# Local Institutions in Economic Development: The Effect of Juntas Auxiliares in Puebla, Mexico

Etienne Ricardez

7/May/2024

## Introduction and context

This is the code I developed for an Economic History course project within the Master's program in Economics at the University of British Columbia. It evaluates the impact of a local institution known as 'Juntas Auxiliares' on the economic development at the municipalities level in the state of Puebla, Mexico. Below is a brief walkthrough of the code used for this project. You can access the full article here: https://drive.google.com/file/d/1q-JoRTzJ4x7ccBQfr_Q87milcnh-A-is/view?usp=drive_link

## Abstract

This study investigates the effect of Juntas Auxiliares, a local governance institutions unique to the state of Puebla, Mexico, on the economic development at the municipality level. I gather information of the number of Juntas Auxiliares in 172 municipalities of Puebla and used administrative public data to measure the impact of Juntas Auxiliares on 2020 household income levels. Using territorial extension of each municipality as the instrument for the number of Juntas Auxiliares to solve for endogeneity and, defending the instrument with 31 falsification tests, I found that Juntas Auxiliares are significantly associated with a 2% increase in household income, indicating their positive contribution to local economic development

## Data

I compiled in one data set the following pieces of information:

- Estimates of municipal income per household, which were available through the National Institute of Statistics and Geography (INEGI). Available in: https://en.www.inegi.org.mx/investigacion/icmm/2020/
- Territoral extension of each municipality in Mexico. Available in: https://www.inegi.org.mx/programas/ccpv/2020/tableros/panorama/
- Latitud, Longitud and Altitude of each municipality in Mexico. Available in (go to "Catálogos Predefinidos"): https://www.inegi.org.mx/app/ageeml/
- Estimates of the indigenous population percentage by municipality in Mexico. Available in: https://www.inpi.gob.mx/indicadores2020/
- And, for the number of Juntas Auxiliares in each municipality of Puebla, I conducted a detailed investigation across official municipal websites, contacted municipal headquarters, and reviewed various academic and journalistic publications. This effort enabled me to compile data on 172 out of the 215 municipalities in the state of Puebla, resulting in a total of 658 Juntas Auxiliares. It is important to note that not every municipality has the same number of Juntas Auxiliares; some have only one, while others have more than ten. This source of variation is exploited in this paper.

```r
data_EH <- read_csv("Econ History - paper.csv")
```

Here are the labels for each variable to clarify the dataset.

```r
library(Hmisc)

label(data_EH$estado) <- "Name of state"
label(data_EH$municipio) <- "Name of municipality"
label(data_EH$juntas) <- "Number of Juntas Auxiliares in the municipality"
label(data_EH$income2020) <- "Average household income of the municipality at the year 2020"
label(data_EH$calculo_ind) <- "Type of calculation of indigenous population. 'm' means 'sample' and 'c' means 'ce
nsus'"
label(data_EH$pob_tot) <- "Total population"
label(data_EH$pob_ind) <- "Indigenous population"
label(data_EH$pob_no_ind) <- "Non-indigenous population"
label(data_EH$no_esp) <- "Not specified"
label(data_EH$por) <- "General percentage, should be 1"
label(data_EH$por_ind) <- "Percentage of indigenous population"
label(data_EH$por_no_ind) <- "Percentage of non-indigenous population"
label(data_EH$por_no_esp) <- "Percentage not specified"
label(data_EH$lat) <- "Latitude"
label(data_EH$lon) <- "Longitude"
label(data_EH$alt) <- "Altitude"
label(data_EH$pob) <- "Population"
label(data_EH$pob_male) <- "Male population"
label(data_EH$pob_female) <- "Female population"
label(data_EH$houses) <- "Number of houses"
label(data_EH$ext) <- "Territorial extension (sq km)"
label(data_EH$porcentaje_ext) <- "What percentage of the territory of the state represents the municipality"
label(data_EH$densidad) <- "Population density (hab/km2)"
label(data_EH$localidades) <- "Number of localities"
```

