Assignment 3 Report

By: Pujan Thakrar and Mark Vandre

Introduction

For this project, we are interested in the impact of immigration on labor-market outcomes. More specifically, we wanted to determine a causal effect of an immigration shock to native labor market outcomes, such as average wage, unemployment rate, labor force participation rate, and the share of employment in manufacturing. This could be used to determine whether or not a large inflow of immigration is more likely to hurt native citizens of a country or if they are more likely to be better off.

PART 1: Constructing Outcomes and Immigration Shock

In this section, we will construct the immigration shock and different native outcome variables by commuting zone.

Assumptions & Variables

We collected 1980 Census and 2007 3-year ACS data from IPUMS. The variables of interest in this analysis were the following: year, person weight, age, citizen, income wage, employment status, state FIPS, PUMA, CNTY98, and industry. We also used GDP deflator data that was provided to us as well as commuting zone data from David Dorn's website.

We then removed any missing data or data that was out of range. In this case, we only considered individuals in the census who were earning an income. For those with a reported income of $75000 in 1980, we applied the top-coding method by multiplying their income by 1.5. In addition to top-coding, we deflated income by dividing by a GDP deflator, which was 42.274 and 92.4825 for the years 1980 and 2007 respectively.

Crosswalk data from David Dorn's data was used to add commuting zones to the data. For 1980, we merge the data by creating a 5-digit code. The first two digits are the state FIPS and the last three digits are the CNTY98 code. For 2007 we do the same thing, however instead of a 5-digit code, we create a 6-digit code, where the first two digits are the state FIPS and the last four digits are the PUMA code. After merging in the commuting zone codes, we also have the 'afactor' which represents the share of the PERWT in that particular commuting zone. Therefore, by multiplying PERWT by the 'afactor', we get the number of people the observation represents for a particular commuting zone. We named this variable 'weight'.

Finally, since we are looking at native and immigrant relationships, we constructed a 'native' variable based on the citizenship status. In this case, we considered all individuals who are either born in America, born abroad of American parents, or naturalized citizens to be considered natives. All other individuals in the census are considered immigrants for the purposes of our analysis.

Constructing Outcomes

To determine the impact of an immigration shock on native outcomes, we picked outcomes that we believed would be the most impacted, namely, native average wages, native unemployment rate, native labor force participation rate, and share of native unemployment in manufacturing.

First, we construct average wage, where we take the total income at the commuting zone – time level for natives and divide by total hours for the respective commuting zone – time bin for natives.

For labor for participation and unemployment rates we sum up the number of unemployed, employed and working age (16 years or older) natives in our sample and aggregate these variables to the commuting zone – year level. This means that we are left with one observation for each commuting zone – time pair. We construct the unemployment rate by taking the number of natives that are unemployed and dividing them by the total number of natives who are unemployed or employed. For the labor force participation rate, we take the total number of natives who are either employed or unemployed, and divide them by the total number of natives who are of working age.

Finally, we follow a very similar process to construct the share of native employment in manufacturing. We first sum up the total number of native individuals working in manufacturing and aggregate to the commuting zone – year level. We then divide the total number of native individuals working in manufacturing in a given commuting zone and at a given time by the total number of employed, native individuals in the same commuting zone and same time period.

Constructing Immigrant Inflow shock

To construct our shock, we use the following equation:

$$x_c = \frac{1}{N_{c,1980}}\left(I_{c,2007} - I_{c,1980}\right)$$

This is the number of new immigrants since 1980 (immigrant inflow given by $I_{c,2007} - I_{c,1980}$) as a share of the native population in 1980 for a given commuting zone ($N_{c,1980}$).

PART 2: Constructing the "Card Instrument"

A concern that one could have regarding the effect of immigrant inflow on native economic outcomes is the probability of endogeneity. In examining the effect of immigrant flows, one could argue as to whether it is immigrant flows causing native outcomes or native outcomes that are causing immigrant flows. In order to address this, we construct a variation of the "Card Instrument."

