Lecture 2 - Replicating Mian and Sufi

UCLA - Econ 221 - Fall 2018

François Geerolf

Contents

Introduction

During this lecture, we do some "hands-on" data work based on the following academic articles:

Saiz, Albert. "The Geographic Determinants of Housing Supply." The Quarterly Journal of Economics 125, no. 3 (August 1, 2010): 1253–96.

Mian, Atif, Kamalesh Rao, and Amir Sufi. "Household Balance Sheets, Consumption, and the Economic Slump." The Quarterly Journal of Economics 128, no. 4 (November 1, 2013): 1687–1726.

Mian, Atif, and Amir Sufi. "What Explains the 2007–2009 Drop in Employment?" Econometrica 82, no. 6 (November 1, 2014): 2197–2223.

1 Data

We attempt to replicate the results in the above papers, using R Statistical Software. You should read this introduction to R first. We load the needed packages: `curl` to download, `tidyverse` to manipulate data, `foreign` to import stata datasets, and `AER` to run IV regressions.

```
rm(list = ls())
pklist <- c("curl", "tidyverse", "foreign", "AER")
source("https://fgeerolf.github.io/code/load-packages.R")
```

1.1 Housing Supply elasticity

```
curl_download("http://real.wharton.upenn.edu/~saiz/SUPPLYDATA.zip",
              destfile = "SUPPLYDATA.zip",
              quiet = FALSE)
# Unzip
unzip("SUPPLYDATA.zip")
unlink("SUPPLYDATA.zip")
unlink("readme.txt")
# Read in dta. Needs package foreign
data.Saiz2010 <- read.dta("HOUSING_SUPPLY.dta")
unlink("HOUSING_SUPPLY.dta")
```

1.2 Crosswalk

```
nber.crosswalk <- read.csv("http://www.nber.org/cbsa-msa-fips-ssa-county-crosswalk/cbsatocountycrosswal
```

Let's glance at the data:

```
nber.crosswalk %>%
  select(msa, cbsa) %>%
  head
```

```
##    msa  cbsa
## 1 5240 33860
## 2 5160    NA
## 3    1    NA
## 4    1 13820
## 5 1000 13820
## 6    1    NA
```

1.3 House Prices

```
fhfa.data <- read.csv("https://www.fhfa.gov/DataTools/Downloads/Documents/HPI/HPI_master.csv")

fhfa.data.monthly <-  fhfa.data %>%
  filter(frequency == "monthly") %>%
  mutate(date = yr + (period - 1)/12) %>%
  select(-yr, -period)


fhfa.data.quarterly <-  fhfa.data %>%
  # Keep only quarterly
  filter(frequency == "quarterly",
         hpi_type == "traditional",
         hpi_flavor == "all-transactions") %>%
  # create new date variable collapsing year and quarter
  mutate(date = yr + (period - 1)/4) %>%
  # remove year and quarter
  select(-yr, -period) %>%
  # Keep only MSA geographies
  filter(level == "MSA") %>%
```

```
  select(place_name, place_id, date, index_nsa, index_sa)

fhfa.data.quarterly.extract <- fhfa.data.quarterly %>%
  filter(date %in% c(2006, 2009)) %>%
  rename(value = index_nsa) %>%
  select(-index_sa) %>%
  mutate(date = paste0("year", date)) %>%
  spread(date, value) %>%
  mutate(houseprice_d1ln = log(year2009) - log(year2006))

# Glance at data
fhfa.data.quarterly.extract %>%
  mutate(fips_msa = place_id %>% paste %>% as.numeric) %>%
  select(fips_msa, houseprice_d1ln) %>%
  head
```

```
##   fips_msa houseprice_d1ln
## 1    10180      0.12489273
## 2    10420     -0.05606812
## 3    10580      0.05390098
## 4    10500      0.06902736
## 5    10540      0.09718197
## 6    10740      0.08403987
```

1.4 Unemployment

```
bls.laus.current <- read.delim("https://download.bls.gov/pub/time.series/la/la.data.1.CurrentS")
bls.laus.series <- read.delim("https://download.bls.gov/pub/time.series/la/la.series")
bls.la.area.type <- read.delim("https://download.bls.gov/pub/time.series/la/la.area_type")
bls.la.05.09 <- read.delim("https://download.bls.gov/pub/time.series/la/la.data.0.CurrentU05-09")
bls.laus.area.code <- read.delim("https://download.bls.gov/pub/time.series/la/la.area")

bls.la.05.09.new <- bls.la.05.09 %>%
  filter(period != "M13") %>%
  mutate(month = period %>% substr(2, 3) %>% as.numeric,
         date = year + (month-1)/12) %>%
  select(-month, - year, -period, - footnote_codes) %>%
  left_join(bls.laus.series, by = "series_id") %>%
  filter(area_type_code == "B") %>%
  filter(date %in% c(2006, 2009),
         measure_code == 3) %>%
  mutate(date = paste0("year", date)) %>%
  select(area_code, date, value) %>%
  spread(date, value) %>%
  mutate(year2006 = year2006 %>% paste %>% as.numeric,
         year2009 = year2009 %>% paste %>% as.numeric,
         unemp_d1 = year2009 - year2006) %>%
  mutate(area_code = area_code %>% paste,
         fips_msa = area_code %>% substr(5, 9) %>% as.numeric)
```

```
## Warning: Column `series_id` joining factors with different levels, coercing
## to character vector
```

```
bls.la.05.09.new %>%
  select(fips_msa, unemp_d1) %>%
  head
```

```
##   fips_msa unemp_d1
## 1    11500      4.8
## 2    12220      4.8
## 3    13820      4.5
## 4    19300      4.6
## 5    19460      5.6
## 6    20020      4.7
```

1.5 Employment

Link to data BLS LAUS: https://download.bls.gov/pub/time.series/la

Measures: https://download.bls.gov/pub/time.series/la/la.measure

```
bls.laus.current <- read.delim("https://download.bls.gov/pub/time.series/la/la.data.1.CurrentS")
bls.laus.series <- read.delim("https://download.bls.gov/pub/time.series/la/la.series")
bls.la.area.type <- read.delim("https://download.bls.gov/pub/time.series/la/la.area_type")
bls.la.05.09 <- read.delim("https://download.bls.gov/pub/time.series/la/la.data.0.CurrentU05-09")
bls.laus.area.code <- read.delim("https://download.bls.gov/pub/time.series/la/la.area")

bls.la.05.09.new.emp <- bls.la.05.09 %>%
  filter(period != "M13") %>%
  mutate(month = period %>% substr(2, 3) %>% as.numeric,
         date = year + (month-1)/12) %>%
  select(-month, - year, -period, - footnote_codes) %>%
  left_join(bls.laus.series, by = "series_id") %>%
  filter(area_type_code == "B") %>%
  filter(date %in% c(2006, 2009),
         measure_code == 5) %>%
  mutate(date = paste0("year", date)) %>%
  select(area_code, date, value) %>%
  spread(date, value) %>%
  mutate(year2006 = year2006 %>% paste %>% as.numeric,
         year2009 = year2009 %>% paste %>% as.numeric,
         emp_d1ln = log(year2009) - log(year2006)) %>%
  mutate(area_code = area_code %>% paste,
         fips_msa = area_code %>% substr(5, 9) %>% as.numeric)
```

```
## Warning: Column `series_id` joining factors with different levels, coercing
## to character vector
```

```
bls.la.05.09.new.emp %>%
  select(fips_msa, emp_d1ln) %>%
  head
```

```
##   fips_msa     emp_d1ln
## 1    11500 -0.035867730
## 2    12220 -0.034828067
## 3    13820 -0.048796762
## 4    19300  0.007244928
## 5    19460 -0.048428114
```

```
## 6    20020 -0.060530114
```

1.6 Labor Force

Link to data BLS LAUS: https://download.bls.gov/pub/time.series/la

Measures: https://download.bls.gov/pub/time.series/la/la.measure

```
bls.laus.current <- read.delim("https://download.bls.gov/pub/time.series/la/la.data.1.CurrentS")
bls.laus.series <- read.delim("https://download.bls.gov/pub/time.series/la/la.series")
bls.la.area.type <- read.delim("https://download.bls.gov/pub/time.series/la/la.area_type")
bls.la.05.09 <- read.delim("https://download.bls.gov/pub/time.series/la/la.data.0.CurrentU05-09")
bls.laus.area.code <- read.delim("https://download.bls.gov/pub/time.series/la/la.area")

bls.la.05.09.new.lf <- bls.la.05.09 %>%
  filter(period != "M13") %>%
  mutate(month = period %>% substr(2, 3) %>% as.numeric,
         date = year + (month-1)/12) %>%
  select(-month, - year, -period, - footnote_codes) %>%
  left_join(bls.laus.series, by = "series_id") %>%
  filter(area_type_code == "B") %>%
  filter(date %in% c(2006, 2009),
         measure_code == 6) %>%
  mutate(date = paste0("year", date)) %>%
  select(area_code, date, value) %>%
  spread(date, value) %>%
  mutate(year2006 = year2006 %>% paste %>% as.numeric,
         year2009 = year2009 %>% paste %>% as.numeric,
         lf_d1ln = log(year2009) - log(year2006)) %>%
  mutate(area_code = area_code %>% paste,
         fips_msa = area_code %>% substr(5, 9) %>% as.numeric)
```

```
## Warning: Column `series_id` joining factors with different levels, coercing
## to character vector
```

```
bls.la.05.09.new.emp %>%
  select(fips_msa, emp_d1ln) %>%
  head
```

```
##   fips_msa     emp_d1ln
## 1    11500 -0.035867730
## 2    12220 -0.034828067
## 3    13820 -0.048796762
## 4    19300  0.007244928
## 5    19460 -0.048428114
## 6    20020 -0.060530114
```

1.7 CBP

```
curl_download("https://www2.census.gov/programs-surveys/cbp/datasets/2009/cbp09co.zip",
              destfile = "cbp09co.zip",
              quiet = FALSE)

unzip("cbp09co.zip")
```

```
unlink("cbp09co.zip")

cbp.2009 <- read.delim("cbp09co.txt", sep = ",")
unlink("cbp09co.txt")
```

```
curl_download("https://www2.census.gov/programs-surveys/cbp/datasets/2006/cbp06co.zip",
              destfile = "cbp06co.zip",
              quiet = FALSE)

unzip("cbp06co.zip")
unlink("cbp06co.zip")

cbp.2006 <- read.delim("cbp06co.txt", sep = ",")
unlink("cbp06co.txt")
```

1.8 Merge into Final Data

```
final <- bls.la.05.09.new %>%
  left_join(bls.la.05.09.new.emp, by = "fips_msa") %>%
  left_join(bls.la.05.09.new.lf, by = "fips_msa") %>%
  select(fips_msa, unemp_d1, emp_d1ln, lf_d1ln) %>%
  left_join(fhfa.data.quarterly.extract %>%
              mutate(fips_msa = place_id %>% paste %>% as.numeric) %>%
              select(fips_msa, houseprice_d1ln))
```

```
## Joining, by = "fips_msa"
```

2 Regressions

2.1 OLS

We run:

$$\Delta_{06-09} U_i = \Delta_{06-09} HP_i + \epsilon_i$$

```
final %>%
  lm(unemp_d1 ~ houseprice_d1ln, data = .) %>%
  summary
```

```
##
## Call:
## lm(formula = unemp_d1 ~ houseprice_d1ln, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.4302  -1.1529  -0.1800   0.9535  11.5579
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.27994    0.09661   33.95   <2e-16 ***
## houseprice_d1ln -6.61945    0.55298  -11.97   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.822 on 355 degrees of freedom
##   (38 observations deleted due to missingness)
## Multiple R-squared:  0.2876, Adjusted R-squared:  0.2856
## F-statistic: 143.3 on 1 and 355 DF,  p-value: < 2.2e-16
```

```
final %>%
  lm(emp_d1ln ~ houseprice_d1ln, data = .) %>%
  summary
```

```
##
## Call:
## lm(formula = emp_d1ln ~ houseprice_d1ln, data = .)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.273318 -0.023605  0.001109  0.026029  0.131301
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.001993   0.002330  -0.855    0.393
## houseprice_d1ln  0.069512   0.013336   5.212 3.17e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04393 on 355 degrees of freedom
##   (38 observations deleted due to missingness)
## Multiple R-squared:  0.07109,    Adjusted R-squared:  0.06847
## F-statistic: 27.17 on 1 and 355 DF,  p-value: 3.172e-07
```

```
final %>%
  lm(lf_d1ln ~ houseprice_d1ln, data = .) %>%
  summary
```

```
##
## Call:
## lm(formula = lf_d1ln ~ houseprice_d1ln, data = .)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.140662 -0.021815 -0.000865  0.022033  0.134941
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.033499   0.001888  17.739   <2e-16 ***
## houseprice_d1ln -0.005143   0.010810  -0.476    0.635
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03561 on 355 degrees of freedom
##   (38 observations deleted due to missingness)
## Multiple R-squared:  0.0006372,  Adjusted R-squared:  -0.002178
## F-statistic: 0.2263 on 1 and 355 DF,  p-value: 0.6345
```

2.2 Matching

Final is fips_msa = 10180 for Abilene, TX

Crosswalk is cbsa = fips_msa

```
nber.crosswalk %>%
  select(fips_msa = cbsa, msa) %>%
  head
```

```
##   fips_msa  msa
## 1    33860 5240
## 2       NA 5160
## 3       NA    1
## 4    13820    1
## 5    13820 1000
## 6       NA    1
```

```
data.Saiz2010 %>%
  rename(msa = msanecma) %>%
  head
```

```
##   msa population                          msaname      WRLURI
## 1  40     126441                 Abilene, TX (MSA)  0.10083067
## 2  80     695781                  Akron, OH (PMSA)  0.06756011
## 3 120     120855                  Albany, GA (MSA) -0.49810675
## 4 160     876129 Albany-Schenectady-Troy, NY (MSA) -0.09178428
## 5 200     714554             Albuquerque, NM (MSA)  0.37347099
## 6 220     126443              Alexandria, LA (MSA) -1.67668915
##   FLAT_SHARE_50_15 S_LAND_50        lu11         lu91         lu92
## 1               99       100 0.008205139 5.833476e-05 0.0012621708
## 2               99       100 0.018677070 2.790428e-02 0.0079509728
## 3              100       100 0.013510201 1.063221e-01 0.0074655255
## 4               81       100 0.021399631 1.864739e-02 0.0032548599
## 5               89       100 0.003276727 2.593219e-03 0.0004627457
## 6              100       100 0.039517049 1.427566e-01 0.0097358040
##       unaval elasticity
## 1 0.01952564   3.094851
## 2 0.06453232   2.585493
## 3 0.12729786   3.392287
## 4 0.23330188   1.700388
## 5 0.11633269   2.112380
## 6 0.19200943   7.148086
```

2.3 IV Regressions

```
final %>%
  left_join(nber.crosswalk %>%
              select(fips_msa = cbsa, msa),
            by = "fips_msa") %>%
  left_join(data.Saiz2010 %>%
              select(msa = msanecma, elasticity)) %>%
  ivreg(unemp_d1 ~ houseprice_d1ln  | elasticity, data = .) %>%
  summary
```

```
## Joining, by = "msa"
```

```
##
## Call:
## ivreg(formula = unemp_d1 ~ houseprice_d1ln | elasticity, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5619  -0.8819  -0.1355   0.6864  11.5840
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.16272    0.06899  45.844   <2e-16 ***
## houseprice_d1ln -4.08364    1.94671  -2.098   0.0363 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.676 on 603 degrees of freedom
## Multiple R-Squared: 0.1815,  Adjusted R-squared: 0.1801
## Wald test:   4.4 on 1 and 603 DF,  p-value: 0.03635
```

```
final %>%
  left_join(nber.crosswalk %>%
              select(fips_msa = cbsa, msa),
            by = "fips_msa") %>%
  left_join(data.Saiz2010 %>%
              select(msa = msanecma, elasticity)) %>%
  lm(houseprice_d1ln ~ elasticity, data = .) %>%
  summary
```

```
## Joining, by = "msa"
```

```
##
## Call:
## lm(formula = houseprice_d1ln ~ elasticity, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87417 -0.03136  0.02299  0.07690  0.21724
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.065042   0.012048  -5.398 9.68e-08 ***
## elasticity   0.027851   0.004256   6.543 1.29e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1315 on 603 degrees of freedom
##   (460 observations deleted due to missingness)
## Multiple R-squared:  0.0663, Adjusted R-squared:  0.06475
## F-statistic: 42.82 on 1 and 603 DF,  p-value: 1.286e-10
```

```
final %>%
  left_join(nber.crosswalk %>%
              select(fips_msa = cbsa, msa),
            by = "fips_msa") %>%
```

```
  left_join(data.Saiz2010 %>%
              select(msa = msanecma, elasticity)) %>%
  select(elasticity) %>%
  summary
```

```
## Joining, by = "msa"
```

```
##    elasticity
##  Min.   : 0.6728
##  1st Qu.: 1.6293
##  Median : 2.4610
##  Mean   : 2.5492
##  3rd Qu.: 3.1189
##  Max.   :12.1480
##  NA's   :446
```

```
final %>%
  left_join(nber.crosswalk %>%
              select(fips_msa = cbsa, msa),
            by = "fips_msa") %>%
  left_join(data.Saiz2010 %>%
              select(msa = msanecma, elasticity)) %>%
  lm(unemp_d1 ~ houseprice_d1ln, data = .) %>%
  summary
```

```
## Joining, by = "msa"
```

```
##
## Call:
## lm(formula = unemp_d1 ~ houseprice_d1ln, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.3229  -0.9963  -0.0805   0.8672  11.6555
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.18794    0.05298   60.17   <2e-16 ***
## houseprice_d1ln -6.77534    0.40804  -16.60   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.639 on 970 degrees of freedom
##   (93 observations deleted due to missingness)
## Multiple R-squared:  0.2213, Adjusted R-squared:  0.2205
## F-statistic: 275.7 on 1 and 970 DF,  p-value: < 2.2e-16
```