Microeconometrics

The general microeconometrics approach

Alex Armand

alex.armand@novasbe.pt

Course structure and evaluation

Microeconometrics aims at

1. giving **empirical content** to **economic relations**
2. using **micro-level** data
3. with a particular focus on **causal inference**

Course structure and evaluation

Grading of the course is divided in three components:

1. **Problem Sets (three)**: 15%
   - Tests what learned in class with practical examples
2. **Group assignment**: 30%
   - Paper replication project
3. **Final exam**: 55%
   - Covers all the material covered in the course
   - Minumum passing grade of 8/20.
   - In accordance with the school norms, there is no procedure for grade improvement after passing a course (no re-sit or second course enrolment)

Material

The course is based on the slides, but we will follow partial material of the following textbook:

- Wooldridge, J., **Econometric Analysis of Cross-Section and Panel Data**. MIT Press, Cambridge, MA.

- **Additional material** can be found here:
  - Angrist, J.D., and J.S. Pischke, **Mostly harmless econometrics: An empiricist's companion**. Princeton University Press.
  - [MORE ADVANCED - PhD level] Cameron, A. C., and P. K. Trivedi, **Microeconometrics: Methods and Applications**. Cambridge University Press, New York, NY.

No requirement of a specific software

The course will present a practical part using:

- STATA

You are free to use other software programs to practice or for the problem sets:

- R
- Matlab
- Etc...

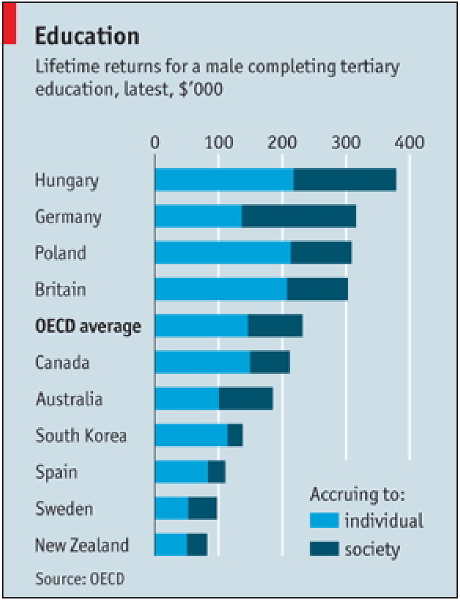
Topics covered

1. General microeconometrics approach
2. Identification in linear models: from cross-section to panel data
3. Non-linear models and the maximum likelihood
4. Latent variable models
5. Censored data
6. Sample selection
7. GMM
8. Statistical learning tools for causal analysis

Summary of today's class

1. Main concepts in microeconometrics
2. General microeconometrics approach
3. The identication problem
    - Parametric versus non-parametric solutions
4. Introduction to non-parametric methods

A problem: returns to education



Causation versus correlation

Interested in the causal effect between two random variables

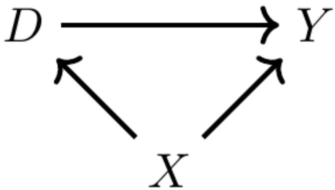
- Variable of interest: D
- Outcome variable: Y

Directed Acyclic Graphs (or DAGs): powerful instrument to understand causation versus correlation

$$D \longrightarrow Y$$

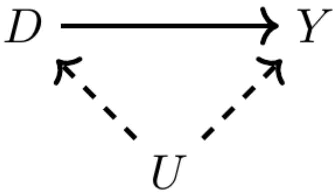
- Direction of the arrow represents the **causal effect**
- Can we recover this relationship?

Some definitions: confounder



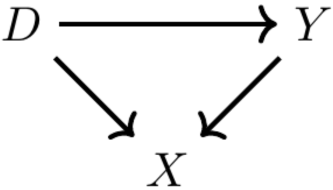
- Two paths from D to Y
  1. **Causal effect**: direct path $D \rightarrow Y$
  2. **Backdoor path**: $D \leftarrow X \rightarrow Y$
     - Spurious correlation between D and Y driven by X
     - X is a **confounder**

Some definitions: observability



- Two paths from D to Y
  1. **Causal effect**: direct path $D \to Y$
  2. **Backdoor path**: $D \leftarrow U \to Y$
     - U <u>unobserved</u> to the researcher (dashed lines)
     - U is a **unobserved confounder (or non-collider)** $\Rightarrow$ backdoor path is **open**

Some definitions: collider



- Two paths from D to Y

  1. **Direct path**: $D \to Y$
  2. **Backdoor path**: $D \to X \leftarrow Y$

     - Two variables cause a third variable along some path
     - X is a **collider** $\Rightarrow$ backdoor path is **closed**

Solving for causation

- Open backdoor paths $\Rightarrow$
  - Systematic non-causal correlations between the variables of interest
- How to solve this problem?
- **Backdoor criterion**: when all backdoor paths are closed, the focus is on causal relationships
  1. **Backdoor path is open**: control for the variable $\Rightarrow$ importance of observability
  2. **Backdoor path is closed**: ignore the variable $\Rightarrow$ controlling would open the path (*bad control*)

Solving for causation

- Open backdoor paths $\Rightarrow$
  - Systematic non-causal correlations between the variables of interest
- How to solve this problem?
- **Backdoor criterion**: when all backdoor paths are closed, the focus is on causal relationships
  1. **Backdoor path is open**: control for the variable $\Rightarrow$ importance of observability
  2. **Backdoor path is closed**: ignore the variable $\Rightarrow$ controlling would open the path (*bad control*)

Solving for causation

- Open backdoor paths $\Rightarrow$
  - Systematic non-causal correlations between the variables of interest

- How to solve this problem?