This instrument is created using by taking the difference in immigrant population within a commuting zone from 1980 to 2007, dividing this difference by the native population within a commuting zone in 1980, and multiplying the result by the share of all immigrants in each commuting zone in 1980. A descriptive equation of the instrument is shown below:

$$z_c = \frac{\sum_s f_{cs,1980}\left(I_{s,2007} - I_{s,1980}\right)}{N_{c,1980}}$$

where: $$f_{cs,1980} = \frac{I_{sc,1980}}{I_{s,1980}}$$

PART 3: Estimation of Immigrant Inflow Effect on Outcomes of Interest

After constructing the instrument, we estimated the following first and second state regressions:

$$\hat{x}_c = \alpha + \eta z_c + \gamma X_c + v_c \tag{1}$$

$$\Delta y_c = \alpha + \beta \hat{x}_c + \gamma X_c + \epsilon_c \tag{2}$$

Equation (1) above shows the form used for the first stage regression where $\hat{x}_c$ is the endogenous immigrant inflow variable to be instrumented and z_c is the instrument described in part 3. Equation (2) shows the second stage regression where Δy is the change in any of our outcome variables of interest and $\hat{x}_c$ is the instrumented endogenous variable from equation (1). In the following tables below, we show the results of our estimated models.

Table 1: First-State Regression

Variables	OLS
	Immigrant Inflow
Card Instrument	0.56
	(0.11)
Constant	-0.004
	(0.005)
Adj R-Squared	0.514
F-Stat	25.28
Observations	24

As we can see, the instrument is significant and our first stage regression has an F-statistic above the common threshold of 10. This suggests that our instrument is relevant.

Next, we show the regression table for the composition-adjusted difference in log average native wages, native unemployment rate, native labor force participation rate, share of native employment in manufacturing for the native population regressed on our instrumented immigrant inflow variable. From these tables we can see that the effect of immigrant inflow has a positive effect of the change in log average wages that is statistically significant in an OLS setting. We can also see that the effect of immigrant inflow on the share of natives employed in manufacturing is negative and statistically significant in an OLS setting. When applying a two stage least squares approach, however, the effect of immigrant inflow becomes statistically significant. If we were to assume significance in the coefficients, the effect of immigrant inflow would be slightly positive for wages, negative for unemployment, positive for labor force participation and negative for the share of employment in manufacturing, all for the native population. Setting aside the statistical insignificance, the overall effects seem logical. One would expect an inflow of people generally to a region to lead to an increase in overall economic activity that would manifest itself in benefits to the native population for both employment opportunity and wages. One would also expect that an inflow of people who have yet to acclimate to a region would most likely work jobs that require less technical training such as manufacturing and may be willing to work those jobs at a lower wage than natives.

Table 2: OLS and 2SLS Regression Results

Variables	OLS			
	Δlog(Wages)	ΔUnemployment Rate	ΔLF Participation	Δ Manufacturing Employment
Immigrant Inflow	2.30	0.27	1.18	-1.15
	(0.850)	(1.869)	(0.476)	(0.549)
Constant	0.79	-0.07	0.04	-0.05
	(0.31)	(0.05)	(0.01)	(0.02)
Adj R-Squared	0.047	-0.044	0.182	0.127
Observations	24	24	24	24
Variables	**2SLS**			
	Δlog(Wages)	ΔUnemployment Rate	ΔLF Participation	Δ Manufacturing Employment
Immigrant Inflow	1.84	-2.22	0.85	-1.07
	(1.16)	(2.66)	(0.66)	(0.75)
Constant	0.26	-0.04	0.04	-0.05
	(0.03)	(0.06)	(0.02)	(0.02)
Adj R-Squared	-0.039	-0.129	0.165	0.127
Observations	24	24	24	24

For the sake of completeness, we include a table with our OLS regressions as before with uninstrumented immigrant inflow replaced with our instrument. In all cases, the coefficient on the instrument is statistically insignificant and significantly less economically significant as compared to the uninstrumented inflow variable.

