Apéndice A

Código utilizado

A.1. Información de los países

A.1.1. Lectura de datos del FMI

```
# FMI Datos de gasto por función versión 2.0
# fuente: https://data.imf.org/  Government Finance Statistics by
    Function of Government (COFOG)

# lee: GFSCOFOG_07-27-2020 23-41-17-45_timeSeries.csv
# crea: FMI_compos_v2.csv, FMI_compos_f_v2.csv, FMI_composc_v2.csv,
     FMI_compos_fc_v2.csv


suppressMessages(library(tidyverse))
suppressMessages(library(compositions))
suppressMessages(library(countrycode))
library(stringr) # función string length

##### Lectura y organización de datos #####
#Directorio
setwd("C:/Users/sgome/Dropbox/#tesis")

prueba<-read.csv(file = "./Data/FMI/COFOG/GFSCOFOG_07-27-2020
    23-41-17-45_timeSeries.csv",nrows = 500,na.strings = "",
                     #header=TRUE,sep = ",",
                   encoding = "UTF-8",check.names = TRUE)
names(prueba)
sapply(prueba, class)

tablaOrig<-read.csv(file = "./Data/FMI/COFOG/GFSCOFOG_07-27-2020
    23-41-17-45_timeSeries.csv",na.strings = "",
                      colClasses = c(rep("factor",9),rep("character"
                          ,48),rep("NULL",3)),encoding = "UTF-8")
```

```
25 names(tablaOrig)
26 names(tablaOrig)[1]<-"Country.Name"
27
28 ##### Preparación de datos #####
29
30 detfin<- tablaOrig %>%
31   select(Country.Code,COFOG.Function.Name,COFOG.Function.Code,Unit.
       Name,Attribute:X2019) %>%# variables a usar
32   filter(Attribute=="Value" , Unit.Name=="Percent of GDP") #%>%#
       tomar sólo los valores en %
33
34
35 # categorías de funciones
36 funciones<-detfin %>% group_by(COFOG.Function.Code,COFOG.Function.
     Name) %>% summarise(n=n()) %>%
37   mutate(nivel=case_when(
38     str_length(COFOG.Function.Code)==2 ~ 1,
39     str_length(COFOG.Function.Code)==4 ~ 2,
40     TRUE ~ 4
41   )) # extraer funciones y sus códigos
42 funciones<-funciones %>% mutate(nivelctm=nivel)
43 funciones$nivelctm[2:10]<-c(1,rep(2,8)) # crear categoría
     personalizada donde gob se detalla en sus subcomponentes
44
45 detfin<- merge(detfin,funciones %>% select(COFOG.Function.Code,nivel
     ,nivelctm))
46
47 # Nombre Gasto reducido
48 nombfun<-detfin %>% filter(nivel==2|nivelctm==2)%>% group_by(COFOG.
     Function.Name) %>% summarise(n=n()) %>%
49   arrange(COFOG.Function.Name)
50 medio<-c("investigación","defensa","economia","educacion","ambiente
     ","gob1","faid","publico",
51           "gobo","gobr","gobg","salud","vivienda","gobd","seguridad"
                ,"cultura","social","gobt")
52 corto<-c("inv","def","eco","edu","amb","gob1","faid","pub","gobo","
     gobr","gobg","sal","viv",
53           "gobd","seg","cul","soc","gobt")
54 nomb2<-data.frame(COFOG.Function.Name=nombfun$COFOG.Function.Name,
     medio=medio,corto=corto)
55
56 detfin<-merge(detfin,nomb2) # agregar nombres reducidos de
     funciones a la base
57 rm(prueba,funciones,nomb2,nombfun,tablaOrig,corto,medio)
58
59 # Base con las 10 categorías primarias del GFSM
60 funlv2<-detfin %>%
61   filter(nivel==2) %>%#usar 10 categorías de nivel2
62   select(-c(COFOG.Function.Code,nivel,nivelctm)) %>%#retirar
       variables ahora innecesarias
63   mutate(Country.Code=as.factor(countrycode(Country.Code,"imf","
       genc3c"))) %>%#reemplazar codigo país IMF por genc3c (ISO3
       no incluye a kosovo)
64   mutate(cont=as.factor(countrycode(Country.Code,"genc3c","
       continent"))) %>%#nueva variable: continente
65   mutate(reg=as.factor(countrycode(Country.Code,"genc3c","region"))
       ) %>%#nueva variable: region World Bank
```

```
66    mutate(cont=ifelse(Country.Code=="XKS","Europe",as.character(cont
          )))%>%# agregar manualmente continete de Kosovo que no está
          incluido en la función
67    mutate_at(vars(5:52), as.numeric) %>%#conversión a valores
68    pivot_longer(cols = starts_with("X"), names_to="year", values_to=
          "gto") %>%# columnas año en una sola variable
69    select(-c(Unit.Name,Attribute)) %>%# retirar Unit Name y Value
          de la base
70    mutate(year=as.numeric(substr(year,2,5))) %>%
71    select(-c(COFOG.Function.Name,medio)) %>%
72    pivot_wider(names_from = corto, values_from = gto) %>%#mover
          variable Función a columnas
73    rename(code=Country.Code) %>%
74    #na.omit() #retirar datos incompletos
75    filter_at(vars(def:soc),any_vars(!is.na(.))) #%>%#al menos una
          observación
76
77  # examino negativos ####
78  temp<-funlv2%>% filter(amb<0|viv<0) %>%
79    mutate(pais=countrycode(code,"genc3c","country.name"))
80  listaneg<-temp%>% group_by(code,pais) %>% summarise(n=n()) # lista
      de países con negativos
81  lista<-c("EST","ISR")
82  temp<-funlv2 %>% filter(code %in% listaneg$code)
83  plot(temp[temp$code=="EST",]$amb)
84  plot(temp[temp$code=="ISR",]$viv)
85
86
87  # retiro negativos y estimo los valores
88  # reemplazar negativos con la media condicional
89  for (i in 1:2){
90    dato<-funlv2 %>% filter(code==lista[i]) %>%
91      unite(col=id,code,year,sep = "'") %>%
92      column_to_rownames(.,var="id") %>%
93      select(def:soc)
94      rellenado=rellena_compv2(dato)
95    funlv2 [funlv2$code==lista[i],5:14] <- rellenado
96  }
97  plot(funlv2[funlv2$code=="EST",]$amb)
98  plot(funlv2[funlv2$code=="ISR",]$viv)
99
100
101 # examinar ceros y NAs ####
102 temp<- funlv2 %>% filter_at(vars(def:soc),any_vars(.==0|is.na(.)))
        %>%
103   group_by(code) %>% summarise(N=n())
104 lista<-as.character(temp$code)
105 temp<-funlv2 %>% filter(code %in% lista) %>%
106   mutate(pais=countrycode(code,"genc3c","country.name")) %>%
107   select(-c(cont,reg)) %>%
108   mutate(nNA=apply(is.na(.),1,sum),ncero=apply(.==0, 1, sum,na.rm=
          TRUE)) %>%#conteo de registros afectados
109   mutate(ntod=nNA + ncero) %>%# total de partes afectadas
110   group_by(code) %>%
111   summarise(N=n(),comp=sum(nNA==0&ncero==0),totNA=sum(nNA),totcero=
          sum(ncero),totamb=sum(ntod))
112 #lista a completar
```

```
lista2<-as.character(temp$code[temp$comp>2])
i<-lista2[9]

funlv2est<-funlv2
for (i in lista2) {
  dato<-funlv2 %>% filter(code==i) %>%
    unite(col=id,code,year,sep = "_") %>%
    column_to_rownames(.,var="id") %>%
    select(def:soc)
  #print(i)
  if(i=="ZAF"){
    rellenado<-rellena_media(dato)
  }else{
    if(i=="MNG"){
      rellenado1<-(rellena_compv2(dato %>%select(-amb))+rellena_
          media(dato %>%select(-amb)))/2 # estima seg ignorando amb
      rellenado2<-(rellena_compv2(dato)+rellena_media(dato))/2
      rellenado<-rellenado2
      rellenado$seg<-rellenado1$seg
    }else{
      if(i=="SGP"){
        rellenado1<-rellena_compv2(dato[c(6:23,26:29),])
        dato[c(6:23,26:29),]<-rellenado1
        rellenado<-rellena_compv2(dato)
      }else{
        rellenado<-rellena_compv2(dato)
      }
    }
  }
  funlv2est [funlv2$code==i,5:14]  <- rellenado
}

# comparación de estimados con serie original
i<-lista2[14]
dato<-funlv2 %>% filter(code==i)
orig<-dato[,apply(dato==0|is.na(dato),2,sum,na.rm=TRUE)>0]
variab<-names(orig)
estim<-funlv2est %>% filter(code==i) %>% select(all_of(variab))
dim(estim)[2]
d<-9
limite<-max(orig[,d],estim[,d],na.rm = TRUE)
plot(x=dato$year,y=t(orig[,d]),ylim = c(0,limite),main = i,xlab = "
    year",ylab = names(estim)[d])
lines(x=dato$year,y=t(estim[,d]))

# Resumen de NAs y ceros restantes ####
temp<-funlv2est %>% filter_at(vars(def:soc),any_vars(.==0|is.na(.))
    ) %>%
  group_by(code) %>% summarise(N=n())
lista<-as.character(temp$code)
temp<-funlv2 %>% filter(code %in% lista) %>%
  mutate(pais=countrycode(code,"genc3c","country.name")) %>%
  select(-c(cont,reg)) %>%
  mutate(nNA=apply(is.na(.),1,sum),ncero=apply(.==0, 1, sum,na.rm=
      TRUE)) %>%#conteo de registros afectados
  mutate(ntod=nNA + ncero) %>%# total de partes afectadas
  group_by(code) %>%
```

```
166   summarise(N=n(),comp=sum(nNA==0&ncero==0),totNA=sum(nNA),totcero=
          sum(ncero),totamb=sum(ntod))
167
168 temp<-funlv2 %>% filter(code %in% lista)
169 temp[temp==0]<-NA
170
171 # guardar archivos ####
172
173 write.csv(funlv2est,file = "./Data/FMI/COFOG/funlv2est.csv",row.
        names = FALSE)
174 funlv2est<-read.csv(file = "./Data/FMI/COFOG/funlv2est.csv",header
        = TRUE)
175
176 # Resumen partes afectadas
177 temp<-funlv2est %>% filter_at(vars(def:soc),any_vars(is.na(.)|.==0))
        %>%
178   group_by(code) %>% summarise(n=n())
179 lista<-as.character(temp$code)
180
181 res_partes<-funlv2est %>%
182   pivot_longer(cols=def:soc,names_to = "funcion",values_to="gto_pib
          ") %>%
183   mutate(nNA=is.na(gto_pib),ncero=ifelse(nNA,FALSE,gto_pib==0)) #
          %>%
184 # resumen por parte
185 res_partes %>% group_by(funcion) %>% summarise(N=n(),totNA=sum(nNA)
        ,totcero=sum(ncero)) %>%
186   mutate(perNA=totNA/N*100,percero=totcero/N*100)
187
188 # resumen por año
189 resanio<-res_partes %>% group_by(year) %>% summarise(N=n(),totNA=
        sum(nNA),totcero=sum(ncero)) %>%
190   mutate(perNA=totNA/N*100,percero=totcero/N*100)
191 rm(temp)
192
193 funlv2<-funlv2est
194 rm(funlv2est)
195
196 # Guardar archivos sin filtrar ####
197 write.csv(funlv2,"./Data/FMI/COFOG/FMI_compos_v2.csv",row.names =
        FALSE)
198 write.csv(funcustom,"./Data/FMI/COFOG/FMI_composc_v2.csv",row.names
        = FALSE)
199
200
201
202 # Tratamiento de ceros, quitar ceros ####
203 funlv2 %>% filter_all(any_vars(is.na(.)))#NA check
204 # buscar ceros y negativos
205 busq<-matrix(0,3,10)
206 rownames(busq)<-c("ceros","negativos","NAs")
207 colnames(busq)<-names(funlv2)[5:14]
208 for (i in 5:14) {
209   busq[1,i-4]<-sum(funlv2[,i]==0,na.rm = TRUE)/dim(funlv2)[1]*100
210   busq[2,i-4]<-sum(funlv2[,i]<0,na.rm = TRUE)/dim(funlv2)[1]*100
211   busq[3,i-4]<-sum(is.na(funlv2[,i]))/dim(funlv2)[1]*100
212 };busq
```

```
213
214 rm(busq)
215 # resumen faltantes viv
216 lista<-funlv2 %>% select(code,viv) %>% filter(viv==0|is.na(viv))%>%
      group_by(code) %>% summarise(n=n()) %>% select(code)
217 lista<-as.character(lista$code) ; lista
218 funlv2 %>% select(code,viv) %>% filter(code %in% lista) %>%
219   mutate(conteo=1,faltante=viv==0|is.na(viv)) %>% group_by(code)
        %>% summarise(tot=sum(conteo),nfal=sum(faltante))
220 # eliminar países con NAs en viv
221 funlv2 <- funlv2 %>% filter(!(code %in% lista))
222
223 # resumen faltantes cul
224 lista<-funlv2 %>% select(code,cul) %>% filter(cul==0|is.na(cul))%>%
      group_by(code) %>% summarise(n=n()) %>% select(code)
225 lista<-as.character(lista$code) ; lista
226 funlv2 %>% select(code,cul) %>% filter(code %in% lista) %>%
227   mutate(conteo=1,faltante=cul==0|is.na(cul)) %>% group_by(code)
        %>% summarise(tot=sum(conteo),nfal=sum(faltante))
228 # eliminar países con NAs en cul
229
230
231
232 # resumen faltantes de defensa
233 lista<-funlv2 %>% select(code,def) %>% filter(def==0|is.na(def))%>%
      group_by(code) %>% summarise(n=n()) %>% select(code)
234 lista<-as.character(lista$code);lista
235 funlv2 %>% select(code,def) %>% filter(code %in% lista) %>%
236   mutate(conteo=1,faltante=def==0|is.na(def)) %>% group_by(code)
        %>% summarise(tot=sum(conteo),nfal=sum(faltante))
237 # HKG y MAC no pueden tener gasto de defensa se podrían promediar
      con CHN (años donde los 3 tengan obs) y eliminar los demás
238 lista<-lista[c(1,3,4,6,7)];lista
239 funlv2<-funlv2 %>%filter(!code%in%lista)
240
241 # resumen faltantes ambiental
242 lista<-funlv2 %>% select(code,amb) %>% filter(amb==0|is.na(amb))%>%
      group_by(code) %>% summarise(n=n()) %>% select(code)
243 lista<-as.character(lista$code) ; lista
244 funlv2 %>% select(code,amb) %>% filter(code %in% lista) %>%
245   mutate(conteo=1,faltante=amb==0|is.na(amb)) %>% group_by(code)
        %>% summarise(tot=sum(conteo),nfal=sum(faltante))
246 # USA tiene presupuesto de gasto ambiental (~0.20% del gdp) se
      podría rellenar con datos externos o ignorar el país
247 lista<-lista[1];lista
248 funlv2<-funlv2 %>% filter(!code%in%lista)
249
250
251 # Posiblemente trabajaremos sin considerar def y amb, entonces se
      pueden eliminar los NAs de los demás
252 # Para análisis que incluya def sumar HKG y MAC con CHN, eliminar
      otros países con NAs
253 # Para analisis que incluya amb conseguir serie de USA, eliminar
      otros países con NAs
254 funlv2 %>% filter_all(any_vars(is.na(.)))#NA check
255
256 # Guardar archivos filtrados ####
```

```
257 fun2fil<-funlv2
258 write.csv(fun2fil,"./Data/FMI/COFOG/FMI_compos_f_v2.csv",row.names
    = FALSE)
259 funcfil<-funcustom %>%na.omit()
260 write.csv(funcfil,"./Data/FMI/COFOG/FMI_compos_fc_v2.csv",row.names
     = FALSE)
261
262 funlv2<-read.csv(file = "./Data/FMI/COFOG/FMI_compos_f_v2.csv",
    header = TRUE)
263 funlv2 %>% filter_all(any_vars(is.na(.)))#NA check
264
265 ### Tabla resumen de # de observaciones por año para cada
    continente ####
266 # funlv2<-read.csv(file = "./Data/FMI/COFOG/FMI_compos_v2.csv",
    header = TRUE)
267 # prefun<-read.csv(file = "./Data/FMI/FMI_compos.csv",header = TRUE
    ) # datos anteriores
268
269
270 # Agregar clasificación WB
271 wbclasif<-read.csv(file = "./Data/wbclasif.csv",header = TRUE) #
    clasificación WB de países
272 wbclasif<-wbclasif %>%group_by(code,year,Nivel) %>% summarise(n=n())
    %>%
273   mutate(code=ifelse(code=="XKX","XKS",code))
274 # funlv2<-funlv2 %>% mutate(cont=countrycode(code,"genc3c","
    continent"))#%>%
275 #   mutate(cont=ifelse(code=="XKS","Europe",cont)) # agregar
    manualmente continete de Kosovo que no está incluido en la
    función
276
277 funlv2clas<-merge(x=funlv2,y=wbclasif %>% select(code,year,Nivel),
    all.x = TRUE) %>%
278   mutate(Nivel=ifelse(code=="USA"&is.na(Nivel),"Alto",Nivel)) #
      corrección de faltantes en la tabla de clasificación
279 funlv2clas %>% filter_all(any_vars(is.na(.)))#NA check
280
281 ny<-funlv2 %>% group_by(year,cont) %>%
282   summarise(n=n_distinct(code)) %>%
283   pivot_wider(names_from = cont,values_from=n) %>%
284   replace_na(list(Americas=0,Europe=0,Asia=0,Africa=0,Oceania=0))
        %>%
285   ungroup() %>%
286   mutate(glob=select(.,Americas:Oceania)%>%rowSums())
287
288 # ggplot(data=ny, aes(x=year,y=glob,color=cont))+
289 #   geom_line()
290 ny<-funlv2 %>% group_by(year,cont) %>%
291   summarise(n=n_distinct(code))
292 ggplot(data=ny, aes(x=year,y=n,color=cont))+
293   geom_line(size=1)
294
295 ## Tabla resumen de # de observaciones por año para cada nivel de
    ingreso ####
296
297 # Tabla resumen de # años por país
298 ny2<-funlv2 %>% group_by(code,cont) %>%
```

```
299   summarise(n=n_distinct(year),ini=min(year),fin=max(year),span=fin
          -ini+1,hueco=span-n) #%%
300   #pivot_wider(names_from = cont,values_from=n) %%
301   #replace_na(list(Americas=0,Europe=0,Asia=0,Africa=0,Oceania=0))
          %%
302   #ungroup() %%
303   #mutate(glob=select(.,Asia:Africa)%%rowSums())
304 library(ggplot2)
305 library(ggsci) # paletas de colores de journals
306
307 ggplot(data = ny2,aes(n,fill=cont))+
308   geom_histogram(binwidth = 2)+
309   theme_bw()+
310   scale_color_locuszoom()
311
312 sum(ny2$n)
313
314 #interpolación de intra-años faltantes ####
315 #revisar continuidad de series ymax-ymin vs #obs
316 resanio<-funlv2clas %%group_by(code) %%summarise(ti=n_distinct(
      year),tmax=max(year),tmin=min(year),maxobs=tmax-tmin+1,huecos=
      maxobs-ti)
317 funlv2clas %%group_by(Nivel,code) %%
318   summarise(ti=n_distinct(year),tmax=max(year),tmin=min(year),
          maxobs=tmax-tmin+1,huecos=maxobs-ti)%%
319   summarise(sti=sum(ti),ttmax=max(tmax),ttmin=min(tmin),shuecos=sum
          (huecos))
320 temp<-resanio %%filter(huecos>0)%%select(code) #filtro de
      países con huecos
321 lista<-as.character(temp$code);lista # lista de países con huecos
322 #regresar huecos en serie como NAs
323 temp<-funlv2clas %%select(-Nivel) %%pivot_longer(cols = def:soc
      ,names_to="fun",values_to="gto")%%
324   pivot_wider(names_from = "year",values_from="gto")%%# esto crea
          NAs en años vacíos
325   pivot_longer(cols=!c(code,cont,reg,fun),names_to = "year",values_
          to="gto") %%#regreso año a una sola variable
326   pivot_wider(names_from = "fun",values_from="gto") %%# regresa
          gasto por función a columnas
327   arrange(code,year) %%
328   merge(resanio%%select(code,tmax,tmin)) %%#agregar información
          previa del máximo tamaño de la serie
329   mutate(year=as.numeric(year)) %%#convertir año a números para
          comparación
330   filter(year<=tmax,year>=tmin)
331
332 # rellenar con interpolación
333 funrell<-temp #inicializar base con huecos interpolados
334
335 for (i in lista) {
336   temp<-funrell %%filter(code==i)
337   serie<-zoo(temp%%select(def:soc),temp$year)
338   serie<-na.approx(serie)
339   funrell[funrell$code==i,5:14]<-serie
340 }
341
342 funrell %%filter_all(any_vars(is.na(.)))#NA check
```

```
resaniocheck<-funrell %>% group_by(code) %>% summarise(ti=n_
    distinct(year),tmax=max(year),tmin=min(year),maxobs=tmax-tmin
    +1,huecos=maxobs-ti)

paisi<-funrell %>% filter(code=="ZAF")
serie<-zoo(paisi %>%select(def:soc),paisi$year)
serie<-na.approx(serie); serie
plot(time(serie),serie$edu,type = "b")

write.csv(funrell,file = "./Data/FMI/COFOG/FMI_compos_fr_v2.csv",
    row.names = FALSE)
```

