

notes

ba_thesis

Frederik Bender Bøeck-Nielsen

7. October 2025

Contents

1. Mastering ‘Metrics – The Path from Cause to Effect (Angrist & Pischke 2015)	1
1.1. Introduction	1
2. Chapter 2: Regression	1
2.1. Make Me a Match, Run Me a Regression	2

1. Mastering ‘Metrics – The Path from Cause to Effect (Angrist & Pischke 2015)

1.1. Introduction

“Economists’ use of data to answer cause-and-effect questions constitutes the field of applied econometrics, known to students and masters alike as *‘metrics*. The tools of the ‘metrics trade are disciplined data analysis, paired with the machinery of statistical inference.” (s. 11).

“As college students majoring in economics, we first learned the

other things equal idea by its Latin name, *ceteris paribus*. Comparisons made under *ceteris paribus* conditions have a causal interpretation.” (s. 13).

“The methods they use – random assignment, regression, instrumental variables, regression discontinuity designs, and differences-in-differences – are the Furious Five of econometric research.” (s. 14)

2. Chapter 2: Regression

“The most basic of these tools is *regression*, which compares treatment and control subjects who have the same observed char-

acteristics. [...]. Regression-based causal inference is predicated on the assumption that when key observed variables have been made equal across treatment and control groups, selection bias from the things we can't see is also mostly eliminated.” (s. 47).

2.1. Make Me a Match, Run Me a Regression

“The regression model in this context is an equation linking the treatment variable to the dependent variable while holding control variables fixed by including them in the model. With only one control variable, A_i , the regression of interest can be written as $Y_i = \alpha + \beta P_i + \gamma A_i + e_i$ (Equation 2.1).

The distinction between the treatment variable, P_i , and the control variable, A_i , in equation (2.1) is conceptual, not formal: there is nothing in equation (2.1) to indicate which is which. Your research question and empirical strategy justify the choice of variables and determine the roles they play. [...] use Greek letters for parameters to distinguish them from the variables in the model. The regression parameters – called *regression coefficients* – are the intercept, α (“alpha”); the causal effect of treatment, β (“beta”); and the effect of being a

group A student, γ (“gamma”).” (s. 57).

“The last component of equation (2.1) is the *residual*, e_i (also called an error term). Residuals are defined as the difference between the observed Y_i , and the fitted values generated by the specific regression model we have in mind. These fitted values are written as: