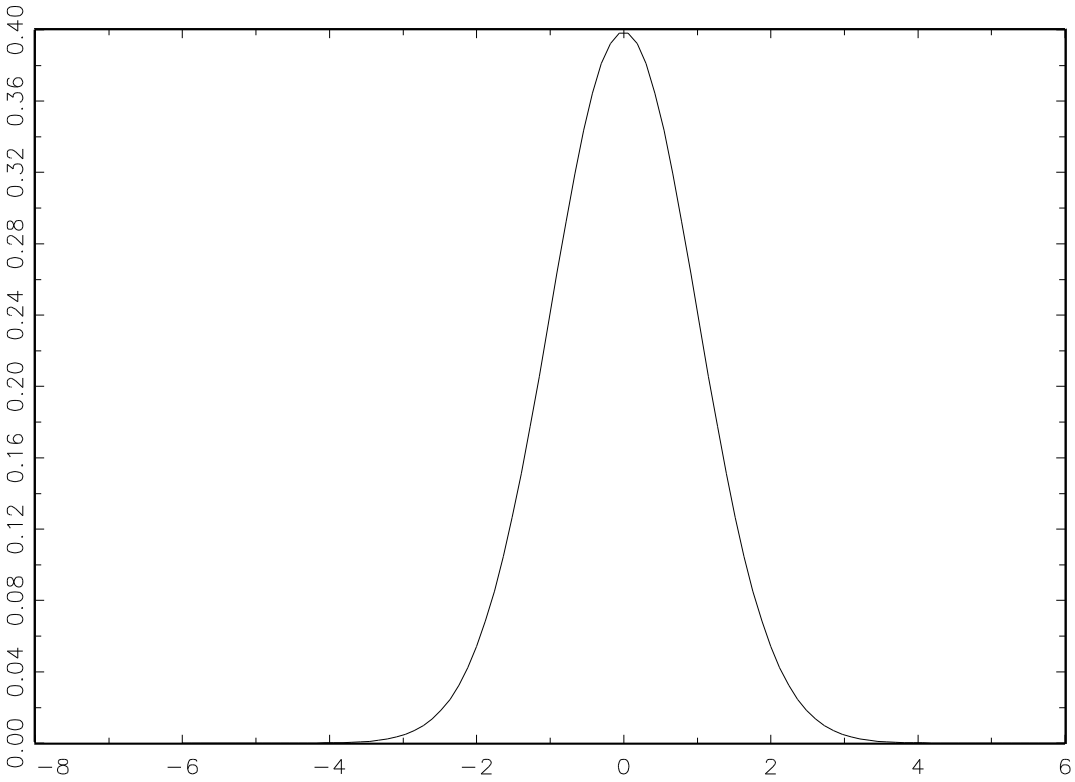
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{u}_i^2}{N-2}$$

- We now replace the variance by its estimated value to obtain a *test statistic*:

$$z^* = \frac{\hat{\beta}_1 - b}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}}}$$

- This test statistic is no longer Normally distributed, but follows the t-distribution with $N-2$ degrees of freedom.

The Student Distribution



Student Distribution with degrees of freedom=$N - 2$=1000.

- We want to accept the null if z^* is "close" to zero

$$z^* = \frac{\hat{\beta}_1 - b}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}}}$$

- How close is close?
- We need to set up an interval in which we agree that z^* is almost zero.

Testing the Hypothesis

- Thus we have that under the null hypothesis:

$$z^* = \frac{\hat{\beta}_1 - b}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}}} \sim t_{N-2}$$

- The next step is to choose the size of the test (significance level). This is the probability that we reject a correct hypothesis. The conventional size is 5%. We say that the size $\alpha = 0.05$
- We now find the *critical values* and $t_{\alpha/2,N}$ and $t_{1-\alpha/2,N}$
  - We *accept* the null hypothesis if the test statistic is between the critical values corresponding to our chosen size.
  - Otherwise we *reject*.
  - The logic of hypothesis testing is that if the null hypothesis is true then the estimate will lie within the critical values $100 * (1 - \alpha)\%$ of the time.

Percentage points of the t distribution

	$\alpha/2$					
df	0.25	0.10	0.05	0.025	0.01	0.005
1	1.000000	3.077684	6.313752	12.7062	31.8205	63.6567
2	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484
3	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091
4	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409
5	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214
6	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743
7	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948
8	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539
9	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984
10	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927
20	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534
30	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000
inf	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583

Confidence Interval

- We have argued that

$$z^* = \frac{\hat{\beta}_1 - b}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{N}(X_i - \bar{X})^2}}} \sim t_{N-2}$$

- This implies that we can construct an interval such that the chance that the true β_1 lies within that interval is some fixed value chosen by us. Call this value $1 - \alpha$.
- For a 95% confidence interval say this would be 0.95.
- From statistical tables we can find critical values such that any random variable which follows a t-distribution falls between these two values with a probability of $1 - \alpha$. Denote these critical values by $t_{\alpha/2,N}$ and $t_{1-\alpha/2,N}$.
- For a t random variable with 10 degrees of freedom and a 95% confidence these values are (2.228,-2.228).
- Thus

$$P(t_{\alpha/2,N} < z^* < t_{1-\alpha/2,N}) = 1 - \alpha$$

- With some manipulation we then get that

$$P\left(\hat{\beta}_1 - s.e.(\hat{\beta}_1) \times t_{\alpha/2,N} < \beta_1 < \hat{\beta}_1 + s.e.(\hat{\beta}_1) \times t_{1-\alpha/2,N}\right) = 1-\alpha$$

- The term in the brackets is the confidence interval.

Example: Confidence Interval

- Log WTP and income.

Variable	Coeff	st.err
β_1	0.23	0.06
β_0	0.42	0.47

- We have 352 observations, so 350 degrees of freedom. At 95% confidence level, $t_{0.05/2,350} = 1.96$

$$P(0.23 - 0.06 * 1.96 < \beta_1 < 0.23 + 0.06 * 1.96) = 0.95$$
$$P(0.11 < \beta_1 < 0.35) = 0.95$$

- The true value has 95% chances of being in [0.11,0.35].

- H_0: $\beta_1 = 0$, H_a: $\beta_1 \neq 0$
  - $z^* = (0.23 - 0)/0.06 = 0.23/0.06 = 3.9$
  - The critical value is again 1.96, at 5%.
  - So z^* is bigger than 1.96, so we reject H_0.

More on Testing

- Do we need the assumption of normality of the error term to carry out inference (hypothesis testing)?

- Under normality our test is exact. This means that the test statistic has exactly a t distribution.

- We can carry out tests based on asymptotic approximations when we have large enough samples.

- To do this we will use Central limit theorem results that state that in large samples weighted averages are distributed as normal variables.

Hypothesis Testing in the Multiple regression model

- Testing that individual coefficients take a specific value such as zero or some other value is done in exactly the same way as with the simple two variable regression model.
- Now suppose we wish to test that a number of coefficients or combinations of coefficients take some particular value.
- In this case we will use the so called "F-test".
- Suppose for example we estimate a model of the form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + u_i$$

- We may wish to test hypotheses of the form:
  - $\{H_0$: $\beta_1 = 0$ and $\beta_2 = 0$ against the alternative that one or more are wrong$\}$.
  - or $\{H_0$: $\beta_1 = 1$ and $\beta_2 - \beta_3 = 0$ against the alternative that one or more are wrong$\}$
  - or $\{H_0$: $\beta_1 + \beta_2 = 1$ and $\beta_0 = 0$ against the alternative that one or more are wrong$\}$.

Definitions

- **The Unrestricted Model:** This is the model without any of the restrictions imposed. It contains all the variables exactly as in the regression of the previous page.
- **The Restricted Model:** This is the model on which the restrictions have been imposed. For example all regressors whose coefficients have been set to zero are excluded and any other restriction has been imposed.
- Example 1: Testing H_0: $\beta_1 = 0$ and $\beta_0 = 0$

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i & \text{unrestricted model} \\ Y_i = \beta_2 X_{i2} + \beta_3 X_{i3} + u_i & \text{restricted model} \end{cases}$$

- Example 2: Testing H_0: $\beta_1 - \beta_2 = 1$ and $\beta_3 = 2$

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + u_i & \text{unrestricted model} \\ Y_i = \beta_0 + \beta_1 X_{i1} + (1 + \beta_1) X_{i2} + 2X_{i3} + u_i & \text{restricted model} \end{cases}$$

and rearranging the restricted model gives:

$$(Y_i - X_{i2} - 2X_{i3}) = \beta_0 + \beta_1 (X_{i1} + X_{i2}) + u_i$$

Intuition of the Test

- Inference will be based on comparing the fit of the restricted and unrestricted regression.
- The unrestricted regression will always fit *at least as well as the restricted one*. The proof is simple: When estimating the model we minimise the residual sum of squares. In the unrestricted model we can always choose the combination of coefficients that the restricted model chooses. Hence the restricted model can never do better than the unrestricted one.
- So the question will be how much improvement in the fit do we get by relaxing the restrictions relative to the loss of precision that follows. The distribution of the test statistic will give us a measure of this so that we can construct a decision rule.

Further Definitions

- Define the Unrestricted Residual Residual Sum of Squares (URSS) as the residual sum of squares obtained from estimating the unrestricted model.
- Define the Restricted Residual Residual Sum of Squares (RRSS) as the residual sum of squares obtained from estimating the restricted model.
- Note that according to our argument above $RRSS \geq URSS$.
- Define the degrees of freedom as $N - k$ where N is the sample size and k is the number of parameters estimated in the unrestricted model (i.e under the alternative hypothesis) (which includes the constant if any).
- Define by q the number of restrictions imposed (in both our examples there were two restrictions imposed.

The F-Statistic

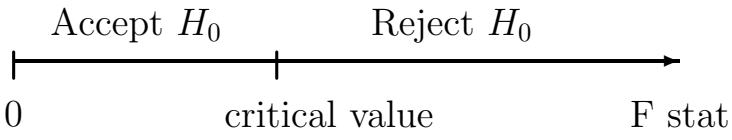
- The Statistic for testing the hypothesis we discussed is

$$F = \frac{(RRSS - URSS)/q}{URSS/(N-k)}$$

$$F = \frac{(R^2 - \tilde{R}^2)/q}{(1-R^2)/(n-k)}$$

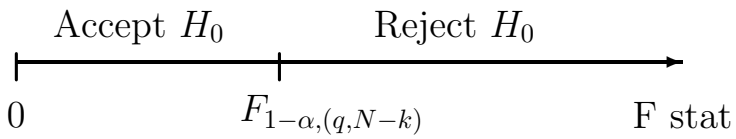
- The test statistic is always positive. We would like this to be "small". The smaller the F-statistic the less the loss of fit due to the restrictions

- Defining "small" and using the statistic for inference we need to know its distribution.



The Distribution of the F-statistic

- As in our earlier discussion of inference we distinguish two cases:
  - **Normally Distributed Errors:** The errors in the regression equation are distributed normally. In this case we can show that under the null hypothesis H_0 the F-statistic is distributed as an **F distribution** with degrees of freedom $(q, N-k)$.
    - The number of restrictions q are the degrees of freedom of the numerator.
    - $N-k$ are the degrees of freedom of the denominator.
    - Since the smaller the test statistic the better and since the test statistic is always positive we only have one critical value. For a test at the level of significance α we choose a critical value of $F_{1-\alpha,(q,N-k)}$.



- When the regression errors are not normal (but satisfy all the other assumptions we have made) we can appeal to the central limit theorem to justify inference.
- In large samples we can show that q times the F statistic is distributed as a random variable with a chi-square distribution:

$$qF \sim \chi^2_{1-\alpha,q}$$

Examples

- Examples of Critical values for 5% tests in a regression model with 6 regressors under the alternative
  - Sample size 18. One restriction to be tested: Degrees of freedom 1, 12: $F_{1-0.05,(1,12)} = 4.75$
  - Sample size 24. Two restrictions to be tested: degrees of freedom 2, 18: $F_{1-0.05,(2,18)} = 3.55$
  - Sample size 21. Three restrictions to be tested: degrees of freedom 3, 15: $F_{1-0.05,(3,15)} = 3.29$

- Examples of Critical values for 5% tests in a regression model with 6 regressors under the alternative. Inference based on large samples:
  - One restriction to be tested: Degrees of freedom 1:

$$\chi^2_{1-\alpha,1} = 3.84$$

  - Two restrictions to be tested: degrees of freedom 2:

$$\chi^2_{1-\alpha,2} = 5.99$$

Summary

- OLS in simple and multiple linear regression models.
- Key assumptions:
  1. The error term is uncorrelated with explanatory variables.
  2. variance of error term is constant (homoskedasticity).
  3. covariance of error term is zero (no autocorrelation).
- Consequences: unbiased coefficients. BLUE. Testing hypothesis.
- Departures from this simple framework:
  - heteroskedasticity.
  - autocorrelation.
  - simultaneity and endogeneity.
  - non linear models.

