ECON 209
Introduction to Econometrics: Honors
Immigration and Inequality

Casper Neo

15 March 2017

1 Introduction

This article is my final project in ECON 209: Honors Econometrics. It is a partial replication and extension of David Card's *Immigration and Inequality*. Like in Card's paper, I estimate the elasticity of substitution between education and income groups using a constant elasticity of substitution (CES) model and linear regression. I also attempt to estimate log-wages with a linear model and produce coverage intervals for the coefficients. Selective inference and multiple testing corrections will be applied to bound the false discovery rate and false coverage rates in the respective models.

Contents

1 Introduction **1**

2 Original Paper **1**

3 Data Overview **2**

4 Multiple Testing Concerns and Selective Inference **3**

5 Elasticity of Substitution **3**

6 Log-Linear Wage Model **3**

7 Conclusion **3**

8 Appendix: My code **3**

2 Original Paper

Immigration and Inequality by David Card was published in 2009 in the American Economic Review. It explores the relationship between immigration status, education, and wages. Card uses the 1980-2000 Census data combined with the 2005-2006 American Community Survey data. In contrast I use the

2007-2015 American Community Survey. Card explores his hypothesis with time series and panel data as well. I select from his paper the following hypothesis to test and replicate.

1. Workers with below high school education are perfect substitutes for those with a high school education.
2. High school-equivalent and college-equivalent workers are imperfect substitutes with an elasticity of substitution between 1.5-2.5;
3. Within education groups, immigrants and natives are imperfect substitutes with an elasticity of substitution on the order of 20

As a main criticism which I will attempt to address is the fact that Card explores multiple hypothesis and does not address the fact that potentially many more hypothesis were considered (implicitly or otherwise) but not included in his paper. For example he may have checked the elasticity of substitution between immigrants and natives, grouped by race, geography or decade of entry.

3 Data Overview

Data from 'https://www.census.gov/programs-surveys/acs/data/pums.html'.

I download the 1-year ACS surveys from 2007-2015 however only few columns are used. I identified the following variables as relevant from the *PUMS Data Dictionary 2011-2015*. I map these variables from the values in the raw data into units usable by my models. Between the 8 year there are 27'725'196 observations. The procedure and models are developed on 10'000 random rows then run on the remaining rows. This is to ensure the hypothesis I choose (a function of the data I develop the procedure from) is independent of the data I use to validate the hypothesis (with p-values and confidence intervals).

Table 1: Data Used

Variable Key	Description
WAGP	Wages or salary income past 12 months
CIT	Citizenship status
NATIVITY	Native or Foreign Born
AGEP	Age
SCHL	Years of Schooling
ESR	Employment Status Recode
DECADE	Decade of entry into the United States
ST	State

4 Multiple Testing Concerns and Selective Inference

5 Elasticity of Substitution

6 Log-Linear Wage Model

7 Conclusion

8 Appendix: My code

Github link: https://github.com/CasperN/Immigration-and-Inequality.git

To download and clean the data, run in a unix terminal *bash getdata.sh*. Note *cleanData.py* should be in the same directory when this happens. This will automatically download and clean the data as I use it, and randomly split the data into a set for hypothesis generation and another for testing.

List of Figures

List of Tables

1 Data Used 2