../Code/1_FMI_gto.R

A.1.2. Lectura de datos de México

```
# Gasto por función de México

#fuente A: Centro de estudios de las finanzas públicas (CEFP)
    https://www.cefp.gob.mx/Pub_Gasto_Estadisticas.htm
#fuente B: SHCP: http://www.shcp.gob.mx/POLITICAFINANCIERA/
    FINANZASPUBLICAS/Estadisticas_Oportunas_Finanzas_Publicas/
    Paginas/unica2.aspx
#fuente C: Transparencia presupuestaria https://www.
    transparenciapresupuestaria.gob.mx/es/PTP/Datos_Abiertos

##### Inicialización de Paquetes #####

library(readxl)
library(tidyverse)
library(compositions)
library(ggplot2)
library(xtable)


##### Lectura y organización de datos #####
#Directorio
setwd("C:/Users/sgome/Dropbox/#tesis")

mex90_02<-read_excel(path = "./Data/CEFP/cfb1.xls",sheet = 1,skip
    =3,n_max = 22,na="-")
mex03_11<-read_excel(path = "./Data/CEFP/cfc1.xls",sheet = 1,skip
    =3,n_max = 30,na="-")
neto80<-read_excel(path = "./Data/CEFP/ca1.xls",sheet = 1,skip=3,n_
    max = 53)
neto07_19<-read_excel(path = "./Data/SHCP/gastoneto.xls",sheet = 1,
    skip = 2,n_max = 26,na="n.d.")
names(neto07_19)[1]<-"Concepto"
mex07_19<-read_excel(path = "./Data/SHCP/gastoprog.xls",sheet = 1,
    skip = 2,n_max = 38,na="n.d.")
names(mex07_19)[1]<-"Concepto"

# Consolidación Gasto Neto
neto07_19$Concepto
```

```
unique(neto07_19$Concepto)
parteA<-neto80 %>% pivot_longer(cols = 2:34,names_to="year",values_to="gto") %>%
  pivot_wider(names_from = "Ramos",values_from="gto") %>%
  select(c(1,45,54,48)) %>% # seleccionar año, gasto programable, gasto neto y cto financiero
  filter(year!="2012 A") %>%
  mutate(year=as.numeric(year))
parteB<-neto07_19[c(1:10,20),] %>% pivot_longer(cols = 2:31,names_to="year",values_to="gto") %>%
  pivot_wider(names_from = "Concepto" ,values_from="gto") %>%
  select(c(1:3,12)) %>% # seleccionar año, gasto programable, gasto neto y cto financiero
  mutate(year=as.numeric(year))
neto<-merge(parteA,parteB,all = TRUE)
names(neto)<-c("year","gtoprog_CEFP","gtoneto_CEFP","gtofin_CEFP","gtoneto_SHCP","gtoprog_SHCP","gtofin_SHCP")
neto<-neto %>% pivot_longer(cols = starts_with("gto"),names_to="serie",values_to="gto") %>%
  separate(serie,into = c("concepto","fuente")) %>%
  pivot_wider(names_from = "concepto",values_from="gto")

ggplot(neto,aes(x=year,y=gtoprog,color=fuente))+geom_point()+
  scale_y_log10()
ggplot(neto,aes(x=year,y=gtoneto,color=fuente))+geom_point()+
  scale_y_log10()
ggplot(neto,aes(x=year,y=gtofin,color=fuente))+geom_point()+
  scale_y_log10()
# son muy similares las series, tomar fuente SHCP (más reciente)
neto<-neto %>% filter(fuente=="CEFP"&year<1990|fuente=="SHCP"&year>=1990)
neto<-neto %>% mutate(noprog=gtoneto-gtoprog)

#Catálogos equivalencias a COFOG FMI ####

cat90<-data.frame(funmex=unique(mex90_02$Concepto),
                  funFMI=c(NA,NA,"edu","sal","soc","eco","soc","viv",NA,"eco","eco","eco","eco",
                           NA,"gob","seg","gob","seg","def","gob","amb","seg"))
cat03<-data.frame(funmex=unique(mex03_11$Concepto),
                  funFMI=c(NA,NA,"edu","sal","soc","viv","viv","soc",NA,"eco","eco","eco","eco","eco","eco","eco",
                           "gob","eco","eco",NA,"gob","gob","def","gob","gob","seg","amb","gob","gob","gob"))
cat07<-data.frame(funmex=unique(mex07_19$Concepto),
                  funFMI=c(NA,NA,"gob","seg","gob","gob","gob","def","seg","gob",NA,"amb","viv","sal","cul","edu",
                           "soc","soc",NA,rep("eco",7),"gob","eco",NA,"eco","gob","gob","soc","gob","gob","eco",
                           "eco","eco"))
# Tabla en el documento: "Comparación entre clasificaciones funcionales del gasto público"
```

```
69 print(xtable(cat07[1:28,]  ,digits=1),include.rownames=FALSE)
70
71 # Traduccion por catálogo a funciones FMI ####
72 parteA<-mex90_02  %>%
73   merge(x=.,y=cat90,by.x="Concepto",by.y="funmex") %>% #inclusión
       catálogo
74   filter(!is.na(funFMI))  %>% #eliminación de totales y subtotales
75   pivot_longer(cols = 2:14,names_to="year",values_to="gto")  %>% #
       año a columna
76   group_by(funFMI,year) %>%
77   summarise(gto=sum(gto,na.rm = TRUE)) %>%
78   pivot_wider(names_from = "funFMI",values_from="gto") %>%
79   mutate(cul=NA)
80
81 # Comparación de totales
82 plot(x=1990:2002,y=mex90_02[1,2:14])
83 lines(x=1990:2002,y=rowSums(parteA[,2:10])) # se respeta el total
84 # Gráfica del gasto neto vs total gasto programable
85 plot(x=1990:2002,y=log(mex90_02[1,2:14]),type="l",col="blue",lwd=2,
     ylim = c(11.5,14.5))
86 lines(x=1990:2002,y=log(neto80[53,12:24]),lwd=2,col="gray")
87
88
89 parteB<-mex03_11  %>%
90   merge(x=.,y=cat03,by.x="Concepto",by.y="funmex") %>% #inclusión
       catálogo
91   filter(!is.na(funFMI))  %>% #eliminación de totales y subtotales
92   pivot_longer(cols = 2:10,names_to="year",values_to="gto")  %>% #
       año a columna
93   group_by(funFMI,year) %>%
94   summarise(gto=sum(gto,na.rm = TRUE)) %>%
95   pivot_wider(names_from = "funFMI",values_from="gto")  %>%
96   mutate(cul=NA)
97
98 plot(x=2003:2011,y=mex03_11[1,2:10])
99 lines(x=2003:2011,y=rowSums(parteB[,2:10])) #chequeo de totales
100
101 parteC<-mex07_19  %>%
102   merge(x=.,y=cat07,by.x="Concepto",by.y="funmex") %>% #inclusión
       catálogo
103   filter(!is.na(funFMI))  %>% #eliminación de totales y subtotales
104   pivot_longer(cols = 2:31,names_to="year",values_to="gto")  %>% #
       año a columna
105   group_by(funFMI,year) %>%
106   summarise(gto=sum(gto,na.rm = TRUE)) %>%
107   pivot_wider(names_from = "funFMI",values_from="gto")  %>%
108   filter_at(vars(amb:viv),any_vars(.>0))
109
110 plot(x=1990:2019,y=mex07_19[1,2:31])
111 lines(x=2007:2019,y=rowSums(parteC[,2:11]))# chequeo de totales
112
113
114
115 # Consolidación de partes A-C####
116 mex<-rbind(parteA,parteB %>% filter(year<=2006),parteC)  %>%
117   mutate(year=as.numeric(year),code="MEX")
118
```

```
# Agregar gasto no prog. como parte del gasto gob.

plot(x=mex$year,y=neto$noprog[dplyr::between(neto$year,1990,2019)],
     type = "b",col="red")
lines(x=mex$year,y=mex$gob,type = "b",col="blue")

neto$gob_exd<-c(rep(NA,10),mex$gob) #gasto en gobierno sin deuda y
     transf
neto$part<-neto$noprog-neto$gtofin #participaciones y adefas
neto$gob<-neto$noprog+neto$gob_exd #gobierno total
neto<-neto %>% mutate(across(gtofin:part,~./gob*100,.names = "p{col
     }"))
neto %>% select(year,gtofin:part) %>%#para gráfica en dinero
  pivot_longer(!year,names_to="serie",values_to="porcentaje") %>%
  ggplot(aes(x=year,y=porcentaje,col=serie))+geom_line(size=1)
medias<-sapply(neto[,10:13],mean,na.rm=T)
neto %>% select(year,pgtofin:ppart) %>%#para gráfica en porcentaje
  pivot_longer(!year,names_to="serie",values_to="porcentaje") %>%
  ggplot(aes(x=year,y=porcentaje,col=serie))+geom_line(size=1)+
  annotate("text",x=1990,y=medias,label=round(medias,1))
plot(x=mex$year,y=neto$noprog[dplyr::between(neto$year,1990,2019)],
     type = "b",col="red")

mex$gob<-mex$gob+neto$noprog[dplyr::between(neto$year,1990,2019)]


# Revisión de NAs y ceros ####
ggplot(mex,aes(x=year,y=gob))+geom_point()

# Rellenado ####
temp<-mex #inicializo rellenado
temp[between(mex$year,2003,2006),2]=NA #descartar valores de gto
     ambiental atípicos
temp<-temp %>% unite(col = "id",code,year,sep="'") %>% column_to_
     rownames(var="id") # poner en formato base FMI
temp<-rellena_compv2(temp) # función rellenar con media condicional
temp<-temp %>% rownames_to_column(var = "id") %>%
  separate(col="id",into = c("code","year")) %>% # regresar
      variables code, year
  mutate(year=as.numeric(year))


mexrell<-mex
mexrell[between(mexrell$year,2003,2006),]$amb<-temp[between(temp$
     year,2003,2006),]$amb
mexrell[between(mexrell$year,1990,2006),]$cul<-temp[between(temp$
     year,1990,2006),]$cul

ggplot(mex,aes(x=year,y=amb))+geom_point()+
  geom_line(data = mexrell,aes(x=year,y=amb))
ggplot(mex,aes(x=year,y=cul))+geom_point()+
  geom_line(data = mexrell,aes(x=year,y=cul))

# Guardar archivo rellenado ####
write.csv(mexrell,file = "./Data/SHCP/mex_compos_r.csv",row.names =
     FALSE)

```

```
166
167
168 # leer datos de PIB ####
169 pib<-data.frame(t(read_excel(path = "./Data/INEGI/PIBT_5.xlsx",
        sheet = 1,skip = 4,n_max = 2))[2:197,])
170 names(pib)<-c("tiempo","pib"); rownames(pib)<-c()
171 pib<-pib %>% mutate(year=rep(1993:2020,times=1,each=7),pib=as.
        numeric(pib)) %>%
172   filter(tiempo=="Anual") %>% na.omit()
173
174 wbdata<-read.csv(file="./Data/World Bank/WDIData.csv", encoding = "
        UTF-8")
175 pibwb<- wbdata %>% filter(Indicator.Name=="GDP (current LCU)") %>%
176   select(Country.Code,c(5:64)) %>%
177   pivot_longer(cols = starts_with("X"), names_to = "year",values_to
          = "pib") %>%
178   mutate(year=as.numeric(substr(year,2,5)),pib = pib /1E6) %>%
179   rename(code=Country.Code)%>%
180   filter(code=="MEX")
181
182 plot(pibwb$year,pibwb$pib,type = "l",col="blue")
183 lines(pib$year,pib$pib,type = "l",col="red")
184
185 # completar base historica con datos del wb
186 pib<-rbind(pib %>% select(year,pib),pibwb %>% select(year,pib) %>%
        filter(year<1993)) %>% arrange(year)
187 seriepib<-pib$pib[pib$year>=1990]
188 # Convertir a porcentaje de PIB ####
189 mexper<-mexrell %>% mutate_at(vars(amb:cul),~./seriepib*100)
190
191 write.csv(mexper,file = "./Data/SHCP/mex_compos_rp.csv",row.names =
        FALSE)
```

