

FUNDAÇÃO GETULIO VARGAS  
SCHOOL OF APPLIED MATHEMATICS

CAIO F. LINS PEIXOTO

**LINEAR INVERSE PROBLEMS AND NONPARAMETRIC  
REGRESSION METHODS IN INSTRUMENTAL VARIABLE  
ESTIMATION**

Rio de Janeiro  
2023

CAIO F. LINS PEIXOTO

**LINEAR INVERSE PROBLEMS AND NONPARAMETRIC  
REGRESSION METHODS IN INSTRUMENTAL VARIABLE  
ESTIMATION**

Bachelor's dissertation presented to  
the School of Applied Mathematics  
(FGV/EMAp) to obtain the Bachelor's  
degree in Applied Mathematics.

Area of Study: Linear Inverse Prob-  
lems, Kernel Methods, Machine Learning,  
Nonparametric Statistics, Instrumental  
Variable Regression.

Advisor: Yuri F. Saporito

Rio de Janeiro

2023

Ficha catalográfica elaborada pela BMHS/FGV

Lins, Caio

Linear Inverse Problems and Nonparametric Regression Methods in Instrumental Variable Estimation/ Caio F. Lins Peixoto. – 2023.  
17f.

Bachelor's Dissertation (Undergraduate) – School of Applied Mathematics.

Advisor: Yuri F. Saporito.  
Includes bibliography.

1. Inverse Problems 2. Kernel Regression 2. Instrumental Variable Regression I. Saporito, Yuri Fahham II. School of Applied Mathematics. III. Linear Inverse Problems and Nonparametric Regression Methods in Instrumental Variable Estimation

CAIO F. LINS PEIXOTO

**LINEAR INVERSE PROBLEMS AND NONPARAMETRIC  
REGRESSION METHODS IN INSTRUMENTAL VARIABLE  
ESTIMATION**

Bachelor's dissertation presented to the School of Applied Mathematics (FGV/EMAp) to obtain the Bachelor's degree in Applied Mathematics.

Area of Study: Linear Inverse Problems, Kernel Methods, Machine Learning, Nonparametric Statistics, Instrumental Variable Regression.

Approved on December —, 2023  
By the organizing committee

---

Yuri F. Saporito  
School of Applied Mathematics

---

Board Member 1  
Institution 1

---

Board Member 2  
Institution 2

I dedicate this thesis to ...

# Acknowledgements

Thanks, ...

*“ Biped! boped! bum! ”*

*Albert Einstein*

# Abstract

Keywords:



# Resumo

Palavras-chave:

# List of Figures

# List of Tables

# Contents

1	INTRODUCTION . . . . .	12
2	APPLICATION IN ECONOMETRICS: INSTRUMENTAL VARIABLE REGRESSION . . . . .	13
2.1	The Role of Instrumental Variables . . . . .	13
3	CONCLUSION . . . . .	15
	References . . . . .	16
	APPENDIX	17

# 1 Introduction

Remember to cite every person ([NEWHEY; POWELL, 2003](#)).

## 2 Application in Econometrics: Instrumental Variable Regression

In this chapter, we turn our attention back to the main application being considered in this thesis: instrumental variable regression. We start by characterizing the type of problem this econometric approach was developed to solve, and then present one of the most well-known and employed estimation procedures for conducting it: two stages least squares (2SLS). Next, we show how an IV regression problem can be formulated as a linear inverse problem and discuss the seminal nonparametric method of Newey and Powell (NEWHEY; POWELL, 2003), followed by a more recent nonparametric approach called Kernel Instrumental Variable (KIV) (SINGH; SAHANI; GRETTON, 2019). We finish with an in-depth analysis of our own method, pointing out its relationship to the others, as well as its strengths and weaknesses.

Is it?

Name of the method.