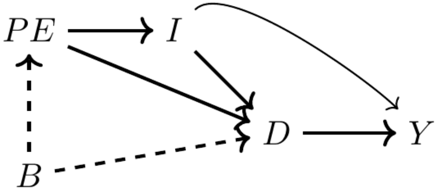
- **Backdoor criterion**: when all backdoor paths are closed, the focus is on causal relationships
  1. **Backdoor path is open**: control for the variable $\Rightarrow$ importance of observability
  2. **Backdoor path is closed**: ignore the variable $\Rightarrow$ controlling would open the path (*bad control*)

Solving for causation

- Open backdoor paths $\Rightarrow$
  - Systematic non-causal correlations between the variables of interest
- How to solve this problem?
- **Backdoor criterion**: when all backdoor paths are closed, the focus is on causal relationships
  1. **Backdoor path is open**: control for the variable $\Rightarrow$ importance of observability
  2. **Backdoor path is closed**: ignore the variable $\Rightarrow$ controlling would open the path (*bad control*)

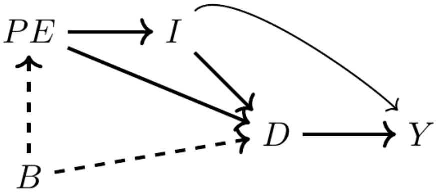
An example

What is the **causal effect of education (D) on earnings (Y)**?



- PE: parental education
- I: family income
- B: background factors (e.g., family environment, mental ability, ...)

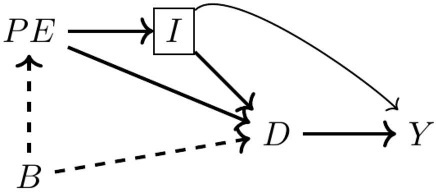
An example



- Four paths between D and Y:
  1. **Causal effect**: $D \rightarrow Y$
  2. **Backdoor path 1**: $D \leftarrow I \rightarrow Y$
  3. **Backdoor path 2**: $D \leftarrow PE \rightarrow I \rightarrow Y$
  4. **Backdoor path 3**: $D \leftarrow B \rightarrow PE \rightarrow I \rightarrow Y$

- How can we isolate the causal effect?

An example

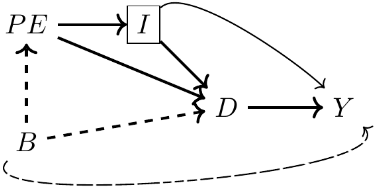


- Identify the open backdoor to close $\Rightarrow$ all through I
- In a linear model, run the following regression controlling for the open backdoor

$$Y_i = \alpha + \beta D_i + \gamma I_i + \epsilon_i$$

  - β can be interpreted as **causal**

An example



- What if you run the same regression?

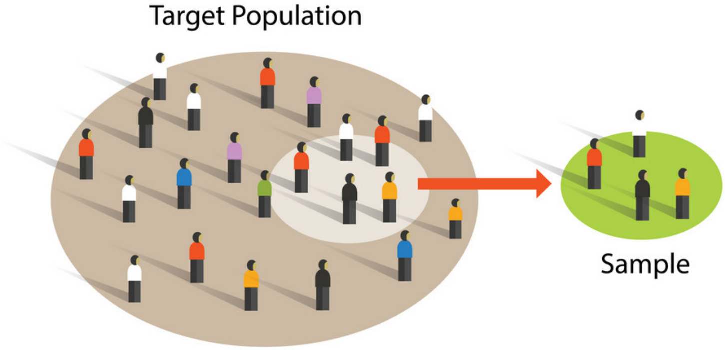
$$Y_i = \alpha + \beta D_i + \gamma I_i + \epsilon_i$$

- One backdoor remains open (causing D to be correlated with ϵ)
- β can be interpreted as **(spurious) correlation** $\Rightarrow$ it captures the causal effect + the backdoor effect through B

Studying causal relationships: population and random sampling

The notion of a **population** is very important

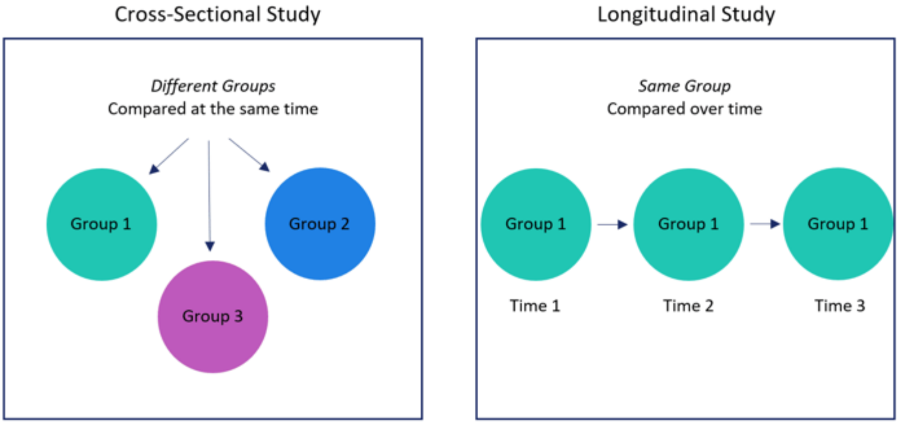
1. Large set of objects of a similar nature - e.g. human beings, households, readings from a measurement device
2. Of interest as a whole



Studying causal relationships: population and random sampling

- Our data are obtained as a **random sample from a specified population of interest**
  - A subset of objects is drawn from a population
  - Each draw is independent and identically distributed (i.i.d.)
- Later on we will focus on violations of random sampling
  - Censoring
  - Missing data
  - Sample selection problems

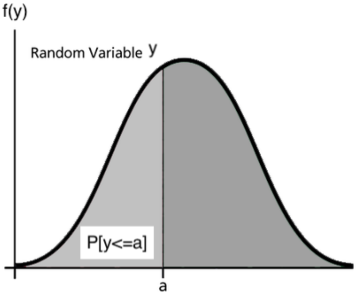
Studying causal relationships: type of data

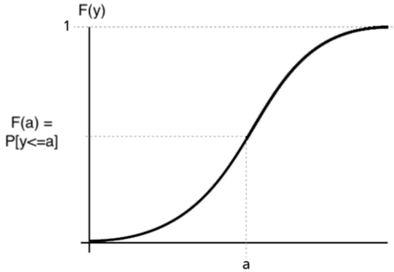


- We will begin by assuming **cross-sectional data**
- We will extend the material to panel (or longitudinal) data or mixed cases (repeated cross sections)

What do we care about outcomes?