../Code/2_SHCP_gto.R

A.1.3. Consolidación de bases

```
1 ##### Organización de datos y consolidación de base#####
2
3 library(tidyverse)
4 library(countrycode) # funcion con códigos de países
5 library(ggrepel) #para text repel
6 library(readxl) #lee excel
7 library(janitor) #para adorn_totals
8 library(xtable) #para imprimir tablas formato LaTeX
9
10 # Agregar México a datos FMI ####
11 #Directorio
12 setwd("C:/Users/sgome/Dropbox/#tesis")
13 fmi<-read.csv(file = "./Data/FMI/COFOG/FMI_compos_fr_v2.csv",
        header = TRUE) %>% select(-c(reg,tmin,tmax)) %>%
14   rename(gob=pub)
15 mex<-read.csv(file = "./Data/SHCP/mex_compos_rp.csv",header = TRUE)
16 mex<-mex %>% mutate(cont=countrycode(code,"genc3c","continent"))
```

```
17
18 fmi<-rbind(fmi,mex)
19 rm(mex)
20
21
22 # Leer variables adicionales ####
23 ## GDP, Miembros FMI, Gasto y COFOG #GDP growth #Debt #Revenues #
       Interest paid on debt #GDP per capita #WB clasif
24
25 # variables del WB ####
26 wbdata<-read.csv(file="./Data/World Bank/WDIData.csv", encoding = "
       UTF-8")
27 lista_variab<-c("Population, total","Population density (people per
       sq. km of land area)","Total natural resources rents (% of GDP
       )",
28                 "GDP, PPP (constant 2011 international $)","GDP per
                      capita, PPP (constant 2011 international $)",
29                 "GDP growth (annual %)","Expense (% of GDP)","
                      Interest payments (% of expense)","Revenue,
                      excluding grants (% of GDP)",
30                 "Tax revenue (% of GDP)","Central government debt,
                      total (% of GDP)",
31                 "New business density (new registrations per 1,000
                      people ages 15-64)","Real interest rate (%)")
32
33 selecWB<-wbdata %>% filter(Indicator.Name%in%lista_variab) %>%
34   select(Country.Code,Indicator.Name,c(5:64)) %>%
35   pivot_longer(cols = starts_with("X"), names_to = "year",values_to
         = "value") %>%
36   mutate(year=as.numeric(substr(year,2,5))) %>%
37   rename(code=Country.Code) %>%
38   pivot_wider(names_from = Indicator.Name,values_from="value") %>%
39   filter_at(vars(3:15),any_vars(!is.na(.))) #%>%#al menos una
         observación
40 income_share<-wbdata %>% filter(str_detect(Indicator.Name,"Income
       share")) %>%
41   select(Country.Code,Indicator.Name,c(5:64)) %>%
42   pivot_longer(cols = starts_with("X"), names_to = "year",values_to
         = "value") %>%
43   mutate(year=as.numeric(substr(year,2,5))) %>%
44   rename(code=Country.Code) %>%
45   pivot_wider(names_from = Indicator.Name,values_from="value") %>%
         na.omit()
46  #filter_at(vars(3:9),any_vars(!is.na(.))) #%>%#al menos una
         observación
47
48
49 # income_share %>% filter_all(any_vars(is.na(.)))#NA check
50 selecWB<-merge(selecWB,income_share,all=TRUE)
51 write.csv(selecWB,file = "./Data/World Bank/selecWB.csv",row.names
       = FALSE)
52 rm(wbdata,income_share)
53
54 # Actualización de variables del WB
55 wbdatax<-read_excel(path = "./Data/World Bank/WDIEXCEL.xlsx",sheet
       = "Data")
56 lista_variabx<-read.csv(file = "./Data/World Bank/milista.csv")
```

```
57
58
59 selecWB<-wbdatax %>%
60   rename(Indicator.Name=`Indicator Name`,Country.Code=`Country Code
       `) %>%
61   filter(Indicator.Name%in%lista_variabx$Indicator.Name) %>%
62   merge(lista_variabx %>% select(Indicator.Name,ShortName)) %>%
63   select(Country.Code,ShortName,c(5:64)) %>%
64   pivot_longer(cols = 3:62, names_to = "year",values_to = "value")
       %>%
65   mutate(year=as.numeric(year)) %>%
66   rename(code=Country.Code) %>%
67   pivot_wider(names_from = ShortName,values_from="value") %>%
68   filter_at(vars(3:15),any_vars(!is.na(.))) #%>%#al menos una
       observación
69
70 write.csv(selecWB,file = "./Data/World Bank/selecWB2.csv",row.names
     = FALSE)
71 names(selecWB)
72
73
74 # Variables del FMI ####
75 #Gasto
76 gto=read_excel(path = "./Data/FMI/Expenditure.xls",na="no data")
77 gto = gto %>%
78   rename(pais=`Expenditure (% of GDP)`) %>%#homologar nombre de
       variable
79   filter(!grepl('Euro|Advanced|Emerging|Developing',pais))%>%
80   mutate(code=countrycode(pais,"country.name","genc3c")) %>%#
       incluir columna de codigos ISO
81   pivot_longer(cols = 2:36,names_to = "year", values_to = "gtof")
       %>%#convertir a tabla larga
82   select(-pais) #%>%
83   #na.omit() # retirar renglones sin dato
84 gto %>% filter(year==2017) %>% summarise(n=n_distinct(code)) #
     chequeo número de paises
85 gdp=read_excel(path = "./Data/FMI/GDP.xls",na="no data")
86 gdp = gdp %>% rename(pais=`GDP, current prices (Billions of U.S.
     dollars)`) %>%#homologar nombre de variable
87   filter(!grepl('Euro|Advanced|Emerging|Developing',pais))%>%
88   mutate(code=countrycode(pais,"country.name","genc3c")) %>% #
       incluir columna de codigos ISO
89   pivot_longer(cols = 2:46, names_to = "year",values_to = "pibf")
       %>%#convertir a tabla larga
90   select(-pais) %>%
91   na.omit() # retirar renglones sin dato
92 #GDP growth
93 growth<-read_excel(path = "./Data/FMI/RealGDPgrowth.xls",na="no
     data")
94 growth = growth %>%
95   rename(pais=`Real GDP growth (Annual percent change)`) %>%#
       homologar nombre de variable
96   filter(!grepl('Euro|Advanced|Emerging|Developing',pais))%>%
97   mutate(code=countrycode(pais,"country.name","genc3c")) %>%#
       incluir columna de codigos ISO
98   pivot_longer(cols = 2:46,names_to = "year", values_to = "cref")
       %>%#convertir a tabla larga
```

```
99    na.omit() %>%# retirar renglones sin dato
100   select(-pais)
101 growth %>% filter(year==2017) %>% summarise(n=n_distinct(code)) #
      chequeo número de paises
102 #Debt
103 deuda= read_excel(path = "./Data/FMI/GeneralGovGrossDebt.xls",na="
      no data")
104 deuda = deuda %>%
105   rename(pais=`General government gross debt (Percent of GDP)`) %>%
        #homologar nombre de variable
106   filter(!grepl('Euro|Advanced|Emerging|Developing',pais))%>%
107   mutate(code=countrycode(pais,"country.name","genc3c")) %>%#
        incluir columna de codigos ISO
108   pivot_longer(cols = 2:46,names_to = "year", values_to = "deuf")
        %>%#convertir a tabla larga
109   na.omit() %>%# retirar renglones sin dato
110   select(-pais)
111 deuda %>% filter(year==2017) %>% summarise(n=n_distinct(code)) #
      chequeo número de paises
112 #Revenues
113 ingr= read_excel(path = "./Data/FMI/Revenue.xls",na="no data")
114 ingr = ingr %>%
115   rename(pais=`Revenue (% of GDP)`) %>%#homologar nombre de
        variable
116   filter(!grepl('Euro|Advanced|Emerging|Developing',pais))%>%
117   mutate(code=countrycode(pais,"country.name","genc3c")) %>%#
        incluir columna de codigos ISO
118   pivot_longer(cols = 2:36,names_to = "year", values_to = "ingf")
        %>%#convertir a tabla larga
119   na.omit() %>%# retirar renglones sin dato
120   select(-pais)
121 ingr %>% filter(year==2017) %>% summarise(n=n_distinct(code)) #
      chequeo número de paises
122 #Interest paid on debt
123 int= read_excel(path = "./Data/FMI/InterestPublicDebt.xls",na="no
      data")
124 int = int %>%
125   rename(pais=`Interest paid on public debt, percent of GDP (% of
        GDP)`) %>%#homologar nombre de variable
126   filter(!grepl('Euro|Advanced|Emerging|Developing',pais))%>%
127   mutate(code=countrycode(pais,"country.name","genc3c")) %>%#
        incluir columna de codigos ISO
128   pivot_longer(cols = 2:213,names_to = "year", values_to = "intf")
        %>%#convertir a tabla larga
129   na.omit() %>%# retirar renglones sin dato
130   select(-pais)
131 int %>% filter(year==2011) %>% summarise(n=n_distinct(code)) #
      chequeo número de paises
132
133 extraFMI<-merge(gto,growth,all = TRUE) %>%
134   merge(gdp,all = TRUE)%>%
135   merge(deuda,all = TRUE) %>%
136   merge(ingr,all = TRUE) %>%
137   merge(int,all = TRUE)
138 write.csv(extraFMI,file = "./Data/FMI/extraFMI2.csv",row.names =
      FALSE)
139 rm(gto,gdp,growth,deuda,ingr,int)
```

```
140
141 # WB clasificación ingreso
142 wbclasif<-read.csv(file = "./Data/wbclasif.csv",header = TRUE) #
      clasificación WB de países
143 wbclasif<-wbclasif %>%select(code,year,Nivel) %>%
144   mutate(code=ifelse(code=="XKX","XKS",code))
145
146 # Combinar todo ####
147 selecWB<-read.csv(file = "./Data/World Bank/selecWB2.csv",header =
      TRUE)
148 extraFMI<-read.csv(file = "./Data/FMI/extraFMI2.csv",header = TRUE)
149
150 consolidado<-fmi %>%
151   merge(y=extraFMI,all.x= TRUE) %>%
152   merge(y=wbclasif %>% select(code,year,Nivel),all.x = TRUE) %>%
153   merge(y=selecWB,all.x = TRUE) %>%
154   mutate(Nivel=ifelse(code=="USA"&is.na(Nivel),"Alto",Nivel)) #
        corrección de faltantes en la tabla de clasificación
155 consolidado %>% filter(is.na(Nivel))
156 rm(extraFMI,selecWB)
157
158
159 #sin filtrar a base GFS
160 consolidado<-fmi %>%
161   merge(y=extraFMI %>% filter(between(year,1970,2018)),all= TRUE)
        %>%
162   merge(y=wbclasif %>% select(code,year,Nivel),all = TRUE) %>%
163   merge(y=selecWB,all = TRUE) %>%
164   mutate(Nivel=ifelse(code=="USA"&is.na(Nivel),"Alto",Nivel)) %>%#
        corrección de faltantes en la tabla de clasificación
165   mutate(pais=countrycode(code,"genc3c","country.name")) %>%
166   filter(!is.na(pais)) %>%
167   select(code,year,pais,Nivel,everything()) %>%
168   mutate(gtofuns=select(.,def:soc) %>% rowSums(),
169          gto2=ifelse(is.na(gtof),gtofuns,gtof)) %>%
170   group_by(code) %>%
171   fill(c(Nivel,cont),.direction ="updown")
172
173
174 rm(extraFMI,selecWB,fmi,wbclasif)
175
176 write.csv(consolidado,file = "./Data/consolidado2b.csv",row.names =
      FALSE)
177
178 # comparar suma gastos con total (FMI) ####
179 gastocomp<-consolidado %>% select(c(code,year,Nivel,pibf,gtof,def:
      soc)) %>%
180   filter_at(vars(gtof:soc),any_vars(!is.na(.))) %>%
181   mutate(gtofun=select(.,def:soc) %>% rowSums(),difgto=gtofun-gtof)
182
183 # Gráfica de diferencia a través del tiempo
184 ggplot(gastocomp %>% filter(!is.na(difgto)),aes(x=year,y=difgto,
      color=Nivel))+
185   geom_point()#+scale_x_continuous(limits = c(1990,2020))
186
187 # Diferencia promedio por nivel
```

```
prom_nivel<-gastocomp %>% filter(!is.na(difgto)) %>% mutate(difgtod
    =difgto*pibf/100) %>% group_by(Nivel,year) %>%
  summarise(sdifgto=sum(difgtod),spib=sum(pibf),N=n_distinct(code),
      med=median(difgto)) %>% mutate(pdifgto=sdifgto/spib*100)
ggplot(prom_nivel,aes(x=year,y=med,color=Nivel))+geom_line()
prom_nivel %>% group_by(Nivel) %>%
  summarise(sdifd=sum(sdifgto),spib=sum(spib),mN=mean(N),med=median
      (med)) %>%
  mutate(pdifgto=sdifd/spib*100)
gastocomp %>% filter(!is.na(difgto)) %>% mutate(difgtod=difgto*pibf
    /100, m=mean(difgtod)) %>% group_by(Nivel) %>%
  summarise(avg=sum(difgtod)/sum(pibf)*100,stddev=sd(difgtod)/sd(
      pibf)*100)

# Zoom Nivel ingreso bajo
ggplot(data=gastocomp %>% filter(!is.na(difgto),Nivel=="Bajo"),aes
    (x=year,y=difgto,color=code))+
  geom_point()+geom_line()

# Zoom Nivel ingreso Medio-bajo
ggplot(data=gastocomp %>% filter(!is.na(difgto),Nivel=="Medio-bajo
    "),aes(x=year,y=difgto,color=code))+
  geom_point()+geom_line()+
  geom_text(aes(label=code))
# LTU 1997:
#CHN 2007:
#MNG 2016: la suma por función parece estar baja siempre,
  #la serie tiene huecos que se rellenaron con medias condicionales
      e interpolacionse
  #En 2016 el gasto el gasto

# Resumen media y sd de la diferencia
temp<-gastocomp %>% group_by(code,Nivel) %>% summarise(m=mean(
    difgto,na.rm = TRUE),sd=sd(difgto,na.rm = TRUE)) %>% na.omit()
# Gráfica media(dif) vs sd(dif) con etiquetas del país
ggplot(temp,aes(m,sd,label=code))+geom_point()+
  geom_text_repel()+
  geom_density2d()
# Gráfica media(dif) vs sd(dif) con colores por nivel de ingreso
ggplot(temp,aes(m,sd))+geom_point(aes(col=Nivel))+
  geom_density2d()

plot(x=gastocomp$year,y=gastocomp$difgto)



# más gráficas de la diferencia por año, por nivel de ingreso, por
    continente, resumen por país


consolidado=read.csv(file = "./Data/consolidado.csv",header = TRUE)
consolidado=read.csv(file = "./Data/consolidado2.csv",header = TRUE
    )
######## Resumen Representatividad #######


# Cuadro de representatividad de gasto y gdp de los datos ####
```

```
234 # tratar de representar gráficamente
235
236 # Total países
237 temp<-consolidado %>% group_by(code) %>% summarise(N=n_distinct(
        year))
238 summary(temp$N)
239 ggplot(temp,aes(x=N))+geom_histogram(binwidth = 1)
240 stem(temp$N)
241
242 # por continente
243 temp<-consolidado %>% group_by(cont) %>% summarise(ni=n_distinct(
        code),ti=n_distinct(year))
244 consolidado %>% group_by(cont,code) %>% summarise(ti=n_distinct(
        year)) %>%
245   summarise(ni=n_distinct(code),mint=min(ti),avgt=mean(ti),mediant=
          median(ti),maxt=max(ti))
246 library(janitor)
247 temp %>% adorn_totals()
248 temp %>% gather(dim,conteo,-cont) %>%
249 ggplot(aes(x=cont,y=conteo,fill=dim))+
250   geom_col(position = "dodge")
251 consolidado %>% filter(cont=="Asia") %>%
252   group_by(code) %>% summarise(t=n_distinct(year)) %>% arrange(t)
          %>%
253   mutate(pais=countrycode(code,"genc3c","country.name"))
254 # quitar Timor Leste de la base pues sólo tiene 1 observación
255 consolidado<- consolidado %>% filter(!code=="TLS")
256
257 # por Nivel de ingreso
258 consolidado %>% group_by(Nivel) %>% summarise(ni=n_distinct(code),
        ti=n_distinct(year)) %>%
259   gather(dim,conteo,-Nivel) %>%
260   ggplot(aes(x=Nivel,y=conteo,fill=dim))+
261   geom_col(position = "dodge")
262 consolidado %>% group_by(Nivel,code) %>% summarise(ti=n_distinct(
        year)) %>%
263   summarise(ni=n_distinct(code),mint=min(ti),avgt=mean(ti),mediant=
          median(ti),maxt=max(ti))
264 consolidado %>% group_by(Nivel,code) %>% summarise(ti=n_distinct(
        year)) %>% filter(ti<=3) %>%
265   mutate(pais=countrycode(code,"genc3c","country.name"))
266
267 consolidado %>% group_by(code) %>% summarise(ti=n_distinct(year))
        %>% filter(ti<=6) %>%
268   mutate(pais=countrycode(code,"genc3c","country.name"))
269
270
271 # completar gto cuando se tiene el desglose ####
272 consolidado=read.csv(file = "./Data/consolidado2.csv",header = TRUE
        )
273
274 consolidado=consolidado %>% mutate(gtofuns=select(.,def:soc) %>%
        rowSums(),
275                                    gto2=ifelse(is.na(gtof),gtofuns,
                                         gtof))
276 #quitar timor leste
277 consolidado <- consolidado %>% filter(!code=="TLS")
```

```
278
279 write.csv(consolidado, file = "./Data/consolidado2.csv",row.names =
        FALSE)
280
281 # Cuadros en documento ####
282
283
284 # 1er Tabla # resumen de observaciones por continente
285 tab<-consolidado %>% group_by(cont,code) %>%
286   summarise(nanios=n_distinct(year),Desde=min(year),Hasta=max(year)
          ) %>%
287   summarise(npaises=n_distinct(code),minanios=min(nanios),maxanios=
          max(nanios),Desde=min(Desde),Hasta=max(Hasta)) %>%
288   adorn_totals("row") #%>%
289 names(tab)<-c("Continente","Países","Mín. Años","Máx. Años","Desde"
        ,"Hasta")
290
291 tab[6,3:6]<-t(c(min(tab[1:5,3]),max(tab[1:5,4]),min(tab[1:5,5]),max
        (tab[1:5,6])))
292 tab
293 print(xtable(tab ,digits=1),include.rownames=FALSE)
294
295 # auxiliar para comentario paises con pocos datos
296 consolidado %>% group_by(code) %>%
297   summarise(nyear=n_distinct(year)) %>%
298   mutate(pais=countrycode(code,"genc3c","country.name")) %>%
299   arrange(nyear)
300
301 # 2da tabla # Resumen de representatividad del Gasto y del PIB
302 tab<-consolidado %>% select(code,year,pibf,gto2,pib,gtop) %>% group
        _by(year) %>%
303  summarise(n=n_distinct(code)) # 2017 tiene más datos
304
305 selecWB<-read.csv(file = "./Data/World Bank/selecWB2.csv",header =
        TRUE)
306 extraFMI<-read.csv(file = "./Data/FMI/extraFMI2.csv",header = TRUE)
307
308 temp<-selecWB %>% select(code,year,pib,gtop,pop) %>%# Datos
        completos para abarcar países fuera del GFS
309   merge(extraFMI %>% select(code,year,pibf,gtof),all = TRUE) %>%
310   merge(consolidado %>% select(code,year,gto2,def),all = TRUE)
311
312
313 temp %>% filter(year==2017) %>% # Chequeo variable población
314   mutate(fmi=is.na(pibf),pais=countrycode(code,"genc3c","country.
          name")) %>%
315   filter(!is.na(pais)) %>%
316   group_by(fmi) %>%
317   summarise(spob=sum(pop/1e6,na.rm = TRUE))
318
319 tab<-temp %>% group_by(year) %>%# Inclusion variable indicadora
        dato disponible
320   filter(year>=1970) %>%
321   mutate(across(pib:gto2,~!is.na(.),.names = "{col}.n")) %>%
322   summarise(across(pib.n:gto2.n,sum))
323
```

```r
temp %>% filter(year==2017) %>%# conteo de observaciones según disponibilidad de gto x función
  mutate(dispf=!is.na(def)) %>%
  group_by(dispf) %>%
  summarise(across(pib:gto2,~sum(!is.na(.))))

tab<-temp %>% filter(year==2017,!is.na(pibf)) %>%# resumen pib, gto en $ y población
  mutate(gto2=ifelse(is.na(gto2),gtof,gto2)) %>%#retomar gasto para países que  no tienen detalle
  mutate(dispf=!is.na(def),gtod=gto2*pibf/100) %>%
  group_by(dispf) %>%
  summarise(spib=sum(pibf,na.rm = TRUE),
            sgto=sum(gtod,na.rm = TRUE),
            spob=sum(pop/1e6,na.rm = TRUE),
            n=n_distinct(code))

tabp<-tab %>% adorn_percentages(denominator = "col") %>%# calcular columnas de porcentajes
  set_names(c("dispf","ppib","pgto","ppob","pn")) %>%
  mutate(across(ppib:pn,~.*100)) %>%
  select(-dispf)
tab2<-  tab %>%# combinar tablas anteriores
  mutate(across(spib:spob,~./1000)) %>%
  cbind(tabp) %>%
  adorn_totals("row") %>%
  select(dispf,spib,ppib,sgto,pgto,n,pn,spob,ppob)

print(xtable(tab2,digits=1),include.rownames=FALSE)

# 3ra tabla # Gasto promedio por continente vs México
#datos méxico

#orden funciones
ord<-data.frame(fun=names(consolidado %>% select(def:soc)),
                orden=c(2,4,9,5,1,7,6,3,8,10))
# datos de méxico y total
temp<-consolidado %>% filter(year==2017,!is.na(def),code=="MEX") %>%
  select(cont,def:soc,gto2,pibf) %>%
  mutate(cont="México",ncode=1,pibf=pibf/1000) %>% bind_rows(
  consolidado %>% filter(year==2017,!is.na(def)) %>%
    select(code,cont,def:soc,gto2,pibf) %>%
    mutate(cont="Total") %>%
    group_by(cont) %>%
    mutate(across(def:gto2,~./100*pibf)) %>%#convertir gto a dinero
    summarise(across(def:pibf,sum),ncode=n_distinct(code)) %>%
    mutate(across(def:gto2,~./pibf*100),pibf=pibf/1000) # regresar a porcentaje, pib en miles de millones USD
  )
tab<-consolidado %>% filter(year==2017,!is.na(def)) %>%
  select(code,year,cont,def:soc,gto2,pibf) %>%
  mutate(across(def:gto2,~./100*pibf)) %>%#convertir gto a dinero
  group_by(cont) %>%
  summarise(across(def:pibf,sum),ncode=n_distinct(code)) %>%
```

```
373   mutate(across(def:gto2,~./pibf*100),pibf=pibf/1000) %>%#
          regresar a porcentaje, pib en miles de millones USD
374   bind_rows(temp) %>%# Pegar datos de México
375   pivot_longer(cols = def:ncode,names_to="var",values_to="val") %>%
          #1er paso para transponer
376   pivot_wider(names_from = cont,values_from=val) %>%#2do paso para
          transponer
377   select(var,México,Africa:Total) %>%
378   arrange(desc(Total))
379
380 tab<-tab[c(4:13,3,1,2),]
381 tab$var[11:13]<-c("Total Gasto","PIB","#Paises")
382 print(xtable(tab,digits = 1),include.rownames = FALSE)
383
384 # Resumen de la clasificación de ingreso #
385 temp<-merge(selecWB %>% mutate(code=ifelse(code=="XKX","XKS",code))
        , extraFMI, all = TRUE) %>%
386   merge(consolidado %>% select(code,year,Nivel,gto2),all = TRUE)
          %>%
387   filter(year==2017) %>%
388   select(code,year,pibpc,pib,pibf,pop,gto2) %>%
389   mutate(fmi=is.na(pibf),pais=countrycode(code,"genc3c","country.
          name"),
390          pop=pop/1e6) %>%
391   filter(!is.na(pais)) #%>%
392 temp %>%
393   summarise(across(pibf:pop,sum,na.rm=TRUE))
394
395 temp %>% ggplot(aes(x=pop))+
396   geom_point(aes(y=pibpc))+
397   geom_point(aes(y=pibf*pop),col="blue")
398   scale_y_log10()+
399   scale_x_continuous(trans = "log")
400
401 tab<-selecWB %>% mutate(code=ifelse(code=="XKX","XKS",code)) %>%
402   merge(extraFMI, all = TRUE) %>%
403   merge(consolidado %>% select(code,year,gto2),all = TRUE) %>%
404   merge(wbclasif,all.x = TRUE) %>%
405   filter(year==2017) %>%
406   select(code,year,Nivel,pibpc,pib,pibf,pop,gtof,gto2) %>%
407   mutate(gto2=ifelse(is.na(gto2),gtof,gto2)) %>%
408   mutate(fmi=is.na(pibf),pais=countrycode(code,"genc3c","country.
          name")) %>%#agregar nombre de país
409   filter(!is.na(pais)) %>%# para quitar entradas que no son países
410   mutate(Nivel=ifelse(code=="TWN","Alto",Nivel)) %>%#recuperar
          nivel de Taiwan que no está en la base
411   group_by(Nivel) %>%
412   mutate(gto2=gto2*pibf/100,pibpc=pibpc*pop,pib=pib/1E9) %>%# re
          escalar variables
413   summarise(across(pibpc:gto2,sum,na.rm = TRUE),n=n()) %>%
414   mutate(pibpc=pibpc/pop,pop=pop/1E6) %>%# regresar a proporciones
415   mutate(across(pibpc:gto2,~./1000)) %>%# re escalar a miles
416   arrange(pibpc) %>%
417   adorn_totals("row") %>%
418   select(Nivel, n,pop,pibpc,pib)
419 tab
420 tab[5,4]<-tab[5,5]/tab[5,3]
```

```
421
422 print(xtable(tab, digits = 1), include.rownames = FALSE)
```

