Regression Analysis With R and Easystats

Maulik Bhatt

21/06/2023

Contents

List of Figures

List of Tables

Preface

Acknowledgement

I would like to express my heartfelt gratitude to the individuals who have played a significant role in the creation of this book on Regression Analysis with R. Their guidance, knowledge, and support have been invaluable throughout this journey.

First and foremost, I extend my deepest appreciation to my esteemed professors, [Professor's Name], [Professor's Name], and [Professor's Name]. Their expertise in statistics and econometrics has shaped my understanding of quantitative analysis and provided a solid foundation for this book. Their dedication to teaching and their unwavering commitment to fostering academic excellence have been instrumental in my growth as a student of econometrics.

I would like to extend special thanks to my colleague and dear friend, [Colleague's Name]. Their introduction to the world of R programming language opened up a world of possibilities for me. Their patience, willingness to share their knowledge, and countless hours spent assisting me with R-related challenges have been indispensable. I am grateful for their unwavering support and the collaborative environment we fostered, which significantly enriched my learning experience.

A heartfelt appreciation goes to the developers of R, an open-source programming language that has revolutionized the field of data analysis and statistical computing. Their tireless efforts in creating and continuously improving R have made it an indispensable tool for researchers and analysts worldwide. Without their dedication, this book would not have been possible.

I would also like to express my gratitude to the developers of the R packages that have been instrumental in the analysis and visualization techniques presented in this book. Their commitment to excellence, innovation, and user-friendly implementations has immensely contributed to the field of regression analysis. Their packages have not only expanded the capabilities of R but have also facilitated the seamless integration of econometric methodologies into practical applications.

Finally, I extend my heartfelt thanks to my family and friends who have supported me throughout this writing process. Their encouragement, understanding, and belief in my abilities have been a constant source of motivation.

To all those mentioned above, and to anyone else who has contributed to this book in any way, I offer my deepest appreciation. Your guidance, knowledge, and support have been invaluable in shaping this work and have helped me fulfill my goal of sharing the beauty and importance of regression analysis with R.

Thank you.

[Your Name]

1. Introduction

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Importance of regression analysis in econometrics*
- *Overview of the book's structure and goals*
- *Introduction to R programming language and its relevance to econometric analysis*

Econometrics is a combination of two words: Econo + Metric. Econo refers to concepts of economics, while metric refers to measurement. Let's take an example of the law of demand from microeconomic theory. We know the demand of a commodity will decrease if the price of the commodity increases. But we don't know how much the demand will decrease, given the increase in the price is one unit (i.e., one Dollar, or one Euro, etc.). In Econometrics, we measure such increase or decrease using various experiments. In these experiments, we observe how much the demand for a commodity increase of decrease in response to a unit change (increase or decrease) in price of the commodity.

1.1. Importance of R for Econometrics and Statistics

Many software are available for statistical and econometric analysis. Some of the popular software include SPSS, SAS, Stata, MATLAB, etc. These are all proprietary software. Free and open source alternatives include python and R, which are programming languages. The users need to type the commands in order to perform various tasks/analysis in these programming languages.

Python is a general purpose programming langauge, while R is a statistical programming language. R also has many packages (i.e., add-ons, which enhance the functionality of R). As of writing this, there are `19697` packages available on CRAN (the Comprehensive R Archive Network, which is the central authority to decide about R programming language). Even more packages are available on `github`, which is an Internet hosting service for software development and version control using the version control system Git. Because of such extensive support, I have chosen R to conduct econometric analysis for this book.

In particular, I will use `easystats` set of packages, which are designed for various statistical analysis tasks. We will see how `easystats` make our tasks easier in R.

I will also use `quarto`, which is a document preparation system, based on pandoc's markdown. The benefit of `quarto` is that we can keep our analysis and prose in the same document. At the end, when we click on the `build` button, it renders the entire book in either pdf or HTML format. It is also possible to render the book in Microsoft Word format, but this output format has got limited support. It is more convenient to design and customize the output in HTML and PDF formats. We will discuss more about this in the [**packages appendix at the end of the book**](#sec-packages). See **?@sec-packages** for more details.