Heteroskedasticity

Definition

- **Definition:** The variance of the residual is not constant across observations:

$$Var(u_i) = \sigma_i^2$$

- In particular the variance of the errors may be a function of explanatory variables:

$$Var(u_i) = \sigma(X_i)^2$$

- Example: Think of food expenditure for example. It may well be that the "diversity of taste" for food is greater for wealthier people than for poor people. So you may find a greater variance of expenditures at high income levels than at low income levels.

Implications of Heteroskedasticity

- Assuming all other assumptions are in place, the assumption guaranteeing unbiasedness of OLS is *not* violated. Consequently **OLS is unbiased** in this model

- However the assumptions required to prove that OLS is efficient are violated. Hence **OLS is not BLUE** in this context

- The formula for the variance of the OLS estimator is no longer valid.

$$Var(\hat{\beta}_1) \neq \frac{1}{N}\frac{\sigma^2}{Var(X)}$$

  Hence we cannot make any inference using the computed standard errors.

- We can devise an efficient estimator by re-weighting the data appropriately to take into account of heteroskedasticity

Testing for Heteroskedasticity

- **Visual inspection of the data**. Graph the residuals $\hat{u}_i$ as a function of explanatory variables. Is there a constant spread across all values of X?

- **White Test**: Extremely general, low power

$$\begin{array}{ll} H_0: & \sigma_i^2 = \sigma^2 \\ H_1: & \text{not} \quad H_0 \end{array}$$

1. Get the residuals from an OLS regression $\hat{u}_i$.
2. Regress $\hat{u}_i^2$ on a constant and $X \otimes X$. (Note $\otimes$ denotes the cross-product of all terms in X. For instance if $X = [X_1, X_2]$ then $X \otimes X = X_1^2 + X_2^2 + X_1 X_2$).
3. Get the R^2 and compute $T.R^2$ which follows a $\chi^2(p-1)$. p is the number of regressors in the auxiliary regression, including the constant.
4. Reject homoskedasticity if $T.R^2 > \chi^2_{1-\alpha}(p-1)$

Testing for Heteroskedasticity

- **Goldfeld-Quandt Test**

  1. Rank observation based on X_j.
  2. Separate in two groups. Low X_j, N_1 values. High X_j, N_2 values. Typically 1/3 and 3/3 observations.
  3. Do the regression on the separate groups. Compute the residuals, $\hat{u}_{1i}$ and $\hat{u}_{2i}$.
  4. Compute

$$f = \frac{\sum_{i=1}^{N_1} \hat{u}_{1i}^2/(N_1 - k)}{\sum_{i=1}^{N_2} \hat{u}_{2i}^2/(N_2 - k)} \qquad \text{or} \qquad f = \frac{\sum_{i=1}^{N_2} \hat{u}_{2i}^2/(N_2 - k)}{\sum_{i=1}^{N_1} \hat{u}_{1i}^2/(N_1 - k)}$$

     whatever is larger than 1. $f \sim F(N_1 - k, N_2 - k)$
  5. Reject homoskedasticity if $f > F(N_1 - k, N_2 - k)$

- **Breusch-Pagan Test:** test if heteroskedasticity is of the form

$$\sigma_i^2 = \sigma^2 F(\alpha_0 + \alpha' Z_i)$$

  1. Compute the OLS regression, get the residuals $\hat{u}_i$.
  2. Compute

$$g_i = \frac{\hat{u}_i^2}{\sum_{i=1}^{N} \hat{u}_i^2/N}$$

  3. regress g_i on a constant and the Z_i.

$$g_i = \gamma_0 + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + \ldots + v_i$$

  4. Compute the Expected Sum of Square (ESS). 0.5*ESS follows a $\chi^2(p)$, where p is the number of variables in Z not including the constant.
  5. Reject homoskedasticity if $0.5 * ESS > \chi^2_{1-\alpha}(p)$.

Generalized Least Squares

- Original model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad Var(u_i) = \sigma_i^2$$

- Divide each term of the equation by σ_i:

$$\begin{aligned} Y_i/\sigma_i &= \beta_0/\sigma_i + \beta_1 X_i/\sigma_i + u_i/\sigma_i \\ \tilde{Y}_i &= \tilde{\beta}_0 + \beta_1 \tilde{X}_i + \tilde{u}_i \end{aligned}$$

  Here $Var(\tilde{u}_i) = 1$.

- Perform an OLS regression of $\tilde{Y}_i$ on $\tilde{X}_i$:

$$\hat{\beta}_{1,GLS} = \frac{\sum_{i=1}^{N} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sigma_i}}{\sum_{i=1}^{N} \frac{(X_i - \bar{X})^2}{\sigma_i}}$$

  The observations are weighted by the inverse of their standard deviation. Observations with a large variance will not contribute much to the determination of $\hat{\beta}_{1,GLS}$.

Properties of GLS

- The GLS estimator is **unbiased**.

- The GLS estimator is the Best Linear Unbiased Estimator (BLUE). In particular, $V(\hat{\beta}_{1,GLS}) \leq V(\hat{\beta}_{1,OLS})$

Feasible GLS

- The only problem is that we do not know σ_i.
- Iterative procedure to compute an estimate: FGLS
  1. Perform an OLS regression on the model:
  $$Y_i = \beta_0 + \beta_1 X_i + u_i$$
  2. Compute the residuals $\hat{u}_i$
  3. Model the square of the residual as a function of the observables, for instance:
  $$\sigma_i^2 = \gamma_0 + \gamma_1 X_i$$
  Estimate γ_0 and γ_1 by an OLS regression:
  $$\hat{u}_i^2 = \gamma_0 + \gamma_1 X_i + v_i$$
  4. Construct $\hat{\sigma}_i^2 = \hat{\gamma}_0 + \hat{\gamma}_1 X_i$ and use it in the GLS formula.

$$\hat{\beta}_{1,FGLS} = \frac{\sum_{i=1}^{N} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\hat{\sigma}_i}}{\sum_{i=1}^{N} \frac{(X_i - \bar{X})^2}{\hat{\sigma}_i}}$$

Robust Standard Errors

- Under heteroskedasticity, the OLS formula for $Var(\hat{\beta}_1)$ is wrong.
- Compute a more correct formula:
  - White (1980)

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^{N} u_i^2 (X_i - \bar{X})^2}{\left(\sum_{i=1}^{N} (X_i - \bar{X})^2 \right)^2}$$

  - Newey-West (1987)

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^{N} u_i^2 (X_i - \bar{X})^2 + \sum_{l=1}^{L} \sum_{i=l+1}^{N} w_l u_i u_{i-l} (X_i - \bar{X})(X_{i-l} - \bar{X})}{\left(\sum_{i=1}^{N} (X_i - \bar{X})^2 \right)^2}$$

    with

$$w_l = \frac{l}{L+1}$$

Autocorrelation

Definition

- **Definition:** The error terms are correlated with each other:

$$cov(u_i, u_j) \neq 0 \qquad i \neq j$$

- With **time series**, the error term at one date can be correlated with the error term the period before:
  - autoregressive process:

$$\begin{array}{ll} \text{order 1 (AR(1)):} & u_i = \rho u_{i-1} + v_t \\ \text{order 2 (AR(2)):} & u_i = \rho_1 u_{i-1} + \rho_2 u_{i-2} + v_t \\ \text{order k (AR(k)):} & u_i = \rho_1 u_{i-1} + \ldots + \rho_k u_{i-k} + v_t \end{array}$$

  - moving average process:

$$\begin{array}{ll} \text{MA(1):} & u_i = v_i + \lambda v_{i-1} \\ \text{MA(2):} & u_i = v_i + \lambda_1 v_{i-1} + \lambda_2 v_{i-2} \\ \text{MA(k):} & u_i = v_i + \lambda_1 v_{i-1} + \ldots + \lambda_k v_{i-k} \end{array}$$

- With **cross-section data**: geographical distance, neighborhood effects...

Implications of Autocorrelation

- Assuming all other assumptions are in place, the assumption guaranteeing unbiasedness of OLS is *not* violated. Consequently **OLS is unbiased** in this model

- However the assumptions required to prove that OLS is efficient are violated. Hence **OLS is not BLUE** in this context

- The formula for the variance of the OLS estimator is no longer valid.

$$Var(\hat{\beta}_1) \neq \frac{1}{N}\frac{\sigma^2}{Var(X)}$$

  Hence we cannot make any inference using the computed standard errors.

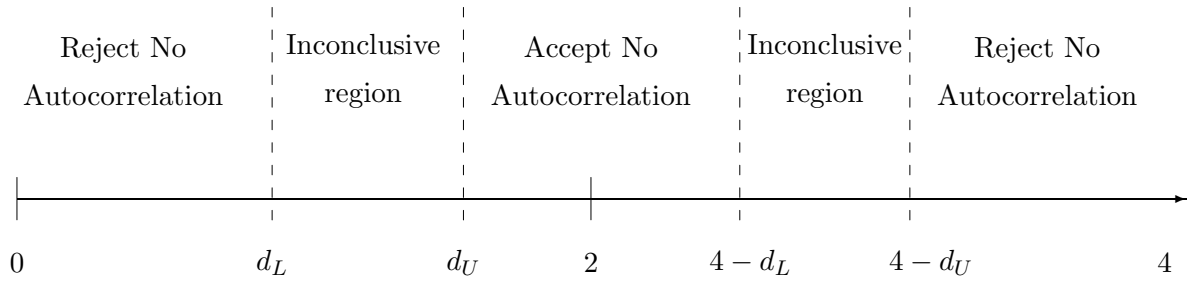
- We can devise an efficient estimator by re-weighting the data appropriately to take into account of autocorrelation.

Testing for Autocorrelation

- **Durbin Watson-Test:** Test for a first order autocorrelation in the residuals. The test relies on several important assumptions:
  - Regression includes a constant.
  - First order autocorrelation for u_i.
  - Regression does not include a lagged dependent variable.
- The test is based on the test statistic:

$$d = \frac{\sum_{i=2}^{N}(u_i - u_{i-1})^2}{\sum_{i=1}^{N} u_i^2} = 2(1-r) - \frac{u_1^2 + u_N^2}{\sum_{i=1}^{N} u_i^2} \qquad \text{with } r = \frac{\sum_{i=2}^{N} u_i u_{i-1}}{\sum_{i=1}^{N} u_i^2}$$
$$\simeq 2(1-r)$$

  Note: that if $|\rho| \leq 1$, $d \in [0, 4]$.
- The test works as following:



- The critical values d_L and d_U depend on the number of observation N.

Testing for Autocorrelation

- **Breusch-Godfrey test:** This test is more general and test for no autocorrelation against an autocorrelation of the form AR(k):

$$u_i = \rho_1 u_{i-1} + \rho_2 u_{i-2} + \cdots + \rho_k u_{i-k} + v_i$$

H_0: $\rho_1 = \cdots = \rho_p = 0$

1. First perform an OLS regression of Y_i on X_i. Get the residuals $\hat{u}_i$.
2. Regress u_i on X_i, $u_{i-1}, \cdots, u_{i-k}$
3. $(N-k).R^2 \sim \chi^2(k)$. Reject H_0 (accept autocorrelation) if $(N-k).R^2$ is larger than the critical value $\chi^2_{1-\alpha}(k)$.

Note: This test works even if no constant or lagged dependent variable.

Estimation under Autocorrelation

- Consider the following model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad u_i = \rho u_{i-1} + v_i$$

- Rewrite $Y_i - \rho Y_{i-1}$:

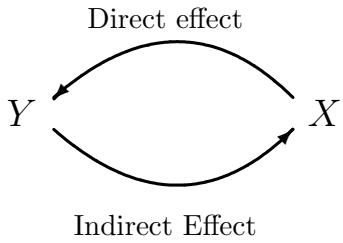
$$Y_i - \rho Y_{i-1} = \beta_0(1-\rho) + \beta_1(X_i - \rho X_{i-1}) + v_i$$

- So if we know ρ, we can be back on familiar grounds.
- If ρ is unknown, then we can do it iteratively:
  1. Estimate the model by OLS as it is. Get $\hat{u}_i$.
  2. Regress $\hat{u}_i = \rho \hat{u}_{i-1} + v_i$, to get $\hat{\rho}$
  3. Transform the model using $\hat{\rho}$ and do OLS.

Simultaneous Equations and Endogeneity

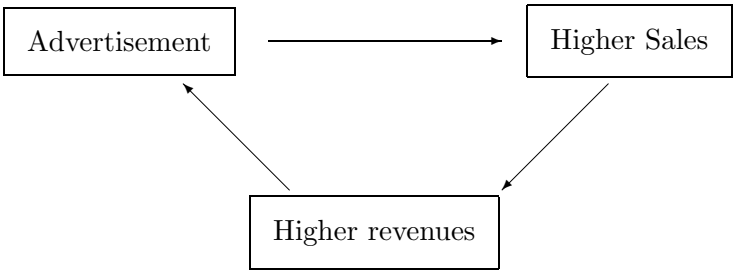
Simultaneity

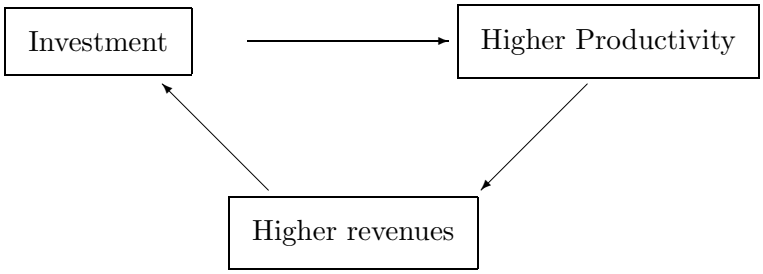
- **Definition:** Simultaneity arises when the causal relationship between Y and X runs both ways. In other words, the explanatory variable X is a function of the dependent variable Y, which in turn is a function of X.

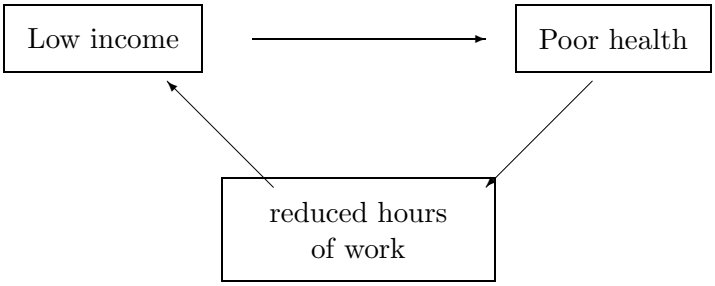


- This arises in many economic examples:
  - Income and health.
  - Sales and advertizing.
  - Investment and productivity.
- What are we estimating when we run an OLS regression of Y on X? Is it the direct effect, the indirect effect or a mixture of both.

Examples


Advertisement
Higher Sales
Higher revenues


Investment
Higher Productivity
Higher revenues


Low income
Poor health
reduced hours
of work

Implications of Simultaneity

- $$\begin{cases} Y_i = \beta_0 + \beta_1 X_i + u_i & \text{(direct effect)} \\ \\ X_i = \alpha_0 + \alpha_1 Y_i + v_i & \text{(indirect effect)} \end{cases}$$

- Replacing the second equation in the first one, we get an equation expressing Y_i as a function of the parameters and the error terms u_i and v_i only. Substituting this into the second equation, we get X_i also as a function of the parameters and the error terms:

$$\begin{cases} Y_i = \dfrac{\beta_0 + \beta_1\alpha_0}{1 - \alpha 1\beta_1} + \dfrac{\beta_1 v_i + u_i}{1 - \alpha_1\beta_1} = B_0 + \tilde{u}_i \\ \\ X_i = \dfrac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1} + \dfrac{v_i + \alpha_1 u_i}{1 - \alpha_1\beta_1} = A_0 + \tilde{v}_i \end{cases}$$

- This is the **reduced form** of our model. In this rewritten model, Y_i is not a function of X_i and vice versa. However, Y_i and X_i are both a function of the two original error terms u_i and v_i.

- Now that we have an expression for X_i, we can compute:

$$\begin{aligned} cov(X_i, u_i) &= cov\left(\frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1} + \frac{v_i + \alpha_1 u_i}{1 - \alpha_1\beta_1}, u_i\right) \\ &= \frac{\alpha_1}{1 - \alpha_1\beta_1}\ Var(u_i) \end{aligned}$$

  which, in general is different from zero. Hence, with simultaneity, our assumption 1 is violated. **An OLS regression of Y_i on X_i will lead to a biased estimate of β_1**. Similarly, an OLS regression of X_i on Y_i will lead to a biased estimate of α_1.

What are we estimating?

- For the model:

$$Y_i = \beta_0 + \beta_1 + X_i + u_i$$

- The OLS estimate is:

$$\begin{aligned} \hat{\beta}_1 &= \beta_1 + \frac{cov(X_i, u_i)}{Var(X_i)} \\ &= \beta_1 + \frac{\alpha_1}{1 - \alpha_1\beta_1}\frac{Var(u_i)}{Var(X_i)} \end{aligned}$$

- So
  - $E\hat{\beta}_1 \neq \beta_1$
  - $E\hat{\beta}_1 \neq \alpha_1$
  - $E\hat{\beta}_1 \neq$ an average of β_1 and α_1.

Identification

- Suppose a more general model:

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_i + \beta_2 T_i + u_i \\ X_i = \alpha_0 + \alpha_1 Y_i + \alpha_2 Z_i + v_i \end{cases}$$

- We have two sorts of variables:
  - **Endogenous:** Y_i and X_i because they are determined within the system. They appear on the right and left hand side.
  - **Exogenous:** T_i and Z_i. They are determined outside of our model, and in particular are not caused by either X_i or Y_i. They appear only on the right-hand-side.

Identification

- The reduced form model can be found by substituting X_i into the first equation, and then finding an expression of Y_i and X_i only as a function of the parameters, the error terms and the exogenous variables.

$$\begin{cases} Y_i = \dfrac{\beta_0 + \beta_1\alpha_0}{1-\alpha_1\beta_1} + \dfrac{\beta_1\alpha_2}{1-\alpha 1\beta_1} Z_i + \dfrac{\beta_2}{1-\alpha_1\beta_1} T_i + \tilde{u}_i \\ \\ X_i = \dfrac{\alpha_0 + \alpha_1\beta_0}{1-\alpha_1\beta_1} + \dfrac{\alpha_2}{1-\alpha_1\beta_1} Z_i + \dfrac{\alpha_1\beta_2}{1-\alpha_1\beta_1} T_i + \tilde{v}_i \end{cases}$$

$$\begin{cases} Y_i = B_0 + B_1 Z_i + B_2 T_i + \tilde{u}_i \\ \\ X_i = A_0 + A_1 Z_i + A_2 T_i + \tilde{v}_i \end{cases}$$

- We can estimate both equations of the reduced form by OLS and get consistent estimates of the reduced form parameters B_0, B_1, B_2, A_0, A_1 and A_2.

- Note that:

$$\begin{array}{ll} \dfrac{B_1}{A_1} = \beta_1 & B_2(1 - \dfrac{B_1 A_2}{A_1 B_2}) = \beta_2 \\ \dfrac{A_2}{B_2} = \alpha_1 & A_1(1 - \dfrac{B_1 A_2}{A_1 B_2}) = \alpha_2 \end{array}$$

  Similarly, one can find expressions for α_0 and β_0.

- Hence, from the reduced form coefficients, we can back out a consistent estimate of the **structural parameters**. We say that in this case they are **identified**.

Rule for Identification

- **Definition:**
  - M: Number of endogenous variables in the model
  - K: Number of predetermined variables in the model
  - m: Number of endogenous variables in a given equation
  - k: Number of predetermined variables in a given equation
- **Order Condition:**(Necessary but not sufficient) In order to have identification in a given equation, one must have

$$K - k \geq m - 1$$

  - Example 1: M=2, K=0:

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_i + u_i & m = 2, k = 0 \text{ not identified} \\ X_i = \alpha_0 + \alpha_1 Y_i + v_i & m = 2, k = 0 \text{ not identified} \end{cases}$$

  - Example 2: M=2, K=1:

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_i + \beta_2 T_i + u_i & m = 2, k = 1 \text{ not identified} \\ X_i = \alpha_0 + \alpha_1 Y_i + v_i & m = 2, k = 0 \ \alpha_1 \text{ identified} \end{cases}$$

  - Example 3: M=2, K=1:

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_i + u_i & m = 2, k = 0 \ \beta_1 \text{ identified} \\ X_i = \alpha_0 + \alpha_1 Y_i + \alpha_2 Z_i + v_i & m = 2, k = 1 \text{ not identified} \end{cases}$$

  - Example 4: M=2, K=2:

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_i + \beta_2 T_i + u_i & m = 2, k = 1 \ \beta_1 \text{ identified} \\ X_i = \alpha_0 + \alpha_1 Y_i + \alpha_2 Z_i + v_i & m = 2, k = 1 \ \alpha_1 \text{ identified} \end{cases}$$

Toward Instrumental Variables

- Consider the following system of equations:

$$\begin{cases} Y_i = \beta_0 + \beta_1 X_i + u_i \\ X_i = \alpha_0 + \alpha_1 Y_i + \alpha_2 Z_i + v_i \end{cases}$$

We are interested in a consistent estimation of β_1. Given the simultaneity, an OLS regression of Y_i on X_i leads to a biased estimate. Applying the identification rule, we know that β_1 can be identified, but not α_1 and α_2. The reduced form is:

$$\begin{cases} Y_i = \dfrac{\beta_0 + \beta_1\alpha_0}{1 - \alpha_1\beta_1} + \dfrac{\beta_1\alpha_2}{1 - \alpha 1\beta_1} Z_i + \tilde{u}_i = B_0 + B_1 Z_i + \tilde{u}_i \\ \\ X_i = \dfrac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1} + \dfrac{\alpha_2}{1 - \alpha_1\beta_1} Z_i + \tilde{v}_i = A_0 + A_1 Z_i + \tilde{v}_i \end{cases}$$

- We can recover $\hat{\beta}_1 = \hat{B}_1/\hat{A}_1$, where $\hat{A}_1$ and $\hat{B}_1$ are obtained by the formula of the OLS regression:

$$\hat{B}_1 = \frac{\sum_{i=1}^{N}(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(Z_i - \bar{Z})^2} \quad \hat{A}_1 = \frac{\sum_{i=1}^{N}(Z_i - \bar{Z})(X_i - \bar{X})}{\sum_{i=1}^{N}(Z_i - \bar{Z})^2}$$

- So an estimator of β_1 is:

$$\hat{\beta}_{1,IV} = \frac{\sum_{i=1}^{N}(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(Z_i - \bar{Z})(X_i - \bar{X})} = \frac{cov(Z_i, Y_i)}{cov(Z_i, X_i)}$$

This is the instrumental variable (IV) estimator, which can be obtain in just one step, without deriving the reduced form model and backing out $\hat{\beta}_1$.

Instrumental Variables

- **Definition:** An instrument for the model $Y_i = \beta_0 + \beta_1 X_i + u_i$ is a variable Z_i which is correlated with X_i but uncorrelated with u_i:

  1. $cov(Z_i, u_i) = 0$
  2. $cov(Z_i, X_i) \neq 0$

- The IV procedure can be seen as a two step estimator within a simultaneous system as seen in previous slide.

- Another way of defining it is from the definition above:

$$\begin{aligned} cov(Z_i, u_i) &= 0 \\ cov(Z_i, Y_i - \beta_0 - \beta_1 X_i) &= 0 \\ cov(Z_i, Y_i) - \beta_1 cov(Z_i, X_i) &= 0 \end{aligned}$$

  Hence

$$\hat{\beta}_{1,IV} = \frac{cov(Z_i, Y_i)}{cov(Z_i, X_i)}$$

Properties of IV

- Under the assumptions listed above, the instrumental variable estimator is **unbiased**:

$$\hat{\beta}_{1,IV} = \beta_1 + \frac{\sum_{i=1}^{N}(Z_i - \bar{Z})u_i}{\sum_{i=1}^{N}(Z_i - \bar{Z})(X_i - \bar{X})}$$

$$E[\hat{\beta}_1|X] = \beta_1 + \frac{\sum_{i=1}^{N} E[(Z_i - \bar{Z})u_i|X]}{\sum_{i=1}^{N}(Z_i - \bar{Z})(X_i - \bar{X})} = \beta_1$$

- The variance of the IV estimator is:

$$Var(\hat{\beta}_1) = \sigma^2 \frac{\sum_{i=1}^{N}(Z_i - \bar{Z})^2}{[\sum_{i=1}^{N}(Z_i - \bar{Z})(X_i - \bar{X})]^2} = \sigma^2 \frac{Var(Z_i)}{cov(Z_i, X_i)^2}$$

  The variance is lower, the lower the variance of Z_i or the higher the covariance between Z_i and X_i.

Examples

- IV is used in a number of cases where the explanatory variable is correlated with the error term (**endogeneity**):
  - measurement error on X.
  - simultaneity.
  - lagged dependent variable and autocorrelation.

