Causal Rosetta

Ahmed H. Darwish

2025-11-21

Table of contents

Preface

The original Rosetta Stone was a breakthrough not because it contained new information, but because it presented the same decree in three different scripts: Hieroglyphs, Demotic, and Ancient Greek. It allowed scholars to finally decipher a language that had been locked away for centuries.

The Causal Rosetta aims to do the same for modern econometrics.

Today, the landscape of causal inference is rich with incredible resources. We have the rigorous theory of *Mostly Harmless Econometrics*, the intuitive storytelling of *The Effect*, and the practical programming of *Causal Inference for the Brave and True*—**all standing in the shadow of the thousand-page encyclopedias by Greene and Wooldridge.** Yet, for a student or practitioner, this abundance can feel overwhelming. Trying to piece together a coherent workflow by jumping between five different textbooks, three coding languages, and endless StackOverflow threads is a recipe for friction, not learning.

As a Teaching Assistant, my job is not to invent new theorems. My job is to translate the "what" into the "how." I see where students get stuck: not on the derivation of the estimator, but on the syntax of the implementation.

This book is a pragmatic synthesis. It is an attempt to blend the wisdom of the giants into a single, accessible manual.

The Philosophy of This Book

- **Synthesis, Not Invention:** I have read the chapters from the competing books so you don't have to. This book aggregates the best explanation for each technique.
- **The "Cheat Sheet" Approach:** We keep the theory light. The heavy mathematical lifting is already done better elsewhere (and I will cite those sources extensively if you wish to dive deeper).
- **Bilingual by Design:** Every concept is implemented side-by-side in **R** and **Python**. Whether you are an econometrics traditionalist or a data science modernist, you belong here.
- **Code-First:** The focus is on the "Do." We start with data and end with a results table.

If the other books are the lectures, think of this book as the lab session.

Enjoy the ride.

1 Introduction

1.1 How Did We Get Here? (Who is Grandpa?)

- **The Origins (The Lab):** Back in the 19th century, researchers in medicine and agriculture invented statistics to solve problems in their domains.
  - *Key Context:* These domains existed inside **laboratories** and controlled fields. If they wanted to know if a fertilizer worked, they simply controlled the environment.
- **The Standardization:** Statistics was not set in stone. In the 20th century, a fierce debate occurred between the giants **Neyman, Pearson, and Fisher**. This solidified the rules of the game (hypothesis testing), resulting in the creation of the first official Statistics Departments.
- **The Ivory Tower:** The graduates of these new departments were experts in math and probability.
  - They began working on abstract theorems and "perfect" distributions.
  - They rarely saw the messy, broken data coming from the real world. This became known as **Mathematical Statistics**.
- **The Wild:** Later on, economists realized they needed statistics too.
  - *The Problem:* They worked in the wild, not the lab. They could not "control" the economy or people's choices.
- **The Birth of Econometrics:** Economists used these tools to **estimate** the magnitude of relationships (e.g., supply and demand) and **test** economic theories.
  - *The Flaw:* Estimation wasn't enough. In the real world, thousands of variables move together. You couldn't isolate the true effect of a policy because you couldn't hold everything else constant.
- **The Distinction:**
  - **Mathematical Statistics:** They invent the math first. They assume a perfect world, derive a theorem, and then check if it fits a lab setting.

  - **Econometrics:** They look at the messy data first. They realize standard testing is impossible, so they invent new techniques to handle the mess.

- **The Destination (Causal Inference):** Causal Inference is the modern extension of econometrics. It is the specific pursuit of identifying the **exact effect** of a treatment, getting results from the wild as if we are in a lab.

1.1.1 The Analogy: The Disease

To understand the difference, imagine a new virus appears:

- **Mathematical Statistics:** Invents a theoretical disease that follows a perfect bell curve. Invents a cure for that theoretical disease. Waits for a patient to actually catch it.
- **Econometrics:** Inspects the sick patients we have right now. Tries to cure them using available data. realizes the data is biased, so they improve the cure.
- **Causal Inference:** Doing everything the econometrician does, but with a laser focus on proving that *it was the cure* that saved the patient, not just luck.

1.2 What Will We Learn?

- **The Anti-Textbook Approach:**
  - Formal books often gatekeep this knowledge behind rigorous Calculus, Linear Algebra, Numerical Analysis, and Probability Theory.
  - By the time you finish the prerequisites, you've forgotten why you wanted to learn this in the first place.

- **The Toolkit:**
  - Here, we will learn the useful stuff **as we go**.
  - Think of this as a backpack containing only what you need for the journey—no heavy, unnecessary theory.

- **The Blend:**
  - We will blend classical **Econometrics** (the foundation) with modern **Causal Inference** (the design) and **Machine Learning** (the power).

2 Summary

The real treasure was the models we learnt along the pages.

References