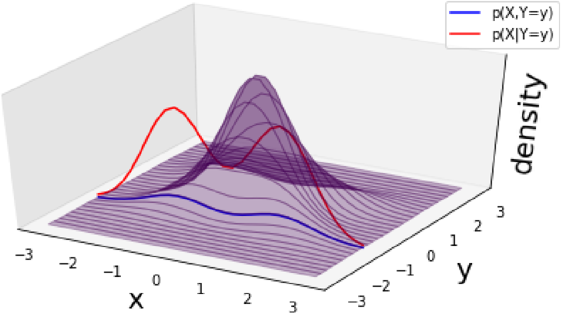
- Density function $f(y)$ (**pdf**) and cumulative distribution function $F(Y)$ (**cdf**)
- Conditional density function $f(y|x)$
- Moments of these distributions (unconditional and conditional means and variance)





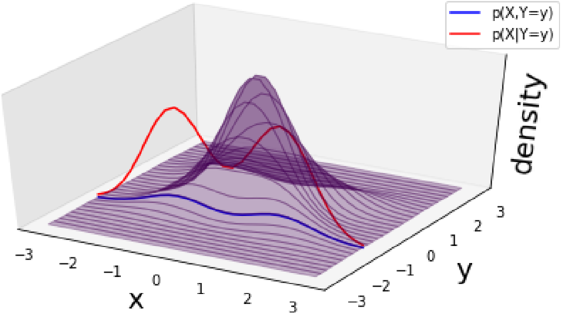
What do we care about outcomes?

- Density function $f(y)$ (**pdf**) and cumulative distribution function $F(Y)$ (**cdf**)
- Conditional density function $f(y|\text{x})$
- Moments of these distributions (unconditional and conditional means and variance)



What do we care about outcomes?

- Density function $f(y)$ (**pdf**) and cumulative distribution function $F(Y)$ (**cdf**)
- Conditional density function $f(y|\mathrm{x})$
- Moments of these distributions (unconditional and conditional means and variance)



General microeconometrics approach

Start with a model for a **conditional density** in the population of interest

$$f(y|\mathrm{x};\beta)$$

- x: vector of (observable) control variables
- β: vector of parameters to be estimated

1. Discuss whether β can be identified and what assumptions are needed
2. Assume we have access to a random sample of size N

$$\{(\mathrm{x}_i, y_i) : i = 1, ..., N\}$$

3. Use a random sample to estimate the population parameter

General microeconometrics approach

Start with a model for a **conditional density** in the population of interest

$$f(y|\mathrm{x}; \beta)$$

- x: vector of (observable) control variables
- β: vector of parameters to be estimated

1. Discuss whether β can be identified and what assumptions are needed
2. Assume we have access to a random sample of size N

$$\{(\mathrm{x}_i, y_i) : i = 1, ..., N\}$$

3. Use a random sample to estimate the population parameter

General microeconometrics approach

Start with a model for a **conditional density** in the population of interest

$$f(y|\mathrm{x};\beta)$$

- x: vector of (observable) control variables
- β: vector of parameters to be estimated

1. Discuss whether β can be identified and what assumptions are needed
2. Assume we have access to a random sample of size N

$$\{(\mathrm{x}_i, y_i) : i = 1, ..., N\}$$

3. Use a random sample to estimate the population parameter

When this approach is feasible?

A statistical model is identifiable if...

...it is theoretically possible to learn the true values underlying parameters with an infinite number of observations from it.

- Types of identification
  1. **Point identification**: underlying parameters are unique values
  2. **Set identification**: underlying parameters are sets of values
  3. **Not identified**: more than one set of parameters generate the same distribution of observations (**observationally equivalent**)

Identification: some examples

With random sampling, identification is purely **a population problem**

- If the model is well specified $\Rightarrow$ it will have similar characteristics to the population when the sample gets larger
- Imagine the following population models:
  1. $y = x\beta$
  2. $y = x\beta + u$
  3. $y = \frac{x\beta}{\lambda} + u$
- Can we identify the parameters β and λ in these cases? What assumptions we need for x and u?

Identification: some examples

With random sampling, identification is purely **a population problem**

- If the model is well specified $\Rightarrow$ it will have similar characteristics to the population when the sample gets larger
- Imagine the following population models:
  1. $y = \mathrm{x}\beta$
  2. $y = \mathrm{x}\beta + u$
  3. $y = \frac{\mathrm{x}\beta}{\lambda} + u$
- Can we identify the parameters β and λ in these cases? What assumptions we need for x and u?

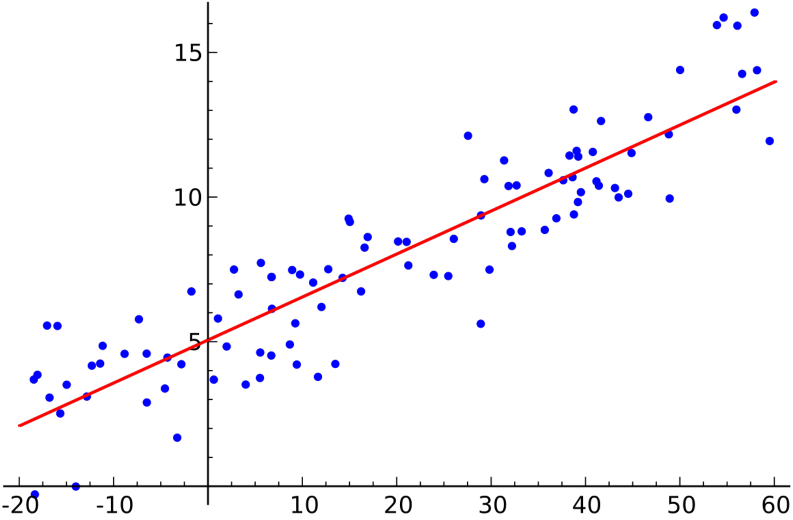
Identification: some examples

With random sampling, identification is purely **a population problem**

- If the model is well specified $\Rightarrow$ it will have similar characteristics to the population when the sample gets larger
- Imagine the following population models:
  1. $y = \mathrm{x}\beta$
  2. $y = \mathrm{x}\beta + u$
  3. $y = \frac{\mathrm{x}\beta}{\lambda} + u$
- Can we identify the parameters β and λ in these cases? What assumptions we need for x and u?

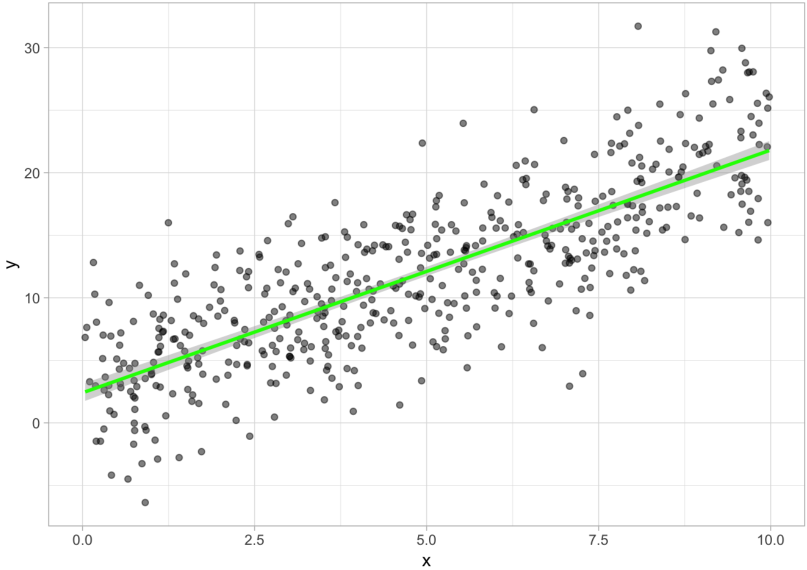
Identification: intuition

Identification Precision (inference)



Identification: intuition

Identification Precision (inference)



The identification problem in the general microeconometrics approach

- Studying $f(y|\mathrm{x};\beta)$ includes two types of unknowns:
  1. What is the function $f()$?
  2. What are the parameters β?
- **Identification problem**: in general we cannot identify both
  1. **Parametric micro-econometrics**
     - Estimate β at the cost of making some assumptions and recovering only some features of $f()$
  2. **Non-parametric micro-econometrics**
     - Recover some features of $f()$ (i.e., an *infinite dimensional feature*)
     - Generally used to look at distributions of a single variable or a relationship between two variables

The identification problem in the general microeconometrics approach

- Studying $f(y|\mathrm{x};\beta)$ includes two types of unknowns:
  1. What is the function $f()$?
  2. What are the parameters β?
- **Identification problem**: in general we cannot identify both
  1. **Parametric micro-econometrics**
     - Estimate β at the cost of making some assumptions and recovering only some features of $f()$
  2. **Non-parametric micro-econometrics**
     - Recover some features of $f()$ (i.e., an *infinite dimensional feature*)
     - Generally used to look at distributions of a single variable or a relationship between two variables

Example: conditional mean of y

We want to estimate $E[y|\mathrm{x}]$ using exogenous covariates x

1. **Parametric**: $E[y|\mathrm{x}] = \alpha + \mathrm{x}\beta$
2. Non-parametric: $E[y|\mathrm{x}] = H(\mathrm{x})$
3. Semi-parametric: $E[y|\mathrm{x}] = G(\alpha + \mathrm{x}\beta)$

Example: conditional mean of y

We want to estimate $E[y|\mathrm{x}]$ using exogenous covariates x

1. **Parametric**: $E[y|\mathrm{x}] = \alpha + \mathrm{x}\beta$
2. **Non-parametric**: $E[y|\mathrm{x}] = H(\mathrm{x})$
3. **Semi-parametric**: $E[y|\mathrm{x}] = G(\alpha + \mathrm{x}\beta)$

Example: conditional mean of y

We want to estimate $E[y|\mathrm{x}]$ using exogenous covariates x

1. **Parametric**: $E[y|\mathrm{x}] = \alpha + \mathrm{x}\beta$
2. **Non-parametric**: $E[y|\mathrm{x}] = H(\mathrm{x})$
3. **Semi-parametric**: $E[y|\mathrm{x}] = G(\alpha + \mathrm{x}\beta)$

Non-parametric methods

1. **Cumulative distribution functions (CDF)**
2. Density (PDF)
3. Conditional means

Empirical CDF

Let y denote a random variable with cdf $F(a) = P(y \leq a)$

$$P(y \leq a) = E\{\mathbb{1}[y \leq a]\}$$

- **Empirical CDF** is the empirical counterpart

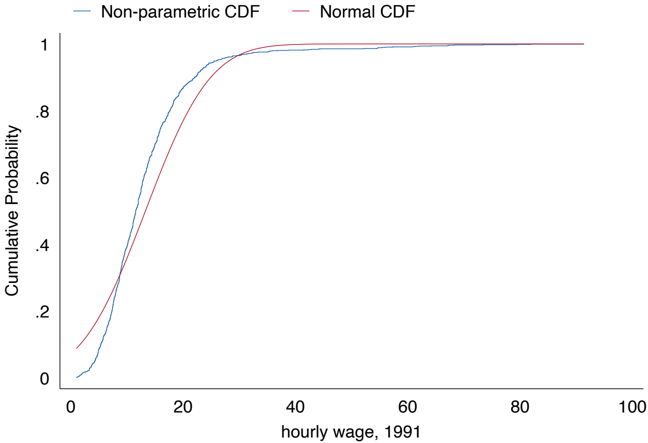
$$\hat{F}(a) = N^{-1} \sum_{i=1}^{N} \mathbb{1}[y_i \leq a]$$

  - Simply the fraction of observations that are $\leq a$
  - $\hat{F}(\cdot)$ is a step function where each step is at a value y_i observed in the sample