Examples: Measurement Errors

- Suppose we are measuring the impact of income, X, on consumption, Y. The true model is:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$\beta_0 = 0, \quad \beta_1 = 1$$

- Suppose we have two measures of income, both with measurement errors.
  - $\check{X}_{1i} = X_i + v_{1i}, \quad s.d.(v_{1i}) = 0.2 * \bar{Y}$
  - $\check{X}_{2i} = X_i + v_{2i}, \quad s.d.(v_{2i}) = 0.4 * \bar{Y}$

  If we use $\check{X}_2$ to instrument $\check{X}_1$, we get:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} (\check{X}_{2i} - \bar{\check{X}}_2)(Y_i - \bar{Y})}{\sum_{i=1}^{N} (\check{X}_{2i} - \bar{\check{X}}_2)(\check{X}_{1i} - \bar{\check{X}}_1)}$$

- Results:

Method	Estimate of β_1
OLS regressing Y on $\check{X}_1$	0.88
OLS regressing Y on $\check{X}_2$	0.68
IV, using $\check{X}_2$ as instrument	0.99

Example: Lagged dependent variable

- Consider the time-series model:

$$Y_i = \beta_0 + \beta_1 Y_{i-1} + \beta_2 X_i + u_i$$

with

$$u_i = v_i + \lambda v_{i-1}$$

where v_i is a i.i.d. shock, and $\text{cov}(X_i, u_i) = 0$.

- In this model:

$$\begin{aligned} cov(Y_{i-1}, u_i) &= cov(\beta_0 + \beta_1 Y_{i-2} + \beta_2 X_{i-1} + u_{i-1}, u_i) \\ &= cov(\beta_0 + \beta_1 Y_{i-2} + \beta_2 X_{i-1} + v_{i-1} + \lambda v_{i-2}, v_i + \lambda v_{i-1}) \\ &= \lambda V(v_{i-1}) \\ &\neq 0 \end{aligned}$$

- A valid instrument is X_{i-1}, as X_{i-1} is correlated with Y_{i-1} and not with u_i.

- The IV estimator is:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_{i-1} - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_{i-1} - \bar{X})(X_i - \bar{X})}$$

More than one Instrument

- The previous slides showed how to use a variable as an instrument. Sometimes, more than one variable can be thought of as an instrument.
- Suppose Z_{1i} and Z_{2i} are two possible instruments for a variable X_i:

$$\begin{cases} cov(Z_{1i}, u_i) = 0 & \hat{\beta}_1 = \dfrac{cov(Z_{1i}, Y_i)}{cov(Z_{1i}, X_i)} \\ \\ cov(Z_{2i}, u_i) = 0 & \hat{\hat{\beta}}_1 = \dfrac{cov(Z_{2i}, Y_i)}{cov(Z_{2i}, X_i)} \end{cases}$$

- How can we combine the two instruments to use the information efficiently?

Intuition of 2SLS

- We can use a linear combination of both instruments:

$$Z_i = \alpha_1 Z_{1i} + \alpha_2 Z_{2i}$$

- The new variable Z_i is still a valid instrument as

$$cov(Z_i, u_i) = cov(\alpha_1 Z_{1i} + \alpha_2 Z_{2i}, u_i) = 0$$

  whatever the weights α_1 and α_2.

- It is up to us to chose the weights so that the covariance between Z_i and X_i is maximal.

- To obtain the best predictor of X_i, a natural way is to run the regression:

$$X_i = \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + w_i$$

  as the OLS maximizes the R^2.

- Once we have obtained $Z_i^* = \hat{\alpha}_1 Z_{1i} + \hat{\alpha}_2 Z_{2i}$, we are back to the case with only one instrumental variable.

$$\hat{\beta}_{1,2SLS} = \frac{\sum_{i=1}^{N}(Z_i^* - \bar{Z}^*)(Y_i - \bar{Y})}{\sum_{i=1}^{N}(Z_i^* - \bar{Z}^*)(X_i - \bar{X})}$$

  This entire procedure is called **two stage least squares** (2SLS).

Exogeneity Test

- Hausman's exogeneity test:

$$H_0 : \text{no endogeneity}$$
$$H_a : \text{endogeneity}$$

- Idea 1: under the null, $\hat{\beta}_{1,OLS} = \hat{\beta}_{1,2SLS}$. Compare both estimate. One can set up a chi-square test.
- Idea 2: under the null, X_i is not correlated with u_i.
- Practical implementation:
  1. First regress X_i on Z_i and get the residual $\hat{v}_i$.
$$X_i = \alpha_0 + \alpha_1 Z_i + v_i$$
  2. Regress
$$Y_i = \beta_0 + \beta_1 X_i + \gamma \hat{v}_i + u_i$$
  3. Test $\gamma = 0$. If $\gamma \neq 0$, then X_i is endogenous.

Qualitative Dependent Variables

Qualitative Variables

- We have already seen cases where the explanatory variable is a dummy variable. We are now interested in models where the dependent variable is of qualitative nature, coded as a dummy variable. We might be interested in analyzing the determinants of:
  - driving to work versus public transportation.
  - living in a flat or a house.
  - investing in England or abroad.
- Y_i takes only two values, 0 or 1.
- This lecture will analyze different models:
  - linear probability model.
  - non linear models (logit, probit).

Linear Probability Model

- This is in fact the same linear model we have studied up to now, under a new name and with some new interpretations.

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad Y_i \in \{0, 1\}$$

- Interpretation:
  - $E[Y_i|X] = \beta_0 + \beta_1 X_i = Prob(Y_i = 1)$
  - $\hat{\beta}_0 + \hat{\beta}_1 X_i$ is the predicted probability that $Y_i = 1$.
- Example: Probability of high education, Sweden.

High educ	Coef.	Std. Err.	t
sex	-.0119405	.0081005	-1.47
age	-.0004545	.0002308	-1.97
Father educ, medium	.241207	.014455	16.69
Father educ, high	.2939869	.01347	21.83
constant	.2395922	.0129522	18.50

Limitations of the LPM

- Non normality of the residuals. Conditional on X_i, the residuals take only two values:

$$u_i = 1 - \beta_0 - \beta_1 X_i \qquad \text{if } Y_i = 1$$

$$u_i = -\beta_0 - \beta_1 X_i \qquad \text{if } Y_i = 0$$

  Normality is not required for the consistency of OLS, but a problem if one wants to make inference in small samples.

- Heteroskedasticity: The error term has not a constant variance:

$$\begin{aligned} V(u_i) &= E(u_i)^2 \\ &= (1 - \beta_0 - \beta_1 X_i)^2 p_i + (1 - p_i)(-\beta_0 - \beta_1 X_i)^2 \\ &= (1 - p_i)^2 p_i + (1 - p_i) p_i^2 \\ &= p_i(1 - p_i) = \sigma_i^2 \end{aligned}$$

  The variance of the residual depends on X_i.

- Predicted probabilities outside [0,1]. The predicted probability of outcome 1 for observation i is $\beta_0 + \beta_1 X_i$. Nothing prevents this to be within [0,1]. Which is problematic for a probability.

- Constant marginal effects. Given the linearity of the model,

$$\partial p_i / \partial X_i = \beta_1$$

  For instance, the effect of a change in income on the probability of choosing to commute by car rather than public transportation is low for low income household- as they use their income for other purposes-, high for middle income households.

- The R^2 is a dubious measure of the goodness of fit with a binary dependent variable.

New models

- These problems call for more complex models.
  - non linear models (in the parameters).
  - model explicitly the qualitative feature.
- estimation more complicated. OLS does not work.
- interpretation of results more complicated.

Structure of the model

- We define a **latent variable** Y_i^*, which is unobserved, but determined by the following model:

$$Y_i^* = \beta_0 + \beta_1 X_i + u_i$$

We observe the variable Y_i which is linked to Y_i^* as:

$$\begin{cases} Y_i = 0 & \text{if } Y^* < 0 \\ Y_i = 1 & \text{if } Y^* \geq 0 \end{cases}$$

- The probability of observing $Y_i = 1$ is:

$$\begin{aligned} p_i = P(Y_i = 1) &= P(Y_i^* \geq 0) \\ &= P(\beta_0 + \beta_1 X_i + u_i \geq 0) \\ &= P(u_i \geq -\beta_0 - \beta_1 X_i) \\ &= 1 - F_u(-\beta_0 - \beta_1 X_i) \end{aligned}$$

where F_u is the cumulative distribution function of the random variable u.

Logit and Probit

- Depending on the distribution of the error term, we have different models:
  - u is normal. This is the **probit** model.

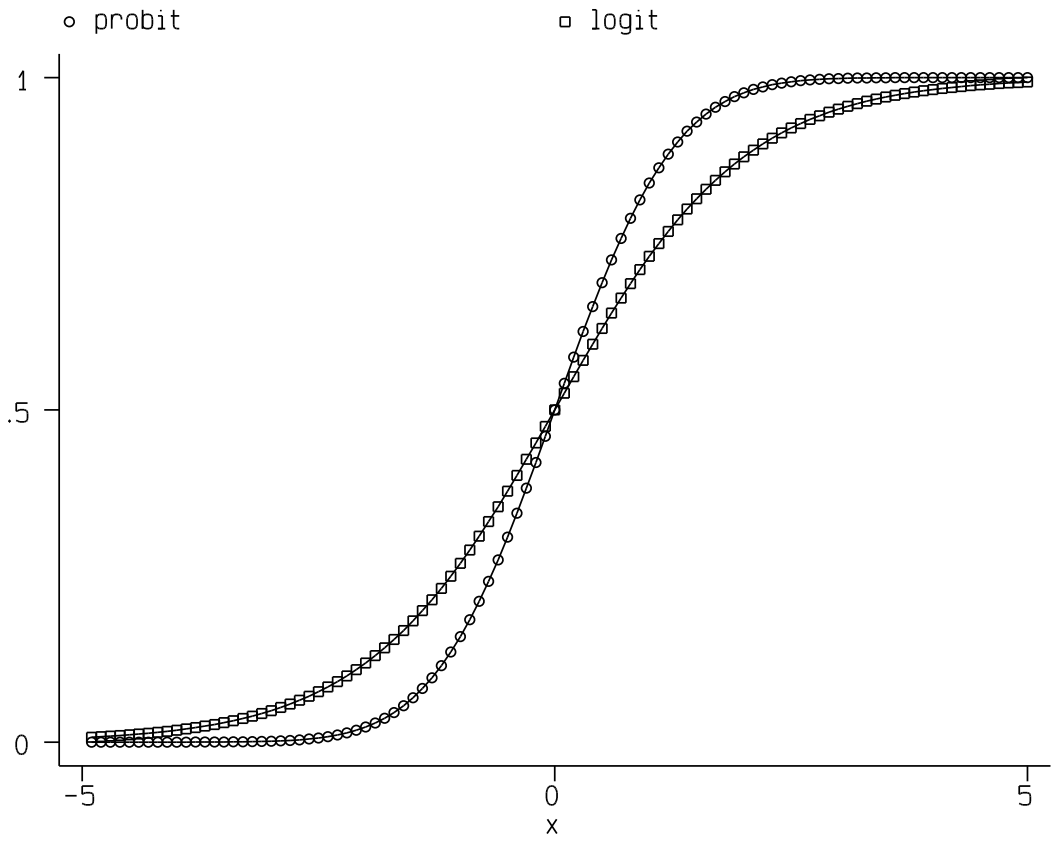
$$p_i = 1 - \Phi(-\beta_0 - \beta_1 X_i) = \Phi(\beta_0 + \beta_1 X_i)$$

  - u is logistic. This is the **logit** model.

$$p_i = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$$

- As both models are non linear, β_1 is *not* the marginal effect of X on Y.

Shape of Logit and Probit Models



Odds-Ratio

- Define the ratio $p_i/(1-p_i)$ as the **odds-ratio**. This is the ratio of the probability of outcome 1 over the probability of outcome 0. If this ratio is equal to 1, then both outcomes have equal probability ($p_i = 0.5$). If this ratio is equal to 2, say, then outcome 1 is twice as likely than outcome 0 ($p_i = 2/3$).
- In the logit model, the log odds-ratio is linear in the parameters:

$$\ln \frac{p_i}{1-p_i} = \beta_0 + \beta_1 X_i$$

- In the logit model, β_1 is the marginal effect of X on the log odds-ratio. A unit increase in X leads to an increase of β_1 % in the odds-ratio.