../Code/3_concentracion.R

A.1.4. Estadística descriptiva

```
1  # paquetes ####
2  library(tidyverse)
3  library(ggrepel)
4  library(ggsci) # paletas de colores de journals
5  library(scales) # para break formatting functions, show_col
6  library(ggplot2)
7  library(countrycode)
8  library(compositions)
9  library(xtable)
10
11 # Funciones
12 nivtrans<-function(x, nivel="Nivel"){# para ordenar niveles de
       ingreso
13   x %>% mutate(Nivel=factor(get(nivel), levels = c("Bajo","Medio-
       bajo","Medio-alto","Alto","México")))
14 }
15 huecos<-function(x){ #resumen países con huecos en su serie
16   x %>%#pivot_longer(!c(code,year,Nivel),names_to="serie") %>%
17     group_by(code) %>%
18     summarise(n=n_distinct(year),miny=min(year),maxy=max(year),
           huecos=maxy-miny-n+1) %>%
19     filter(huecos>0)
20 }
21
22 obsres<-function(x){# evolución de total de observaciones tras
       añadir cada serie
23   obs<-rep(0,length(names(x)))
24   names(obs)<-names(x)
25
26   for (i in names(x)) {
27     temp<-x %>% select(code:as.name(i))
28     obs[i]<-dim(temp %>% na.omit())[1]
29   }
30   res<-as.data.frame(obs)
31   res$delta<-c(0,diff(res$obs))
32   return(res)
33 }
34
35 codeobsres<-function(x){# resumen de países ordenado por número de
       años obs (menor a mayor)
36   x %>% group_by(code) %>% summarise(n=n_distinct(year)) %>%
       arrange(n)
37 }
38 codemed<-function(codigo,metodo=TRUE){ #media composicional por
       país
39   dat<-amalgA %>%as.data.frame() %>%
40     rownames_to_column("id") %>%
```

```
41      separate(col="id",into = c("code","year")) %>% select(-year)
            %>%
42      filter(code==codigo) %>% select(-code) %>%
43      acomp()
44    resultado<-mean(dat,robust=metodo) %>% as.matrix()
45    return(resultado)
46  }
47
48
49  # Directorio
50  setwd("C:/Users/sgome/Dropbox/#tesis")
51
52  # Lectura de archivos new ####
53
54  consolidado=read.csv(file = "./Data/consolidado2b.csv",header =
        TRUE) %>%#sin filtrar NAs
55    nivtrans()
56
57
58  minval<-consolidado %>% filter(across(def:soc,~.>0)) %>% summarise(
        across(def:soc,~min(.)/10,na.rm=TRUE)) %>%
59    pivot_longer(cols = def:soc,names_to="fun",values_to="min")
60
61
62  # Logratio Analysis ####
63  funciones <- consolidado %>%
64    mutate(Nivel=ifelse(code=="TWN","Alto",as.character(Nivel))) %>%
          #completar Taiwan
65    mutate(cont=countrycode(code,"genc3c","continent")) %>%#
          completar continente
66    mutate(cont=ifelse(code=="XKS","Europe",cont)) %>%#completar
          continete para Kosovo
67    #mutate(Nivel=ifelse(code=="MEX","México",Nivel)) %>%#México en
          categoría separada
68    mutate(across(c(def:soc),~ifelse(.==0,min(minval$min),.))) %>%#
          imputación valores pequeños
69    select(code:soc,gto2,cre) %>% filter(across(code:gto2,~!is.na(.))
          ) %>% nivtrans()
70  obsres(funciones)
71  codeobsres(funciones)
72  levels(funciones$Nivel)
73
74  # 1 Descriptive ####
75  gfun_comp<-funciones %>% unite("id",code:year,sep = "'") %>% column
        _to_rownames("id") %>%
76    select(def:soc) %>% acomp()
77
78  gfun_sub<-funciones %>% unite("id",code:year,sep = "'") %>% column_
        to_rownames("id") %>%
79    mutate(otros=seg+viv+cul+amb+def) %>%
80    #select(soc,gob,edu,sal,eco,otros) %>%# con amalgamación otros
81    select(soc,gob,edu,sal,eco) %>%# subcomposición
82    acomp()
83
84  # 2.2 Variable Selection ####
85
86  orden<-mean(gfun_comp) %>% sort(decreasing = T) %>% names()
```

```
87
88 varD<-variation(gfun_comp[,orden])
89 varD
90
91 tabla<-varD %>% as.data.frame() %>% bind_rows(mean(gfun_comp)) %>%
92   bind_rows(cumsum(mean(gfun_comp[,orden])))
93 print(xtable(tabla,digits=2))
94 #min(varD);max(varD)
95 #summary(gfun_comp)
96
97 vardend<-varD %>% as.dist() %>% hclust(method = "ward.D") %>%
98   as.dendrogram()
99
100 pdf("./TeX/Fig/4.3.2_1_vardend.pdf",width = 7,height = 3.5)
101 par(mar=c(2,3,1,1))
102 plot(vardend)
103 dev.off()
104
105
106 # correlaciones con crecimiento pib
107 head(amalgA)
108 head(amalgA[,1:3])
109 indice<-!is.na(funciones$cre)
110
111 partes<-names(gfun_comp)
112 sapply(partes,function(x) cor(gfun_comp[indice,x],funciones[indice
      ,]$cre)) %>% sort(decreasing = T)
113 head(alr(gfun_comp[indice,c(1:4,6:10,5)])$"soc")
114 cor(gfun_comp[indice,"def"],funciones[indice,]$cre)
115 partes[-5]
116
117 temp<-alr(gfun_comp[indice,c(1:4,6:10,5)]) %>% as.data.frame()
118 sapply(partes[-5],function(x) cor(temp[,x],
119                                   funciones[indice,]$cre)) %>% sort
                                       (decreasing = T)
120
121
122 # amalgama para simplificar ####
123
124 amalgA<-groupparts(gfun_comp,"resto"=c("seg","def","amb","cul","viv
      "))
125 orden<-(mean(amalgA) %>% sort(decreasing = T) %>% names())[c
      (1,3:6,2)]
126 amalgA<-amalgA[,orden]
127 mean(amalgA)
128 # resumen variables seleccionadas ####
129
130 summary(alr(amalgA))
131 library(compositions)
132
133
134
135 png("./TeX/Fig/4.3.2_2_compscatter.png",width = 1080,height = 1080)
136
137 plot(amalgA,margin = "gob",cex=0.5)
138 r = sqrt( qchisq(p=0.95,df=2) )
139 mn = mean(amalgA,robust = T)
```

```
vr = var(amalgA,robust = T)
ellipses(mean=mean(amalgA), var=var(amalgA), r=r, lwd=2 ,col="green")
ellipses(mean=mn, var=vr, r=r, lwd=2 ,col="red")
dev.off()

library(psych)
tabla<-describe(alr(amalgA))[,c(3:5,8:12)]
tabla
print(xtable(tabla))

# normalidad preliminar ####
library(MVN)
mvn(alr(amalgA),univariateTest = "Lillie",mvnTest = "dh",desc = T)
mvn(alr(amalgA)[,2:4],univariateTest = "Lillie",mvnTest = "dh",desc = T)

qchisq(.95,2)

temp<-mvn(alr(amalgA)[,c(2,1)],univariateTest = "Lillie",mvnTest = "dh",desc = F,multivariatePlot = "contour")
mvn(alr(amalgA)[,c(1,2)],univariateTest = "Lillie",mvnTest = "dh",desc = F,multivariatePlot = "contour")
mvn(alr(amalgA)[,c(2,3)],univariateTest = "Lillie",mvnTest = "dh",desc = F,multivariatePlot = "contour")
mvn(alr(amalgA)[,c(2,4)],univariateTest = "Lillie",mvnTest = "dh",desc = F,multivariatePlot = "contour")
mvn(alr(amalgA)[,c(3,4)],univariateTest = "Lillie",mvnTest = "dh",desc = F,multivariatePlot = "contour")

X<-amalgA[,c(2:4,6)]
head(amalgA)
head(alr(X))
mvn(alr(X),univariateTest = "Lillie",mvnTest = "hz",desc = F)
library(robCompositions)
out<-outCoDa(X)$outlierIndex
mvn(alr(X)[!out,c(2,3)],univariateTest = "Lillie",mvnTest = "hz",desc = F,multivariatePlot = "contour")

# media de paises ####
lpaises<-unique(funciones$code)
nobs<-funciones %>% group_by(code) %>% summarise(n=n())
lpaisesA<-nobs$code[nobs$n>=2*(ncol(amalgA)-1)] # con 5 o más datos
lpaisesB<-nobs$code[nobs$n<2*(ncol(amalgA)-1)] # con menos de 5 datos
media<-sapply(lpaises, function(x) if (x%in%lpaisesA) codemed(x,metodo = T) else codemed(x,metodo=F)) %>% t() %>%
  as.data.frame()


paises<-read.csv("./Data/nomb_paises.csv",encoding = "UTF-8") %>%
  merge( funciones %>%group_by(code) %>% mutate(Nivel=last(Nivel)) %>% select(code,Nivel) %>% unique()) %>%
  mutate(ni=Nivel)
levels(paises$ni)<-list(B="Bajo",MB="Medio-bajo",MA="Medio-alto",A="Alto")

```

```
rownames(media)<-paises$esp
names(media)<-names(amalgA)
head(media)
mean(media %*% acomp)

write.csv(media,"./Data/FMI/mediapaises.csv")


# PCA
pcagf<-princomp(media,cor = T)
plot(pcagf)
loadings(pcagf)
plot(pcagf,type="lines")
biplot(pcagf,scale = 1)

library(robCompositions)
pcagf<-pcaCoDa(media,method = "classical")
summary(pcagf)
set.seed(1906)
pcagf<-pcaCoDa(media,method = "robust")
summary(pcagf)
biplot(pcagf,xlabs=rownames(media))

n=4
set.seed(1906)
res.km <- kmeans(scale(alr(media)), centers=n)

# alr PCA ####
n=6
res.pca <- prcomp(alr(media),  scale = TRUE)
#biplot(res.pca)

# Coordinates of individuals
library(factoextra) # para get_pca_ind
ind.coord <- as.data.frame(get_pca_ind(res.pca)$coord)
# Add clusters obtained using the dendrogram cutree

dendpai<-media %*% alr() %*% dist() %*% hclust(method = "ward.D")
ind.coord$Grupo<-factor(cutree(dendpai,n))
# Add Species groups from the original data sett
ind.coord$code=paises$code; ind.coord$pais=paises$esp
ind.coord<-ind.coord %*%
  merge( funciones %*%group_by(code) %*% mutate(Nivel=last(Nivel))
       %*% select(code,Nivel) %*% unique())
# Percentage of variance explained by dimensions
eigenvalue <- round(get_eigenvalue(res.pca), 1)
variance.percent <- eigenvalue$variance.percent
head(eigenvalue)

# Coeficientes de PCA ####
# Helper function
#::::::::::::::::::::::::::::::::::::::::
var_coord_func <- function(loadings, comp.sdev){
  loadings*comp.sdev
}
# Compute Coordinates
#::::::::::::::::::::::::::::::::::::::::
```

```
loadings <- res.pca$rotation
sdev <- res.pca$sdev
var.coord <- t(apply(loadings, 1, var_coord_func, sdev))
head(var.coord[, 1:4])
print(xtable(var.coord))

#datos transformados completos vs Grupo
fgrupo<-alr(amalgA) %>% as.data.frame() %>%mutate(code=funciones$
    code,year=funciones$year) %>%
  merge(ind.coord %>% select(code,Grupo))

#Paises por grupo
temp<-ind.coord %>% group_by(Grupo) %>% summarise(N=n())
ndato<-temp %>% merge(fgrupo %>% group_by(Grupo) %>% summarise(Ti=n
    ())) %>%
  merge(ind.coord %>% group_by(Grupo) %>% summarise(across(Dim.1:
      Dim.2,mean))) %>%
  mutate(etiq=paste(N,"paises"))
#Ajustes para visualizar etiquetas de grupo
ndato$Dim.1_adj<-ndato$Dim.1+c(0.5,-0.25,-0.5,-1,0.5,-0.5)
ndato$Dim.2_adj<-ndato$Dim.2+c(-0.5,0.95,0.75,1,0.5,-0.5)

write.csv(ind.coord,"./Data/FMI/mediaspaisesPCA.csv",row.names = F)

library(ggpubr)
pcagrupo<-ggscatter(
  ind.coord, x = "Dim.1", y = "Dim.2",
  color = "Grupo", palette = "npg", ellipse = TRUE, ellipse.type =
      "convex",
  shape = "Nivel", size = 2, rug=F, legend = "right", ggtheme =
      theme_bw(),
  show.legend.text = TRUE,
  xlab = paste0("Componente 1 (", variance.percent[1], "%)" ),
  ylab = paste0("Componente 2 (", variance.percent[2], "%)" )
) +
  stat_mean(aes(col=Grupo), size = 3)+
  coord_fixed(xlim = c(-6,3),ylim = c(-4,2))+#default xlim = c
      (-5.1,2.46),ylim = c(-3.88,1.65)
  #geom_text(aes(Dim.1,Dim.2,label=paises$esp))
  geom_text(data = subset(ind.coord,code=="MEX"),aes(Dim.1,Dim.2,
      label=pais),nudge_x = -0.2,nudge_y = -0.2)+
  geom_text(data=ndato, aes(Dim.1_adj,Dim.2_adj,label=etiq))
pcagrupo
ggsave(filename ="./TeX/Fig/4.3.2_4_PCAmedia.pdf",width = 15,height
    = 10,units = "cm" )


# Dendrograma circular ####

#https://cran.r-project.org/web/packages/dendextend/vignettes/
    dendextend.html
library(circlize)
library(dendextend)

ordend<-dendextend::get_nodes_attr(dend,"label")
ordend<-ordend[!is.na(ordend)] %>% as.data.frame()
names(ordend)<-"esp"
```

```
289 ordend$pos<-c(rep("r",17),rep("l",33),rep("r",17))
290
291 paises <- paises %>% merge(ordend) %>%
292   mutate(paisN=ifelse (pos=="r", paste(esp,ni,sep=" "),paste(ni,
          esp,sep=" "))) %>%
293   arrange(code)
294
295
296 paleta<-pal_npg()(n)[c(1,3,4,2,6,5)]
297
298 dend <- dendpai %>% as.dendrogram %>%
299   set("branches_lwd", 2) %>%
300   set("branches_lty", 1) %>%
301   set("labels_cex", 1) %>%
302   color_branches(k=n,col =paleta) %>%
303   color_labels(k=n,col =paleta)
304
305 par(mar= rep(0,4))
306 pdf("./TeX/Fig/4.3.2_3_dendro.pdf",width=7, height=5)
307 circlize_dendrogram(dend,facing = "outside",
308                     labels_track_height = .4,
309                     dend_track_height = 0.5)
310 dev.off()
311
312
313 # Datos completos en PCA ####
314 #http://www.sthda.com/english/articles/31-principal-component-
       methods-in-r-practical-guide/118-principal-component-analysis-
       in-r-prcomp-vs-princomp/
315
316
317 ind.sup.coord <- predict(res.pca, newdata = fgrupo) %>%
318   as.data.frame() %>%
319   mutate(code=funciones$code,year=funciones$year,Nivel=funciones$
          Nivel) %>%
320   merge(ind.coord %>% select(code,Grupo))
321 names(ind.sup.coord)[2:6]<-paste("Dim",1:5,sep = ".")
322
323 head(ind.sup.coord)
324
325
326 pcafullgrupo<-ggscatter(
327   ind.sup.coord, x = "Dim.1", y = "Dim.2",
328   color = "Grupo", palette = "npg", #ellipse = TRUE, ellipse.type =
          "convex",
329   shape = "Nivel", size = 2, rug=F, legend = "right", ggtheme =
          theme_bw(),
330   show.legend.text = TRUE,
331   xlab = paste0("Componente 1 (", variance.percent[1], "%)" ),
332   ylab = paste0("Componente 2 (", variance.percent[2], "%)" )
333 ) +
334   #coord_cartesian(xlim = c(-5.5,3))+
335   #stat_mean(aes(col=Grupo), size = 3)#+
336   coord_fixed(xlim = c(-6,3),ylim = c(-4,2))#+#default xlim = c
          (-5.1,2.46),ylim = c(-3.88,1.65)
337 pcafullgrupo
338
```

```
temp<-ggplot_build(pcagrupo)
# Normalidad datos agrupados ####

temp<-ind.sup.coord %>% select(Dim.1,Dim.2,Grupo)
mvn(data = temp, subset = "Grupo",
    univariateTest = "Lillie",mvnTest = "dh",desc = F)#$
        multivariateNormality

temp<-fgrupo %>% select(soc:resto,Grupo)
mvn(data = temp, subset = "Grupo",
    univariateTest = "Lillie",mvnTest = "dh",desc = F)

temp<-fgrupo %>% select(edu:sal,Grupo)
mvn(data = temp, subset = "Grupo",
    univariateTest = "Lillie",mvnTest = "dh",desc = F)


# Distribuciones marginales
fgrupo %>%#select(soc:resto,Grupo) %>%
  pivot_longer(cols = soc:resto,names_to="fun",values_to="val") %>%
  mutate(fun=factor(fun,levels = c("soc","edu","eco","sal","resto")
      )) %>%
  ggplot(aes(x=val))+
  geom_density(size=1)+
  facet_wrap(~fun)+
  coord_cartesian(xlim = c(-4,2))

library(GGally)
fgrupo %>%
  ggpairs(columns = 2:6, #ggplot2::aes(color=Grupo),
          diag = list(discrete="barDiag",
                      continuous = wrap("densityDiag", alpha=0.5 )))
                          +
  scale_color_npg()+
  scale_fill_npg()
# Curvas Densidad en alr o ternaria
plot(amalgA[,c(1:4,6)],margin = "gob",pca = T,col.pca = "blue",
    robust = T)
plot(amalgA[,c(1:4,6)],margin = "gob",pca = T,col.pca = "blue",
    robust = F)


#normality ###
#library(MVN)
library(robCompositions)

subn<-names(fgrupo)[c(2:6,8)] #todas las variables
subn<-names(fgrupo)[c(3:5,8)] # sin soc o el resto
mvn(data = fgrupo[,c(2:6,8)], subset = "Grupo",
    univariateTest = "Lillie",mvnTest = "dh",desc = F)$
        multivariateNormality

set.seed(1511)
sapply(1:6,function(x) norm_dendgrupo2(fgrupo,grupo = x,out = T,
    colu = 1:5))

set.seed(1906)
```

```
sapply(1:6,function(x) norm_dendgrupo2(fgrupo,grupo = x,out = T,
    colu = 2:4))


X<-amalgA[fgrupo$Grupo==4,c(2:4,6)]
plot(X,margin = "gob")
out<-outCoDa(X);out
outlier<-out$outlierIndex
rownames(X)[out$outlierIndex]
plot(out)
head(X)
plot(X,margin="gob",col=1+as.numeric(out$outlierIndex))


pruebas<-c("mardia","hz","royston","dh")
temp<-sapply(pruebas,function(x) mvn(data = alr(X[!outlier,]),
                                     univariateTest = "Lillie",mvnTest =
                                         x,desc = T)$
                                         multivariateNormality)
temp<-mvn(data = alr(X[!outlier,]),
          univariateTest = "Lillie",mvnTest = "royston",desc = T)
temp %>% as.data.frame()


#1-pchisq(estadístico,df)
# grupo, prueba, estadístico, p value, MVN

pcagrupo
pruebas<-c("mardia","hz","royston","dh")

X<-amalgA[fgrupo$Grupo==4,c(2:4,6)]
set.seed(1906)
out<-outCoDa(X)
rownames(X)[out$outlierIndex]

mvnpruebas<-sapply(1:6, function(k){
  X<-amalgA[fgrupo$Grupo==k,c(2:4,6)]
  set.seed(1906)
  out<-outCoDa(X)
  #if(k==4) out<-outCoDa(X[,c(2:4,6)])
  outlier<-out$outlierIndex
  res<-sapply(pruebas,
              function(x){
                prueba<-mvn(data =alr(X[!outlier,]),
                            univariateTest = "Lillie",
                            mvnTest = x,desc = T)$
                                multivariateNormality
                if (x=="mardia") mvn<- prueba$Result[3] else mvn<-
                    prueba$MVN
                return(mvn)

              })
  return(res)
})
mvnpruebas

mvnpruebas<-sapply(1:6, function(k){
```

```
X<-amalgA[fgrupo$Grupo==k,]
set.seed(1115)
out<-outCoDa(X)
if(k==3|k==4) out<-outCoDa(X[,c(2:4,6)])
outlier<-out$outlierIndex
nout<-c(sum(outlier),sum(outlier)/nrow(X)*100)
names(nout)<-c("atipicos","porcen")
res<-sapply(pruebas,
            function(x){
              prueba<-mvn(data =alr(X[!outlier,c(2:4,6)]),
                   univariateTest = "Lillie",
                   mvnTest = x,desc = T)$multivariateNormality
              if (x=="mardia") mvn<- prueba$Result[3] else mvn<-
                   prueba$MVN
              return(mvn)

  })
#res<-c(nout,res)
return(res)
})
mvnpruebas

nout<-c(sum(outlier),sum(outlier)/nrow(X)*100)
names(nout)<-c("atipicos","porcen")
nout
# grupo, prueba (dh), estadístico, p value, MVNs
# subcomposición edu, eco, sal resumen pruebas ####
library(robCompositions)
library(MVN)
k=1
for(k in 1:6){
X<-amalgA[fgrupo$Grupo==k,]
set.seed(1115)
out<-outCoDa(X)
#out
#plot(out$mahalDist) # distancia mahalanobis
#text(1:nrow(X),out$mahalDist,rownames(X))
#text(1:nrow(X),out$mahalDist*out$outlierIndex,rownames(X),col="red
    ")
if(k==4) out<-outCoDa(X[,c(2:4,6)])
outlier<-out$outlierIndex
prueba<-mvn(data =alr(X[!outlier,c(2:4,6)]),
            univariateTest = "Lillie",
            mvnTest = "dh",desc = T)

multi<-prueba$multivariateNormality
multi$observaciones<-nrow(X)
multi$atipico<-sum(outlier)
multi$per<-multi$atipico/multi$observaciones*100

marginal<-prueba$univariateNormality
# extrae p-values
pvalues<-trimws(marginal[,4])

# si rechaza H0 pone asterísco junto a p-value
nrechaza<-trimws(t(marginal[,5]))
nrechaza[nrechaza=="YES"]<-" "
```

```
nrechaza[nrechaza=="NO"]<-"*"
univ<- paste(pvalues,nrechaza,sep = "")

multi<-cbind(multi,t(univ))
# guardar: mahalDist, indicador atípico
temp<-data.frame(id=rownames(X),mD=out$mahalDist,outlier=out$
    outlierIndex)

if(k==1){
  res<- multi
  atipicos<-temp
}
else {
  res<-rbind(res,multi)
  atipicos<-rbind(atipicos,temp)
  }
}
rownames(res)<-1:6
names(res)[9:11]<-c("edu","eco","sal")

res
print(xtable(res[,c(2:4,6:11)],digits = 2))
#incluir resumen de outliers en base transformada alr
atipicos<-atipicos %>% separate(id,c("code","year"))
fgrupo<-fgrupo %>% merge(atipicos)
write.csv(fgrupo,"./Data/gto_conglo.csv",row.names = F)
#resumen mD
fgrupo %>% group_by(Grupo) %>% summarise(across(mD:outlier,list("m"
    =mean,"sd"=sd,"max"=max,"min"=min)))

# No normalidad de soc y resto ####
#pruebas
for(k in 1:6){
  X<-amalgA[fgrupo$Grupo==k,]
  set.seed(1115)
  out<-outCoDa(X)
  outlier<-out$outlierIndex
  prueba<-mvn(data =alr(X[!outlier,]),
              univariateTest = "Lillie",
              mvnTest = "dh",desc = T)

  multi<-prueba$multivariateNormality
  multi$observaciones<-nrow(X)
  multi$atipico<-sum(outlier)
  multi$per<-multi$atipico/multi$observaciones*100

  marginal<-prueba$univariateNormality
  # extrae p-values
  pvalues<-trimws(marginal[,4])

  # si rechaza H0 pone asterísco junto a p-value
  nrechaza<-trimws(t(marginal[,5]))
  nrechaza[nrechaza=="YES"]<-" "
  nrechaza[nrechaza=="NO"]<-"*"
  univ<- paste(pvalues,nrechaza,sep = "")

  multi<-cbind(multi,t(univ))
```

```
551
552   if(k==1) {
553     res<-multi
554     }
555   else{
556     res<-rbind(res,multi)
557     }
558 }
559
560 names(res)[9:13]<-c("soc","edu","eco","sal","resto")
561
562 res
563 print(xtable(res[,c(2:4,6:9,13)],digits = 2))
564 #sesgo / kurtosis
565 library(psych)
566 describeBy(fgrupo[,2:6],fgrupo$Grupo)
```

