main_analysis.R

Poertnem

2023-01-20

```
rm(list=ls())
library(plyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stargazer)
```

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

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(ggplot2)
library(stringr)
library(ri2)
```

```
## Loading required package: randomizr
```

```
## Loading required package: estimatr
```

```
library(boot)
library(lfe)
```

```
## Loading required package: Matrix
```

```
library(sjPlot)
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

```
## The following objects are masked from 'package:Matrix':
```

```
##
##     expand, pack, unpack
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(foreign)
library(Hmisc)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'lattice'
```

```
## The following object is masked from 'package:boot':
##
##     melanoma
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:boot':
##
##     aml
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
```

```
## The following objects are masked from 'package:plyr':
##
##     is.discrete, summarize
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
library(sna)
```

```
## Loading required package: statnet.common
```

```
##
## Attaching package: 'statnet.common'
```

```
## The following objects are masked from 'package:base':
##
##     attr, order
```

```
## Loading required package: network
```

```
##
## 'network' 1.17.1 (2021-06-12), part of the Statnet Project
## * 'news(package="network")' for changes since last version
## * 'citation("network")' for citation information
## * 'https://statnet.org' for help, support, and other information

##
## Attaching package: 'network'

## The following object is masked from 'package:Hmisc':
##
##     is.discrete

## The following object is masked from 'package:plyr':
##
##     is.discrete

## sna: Tools for Social Network Analysis
## Version 2.6 created on 2020-10-5.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##  For citation information, type citation("sna").
##  Type help(package="sna") to get started.
library(gridExtra)

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
library(grDevices)
library(sandwich)
library(reshape2)

##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths
library(parallel)
library(xtable)

##
## Attaching package: 'xtable'

## The following objects are masked from 'package:Hmisc':
##
##     label, label<-
library(Cairo)

###set WD:
#setwd() #set to your local working directory
setwd("~/Google Drive/Book Project/Tombola Papers/Party Building Paper/JOP/Replication Data")

##########
```

```r
# Preparing the data:
##########

###load data with 2015 MORENA PR list candidates:
tombola <- read.csv("MORENA_list.csv", header=T, sep=",")

###data cleaning:
tombola$fecha_nacimiento <- as.Date(tombola$fecha_nacimiento, "%d/%m/%Y")
tombola$fecha_afiliacion <- as.Date(tombola$fecha_afiliacion, "%d/%m/%Y")
tombola$age <- as.numeric(floor((as.Date("7/6/2015","%d/%m/%Y") - tombola$fecha_nacimiento)/365)) #calc

#subset to lottery candidates:
tombola <- subset(tombola, tombola$internal==1)


###initial balance tests:
#Table 2:

balance.tests.diputado <- rbind(
  c("Female",t.test(tombola$genero[tombola$diputado_dummy==0],tombola$genero[tombola$diputado_dummy==1]
  c("tiempo_de_residencia",t.test(tombola$tiempo_de_residencia[tombola$diputado_dummy==0],tombola$tiempo
  c("Age",t.test(tombola$age[tombola$diputado_dummy==0],tombola$age[tombola$diputado_dummy==1])$estimat

colnames(balance.tests.diputado) <- c("Variable","Control","Treatment","Diff. means","p-value","n")
balance.tests.diputado[,2] <- round(as.numeric(balance.tests.diputado[,2]),4)
balance.tests.diputado[,3] <- round(as.numeric(balance.tests.diputado[,3]),4)
balance.tests.diputado[,4] <- round(as.numeric(balance.tests.diputado[,4]),4)
balance.tests.diputado[,5] <- round(as.numeric(balance.tests.diputado[,5]),4)

balance.tests.diputado[,1] <- c("Female","Years of Residence","Age")

balance.tests.diputado
```

```
##      Variable             Control   Treatment Diff. means p-value  n
## [1,] "Female"             "0.5128"  "0.5"     "-0.0128"   "0.9222" "135"
## [2,] "Years of Residence" "4.0983"  "4.75"    "0.6517"    "0.4773" "135"
## [3,] "Age"                "45.6496" "44.3889" "-1.2607"   "0.7084" "135"
```

```r
#Table 2:
stargazer(balance.tests.diputado, out="Table2.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:32
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}} cccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Variable & Control & Treatment & Diff. means & p-value & n \\
## \hline \\[-1.8ex]
## Female & 0.5128 & 0.5 & -0.0128 & 0.9222 & 135 \\
## Years of Residence & 4.0983 & 4.75 & 0.6517 & 0.4773 & 135 \\
## Age & 45.6496 & 44.3889 & -1.2607 & 0.7084 & 135 \\
```

```
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

```
#F-Test:
f_test_balance_diputado <- lm(diputado_dummy ~ genero + tiempo_de_residencia + age, data=tombola)
summary(f_test_balance_diputado)
```

```
##
## Call:
## lm(formula = diputado_dummy ~ genero + tiempo_de_residencia +
##     age, data = tombola)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.1947 -0.1463 -0.1248 -0.1031  0.9096
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.176187   0.114033   1.545    0.125
## genero               -0.007557   0.059310  -0.127    0.899
## tiempo_de_residencia  0.007268   0.008496   0.855    0.394
## age                  -0.001526   0.002410  -0.633    0.528
##
## Residual standard error: 0.3439 on 131 degrees of freedom
## Multiple R-squared:  0.006757,   Adjusted R-squared:  -0.01599
## F-statistic: 0.2971 on 3 and 131 DF,  p-value: 0.8275
```

```
###merge with 2018 electoral data:
load("elections_2018_municipio.Rda")

#some data cleaning:
elections_2018_municipio <- as.data.frame(elections_2018_municipio)
colnames(elections_2018_municipio)[2] <- "municipio_residencia"
tombola$municipio_residencia <- as.character(tombola$municipio_residencia)
elections_2018_municipio$municipio_residencia <- as.character(elections_2018_municipio$municipio_reside

for (r in 1:nrow(tombola)){
  tombola$municipio_residencia[r] <- toupper(tombola$municipio_residencia[r])
}

#code state names consistently with other dataset:
elections_2018_municipio$estado_short <- elections_2018_municipio$estado
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="YUCATAN"] <- "YUC"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="COAHUILA"] <- "COAH"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="OAXACA"] <- "OAX"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="CHIAPAS"] <- "CHIS"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="PUEBLA"] <- "PUE"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="GUANAJUATO"] <- "GTO"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="MEXICO"] <- "MEX"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="NAYARIT"] <- "NAY"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="GUERRERO"] <- "GRO"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="JALISCO"] <- "JAL"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="HIDALGO"] <- "HGO"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="VERACRUZ"] <- "VER"
```

```
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="SONORA"] <- "SON"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="TLAXCALA"] <- "TLAX"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="MICHOACAN"] <- "MICH"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="NUEVO LEON"] <- "NL"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="AGUASCALIENTES"] <- "AGS"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="SINALOA"] <- "SIN"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="SAN LUIS POTOSI"] <- "SLP"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="CHIHUAHUA"] <- "CHIH"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="TAMAULIPAS"] <- "TAMPS"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="CIUDAD DE MEXICO"] <- "CDM
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="MORELOS"] <- "MOR"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="QUERETARO"] <- "QRO"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="ZACATECAS"] <- "ZAC"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="COLIMA"] <- "COL"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="QUINTANA ROO"] <- "QROO"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="TABASCO"] <- "TAB"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="CAMPECHE"] <- "CAM"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="DURANGO"] <- "DGO"
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="BAJA CALIFORNIA SUR"] <- "
elections_2018_municipio$estado_short[elections_2018_municipio$estado_short=="BAJA CALIFORNIA"] <- "BC"

#unique state-municipality id:
tombola$id_mun <- paste(tombola$estado_residencia, tombola$municipio_residencia, sep="_")
elections_2018_municipio$id_mun <- paste(elections_2018_municipio$estado_short, elections_2018_municipio

#number of candidates per municipality in treatment or control:
mun <- tombola %>%
  group_by(id_mun) %>%
  count(internal)

#number of candidates per municipality in treatment:
mun_t <- subset(tombola, diputado_dummy==1) %>%
  group_by(id_mun) %>%
  count(diputado_dummy)

#number of candidates per municipality in control:
mun_c <- subset(tombola, diputado_dummy==0) %>%
  group_by(id_mun) %>%
  count(diputado_dummy)

municipalities <- left_join(elections_2018_municipio,mun)
```

```
## Joining, by = "id_mun"
```

```
names(municipalities)[20] <- "tombola"
municipalities <- municipalities[,c(1:18,20)]

municipalities <- left_join(municipalities,mun_t)
```

```
## Joining, by = "id_mun"
```

```
names(municipalities)[21] <- "deputy"
municipalities <- municipalities[,c(1:19,21)]

municipalities <- left_join(municipalities,mun_c)
```

```
## Joining, by = "id_mun"
```

```
names(municipalities)[22] <- "candidate"
municipalities <- municipalities[,c(1:20,22)]

municipalities_full <- municipalities
municipalities_full$deputy <- ifelse(is.na(municipalities_full$deputy),0,municipalities_full$deputy)
municipalities_full$candidate <- ifelse(is.na(municipalities_full$candidate),0,municipalities_full$candi

#coding main treatment variable (elected deputy):
municipalities_full$diputado_dummy <- ifelse(municipalities_full$deputy > 0, 1, 0)

municipalities <- subset(municipalities, !is.na(tombola))
municipalities$deputy <- ifelse(is.na(municipalities$deputy),0,municipalities$deputy)
municipalities$candidate <- ifelse(is.na(municipalities$candidate),0,municipalities$candidate)

municipalities$diputado_dummy <- ifelse(municipalities$deputy > 0, 1, 0)


##add information other other deputies from that municipality:

#municipalities that were (also) represented by other deputy:
mun_mult_elect_1 <- unique(subset(tombola, mult_elect==1)$id_mun)

#municipalities that were not represented by other deputy:
mun_mult_elect_0 <- unique(subset(tombola, mult_elect==0)$id_mun)

municipalities$mult_elect <- ifelse(municipalities$id_mun %in% mun_mult_elect_0, 0, NA) #municipalities
municipalities$mult_elect <- ifelse(municipalities$id_mun %in% mun_mult_elect_1, 1, municipalities$mult_