Just verifying that all labels are okay:

```r
sapply(data_EH, label)
```

```
##                                                                                     estado
##                                                                            "Name of state"
##                                                                                  municipio
##                                                                     "Name of municipality"
##                                                                                     juntas
##                                          "Number of Juntas Auxiliares in the municipality"
##                                                                                 income2020
##                            "Average household income of the municipality at the year 2020"
##                                                                                calculo_ind
## "Type of calculation of indigenous population. 'm' means 'sample' and 'c' means 'census'"
##                                                                                    pob_tot
##                                                                         "Total population"
##                                                                                    pob_ind
##                                                                    "Indigenous population"
##                                                                                 pob_no_ind
##                                                                "Non-indigenous population"
##                                                                                     no_esp
##                                                                            "Not specified"
##                                                                                        por
##                                                          "General percentage, should be 1"
##                                                                                    por_ind
##                                                      "Percentage of indigenous population"
##                                                                                 por_no_ind
##                                                  "Percentage of non-indigenous population"
##                                                                                 por_no_esp
##                                                                 "Percentage not specified"
##                                                                                        lat
##                                                                                 "Latitude"
##                                                                                        lon
##                                                                                "Longitude"
##                                                                                        alt
##                                                                                 "Altitude"
##                                                                                        pob
##                                                                               "Population"
##                                                                                   pob_male
##                                                                          "Male population"
##                                                                                 pob_female
##                                                                        "Female population"
##                                                                                     houses
##                                                                         "Number of houses"
##                                                                                        ext
##                                                            "Territorial extension (sq km)"
##                                                                             porcentaje_ext
##               "What percentage of the territory of the state represents the municipality"
##                                                                                   densidad
##                                                             "Population density (hab/km2)"
##                                                                                localidades
##                                                                     "Number of localities"
```

## Methodology

First, let's measure the impact of the number of Juntas Auxiliares on household income using simple OLS with different model specifications.

```r
OLS_a <- lm(income2020 ~ juntas, data = data_EH)
OLS_b <- lm(income2020 ~ juntas + lat + lon + alt, data = data_EH)
OLS_c <- lm(income2020 ~ juntas + lat + lon + alt + por_ind, data = data_EH)
OLS_d <- lm(income2020 ~ juntas + lat + lon + alt + por_ind + juntas:por_ind, data = data_EH)
```

```r
mean_dependent_var <- mean(data_EH$income2020[!is.na(data_EH$juntas)])
```

```r
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```r
stargazer(OLS_a, OLS_b, OLS_c, OLS_d, type = "text",
          title = "OLS - Results",
          align = TRUE,
          header = FALSE,
          dep.var.labels.include = FALSE,
          covariate.labels = c("Juntas", "Latitude", "Longitude", "Altitude", "Percentage Indigenous", "Juntas:Pe
rcentage Indigenous"),
          omit.stat = c("LL", "ser", "f"),
          model.names = FALSE,
          omit = "Constant",
          add.lines = list(c("Mean of Dependent Variable", format(mean_dependent_var, digits=4), format(mean_depe
ndent_var, digits=4), format(mean_dependent_var, digits=4), format(mean_dependent_var, digits=4), format(mean_dep
endent_var, digits=4))))
```

```
##
## OLS - Results
## ==================================================================================
##                                              Dependent variable:
##                              -----------------------------------------------------
##                                 (1)          (2)           (3)            (4)
## ----------------------------------------------------------------------------------
## Juntas                       546.444***  463.832***     447.701***   1,056.982***
##                              (136.241)    (127.990)     (114.213)      (221.127)
##
## Latitude                                -1,645.655**     -545.024      -524.640
##                                           (666.220)     (617.188)      (600.866)
##
## Longitude                               -3,134.837***    -316.263      -300.800
##                                          (1,014.811)   (1,000.496)     (973.995)
##
## Altitude                                   2.491***        0.066          0.231
##                                            (0.713)       (0.734)        (0.717)
##
## Percentage Indigenous                                 -11,291.060*** -6,419.207***
##                                                        (1,705.749)    (2,256.844)
##
## Juntas:Percentage Indigenous                                         -1,397.614***
##                                                                        (438.453)
##
## ----------------------------------------------------------------------------------
## Mean of Dependent Variable     29230        29230         29230          29230
## Observations                    172          172           172            172
## R2                             0.086        0.250         0.406          0.441
## Adjusted R2                    0.081        0.232         0.389          0.421
## ==================================================================================
## Note:                                                  *p<0.1; **p<0.05; ***p<0.01
```