APPLICATION: empirical CDF of wage

- Compare empirical CDF with normal distribution

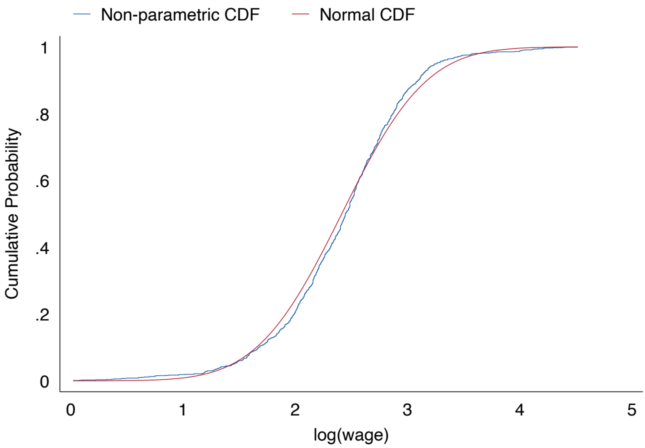
```
use htv.dta  cdfplot wage
```



APPLICATION: empirical CDF of log-wage

- Compare empirical CDF with normal distribution

```
cdfplot lwage
```



Some useful non-parametric methods

1. Cumulative distribution functions (CDF)
2. **Density (PDF)**
   - Centered histogram
   - Density estimation
3. Conditional means

Centered histogram

Let y_i have a continuous distribution with CDF $F(\cdot)$

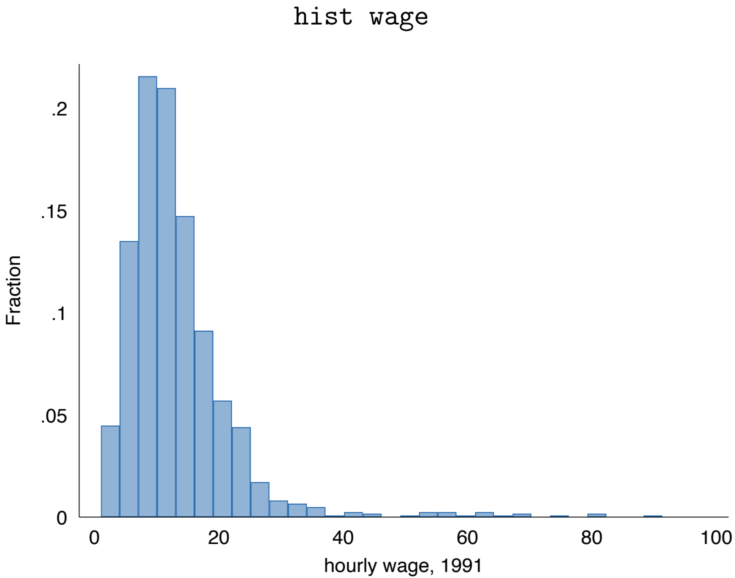
- By definition the PDF = derivative of the CDF

$$f(y) \equiv \frac{dF(y)}{dy} = \lim_{h \to 0} \frac{F(y+h) - F(y-h)}{2h}$$

- **(Centered) histogram**: choose a small **bandwidth** h and plug in the sample CDF in place of F

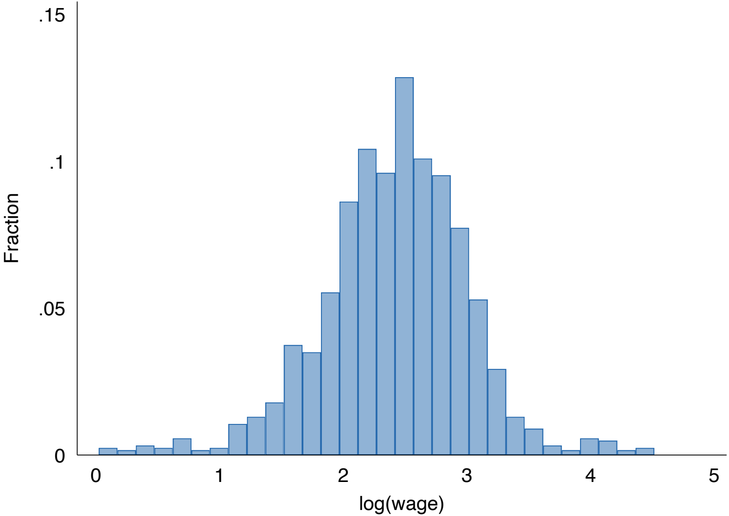
$$\begin{aligned} \hat{f}(y) &= N^{-1} \sum_{i=1}^{N} \frac{1}{2h} (\mathbb{1}[y_i \leq y + h] - \mathbb{1}[y_i \leq y - h]) \\ &= N^{-1} \sum_{i=1}^{N} \frac{1}{2h} (\mathbb{1}[\frac{y_i - y}{h} \leq 1] - \mathbb{1}[\frac{y_i - y}{h} \leq -1]) \end{aligned}$$

APPLICATION: wage distribution



APPLICATION: log wage distribution

hist lwage



Density estimation

- Rewrite the centered histogram estimator of $f(x)$ as

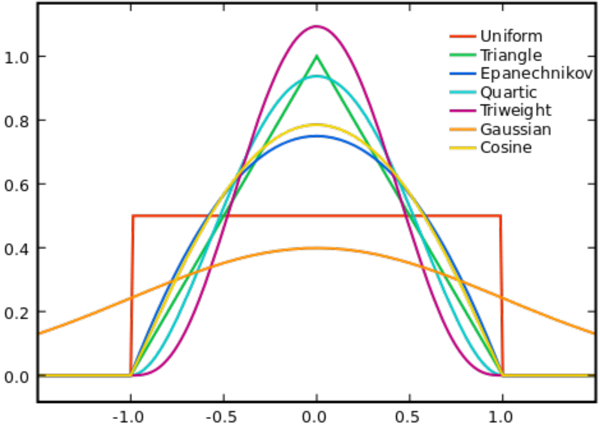
$$\hat{f}(y) = N^{-1} \sum_{i=1}^{N} \frac{1}{h} k \left(\frac{y_i - y}{h} \right)$$