#states in each multi-state constituency:
circun_1 <- unique(subset(tombola, circunscripcion==1)$estado_residencia)
circun_2 <- unique(subset(tombola, circunscripcion==2)$estado_residencia)
circun_3 <- unique(subset(tombola, circunscripcion==3)$estado_residencia)
circun_4 <- unique(subset(tombola, circunscripcion==4)$estado_residencia)
circun_5 <- unique(subset(tombola, circunscripcion==5)$estado_residencia)


municipalities$circunscripcion <- ifelse(municipalities$estado_short %in% circun_1, 1, NA)
municipalities$circunscripcion <- ifelse(municipalities$estado_short %in% circun_2, 2, municipalities$ci
municipalities$circunscripcion <- ifelse(municipalities$estado_short %in% circun_3, 3, municipalities$ci
municipalities$circunscripcion <- ifelse(municipalities$estado_short %in% circun_4, 4, municipalities$ci
municipalities$circunscripcion <- ifelse(municipalities$estado_short %in% circun_5, 5, municipalities$ci

###add election results from 2012 (baseline):
load("elections_2012_municipio.Rda")

#some data cleaning:
elections_2012_municipio <- as.data.frame(elections_2012_municipio)

#create a unique ID for each municipality:
elections_2012_municipio$id <- paste0(elections_2012_municipio$id_estado,elections_2012_municipio$munici
municipalities$id <- paste0(municipalities$id_estado,municipalities$municipio)
```

```
municipalities_full$id <- paste0(municipalities_full$id_estado,municipalities_full$municipio)

#merge the data sets:
municipalities <- join(municipalities,elections_2012_municipio,by="id",type="left")
municipalities_full <- join(municipalities_full,elections_2012_municipio,by="id",type="left")

#municipalities with valid lottery candidates:
valid_tombola <- subset(municipalities, !is.na(tombola))

#municipalities with valid deputy:
valid_deputy <- subset(municipalities, !is.na(deputy))

#municipalities with valid candidate:
valid_candidate <- subset(municipalities, !is.na(candidate))

#municipalities with valid diputado_dummy:
valid_diputado_dummy <- subset(municipalities, !is.na(diputado_dummy))

#number of candidates per constituency:
const_tombola <- valid_diputado_dummy %>%
  group_by(circunscripcion) %>%
  summarise(const_tombola = sum(tombola))

#number of elected deputies per constituency:
const_deputy <- valid_diputado_dummy %>%
  group_by(circunscripcion) %>%
  summarise(const_deputy = sum(deputy))

valid_diputado_dummy <- left_join(valid_diputado_dummy,const_tombola)
```

```
## Joining, by = "circunscripcion"
```

```
valid_diputado_dummy <- left_join(valid_diputado_dummy,const_deputy)
```

```
## Joining, by = "circunscripcion"
```

```
##probability of treatment assignment:
#for any given candidate: number of candidates elected from a list / length of list
#for any given municipality: prob of a candidate getting elected * number of candidates

valid_diputado_dummy$prob <- (valid_diputado_dummy$const_deputy/valid_diputado_dummy$const_tombola) * va
valid_diputado_dummy$ipw <- ifelse(valid_diputado_dummy$diputado_dummy==1, 1 / valid_diputado_dummy$prob

save(valid_diputado_dummy, file="valid_diputado_dummy.RDa") #save data to use for additional analyses


###Merge with party membership data:

#load membership data:
load("padron.RDa")
padron$fecha_afiliacion <- as.Date(padron$fecha_afiliacion, "%Y-%m-%d")

#recode dates of joining the party:
padron$pre_tombola <- ifelse(padron$fecha_afiliacion < as.Date("2015-06-07","%Y-%m-%d"), 1, 0)
padron$new2012 <- ifelse(padron$fecha_afiliacion > as.Date("2011-12-31","%Y-%m-%d") & padron$fecha_afili
```

```
padron$new2013 <- ifelse(padron$fecha_afiliacion > as.Date("2012-12-31","%Y-%m-%d") & padron$fecha_afili
padron$new2014 <- ifelse(padron$fecha_afiliacion > as.Date("2013-12-31","%Y-%m-%d") & padron$fecha_afili
padron$new2015 <- ifelse(padron$fecha_afiliacion > as.Date("2015-06-06","%Y-%m-%d") & padron$fecha_afili
padron$new2016 <- ifelse(padron$fecha_afiliacion > as.Date("2015-12-31","%Y-%m-%d") & padron$fecha_afili
padron$new2017 <- ifelse(padron$fecha_afiliacion > as.Date("2016-12-31","%Y-%m-%d") & padron$fecha_afili
padron$new2018 <- ifelse(padron$fecha_afiliacion > as.Date("2017-12-31","%Y-%m-%d") & padron$fecha_afili

#calculate number of party members by municipality:
padron_mun <-
padron %>%
  group_by(entidad, mun) %>%
  summarise(pre_tombola = sum(pre_tombola,na.rm=T), new2012 = sum(new2012,na.rm =T), new2013 = sum(new2
            new2017 = sum(new2017,na.rm=T), new2018 = sum(new2018,na.rm=T))
```

```
## `summarise()` has grouped output by 'entidad'. You can override using the
## `.groups` argument.
```

```
#correcting some encoding errors:
padron_mun$mun[padron_mun$mun=="GUSTAVO A MADERO"] <- "GUSTAVO A. MADERO"

padron_mun$mun_id <- paste0(padron_mun$entidad,"_",padron_mun$mun)

#merge with valid_diputado_dummy:
valid_diputado_dummy$mun_id <- paste0(valid_diputado_dummy$estado,"_",valid_diputado_dummy$municipio_res
valid_diputado_dummy <- left_join(valid_diputado_dummy,padron_mun, by= "mun_id")


###recode a few variables for analysis:
#number total MORENA members by end of 2018:
padron$total <- padron$new2012 + padron$new2013 + padron$new2014 + padron$new2015 + padron$new2016 + pad
sum(padron$total, na.rm = T)
```

```
## [1] 440087
```

```
#calculate overall membership rate:
sum(padron$total, na.rm = T)/sum(municipalities_full$lista_nominal_deputies)
```

```
## [1] 0.004948059
```

```
#Proportion of registered voters who are affiliated with the party PRIOR to the lottery in percent:
valid_diputado_dummy$pre_tombola_percent <- valid_diputado_dummy$pre_tombola / valid_diputado_dummy$list

#Proportion of registered voters who are affiliated with the party PRIOR to the lottery per 100,000:
valid_diputado_dummy$pre_tombola_per <- valid_diputado_dummy$pre_tombola / valid_diputado_dummy$lista_no

#Number of voters who affiliated with the party AFTER to the lottery:
valid_diputado_dummy$combined_padron <- valid_diputado_dummy$new2015 + valid_diputado_dummy$new2016 + va

#Proportion of registered voters who affiliated with the party AFTER to the lottery in percent:
valid_diputado_dummy$combined_padron_percent <- valid_diputado_dummy$combined_padron / valid_diputado_du

#Proportion of registered voters who affiliated with the party AFTER to the lottery per 100,000:
valid_diputado_dummy$combined_padron_per <- valid_diputado_dummy$combined_padron / valid_diputado_dummy$

#Combined proportion of registered voters who affiliated with the party before or after the lottery in p
valid_diputado_dummy$combined_padron_total_percent <- valid_diputado_dummy$combined_padron_percent + val
```

```r
#Combined proportion of registered voters who affiliated with the before or after the lottery per 100,0
valid_diputado_dummy$combined_padron_total_per <- valid_diputado_dummy$combined_padron_per + valid_diput


##########
# MAIN ANALYSES:
##########

###Local party membership (in percent):

##Effect on proportion of registered voters who affiliated with the party at baseline (i.e., BEFORE to

#IPW:
pre_tombola_percent_ipw_cl <- felm(pre_tombola_percent ~ diputado_dummy | 0 | 0 | circunscripcion + tomb
pre_tombola_percent_ipw <- felm(pre_tombola_percent ~ diputado_dummy | 0 | 0 | 0, data=valid_diputado_du

#FE Model:
pre_tombola_percent_fe <- felm(pre_tombola_percent ~ diputado_dummy | tombola + circunscripcion | 0 | es


##Effect on proportion of registered voters who affiliated with the party AFTER to the lottery in percen

#IPW:
combined_padron_percent_ipw_cl <- felm(combined_padron_percent ~ diputado_dummy | 0 | 0 | circunscripcio
combined_padron_percent_ipw <- felm(combined_padron_percent ~ diputado_dummy | 0 | 0 | 0, data=valid_dip

#FE Model:
combined_padron_percent_fe <- felm(combined_padron_percent ~ diputado_dummy | tombola + circunscripcion


##Effect on proportion of registered voters who affiliated with the party AFTER to the lottery in percen
##controlling for the number of party members before the lottery:

#IPW:
combined_padron_percent_control_ipw_cl <- felm(combined_padron_percent ~ diputado_dummy + pre_tombola_pe
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```r
combined_padron_percent_control_ipw <- felm(combined_padron_percent ~ diputado_dummy  + pre_tombola_perc

#FE Model:
combined_padron_percent_control_fe <- felm(combined_padron_percent ~ diputado_dummy + pre_tombola_percen


##Effect on total proportion of registered voters who affiliated by the end of 2018 in percent:

#IPW:
combined_padron_total_percent_ipw_cl <- felm(combined_padron_total_percent ~ diputado_dummy | 0 | 0 | ci
combined_padron_total_percent_ipw <- felm(combined_padron_total_percent ~ diputado_dummy | 0 | 0 | 0, da

#FE Model:
combined_padron_total_percent_fe <- felm(combined_padron_total_percent ~ diputado_dummy  | tombola + ci
```

```
##Effect on total proportion of registered voters who affiliated by the end of 2018 in percent:
##controlling for the number of party members before the lottery:

#IPW:
combined_padron_total_control_percent_ipw_cl <- felm(combined_padron_total_percent ~ diputado_dummy + pr
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
combined_padron_total_control_percent_ipw <- felm(combined_padron_total_percent ~ diputado_dummy + pre_t

#FE Model:
combined_padron_total_control_percent_fe <- felm(combined_padron_total_percent ~ diputado_dummy + pre_to


###Local party membership (rescaled -- per 100,000 registered votes):

##Effect on proportion of registered voters who are affiliated with the party PRIOR to the lottery per

#IPW:
pre_tombola_per_ipw_cl <- felm(pre_tombola_per ~ diputado_dummy | 0 | 0 | circunscripcion + tombola, da
pre_tombola_per_ipw <- felm(pre_tombola_per ~ diputado_dummy | 0 | 0 | 0, data=valid_diputado_dummy, we

#FE Model:
pre_tombola_per_fe <- felm(pre_tombola_per ~ diputado_dummy | tombola + circunscripcion | 0 | estado + t


##Effect on proportion of registered voters who affiliated with the party AFTER to the lottery per 100,

