# Econometrics

Mimi

2024-10-07

# Contents

# About

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports; for example, a math equation $a^2 + b^2 = c^2$.

## 0.1 Usage

Each **bookdown** chapter is an .Rmd file, and each .Rmd file can contain one (and only one) chapter. A chapter *must* start with a first-level heading: `# A good chapter`, and can contain one (and only one) first-level heading.

Use second-level and higher headings within chapters like: `## A short section` or `### An even shorter section`.

The `index.Rmd` file is required, and is also your first book chapter. It will be the homepage when you render the book.

## 0.2 Render book

You can render the HTML version of this example book without changing anything:

1. Find the **Build** pane in the RStudio IDE, and
2. Click on **Build Book**, then select your output format, or select "All formats" if you'd like to use multiple formats from the same book source files.

Or build the book from the R console:

```
bookdown::render_book()
```

To render this example to PDF as a `bookdown::pdf_book`, you'll need to install XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): https://yihui.org/tinytex/.

## 0.3 Preview book

As you work, you may start a local server to live preview this HTML book. This preview will update as you edit the book when you save individual .Rmd files. You can start the server in a work session by using the RStudio add-in "Preview book", or from the R console:

```
bookdown::serve_book()
```

# Chapter 1

# Econometrics

- Econometrics: Using statistical methods for estimating relationships in economics. It can be used to forecast and policy evaluation.
  - It uses empirical data to test assumptions, relationships and theories.

## How to start?

- What is the question of interest?
- An economic model may be needed to test economic theories; they are equations stating the relationship you are looking for. It determines the variables that should be included. Intuition can be also a starting point:
  - The relationship between wages and years of education: $wages = f(y\_educ)$. It is expected that years of education increase wages.
- The econometric model specify the function of the economic model: Wages depend on the years of education and a term u.

$$wages = \beta_0 + \beta_1 y\_educ + u$$

- However, not every factor that affects wages can be observed or measured. The term u represent such unobserved factors, and the measuring error of the included variables. It can never be eliminated entirely. It is also called error term or disturbance

- $\beta_n$ are the parameters. They give information about the relationship between the independent and dependent variable.
- After establishing the econometric model, we can start doing hypothesis about the parameters. Is the relationship positive, negative or zero?
- Then, data is required to estimate the parameters of the established model. There are different types of data structures:

## Cross-Sectional Data

- Sample of variety of units taken in a point in time. Minor timing differences in collecting the data are ignored.
- The order of the observations doesn't matter.
- Random Sampling is needed to get better results. It means the observations drawed are independent. However, it is not always appropriate to make this assumption.
- Random Sampling is violated when population is not large enough, so the observations are not independent draws. It those cases, the same methodology has to be refined.
- In some cases, Random Sampling can be checked if the descriptive statistics are balanced; for example, if we are making a comparison between women and men, the sample should be compound 50% women and 50% men, approximately.
- An example of cross-sectional data using **dplyr** package [Wickham et al., 2023] is shown below

```
## # A tibble: 87 x 14
##    name     height  mass hair_color skin_color eye_color birth_year sex   gender
##    <chr>     <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
##  1 Luke Sk~    172    77 blond      fair       blue            19   male  mascu~
##  2 C-3PO       167    75 <NA>       gold       yellow         112   none  mascu~
##  3 R2-D2        96    32 <NA>       white, bl~ red             33   none  mascu~
##  4 Darth V~    202   136 none       white      yellow          41.9 male  mascu~
##  5 Leia Or~    150    49 brown      light      brown           19   fema~ femin~
##  6 Owen La~    178   120 brown, gr~ light      blue            52   male  mascu~
##  7 Beru Wh~    165    75 brown      light      blue            47   fema~ femin~
##  8 R5-D4        97    32 <NA>       white, red red             NA   none  mascu~
##  9 Biggs D~    183    84 black      light      brown           24   male  mascu~
## 10 Obi-Wan~    182    77 auburn, w~ fair       blue-gray       57   male  mascu~
## # i 77 more rows
## # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>
```

- Some variables can correspond to a different time period, but have a relationship with the dependent variable, so they must be included. That is not going to lead to special problems in the analysis of the data.