Marginal Effects

- Logit model:

$$\begin{aligned}\frac{\partial p_i}{\partial X_i} &= \frac{\beta_1 \exp(\beta_0+\beta_1 X_i)(1+\exp(\beta_0+\beta_1 X_i)) - \beta_1 \exp(\beta_0+\beta_1 X_i)^2}{(1+\exp(\beta_0+\beta_1 X_i))^2} \\ &= \frac{\beta_1 \exp(\beta_0+\beta_1 X_i)}{(1+\exp(\beta_0+\beta_1 X_i))^2} \\ &= \beta_1 p_i(1-p_i)\end{aligned}$$

  A one unit increase in X leads to an increase of $\beta_1 p_i(1-p_i)$

- Probit model:

$$\frac{\partial p_i}{\partial X_i} = \beta_1 \phi(\beta_0+\beta_1 X_i)$$

  A one unit increase in X leads to an increase of $\beta_1 \phi(\beta_0+\beta_1 X_i)$

Estimation

- Both logit and probit are non linear models. The estimation is done by maximum likelihood. The likelihood of observing $Y_i = 1$, $i \in N_1$ and $Y_j = 0$, $j \in N_0$ is:

$$L = \prod_{i \in N_1} Prob(Y_i = 1) \prod_{j \in N_0} Prob(Y_j = 0)$$

- The optimal parameters are found by maximizing the likelihood with respect to β_0 and β_1:

$$\begin{cases} \dfrac{\partial L}{\partial \beta_0} = 0 \\ \\ \dfrac{\partial L}{\partial \beta_1} = 0 \end{cases}$$

  This is a (non linear) system with 2 unknowns and 2 equations, which can be solved numerically.

Example

- We have data from households in Kuala Lumpur (Malaysia) describing household characteristics and their concern about the environment. The question is

  "Are you concerned about the environment? Yes / No".

  We also observe their age, sex (coded as 1 men, 0 women), income and quality of the neighborhood measured as air quality. The latter is coded with a dummy variable *smell*, equal to 1 if there is a bad smell in the neighborhood. The model is:

  $$Concern_i = \beta_0 + \beta_1 age_i + \beta_2 sex_i + \beta_3 \log income_i + \beta_4 smell_i + u_i$$

- We estimate this model with three specifications, LPM, logit and probit:

Probability of being concerned by Environment

Variable	LPM		Logit		Probit	
	Est.	t-stat	Est.	t-stat	Est.	t-stat
age	.0074536	3.9	.0321385	3.77	.0198273	3.84
sex	.0149649	0.3	.06458	0.31	.0395197	0.31
log income	.1120876	3.7	.480128	3.63	.2994516	3.69
smell	.1302265	2.5	.5564473	2.48	.3492112	2.52
constant	-.683376	-2.6	-5.072543	-4.37	-3.157095	-4.46
	Some Marginal Effects					
Age	.0074536		.0077372		.0082191	
log income	.1120876		.110528		.1185926	
smell	.1302265		.1338664		.1429596	

Multinomial Logit

- The logit model was dealing with two qualitative outcomes. This can be generalized to multiple outcomes:
  - choice of transportation: car, bus, train...
  - choice of dwelling: house, apartment, social housing.
- The multinomial logit: Denote the outcomes as $A, B, C...$ and p_A the probability of outcome A.

$$p_A = \frac{\exp(\beta_0^A + \beta_1^A X_i)}{\exp(\beta_0^A + \beta_1^A X_i) + \exp(\beta_0^B + \beta_1^B X_i) + \exp(\beta_0^C + \beta_1^C X_i)}$$
$$p_B = \frac{\exp(\beta_0^B + \beta_1^B X_i)}{\exp(\beta_0^A + \beta_1^A X_i) + \exp(\beta_0^B + \beta_1^B X_i) + \exp(\beta_0^C + \beta_1^C X_i)}$$
$$p_C = \frac{\exp(\beta_0^C + \beta_1^C X_i)}{\exp(\beta_0^A + \beta_1^A X_i) + \exp(\beta_0^B + \beta_1^B X_i) + \exp(\beta_0^C + \beta_1^C X_i)}$$

Identification

- If we multiply all the coefficients by a factor λ this does not change the probabilities p_A, p_B and p_C, as the factor cancel out. This means that there is under identification. We have to normalize the coefficients of one outcome, say, C to zero. All the results are interpreted as deviations from the baseline choice.

$$
\begin{aligned}
p_A &= \frac{\exp(\beta_0^A + \beta_1^A X_i)}{\exp(\beta_0^A + \beta_1^A X_i) + \exp(\beta_0^B + \beta_1^B X_i) + 1} \\
p_B &= \frac{\exp(\beta_0^B + \beta_1^B X_i)}{\exp(\beta_0^A + \beta_1^A X_i) + \exp(\beta_0^B + \beta_1^B X_i) + 1} \\
p_C &= \frac{1}{\exp(\beta_0^A + \beta_1^A X_i) + \exp(\beta_0^B + \beta_1^B X_i) + 1}
\end{aligned}
$$

- We can express the logs odds-ratio as:

$$
\begin{aligned}
\ln \frac{p_A}{p_C} &= \beta_0^A + \beta_1^A X_i \\
\ln \frac{p_B}{p_C} &= \beta_0^B + \beta_1^B X_i
\end{aligned}
$$

- The odds-ratios of choice A versus C are only expressed as a function of the parameters of choice A, but not of those for choice B: Independence of Irrelevant Alternatives (IIA).

Independence of Irrelevant Alternatives

- Consider travelling choices, by car or with a red bus. Assume for simplicity that the choice probabilities are equal:

$$P(car) = P(\text{red bus}) = 0.5 \quad \Longrightarrow \quad \frac{P(car)}{P(\text{red bus})} = 1$$

- Suppose we introduce a blue bus, (almost) identical to the red bus. The probability that individuals will choose the blue bus is therefore the same as for the red bus and the odd ratio is:

$$P(\text{blue bus}) = P(\text{red bus}) \quad \Longrightarrow \quad \frac{P(\text{blue bus})}{P(\text{red bus})} = 1$$

- However, the IIA implies that odds ratios are the same whether of not another alternative exists. The only probabilities for which the three odds ratios are equal to one are:

$$P(car) = P(\text{blue bus}) = P(\text{red bus}) = 1/3$$

  However, the prediction we ought to obtain is:

$$P(\text{red bus}) = P(\text{blue bus}) = 1/4 \quad P(car) = 0.5$$

Marginal Effects: Multinomial Logit

- β_1^A can be interpreted as the marginal effect of X on the log odds-ratio of choice A to the baseline choice.
- The marginal effect of X on the probability of choosing outcome A can be expressed as:

$$\frac{\partial p_A}{\partial X_i} = p_A[\beta_1^A - (p_A\beta_1^A + p_B\beta_1^B + p_C\beta_1^C)]$$

  Hence, the marginal effect on choice A involves not only the coefficients relative to A but also the coefficients relative to the other choices.

- Note that we can have $\beta_1^A < 0$ and $\partial p_A/\partial X_i > 0$ or vice versa.

  Due to the non linearity of the model, the sign of the coefficients does not indicate the direction nor the magnitude of the effect of a variable on the probability of choosing a given outcome. One has to compute the marginal effects.

Example

- We analyze here the choice of dwelling: house, apartment or low cost flat, the latter being the baseline choice. We include as explanatory variables the age, sex and log income of the head of household:

Variable	Estimate	Std. Err.	Marginal Effect
	Choice of House		
age	.0118092	.0103547	-0.002
sex	-.3057774	.2493981	-0.007
log income	1.382504	.1794587	0.18
constant	-10.17516	1.498192	
	Choice of Apartment		
age	.0682479	.0151806	0.005
sex	-.89881	.399947	-0.05
log income	1.618621	.2857743	0.05
constant	-15.90391	2.483205	

Ordered Models

- In the multinomial logit, the choices were not ordered. For instance, we cannot rank cars, busses or train in a meaningful way. In some instances, we have a natural ordering of the outcomes even if we cannot express them as a continuous variable:
  - Yes / Somehow / No.
  - Low / Medium / High
- We can analyze these answers with ordered models.

Ordered Probit

- We code the answers by arbitrary assigning values:

$$Y_i = 0 \text{ if No}, \quad Y_i = 1 \text{ if Somehow}, \quad Y_i = 2 \text{ if Yes}$$

- We define a latent variable Y_i^* which is linked to the explanatory variables:

$$Y_I^* = \beta_0 + \beta_1 X_i + u_i$$

$$\begin{aligned} Y_i &= 0 \qquad \text{if } Y_i^* < 0 \\ Y_i &= 1 \qquad \text{if } Y_i^* \in [0, \mu[\\ Y_i &= 2 \qquad \text{if } Y_i^* \geq \mu \end{aligned}$$

  μ is a threshold and an auxiliary parameter which is estimated along with β_0 and β_1.

- We assume that u_i is distributed normally.

- The probability of each outcome is derived from the normal cdf:

$$\begin{aligned} P(Y_i = 0) &= \Phi(-\beta_0 - \beta_1 X_i) \\ P(Y_i = 1) &= \Phi(\mu - \beta_0 - \beta_1 X_i) - \Phi(-\beta_0 - \beta_1 X_i) \\ P(Y_i = 2) &= 1 - \Phi(\mu - \beta_0 - \beta_1 X_i) \end{aligned}$$

Ordered Probit

- Marginal Effects:

$$\frac{\partial P(Y_i = 0)}{\partial X_i} = -\beta_1 \phi(-\beta_0 - \beta_1 X_i)$$
$$\frac{\partial P(Y_i = 1)}{\partial X_i} = \beta_1 \left(\phi(\beta_0 + \beta_1 X_i) - \phi(\mu - \beta_0 - \beta_1 X_i)\right)$$
$$\frac{\partial P(Y_i = 2)}{\partial X_i} = \beta_1 \phi(\mu - \beta_0 - \beta_1 X_i)$$

- Note that if $\beta_1 > 0$, $\partial P(Y_i = 0)/\partial X_i < 0$ and $\partial P(Y_i = 2)/\partial X_i > 0$:
  - If X_i has a positive effect on the latent variable, then by increasing X_i, fewer individuals will stay in category 0.
  - Similarly, more individuals will be in category 2.
  - In the intermediate category, the fraction of individual will either increase or decrease, depending on the relative size of the inflow from category 0 and the outflow to category 2.

Limited Dependent Variable Models

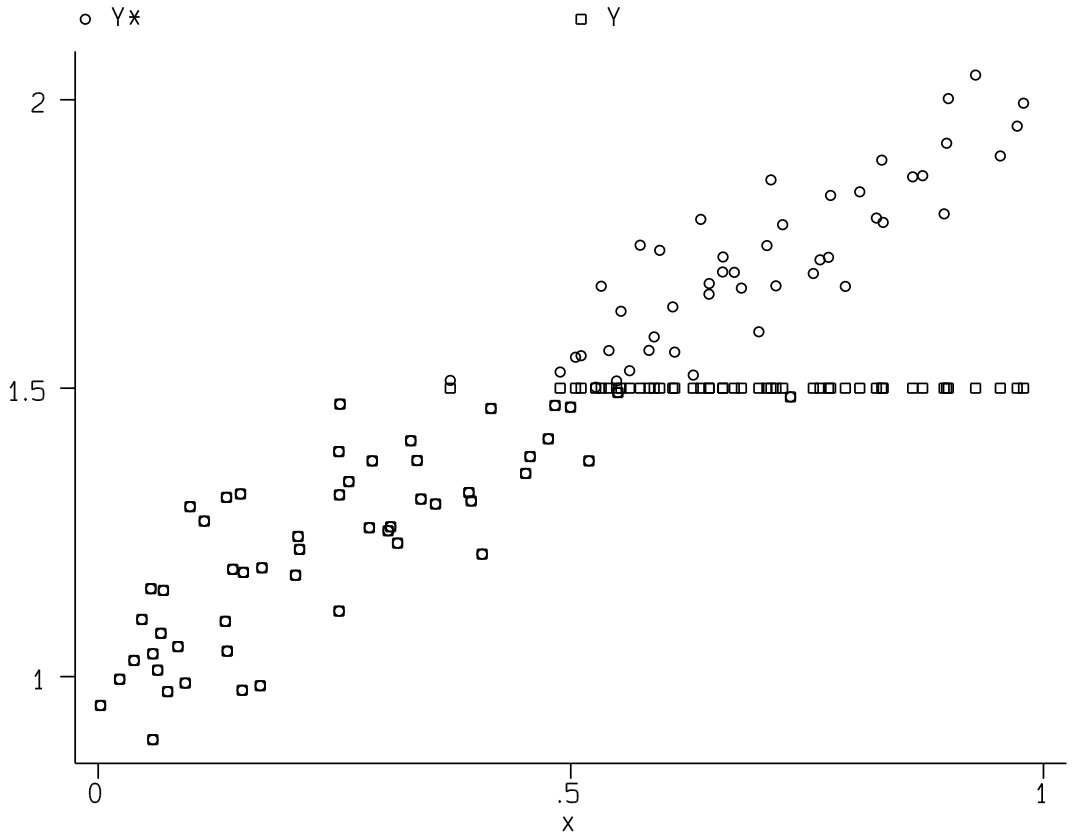
Tobit Model

- Structure of the model:

$$Y_i^* = \beta_0 + \beta_1 X_i + u_i$$

$$\begin{aligned} Y_i &= Y_i^* \qquad \text{if } Y_i^* \leq \mu \\ Y_i &= \mu \qquad \text{if } Y_i^* > \mu \end{aligned}$$

- Example: $\mu = 1.5$, $\beta_0 = 1$, $\beta_1 = 1$.