../Code/4_pais_conglo.R

A.1.5. Modelos de panel e inferencia estadística

```
1  # Regresiones
2
3  # paquetes ####
4  library(tidyverse)
5  library(plm)
6  library(ggrepel)
7  library(ggsci) # paletas de colores de journals
8  library(scales) # para break formatting functions
9  library("lmtest") #para coeftest
10 library(ggplot2)
11 library(countrycode)
12 library(compositions)
13 library(e1071) #para sesgo y kurtosis
14 library(stargazer)
15 library(xtable)
16 library(MVN)
17 library(janitor)
18
19 nivtrans<-function(x){# para ordenar niveles de ingreso
20   x %>% mutate(Nivel=factor(Nivel,levels = c("Bajo","Medio-bajo","
       Medio-alto","Alto","México")))
21 }
22
23 # datos
24 setwd("C:/Users/sgome/Dropbox/#tesis")
25 consolidado=read.csv(file = "./Data/consolidado2b.csv",header =
       TRUE) %>%#sin filtrar NAs
26   nivtrans()
27 paises<-read.csv("./Data/nomb_paises.csv",encoding = "UTF-8")
28 idhs<-read.csv("./Data/UN/variaciones/ind2.csv") # variantes
       índices desarrollo
29
30 minval<-consolidado %>% filter(across(def:soc,~.>0)) %>% summarise(
       across(def:soc,~min(.)/10,na.rm=TRUE)) %>%
```

```
  pivot_longer(cols = def:soc,names_to="fun",values_to="min")
lsfun<-list("media"=mean,"dev"=sd,
            "p25"=function(x,na.rm=T) quantile(x,probs = 0.25,na.rm
                =),
            "p75"=function(x,na.rm=T) quantile(x,probs = 0.75,na.rm
                =),
            "mín"=min,"máx"=max,
            "sesgo"=skewness,"kurtosis"=kurtosis)

# Datos con PIB ####
funciones <- consolidado %>% filter(code!="TLS") %>%# quitar pais
    con pocas observaciones
  mutate(Nivel=ifelse(code=="TWN","Alto",as.character(Nivel))) %>%
      #completar Taiwan
  mutate(cont=countrycode(code,"genc3c","continent")) %>%#
      completar continente
  mutate(cont=ifelse(code=="XKS","Europe",cont)) %>%#completar
      continete para Kosovo
  #mutate(Nivel=ifelse(code=="MEX","México",Nivel)) %>%#México en
      categoría separada
  mutate(across(c(def:soc),~ifelse(.==0,min(minval$min),.))) %>%#
      imputación valores pequeños
  select(code:soc,gto2,cre,pibpc) %>% filter(across(code:gto2,~!is.
      na(.))) %>% nivtrans()
head(funciones)
levels(funciones$Nivel)


fgrupo<-read.csv("./Data/gto_conglo.csv") %>%# grupos y resumen de
    atípicos por grupo
  mutate(peso=ifelse(outlier,1/mD,1),Grupo=as.factor(Grupo))
levels(fgrupo$Grupo)

fgrupo %>%group_by(outlier) %>% summarise(across(mD,lsfun))
fgrupo %>%group_by(outlier) %>% summarise(across(peso,lsfun))

# g_t
temp<-funciones %>% select(code,year,cre,pibpc) %>%
  pdata.frame() %>%
  group_by(code) %>%
  mutate(g=100*pibpc/dplyr::lag(pibpc)-100)

plot(temp$cre,temp$g)

g_dat<-funciones %>% select(code,year,Nivel,pibpc,cre,gto2,def:soc)
    %>%
  mutate(resto=seg+def+amb+cul+viv) %>% select(-c(seg,def,amb,cul,
      viv)) %>%#amalgama
  mutate(across(eco:resto,~log(./gob))) %>% select(-gob) %>%#
      logcocientes
  merge(fgrupo %>% select(code,year,Grupo:peso)) %>%# variables de
      grupos
  mutate(gtosq=gto2^2) %>% filter(year>=1990) %>%# gasto^2 y
      filtro año
  filter(!is.na(cre)) %>%# sólo datos disponibles
  mutate(d92=ifelse(year==1992,1,0),d93=ifelse(year==1993,1,0),
```

```
72          d08=ifelse(year==2008,1,0),d09=ifelse(year==2009,1,0)) %>%
                #indicadoras crisis
73   pdata.frame()
74 pdim(g_dat)
75
76 g_mco<-plm(cre~gto2+soc+edu+eco+sal+resto,g_dat,model = "pooling")
       # ols
77 g_int<-plm(cre~gto2+soc+edu+eco+sal+resto,g_dat,model = "within",
       effect = "individual") # within
78 g_mcg<-plm(cre~gto2+soc+edu+eco+sal+resto,g_dat,model = "random",
       effect = "twoways") # gls
79
80 lmodel<-list(g_mco,g_int,g_mcg)
81 fnames<-c("soc","edu","eco","sal","resto")
82 sgama<-sapply(lmodel,function(x) sum(x$coefficients[fnames]))
83
84 stargazer(g_mco,g_int,g_mcg, type="text",
85           digits = 2,
86           title = "Regresión de panel para el crecimiento económico
                ",
87           column.labels = c("$\\hat{\\und\\gamma}_{\\text{MCO}}$",
88                             "$\\hat{\\und\\gamma}_{\\text{W}}$",
89                             "$\\hat{\\und\\gamma}_{\\text{MCG}}$"),
90           single.row = T,
91           dep.var.caption = "Variable dependiente: $g_t$",
92           omit.stat = "F",
93           #dep.var.labels = "$g_t$",
94           label = "tab:gmod0",
95           dep.var.labels.include = F,
96           covariate.labels=c("gto","$\\log$(soc/gob)","$\\log$(edu/
                gob)",
97                              "$\\log$(eco/gob)","$\\log$(sal/gob)",
98                              "$\\log$(resto/gob)","ordenada"))
99
100 coeftest(g_mco,vcov. = vcovHC)
101 coeftest(g_mco2,vcov. = vcovHC)
102 coeftest(g_int,vcov. = vcovHC)
103 coeftest(g_int2,vcov. = vcovHC)
104 coeftest(g_mcg,vcov. = vcovHC)
105 coeftest(g_mcg2,vcov. = vcovHC)
106
107 coeftest(g_mco)
108 coeftest(g_mco2)
109 coeftest(g_int2)
110 fixef(g_int2)
111
112 # Coeficientes ####
113 # variables x Nivel
114 consolidado %>% filter(code!="TWN",year==2017) %>% group_by(Nivel)
       %>%
115   summarise(gto=mean(gto2,na.rm = T),
116             g=mean(cre,na.rm = T),
117             n=n_distinct(code)) #%>% adorn_totals()
118
119
120 g_mco$coefficients
121
```

```
sum(g_mco$coefficients[fnames])
sum(g_int$coefficients[fnames])
sum(g_mcg$coefficients[fnames])

mean(g_dat$cre)
gm<-g_dat %>% filter(code=="MEX",year==2019) %>% select(eco:resto,
    gto2)
gm<-funciones %>% filter(code=="MEX",year==2019) %>%
  select(def:soc,gto2) %>% mutate(resto=def+amb+viv+seg+cul) %>%
      select(-c(def,amb,viv,seg,cul)) %>%
  relocate(gto2,.after=resto)
gm

logob<-function(x) x %>% mutate(across(eco:resto,~log(./gob))) %>%
  select(-gob)
logob(gm)
predict(g_mco,logob(gm))-
predict(g_mco,logob(gm+c(0,1,0,0,-1,0,0)))

predict(g_mco,logob(gm))-
  predict(g_mco,logob(gm+c(0,0,-1,1,0,0,0)))

fitted.values(g_mco)
# g plots ####
# Por nivel de ingreso
opciones<-list(theme_bw()+xlab("")+ylab(""),
                  coord_cartesian(ylim = c(0,6)),
                  scale_fill_locuszoom(),
                  theme(legend.position = "none",axis.text.x =
                      element_text(angle = 90),
                        axis.text.y = element_blank(),axis.ticks=
                            element_blank()))
temp<-consolidado %>% filter(year>=1990,!is.na(cre)) %>% select(
    code,pais,year,Nivel,cont,cre) %>%
  mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel))
      %>% nivtrans()
temp %>% group_by(Nivel) %>% summarise(n=n_distinct(code)) %>%
    adorn_totals()

temp %>% mutate(cont=ifelse(is.na(cont),countrycode(code,"genc3c","
    continent"),cont)) %>%
  group_by(cont,Nivel) %>% summarise(n=n_distinct(code)) %>% adorn_
      totals()
a<-consolidado %>% filter(year>=1990,!is.na(cre)) %>% select(code,
    year,Nivel,cre) %>%
  mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
      %>% nivtrans() %>%#México en categoría separada
  group_by(Nivel) %>%
  summarise(g=mean(cre,na.rm = T),N=n_distinct(code),n=n()/N,g2=100
      *geometricmean(1+cre/100)-100) %>%
  ggplot(aes(x=reorder(Nivel,g),y=g,fill=Nivel)) +
  geom_bar(stat = "identity")+ #coord_flip()+
  geom_text(aes(label=round(g,2)),nudge_y = .5,angle=90,size=3)+
  theme_bw()+xlab("")+ylab("")+
  coord_cartesian(ylim = c(0,6))+
  scale_fill_locuszoom()+
```

```
166   theme(legend.position = "none",axis.text.x = element_text(angle =
          90),
167           axis.text.y = element_blank(),axis.ticks=element_blank())
168
169 # Por continente
170 temp<-consolidado %>% filter(year>=1990,!is.na(cre)) %>%
171   mutate(cont=ifelse(is.na(cont),countrycode(code,"genc3c","
          continent"),cont))
172 temp %>% group_by(cont) %>% summarise(n=n_distinct(code)) %>% adorn
      _totals()
173 b<-consolidado %>% filter(year>=1990,!is.na(cre)) %>%
174   mutate(cont=ifelse(is.na(cont),countrycode(code,"genc3c","
          continent"),cont)) %>%
175   select(code,year,cont,cre) %>%
176   group_by(cont) %>%
177   summarise(g=mean(cre,na.rm = T),N=n_distinct(code),n=n()/N,g2=100
          *geometricmean(1+cre/100)-100) %>%
178   ggplot(aes(x=reorder(cont,g),y=g,fill=cont)) +
179   geom_bar(stat = "identity")+# coord_flip()+
180   geom_text(aes(label=round(g,2)),nudge_y = .5,angle=90,size=3)+
181   theme_bw()+xlab("")+ylab("")+
182   coord_cartesian(ylim = c(0,6))+
183   scale_fill_aaas()+
184   theme(legend.position = "none",axis.text.x = element_text(angle =
          90),
185           axis.text.y = element_blank(),axis.ticks=element_blank())
186
187 # Por grupo gasto público
188 g_dat %>% summarise(n=n_distinct(code))
189 c<-g_dat %>% filter(as.numeric(as.character(year))>=1990,!is.na(cre
      )) %>% select(code,year,Grupo,cre) %>%
190   group_by(Grupo) %>%
191   summarise(g=mean(cre,na.rm = T),N=n_distinct(code),n=n()/N,g2=100
          *geometricmean(1+cre/100)-100) %>%
192   ggplot(aes(x=reorder(paste("Grupo",as.numeric(Grupo)),g),y=g,fill
          =Grupo)) +
193   geom_bar(stat = "identity")+ #coord_flip()+
194   geom_text(aes(label=round(g,2)),nudge_y = .5,angle=90,size=3)+
195   theme_bw()+xlab("")+ylab("")+
196   coord_cartesian(ylim = c(0,6))+
197   scale_fill_npg()+
198   theme(legend.position = "none",axis.text.x = element_text(angle =
          90),
199           axis.text.y = element_blank(),axis.ticks=element_blank())
200
201
202 library(cowplot)
203 allplotslist <- align_plots(a,b,c, align = "hv")
204 library(ggpubr)
205
206 ggarrange(allplotslist[[1]],allplotslist[[2]],allplotslist[[3]],
      ncol=3)
207 ggsave("./TeX/Fig/441_0_ggrupos.pdf",width = 14,height = 8,units =
      "cm")
208
209 g_dat %>% group_by(Grupo) %>% summarise(across(cre:resto,mean)) #
      promedio var por grupo
```

```
210
211 # variables promedio por grupo
212 funciones %>% select(code,year,Nivel,cre,gto2,def:soc) %>%
213   mutate(resto=seg+def+amb+cul+viv) %>% select(-c(seg,def,amb,cul,
        viv)) %>%#amalgama
214   #mutate(across(eco:resto,~log(./gob))) %>% select(-gob) %>%#
        logcocientes
215   merge(fgrupo %>% select(code,year,Grupo:peso)) %>%# variables de
         grupos
216   mutate(gtosq=gto2^2) %>% filter(year>=1990) %>%# gasto^2 y
        filtro año
217   filter(!is.na(cre)) %>%# sólo datos disponibles
218   mutate(d92=ifelse(year==1992,1,0),d93=ifelse(year==1993,1,0),
219          d08=ifelse(year==2008,1,0),d09=ifelse(year==2009,1,0)) %>%
                #indicadoras crisis
220   pdata.frame() %>%
221   group_by(Grupo) %>% summarise(across(cre:resto,mean)) %>%
222   xtable(caption = "Gasto y crecimiento de los países por grupos",
223          label="ggasto_res",digits=1) %>% print(include.rownames=F)
224
225 # Individual effects ####
226
227 #within_intercept Overall Intercept for Within Models Along its
      Standard Error ###
228
229 within_intercept(g_int)
230
231 fixef(g_int,type="dmean")
232
233 mean(fixef(g_int))
234 data.frame(code=names(fixef(g_int,type = "dmean")),fijo=fixef(g_int
      ),alea=ranef(g_mcg)) %>%
235 ggplot(aes(x=reorder(code,fijo),y=fijo))+
236   geom_bar(stat = "identity",position = "dodge")+
237   theme(axis.text.x = element_text(angle=90))
238
239 data.frame(code=names(fixef(g_int)),fijo=fixef(g_int,type = "dmean"
      ),alea=ranef(g_mcg)) %>%
240   merge(g_dat %>% group_by(code) %>% summarise(g=mean(cre))) %>%
241   merge(paises %>% select(code,esp)) %>%
242   ggplot(aes(x=fijo,y=alea,label=esp))+geom_point(aes(size=g)) +
243   geom_text_repel(size=3)
244
245 # Time effects ####
246
247 #ranef.plm Extract the Random Effects
248 ranef(g_mcg)
249 data.frame(year=1990:2019,tempo=ranef(g_mcg,effect = "time")) %>%
250   merge(g_dat %>% group_by(year) %>% summarise(g=mean(cre))) %>%
251   pivot_longer(cols = tempo:g,names_to="serie",values_to="valor")
        %>%
252   ggplot(aes(x=year,y=valor,col=serie))+geom_line()
253
254
255 # %Observaciones totales, países N, años T_i *
256 pdim(g_dat)
257
```

```
# Pruebas ####
# I. presencia de efectos
pFtest(update(g_int2,effect="time"),g_mco2)
pFtest(update(g_int2,effect="twoways"),g_mco2)
pFtest(update(g_int2,effect="twoways"),g_int2)
pFtest(g_int2,g_mco)
pFtest(g_int2,g_mco)

extp<-function(prueba,param=2){# extrae tabla con resultados de la prueba
  if (param==2) {
    valores=data.frame(prueba$statistic,names(prueba$statistic),
        prueba$p.value,prueba$method,
              prueba$parameter[[1]],prueba$parameter[[2]],
              prueba$alternative)
    names(valores)<-c("Estadístico","Dist.","p-value","Tipo","df1",
        "df2","Hipótesis alt.")
    }
  else{
    if(param==1){
      valores=data.frame(prueba$statistic,names(prueba$statistic),
          prueba$p.value,prueba$method,
                         prueba$parameter[[1]],
                         prueba$alternative)
      names(valores)<-c("Estadístico","Dist.","p-value","Tipo","df1
          ","Hipótesis alt.")
    }
    else{
    valores=data.frame(prueba$statistic,names(prueba$statistic),
        prueba$p.value,prueba$method,
              prueba$alternative)
    names(valores)<-c("Estadístico","Dist.","p-value","Tipo","
        Hipótesis alt.")
    }
  }
return(valores)
}


# F test for ind / time effects *
for (i in c("individual","time","twoways")){
  temp<-extp(pFtest(update(g_int,effect=i),g_mco))
  temp$efecto<-i
  if (i=="individual") res<-temp else res<-rbind(res,temp)
}
pruf<-res

# plmtest Lagrange FF Multiplier Tests for Panel Models *
for (i in c("individual","time","twoways")){
  temp<-extp(plmtest(g_mco,effect = i),param = F)
  temp$efecto<-i
  if (i=="individual") res<-temp else res<-rbind(res,temp)
}
prulm<-res

tab<-bind_rows(pruf,prulm) %>%
```

```
307   mutate(across(df1:df2,~ifelse(is.na(.),"",formatC(.,format = "f",
          big.mark = ",",digits = 0)))) %>%
308   mutate(`p-value`=format(`p-value`,format="e",digits = 2)) %>%
309   mutate(Tipo=c("F","","","Honda","","")) %>% dplyr::select(Tipo,
          everything()) %>%
310   mutate(efecto=rep(c("individual","temporal","ambos"),2))
311
312 print(xtable(tab %>% dplyr::select(-`Hipótesis alt.`),
313              caption="Pruebas de la presencia de efectos",
314              label="tab:g0effects"), include.rownames=FALSE)
315
316 # II. Correlación de efectos
317 # %Hausman: cor(X,eta) *
318 phtest(g_int,g_mcg)
319 # H0:RE (mcg) vs H1:FE (int)/ H0 not rejected / testing if cor(e,X)
      <>0 H0: cor(e,X)=0
320
321 head(g_dat)
322 head(g_dat[,c(fnames,"gto2","cre")])
323 tab<-sapply(c(fnames,"gto2","cre"), function (x) cor(fixef(g_int),
      between(g_dat[,x])))
324
325 print(xtable(as.data.frame(t(tab))),include.rownames = F)
326
327 consolidado %>% filter(code!="TWN",year>=1990,!is.na(cre)) %>%
328   mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
          %>% nivtrans() %>%
329   summarise(n=n_distinct(code))
330
331 # Resumen de gasto total y g por nivel con todos los datos
332 temp<-consolidado %>% filter(code!="TWN",year>=1990) %>%
333   mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
          %>% nivtrans() %>%#México en categoría separada
334   group_by(Nivel) %>%
335   summarise(gto=mean(gto2,na.rm = T),
336             g=mean(cre,na.rm = T),
337             n=n_distinct(code))
338 temp %>% adorn_totals()
339
340 # Tabla resumen por Nivel
341 funciones %>% dplyr::select(code,year,Nivel,cre,gto2,def:soc) %>%
342   mutate(resto=seg+def+amb+cul+viv) %>% dplyr::select(-c(seg,def,
          amb,cul,viv)) %>%#amalgama
343   mutate(gtosq=gto2^2) %>% filter(year>=1990) %>%# gasto^2 y
          filtro año
344   mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
          %>% nivtrans() %>%#México en categoría separada
345   pdata.frame() %>%
346   group_by(Nivel) %>% summarise(across(eco:resto,mean)) %>%
347   mutate(gto=temp$gto,g=temp$g) %>% relocate(g:gto,.after=Nivel)
          %>%
348   xtable(caption = "Promedio del crecimiento económico y del gasto
          de los países",
349          label="gnivel_res",digits=1) %>% print(include.rownames=F)
350
351
352
```

```
# III. Autocorrelación
# %AR: Autocorrelación de los residuales *
temp<-pbgtest(g_mco)
pbgtest(g_int)
pbgtest(g_mcg)
#pbltest Baltagi and Li Serial Dependence Test For Random Effects Models
temp<-pbltest(g_mcg)

#pdwtest Durbin-Watson Test for Panel Models
pdwtest(g_mco)
pdwtest(g_int)
pdwtest(g_mcg)

tab<-rbind(extp(pbgtest(g_mco),1),
      extp(pbgtest(g_int),1),
      extp(pbgtest(g_mcg),1),
      extp(pbltest(g_mcg),1)) %>%
  bind_rows(rbind(
      extp(pdwtest(g_mco),0),
      extp(pdwtest(g_int),0),
      extp(pdwtest(g_mcg),0))) %>%
  mutate(Tipo=c(rep("Breusch-Godfrey",3),"Baltagi-Li",rep("Durbin-Watson",3))) %>%
  mutate(`p-value`=format(`p-value`,format="e",digits = 2)) %>%
  mutate(df1=ifelse(is.na(df1),"",format(df1,digits = 0))) %>%
  mutate(Modelo=c("MCO","W","MCG","MCG","MCO","W","MCG")) %>%
  dplyr::select(Tipo,everything(),-`Hipótesis alt.`)

tab
print(xtable(tab,digits = 2),include.rownames=FALSE)
# library(MASS)
consolidado %>% filter(dplyr::between(cre,-10,10)) %>%
  group_by(code) %>% mutate(lg=dplyr::lag(cre)) %>%
  filter(!is.na(lg),!is.na(cre)) %>%
  ggplot(aes(x=lg,y=cre))+
  geom_hline(yintercept = 0,col="darkgray")+geom_vline(xintercept = 0,col="darkgray")+
  geom_point(alpha=0.5,size=0.5)+
  geom_density_2d(size=1,col=3)+
  theme_bw()+scale_color_locuszoom()+