However, this model presents an endogeneity problem: higher or lower income may cause the community to demand more Juntas Auxiliares and Juntas Auxiliares may cause conditions that affect income. In other words, there could be a reverse causality issue.

## Solving the endogenity: IV

In order to solve the endogeneity problem, I use territorial extension of each municipality as instrument for the number of Juntas Auxiliares.

To analyse the relationship between Juntas Auxiliares and Territorial extension, here are the results of the first stage.

```r
First_a <- lm(juntas ~ ext, data = data_EH)
First_b <- lm(juntas ~ ext + lat + lon + alt, data = data_EH)
First_c <- lm(juntas ~ ext + lat + lon + alt + por_ind, data = data_EH)
```

```r
mean_dependent_var_re <- mean(data_EH$juntas[!is.na(data_EH$juntas)])
```

```r
stargazer(First_a, First_b, First_c, type = "text",
          title = "First Stage - Results",
          align = TRUE,
          header = FALSE,
          dep.var.labels.include = FALSE,
          covariate.labels = c("Juntas", "Latitude", "Longitude", "Altitude", "Percentage Indigenous", "Juntas:Pe
rcentage Indigenous"),
          omit.stat = c("LL", "ser"),
          model.names = FALSE,
          add.lines = list(c("Mean of Dependent Variable", format(mean_dependent_var_re, digits=4), format(mean_d
ependent_var_re, digits=4), format(mean_dependent_var_re, digits=4), format(mean_dependent_var_re, digits=4), for
mat(mean_dependent_var_re, digits=4))))
```

```
##
## First Stage - Results
## ====================================================================================================
##                                                        Dependent variable:
##                              ----------------------------------------------------------------------
##                                       (1)                     (2)                     (3)
## ----------------------------------------------------------------------------------------------------
## Juntas                             0.009***                0.010***                0.010***
##                                    (0.002)                 (0.002)                 (0.002)
##
## Latitude                                                   1.242***                1.190***
##                                                            (0.370)                 (0.382)
##
## Longitude                                                  -1.511***               -1.670***
##                                                            (0.549)                 (0.614)
##
## Altitude                                                   0.001**                 0.001**
##                                                           (0.0004)                (0.0004)
##
## Percentage Indigenous                                                               0.616
##                                                                                    (1.058)
##
## Juntas:Percentage Indigenous       2.196***               -171.371***             -186.430***
##                                    (0.397)                 (55.876)                (61.670)
##
## ----------------------------------------------------------------------------------------------------
## Mean of Dependent Variable           3.826                   3.826                   3.826
## Observations                          172                     172                     172
## R2                                   0.141                   0.225                   0.227
## Adjusted R2                          0.136                   0.207                   0.204
## F Statistic                27.875*** (df = 1; 170) 12.142*** (df = 4; 167) 9.743*** (df = 5; 166)
## ====================================================================================================
## Note:                                                                      *p<0.1; **p<0.05; ***p<0.01
```

From the first stage is important to mention the following: the instrument has a lot of predictable power for the endogenous variable, especially without any control. With an F-statistic of 27.87, it can be argued that this is an strong instrument.