- $k(\cdot)$ is a **kernel function** $\Rightarrow$ weights observations
  - Choose alternative kernel to **increase smoothness**
  - For uniform density on the interval $[-1, 1]$

$$k(u) \equiv \frac{1}{2} \mathbb{1}[-1 < u \leq 1]$$

Choice of kernel: many options available

- $k(\cdot)$ is typically a **symmetric density about zero**



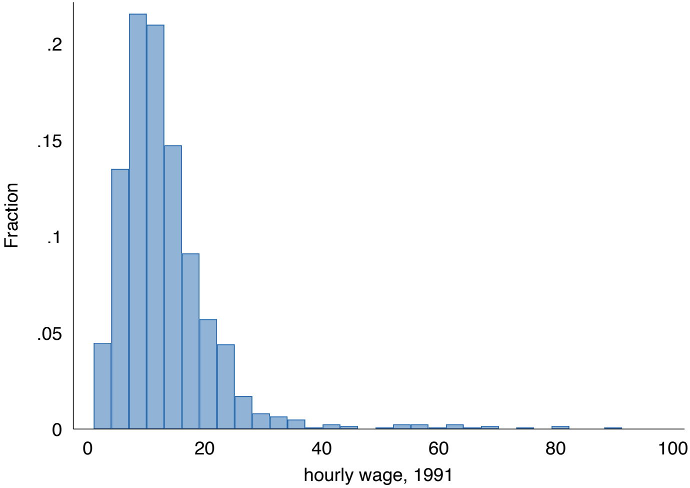
Choice of bandwidth

1. **Guess and experiment**
2. **Rules based on experience or optimality** for common distributions
3. Use data-driven methods, such as **cross validation**
    - Evaluate quality of the bandwidth by looking at how well the resulting estimator forecasts in the given sample

APPLICATION: wage distribution

Histogram Density

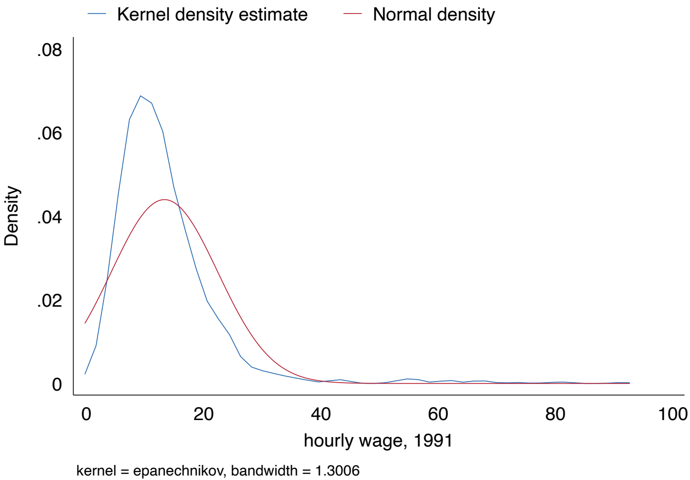
```
hist wage
```



APPLICATION: wage distribution

Histogram **Density**

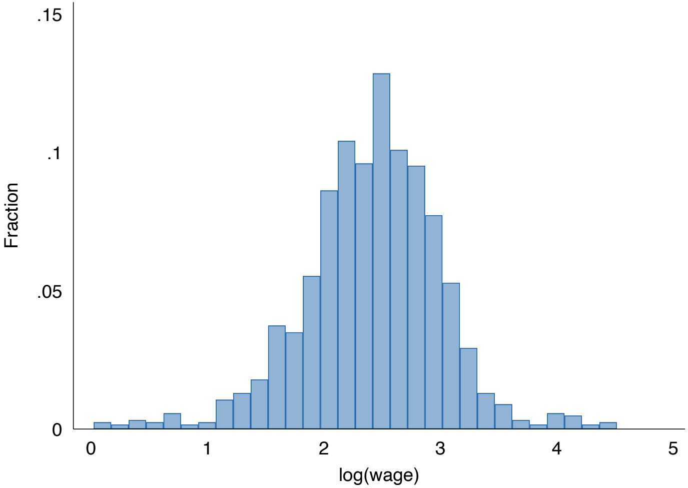
```
kdensity wage, kernel(epan) normal
```



APPLICATION: log-wage distribution

Histogram Density

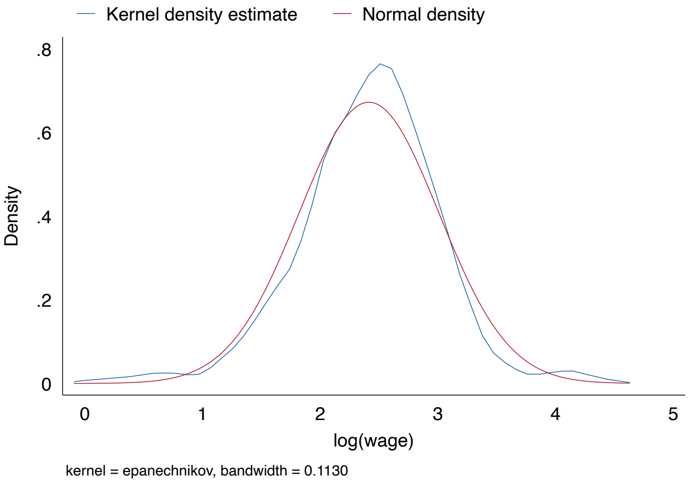
`hist lwage`



APPLICATION: log-wage distribution

Histogram **Density**

```
kdensity lwage, kernel(epan) normal
```



Some useful non-parametric methods

1. Cumulative distribution functions (CDF)
2. Density (PDF)
3. Conditional means
    - **Local smoothing or kernel regression**

Local smoothing (kernel regression)

We want to estimate $m(x) = E(y_i|x_i = x)$ from a random sample (x_i, y_i)

- **Kernel estimators** are weighted averages of the y_i

$$\hat{m}(x) = \frac{\sum_{i=1}^{N} k(\frac{x_i - x}{h}) y_i}{\sum_{i=1}^{N} k(\frac{x_i - x}{h})} \equiv \sum_{i=1}^{N} w_{N,i}(x) y_i$$

- $w_{N,i}(x)$ are weights that give greater weight to x_i closer to x
  - Non-negative and sum to unity
  - You can choose $k(\cdot)$ so that observations far enough away receive zero-weight
    - Epanechnikov, rectangular or triangular

APPLICATION: wage and ability

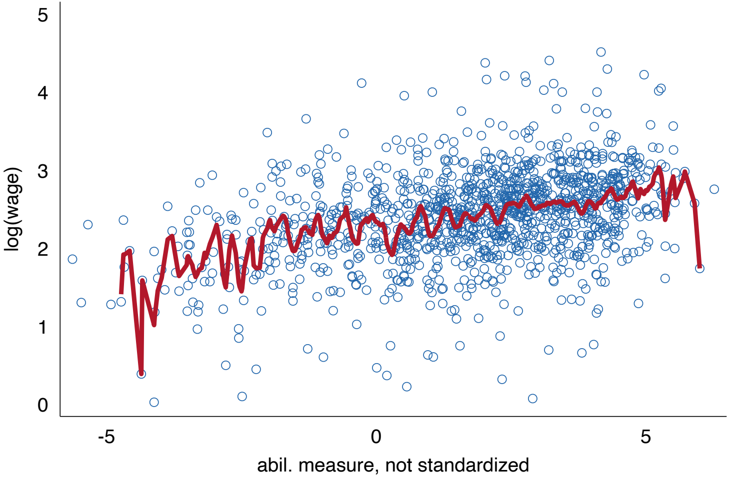
- **Choice of bandwidth** **0.05** 0.20 0.50 2.00

```
npregress kernel lwage abil, bwidth(0.05 0.05, copy)
kernel(epanechnikov)
npgraph
```

```
Local-linear regression                    Number of obs      =      1,225
Kernel   : epanechnikov                    E(Kernel obs)      =         61
Bandwidth: cross validation                R-squared          =     0.2005
--------------------------------------------------------------------------
       lwage |   Estimate
-------------+------------------------------------------------------------
Mean         |
       lwage |   2.415723
-------------+------------------------------------------------------------
Effect       |
        abil |   .0967772
--------------------------------------------------------------------------
Note: Effect estimates are averages of derivatives.
```

APPLICATION: wage and ability

- **Choice of bandwidth** **0.05** 0.20 0.50 2.00



Local-linear estimates
kernel = epanechnikov bandwidth = .05

APPLICATION: wage and ability

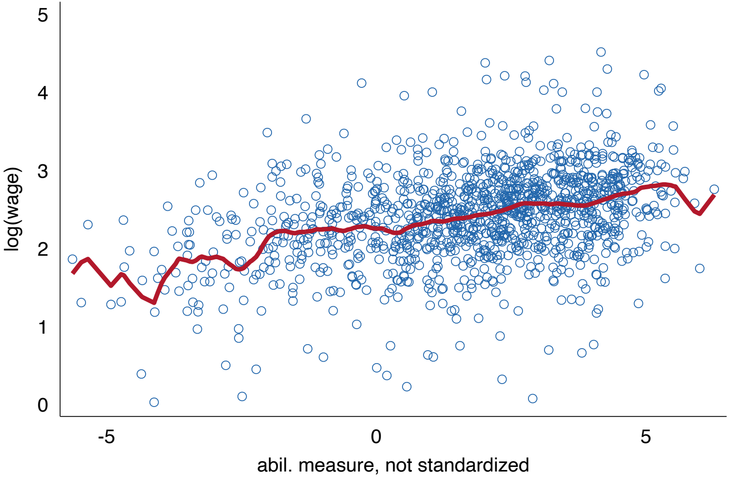
- **Choice of bandwidth** 0.05 **0.20** 0.50 2.00

```
npregress kernel lwage abil, bwidth(0.20 0.20, copy)
kernel(epanechnikov)
npgraph
```

```
Local-linear regression                    Number of obs      =        1,230
Kernel   : epanechnikov                    E(Kernel obs)      =          246
Bandwidth: cross validation                R-squared          =       0.1535
------------------------------------------------------------------------------
       lwage |   Estimate
-------------+----------------------------------------------------------------
Mean         |
       lwage |    2.41346
-------------+----------------------------------------------------------------
Effect       |
        abil |   .1127679
------------------------------------------------------------------------------
Note: Effect estimates are averages of derivatives.
```

APPLICATION: wage and ability

- **Choice of bandwidth** 0.05 **0.20** 0.50 2.00



Local-linear estimates
kernel = epanechnikov bandwidth = .2

APPLICATION: wage and ability

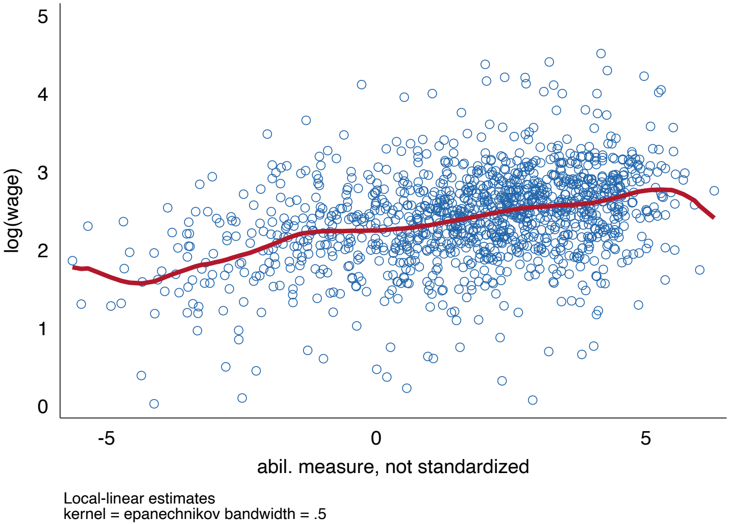
- **Choice of bandwidth** 0.05 0.20 **0.50** 2.00

```
npregress kernel lwage abil, bwidth(0.50 0.50, copy)
kernel(epanechnikov)
npgraph
```

```
Local-linear regression                    Number of obs      =      1,230
Kernel   : epanechnikov                    E(Kernel obs)      =        615
Bandwidth: cross validation                R-squared          =     0.1439
------------------------------------------------------------------------------
       lwage |   Estimate
-------------+----------------------------------------------------------------
Mean         |
       lwage |   2.413593
-------------+----------------------------------------------------------------
Effect       |
        abil |   .1036292
------------------------------------------------------------------------------
Note: Effect estimates are averages of derivatives.
```