Table 3: OLS Regression with Instrument

Variables	OLS			
	Δlog(Wages)	ΔUnemployment Rate	ΔLF Participation	ΔManufacturing Employment
Instrument	1.03	-1.24	0.48	-0.60
	(0.72)	(1.41)	(0.40)	(0.44)
Constant	0.26	-0.03	0.04	-0.04
	(0.03)	(0.07)	(0.02)	(0.02)

Adj R-Squared	-0.045	-0.009	0.018	0.036
Observations	24	24	24	24

Conclusion

In conclusion, we find that with our OLS regression, immigration inflow has a statistically significant positive relationship with most native outcomes with the exception of employment in manufacturing. More specifically, we find that a unit increase in immigration inflow is associated with an increase in wages and labor force participation rates for natives, while it is associated with a decrease in employment manufacturing. There is not statistically significant effect on the unemployment rate. By implementing the Card instrument to determine a causal relationship between variables, we get that there is actually no causal effect of immigration inflow on any native outcome due to the lack statistical significance.

References

https://www.ipums.org/

http://www.ddorn.net/data.htm

APPENDIX

```
##############
#load in data#
##############

library(ivpack)
library(ggplot2)
library(utils)
library(plyr)
library(haven)
library(dplyr)
library(AER)


#set wd
setwd("C:/Users/Pujan2/Documents/UCLA Course/Quarter C/ECON 424")

#read in ipums data
gz = gzfile("usa_00118.csv.gz")
acs = read.csv(gz, header = T)
save(acs,file="acs.Rdata")

#read in czone data
CZ_1980 <- read_dta("cw_ctygrp1980_czone_corr.dta")
CZ_2000 <- read_dta("cw_puma2000_czone.dta")

#gdp deflator

GDPDEF <- read.csv(file="GDPDEF.csv", header=TRUE, sep=",")
GDPDEF$YEAR = as.numeric(substr(GDPDEF$DATE, 0, 4) )
############
#clean data#
############

#merge cz data
acs$STATEFIP=formatC(acs$STATEFIP,width=2,flag="0")
acs$PUMA=formatC(acs$PUMA,width=4,flag="0")
acs$CNTYGP98=formatC(acs$CNTYGP98,width=3,flag="0")
acs$puma2000=paste(acs$STATEFIP,acs$PUMA,sep = "")
CZ_1980$ctygrp1980= formatC(CZ_1980$ctygrp1980,width=5,flag = "0")
acs$ctygrp1980=paste(acs$STATEFIP,acs$CNTYGP98,sep = "")

acs_2007 = acs[acs$YEAR==2007,]
acs_1980 = acs[acs$YEAR==1980,]

rm(acs)
acs_2007 = merge(acs_2007,CZ_2000,by="puma2000")
acs_1980 = merge(acs_1980,CZ_1980,by="ctygrp1980")
save(acs_2007,file="acs2007.Rdata")
save(acs_1980,file="acs1980.Rdata")
```

```r
acs = rbind(acs_2007,acs_1980)


#rm(acs_2007,acs_1980)

acs$weight = acs$PERWT*acs$afactor
acs=acs[,-c(1,3,4,5,10,15,16)]

#ct-bin
acs$ct_bin = paste(acs$YEAR,acs$czone,sep = "")
acs = acs[acs$INCWAGE!=999998 & acs$INCWAGE!=999999,]

save(acs,file="acs1.RData")

################################
#construct native average wages#
################################

#natives 0,1,2

acs$native =  mapply(function(x){ifelse(x==0|x==1|x==2,1,0)}, acs$CITIZEN)

acs = merge(acs,GDPDEF,by="YEAR")

acs[acs$INCWAGE==75000 & acs$YEAR==1980,"INCWAGE"]=1.5*acs[acs$INCWAGE==75000 & acs$YEAR==1980,"INCWAGE

acs$INCWAGE = acs$INCWAGE/acs$GDPDEF