#IPW:
combined_padron_per_ipw_cl <- felm(combined_padron_per ~ diputado_dummy | 0 | 0 | circunscripcion + tomb
combined_padron_per_ipw <- felm(combined_padron_per ~ diputado_dummy | 0 | 0 | 0, data=valid_diputado_du

#FE Model:
combined_padron_per_fe <- felm(combined_padron_per ~ diputado_dummy | tombola + circunscripcion | 0 | 0


##Effect on proportion of registered voters who affiliated with the party AFTER to the lottery per 100,
##controlling for the number of party members before the lottery:

#IPW:
combined_padron_per_control_ipw_cl <- felm(combined_padron_per ~ diputado_dummy + pre_tombola_per | 0 |
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
combined_padron_per_control_ipw <- felm(combined_padron_per ~ diputado_dummy  + pre_tombola_per | 0 | 0

#FE Model:
combined_padron_per_control_fe <- felm(combined_padron_per ~ diputado_dummy + pre_tombola_per | tombola
```

```r
##Effect on total proportion of registered voters who affiliated by the end of 2018 per 100,000:

#IPW:
combined_padron_total_per_ipw_cl <- felm(combined_padron_total_per ~ diputado_dummy | 0 | 0 | circunscri
combined_padron_total_per_ipw <- felm(combined_padron_total_per ~ diputado_dummy | 0 | 0 | 0, data=valid

#FE Model:
combined_padron_total_per_fe <- felm(combined_padron_total_per ~ diputado_dummy  | tombola + circunscrip


###Combined results:

#function to adjust p values for one tailed tests:
one_tailed <- function(x){
  p_one_tailed <- x / 2
  return(p_one_tailed)
}

#calculate percent change relative to control:
percent_change_combined <- paste0("+",round(combined_padron_percent_ipw_cl$coefficients[2]/combined_padr
percent_change_total <- paste0("+",round(combined_padron_total_percent_ipw_cl$coefficients[2]/combined_p


##Table 4:
stargazer(combined_padron_per_ipw_cl, combined_padron_total_per_ipw_cl, digits=2,
          dep.var.caption = "Outcome:",
          dep.var.labels = c("New Membership","Total Membership"),
          covariate.labels = c("Treatment Effect","Constant (Control Mean)"),
          type="latex",
          apply.p = one_tailed,
          add.lines = list(c("Percent Change (Rel. to Control)", percent_change_combined, percent_change
          keep.stat=c("n"),
          notes.append = T,
          no.space=TRUE,
          out="Table4.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:35
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{2}{c}{Outcome:} \\
## \cline{2-3}
## \\[-1.8ex] & New Membership & Total Membership \\
## \\[-1.8ex] & (1) & (2)\\
## \hline \\[-1.8ex]
##  Treatment Effect & 130.40$^{**}$ & 198.11$^{*}$ \\
##   & (67.70) & (143.45) \\
##   Constant (Control Mean) & 92.25$^{**}$ & 503.37$^{***}$ \\
```

```
##   & (51.10) & (62.73) \\
##  \hline \\[-1.8ex]
## Percent Change (Rel. to Control) & +141% & +39% \\
## Observations & 102 & 102 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{2}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
##Table D1:
stargazer(combined_padron_per_ipw, combined_padron_total_per_ipw, digits=2,
          dep.var.caption = "Outcome:",
          dep.var.labels = c("New Membership","Total Membership"),
          covariate.labels = c("Treatment Effect","Constant (Control Mean)"),
          type="latex",
          apply.p = one_tailed,
          add.lines = list(c("Percent Change (Rel. to Control)", percent_change_combined, percent_change
          keep.stat=c("n"),
          notes.append = T,
          no.space=TRUE,
          out="TableD1.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:35
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{2}{c}{Outcome:} \\
## \cline{2-3}
## \\[-1.8ex] & New Membership & Total Membership \\
## \\[-1.8ex] & (1) & (2)\\
## \hline \\[-1.8ex]
##  Treatment Effect & 130.40$^{**}$ & 198.11$^{**}$ \\
##   & (58.83) & (110.56) \\
##   Constant (Control Mean) & 92.25$^{**}$ & 503.37$^{***}$ \\
##   & (40.83) & (76.72) \\
##  \hline \\[-1.8ex]
## Percent Change (Rel. to Control) & +141% & +39% \\
## Observations & 102 & 102 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{2}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
##Table 5:
stargazer(pre_tombola_per_ipw_cl, combined_padron_per_control_ipw_cl, digits=2,
          dep.var.caption = "Outcome:",
          dep.var.labels = c("Baseline Membership","New Membership"),
          covariate.labels = c("Treatment Effect","Baseline Membership","Constant (Control Mean)"),
```

```
          type="latex",
          apply.p = one_tailed,
          keep.stat=c("n"),
          notes.append = T,
          no.space=TRUE,
          out="Table5.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:35
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{2}{c}{Outcome:} \\
## \cline{2-3}
## \\[-1.8ex] & Baseline Membership & New Membership \\
## \\[-1.8ex] & (1) & (2)\\
## \hline \\[-1.8ex]
##  Treatment Effect & 67.72 & 116.64$^{*}$ \\
##   & (144.07) & (74.19) \\
##   Baseline Membership &  & 0.20$^{**}$ \\
##   &  & (0.11) \\
##   Constant (Control Mean) & 411.12$^{***}$ & 8.70 \\
##   & (44.40) & (47.10) \\
##  \hline \\[-1.8ex]
## Observations & 102 & 102 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{2}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
##Table D2:
stargazer(pre_tombola_per_ipw, combined_padron_per_control_ipw, digits=2,
          dep.var.caption = "Outcome:",
          dep.var.labels = c("Baseline Membership","New Membership"),
          covariate.labels = c("Treatment Effect","Baseline Membership","Constant (Control Mean)"),
          type="latex",
          apply.p = one_tailed,
          keep.stat=c("n"),
          notes.append = T,
          no.space=TRUE,
          out="TableD2.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:35
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcc}
## \\[-1.8ex]\hline
```