  xlab(expression(g[t-1]))+ylab(expression(g[t]))
ggsave(filename = "./TeX/Fig/441_01_glag.pdf",width = 10,height = 8,units = "cm")

cuads<-read.csv("./Rcode/cuadrantes.csv")
consolidado %>% group_by(code) %>% mutate(lg=dplyr::lag(cre)) %>%
  filter(!is.na(lg),!is.na(cre)) %>% ungroup() %>% select(code:cont,lg,cre) %>%
  mutate(xsign=sign(lg),ysign=sign(cre),id=paste(xsign,ysign,sep = "")) %>%
  merge(cuads %>% select(id,cuadrante)) %>% select(-c(id)) %>%
  group_by(cuadrante) %>% summarise(n=n_distinct(code),m=n())

temp<-consolidado %>% filter(!is.na(cre))
auto<-by(temp$cre, temp$code, function(i) { acf(i,  plot = FALSE)$acf })
```

```
temp<-matrix(NA,nrow = 209,ncol = 19)
for (i in 1:209) {
  temp[i,1]<-names(auto)[i]
  m<-length(auto[[i]])
  temp[i,2:(m+1)]<-auto[[i]]
}
colnames(temp)<-c("code",0:17)
auto<-temp %>% as.data.frame() %>%
  pivot_longer(cols = `0`:`17`,names_to="lag",values_to="cor") %>%
  na.omit() %>% mutate(cor=as.numeric(cor),lag=as.numeric(lag))

auto %>% filter(lag <=15) %>%
  ggplot(aes(x=lag,y=cor,group=lag)) +
  geom_boxplot()+theme_bw()+
  ylab("")

ggsave("./TeX/Fig/441_02_acfbox.pdf",width = 10,height = 8,units =
    "cm")



# g system GMM ####

fsignif=function(x,lvl=c(0.1,0.05,0.01)){
  y<-as.numeric(x)
  codigos = (y<lvl[1]) + (y<lvl[2]) + (y<lvl[3])
  codigos[codigos==3]<-"$^{***}$"
  codigos[codigos==2]<-"$^{**}$"
  codigos[codigos==1]<-"$^{*}$"
  res=paste(x,replace_na(codigos,""),sep="")
  return(res)}

eqsgmm<-cre ~ lag(cre) + gto2+
  soc+edu+eco+sal+resto |
  lag(cre, 2:99) | # system GMM
  lag(gto2,2)+lag(soc, 2)+lag(edu,2)+lag(eco,2)+lag(sal,2)+lag(
      resto,2)

eqsgmm2<-cre ~ lag(cre) + gto2+
  soc+edu+eco+sal+resto+d08+d09 |
  lag(cre, 2:99) | # system GMM
  lag(gto2,2)+lag(soc, 2)+lag(edu,2)+lag(eco,2)+lag(sal,2)+lag(
      resto,2)+lag(d08,2)+lag(d09,2)


base <- pgmm(eqsgmm,g_dat,index=c("code", "year"),model="twosteps",
    #subset = Nivel!="Bajo",
              effect="twoways",transformation = "ld",collapse = TRUE
                  )

base.ind<-pgmm(eqsgmm,g_dat,index=c("code", "year"),model="twosteps
    ",#subset = Nivel!="Bajo",
               effect="individual",transformation = "ld",collapse =
                   TRUE)

baseic<-pgmm(eqsgmm2,g_dat,index=c("code", "year"),model="twosteps"
    ,#subset = Nivel!="Bajo",
```

```
452                 effect="individual",transformation = "ld",collapse =
                        TRUE)
453
454 sbase<-summary(base);sbasei<-summary(base.ind);sbaseic<-summary(
        baseic)
455
456 lmodel<-list(sbasei,sbaseic,sbase)
457 vsarg<-format(sapply(lmodel, function(x) x$sargan$p.value),digits =
        1)
458 ar1<-format(sapply(lmodel, function(x) x$m1$p.value),digits = 2)
459 ar2<-format(sapply(lmodel, function(x) x$m2$p.value),digits = 2)
460 waldmu<-c("","",format(sbase$wald.td$p.value[[1]],digits = 2))
461
462 stargazer(base.ind,baseic,base,#type = "text",
463           digits = 2,
464           title = "sys-GMM para el crecimiento económico",
465           label = "tab:gsGMM",
466           column.labels = c("Sin $\\mu$", "Sin $\\mu$ + Ind. crisis
                ","Con $\\mu$"),
467           single.row = T,
468           dep.var.caption = "Variable dependiente: $g_t$",
469           #dep.var.labels = "$g_t$",
470           dep.var.labels.include = F,
471           # covariate.labels=c("$\\rho$","gto","$\\log$(soc/gob)","
                $\\log$(edu/gob)",
472           #                   "$\\log$(eco/gob)","$\\log$(sal/gob)
                ",
473           #                   "$\\log$(resto/gob)","Ind. 2008","
                Ind. 2009"),
474           keep.stat = "f",
475           add.lines = list(#c("Observaciones",rep("1,296",3)),
476                            c("Países",rep(65,3)),
477                            c("Instrumentos",41,43,69),
478                            c("Sargan p-value",vsarg %>% fsignif()),
479                            c("AR(1) p-value",ar1 %>% fsignif()),
480                            c("AR(2) p-value",ar2 %>% fsignif()),
481                            c("Wald $\\mu_t$ p-value",waldmu %>%
                                fsignif())
482                            ))
483
484
485 # Covariance
486 sqrt(diag(vcov(base)))[1:7]
487 sqrt(diag(vcovHC(base)))[1:7]
488
489 #sargan
490 sargan(base)
491 sargan(base.ind)
492 sargan(difer)
493 sargan(difer.ind)
494
495 #serial correlation
496 mtest(base,vcov = vcovHC)
497
498 # Gráficas auxiliares ####
499 # efectos temporales
500 # efectos temporales
```

```
501
502 tiempo<-base$coefficients[[2]][9:36]
503 tiempo<-data.frame(year=names(tiempo),mu=tiempo)
504 base$vcov[9:36,9:36]
505 #plot(as.numeric(names(tiempo)),tiempo,type="b")
506 tiempo %>% ggplot(aes(x=as.numeric(year),y=mu))+geom_line(size=1,
      col=1)+
507   #geom_smooth(formula = 'y ~ poly(x,3)',method = 'glm' )+
508   theme_bw()+
509   xlab('Año')+ylab('')+scale_x_continuous(n.breaks = 6)
510 ggsave("./TeX/Fig/441_1_temporal.pdf",width = 12,height = 4.5,units
      = "cm")
511
512 coeftest(base)
513
514 # con s.e. de mu
515 tiempo %>% ggplot(aes(x=as.numeric(year),y=mu))+geom_line(size=1,
      col=1)+
516   geom_ribbon(aes(ymin=mu-sqrt(diag(base$vcov))[9:36],ymax=mu+sqrt(
        diag(base$vcov))[9:36]),alpha=0.5)
517 mean(sqrt(diag(base$vcov))[9:36])
518 mean(tiempo$mu)
519 # con boxplot de crecimiento
520 g_dat %>% ggplot(aes(x=as.numeric(as.character(year)),y=cre,group=
      year)) + geom_boxplot()+
521   geom_line(data=tiempo,aes(x=as.numeric(year),y=mu,group=1),size
        =1,col=4)+
522   theme_bw()+xlab('Año')+ylab('')+scale_x_continuous(n.breaks = 6)+
523   coord_cartesian(ylim=c(-10,10),xlim = c(1991,2019))
524 ggsave("./TeX/Fig/441_1_temporal2.pdf",width = 12,height = 4.5,
      units = "cm")
525
526 # Crisis por nivel
527 consolidado %>% filter(!is.na(cre)) %>%
528   mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
        %>% nivtrans() %>%#México en categoría separada
529   group_by(Nivel) %>% filter(year %in% c
        (1992,1993,1995,2001,2008,2009)) %>%
530   dplyr::select(code,year,Nivel,cre) %>%
531   pivot_wider(names_from = year,values_from=cre) %>%
532   summarise(across('1992':'2009',mean,na.rm=T))
533 # Crisis por grupos ####
534 # Crisis 2008 por Nivel
535 is_outlier <- function(x) {
536   return(x < quantile(x, 0.25) - 1.5 * IQR(x) | x > quantile(x,
        0.75) + 1.5 * IQR(x))
537 }
538
539 consolidado %>% filter(year==2009,!is.na(cre)) %>%
540   mutate(Nivel=ifelse(code=="TWN","Alto",as.character(Nivel))) %>%
        #completar Taiwan
541   #mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
        %>% nivtrans() %>%#México en categoría separada
542   nivtrans() %>% group_by(Nivel) %>%
543   #mutate(outlier = ifelse(is_outlier(cre),pais,"")) %>%
544   ggplot(aes(x=Nivel,y=cre)) + geom_boxplot()+
545   theme_bw()+xlab("")+ylab("")+
```

```
546   geom_point(data = subset(consolidado,code=="MEX"&year==2009),aes
          (3,cre),col="blue")+
547   geom_text(data = subset(consolidado,code=="MEX"&year==2009),aes
          (2.5,cre,label="México"))
548 ggsave("./TeX/Fig/441_12_crisis09.pdf",width = 8,height = 8,units =
      "cm")
549
550 consolidado %>% filter(year==2008,!is.na(cre)) %>%
551   mutate(Nivel=ifelse(code=="TWN","Alto",as.character(Nivel))) %>%
          #completar Taiwan
552   mutate(cont=countrycode(code,"genc3c","continent")) %>%#
          completar continente
553   mutate(region=countrycode(code,"genc3c","region")) %>%
554   #dplyr::select(code,year,cont,region,cre) %>% arrange(cre)
555   #mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
          %>% nivtrans() %>%#México en categoría separada
556   nivtrans() %>%#group_by(cont) %>%
557   #mutate(outlier = ifelse(is_outlier(cre),pais,"")) %>%
558   ggplot(aes(x=Nivel,y=cre)) + geom_boxplot()+
559   theme_bw()+xlab("")+ylab("")+#coord_cartesian(ylim = c(-10,20))
560   geom_point(data = subset(consolidado,code=="MEX"&year==2008),aes
          (3,cre),col="blue")+
561   geom_text(data = subset(consolidado,code=="MEX"&year==2008),aes
          (2.5,cre,label="México"))
562 ggsave("./TeX/Fig/441_12_crisis08.pdf",width = 8,height = 8,units =
      "cm")
563
564 # Crecimiento por Nivel
565 consolidado %>% filter(!is.na(cre)) %>%
566   mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
          %>% nivtrans() %>%#México en categoría separada
567   group_by(Nivel,year) %>% summarise(g=mean(cre,na.rm = T)) %>%
568   ggplot(aes(x=year,y=g,col=Nivel))+geom_line()+scale_color_
          locuszoom()
569
570 # fitted values
571 ajust<-baseic$fitted.values %>% as.data.frame() %>%
572   mutate(year=names(base$residuals[[1]]),
573          model=c(rep('d',28),rep('l',29))) %>%
574   pivot_longer(!c(year,model),names_to='code',values_to='fitted')
          %>%
575   mutate(fitted=na_if(fitted,0))
576 ajust %>% ggplot(aes(x=as.numeric(year),y=fitted))+geom_point()
577 # errores ####
578
579 lsfun<-list("media"=mean,#"dev"=sd,
580             "plo"=function(x,na.rm=T) quantile(x,probs = 0.25,na.rm
                  ),
581             "pup"=function(x,na.rm=T) quantile(x,probs = 0.75,na.rm
                  ))
582
583 todo<-data.frame(year=NA,resid_media=NA,resid_plo=NA,resid_pup=NA,
      model=NA)
584 modelos<-list("(3) Temp"=base,"(1) Indiv"=base.ind,"(2) Indiv+
      crisis"=baseic)
585
586 for (i in 1:length(modelos)) {
```

```
587   errores<-modelos[[i]]$residuals %>% as.data.frame() %>%
588     mutate(year=names(base$residuals[[1]]),
589            model=c(rep('d',28),rep('l',29))) %>%
590     pivot_longer(!c(year,model),names_to='code',values_to='resid')
          %>%
591     mutate(resid=na_if(resid,0)) %>% na.omit()
592   temp<-errores %>% group_by(year) %>% summarise(across(resid,lsfun
        )) %>% mutate(modelo=names(modelos)[i])
593   if (i==1) todo=temp else todo=rbind(todo,temp)
594 }
595
596 todo %>% ggplot(aes(x=as.numeric(year),y=resid_media))+
597   geom_ribbon(aes(ymin=resid_plo,ymax=resid_pup,fill=modelo),alpha
        =0.2)+
598   geom_line(aes(col=modelo),size=1)+theme_bw()+
599   xlab("Año")+ylab("")+labs(col="",fill="")+
600   theme(legend.position = c(.45,.95),
601         legend.background = element_rect(fill = "transparent"))+
602   guides(fill=guide_legend(nrow=1))+scale_x_continuous(n.breaks =
        6)
603 ggsave("./TeX/Fig/441_2_errores.pdf",width = 12,height = 4.5,units
      = "cm",bg="transparent")
604
605
606
607
608 temp %>% ggplot(aes(x=as.numeric(year),y=resid_media))+
609   geom_ribbon(aes(ymin=resid_plo,ymax=resid_pup),fill="grey90")+
        geom_line(size=1)
610
611 errores %>% ggplot(aes(x=as.numeric(year),y=resid))+geom_point()
612
613 temp <- merge(ajust,errores) %>% na.omit() %>%
614   merge(g_dat %>% select(code,year,outlier,Grupo,cre,Nivel),all.x =
        T)
615 temp %>%
616   filter(!outlier) %>%
617   ggplot(aes(x=fitted,y=resid))+geom_point(alpha=0.5)+
618   geom_density2d()
619
620 temp%>%
621   ggplot(aes(x=fitted,y=cre))+
622   geom_point(aes(col=Nivel))+#geom_smooth()+
623   coord_fixed()+scale_color_locuszoom()+geom_density2d()
624
625 temp %>% filter(!outlier) %>%
626   select(fitted,resid,Nivel) %>%
627   mvn(mvnTest = 'dh', subset = "Nivel",desc = F)
628
629 # Coeficientes ####
630 gm<-function(c="MEX",y=2018){
631   res<-funciones %>% mutate(lg=dplyr::lag(cre)) %>%
632   filter(code==c,year==y) %>%
633   dplyr::select(def:soc,gto2,cre,lg) %>%
634   mutate(resto=def+amb+viv+seg+cul) %>%
635   dplyr::select(-c(def,amb,viv,seg,cul)) %>%
636   relocate(lg,gto2,soc,edu,eco,sal,resto)
```

```
637  return(res)
638 }
639 gm()
640
641 logob<-function(x) x %>% mutate(across(soc:resto,~log(./gob))) %>%
642   dplyr::select(-gob)
643 gm("MEX",2018)
644 logob(gm("MEX",2018))
645
646 base.ind$coefficients[[2]]
647 base.ind$fitted.values[c("2018","2019"),"MEX"]
648
649 temp<-base.ind$fitted.values[29:57,"MEX"]
650
651
652 sum(logob(gm("MEX",1992))[1:7]*base.ind$coefficients[[2]])
653 temp<-funciones %>% mutate(lg=dplyr::lag(cre)) %>%
654   filter(code=="MEX",dplyr::between(year,1991,2019)) %>%
655   unite("id",code:year,sep=".") %>%
656   dplyr::select(id,def:soc,gto2,cre,lg) %>%
657   mutate(resto=def+amb+viv+seg+cul) %>%
658   dplyr::select(-c(def,amb,viv,seg,cul)) %>%
659   column_to_rownames("id") %>%
660   relocate(lg,gto2,soc,edu,eco,sal,resto)
661
662 temp$mfitted<-base.ind$fitted.values[29:57,"MEX"]
663 temp$calres<-temp$cre-temp$calculos
664 temp$mresid<-base.ind$residuals[["MEX"]][29:57]
665
666 # cambios en modelo individual
667 (logob(gm("MEX",2019)+c(0,0,0,0.5,0,-0.5,0,0,0))[1:7]*base.ind$
      coefficients[[2]] ) %>% sum()-
668 (logob(gm("MEX",2019)+c(0,0,0,0,0,0,0,0,0))[1:7]*base.ind$
      coefficients[[2]] ) %>% sum()
669
670 base.ind$coefficients[[2]]["edu"]/3.053979*0.5-
671 base.ind$coefficients[[2]]["sal"]/2.37477*0.5
672
673 # cambios en modelo con indicadoras
674 (logob(gm("MEX",2019)+c(0,0,0,0.5,0,-0.5,0,0,0))[1:7]*baseic$
      coefficients[[2]] ) %>% sum()-
675   (logob(gm("MEX",2019)+c(0,0,0,0,0,0,0,0,0))[1:7]*baseic$
        coefficients[[2]] ) %>% sum()
676
677 baseic$coefficients[[2]]["edu"]/3.053979*0.5-
678   baseic$coefficients[[2]]["sal"]/2.37477*0.5
679
680 # Cuadro resumen sustituciones ####
681 fnames<-c("edu","sal","resto")
682 gmex19<-gm("MEX",2019)[fnames]
683 base.ind$coefficients[[2]]
684
685 inicial<-gm("MEX",2019)
686 rebal<-inicial
687 rebal[fnames[1]]<-gm("MEX",2019)[fnames[1]]+cambios[1]
688 rebal[fnames[2]]<-gm("MEX",2019)[fnames[2]]-cambios[1]
689 inicial
```

```
690 rebal
691
692 sum(logob(inicial)[1:7]*base.ind$coefficients[[2]])-sum(logob(rebal
        )[1:7]*base.ind$coefficients[[2]])
693
694 cambios<-c(0.1,0.5,1)
695 reasignaciones<-function(modelo){
696 temp<-lapply(cambios, function(x){
697   res=matrix(0,3,3)
698   for(j in 1:3){
699     for(k in 1:3){
700       inicial<-gm("MEX",2019)
701       rebal<-inicial
702       rebal[fnames[j]]<-rebal[fnames[j]]+x
703       rebal[fnames[k]]<-rebal[fnames[k]]-x
704       res[j,k]<-sum(logob(rebal)[1:7]*modelo$coefficients[[2]])-sum
              (logob(inicial)[1:7]*modelo$coefficients[[2]])
705     }
706   }
707   return(res)
708 })
709 matriz<-rbind(temp[[1]],temp[[2]],temp[[3]])
710 return(matriz)
711 }
712
713 tab<-cbind(reasignaciones(base.ind),reasignaciones(baseic)) %>% as.
        data.frame()
714 colnames(tab)<-paste(rep(c("Ind","Ind+Cri"),each=3),rep(fnames,2))
715 tab$fun<-rep(fnames,3)
716 tab$delta<-rep(cambios,each=3)
717
718 tab<-tab %>% relocate(delta,fun,`Ind edu`)
719
720 tab %>% xtable(caption = "Efectos de la reasignación del gasto",
        label="tab:gsGMM_delta") %>%
721   print(include.rownames=F)
722
723 # correlacion con dependiente
724 with(temp %>% filter(model=='l'),cor(cre,fitted)^2)
725 with(temp %>% filter(model=='d'),cor(cre,fitted)^2)
726
727
728 #Datos con IDH ####
729 funciones <- consolidado %>% filter(code!="TLS") %>%# quitar pais
        con pocas observaciones
730   mutate(Nivel=ifelse(code=="TWN","Alto",as.character(Nivel))) %>%
        #completar Taiwan
731   mutate(cont=countrycode(code,"genc3c","continent")) %>%#
        completar continente
732   mutate(cont=ifelse(code=="XKS","Europe",cont)) %>%#completar
        continete para Kosovo
733   #mutate(Nivel=ifelse(code=="MEX","México",Nivel)) %>%#México en
        categoría separada
734   mutate(across(c(def:soc),~ifelse(.==0,min(minval$min),.))) %>%#
        imputación valores pequeños
735   merge(idhs %>% dplyr::select(code,year,HDI),all.x = TRUE) %>%
```

```
736   dplyr::select(code:soc,gto2,HDI,pibpc) %>% filter(across(code:
          gto2,~!is.na(.))) %>% nivtrans()
737 head(funciones)
738 levels(funciones$Nivel)
739
740 #h_t ####
741 h_dat<-funciones %>% dplyr::select(code,year,Nivel,pibpc,HDI,gto2,
        def:soc) %>%
742   mutate(resto=seg+def+amb+cul+viv) %>% dplyr::select(-c(seg,def,
          amb,cul,viv)) %>%#amalgama
743   mutate(across(eco:resto,~log(./gob))) %>% dplyr::select(-gob) %>%
           # logcocientes
744   merge(fgrupo %>% dplyr::select(code,year,Grupo:peso)) %>%#
          variables de grupos
745   group_by(code) %>%
746   mutate(gtosq=gto2^2,h=100*(HDI/dplyr::lag(HDI)-1)) %>% filter(
          year>=1990) %>%# gasto^2, h y filtro año
747   filter(!is.na(h)) %>%# sólo datos disponibles
748   mutate(d92=ifelse(year==1992,1,0),d93=ifelse(year==1993,1,0),d94=
          ifelse(year==1994,1,0),
749          d08=ifelse(year==2008,1,0),d09=ifelse(year==2009,1,0)) %>%
                #indicadoras crisis
750   pdata.frame()
751
752
753 class(h_dat$h)
754 h_dat %>% summarise(n=n_distinct(code))
755
756 h_dat %>% filter(h==0)
757 # codeobsres(h_dat)
758
759 # OLS, Within, GLS ####
760 h_mco<-plm(h~gto2+soc+edu+eco+sal+resto,h_dat,model = "pooling") #
        ols
761 h_int<-plm(h~gto2+soc+edu+eco+sal+resto,h_dat,model = "within",
        effect = "individual") # within
762 h_mcg<-plm(h~gto2+soc+edu+eco+sal+resto,h_dat,model = "random",
        effect = "individual") # gls
763
764
765 #presencia de efectos
766 pres<-c(format(pFtest(update(h_int,effect="twoways"),h_mco)$p.value
        ,digits = 2),"","")
767 #Hausmant test, correlación de efectos
768 hcor<-c("","",format(phtest(h_int,h_mcg)$p.value,digits = 2))
769 #AR(1)
770 ar1<-sapply(list(h_mco,h_int,h_mcg),function(x) format(pbgtest(x)$p
        .value,digits = 2))
771 #AR(2)
772 ar2<-sapply(list(h_mco,h_int,h_mcg),function(x) format(pbgtest(x,