Tobit Model

- Marginal Effect:
  - Marginal effect of X_i on Y_i^*:

$$\partial Y_i^* / \partial X_i = \beta_1$$

  - Marginal effect of X_i on Y_i:

$$\frac{\partial Y_i}{\partial X_i} = \beta_1 \Phi(\frac{\beta_0 + \beta_1 X_i}{\sigma})$$

    Because of the truncation, note that $\partial Y_i / \partial X_i < \partial Y_i^* / \partial X_i$

Example: WTP

- The WTP is censored at zero. We can compare the two regressions:

  OLS: $WTP_i = \beta_0 + \beta_1 lny + \beta_2 age_i + \beta_3 smell_i + u_i$

  Tobit: $WTP_i^* = \beta_0 + \beta_1 lny + \beta_2 age_i + \beta_3 smell_i + u_i$
  $WTP_i = WTP_i^* \quad \text{if} \quad WTP_i^* > 0$
  $WTP_i = 0 \quad \text{if} \quad WTP_i^* < 0$

	OLS		Tobit		
Variable	Estimate	t-stat	Estimate	t-stat	Marginal effect
lny	2.515	2.74	2.701	2.5	2.64
age	-.1155	-2.00	-.20651	-3.0	-0.19
sex	.4084	0.28	.14084	0.0	.137
smell	-1.427	-0.90	-1.8006	-0.9	-1.76
constant	-4.006	-0.50	-3.6817	-0.4	

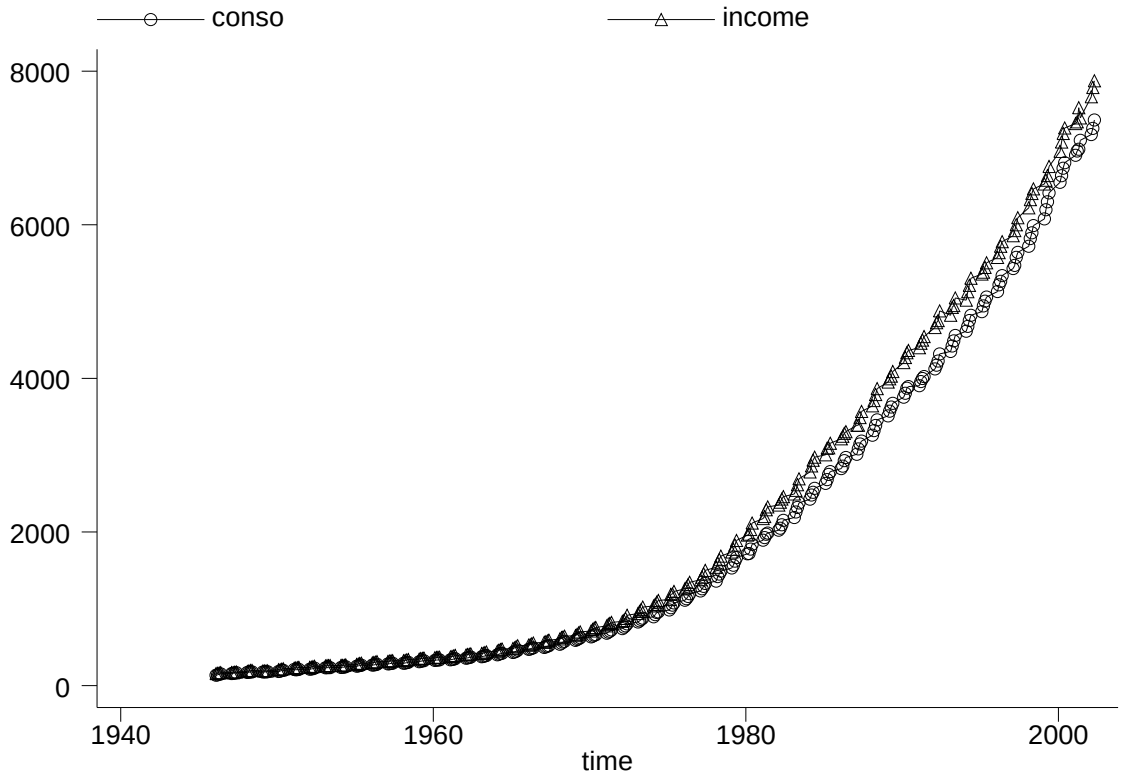
Time Series

Time Series

- The data set describes a phenomenon over time.
- Usually macro-economic series.
  - Temperature and carbon dioxide over time.
  - Unemployment over time.
  - Financial series over time.
- We want to describe and maybe forecast the evolution of this phenomenon.
  - evaluate the influence of explanatory variables.
  - evaluate the short-run effect of policy variables.
  - evaluate the long-run effect of policy variables.

Examples

- Disposable income and consumption in US:



Stationarity

- A series of data is said to be stationary if its mean and variance are constant across time periods.

$$E(Y_t) = \mu_y \qquad \text{for all } t$$

$$V(Y_t) = \sigma_y^2 \qquad \text{for all } t$$

$$cov(Y_t, Y_{t+k}) = \gamma_k \quad \text{for all } t$$

- A series is said to be non stationary if either the mean or the variance is varying with time.
  - Changing mean:

$$Y_t = \beta_0 + \beta_1 t + u_t$$

  - Changing variance:

$$Y_t = \beta_0 + \beta_1 X_t + u_t \qquad V(u_t) = t\sigma^2$$

- We will first study the case where the data is stationary.

Autoregressive Process

- AR(p) models:

$$Y_t = \mu + +\rho_1 Y_{t-1} + \ldots + \rho_p Y_{t-p} + u_t$$

  - Simplest form, AR(1):

$$Y_t = \mu + \rho_1 Y_{t-1} + u_t$$

$$\begin{cases} E(Y_t) = \dfrac{\mu}{1-\rho_1} \\ \\ V(Y_t) = \dfrac{\sigma^2}{1-\rho_1^2} \\ \\ cov(Y_t, Y_{t-k}) = \sigma^2 \dfrac{\rho_1^k}{1-\rho_1^2} \end{cases}$$

  - For the process to be stationary, we must have $|\rho_1| < 1$
  - Estimation: OLS.

Example

- Detrended Consumption in US over time (1946-2002):

$$c_t = \mu + \rho_1 c_{t-1} + \ldots \rho_p c_{t-p} + u_t$$

lags	Coefficient (s.e.)			
$t-1$	0.98 (0.01)	0.81 (0.1)	0.75 (0.1)	0.64 (0.07)
$t-2$		0.17 (0.1)	-0.05 (0.12)	-0.04 (0.09)
$t-3$			0.28 (0.08)	0.02 (0.09)
$t-4$				0.37 (0.08)
constant	-0.008 (0.07)	-0.01 (0.1)	-0.004 (0.13)	
Log lik.	548	552	561	577

AR: Implied Dynamics

- Suppose we have the following model:

$$Y_t = \mu + \rho_1 Y_{t-1} + \beta X_t + u_t$$

- Compare two scenarios starting in period t:
  - Scenario 1: X_t is constant for all future t.
  - Scenario 2: X_t is increased by 1 unit in t and then goes back to previous level.
- Compute Y_t^2 and Y_t^1:

$$\begin{aligned} Y_t^2 - Y_t^1 &= \beta \\ Y_{t+1}^2 - Y_{t+1}^1 &= \rho_1 \beta \\ Y_{t+2}^2 - Y_{t+2}^1 &= \rho_1^2 \beta \\ &\vdots \\ Y_{t+k}^2 - Y_{t+k}^1 &= \rho_1^k \beta \end{aligned}$$

  - β measures the immediate impact of X_t on Y_t.
  - $\beta\rho_1^k$ measures the impact after k periods of X_t on Y_t.
  - $\beta/(1-\rho_1)$ measures the long-run impact of X_t on Y_t.
- similar (and more complex) calculations can be done in the case of an AR(p) process.

Moving Average

- Y_t is a weighted sum of lagged i.i.d. shocks.

$$Y_t = \mu + u_t + \lambda_1 u_{t-1} + \ldots + \lambda_q u_{t-q}$$

- Simplest case, MA(1):

$$Y_t = \mu + u_t + \lambda_1 u_{t-1}$$

- Moments:

$$\begin{aligned} EY_t &= \mu \\ VY_t &= E(Y_t - \mu)^2 = (1 + \lambda^2)\sigma^2 \\ \gamma_1 &= cov(Y_t, Y_{t-1}) = \lambda\sigma^2 \\ \gamma_j &= cov(Y_t, Y_{t-j}) = 0 \end{aligned}$$

- Estimation: Maximum Likelihood.

- Example: MA(1) applied to US consumption:

Variable	Coefficient	s.e
μ	720	28
λ_1	0.83	0.6

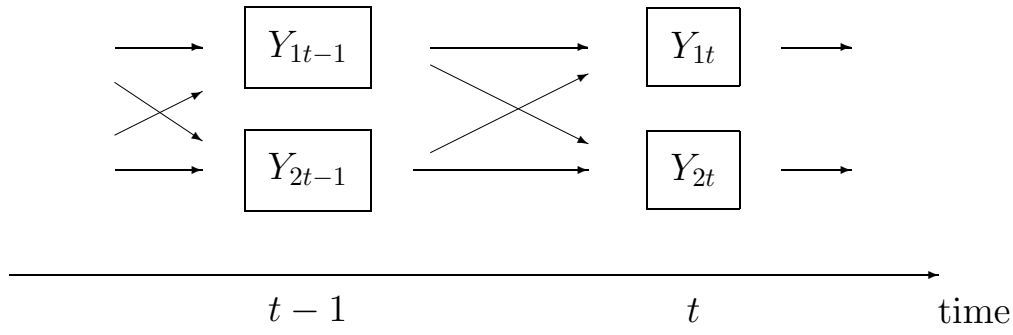
Example

- (Detrended) Log consumption in US, 1946-2002:

lags	Coefficient (s.e.)			
$t-1$	0.77 (0.05)	1.32 (0.03)	1.59 (0.07)	1.22 (0.07)
$t-2$		0.17 (0.1)	1.04 (0.03)	1.37 (0.09)
$t-3$			0.47 (0.08)	1.18 (0.08)
$t-4$				0.59 (0.08)
constant	-0.001 (0.007)	-0.01 (0.1)	-0.001 (0.01)	
Log lik.	286	409	452	470

Vector Autoregression (VAR)

- Model several outcomes jointly, as function of past values.
- e.g.: GDP, inflation, unemployment...
- Model:



- Analytically (VAR(1)):

$$\begin{cases} Y_{1t} = \rho_{11} Y_{1t-1} + \rho_{12} Y_{2t-1} + u_{1t} \\ \\ Y_{2t} = \rho_{21} Y_{1t-1} + \rho_{22} Y_{2t-1} + u_{2t} \end{cases}$$

Example

- VAR(2) for (detrended) log consumption and log income:

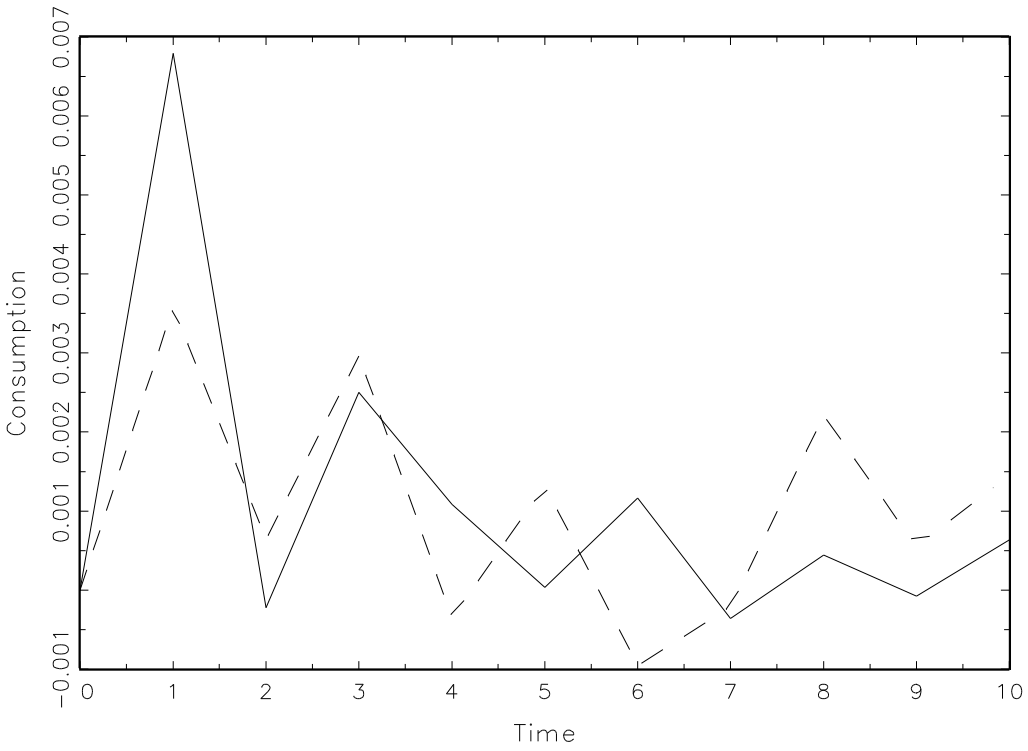
Var	Income	Consumption
UY(-1)	0.710221	0.032335
	(0.13817)	(0.13103)
UY(-2)	0.363859	0.125324
	(0.13941)	(0.13221)
UC(-1)	0.115288	0.754669
	(0.14562)	(0.13810)
UC(-2)	-0.206404	0.074719
	(0.14355)	(0.13613)
Constant	-0.001497	-0.001211
	(0.00148)	(0.00140)

Impulse Response Function

- VAR results can be difficult to interpret.
- What are the long-run effects of variables?
- How is the dynamic of Y_t^1 after a shock?
- Impulse response function measures this dynamic:
  - analyzes the predicted response of one of the dependent variables to a shock through time.
  - compare to a baseline with no shock.