```
## \hline \\[-1.8ex]
##  & \multicolumn{2}{c}{Outcome:} \\
## \cline{2-3}
## \\[-1.8ex] & Baseline Membership & New Membership \\
## \\[-1.8ex] & (1) & (2)\\
## \hline \\[-1.8ex]
##  Treatment Effect & 67.72 & 116.64$^{**}$ \\
##   & (78.93) & (57.10) \\
##   Baseline Membership &  & 0.20$^{***}$ \\
##   &  & (0.07) \\
##   Constant (Control Mean) & 411.12$^{***}$ & 8.70 \\
##   & (54.77) & (49.36) \\
##  \hline \\[-1.8ex]
## Observations & 102 & 102 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{2}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

```
###Downstream effects on vote choice:

##vote share MORENA:

#IPW:
share_morena_ipw_cl <- felm(share_morena_deputies ~ diputado_dummy | 0 | 0 | circunscripcion + tombola,

#FE Model:
share_morena_fe <- felm(share_morena_deputies ~ diputado_dummy | tombola + circunscripcion | 0 | estado

##vote share PRD:

#IPW:
share_prd_ipw_cl <- felm(share_pan_prd_deputies ~ diputado_dummy | 0 | 0 | circunscripcion + tombola, da

#FE Model:
share_prd_fe <- felm(share_pan_prd_deputies ~ diputado_dummy | tombola + circunscripcion | 0 | estado +


##vote share PRI:

#IPW:
share_pri_ipw_cl <- felm(share_pri_deputies ~ diputado_dummy | 0 | 0 | circunscripcion + tombola, data=
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
#FE Model:
share_pri_fe <- felm(share_pri_deputies ~ diputado_dummy | tombola + circunscripcion | 0 | estado, data=


##effect on margin-of-victory for MORENA at municipal level (over strongest other party):

for(l in 1:nrow(valid_diputado_dummy)){
```

```
  valid_diputado_dummy$MoV_MORENA_deputies[l] <- valid_diputado_dummy$share_morena_deputies[l] - max(c(v
}

#IPW:
MoV_morena_ipw_cl <- felm(MoV_MORENA_deputies ~ diputado_dummy | 0 | 0 | circunscripcion + tombola, data

#FE Model:
MoV_morena_fe <- felm(MoV_MORENA_deputies ~ diputado_dummy | tombola + circunscripcion | 0 | estado + to


##Effect on margin-of-victory for MORENA over PRD/PAN at municipal level:

for(l in 1:nrow(valid_diputado_dummy)){
  valid_diputado_dummy$MoV_MORENA_PRD_deputies[l] <- valid_diputado_dummy$share_morena_deputies[l] - val
}

#IPW:
MoV_morena_prd_ipw_cl <- felm(MoV_MORENA_PRD_deputies ~ diputado_dummy | 0 | 0 | circunscripcion + tombo

#FE Model:
MoV_morena_prd_fe <- felm(MoV_MORENA_PRD_deputies ~ diputado_dummy | tombola + circunscripcion | 0 | est


###Combine all results:

##main results (with clustered SEs):
findings_cl <- as.data.frame(rbind(rep(NA,10),rep(NA,10),rep(NA,10),rep(NA,10),rep(NA,10)))

findings_cl[1,] <- c("share_morena",summary(share_morena_ipw_cl)$coefficients[2,c(1,2,4)],summary(share_
findings_cl[2,] <- c("share_prd",summary(share_prd_ipw_cl)$coefficients[2,c(1,2,4)],summary(share_prd_ip
findings_cl[3,] <- c("share_pri",summary(share_pri_ipw_cl)$coefficients[2,c(1,2,4)],summary(share_pri_ip
findings_cl[4,] <- c("MoV_morena",summary(MoV_morena_ipw_cl)$coefficients[2,c(1,2,4)],summary(MoV_morena
findings_cl[5,] <- c("MoV_morena_prd",summary(MoV_morena_prd_ipw_cl)$coefficients[2,c(1,2,4)],summary(Mo
findings_cl[6,] <- c("padron_increase",summary(combined_padron_per_ipw_cl)$coefficients[2,c(1,2,4)],summ
findings_cl[7,] <- c("padron_total",summary(combined_padron_total_per_ipw_cl)$coefficients[2,c(1,2,4)],s
findings_cl[8,] <- c("padron_increase_control",summary(combined_padron_per_control_ipw_cl)$coefficients


colnames(findings_cl) <- c("Outcome","ATE (IPW)", "SE (IPW)","p-value (IPW)","Intercept (IPW)","n (IPW)"
findings_cl[,2] <- as.numeric(findings_cl[,2])
findings_cl[,3] <- as.numeric(findings_cl[,3])
findings_cl[,4] <- as.numeric(findings_cl[,4])
findings_cl[,5] <- as.numeric(findings_cl[,5])
findings_cl[,6] <- as.numeric(findings_cl[,6])
findings_cl[,7] <- as.numeric(findings_cl[,7])
findings_cl[,8] <- as.numeric(findings_cl[,8])
findings_cl[,9] <- as.numeric(findings_cl[,9])
findings_cl[,10] <- as.numeric(findings_cl[,10])

findings_cl
```

```
##                  Outcome    ATE (IPW)     SE (IPW) p-value (IPW)
## 1            share_morena   0.05095607 2.682944e-02  6.041229e-02
## 2               share_prd  -0.04719793 3.112350e-02  1.325549e-01
```

```
## 3              share_pri  -0.01144321 2.364679e-03  4.744803e-06
## 4             MoV_morena   0.08870984 3.738690e-02  1.956847e-02
## 5          MoV_morena_prd   0.09815400 5.871950e-02  9.773485e-02
## 6         padron_increase 130.39948877 6.770030e+01  5.692801e-02
## 7            padron_total 198.11464071 1.434528e+02  1.703440e-01
## 8 padron_increase_control 116.63842842 7.418887e+01  1.190989e-01
##   Intercept (IPW) n (IPW)      ATE (FE)      SE (FE) p-value (FE) n (FE)
## 1       0.4707472     102   0.034897441   0.01381598   0.08573364    102
## 2       0.2890946     102  -0.040579070   0.02267344   0.17143737    102
## 3       0.2277320     102   0.006419836   0.02500188   0.79792036    102
## 4       0.1498878     102   0.060233147   0.01762515   0.04192103    102
## 5       0.1816526     102   0.075476511   0.03908116   0.14897971    102
## 6      92.2520518     102 113.206336348  55.85793653   0.04555829    102
## 7     503.3740043     102 179.344404871 137.61360366   0.19570805    102
## 8       8.7039266     102 102.573945941   0.16076052   0.05052300    102
```

```
##share_morena:
results_share_morena <- as.data.frame(rbind("Control","Treatment"))
results_share_morena$Estimates <- rbind(findings_cl[1,]$`Intercept (IPW)`*100, (findings_cl[1,]$`Interc
colnames(results_share_morena)[1] <- "Group"
results_share_morena$Estimates <- round(results_share_morena$Estimates,2)
results_share_morena
```

```
##       Group Estimates
## 1   Control     47.07
## 2 Treatment     52.17
```

```
##Figure 5:
pd <- position_dodge(0.1)
pdf("fg5.pdf", width = 5, height = 5)
share_morena_figure <- ggplot(results_share_morena, aes(x=factor(Group), y=Estimates, fill=factor(Group)
  geom_bar(position=pd, stat="identity", width=0.45, alpha=0.9) +
  geom_text(aes(label = Estimates), size = 4.0, vjust=-0.5, hjust = 0.75, nudge_x = 0.05) +
  ggtitle(" ") +
  xlab("") +
  ylab("Vote Share (%)") +
  ylim(0, 100) +
  theme_minimal() +
  theme(axis.text.y=element_text(colour = "gray30", size=11), axis.ticks.y=element_blank()) +
  theme(axis.text.x=element_text(colour = "gray30", size=11), axis.ticks.x=element_line(colour="gray30"
share_morena_figure<- share_morena_figure + geom_segment(aes(x=1, y=70, xend=1.2, yend=73), colour="dark
  geom_segment(aes(x=1.2, y=73, xend=1.8, yend=73), colour="darkgray") +
  geom_segment(aes(x=1.8, y=73, xend=2, yend=70), colour="darkgray")
share_morena_figure <- share_morena_figure + annotate("text", x=1.5, y=76, label="+5.10%p (p=0.0302)", s

share_morena_figure <- share_morena_figure + scale_fill_manual(values=c("cornflowerblue", "darkred"))
share_morena_figure <- share_morena_figure + theme(legend.position="none")
print(share_morena_figure)
dev.off()
```

```
## pdf
##   2
```

```
##Table with vote choice findings:
results_table <- findings_cl[c(2:5),]
results_table$Control <- results_table$`Intercept (IPW)`
```

```
results_table$Treatment <- results_table$`Intercept (IPW)` + results_table$`ATE (IPW)`

results_table <- results_table[,c(1,11:12,2:4,6)]
results_table$Control <- round(results_table$Control,4)
results_table$Treatment <- round(results_table$Treatment,4)
results_table$`ATE (IPW)` <- round(results_table$`ATE (IPW)`,4)
results_table$`SE (IPW)` <- round(results_table$`SE (IPW)`,4)
results_table$`p-value (IPW)` <- round(results_table$`p-value (IPW)`/2,4)

results_table$Outcome <- c("Vote Share for PRD","Vote Share for PRI","MoV: MORENA over Runner-up", "MoV
names(results_table)[c(4:7)] <- c("ATE","SE","p-value","n")

results_table
```

```
##                       Outcome Control Treatment     ATE     SE p-value   n
## 2         Vote Share for PRD  0.2891    0.2419 -0.0472 0.0311  0.0663 102
## 3         Vote Share for PRI  0.2277    0.2163 -0.0114 0.0024  0.0000 102
## 4 MoV: MORENA over Runner-up  0.1499    0.2386  0.0887 0.0374  0.0098 102
## 5   MoV: MORENA over PRD/PAN  0.1817    0.2798  0.0982 0.0587  0.0489 102
```

```
#Table 6:
stargazer(as.matrix(results_table), rownames = F,
          out="Table6.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:36
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}} ccccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Outcome & Control & Treatment & ATE & SE & p-value & n \\
## \hline \\[-1.8ex]
## Vote Share for PRD & 0.2891 & 0.2419 & -0.0472 & 0.0311 & 0.0663 & 102 \\
## Vote Share for PRI & 0.2277 & 0.2163 & -0.0114 & 0.0024 & 0.0000 & 102 \\
## MoV: MORENA over Runner-up & 0.1499 & 0.2386 &  0.0887 & 0.0374 & 0.0098 & 102 \\
## MoV: MORENA over PRD/PAN & 0.1817 & 0.2798 &  0.0982 & 0.0587 & 0.0489 & 102 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

```
##Table with replication of findings with FEs:
results_table_fe <- findings_cl[c(1:5),]

results_table_fe <- results_table_fe[,c(1,7:10)]
results_table_fe$`ATE (FE)` <- round(results_table_fe$`ATE (FE)`,4)
results_table_fe$`SE (FE)` <- round(results_table_fe$`SE (FE)`,4)
results_table_fe$`p-value (FE)` <- round(results_table_fe$`p-value (FE)`/2,4)

results_table_fe$Outcome <- c("Vote Share for MORENA","Vote Share for PRD","Vote Share for PRI","MoV: M
names(results_table_fe)[c(2:5)] <- c("ATE","SE","p-value","n")

results_table_fe
```

```
##                       Outcome     ATE     SE p-value   n
## 1      Vote Share for MORENA  0.0349 0.0138  0.0429 102
## 2         Vote Share for PRD -0.0406 0.0227  0.0857 102
## 3         Vote Share for PRI  0.0064 0.0250  0.3990 102
## 4 MoV: MORENA over Runner-up  0.0602 0.0176  0.0210 102
## 5   MoV: MORENA over PRD/PAN  0.0755 0.0391  0.0745 102
```

```
#Table E2:
stargazer(as.matrix(results_table_fe), rownames = F,
          out="TableE2.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:36
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}} ccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Outcome & ATE & SE & p-value & n \\
## \hline \\[-1.8ex]
## Vote Share for MORENA &  0.0349 & 0.0138 & 0.0429 & 102 \\
## Vote Share for PRD & -0.0406 & 0.0227 & 0.0857 & 102 \\
## Vote Share for PRI &  0.0064 & 0.0250 & 0.3990 & 102 \\
## MoV: MORENA over Runner-up &  0.0602 & 0.0176 & 0.0210 & 102 \\
## MoV: MORENA over PRD/PAN &  0.0755 & 0.0391 & 0.0745 & 102 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

```
##Table with replication of main findings with FEs:
padron_table_fe <- findings_cl[c(6:7),]

padron_table_fe <- padron_table_fe[,c(1,7:10)]
padron_table_fe$`ATE (FE)` <- round(padron_table_fe$`ATE (FE)`,4)
padron_table_fe$`SE (FE)` <- round(padron_table_fe$`SE (FE)`,4)
padron_table_fe$`p-value (FE)` <- round(padron_table_fe$`p-value (FE)`/2,4)

padron_table_fe$Outcome <- c("Total MORENA Members","New MORENA Members")
names(padron_table_fe)[c(2:5)] <- c("ATE","SE","p-value","n")

padron_table_fe
```

```
##                Outcome      ATE       SE p-value   n
## 6 Total MORENA Members 113.2063  55.8579  0.0228 102
## 7   New MORENA Members 179.3444 137.6136  0.0979 102
```

```
#Table E1:
stargazer(as.matrix(padron_table_fe), rownames = F,
          out="TableE1.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:36
## \begin{table}[!htbp] \centering
##   \caption{}
```

```
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}} ccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Outcome & ATE & SE & p-value & n \\
## \hline \\[-1.8ex]
## Total MORENA Members & 113.2063 &  55.8579 & 0.0228 & 102 \\
## New MORENA Members & 179.3444 & 137.6136 & 0.0979 & 102 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

```
###Effect of access to office on local party institutionalization by whether municipality was (also) re

##Effect on proportion of registered voters who affiliated with the party AFTER to the lottery per 100,
combined_padron_per_mult_elect_0_ipw_cl <- felm(combined_padron_per ~ diputado_dummy | 0 | 0 | circunsc
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
combined_padron_per_mult_elect_1_ipw_cl <- felm(combined_padron_per ~ diputado_dummy | 0 | 0 | circunsc
combined_padron_per_mult_elect_interact_ipw_cl <- felm(combined_padron_per ~ diputado_dummy * mult_elec
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
##Effect on total proportion of registered voters who affiliated by the end of 2018 per 100,000:
combined_padron_total_per_mult_elect_0_ipw_cl <- felm(combined_padron_total_per ~ diputado_dummy | 0 | 
combined_padron_total_per_mult_elect_1_ipw_cl <- felm(combined_padron_total_per ~ diputado_dummy | 0 | 
combined_padron_total_per_mult_elect_interact_ipw_cl <- felm(combined_padron_total_per ~ diputado_dummy
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
##combining results:
findings_mult_elect_cl <- as.data.frame(rbind(rep(NA,6),rep(NA,6),rep(NA,6),rep(NA,6),rep(NA,6),rep(NA,

findings_mult_elect_cl[1,] <- c("combined_padron_per_mult_elect_0",summary(combined_padron_per_mult_ele
findings_mult_elect_cl[2,] <- c("combined_padron_per_mult_elect_1",summary(combined_padron_per_mult_ele
findings_mult_elect_cl[3,] <- c("combined_padron_per_mult_elect_interact",summary(combined_padron_per_m

findings_mult_elect_cl[4,] <- c("combined_padron_total_per_mult_elect_0",summary(combined_padron_total_
findings_mult_elect_cl[5,] <- c("combined_padron_total_per_mult_elect_1",summary(combined_padron_total_
findings_mult_elect_cl[6,] <- c("combined_padron_total_per_mult_elect_interact",summary(combined_padron

findings_mult_elect_cl$other_mp <- c("No","Yes","Difference","No","Yes","Difference")

findings_mult_elect_cl <- findings_mult_elect_cl[,c(1,7,2:4,6)]

colnames(findings_mult_elect_cl) <- c("Outcome","Other MP","ATE", "SE","p-value","n")
findings_mult_elect_cl[,3] <- as.numeric(findings_mult_elect_cl[,3])
findings_mult_elect_cl[,4] <- as.numeric(findings_mult_elect_cl[,4])
findings_mult_elect_cl[,5] <- as.numeric(findings_mult_elect_cl[,5])
findings_mult_elect_cl[,6] <- as.numeric(findings_mult_elect_cl[,6])
```