        order=2)$p.value,digits = 2))
773
774 stargazer(h_mco,h_int,h_mcg,digits = 2, #type="text",
775           title = "Regresión de panel para desarrollo",
776           label = "tab:hmod0",
777           column.labels = c("$\\hat{\\und\\gamma}_{\\text{MCO}}$",
778                             "$\\hat{\\und\\gamma}_{\\text{W}}$",
```

```
                                    "$\\hat{\\und\\gamma}_{\\text{MCG}}$"),
          single.row = T,
          dep.var.caption = "Variable dependiente: $h_t$",
          omit.stat = "F",
          dep.var.labels.include = F,
          covariate.labels=c("gto","$\\log$(soc/gob)","$\\log$(edu/
              gob)",
                             "$\\log$(eco/gob)","$\\log$(sal/gob)",
                             "$\\log$(resto/gob)","ordenada"),
          add.lines = list(c("F (efectos) p-val",pres %>% fsignif()
              ),
                           c("Hausman p-value",hcor %>% fsignif()),
                           c("AR(1) p-value",ar1 %>% fsignif()),
                           c("AR(2) p-value",ar2 %>% fsignif()))
          )

summary(h_mco)
coeftest(h_mco)
coeftest(h_int)
coeftest(h_mcg)

coeftest(h_int2)

temp<-ranef(h_mcg,effect = "time")
plot(names(temp),temp,type="b")


# h plots ####
# Por nivel de ingreso
temp<-consolidado %>% merge(idhs %>% dplyr::select(code,year,HDI),
    all.x = TRUE) %>%
  filter(year>=1990,!is.na(HDI)) %>% dplyr::select(code,pais,year,
      Nivel,cont,HDI) %>%
  mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel))
      %>% nivtrans()
temp %>% group_by(Nivel) %>% summarise(n=n_distinct(code)) %>%
    adorn_totals()

temp %>% mutate(cont=ifelse(is.na(cont),countrycode(code,"genc3c","
    continent"),cont)) %>%
  group_by(cont,Nivel) %>% summarise(n=n_distinct(code)) %>% adorn_
      totals()
a<-consolidado %>% merge(idhs %>% dplyr::select(code,year,HDI),all.
    x = TRUE) %>%
  filter(year>=1990,!is.na(HDI)) %>% dplyr::select(code,year,Nivel,
      HDI) %>%
  mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
      %>% nivtrans() %>%#México en categoría separada
  group_by(code) %>% mutate(h=100*(HDI/dplyr::lag(HDI)-1)) %>%
      ungroup %>%#cálculo tasa h
  group_by(Nivel) %>%
  summarise(g=mean(HDI,na.rm = T),h=mean(h,na.rm = T),N=n_distinct(
      code),n=n()/N) %>%
  ggplot(aes(x=reorder(Nivel,h),y=h,fill=Nivel)) +
  geom_bar(stat = "identity")+ #coord_flip()+
  geom_text(aes(label=round(h,2)),nudge_y = .1,angle=90,size=3)+
  theme_bw()+xlab("")+ylab("")+
```

```
823   coord_cartesian(ylim = c(0,1.5))+
824   scale_fill_locuszoom()+
825   theme(legend.position = "none",axis.text.x = element_text(angle =
          90),
826         axis.text.y = element_blank(),axis.ticks=element_blank())
827
828 # Por continente
829 temp<-consolidado %>% merge(idhs %>% dplyr::select(code,year,HDI),
      all.x = TRUE) %>%
830   filter(year>=1990,!is.na(HDI)) %>%
831   mutate(cont=ifelse(is.na(cont),countrycode(code,"genc3c","
          continent"),cont))
832 temp %>% group_by(cont) %>% summarise(n=n_distinct(code)) %>% adorn
      _totals()
833
834 b<-consolidado %>% merge(idhs %>% dplyr::select(code,year,HDI),all.
      x = TRUE) %>%
835   filter(year>=1990,!is.na(HDI)) %>%
836   mutate(cont=ifelse(is.na(cont),countrycode(code,"genc3c","
          continent"),cont)) %>%
837   dplyr::select(code,year,cont,HDI) %>%
838   group_by(code) %>% mutate(h=100*(HDI/dplyr::lag(HDI)-1)) %>%
          ungroup %>%#cálculo tasa h
839   group_by(cont) %>%
840   summarise(HDI=mean(HDI,na.rm = T),N=n_distinct(code),n=n()/N,h=
          mean(h,na.rm = T)) %>%
841   ggplot(aes(x=reorder(cont,h),y=h, fill=cont)) +
842   geom_bar(stat = "identity")+# coord_flip()+
843   geom_text(aes(label=round(h,2)),nudge_y = .1,angle=90,size=3)+
844   theme_bw()+xlab("")+ylab("")+
845   coord_cartesian(ylim = c(0,1.5))+
846   scale_fill_aaas()+
847   theme(legend.position = "none",axis.text.x = element_text(angle =
          90),
848         axis.text.y = element_blank(),axis.ticks=element_blank())
849
850
851 # Por grupo gasto público
852 h_dat %>% summarise(n=n_distinct(code))
853 c<-h_dat %>% filter(as.numeric(as.character(year))>=1990,!is.na(HDI
      )) %>% dplyr::select(code,year,Grupo,HDI,h) %>%
854   #mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
          %>% nivtrans() %>%#México en categoría separada
855   group_by(Grupo) %>%
856   summarise(HDI=mean(HDI,na.rm = T),N=n_distinct(code),n=n()/N,h=
          mean(h,na.rm = T)) %>%
857   ggplot(aes(x=reorder(paste("Grupo",as.numeric(Grupo)),h),y=h, fill
          =Grupo)) +
858   geom_bar(stat = "identity")+ #coord_flip()+
859   geom_text(aes(label=round(h,2)),nudge_y = .1,angle=90,size=3)+
860   theme_bw()+xlab("")+ylab("")+
861   coord_cartesian(ylim = c(0,1.5))+
862   scale_fill_npg()+
863   theme(legend.position = "none",axis.text.x = element_text(angle =
          90),
864         axis.text.y = element_blank(),axis.ticks=element_blank())
865
```

```r

library(cowplot)
allplotslist <- align_plots(a,b,c, align = "hv")
library(ggpubr)

ggarrange(allplotslist[[1]], allplotslist[[2]], allplotslist[[3]],
    ncol=3)
ggsave("./TeX/Fig/442_0_hgrupos.pdf",width = 14,height = 8,units =
    "cm")

g_dat %>% group_by(Grupo) %>% summarise(across(HDI:resto,mean)) #
    promedio var por grupo

# variables promedio por grupo
funciones %>% dplyr::select(code,year,Nivel,HDI,gto2,def:soc) %>%
  mutate(resto=seg+def+amb+cul+viv) %>% dplyr::select(-c(seg,def,
      amb,cul,viv)) %>%#amalgama
  group_by(code) %>% mutate(h=100*(HDI/dplyr::lag(HDI)-1)) %>%
      ungroup %>%#cálculo tasa h
  #mutate(across(eco:resto,~log(./gob))) %>% dplyr::select(-gob)
      %>%# logcocientes
  merge(fgrupo %>% dplyr::select(code,year,Grupo:peso)) %>%#
      variables de grupos
  mutate(gtosq=gto2^2) %>% filter(year>=1990) %>%# gasto^2 y
      filtro año
  filter(!is.na(HDI)) %>%# sólo datos disponibles
  pdata.frame() %>%
  group_by(Grupo) %>% summarise(across(HDI:h,mean,na.rm=T)) %>%
  xtable(caption = "Gasto y desarrollo de los países por grupos",
         label="hgasto_res",digits=1) %>% print(include.rownames=F)

# Pruebas ####
# I. presencia de efectos


# F test for ind / time effects *
for (i in c("individual","time","twoways")){
  temp<-extp(pFtest(update(h_int,effect=i),h_mco))
  temp$efecto<-i
  if (i=="individual") res<-temp else res<-rbind(res,temp)
}
pruf<-res

# plmtest Lagrange FF Multiplier Tests for Panel Models *
for (i in c("individual","time","twoways")){
  temp<-extp(plmtest(h_mco,effect = i),param = F)
  temp$efecto<-i
  if (i=="individual") res<-temp else res<-rbind(res,temp)
}
prulm<-res


tab<-bind_rows(pruf,prulm) %>%
  mutate(across(df1:df2,~ifelse(is.na(.),"",formatC(.,format = "f",
      big.mark = ",",digits = 0)))) %>%
  mutate(`p-value`=format(`p-value`,format="e",digits = 2)) %>%
```

```
913   mutate(Tipo=c("\\multirow{3}{*}{F}","","","\\multirow{3}{*}{Honda
          }","","")) %>%
914   dplyr::select(Tipo,everything()) %>%
915   mutate(efecto=rep(c("individual","temporal","ambos"),2))
916 tab
917 print(xtable(tab %>% dplyr::select(-`Hipótesis alt.`)), include.
      rownames=FALSE)
918
919 # II. Correlación de efectos
920 # %Hausman: cor(X,eta) *
921 phtest(h_int,h_mcg)
922 # H0:RE (mcg) vs H1:FE (int)/ H0 not rejected / testing if cor(e,X)
      <>0 H0: cor(e,X)=0
923 fnames<-c("soc","edu","eco","sal","resto","gto2","h")
924 head(h_dat)
925 head(h_dat[,fnames])
926 tab<-sapply(names(h_dat[,fnames]), function (x) cor(fixef(h_int),
      between(h_dat[,x]))) %>% t()
927 tab
928 # temp<-update(h_int,effect="time")
929 # sapply(names(h_dat[,c(3:7,13,12)]), function (x) cor(fixef(temp,"
      time"), between(h_dat[,x],"time")))
930 rownames(tab)<-"correlación"
931 print(xtable(as.data.frame((tab)),
932               caption = "Correlación entre $\\und \\eta_i$ y las
                   variables del modelo",
933               label="tab:h0coreta"),include.rownames = T)
934
935 consolidado %>% merge(idhs %>% dplyr::select(code,year,HDI),all =
      TRUE) %>%
936   filter(year>=1990,!is.na(HDI)) %>%
937   mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel))
          %>% nivtrans() %>%
938   summarise(n=n_distinct(code))
939
940 # Resumen de gasto total y h por nivel con todos los datos
941 temp<-consolidado %>%
942   merge(idhs %>% dplyr::select(code,year,HDI),all = TRUE) %>%
943   group_by(code) %>% mutate(h=100*(HDI/dplyr::lag(HDI)-1)) %>%
          ungroup %>%#cálculo tasa h
944   filter(year>=1990) %>% dplyr::select(code:cont,gto2,HDI,h) %>%
945   filter(code!="TWN",code!="XWB") %>%#filtrar Taiwán y Palestina
          por falta de datos
946   mutate(pais=countrycode(code,"genc3c","country.name")) %>%#
          completar nombre de pais
947   mutate(Nivel=ifelse(code=="SRB","Medio-alto",as.character(Nivel))
          ) %>%
948   mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
          %>% nivtrans() %>%#México en categoría separada
949   group_by(Nivel) %>%
950   summarise(gto=mean(gto2,na.rm = T),
951             HDI=mean(HDI,na.rm = T),h=mean(h,na.rm = T),
952             n=n_distinct(code))
953 temp %>% adorn_totals()
954
955 # Tabla resumen por Nivel
956 funciones %>% dplyr::select(code,year,Nivel,HDI,gto2,def:soc) %>%
```

```
957   mutate(resto=seg+def+amb+cul+viv) %>% dplyr::select(-c(seg,def,
          amb,cul,viv)) %>%#amalgama
958   mutate(gtosq=gto2^2) %>% filter(year>=1990) %>%# gasto^2 y
          filtro año
959   mutate(Nivel=ifelse(code=="MEX","México",as.character(Nivel)))
          %>% nivtrans() %>%#México en categoría separada
960   pdata.frame() %>%
961   group_by(Nivel) %>% summarise(across(eco:resto,mean,na.rm=T)) %>%
962   mutate(gto=temp$gto,h=temp$h,IDH=format(temp$HDI,digits=2)) %>%
          relocate(IDH:gto,.after=Nivel) %>%
963   xtable(caption = "Promedio del crecimiento en IDH y del gasto de
          los países",
964          label="tab:hnivel_res",digits=1) %>% print(include.
                 rownames=F)
965
966 # III. Autocorrelación
967 # %AR: Autocorrelación de los residuales *
968 pbgtest(h_mco)
969 pbgtest(h_int)
970 pbgtest(h_mcg)
971 #pbltest Baltagi and Li Serial Dependence Test For Random Effects
        Models
972 pbltest(h_mcg)
973
974 #pdwtest Durbin-Watson Test for Panel Models
975 pdwtest(h_mco)
976 pdwtest(h_int)
977 pdwtest(h_mcg)
978
979 tab<-rbind(extp(pbgtest(h_mco),1),
980            extp(pbgtest(h_int),1),
981            extp(pbgtest(h_mcg),1),
982            extp(pbltest(h_mcg),1)) %>%
983   bind_rows(rbind(
984     extp(pdwtest(h_mco),0),
985     extp(pdwtest(h_int),0),
986     extp(pdwtest(h_mcg),0))) %>%
987   mutate(Tipo=c(rep("Breusch-Godfrey",3),"Baltagi-Li",rep("Durbin-
          Watson",3))) %>%
988   mutate(`p-value`=format(`p-value`,format="e",digits = 2)) %>%
989   mutate(df1=ifelse(is.na(df1),"",format(df1,digits = 0))) %>%
990   mutate(Modelo=c("MCO","W","MCG","MCG","MCO","W","MCG")) %>%
991   dplyr::select(Tipo,everything(),-`Hipótesis alt.`)
992
993 tab
994 print(xtable(tab,digits = 2,
995              caption="Pruebas de autocorrelación para $\\und\\nu$",
996              label="tab:h0auto"),include.rownames=FALSE)
997
998
999
1000
1001 # h system GMM ####
1002 h_dat %>% ggplot(aes(log(HDI),h))+geom_point()
1003
1004 eqsgmm<-h ~ lag(HDI) + gto2+
1005   soc+edu+eco+sal+resto |
```

```
1006   lag(h, 2:99) | # exógenas en ef. ind. e idios.
1007   lag(gto2,2)+lag(soc, 2)+lag(edu,2)+lag(eco,2)+lag(sal,2)+lag(
           resto,2)
1008
1009 eqsgmm2<-h ~ lag(HDI) + gto2+
1010   soc+edu+eco+sal+resto+d08+d09 |
1011   lag(h, 2:99)| # exógenas en ef. ind. e idios.
1012   lag(gto2,2)+lag(soc, 2)+lag(edu,2)+lag(eco,2)+lag(sal,2)+lag(
           resto,2) +lag(d08,2)+lag(d09,2)
1013
1014 base <- pgmm(eqsgmm,h_dat,index=c("code", "year"),model="twosteps",
1015                effect="twoways",transformation = "ld",collapse = TRUE
                       )
1016 base.ind<-pgmm(eqsgmm,h_dat,index=c("code", "year"),model="twosteps
         ",
1017                  effect="individual",transformation = "ld",collapse =
                         TRUE)
1018 baseic<-pgmm(eqsgmm2,h_dat,index=c("code", "year"),model="twosteps"
         ,
1019                effect="individual",transformation = "ld",collapse =
                       TRUE)
1020
1021 sbase<-summary(base)
1022 sbasei<-summary(base.ind)
1023 sbaseic<-summary(baseic)
1024
1025
1026
1027 lmodel<-list(sbasei,sbaseic,sbase)
1028 vsarg<-format(sapply(lmodel, function(x) x$sargan$p.value),digits =
         1)
1029 ar1<-format(sapply(lmodel, function(x) x$m1$p.value),digits = 1)
1030 ar2<-format(sapply(lmodel, function(x) x$m2$p.value),digits = 1)
1031 waldmu<-c("","",format(sbase$wald.td$p.value[[1]],digits = 1))
1032
1033 stargazer(base.ind,baseic,base,# type="text",
1034           digits = 2,
1035           title = "sys-GMM para el desarrollo",
1036           label = "tab:hsGMM",
1037           column.labels = c("Sin $\\mu$", "Sin $\\mu$ + Ind. crisis
                 ","Con $\\mu$"),
1038           dep.var.caption = "Variable dependiente: $h_t$",
1039           single.row = T,
1040           dep.var.labels.include = F,
1041           # covariate.labels=c("$IDH_{t-1}$","gto","$\\log$(soc/gob
                 )","$\\log$(edu/gob)",
1042           #                    "$\\log$(eco/gob)","$\\log$(sal/gob)
                 ",
1043           #                    "$\\log$(resto/gob)","Ind. 2008","
                 Ind. 2009"),
1044           omit.stat = "n",
1045           add.lines = list(#c("Observaciones",rep("1,296",3)),
1046             c("Países",rep(64,3)),
1047             c("Instrumentos",41,43,69),
1048             c("Sargan p-value",vsarg %>% fsignif()),
1049             c("AR(1) p-value",ar1 %>% fsignif()),
1050             c("AR(2) p-value",ar2 %>% fsignif()),
```

```
1051              c("Wald $\\mu_t$ p-value",waldmu %>% fsignif())
1052            ))
1053
1054 # Coeficientes ####
1055 #Martins coefficients
1056 (0.1*0.0872-0.1^2*0.381)*100 #educación
1057
1058 # Cifras México
1059 3.03/25.67
1060 (3.03+2.57)/25.67
1061 (3.03+2.57)
1062
1063 hm<-function(c="MEX",y=2018){
1064   res<-funciones %>% mutate(h=100*(HDI/dplyr::lag(HDI)-1),lH=dplyr
         ::lag(HDI)) %>%
1065     filter(code==c,year==y) %>%
1066     dplyr::select(def:soc,gto2,h,lH) %>%
1067     mutate(resto=def+amb+viv+seg+cul) %>%
1068     dplyr::select(-c(def,amb,viv,seg,cul)) %>%
1069     relocate(lH,gto2,soc,edu,eco,sal,resto)
1070   return(res)
1071 }
1072 hm()
1073
1074 base.ind$fitted.values[26+27,"MEX"]
1075
1076 # cambios en modelo individual
1077 (logob(hm("MEX",2018)+c(0,0,0,2,0,-2,0,0,0))[1:7]*base.ind$
       coefficients[[2]]) %>% sum()-
1078   (logob(hm("MEX",2018)+c(0,0,0,0,0,0,0,0,0))[1:7]*base.ind$
         coefficients[[2]]) %>% sum()
1079
1080 base.ind$coefficients[[2]]["edu"]/3.026974*2-
1081   base.ind$coefficients[[2]]["sal"]/2.399094*2
1082
1083 # cambios en modelo con indicadoras
1084 (logob(gm("MEX",2019)+c(0,0,0,0.5,0,-0.5,0,0,0))[1:7]*baseic$
       coefficients[[2]]) %>% sum()-
1085   (logob(gm("MEX",2019)+c(0,0,0,0,0,0,0,0,0))[1:7]*baseic$
         coefficients[[2]]) %>% sum()
1086
1087
1088 # Gráficas auxiliares ####
1089 # efectos temporales
1090 tiempo<-base$coefficients[[2]][9:34]
1091 plot(as.numeric(names(tiempo)),tiempo,type="b")
1092 tiempo<-data.frame(year=names(tiempo),mu=tiempo)
1093
1094 tiempo %>% ggplot(aes(x=as.numeric(year),y=mu))+geom_point()+
1095   geom_smooth(formula = 'y ~ poly(x,3)',method = 'glm')+
1096   theme_bw()+
1097   xlab('Año')+ylab('')
1098 ggsave("./TeX/Fig/442_1_temporal.pdf",width = 12,height = 6,units =
       "cm")
1099
1100 coef(base)
1101
```

```r
# h dat
h_dat %>% ggplot(aes(x=1990+as.numeric(year),y=h))+geom_point()

# fitted values
ajust<-baseic$fitted.values %>% as.data.frame() %>%
  mutate(year=names(base$residuals[[1]]),
         model=c(rep('d',26),rep('l',27))) %>%
  pivot_longer(!c(year,model),names_to='code',values_to='fitted') %>%
  mutate(fitted=na_if(fitted,0)) %>% na.omit()
ajust %>% ggplot(aes(x=as.numeric(year),y=fitted))+geom_point()
# errores ####

lsfun<-list("media"=mean,#"dev"=sd,
            "plo"=function(x,na.rm=T) quantile(x,probs = 0.25,na.rm),
            "pup"=function(x,na.rm=T) quantile(x,probs = 0.75,na.rm))


modelos<-list("(3) Temp"=base,"(1) Indiv"=base.ind,"(2) Indiv+crisis"=baseic)

for (i in 1:length(modelos)) {
  errores<-modelos[[i]]$residuals %>% as.data.frame() %>%
    mutate(year=names(base$residuals[[1]]),
           model=c(rep('d',26),rep('l',27))) %>%
    pivot_longer(!c(year,model),names_to='code',values_to='resid') %>%
    mutate(resid=na_if(resid,0)) %>% na.omit()
  temp<-errores %>% group_by(year) %>% summarise(across(resid,lsfun)) %>% mutate(modelo=names(modelos)[i])
  if (i==1) todo=temp else todo=rbind(todo,temp)
}

todo %>% ggplot(aes(x=as.numeric(year),y=resid_media))+
  geom_ribbon(aes(ymin=resid_plo,ymax=resid_pup,fill=modelo),alpha=0.2)+
  geom_line(aes(col=modelo),size=1)+theme_bw()+
  xlab("Año")+ylab("")+labs(col="",fill="")+
  theme(legend.position = c(.45,.95),
        legend.background = element_rect(fill = "transparent"))+
  guides(fill=guide_legend(nrow=1))
ggsave("./TeX/Fig/442_2_errores.pdf",width = 12,height = 5,units = "cm",bg="transparent")


summary(base)
coeftest(base)
```