In other concerns, there might be a serious problem regarding territorial extent as an instrument; it might be that territorial extent affects income, potentially violating the exogeneity and excludability condition.

## Defending the instrument: 31 falsification tests

In order to defend the instrument condition I had the following idea: if territorial extension of another state's municipalities doesn't have an impact on income, I can argue the validity of the instrument.

However, I could test this idea on more than one state since I have data for municipalities from all 32 states of Mexico. Therefore, I decided to conduct 31 falsification tests.

```r
library(broom)
library(dplyr)
```

```r
lista_estados <- split(data_EH, data_EH$estado)

modelos <- lapply(lista_estados, function(subconjunto) {
  lm(income2020 ~ ext, data = subconjunto)
})

obs_por_modelo <- sapply(modelos, function(x) length(x$residuals))

tidy_modelos <- lapply(names(modelos), function(nombre) {
  modelo_tidy <- tidy(modelos[[nombre]])
  modelo_tidy <- modelo_tidy[modelo_tidy$term == "ext",]
  modelo_tidy$Observations <- obs_por_modelo[nombre]
  modelo_tidy$Estado <- nombre
  modelo_tidy
})

resultados <- do.call(rbind, tidy_modelos)

resultados_finales <- resultados %>%
  select(Estado, Coefficient = estimate, Std.Errors = std.error, p.values = p.value, Observations)
```

```r
library(pander)
pander(resultados_finales)
```

| Estado | Coefficient | Std.Errors | p.values | Observations |
|---|---|---|---|---|
| 00 Estados Unidos Mexicanos | NA | NA | NA | 1 |
| 01 Aguascalientes | 2.421 | 1.629 | 0.1681 | 12 |
| 02 Baja California | 0.04188 | 0.07986 | 0.6277 | 6 |
| 03 Baja California Sur | 0.04683 | 0.1225 | 0.7217 | 6 |
| 04 Campeche | 0.1698 | 0.1998 | 0.4153 | 12 |
| 05 Coahuila de Zaragoza | 0.04961 | 0.05754 | 0.3941 | 39 |
| 06 Colima | 1.508 | 1.542 | 0.3536 | 11 |
| 07 Chiapas | 0.08333 | 0.08438 | 0.3254 | 119 |
| 08 Chihuahua | 0.08165 | 0.0439 | 0.06738 | 68 |
| 09 Ciudad de México | -3.981 | 9.918 | 0.6938 | 17 |
| 10 Durango | 0.1032 | 0.06391 | 0.1146 | 40 |
| 11 Guanajuato | 0.2179 | 0.2315 | 0.3516 | 47 |
| 12 Guerrero | 0.06745 | 0.1566 | 0.6679 | 82 |
| 13 Hidalgo | 0.245 | 0.3493 | 0.485 | 85 |
| 14 Jalisco | 0.1389 | 0.09011 | 0.1257 | 126 |
| 15 México | 0.2814 | 0.4115 | 0.4953 | 126 |
| 16 Michoacán de Ocampo | 0.08306 | 0.131 | 0.5273 | 114 |
| 17 Morelos | 1.212 | 1.42 | 0.3997 | 34 |
| 18 Nayarit | 0.04559 | 0.4034 | 0.9112 | 21 |
| 19 Nuevo León | 0.1615 | 0.2423 | 0.5082 | 52 |
| 20 Oaxaca | 0.07271 | 0.07528 | 0.3345 | 571 |
| 21 Puebla | 0.327 | 0.1786 | 0.06854 | 218 |
| 22 Querétaro | 1.298 | 1.111 | 0.2588 | 19 |
| 23 Quintana Roo | 0.06822 | 0.212 | 0.7542 | 12 |
| 24 San Luis Potosí | 0.2249 | 0.1277 | 0.08357 | 59 |
| 25 Sinaloa | 0.1565 | 0.1573 | 0.3336 | 19 |
| 26 Sonora | 0.1084 | 0.05097 | 0.03694 | 73 |
| 27 Tabasco | 0.08251 | 0.3435 | 0.8132 | 18 |
| 28 Tamaulipas | 0.1345 | 0.122 | 0.2766 | 44 |
| 29 Tlaxcala | 0.8856 | 1.516 | 0.5614 | 61 |
| 30 Veracruz de Ignacio de la Llave | 0.08784 | 0.09592 | 0.3609 | 213 |
| 31 Yucatán | 0.3971 | 0.1599 | 0.01459 | 107 |
| 32 Zacatecas | 0.06735 | 0.08735 | 0.4439 | 59 |