### 2.1 The Role of Instrumental Variables

Our first goal is to introduce the problem of endogenous covariates. The structural equation we consider is the following:

$$Y = h^*(X) + \varepsilon, \quad (2.1)$$

where  $X$  is a  $d$ -dimensional vector of explanatory variables,  $Y$  is the scalar response,  $\varepsilon$  is a zero mean noise and the function  $h^*$  is the structural parameter we would like to estimate. The simplest estimation method for this model specification — and, therefore, one we would like to be able to use — is ordinary least squares (OLS), which works by finding, within a given class of functions  $\mathcal{H}$ , the element which minimizes the mean squared error:

$$\hat{h} = \arg \min_{h \in \mathcal{H}} \mathbb{E}[(Y - h(X))^2]. \quad (2.2)$$

A reasonable and ample choice for  $\mathcal{H}$  is the set of all square-integrable functions of  $X$ , that is, such that  $\mathbb{E}[h(X)^2] < \infty$ . Under this choice, we recover the conditional expectation of  $Y$  given  $X$ , i.e.,  $\hat{h}(X) = \mathbb{E}[Y | X]$ . Expanding  $Y$  through (2.1), we find that  $\hat{h}(X) = h^*(X) + \mathbb{E}[\varepsilon | X]$ . Hence, if  $\mathbb{E}[\varepsilon | X]$  is not identically null, we have introduced bias in our estimation.

This is one of the problems which appear when  $\mathbb{E}[\varepsilon | X] \neq 0$ , or, more generally, when  $X$  and  $\varepsilon$  are correlated in some way. When this happens, we say that  $X$  is *endogenous*. There are several causes for endogenous covariates, the most common of which are (WOOLDRIDGE, 2001):

Provide one example for each

**Omitted Variables** This means  $\varepsilon$  can be decomposed as  $g^*(W) + \eta$ , where  $\mathbb{E}[\eta \mid X, W] = 0$  a.s. and  $X$  and  $W$  are correlated. Hence, when we don't observe  $W$  and leave it to the error term, we end up estimating  $h^*(X) + \mathbb{E}[g^*(W) \mid X]$ .

**Measurement Error** If we are unable to exactly measure one of the covariates,  $X_k$ , and instead measure  $X'_k$  subject to some stochastic error, by using  $X'_k$  in our regression instead of  $X_k$  we are delegating to  $\varepsilon$  some measure of the difference between  $X_k$  and  $X'_k$ . Depending on how these two variables are related, we may introduce endogeneity.

**Simultaneity** Simultaneity arises when one covariate  $X_k$  is determined simultaneously with  $Y$ . For example, if we are regressing neighborhood murder rates using the size of the local task force as a covariate, there is a simultaneity problem, since larger murder rates in a place cause a larger task force to be allocated there.

Bias in the estimation procedure is only one of the problems which arise when there are endogenous covariates. It's well known that the OLS estimate for linear regression fails to be consistent if any one of the covariates is endogenous ([WOOLDRIDGE, 2001](#)). To overcome endogeneity a few approaches exist, but by far the one most used by empirical economic research is instrumental variable estimation ([WOOLDRIDGE, 2001](#)).

### 3 Conclusion



# References

NEWHEY, Whitney K.; POWELL, James L. Instrumental Variable Estimation of Nonparametric Models. **Econometrica**, v. 71, n. 5, p. 1565–1578, 2003. DOI: <http://dx.doi.org/10.1111/1468-0262.00459>.

SINGH, Rahul; SAHANI, Maneesh; GRETTON, Arthur. Kernel Instrumental Variable Regression. In\_\_\_\_\_. **Advances in Neural Information Processing Systems**. [S.l.]: Curran Associates, Inc., 2019. v. 32. Available from: <[https://proceedings.neurips.cc/paper\\_files/paper/2019/file/17b3c7061788dbe82de5abe9f6fe22b3-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/17b3c7061788dbe82de5abe9f6fe22b3-Paper.pdf)>.

WOOLDRIDGE, Jeffrey M. **Econometric Analysis of Cross Section and Panel Data**. [S.l.]: The MIT Press, 2001. ISBN 9780262232197.

## Appendix