APPLICATION: wage and ability

- **Choice of bandwidth** 0.05 0.20 **0.50** 2.00



APPLICATION: wage and ability

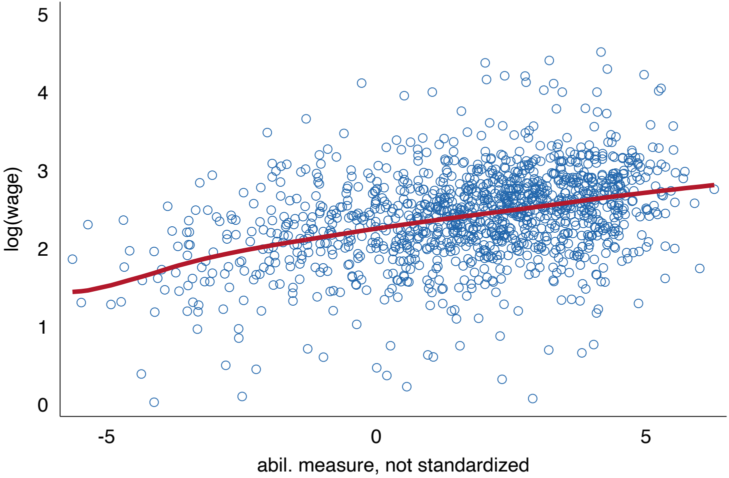
- **Choice of bandwidth** 0.05 0.20 0.50 **2.00**

```
npregress kernel lwage abil, bwidth(2.00 2.00, copy)
kernel(epanechnikov)
npgraph
```

```
Local-linear regression                    Number of obs      =      1,230
Kernel    : epanechnikov                   E(Kernel obs)      =      1,230
Bandwidth: cross validation                R-squared          =     0.1360
------------------------------------------------------------------------------
       lwage |   Estimate
-------------+----------------------------------------------------------------
Mean         |
       lwage |   2.410087
-------------+----------------------------------------------------------------
Effect       |
        abil |   .0962977
------------------------------------------------------------------------------
```

APPLICATION: wage and ability

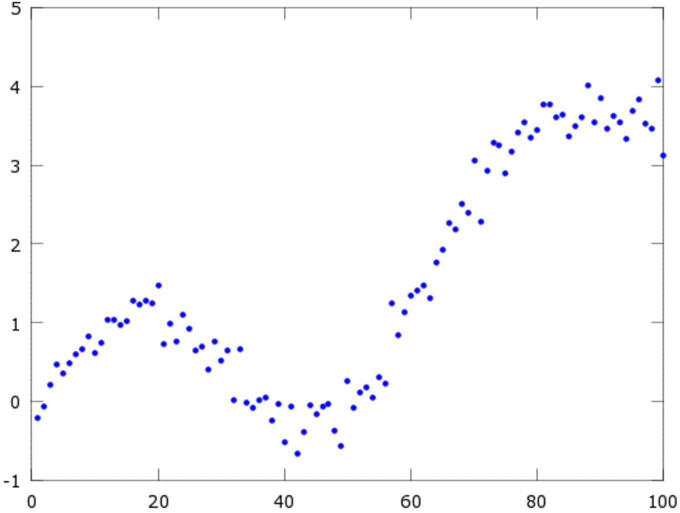
- **Choice of bandwidth** 0.05 0.20 0.50 **2.00**



Local-linear estimates
kernel = epanechnikov bandwidth = 2

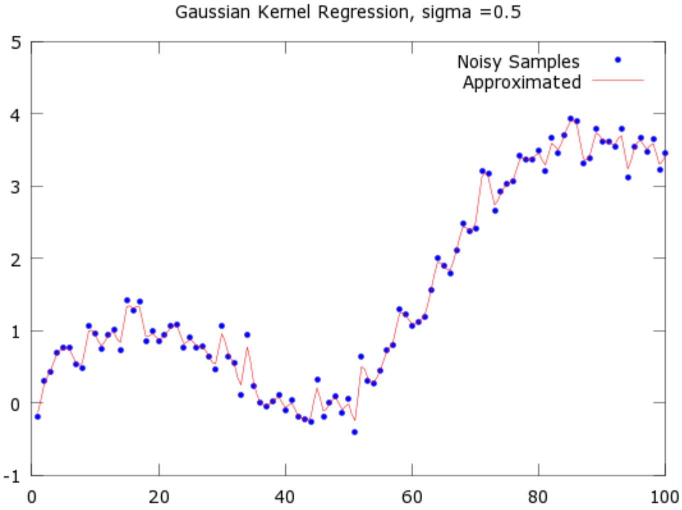
Local smoothing with non-linear relationships

- **Observed data** Kernel w/ bandwidth 0.5 5 50



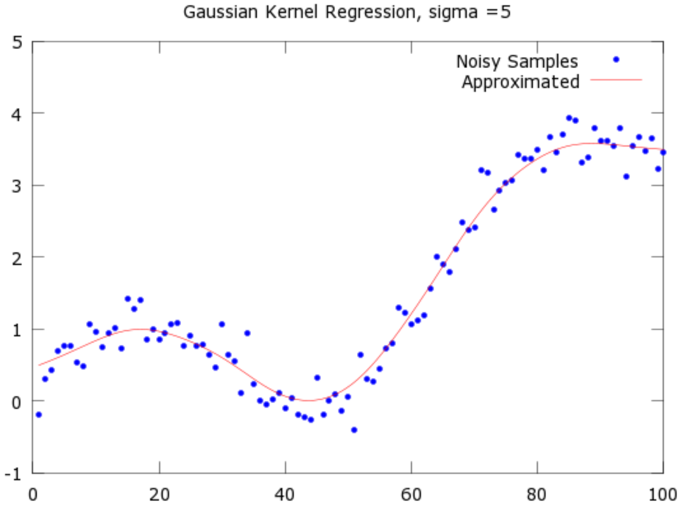
Local smoothing with non-linear relationships

- Observed data **Kernel w/ bandwidth 0.5** 5 50



Local smoothing with non-linear relationships

- Observed data **Kernel w/ bandwidth** 0.5 **5** 50



Local smoothing with non-linear relationships

- Observed data   Kernel w/ bandwidth 0.5 5 50