Only 5 out of the 32 Mexican states exhibit a statistically significant relationship between territorial extension and income household (Chihuahua, Puebla, San Luis Potosí, Sonora, and Yucatán), suggesting that, in most cases, territorial extension does not impact income.

This evidence supports the validity of the instrument, reinforcing the idea of using territorial extent as an instrumental variable for number of Juntas Auxiliares. The lack of widespread significance across the states reinforces the argument that observable and unobserved factors of income are unlikely to be systematically correlated with territorial extension, allowing for a more confident interpretation of the causal effects in the analysis.

## Using the IV

```r
library(AER)
```

```r
modelo_iv1 <- ivreg(income2020 ~ juntas | ext, data = data_EH)
modelo_iv2 <- ivreg(income2020 ~ juntas + lat + lon + alt | ext + lat + lon + alt, data = data_EH)
modelo_iv3 <- ivreg(income2020 ~ juntas + lat + lon + alt + por_ind | ext + lat + lon + alt + por_ind, data = dat
a_EH)
```

```r
stargazer(modelo_iv1, modelo_iv2, modelo_iv3, type = "text",
          title = "IV - Results",
          align = TRUE,
          header = FALSE,
          dep.var.labels.include = FALSE,
          covariate.labels = c("Juntas", "Latitude", "Longitude", "Altitude", "Percentage Indigenous"),
          omit.stat = c("LL", "ser"),
          model.names = FALSE,
          omit = "Constant",
          add.lines = list(c("Mean of Dependent Variable", format(mean_dependent_var, digits=4), format(mean_depe
ndent_var, digits=4), format(mean_dependent_var, digits=4), format(mean_dependent_var, digits=4), format(mean_dep
endent_var, digits=4))))
```

```
##
## IV - Results
## =================================================================
##                                      Dependent variable:
##                              -------------------------------------
##                                 (1)        (2)           (3)
## -----------------------------------------------------------------
## Juntas                       937.316**  705.685**      449.370*
##                              (371.676)  (303.113)     (266.691)
##
## Latitude                              -1,823.697**     -546.303
##                                         (702.896)     (644.266)
##
## Longitude                             -2,828.730***    -314.284
##                                        (1,082.700)   (1,040.494)
##
## Altitude                                 2.269***        0.065
##                                          (0.763)       (0.762)
##
## Percentage Indigenous                               -11,290.530***
##                                                      (1,707.478)
##
## -----------------------------------------------------------------
## Mean of Dependent Variable     29230      29230         29230
## Observations                    172        172           172
## R2                             0.042      0.234         0.406
## Adjusted R2                    0.037      0.215         0.389
## =================================================================
## Note:                                 *p<0.1; **p<0.05; ***p<0.01
```

## Conclusion

The positive and statistically significant coefficients for Juntas Auxiliares across different model specifications and within the OLS and the IV approach, provides solid evidence to argue that local institutions can be effective in driving economic development.

Overall, the research suggests that Juntas Auxiliares play an influential role in municipal governance and economic outcomes, validating their importance in Puebla's social and economic framework. The examination of the exogeneity of territorial extent as an instrument though 31 falsification tests solidify these findings, offering a robust argument for the selected econometric approach.

The comprehensive data, the careful application of statistical methods, and the discussion of potential limitations all contribute to the strength of this study's conclusions: is better to build bridges.