## Time series data

- This type of data consist on a variable or plenty variables over time. Past events can influence the future. That is the expected behavior of series like stocks or GDP.

- Unlike in cross-sectional data, time series data order matters: the chronology of events holds important information for the analysis. Even lags can hold useful information.

- Observations rarely can be assumed to be independent across time.

- Cross-section methodologies can be used in time series; however, due to the nature of time series, like trends, modifications have been made in order to study time series.

- Data frequency is one of the characteristics of the data collected in time series. The most common ones are daily, weekly, monthly and annually.

- An example of a time series from the **datasets** package [**?**] is displayed below:

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1949 112 118 132 129 121 135 148 148 136 119 104 118
## 1950 115 126 141 135 125 149 170 170 158 133 114 140
## 1951 145 150 178 163 172 178 199 199 184 162 146 166
## 1952 171 180 193 181 183 218 230 242 209 191 172 194
## 1953 196 196 236 235 229 243 264 272 237 211 180 201
## 1954 204 188 235 227 234 264 302 293 259 229 203 229
## 1955 242 233 267 269 270 315 364 347 312 274 237 278
## 1956 284 277 317 313 318 374 413 405 355 306 271 306
## 1957 315 301 356 348 355 422 465 467 404 347 305 336
## 1958 340 318 362 348 363 435 491 505 404 359 310 337
## 1959 360 342 406 396 420 472 548 559 463 407 362 405
## 1960 417 391 419 461 472 535 622 606 508 461 390 432
```

- Another characteristic of time series is seasonality. It shows effects like the weather.

## Pooled Cross Sections

- Cross-section and time series features.
- For example, a survey conducted in different years to different random samples.
- It helps analyzing the effects of policies, having measured the variables before and after it was implemented.
- The order of the data is not important. However, corresponding year of the information should be tracked.
- The analysis is similar to the cross section data. Although, it is important to account for secular differences in the variables.
- How a key relationship has changed over time?
- Below there is an example of these type of data:

```
##   observation year     price   income
## 1           1 2022 177935.42 73153.77
## 2           2 2022  71976.22 77992.54
## 3           3 2022 103525.42 69557.17
## 4           4 2022  88491.13 69557.17
## 5           5 2022  88491.13 73153.77
## 6           6 2022  71976.22 73153.77
```

```
##     observation year     price   income
## 495         495 2023 103525.42 77992.54
## 496         496 2023  88491.13 73153.77
## 497         497 2023 177935.42 69557.17
## 498         498 2023 177935.42 75404.00
## 499         499 2023 177935.42 69557.17
## 500         500 2023 177935.42 77992.54
```

## Panel or Longitudinal Data

- Every cross-sectional observation is a time series.
- The same individuals are tracked in time, unlike in pooled cross-section data. The same units are included.
- An example is provided using the **gapminder** package [Bryan, 2023]