acs$THRS = acs$UHRSWORK*acs$WKSWORK1


acs$avWage_N = acs$weight*acs$INCWAGE*acs$native

acs$THRS_N = acs$weight*acs$THRS*acs$native

ct_data = aggregate(list(acs$avWage_N,acs$THRS_N),by=list(ct_bin = acs$ct_bin),FUN=sum)

colnames(ct_data) = paste(c("ct_bin","N_avg_wage","N_thrs"))


ct_data$N_avg_wage = ct_data$N_avg_wage/ct_data$N_thrs


save(acs,file="acs2.Rdata")
save(ct_data,file="ct.Rdata")

################################################
#construct native unemployment & lab force rate#
################################################

acs$workingAge = acs$native*acs$weight
acs[acs$AGE<16,"workingAge"]=0
```

```r
acs$UNEMP_N = acs$native*acs$weight
acs[acs$EMPSTAT!=2,"UNEMP_N"]=0

acs$EMP_N = acs$native*acs$weight
acs[acs$EMPSTAT!=1,"EMP_N"]=0


native = aggregate(list(acs$workingAge,acs$UNEMP_N,acs$EMP_N),by=list(bin = acs$ct_bin),FUN=sum)
colnames(native) = paste(c("ct_bin","N_working_age","UNEMP_N","EMP_N"))

native$UR_N = native$UNEMP_N/(native$UNEMP_N+native$EMP_N)
native$LFR_N = (native$UNEMP_N+native$EMP_N)/native$N_working_age

native1 = native[,c(1,5,6)]
ct_data = merge(ct_data,native1,by="ct_bin")

#######################################################
#construct share of native employment in manufacturing#
#######################################################

acs$Manu = acs$native*acs$weight
acs[acs$IND<100 & acs$IND>399 & acs$YEAR==1980,"Manu"]=0
acs[acs$IND<1000 & acs$IND>4000 & acs$YEAR==2007,"Manu"]=0

native2 = aggregate(list(acs$Manu,acs$EMP_N),by=list(bin = acs$ct_bin),FUN=sum)

colnames(native2) = paste(c("ct_bin","Manu","EMP_N"))

native2$N_emp_manu= native2$Manu/native2$EMP_N

native3 = native2[,c(1,4)]

ct_data = merge(ct_data,native3,by="ct_bin")

save(ct_data,file="ct1.Rdata")


##############################################
#construct immigrant inflow by commuting zone#
##############################################

acs$immigrant = mapply(function(x){ifelse(x==1,0,1)}, acs$native)

immigrant = aggregate(acs$immigrant,by=list(bin = acs$ct_bin),FUN=sum)

colnames(immigrant) = paste(c("ct_bin","immigrant_pop"))

immigrant$YEAR = gsub("([0-9][0-9][0-9][0-9]).*","\\1",immigrant$ct_bin)
immigrant$czone = gsub("[0-9][0-9][0-9][0-9](.*)","\\1",immigrant$ct_bin)

immigrant1 = immigrant %>%
  group_by(czone) %>%
  mutate(dImm = immigrant_pop - lag(immigrant_pop))%>%
```

```
  ungroup()

#immigrant1[immigrant1$czone==10301,]
immigrant1 = immigrant1[,c(1,5)]
immigrant1 = immigrant1[!is.na(immigrant1$dImm),]

immigrant1$YEAR = gsub("([0-9][0-9][0-9][0-9]).*","\\1",immigrant1$ct_bin)
immigrant1$czone = gsub("[0-9][0-9][0-9][0-9](.*)","\\1",immigrant1$ct_bin)

immigrant1 = immigrant1[,c(2,4)]
#get natives in 1980

native_sub = acs[acs$YEAR==1980,]

native_sub$native_pop = native_sub$weight*native_sub$native

native4 = aggregate(native_sub$native_pop,by=list(bin = native_sub$czone),FUN=sum)
colnames(native4) = paste(c("czone","native_1980"))


inflow = merge(immigrant1,native4,by="czone")

inflow$inflow = inflow$dImm/inflow$native_1980