2. Introduction to Econometrics and Regression Analysis

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Understanding the basics of econometrics*
- *Role of regression analysis in econometric modeling*
- *Overview of the regression analysis process*
- *Introduction to R for econometric analysis*

Basics of Econometrics

Role of Regression

Introduction to R

3. Getting Started with R for Econometrics

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Introduction to R programming language and its ecosystem*
- *Setting up the R environment and installing necessary packages*
- *Loading and manipulating data in R for econometric analysis*
- *Exploring data visualization techniques using R*

4. Simple Linear Regression

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Understanding the principles of simple linear regression*
- *Performing simple linear regression in R for econometric analysis*
- *Interpreting regression results in the context of economic variables*
- *Assessing model assumptions and addressing violations*
- *Practical examples and exercises using R*

- Understanding the principles of simple linear regression
- Performing simple linear regression in R for econometric analysis
- Interpreting regression results in the context of economic variables
- Assessing model assumptions and addressing violations
- Practical examples and exercises using R

5. Multiple Linear Regression

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Extending regression analysis to multiple independent variables*
- *Building and interpreting multiple linear regression models in R*
- *Handling multicollinearity and selecting significant predictors in an economic context*
- *Model evaluation and diagnostics in econometric regression*
- *Application of multiple linear regression in economic analysis using R*

- Extending regression analysis to multiple independent variables
- Building and interpreting multiple linear regression models in R
- Handling multicollinearity and selecting significant predictors in an economic context
- Model evaluation and diagnostics in econometric regression
- Application of multiple linear regression in economic analysis using R

6. Regression Analysis with Dummy Variables

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Incorporating categorical variables in regression analysis*
- *Creating and interpreting dummy variables in R*
- *Dummy variable pitfalls and remedies in econometric modeling*
- *Examples and case studies of dummy variable regression in economics using R*

- Incorporating categorical variables in regression analysis
- Creating and interpreting dummy variables in R
- Dummy variable pitfalls and remedies in econometric modeling
- Examples and case studies of dummy variable regression in economics using R

7. Heteroscedasticity and Robust Regression

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Understanding heteroscedasticity and its implications*
- *Addressing heteroscedasticity using robust regression techniques in R*
- *Interpreting robust regression results in an economic context*
- *Practical examples and exercises showcasing robust regression in econometrics*

- Understanding heteroscedasticity and its implications
- Addressing heteroscedasticity using robust regression techniques in R
- Interpreting robust regression results in an economic context
- Practical examples and exercises showcasing robust regression in econometrics

8. Time Series Regression

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Introduction to time series data in econometrics*
- *Time series regression models in R for economic analysis*
- *Dealing with autocorrelation and lagged variables*
- *Forecasting with time series regression models in R*
- *Applications of time series regression in economic forecasting*

- Introduction to time series data in econometrics
- Time series regression models in R for economic analysis
- Dealing with autocorrelation and lagged variables
- Forecasting with time series regression models in R
- Applications of time series regression in economic forecasting

9. Introduction to Logistic Regression

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Basics of logistic regression in econometrics*
- *Estimating logistic regression models in R*
- *Interpreting logistic regression coefficients and odds ratios*
- *Applications of logistic regression in economic research using R*

- Basics of logistic regression in econometrics
- Estimating logistic regression models in R
- Interpreting logistic regression coefficients and odds ratios
- Applications of logistic regression in economic research using R

10. Model Evaluation and Selection

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Evaluating model performance and goodness-of-fit measures in econometrics*
- *Validation techniques for econometric regression models*
- *Comparing and selecting models using information criteria*
- *Cross-validation and bootstrapping for robust model assessment in econometrics*

- Evaluating model performance and goodness-of-fit measures in econometrics
- Validation techniques for econometric regression models
- Comparing and selecting models using information criteria
- Cross-validation and bootstrapping for robust model assessment in econometrics

11. Practical Tips and Resources for Econometric Regression

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Data preparation and preprocessing tips for econometric analysis*
- *Handling missing data and outliers in regression analysis*
- *Dealing with endogeneity and instrumental variables*
- *Additional resources for further learning and practice in econometrics with R*

- Data preparation and preprocessing tips for econometric analysis
- Handling missing data and outliers in regression analysis
- Dealing with endogeneity and instrumental variables
- Additional resources for further learning and practice in econometrics with R

12. Conclusion

Status
This chapter is currently a dumping ground for ideas, and we don't recommend reading it.

Learning Objectives:
In this chapter, you will learn to

- *Summary of the key concepts covered in the book*
- *Importance of regression analysis in econometrics and economic research*
- *Encouragement for further exploration and application of econometric regression using R*

- Summary of the key concepts covered in the book
- Importance of regression analysis in econometrics and economic research
- Encouragement for further exploration and application of econometric regression using R

A. Appendix A: R packages for econometric regression analysis and additional resources

This book could never be completed without using many packages. The most notable of the packages include tidyverse, easystats, AER, lmtest, fpp3, gujarati5sie, etc. These packages can be installed using the following commands in R.

```
RA_packages <- c("AER","easystats","fpp3","lmtest","ggthemes","gt","gtsummary","patchwork","
install.packages(RA_packages)
```

The commands above install the packages into your R system. However, the functionality of these packages are not added into your R session just because you installed these packages. You have to load the required packages specifically whenever you need them.