Impulse Response Function

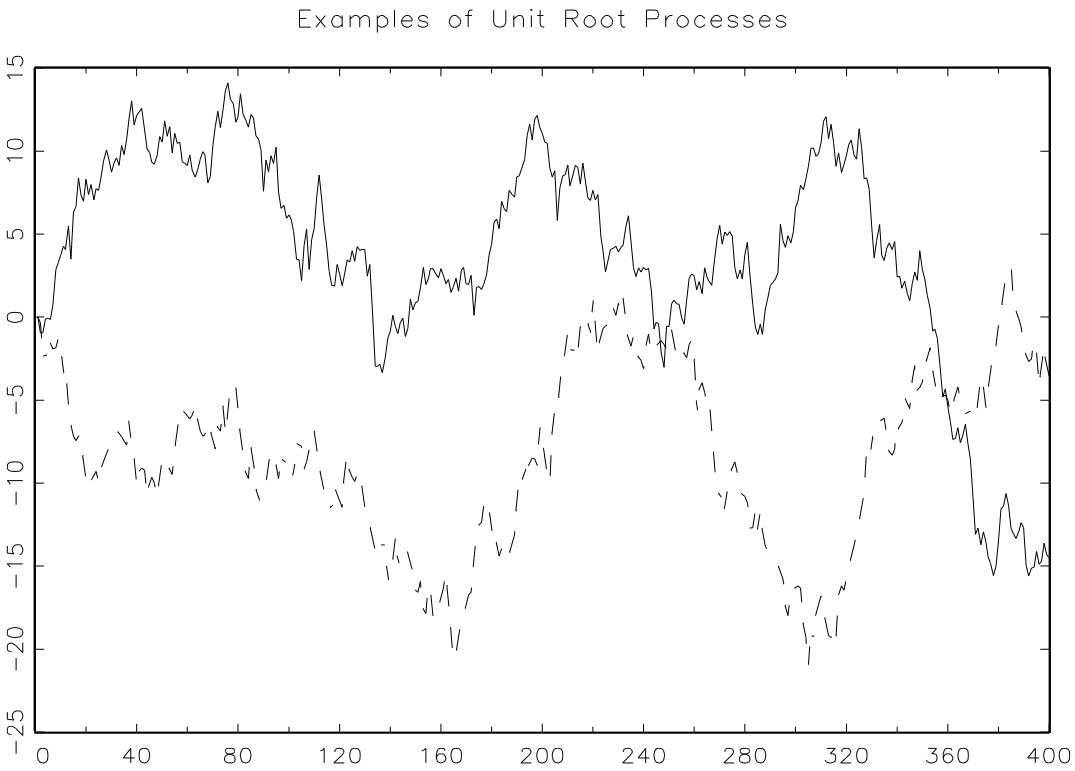
- Response of consumption to one standard deviation shock to consumption and income.



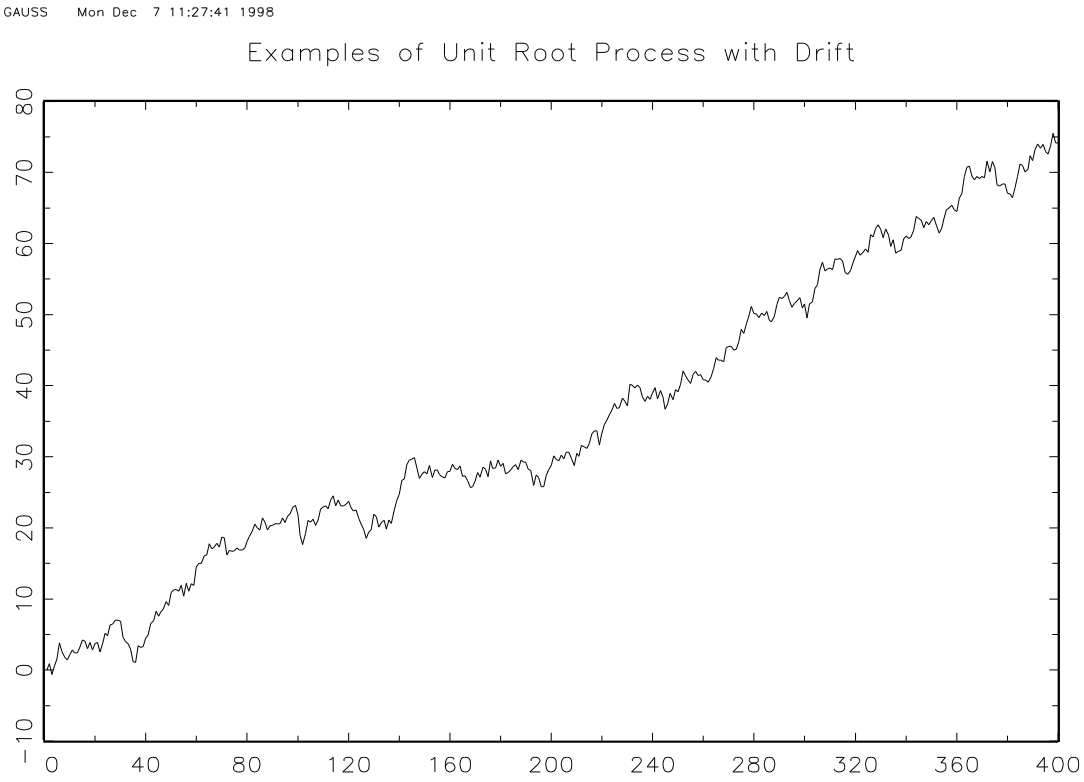
Non Stationary Series

- Trend Stationary Process:
  - Linear trend:
$$y_t = \mu_0 + \mu_1 t + \varepsilon_t$$
  - Exponential trend:
$$y_t = e^{\mu_0 + \mu_1 t + \varepsilon_t}$$
- Stochastic Trend
$$y_t = y_{t-1} + \varepsilon_t$$
$$y_t = \sum_{j=0}^{t} \varepsilon_j$$
  - The variance of y_t is growing over time:
$$V(y_t) = t\sigma^2$$
  - But the unconditional mean is zero: $Ey_t = 0$. Second order non stationarity.
- Stochastic trend with drift:
$$y_t = \mu + y_{t-1} + \varepsilon_t$$
$$y_t = t\mu + \sum_{j=0}^{t} \varepsilon_j$$
Now, $Ey_t = t.\mu$ and $V(y_t) = t\sigma^2$. Both the mean and variance are drifting with time.

Examples


Examples of Unit Root Processes
15
10
5
0
−5
−10
−15
−20
−25
0
40
80
120
160
200
240
280
320
360
400

Examples


GAUSS Mon Dec 7 11:27:41 1998
Examples of Unit Root Process with Drift
80
70
60
50
40
30
20
10
0
−10
0
40
80
120
160
200
240
280
320
360
400

Spurious Regression

- Suppose we have two completely unrelated series, Y_t^1 and Y_t^2 with a stochastic trend:

$$\begin{cases} Y_t^1 = Y_{t-1}^1 + u_{1t} \\ \\ Y_t^2 = Y_{t-1}^2 + u_{2t} \end{cases}$$

- A regression of Y_t^1 on Y_t^2 should give a coefficient of zero.
- However, this is not the case:

Variable	Estimate	T-stat
const	-5.53	-15
Y_t^2	-0.83	-17
$R^2 = 0.4$ $DW = 0.02$		

- spurious correlation. Because the two series have a stochastic trend, OLS is picking up this apparent correlation.
- In fact, the t test is not valid under non-stationarity.
- Risky to regress two variables which have are non stationary.

Back to Stationarity

- Suppose we want to estimate the relationship between two variables:

$$Y_t^1 = \alpha_0 + \alpha_1 Y_t^2 + v_t$$

- Suppose Y_t^1 and Y_t^2 are non-stationary. This regression will be a spurious one, in the sense that we might find an α_1 different from zero even if both series are unrelated.

- To test whether these variables are correlated, we need to make them stationary.

- Procedures to "stationarise" a series. Depends on the nature of non stationarity:

  - trend stationary.

$$\begin{cases} Y_t^1 = a_0 + a_1 t + u_{1t} \\ Y_t^2 = b_0 + b_1 t + u_{2t} \end{cases}$$

  - stochastic trend.

$$\begin{cases} Y_t^1 = Y_{t-1}^1 + u_{1t} \\ Y_t^2 = Y_{t-1}^2 + u_{2t} \end{cases}$$

How to Make a Series Stationary

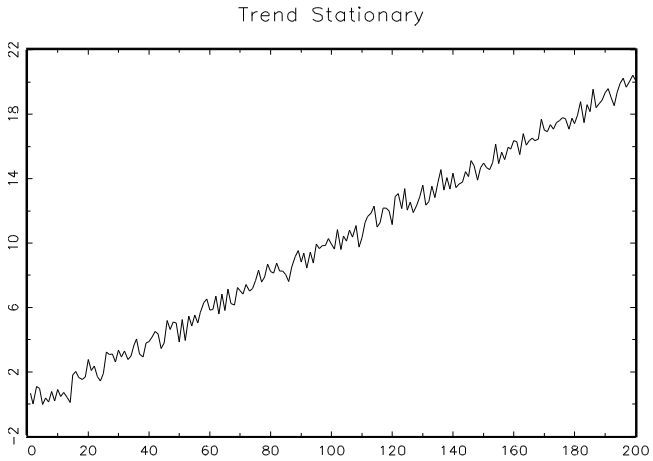
- Trend stationary process: remove the trend by regressing the dependent variable on a trend and taking residuals. Two step approach:
  1. Regress Y_t^1 and Y_t^2 on t and a constant.
  2. Predict the residuals $\hat{u}_{1t}$ and $\hat{u}_{2t}$.
  3. Regress $\hat{u}_{1t}$ on $\hat{u}_{2t}$.