save(inflow,file= "inflow.Rdata")

###########################
#construct Card Instrument#
###########################

inflow_national = aggregate(immigrant$immigrant_pop,by=list(bin = immigrant$YEAR),FUN=sum)

inf_nat = inflow_national[2,2]-inflow_national[1,2]


immigrant_sub=acs[acs$YEAR==1980,]

immigrant_sub$immigrant_pop = immigrant_sub$immigrant*immigrant_sub$weight


immigrant2 = aggregate(immigrant_sub$immigrant_pop,by=list(bin = immigrant_sub$czone),FUN=sum)

colnames(immigrant2) = paste(c("czone","immigrant_1980"))

share_imm_cz = immigrant2

share_imm_cz$share = share_imm_cz$immigrant_1980/inflow_national[1,2]

share_imm_cz$numerator = share_imm_cz$share*inf_nat

instrument = inflow[,c(1,3)]

share_numerator = share_imm_cz[,c(1,4)]
```

```r
instrument = merge(instrument,share_numerator,by="czone")

instrument$Z_c = instrument$numerator/instrument$native_1980

save(instrument, file="instrument.RData")


#################
#regression data#
#################

ct_data$N_avg_wage = log(ct_data$N_avg_wage)
ct_data$UR_N = log(ct_data$UR_N)
ct_data$LFR_N = log(ct_data$LFR_N)
ct_data$N_emp_manu = log(ct_data$N_emp_manu)

ct_data$YEAR = gsub("([0-9][0-9][0-9][0-9]).*","\\1",ct_data$ct_bin)
ct_data$czone = gsub("[0-9][0-9][0-9][0-9](.*)","\\1",ct_data$ct_bin)

outcome = ct_data %>%
  group_by(czone) %>%
  mutate(dAVWAGE = N_avg_wage - lag(N_avg_wage),
         dUNEMP = UR_N - lag(UR_N),
         dLFR = LFR_N - lag(LFR_N),
         dManu_emp = N_emp_manu - lag(N_emp_manu))%>%
  ungroup()

outcome1 = outcome[!is.na(outcome$dAVWAGE),]

save(outcome1,file = "outcome1.Rdata")

#inflow$inflow and instrume$Z_ct

reg_data = outcome1[,c(7:11)]

inflow_reg=inflow[,c(1,4)]

reg_data=merge(reg_data,inflow_reg,by="czone")

instrument_reg=instrument[,c(1,4)]

reg_data=merge(reg_data,instrument_reg,by="czone")

save(reg_data,file="reg_data.Rdata")


#############
#Regressions#
#############


reg_Wage = lm(reg_data$dAVWAGE~reg_data$inflow)
summary(reg_Wage)
```

```r
reg_Unemp = lm(reg_data$dUNEMP~reg_data$inflow)
summary(reg_Unemp)

reg_LF = lm(reg_data$dLFR~reg_data$inflow)
summary(reg_LF)

reg_Manu = lm(reg_data$dManu_emp~reg_data$inflow)
summary(reg_Manu)


#first stage

first_stage = lm(reg_data$inflow~reg_data$Z_c)
summary(first_stage)

#ols with instrument


reg_Wage_fs = lm(reg_data$dAVWAGE~reg_data$Z_c)
summary(reg_Wage_fs)

reg_Unemp_fs = lm(reg_data$dUNEMP~reg_data$Z_c)
summary(reg_Unemp_fs)

reg_LF_fs = lm(reg_data$dLFR~reg_data$Z_c)
summary(reg_LF_fs)

reg_Manu_fs = lm(reg_data$dManu_emp~reg_data$Z_c)
summary(reg_Manu_fs)


#2sls

reg_Wage_iv = ivreg(reg_data$dAVWAGE~reg_data$inflow|.-reg_data$inflow+reg_data$Z_c)
summary(reg_Wage_iv)

reg_Unemp_iv = ivreg(reg_data$dUNEMP~reg_data$inflow|.-reg_data$inflow+reg_data$Z_c)
summary(reg_Unemp_iv)

reg_LF_iv = ivreg(reg_data$dLFR~reg_data$inflow|.-reg_data$inflow+reg_data$Z_c)
summary(reg_LF_iv)

reg_Manu_iv = ivreg(reg_data$dManu_emp~reg_data$inflow|.-reg_data$inflow+reg_data$Z_c)
summary(reg_Manu_iv)
```