```
## # A tibble: 6 x 6
##   country     continent  year lifeExp      pop gdpPercap
##   <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
## 1 Afghanistan Asia       1952    28.8  8425333      779.
## 2 Afghanistan Asia       1957    30.3  9240934      821.
## 3 Afghanistan Asia       1962    32.0 10267083      853.
## 4 Afghanistan Asia       1967    34.0 11537966      836.
## 5 Afghanistan Asia       1972    36.1 13079460      740.
## 6 Afghanistan Asia       1977    38.4 14880372      786.
```

- An advantage of these type of data is that having the same observations in time makes it possible to control by unobserved characteristics of the individuals.
  - We are comparing the same individuals over time, so their unobserved characteristics are 'included' in the analysis.
  - It can facilitate causal inference having more than one observation.
  - The behavior of lags can be studied to look for the effects of a policy after a time has passed.

## 1.1

- Without the ceteris paribus assumption, the causal effect that we are looking for is still unknown.
  - Holding the other variables fixed is crucial to determine the link between an independent variable and the dependent variable.

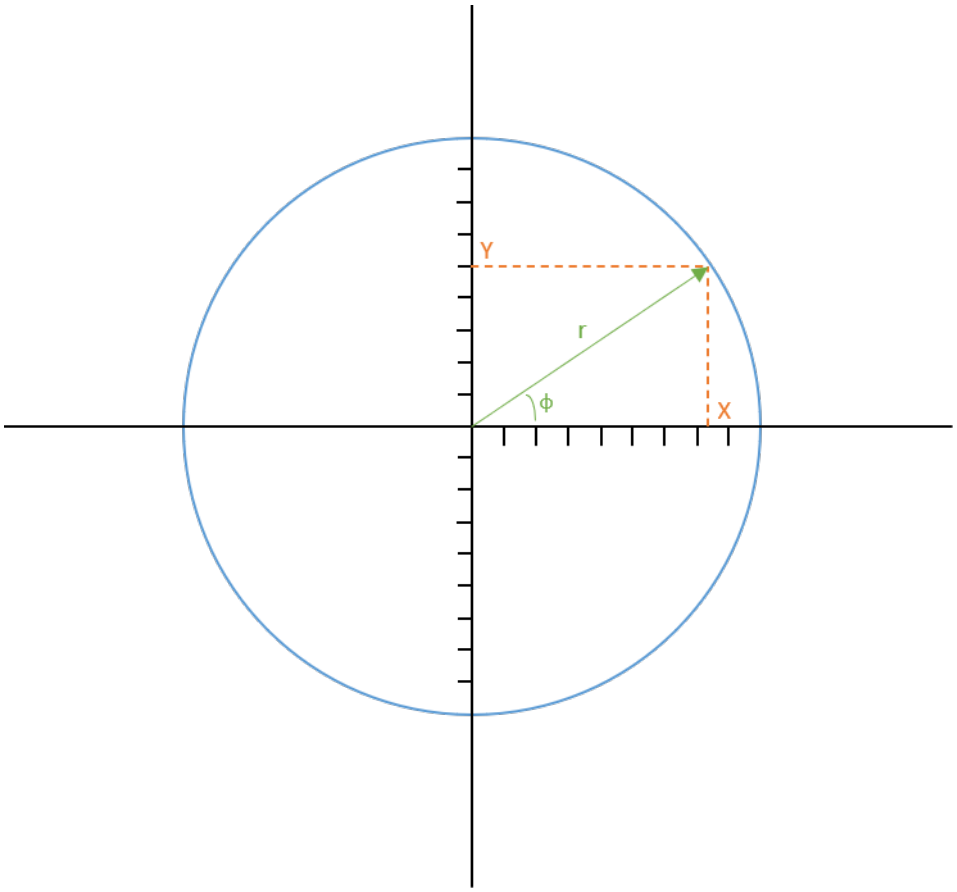
# Chapter 2

# Spatial Econometrics

Chapter 1 gives initial concepts and definitions to start working econometrics. Nonetheless, the constant evolution of methodologies have brought new intriguing issues, like spacial econometrics.

- Spacial data: They have spacial references, like coordinates, longitude and latitude.
- Is it a random process? The variable to study can have spillover effects, where the variable also affects its neighbors. It is not random.
- Geometries are build up with points, that are coordinates in a space from 2 to 4 dimensions:
  - $X$ and $Y$.
  - $Z$ denoting height.
  - $M$ denoting some measurement associated to the point, like the time or the measurement error.
- There are four possible cases:
  - Bidimensional points refer to $X$ and $Y$: east and north (longitude and latitude). $XY$.
  - Tridimensionals: $XYZ$.
  - Tridimensionals: $XYM$.
  - Four dimensions: $XYZM$.

Figure 2.1: Polar and Cartesian Coordinates



# Introduction to Coordinates

## Polar and cartesian coordinates

Figure 2.1 shows

# Exploratory Spacial Data Analysis (ESDA)

- It is another way to call spacial statistics methods.
- ESDA is the starting point to prove if spacial correlation is present.
- Conclusions from the ESDA, the hypothesis of spacial randomness can be rejected, to look for spacial conglomerates. It also helps with the spacial specification of the models.
- There are two types of measurements:
  - Global measurements: Indicator of spacial autocorrelation or general similarity between regions. Its disadvantage is to average.
  - Local measurements: These statistics are determined for each region. There is no generalization of the area. Makes possible to compare between regions if they are similar or different.

## Spacial autocorrelation

- Spacial autocorrelation: Happens when variables are correlated with the location. It means that observations that are closer have more similarities in between than the distant ones. That means that knowing the values of a location can help predicting the near ones.
- The degree of autocorrelation is another concept that can be looked for. It is the distance from where the observations become independent.
- Global autocorrelation can be measured using Moran's I statistics and Geary's C.
- The most popular local measurements are part of the Local Indicator of Spacial Association (LISA). It includes Moran's $I_i$ statistics and Geary's $C_i$ local statistics.
- Global statistics can identify conglomerates and spacial relationships just for all the system, but they can be disaggregated into local statistics to detect local spacial relationship paterns between regions.

### Moran's I

- $H_0$: No spacial autocorrelation; i.e. the values are randomly distributed.
- $I > 0$: Positive spacial autocorrelation. The values to the distance are similar.
- $I < 0$: Negative spacial autocorrelation. The values to the distance are different.
- $I = 0$: The values to the distance are randomly distributed.