Imagine you are building a home. You have completed electrification in your home. But just because you have a fan or an AC in your home doesn't mean they start automatically. You have to switch on the appliance whenever you need it. Similarly, you will have to load the R packages into your session whenever you need those packages. You can load the package using the `library(package_name)` command. For example, the command to load `easystats` set of packages would be

```
library(easystats)
```

When you run the `install.packages(package_name)` command, R installs the package from Comprehensive R Archive Network (CRAN), which is the highest authority to decide about R. However, there are many packages, which are not available on CRAN. These packages can be downloaded from code sharing platform Github.

To install packages from Github, you need to install one of the three packages first: `remotes`, `devtools` or `pak`. After that, you can easily install packages from Github also. Apart from such packages, the development versions of regular packages (which are available on CRAN) can also be downloaded from Github. For example, if you want to install the package "gujarati5sie", which is a package containing data from the book "Basic Econometrics" written by Damodar Gujarati and others, then you can do it as follows:

```
#Install devtools, remotes or pak
install.packages("remotes")#or install.packages("devltools")#or install.packages("pak")
#Load the package
library(remotes)#or library(devtools) or library(pak)
#download the gujarati5sie package using one of the above packages
remotes::install_github("bhattmaulik/Gujarati5sie")#or devtools::install_github("bhattmaulik
```

Appendix B: Data sets used in the book's examples

Appendix C: How to Build This Book Locally

If you want to build this book locally on your computer, please download the entire code from Github. The first step is to visit the website `www.github.com/bhattmaulik/RegressionAnalysis`. Here, you will find the option to download the entire book in a zip file, as shown below.

```
knitr::include_graphics(here::here("download-book.png"))
```

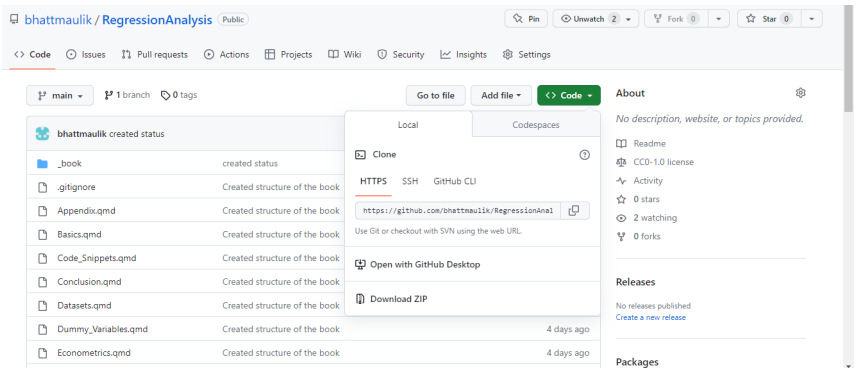

Figure C.1.: Download the book from Github

After you download the zip file, unzip it. This will create a folder in your computer. Within this folder, double click on the "RegressionAnalysis.Rproj" file. This will open the whole project in RStudio.

After opening this project, go to "Build" pane in RStudio. This pane is generally on the top right of RStudio along with Environment, History, Connections, Git and Tutorial. In the build pane, click on "`Render Book`" and select "`HTML format`".

You can also choose to build the book in PDF format, but for that you will need additional software called `tinytex`. In order to build a book in PDF format, R also needs `LaTeX` compiler. There are two popular `LaTeX` compilers: TeX Live and MikTeX. However, they have their own set of problems for R users. The `LaTeX` compiler `tinytex` attempts to solve many of them.

If you want to build the PDF book, you can install `tinytex` through terminal in RStudio using the command `quarto install tinytex`.

```
knitr::include_graphics(here::here("build-book.png"))
```

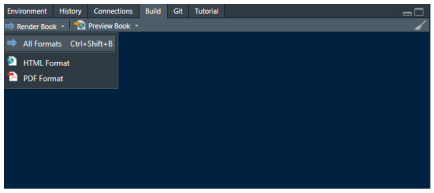

Figure C.2.: Build book using tinytex

If you can't install it this way, or if you want to install it through traditional way using R code, you can use the following code:

```
install.packages("tinytex")
library(tinytex)
tinytex::install_tinytex()
#if you want to install all the LaTeX packages, you can modify the command to tinytex::insta
```

This confuses some readers because there are two `tinytex` in this code: the first tinytex is the `tinytex` R package. The other `tinytex` is the `LaTeX` compiler `tinytex`. So, when we write `library(tinytex)`, we are calling the R package tinytex. And when we use the command `install_tinytex`, we are installing the `LaTeX` compiler `tinytex` using the R package `tinytex`. The benefit of this approach is that you get to select which bundle you want to install. By default, you get to install the bundle `TinyTeX-1`, which contains only the most necessary LaTeX packages. But if you choose the bundle `TinyTeX-2`, you can download all the `LaTeX` packages.