```r
findings_mult_elect_cl$`p-value` <- findings_mult_elect_cl$`p-value` #one-tailed tests

findings_mult_elect_cl$ATE <- round(findings_mult_elect_cl$ATE,4)
findings_mult_elect_cl$SE <- round(findings_mult_elect_cl$SE,4)
findings_mult_elect_cl$`p-value` <- round(findings_mult_elect_cl$`p-value`,4)

findings_mult_elect_cl$Outcome <- c("New Members","New Members","New Members","Total Members","Total Mem

findings_mult_elect_cl
```

```
##         Outcome   Other MP       ATE       SE p-value   n
## 1   New Members         No  192.6219  22.2192  0.0000  75
## 2   New Members        Yes   88.7294 103.6066  0.3999  27
## 3   New Members Difference -103.8925 122.4043  0.3981 102
## 4 Total Members         No  182.6354  92.3761  0.0518  75
## 5 Total Members        Yes  329.6595 131.5554  0.0191  27
## 6 Total Members Difference  147.0241 169.2624  0.3872 102
```

```r
#Table G1:
stargazer(as.matrix(findings_mult_elect_cl), rownames = F,
          out="TableG1.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:36
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}} cccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Outcome & Other MP & ATE & SE & p-value & n \\
## \hline \\[-1.8ex]
## New Members & No &  192.6219 &  22.2192 & 0.0000 &  75 \\
## New Members & Yes &   88.7294 & 103.6066 & 0.3999 &  27 \\
## New Members & Difference & -103.8925 & 122.4043 & 0.3981 & 102 \\
## Total Members & No &  182.6354 &  92.3761 & 0.0518 &  75 \\
## Total Members & Yes &  329.6595 & 131.5554 & 0.0191 &  27 \\
## Total Members & Difference &  147.0241 & 169.2624 & 0.3872 & 102 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

```r
###Balance statistics for other municipality characteristics:
### drawing on census & CONEVAL data:

##import data with information on municipality characteristics:
mun_charact <- read.csv("mun_charact.csv", header=T, row.names=1, sep=",") #import data with informatio

##some data cleaning:
mun_charact$uid_join <- paste0(mun_charact$estado,"_",mun_charact$municipio) #create join variable to m

valid_diputado_dummy$uid_join <- paste0(valid_diputado_dummy$estado,"_",valid_diputado_dummy$municipio_i
valid_diputado_dummy <- left_join(valid_diputado_dummy,mun_charact, by="uid_join",copy=F)
```

```
names(valid_diputado_dummy)[names(valid_diputado_dummy)=="X.sector.primario"] <- "sector.primario"
names(valid_diputado_dummy)[names(valid_diputado_dummy)=="X.sector.comercio"] <- "sector.comercio"
names(valid_diputado_dummy)[names(valid_diputado_dummy)=="X.sector.servicios"] <- "sector.servicios"
names(valid_diputado_dummy)[names(valid_diputado_dummy)=="X.Trabajadores.de.la.Industria"] <- "Trabajad

##Industrial workers:
Trabajadores.de.la.Industria_ipw <- felm(Trabajadores.de.la.Industria ~ diputado_dummy | 0 | 0 | circuns
Trabajadores.de.la.Industria_fe <- felm(Trabajadores.de.la.Industria ~ diputado_dummy | tombola + circu

##Industrial workers:
sector.primario_ipw <- felm(sector.primario ~ diputado_dummy | 0 | 0 | circunscripcion + tombola, data=
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
sector.primario_fe <- felm(sector.primario ~ diputado_dummy | tombola + circunscripcion | 0 | estado.x

##Commercial sector workers:
sector.comercio_ipw <- felm(sector.comercio ~ diputado_dummy | 0 | 0 | circunscripcion + tombola, data=
sector.comercio_fe <- felm(sector.comercio ~ diputado_dummy | tombola + circunscripcion | 0 | estado.x

##Service sector workers:
sector.servicios_ipw <- felm(sector.servicios ~ diputado_dummy | 0 | 0 | circunscripcion + tombola, dat
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
sector.servicios_fe <- felm(sector.servicios ~ diputado_dummy | tombola + circunscripcion | 0 | estado.

##Different age segments of the population:

#creating variables for proportion of population ages 0-29, 30-49, and 50+:
valid_diputado_dummy$poblacion.00.29.anos <- valid_diputado_dummy$poblacion.00.04.anos + valid_diputado
valid_diputado_dummy$poblacion.30.49.anos <- valid_diputado_dummy$poblacion.30.34.anos + valid_diputado
valid_diputado_dummy$poblacion.50.anos.y.mas <- valid_diputado_dummy$poblacion.50.54.anos + valid_diputa
valid_diputado_dummy$total_population <- valid_diputado_dummy$poblacion.00.29.anos + valid_diputado_dumm
valid_diputado_dummy$poblacion.00.29.anos <- valid_diputado_dummy$poblacion.00.29.anos / valid_diputado
valid_diputado_dummy$poblacion.30.49.anos <- valid_diputado_dummy$poblacion.30.49.anos / valid_diputado
valid_diputado_dummy$poblacion.50.anos.y.mas <- valid_diputado_dummy$poblacion.50.anos.y.mas / valid_di

#analysis:
poblacion.00.29.anos_ipw <- felm(poblacion.00.29.anos ~ diputado_dummy | 0 | 0 | circunscripcion + tombo
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
poblacion.00.29.anos_fe <- felm(poblacion.00.29.anos ~ diputado_dummy | tombola + circunscripcion | 0 |

poblacion.30.49.anos_ipw <- felm(poblacion.30.49.anos ~ diputado_dummy | 0 | 0 | circunscripcion + tombo
poblacion.30.49.anos_fe <- felm(poblacion.30.49.anos ~ diputado_dummy | tombola + circunscripcion | 0 |

poblacion.50.anos.y.mas_ipw <- felm(poblacion.50.anos.y.mas ~ diputado_dummy | 0 | 0 | circunscripcion
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
poblacion.50.anos.y.mas_fe <- felm(poblacion.50.anos.y.mas ~ diputado_dummy | tombola + circunscripcion

##Income vulnerable population (percent):
Porcentaje.de.poblacion.vulnerable.por.ingresos_ipw <- felm(Porcentaje.de.poblacion.vulnerable.por.ingre
Porcentaje.de.poblacion.vulnerable.por.ingresos_fe <- felm(Porcentaje.de.poblacion.vulnerable.por.ingres


##combine results:
balance <- as.data.frame(rbind(rep(NA,10),rep(NA,10),rep(NA,10),rep(NA,10),rep(NA,10),rep(NA,10),rep(NA

balance[1,] <- c("poblacion.00.29.anos",summary(poblacion.00.29.anos_ipw)$coefficients[2,c(1,2,4)],summa
balance[2,] <- c("poblacion.30.49.anos",summary(poblacion.30.49.anos_ipw)$coefficients[2,c(1,2,4)],summa
balance[3,] <- c("poblacion.50.anos.y.mas",summary(poblacion.50.anos.y.mas_ipw)$coefficients[2,c(1,2,4)
balance[4,] <- c("sector.primario",summary(sector.primario_ipw)$coefficients[2,c(1,2,4)],summary(sector
balance[5,] <- c("Trabajadores.de.la.Industria",summary(Trabajadores.de.la.Industria_ipw)$coefficients[
balance[6,] <- c("sector.comercio",summary(sector.comercio_ipw)$coefficients[2,c(1,2,4)],summary(sector
balance[7,] <- c("sector.servicios",summary(sector.servicios_ipw)$coefficients[2,c(1,2,4)],summary(secto
balance[8,] <- c("Porcentaje.de.poblacion.vulnerable.por.ingresos",summary(Porcentaje.de.poblacion.vulne
balance[9,] <- c("MORENA Members (Baseline)",summary(pre_tombola_percent_ipw_cl)$coefficients[2,c(1,2,4


colnames(balance) <- c("Outcome","ATE (IPW)", "SE (IPW)","p-value (IPW)","Intercept (IPW)","n (IPW)","A
balance[,2] <- as.numeric(balance[,2])
balance[,3] <- as.numeric(balance[,3])
balance[,4] <- as.numeric(balance[,4])
balance[,5] <- as.numeric(balance[,5])
balance[,6] <- as.numeric(balance[,6])
balance[,7] <- as.numeric(balance[,7])
balance[,8] <- as.numeric(balance[,8])
balance[,9] <- as.numeric(balance[,9])
balance[,10] <- as.numeric(balance[,10])

balance
```

```
##                                       Outcome     ATE (IPW)    SE (IPW)
## 1                          poblacion.00.29.anos  0.0046051578 0.019700908
## 2                          poblacion.30.49.anos -0.0120887119 0.009933246
## 3                        poblacion.50.anos.y.mas  0.0074835541 0.026150553
## 4                              sector.primario -0.3652503616 0.630412066
## 5                  Trabajadores.de.la.Industria -0.3752941724 2.087310066
## 6                              sector.comercio  2.8914432972 1.725461805
## 7                            sector.servicios  1.4872284936 1.960232276
## 8 Porcentaje.de.poblacion.vulnerable.por.ingresos  1.2894034245 1.152577186
## 9                      MORENA Members (Baseline)  0.0006771515 0.001440654
##   p-value (IPW) Intercept (IPW) n (IPW)      ATE (FE)    SE (FE) p-value (FE)
## 1     0.8156542      0.38233467     102 -0.0003534719 0.06741676    0.9961458
## 2     0.2264704      0.26187058     102 -0.0152162782 0.03574706    0.6990428
## 3     0.7753384      0.35579476     102  0.0155697501 0.10228519    0.8886752
## 4     0.5636334     21.32097601     102 -1.6540474603 6.49991218    0.8155768
## 5     0.8576750     21.76771962     102 -0.0154542983 0.64832909    0.9824794
## 6     0.0969112     16.06660839     102  1.7624126792 1.22156635    0.2447849
```

```
## 7     0.4498165     34.78110153     102  1.5027517258 5.48200113    0.8017804
## 8     0.2659426      7.66884707     102  1.3073800198 0.62162399    0.1261841
## 9     0.6393567      0.00411122     102  0.0006613807 0.00135144    0.6581155
##   n (FE)
## 1    102
## 2    102
## 3    102
## 4    102
## 5    102
## 6    102
## 7    102
## 8    102
## 9    102
```

```
##table with balance tests:
balance_ipw <- balance
balance_ipw$Control <- balance_ipw$`Intercept (IPW)`
balance_ipw$Treatment <- balance_ipw$`Intercept (IPW)` + balance_ipw$`ATE (IPW)`

balance_ipw <- balance_ipw[,c(1,11:12,2:4,6)]
balance_ipw[c(4:8),c(2:5)] <- balance_ipw[c(4:8),c(2:5)]/100
balance_ipw$Control <- round(balance_ipw$Control,4)
balance_ipw$Treatment <- round(balance_ipw$Treatment,4)
balance_ipw$`ATE (IPW)` <- round(balance_ipw$`ATE (IPW)`,4)
balance_ipw$`SE (IPW)` <- round(balance_ipw$`SE (IPW)`,4)
balance_ipw$`p-value (IPW)` <- round(balance_ipw$`p-value (IPW)`,4)

balance_ipw$Outcome <- c("Population Ages 0-29","Population Ages 30-49","Population Ages 50+","Primary S
                         "Industrial Workers","Commercial Sector Workers","Service Sector Workers","Inco
                         "MORENA Members (Baseline)")

names(balance_ipw)[c(4:7)] <- c("ATE","SE","p-value","n")

balance_ipw
```

```
##                        Outcome Control Treatment     ATE     SE p-value   n
## 1         Population Ages 0-29  0.3823    0.3869  0.0046 0.0197  0.8157 102
## 2        Population Ages 30-49  0.2619    0.2498 -0.0121 0.0099  0.2265 102
## 3          Population Ages 50+  0.3558    0.3633  0.0075 0.0262  0.7753 102
## 4       Primary Sector Workers  0.2132    0.2096 -0.0037 0.0063  0.5636 102
## 5           Industrial Workers  0.2177    0.2139 -0.0038 0.0209  0.8577 102
## 6    Commercial Sector Workers  0.1607    0.1896  0.0289 0.0173  0.0969 102
## 7       Service Sector Workers  0.3478    0.3627  0.0149 0.0196  0.4498 102
## 8 Income Vulnerable Population  0.0767    0.0896  0.0129 0.0115  0.2659 102
## 9    MORENA Members (Baseline)  0.0041    0.0048  0.0007 0.0014  0.6394 102
```

```
#Table 3:
stargazer(as.matrix(balance_ipw), rownames = F,
          out="Table3.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:37
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
```

```
## \begin{tabular}{@{\extracolsep{5pt}} ccccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Outcome & Control & Treatment & ATE & SE & p-value & n \\
## \hline \\[-1.8ex]
## Population Ages 0-29 & 0.3823 & 0.3869 &  0.0046 & 0.0197 & 0.8157 & 102 \\
## Population Ages 30-49 & 0.2619 & 0.2498 & -0.0121 & 0.0099 & 0.2265 & 102 \\
## Population Ages 50+ & 0.3558 & 0.3633 &  0.0075 & 0.0262 & 0.7753 & 102 \\
## Primary Sector Workers & 0.2132 & 0.2096 & -0.0037 & 0.0063 & 0.5636 & 102 \\
## Industrial Workers & 0.2177 & 0.2139 & -0.0038 & 0.0209 & 0.8577 & 102 \\
## Commercial Sector Workers & 0.1607 & 0.1896 &  0.0289 & 0.0173 & 0.0969 & 102 \\
## Service Sector Workers & 0.3478 & 0.3627 &  0.0149 & 0.0196 & 0.4498 & 102 \\
## Income Vulnerable Population & 0.0767 & 0.0896 &  0.0129 & 0.0115 & 0.2659 & 102 \\
## MORENA Members (Baseline) & 0.0041 & 0.0048 &  0.0007 & 0.0014 & 0.6394 & 102 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

```
#F-Test:
f_test_balance <- felm(diputado_dummy ~ poblacion.30.49.anos + poblacion.50.anos.y.mas + sector.primari
```

```
## Warning in newols(mm, nostats = nostats[1], exactDOF = exactDOF, onlyse =
## onlyse, : Negative eigenvalues set to zero in multiway clustered variance
## matrix. See felm(...,psdef=FALSE)
```

```
#omitted age category: poblacion.00.29.anos
summary(f_test_balance)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```

```
##
## Call:
##    felm(formula = diputado_dummy ~ poblacion.30.49.anos + poblacion.50.anos.y.mas +      sector.prima
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8063 -0.5508 -0.4489 -0.2633  2.2579
##
## Coefficients:
##                                              Estimate Cluster s.e. t value
## (Intercept)                                 -0.292186     0.303438  -0.963
## poblacion.30.49.anos                        -0.412724     0.353984  -1.166
## poblacion.50.anos.y.mas                      0.066358     0.229132   0.290
## sector.primario                              0.006395     0.004907   1.303
## sector.comercio                              0.017867     0.007251   2.464
## sector.servicios                             0.003694     0.006404   0.577
## Porcentaje.de.poblacion.vulnerable.por.ingresos  0.022081     0.017291   1.277
## pre_tombola_percent                         21.234027    14.175997   1.498
##                                             Pr(>|t|)
## (Intercept)                                   0.3381
## poblacion.30.49.anos                          0.2466
## poblacion.50.anos.y.mas                       0.7728
## sector.primario                               0.1957
## sector.comercio                               0.0156 *
## sector.servicios                              0.5654
```

```
## Porcentaje.de.poblacion.vulnerable.por.ingresos   0.2047
## pre_tombola_percent                               0.1375
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.683 on 94 degrees of freedom
## Multiple R-squared(full model): 0.1032   Adjusted R-squared: 0.03637
## Multiple R-squared(proj model): 0.1032   Adjusted R-squared: 0.03637
## F-statistic(full model, *iid*):1.545 on 7 and 94 DF, p-value: 0.1619
## F-statistic(proj model): 3.739e+05 on 7 and 3 DF, p-value: 6.666e-09
```

```
##table with balance tests with FE:
balance_fe <- balance
balance_fe <- balance_fe[,c(1,7:10)]
balance_fe[c(4:8),c(2:3)] <- balance_fe[c(4:8),c(2:3)]/100

balance_fe$`ATE (FE)` <- round(balance_fe$`ATE (FE)`,4)
balance_fe$`SE (FE)` <- round(balance_fe$`SE (FE)`,4)
balance_fe$`p-value (FE)` <- round(balance_fe$`p-value (FE)`,4)

balance_fe$Outcome <- c("Population Ages 0-29","Population Ages 30-49","Population Ages 50+","Primary Se
                        "Industrial Workers","Commercial Sector Workers","Service Sector Workers","Incom
                        "MORENA Members (Baseline)")
names(balance_fe)[c(2:5)] <- c("ATE","SE","p-value","n")
balance_fe
```

```
##                         Outcome     ATE     SE p-value   n
## 1         Population Ages 0-29 -0.0004 0.0674  0.9961 102
## 2        Population Ages 30-49 -0.0152 0.0357  0.6990 102
## 3          Population Ages 50+  0.0156 0.1023  0.8887 102
## 4       Primary Sector Workers -0.0165 0.0650  0.8156 102
## 5           Industrial Workers -0.0002 0.0065  0.9825 102
## 6    Commercial Sector Workers  0.0176 0.0122  0.2448 102
## 7       Service Sector Workers  0.0150 0.0548  0.8018 102
## 8 Income Vulnerable Population  0.0131 0.0062  0.1262 102
## 9    MORENA Members (Baseline)  0.0007 0.0014  0.6581 102
```

```
#Table C1:
stargazer(as.matrix(balance_fe), rownames = F,
          out="TableC1.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:29:37
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}} ccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Outcome & ATE & SE & p-value & n \\
## \hline \\[-1.8ex]
## Population Ages 0-29 & -0.0004 & 0.0674 & 0.9961 & 102 \\
## Population Ages 30-49 & -0.0152 & 0.0357 & 0.6990 & 102 \\
## Population Ages 50+ &  0.0156 & 0.1023 & 0.8887 & 102 \\
## Primary Sector Workers & -0.0165 & 0.0650 & 0.8156 & 102 \\
```

```
## Industrial Workers & -0.0002 & 0.0065 & 0.9825 & 102 \\
## Commercial Sector Workers &  0.0176 & 0.0122 & 0.2448 & 102 \\
## Service Sector Workers &  0.0150 & 0.0548 & 0.8018 & 102 \\
## Income Vulnerable Population &  0.0131 & 0.0062 & 0.1262 & 102 \\
## MORENA Members (Baseline) &  0.0007 & 0.0014 & 0.6581 & 102 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

```
##Equivalance tests:

##adapting code from Hartman and Hidalgo (2018)

####################
#### Functions ####
####################
equiv.t.test <- function(x, y, alpha = .1, epsilon = .2, std.err = "nominal", cluster.x = NULL, cluster
  x = x[!is.na(x)]
  y = y[!is.na(y)]


  dbar <- mean(x) - mean(y)
  m <- as.double(length(x))
  n <- as.double(length(y))
  N <- m+n
  x.var <-  var(x)
  y.var <- var(y)
  non.cent <- (m*n*epsilon^2)/N
  critical.const <- sqrt(qf(alpha,1,N-2,non.cent))

  se = sqrt((m-1)*x.var + (n-1)*y.var) / sqrt(m*n * (N-2)/N)
  df = N - 2

  t.stat <-  dbar / se
  p = pf(abs(t.stat)^2, 1, df , non.cent)
  obs_smd = (mean(x) - mean(y)) / sd(y)
  inverted <- try(uniroot(function(x) pf(abs(t.stat)^2, 1, N-2, ncp = (m*n*x^2)/N) - ifelse(pf(abs(t.sta
  if(class(inverted) == "try-error") {
    inverted = NA
  }
  rej = abs(t.stat) <= critical.const
  return(list(t.stat = t.stat, critical.const = critical.const, power = 2*pt(critical.const, N-2)-1, re
}



# a helper function to run the equivalence tests on all balance covariates
run_equiv = function(X, Tr, epsilon.method = "std.effect", Y = NULL,
                     custom.epsilon = NULL, std.err = "nominal", type = "equiv.t.test",
                     fdr_correct = FALSE, cluster.x = NULL, cluster.y = NULL) {
  switch(epsilon.method,
         std.effect = {
           tol = abs(mean(Y[Tr == 1], na.rm = TRUE) - mean(Y[Tr == 0], na.rm = TRUE)) / sd(Y[Tr == 0], r
         },
```

```
        custom = {
          if(is.null(custom.epsilon)) stop("ERROR: Must enter a custom epsilon value to use the 'custo
          tol = custom.epsilon
        },
        strict = {
          tol = 0.36
        },
        liberal = {
          tol = 0.74
        },
        stop("ERROR: 'epsilon.method' not set to a valid option")
)

ranges = rep(tol, ncol(X))
names(ranges) = names(X)
print(ranges)

# conduct equiv test for each variable
tests = do.call("rbind", lapply(names(X), function(var) {
  print(var)

  res = equiv.t.test(X[Tr == 1, var], X[Tr == 0, var], epsilon = ranges[var], std.err = std.err, clus
  res$stat = res$t.stat
  res$obs_diff = (mean(X[Tr == 1, var], na.rm = TRUE) - mean(X[Tr == 0, var], na.rm = TRUE))
  res$obs_smd = ((mean(X[Tr == 1, var], na.rm = TRUE) - mean(X[Tr == 0, var], na.rm = TRUE))) / sd(X[
  res$sd = sd(X[Tr == 0, var], na.rm = TRUE)

  return(c(res$inverted
           , round(res$p, 4)
           , res$obs_smd
           , res$obs_diff
           , res$power
           , res$sd))

}))

p.vals = unlist(tests[,2])
names(p.vals) = names(ranges)

power = unlist(tests[,5])
names(power) = names(ranges)

inverted = unlist(tests[,1])
names(inverted) = names(ranges)

sd = unlist(tests[,6])
names(sd) = names(ranges)

# return to scale of var
inverted.scaled = unlist(lapply(names(inverted), function(var) {
  return(inverted[var] * sd[var])
}))
names(inverted.scaled) = names(ranges)
```

```
  observed.smd = unlist(tests[,3])
  names(observed.smd) = names(ranges)

  observed.diff = unlist(tests[,4])
  names(observed.diff) = names(ranges)

  # conduct BH FDR adjustment
  if(fdr_correct) {
    p.vals = p.adjust(p.vals, method = "BH")
  }

  return(list(tol = ranges, inverted = inverted, inverted.scaled = inverted.scaled, p.vals = p.vals, obs
}

## a helper function to generate the plot
generate_plot = function(equiv_tests, panel.widths=c(1, 1, 5, 1, 1), display.names = NULL, var.rounding
  .e <- environment()
  if(!is.null(display.names)) {
    equiv_tests$names = factor(display.names, levels = rev(display.names))
  } else {
    equiv_tests$names = factor(row.names(equiv_tests), levels = rev(row.names(equiv_tests)))
  }
  equiv_tests$const = rep(1, nrow(equiv_tests))
  equiv_tests = equiv_tests[nrow(equiv_tests):1, ]

  g = ggplot(equiv_tests, aes(x = names) )
  print(length(unique(equiv_tests$tol)))
  g = g + geom_linerange(aes(ymin = -1 * inverted, ymax = inverted), size = 5, color = "darkgray", alpha
  if(length(unique(equiv_tests$tol)) > 1){
    g = g + geom_linerange(aes(ymin = -1 * tol, ymax = tol), size = 10, color = 'gray')
  } else {
    print("here")
    print(unique(equiv_tests$tol))
    lines = unique(equiv_tests$tol)
    g = g + geom_hline(yintercept = c(-1 * lines, lines), linetype = 2, size = 0.75)
  }
  g = g  + scale_shape_identity() + geom_point(aes(y = observed.smd), color = "black", shape = 18, size
  g = (g + coord_flip()
       + theme(
         axis.text.y = element_blank()
         , axis.ticks.y = element_blank()
         , axis.title.y = element_blank()
         , axis.text.x = element_text(size = 14)
         , axis.title.x = element_text(size = 14))
       + labs(x=NULL, y = paste("Equivalence Range (in standard deviations \u03C3)"), font=5) + theme_bw
  )
  if(length(unique(equiv_tests$tol)) == 1) {
    g = g + ggtitle(paste0("Equivalence Tests  \n \n", "Equivalence Range: +/- ", round(unique(equiv_tes
  } else {
    g = g + ggtitle("Equivalence Tests  \n \n \n")
  }

  g_inv = ggplot(equiv_tests, aes(x = names, y = const, label = round(inverted.scaled, var.rounding)),
```

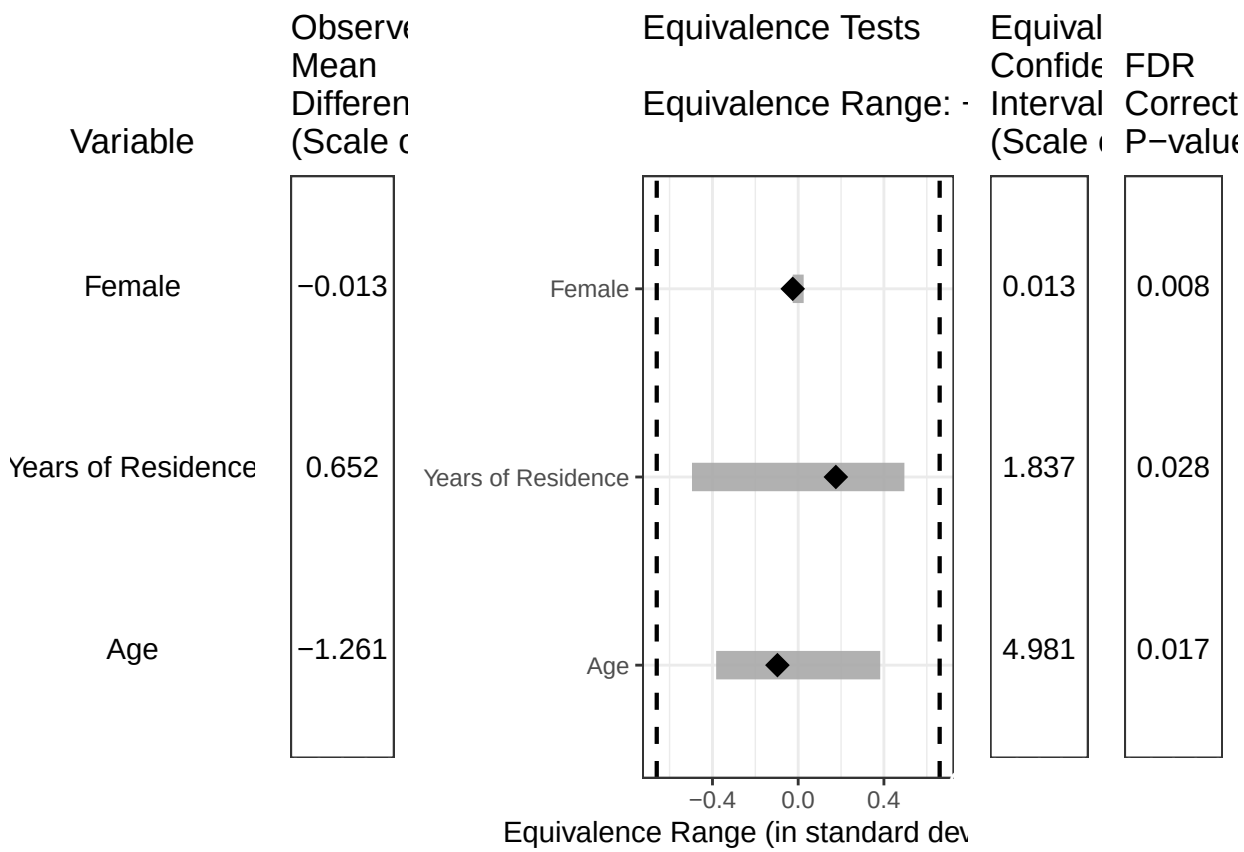
```
g_inv = g_inv + geom_text()
g_inv = (g_inv + theme_bw() + coord_flip()
         + theme(panel.grid.minor=element_blank()
                 , panel.grid.major=element_blank()
                 #, axis.line.x = element_blank()
                 , axis.text.x = element_text(color = "white", size = 14)
                 , axis.ticks.x = element_line(color = "white")
                 , axis.text.y = element_blank()
                 , axis.ticks.y = element_blank()
                 , axis.title.x = element_text(size = 14)
         )
         + ylim(1-.05, 1.05)
         + ggtitle("Equivalence\nConfidence\nInterval (+/-)\n(Scale of Var)")
         + labs(y = " ", x = NULL)
)

g_obs = ggplot(equiv_tests, aes(x = names, y = const, label = round(observed.diff, var.rounding)), en
g_obs = g_obs + geom_text()
g_obs = (g_obs + theme_bw() + coord_flip()
         + theme(panel.grid.minor=element_blank()
                 , panel.grid.major=element_blank()
                 #, axis.line.x = element_blank()
                 , axis.text.x = element_text(color = "white", size = 14)
                 , axis.ticks.x = element_line(color = "white")
                 , axis.text.y = element_blank()
                 , axis.ticks.y = element_blank()
                 , axis.title.x = element_text(size = 14)
         )
         + ylim(1-.05, 1.05)
         + ggtitle("Observed\nMean\nDifference\n(Scale of Var)")
         + labs(y = " ", x = NULL)
)

g_pval = ggplot(equiv_tests, aes(x = names, y = const, label = round(p.vals, pval.rounding)), environ
g_pval = g_pval + geom_text()
g_pval = (g_pval + theme_bw() + coord_flip()
          + theme(panel.grid.minor=element_blank()
                  , panel.grid.major=element_blank()
                  #, axis.line.x = element_blank()
                  , axis.text.x = element_text(color = "white", size = 14)
                  , axis.ticks.x = element_line(color = "white")
                  , axis.text.y = element_blank()
                  , axis.ticks.y = element_blank()
                  , axis.title.x = element_text(size = 14)
          )
          + ylim(1-.05, 1.05)
          + ggtitle(ifelse(fdr_correct, "\nFDR\nCorrected\nP-value", "\n\n\nP-value"))
          + labs(y = " ", x = NULL)
)

g_var = ggplot(equiv_tests, aes(x = names, y = const, label = names))
g_var = g_var + geom_text()
g_var = (g_var + theme_bw() + coord_flip()
```