```
moran()
moran.test()
moran.mc()
moran.plot()
```

### Geary's C

- $H_0$: No spacial autocorrelation; i.e. the values are randomly distributed.
- $0 < C > 1$: Positive spacial autocorrelation. The values to the distance are similar.
- $1 < C < 2$: Negative spacial autocorrelation. The values to the distance are different.

```
geary()
geary.test()
geary.mc()
```

## Assumptions

- Estationarity. The analized data must be normal distributed with constant mean and variance.

  –

# Chapter 3

# Parts

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART\*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

# Chapter 4

# Footnotes and citations

## 4.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one [1].

## 4.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package [Xie, 2024] (check out the last code chunk in index.Rmd to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** [Xie, 2015] (this citation was added manually in an external file book.bib). Note that the `.bib` files need to be listed in the index.Rmd with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations: https://rstudio.github.io/visual-markdown-editing/#/citations

[1] This is a footnote.

# Chapter 5

# Blocks

## 5.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{5.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (5.1).

## 5.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem 5.1.

**Theorem 5.1.** *For a right triangle, if c denotes the* length *of the hypotenuse and a and b denote the lengths of the* ***other*** *two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html.

## 5.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html

# Chapter 6

# Sharing your book

## 6.1 Publishing

HTML books can be published online, see: https://bookdown.org/yihui/bookdown/publishing.html

## 6.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

## 6.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `gitbook` uses the same social sharing data across all chapters in your book- all links shared will look the same.

Specify your book's source repository on GitHub using the `edit` key under the configuration options in the `_output.yml` file, which allows users to suggest an edit by linking to a chapter's source file.

Read more about the features of this output format here:

https://pkgs.rstudio.com/bookdown/reference/gitbook.html

Or use:

```
?bookdown::gitbook
```

# Bibliography

Jennifer Bryan. *gapminder: Data from Gapminder*, 2023. URL https://CRAN.R-project.org/package=gapminder. R package version 1.0.0.

Hadley Wickham, Romain Francois, Lionel Henry, Kirill Muller, and Davis Vaughan. *dplyr: A Grammar of Data Manipulation*, 2023. URL https://CRAN.R-project.org/package=dplyr. R package version 1.1.4.

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL http://yihui.org/knitr/. ISBN 978-1498716963.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2024. URL https://github.com/rstudio/bookdown. R package version 0.40.