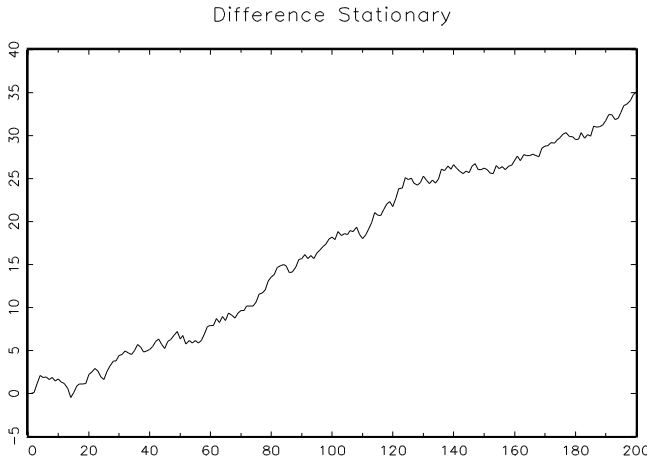
$$\hat{u}_{1t} = \alpha_1 \hat{u}_{2t} + v_t$$

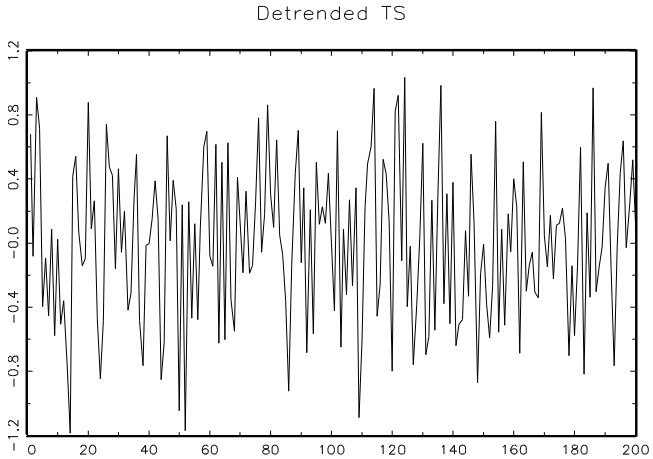
- Stochastic trend: Taking first differences gives a stationary process:

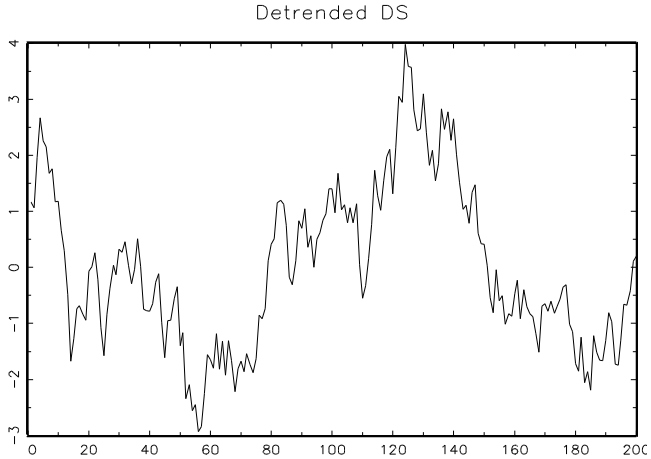
$$Y_t^1 - Y_{t-1}^1 = \alpha_1 (Y_t^2 - Y_{t-1}^2) + v_t - v_{t-1}$$

Examples

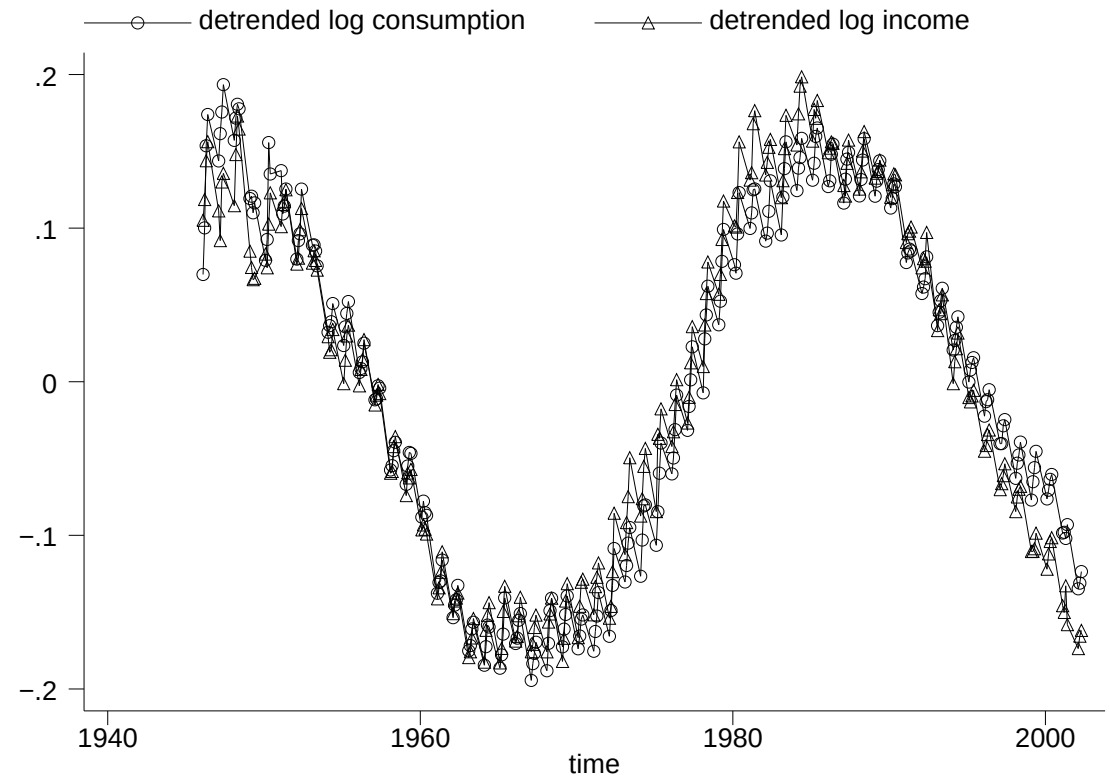

Trend Stationary
22
18
14
10
6
2
−2
0
20
40
60
80
100
120
140
160
180
200

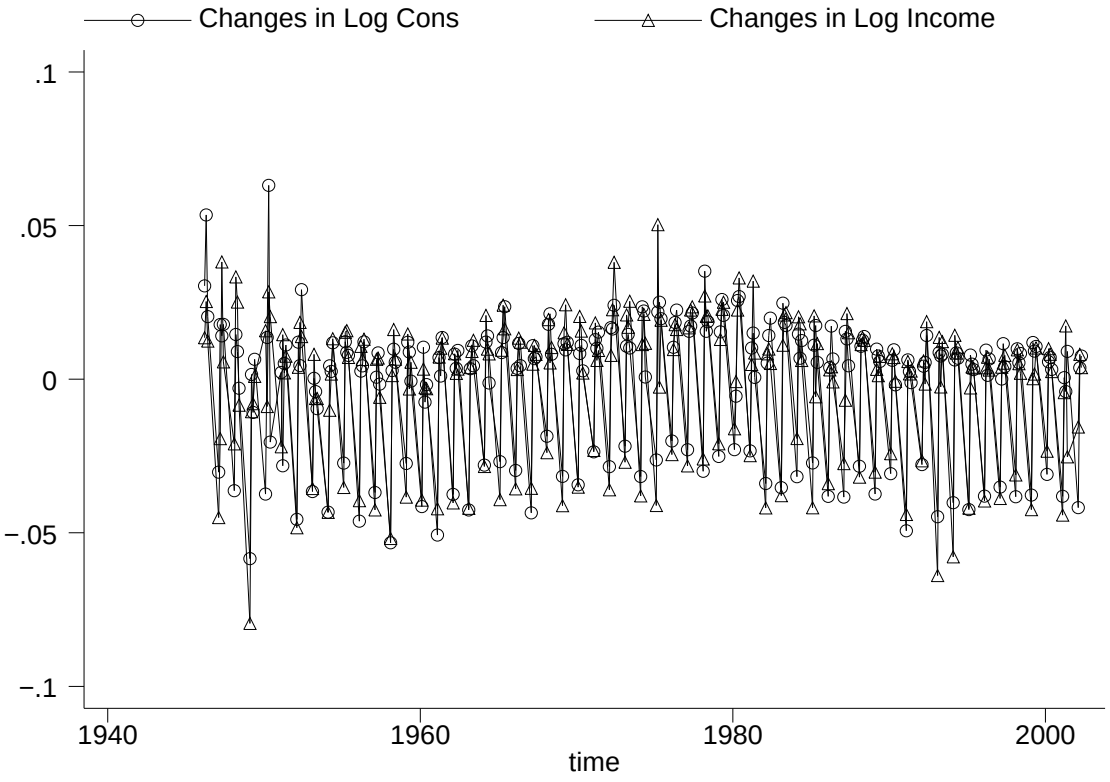

Difference Stationary
40
35
30
25
20
15
10
5
0
−5
0
20
40
60
80
100
120
140
160
180
200


Detrended TS
1.2
0.8
0.4
−0.0
−0.4
−0.8
−1.2
0
20
40
60
80
100
120
140
160
180
200


Detrended DS
4
3
2
1
0
−1
−2
−3
0
20
40
60
80
100
120
140
160
180
200

Example: US Consumption and Income

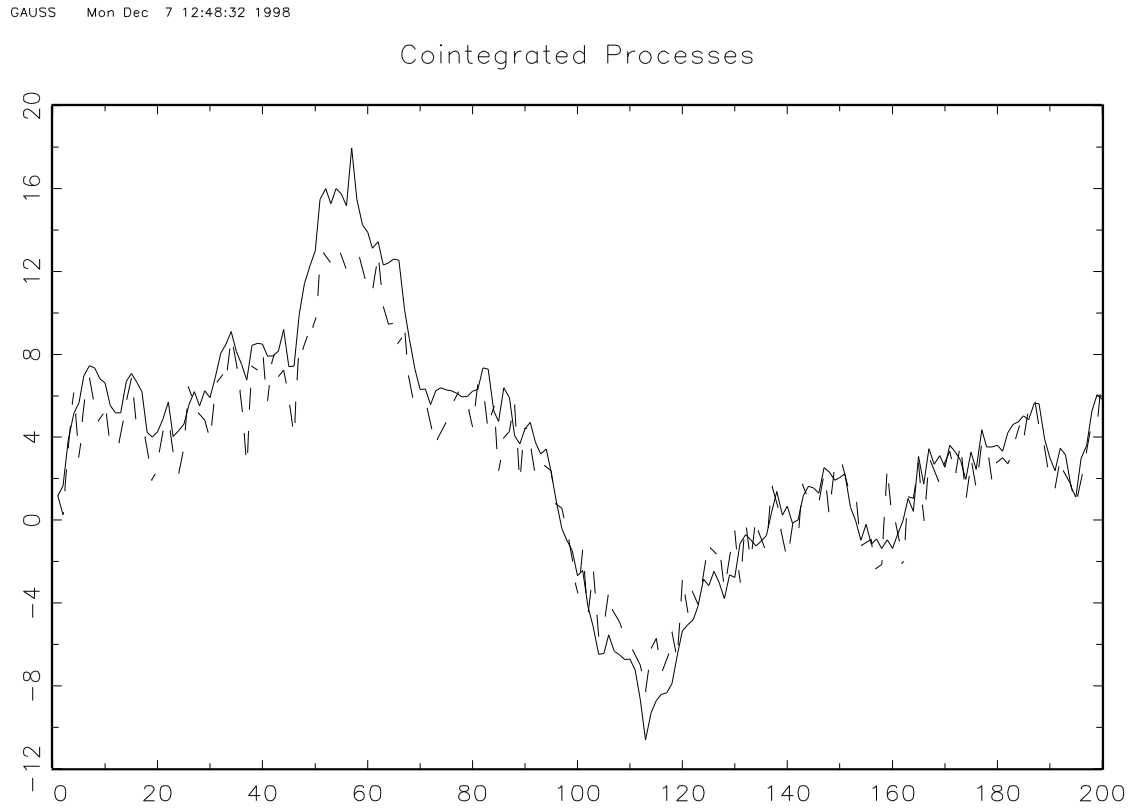




Cointegration

- Suppose Y_t^1 and Y_t^2 are non stationary.
- Definition: Y_t^1 and Y_t^2 are said to be cointegrated if there exits a coefficient α_1 such that $Y_t^1 - \alpha_1 Y_t^2$ is stationary.
- An OLS regression identifies the cointegration relationship.
- α_1 represents the long-run relationship between Y_t^1 and Y_t^2.

Example


GAUSS Mon Dec 7 12:48:32 1998
Cointegrated Processes
20
16
12
8
4
0
−4
−8
−12
0
20
40
60
80
100
120
140
160
180
200

Final Lecture Exercise

- We want to investigate the effect of tobacco on life expectancy. We have two data sets describing smoking and life expectancy or age at death:
  - DATASET 1: A cross section data set which describes average smoking and life expectancy across 50 countries in 1980.
  - DATASET 2: A cross section data set of about 40000 Swedish individuals. They were interviewed between 1980 and 1997. The data set reports the age, sex, education level, household size, occupation, region of living, health measures, as well as several variables describing smoking(dummy for ever smoker, duration of smoking habit, current number of cigarette smoked). The data set reports the age at death if death occurred before 1999.
- The object of this class is to investigate and quantify the effect of smoking on life expectancy. To this end, I propose to attack the problem in the following way:
  - What is our model? (simplest first).
  - How does it perform on the two data sets?
  - What is wrong with the simple approach?
  - How can we improve it?

Data Set 1

Variable	Description
country	country name.
ET0	life expectancy, women
ET1	life expectancy, men
q80	average per capita cigarettes per year in 1980
q90	average per capita cigarettes per year in 1980
gdp	GDP per capita
pop	population size
gini	Gini coefficient (measure of income inequality)
europe	dummy for European country
asia	dummy for Asian country
america	dummy for American country
devel	dummy for developing country
rprice	relative price of cigarettes

Data Set 2 (Not all variables are listed here)

Variable	Description
aad	age at death
smoker	dummy for ever smoker
Ysmoke	duration of smoking habit (in years)
Qcgtte	number of cigarettes currently smoked
age	age of individual
sex	dummy for men
Educ1-Educ3	dummies for level of education
hsize	household size
matri	marital status (single, married, divorced, widowed)
region	region of living
ypc	household income per capita
Ghealth	Self assessed health (good, fair or poor)
Alc1-Alc3	dummy for alcohol drinking (low, moderate, high)
height	height in centimeters