```
          + theme(panel.grid.minor=element_blank()
                  , panel.grid.major=element_blank()
                  #, axis.line.x = element_blank()
                  , axis.text.x = element_text(color = "white", size = 14)
                  , axis.ticks.x = element_line(color = "white")
                  , axis.text.y = element_blank()
                  , axis.ticks.y = element_blank()
                  , panel.border = element_blank()
                  , axis.title.x = element_text(size = 14)
          )
          + ylim(1-.05, 1.05)
          + ggtitle("\n \n \nVariable") + theme(plot.title = element_text(hjust = 0.5))
          + labs(y = " ", x = NULL)
  )

  grid.arrange(g_var, g_obs, g, g_inv, g_pval, ncol=5, nrow=1, widths= panel.widths, heights=c(4))
}


# #candidate characteristics:

replicate_strict = as.data.frame(run_equiv(X = subset(tombola, select = c("genero", "tiempo_de_residenci
                                           Tr = tombola$diputado_dummy, epsilon.method = "custom", custo
                                           fdr_correct = TRUE))
```

```
##               genero tiempo_de_residencia                 age
##                 0.66                 0.66                0.66
## [1] "genero"
## [1] "tiempo_de_residencia"
## [1] "age"
```

```
#Figure A1:
CairoPDF(file = "FigA1.pdf", width = 12, height = 6)
generate_plot(replicate_strict, panel.widths = c(2.5, 1.25, 5, 1.2, 1.2),
              display.names = c("Female", "Years of Residence", "Age"),
              pval.rounding = 3, fdr_correct = TRUE)
```

```
## [1] 1
## [1] "here"
## [1] 0.66
```




```
dev.off()
```

```
## pdf
##   2
```

```
###Permutation inference (with controls for municipality size):

#subset valid_diputado_dummy to remove one observation with missing value on covariate:
valid_diputado_dummy_c <- subset(valid_diputado_dummy, !is.na(total_lista_nominal_2012))

# Declare randomization procedure
declaration <- declare_ra(N = length(valid_diputado_dummy_c$diputado_dummy), m = sum(valid_diputado_dumm

#effect on new members:
##combined_padron_per_c:
Z <- valid_diputado_dummy_c$diputado_dummy
Y <- valid_diputado_dummy_c$combined_padron_per
C <- valid_diputado_dummy_c$total_lista_nominal_2012
dat <- data.frame(Y, Z, C)
set.seed(12102019)

# Conduct Randomization Inference
combined_padron_per_c <- conduct_ri(
  formula = Y ~ Z + C,
  declaration = declaration,
  sharp_hypothesis = 0,
  data = dat,
  sims = 50000,
```

```
  p = "two-tailed"
)
summary(combined_padron_per_c)
```

```
##   term estimate two_tailed_p_value
## 1    Z 216.7163            0.00442
```

```
#effect on total members:
##combined_padron_total_per_c:
Z <- valid_diputado_dummy_c$diputado_dummy
Y <- valid_diputado_dummy_c$combined_padron_total_per
C <- valid_diputado_dummy_c$total_lista_nominal_2012
dat <- data.frame(Y, Z, C)
set.seed(12102019)

# Conduct Randomization Inference
combined_padron_total_per_c <- conduct_ri(
  formula = Y ~ Z + C,
  declaration = declaration,
  sharp_hypothesis = 0,
  data = dat,
  sims = 50000,
  p = "two-tailed"
)
summary(combined_padron_total_per_c)
```

```
##   term estimate two_tailed_p_value
## 1    Z 433.6942            0.00838
```

```
##share_morena_deputies_c:
Z <- valid_diputado_dummy_c$diputado_dummy
Y <- valid_diputado_dummy_c$share_morena_deputies
C <- valid_diputado_dummy_c$total_lista_nominal_2012
dat <- data.frame(Y, Z, C)
set.seed(12102019)

# Conduct Randomization Inference
share_morena_deputies_c <- conduct_ri(
  formula = Y ~ Z + C,
  declaration = declaration,
  sharp_hypothesis = 0,
  data = dat,
  sims = 50000,
  p = "two-tailed"
)
summary(share_morena_deputies_c)
```

```
##   term   estimate two_tailed_p_value
## 1    Z 0.06274493            0.09766
```

```
##share_pan_prd_deputies_c:
Z <- valid_diputado_dummy_c$diputado_dummy
Y <- valid_diputado_dummy_c$share_pan_prd_deputies
C <- valid_diputado_dummy_c$total_lista_nominal_2012
dat <- data.frame(Y, Z, C)
set.seed(12102019)
```

```r
# Conduct Randomization Inference
share_pan_prd_deputies_c <- conduct_ri(
  formula = Y ~ Z + C,
  declaration = declaration,
  sharp_hypothesis = 0,
  data = dat,
  sims = 50000,
  p = "two-tailed"
)
summary(share_pan_prd_deputies_c)
```

```
##   term    estimate two_tailed_p_value
## 1    Z -0.07024376            0.03764
```

```r
##share_pri_deputies_c:
Z <- valid_diputado_dummy_c$diputado_dummy
Y <- valid_diputado_dummy_c$share_pri_deputies
C <- valid_diputado_dummy_c$total_lista_nominal_2012
dat <- data.frame(Y, Z, C)
set.seed(12102019)

# Conduct Randomization Inference
share_pri_deputies_c <- conduct_ri(
  formula = Y ~ Z + C,
  declaration = declaration,
  sharp_hypothesis = 0,
  data = dat,
  sims = 50000,
  p = "two-tailed"
)
summary(share_pri_deputies_c)
```

```
##   term    estimate two_tailed_p_value
## 1    Z 0.008459912            0.66214
```

```r
##MoV_MORENA_deputies_c:
Z <- valid_diputado_dummy_c$diputado_dummy
Y <- valid_diputado_dummy_c$MoV_MORENA_deputies
C <- valid_diputado_dummy_c$total_lista_nominal_2012
dat <- data.frame(Y, Z, C)
set.seed(12102019)

# Conduct Randomization Inference
MoV_MORENA_deputies_c <- conduct_ri(
  formula = Y ~ Z + C,
  declaration = declaration,
  sharp_hypothesis = 0,
  data = dat,
  sims = 50000,
  p = "two-tailed"
)
summary(MoV_MORENA_deputies_c)
```

```
##   term  estimate two_tailed_p_value
## 1    Z 0.1148086            0.07642
```

```r
##MoV_MORENA_PRD_deputies_c:
Z <- valid_diputado_dummy_c$diputado_dummy
Y <- valid_diputado_dummy_c$MoV_MORENA_PRD_deputies
C <- valid_diputado_dummy_c$total_lista_nominal_2012
dat <- data.frame(Y, Z, C)
set.seed(12102019)

# Conduct Randomization Inference
MoV_MORENA_PRD_deputies_c <- conduct_ri(
  formula = Y ~ Z + C,
  declaration = declaration,
  sharp_hypothesis = 0,
  data = dat,
  sims = 50000,
  p = "two-tailed"
)
summary(MoV_MORENA_PRD_deputies_c)
```

```
##   term  estimate two_tailed_p_value
## 1    Z 0.1329887            0.05322
```

```r
##combining results (RI):
findings_RI_c <- as.data.frame(rbind(rep(NA,3),rep(NA,3),rep(NA,3),rep(NA,3)))

findings_RI_c[1,] <- c("combined_padron_per_c",summary(combined_padron_per_c))[c(1,3:4)]
findings_RI_c[2,] <- c("combined_padron_total_per_c",summary(combined_padron_total_per_c))[c(1,3:4)]
findings_RI_c[3,] <- c("share_morena_deputies_c",summary(share_morena_deputies_c))[c(1,3:4)]
findings_RI_c[4,] <- c("share_pan_prd_deputies_c",summary(share_pan_prd_deputies_c))[c(1,3:4)]
findings_RI_c[5,] <- c("share_pri_deputies_c",summary(share_pri_deputies_c))[c(1,3:4)]
findings_RI_c[6,] <- c("MoV_MORENA_deputies_c",summary(MoV_MORENA_deputies_c))[c(1,3:4)]
findings_RI_c[7,] <- c("MoV_MORENA_PRD_deputies_c",summary(MoV_MORENA_PRD_deputies_c))[c(1,3:4)]

colnames(findings_RI_c) <- c("Outcome","RI ATE", "RI p-value")

findings_RI_c
```

```
##                       Outcome        RI ATE RI p-value
## 1       combined_padron_per_c 216.716320250    0.00442
## 2 combined_padron_total_per_c 433.694235261    0.00838
## 3     share_morena_deputies_c   0.062744929    0.09766
## 4    share_pan_prd_deputies_c  -0.070243761    0.03764
## 5        share_pri_deputies_c   0.008459912    0.66214
## 6       MoV_MORENA_deputies_c   0.114808640    0.07642
## 7   MoV_MORENA_PRD_deputies_c   0.132988690    0.05322
```

```r
##table with replication of main findings: randomization inference:
RI_table <- findings_RI_c

RI_table$`RI ATE` <- round(RI_table$`RI ATE`,4)
RI_table$`RI p-value` <- round(RI_table$`RI p-value`,4)

RI_table$Outcome <- c("New Members","Total Members","Vote Share for MORENA","Vote Share for PRD","Vote S
names(RI_table)[c(2:3)] <- c("ATE","p-value")

#Table F1:
```

```
stargazer(as.matrix(RI_table)[c(1:2),], rownames = F,
          type="latex", out = "TableF1.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:31:56
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}} ccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Outcome & ATE & p-value \\
## \hline \\[-1.8ex]
## New Members & 216.7163 & 0.0044 \\
## Total Members & 433.6942 & 0.0084 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```

```
#Table F2:
stargazer(as.matrix(RI_table)[c(3:7),], rownames = F,
          type="latex", out = "TableF2.tex")
```

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Fri, Jan 20, 2023 - 09:31:56
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}} ccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
## Outcome & ATE & p-value \\
## \hline \\[-1.8ex]
## Vote Share for MORENA &   0.0627 & 0.0977 \\
## Vote Share for PRD &  -0.0702 & 0.0376 \\
## Vote Share for PRI &   0.0085 & 0.6621 \\
## MoV: MORENA over Runner-up &   0.1148 & 0.0764 \\
## MoV: MORENA over PRD/PAN &   0.1330 & 0.0532 \\
## \hline \\[-1.8ex]
## \end{tabular}
## \end{table}
```