../Code/6_linearmodels.R

A.2. Información entidades federativas

```
1
2  # paquetes
3  library(tidyverse)
4  library(ggsci) # paletas de colores de journals
5  library(ggplot2)
6  library(compositions)
7  library(e1071) #para sesgo y kurtosis
8  library(xtable)
9  library(readxl)
10 library(janitor)
11 library(plm)
12 library(lmtest)
13 library(stargazer)
14 library(ggrepel)
15
16 # datos ####
17 setwd("C:/Users/sgome/Dropbox/#tesis")
18
19 #2015
20 pef15<-read_excel(path = "./Data/PEF/PEF2015_AC01.xlsx",sheet = 2)
21 #2016
22 pef16<-read_excel(path = "./Data/PEF/PEF2016_AC01.xlsx",sheet = 2)
23 #2017
24 pef17<-read_excel(path = "./Data/PEF/pef_ac01_2017.xlsx",sheet = 2)
25 #2018
26 pef18<-read_excel(path = "./Data/PEF/PEF_2018.xlsx",sheet = 1)
27 #2019
28 pef19<-read_excel(path = "./Data/PEF/PEF_2019.xlsx",sheet = 1)
29 #2020
30 pef20<-read_excel(path = "./Data/PEF/PEF_2020.xlsx",sheet = 1)
31 #2021
32 pef21<-read_excel(path = "./Data/PEF/PPEF_2021.xlsx",sheet = 1)
33
34 #selección de variables ####
35 names(pef15) #c(1,9,24,25,27)
36 names(pef16) #c(1,9,26,27,29)
37 names(pef17) #c(1,9,26,27,29)
38 names(pef18) #c(1,9,30,31,33)
39 names(pef19) #c(1,9,30,31,33)
40 names(pef20) #c(1,9,30,31,33)
41 names(pef21) #c(1,9,30,31,33)
42
43
44 clmns<-c(c(1,3,7,9,24,25,27),c(1,3,7,9,26,27,29),c
       (1,3,7,9,26,27,29),
45           c(1,3,7,9,30,31,33),c(1,3,7,9,30,31,33),c
                (1,3,7,9,30,31,33),
46           c(1,3,7,9,30,31,33)) %>%
47   matrix(nrow = 7,ncol=7) %>% t()
48 clnmb<-c("year","ramo","grf","fnmx","codef","entfed","gto")
49 colnames(clmns)<-clnmb
50
51 pef15 %>% dplyr::select(clmns[1,]) %>% head()
52 # pef16 %>% dplyr::select(clmns[2,]) %>% head()
53 # pef17 %>% dplyr::select(clmns[3,]) %>% head()
54 # pef18 %>% dplyr::select(clmns[4,]) %>% head()
55 # pef19 %>% dplyr::select(clmns[5,]) %>% head()
```

```
# pef20 %>% dplyr::select(clmns[6,]) %>% head()
# pef21 %>% dplyr::select(clmns[7,]) %>% head()

todo<-list(pef15,pef16,pef17,pef18,pef19,pef20,pef21)
cont=0
for(i in todo){
  temp<- i %>% dplyr::select(clmns[cont+1,])
  names(temp)<-clnmb
  temp<-temp %>%
    filter(ramo!="Aportaciones a Seguridad Social"|fnmx!="
        Protección Social") %>%
    group_by(year,ramo,grf,fnmx,codef,entfed) %>% summarise(gto=sum
        (gto)) %>% ungroup()
  if (cont==0) PEF<-temp else PEF<-PEF %>% bind_rows(temp)
  cont=cont+1
}
head(PEF)
PEF %>% group_by(year) %>% summarise(gto=sum(gto)/1e9)

unique(PEF$year)
unique(PEF$ramo)
unique(PEF$grf)
unique(PEF$entfed)
unique(PEF$fnmx)

# rm(i,todo,pef16,pef17,pef18,pef19,pef20,pef21,todo,clmns,clnmb)
# write.csv(PEF, file = "./Data/PEF/resumenPEF.csv",row.names = F)


PEF<-read.csv("./Data/PEF/resumenPEF.csv")


# catálogos ####
# Entidades Federativas
original<-unique(PEF$entfed)
abrevia<-read.csv("./Data/PEF/edo_codigos.csv",encoding = "UTF-8")
cat_entidad<-data.frame(entfed=original,
                        estado=ifelse(grepl("Distribuible|
                            Extranjero",original),"ND", # no
                            distribuible
                                      ifelse(original=="Distrito
                                          Federal","Ciudad de
                                          México", # nuevo nombre
                                          cdmx
                                             ifelse(original=="
                                                 Estado de México",
                                                 "México",original)
                                          ))) %>%# nombre
                                          original
  merge(abrevia %>% dplyr::select(Estado,ISO3),by.x = "estado",by.y
      = "Estado",all.x = T) %>%
  mutate(ISO3=ifelse(estado=="ND","ND",ISO3))

# Funciones México a FMI
original<-unique(PEF$fnmx)
temp<-unique(PEF %>% dplyr::select(grf,fnmx))
original
```

```
cat_funcion<-data.frame(fnmx=original,
                        #corto=original,
                        fmi=c("gob","eco","gob","edu","gob","sal","soc","viv","seg","eco",
                              "eco","seg","eco","eco","def","gob","eco","cul","gob","gob",
                              "soc","gob","amb","gob","gob","gob","eco","eco"))
rm(original,abrevia,temp)
# Traducción base PEF a catálogos
peftrad<-PEF %>%
  merge(cat_funcion) %>%
  #group_by(year,entfed,fmi) %>% summarise(gto=sum(gto)) %>% ungroup() %>%
  merge(cat_entidad) %>%
  group_by(ISO3,estado,year,fmi) %>% summarise(gto=sum(gto)) %>% ungroup()
head(peftrad)
peftrad %>% group_by(year) %>% summarise(gto=sum(gto)/1e9)

# Gasto x estado ####
gedo<-peftrad %>%#filter(ISO3!="ND") %>%
  dplyr::select(-estado) %>%#group_by(ISO3,year) %>% summarise(gto=sum(gto)) %>%#suma de funciones
  pivot_wider(names_from = ISO3,values_from=gto) %>% relocate(ND,.after=ZAC) %>%# estados a columnas, separar No distribuible
  #column_to_rownames("year") %>% acomp() %>% unclass() %>% as.data.frame() %>%
  #rownames_to_column("year") %>%# composición de estados
  mutate(tot=dplyr::select(.,AGU:ZAC) %>% rowSums(na.rm = T),across(AGU:ZAC,~./tot)) %>% dplyr::select(-tot) %>%#%total
  pivot_longer(AGU:ZAC,names_to = "ISO3",values_to="gtop") %>%# gasto en % del total
  mutate(gtop=replace_na(gtop,0)) # ceros en NA
  #group_by(ISO3) %>% summarise(gto=geometricmean(gto)) %>% arrange(desc(gto))
# gedo %>% group_by(year) %>% summarise(gtop=sum(gtop))

# prorratear No distribuible entre estados por Gasto####
pefprorr<-peftrad %>% merge(gedo,all.x = T) %>%#agregar porcentajes
  mutate(gtoorig=gto,gto=gto+ND*gtop) %>% filter(ISO3!="ND") # agregar gasto ND prorrateado a gasto, guardar orig
total<-pefprorr %>% group_by(year,fmi) %>% summarise(gto=sum(gto),gtoorig=sum(gtoorig)) %>%
  mutate(estado="Total",ISO3="TOT") # crear tabla con gasto total
total %>% group_by(year) %>% summarise(gto=sum(gto)/1e9,gtoorig=sum(gtoorig)/1e9) #check gastos totales

peftrad<-pefprorr %>% dplyr::select(-c(ND,gtop)) %>%# retirar auxiliares de prorrateo
  bind_rows(total)
rm(pefprorr,gedo) #borrar auxiliares

# Resumenes gasto ####
# Total por años en millones de pesos corrientes
```

```
141 peftrad %>% filter(ISO3!="TOT") %>% group_by(year) %>% summarise(
      gto=sum(gto)/1e6) %>%
142   ggplot(aes(x=year,y=gto)) + geom_point()+geom_text(aes(x=year,y=
        gto,label=round(gto/1e6,2)))
143
144 # Gasto en pesos constantes ####
145
146 # Gasto per cápita ####
147
148 # Gasto en % de PIB (sólo totales) ####
149 # porcentaje de PIB total
150 pib<-read_excel("./Data/PEF/PIBT_5.xlsx",sheet=3)
151 inf<-c(0.04,0.03) #inflación esperada 20-21
152 g<-c(-0.1,0.03) # tasa pib real esperada 20-21
153 proy<-data.frame(year=2020:2021,PIBcorr=c(NA,NA))
154 pib<-pib %>% bind_rows(proy)
155 pib$PIBcorr[6]<-pib$PIBcorr[5]*(1+g[1])*(1+inf[1]) # estimación
      simple PIB 2020
156 pib$PIBcorr[7]<-pib$PIBcorr[6]*(1+g[2])*(1+inf[2]) # estimación
      simple PIB 2021
157
158 # Gráfica Total por años en porcentaje de PIB
159 peftrad %>% filter(ISO3!="TOT") %>% group_by(year) %>% summarise(
      gto=sum(gto)/1e6) %>%
160   merge(pib) %>% mutate(gtop=gto/PIBcorr*100) %>%
161   ggplot(aes(x=year,y=gtop)) + geom_point() +geom_text(aes(x=year,y
        =gtop,label=round(gtop,2)))
162 rm(inf,g,proy)
163
164 # tamaño promedio funciones
165 peftrad %>% filter(ISO3!="TOT") %>%#eliminar total para no
      duplicar
166   group_by(year,fmi) %>% summarise(gto=sum(gto)/1e6) %>%# suma por
        año-función
167   merge(pib) %>% mutate(gtop=gto/PIBcorr*100) %>%# gasto en % de
        PIB
168   group_by(fmi) %>% summarise(gtop=mean(gtop)) %>%# media
        aritmetica del gasto por función
169   arrange(desc(gtop)) #ordenar
170
171 # Tabla gasto (total Mex) por función en %PIB para cada año
172 peftrad %>% filter(ISO3!="TOT") %>% group_by(year,fmi) %>%
      summarise(gto=sum(gto)/1e6) %>%
173   merge(pib) %>% mutate(gtop=gto/PIBcorr*100) %>% dplyr::select(fmi
        ,year,gtop) %>%
174   pivot_wider(names_from = year,values_from=gtop) %>% arrange(desc
        (`2017`)) %>%
175   adorn_totals("row")
176 # más o menos se parece a lo de SHCP, 3% de diferencia en social
177 # con el prorrateado se parece mas el nivel x funcion y orden de
      funciones pero el total no cuadra con neto SHCP
178
179 # Gráfica trends funciones
180 peftrad %>% group_by(year,fmi) %>% summarise(gto=sum(gto)/1e6) %>%
181   merge(pib) %>% mutate(gtop=gto/PIBcorr*100) %>%
182   filter(fmi%in%c("soc","gob","eco","edu","sal")) %>%
183   ggplot(aes(x=year,y=gtop,col=fmi)) + geom_line(size=1)
```

```
184
185
186 # Detalle de social que crece rápido
187 pefsoc<-PEF %>%
188   merge(cat_funcion) %>%
189   group_by(year,entfed,fmi,ramo) %>% summarise(gto=sum(gto)) %>%
          ungroup() %>%
190   filter(fmi=="soc") %>%
191   merge(cat_entidad %>% dplyr::select(entfed,ISO3)) %>% relocate(
          ISO3,entfed) %>%
192   group_by(year,ramo) %>% summarise(gto=sum(gto)/1e6) %>%
193   merge(pib) %>% mutate(gtop=gto/PIBcorr*100)
194 temp<-pefsoc %>% group_by(year) %>% summarise(gto=sum(gtop)) #
      resumen porcentajes
195 temp[7,2]-temp[1,2]  #incremento
196
197 temp<-pefsoc %>% dplyr::select(year,ramo,gtop) %>%
198   pivot_wider(names_from = ramo,values_from=gtop)
199 temp[is.na(temp)]<-0
200 t(temp[7,2:13]-temp[1,2:13]) %>% round(2) %>% as.data.frame() %>%
      arrange(desc(V1))
201 # IMSS, Bienestar e ISSTE traen aumento
202 rm(pefsoc)
203
204 # subcomp. funciones por estado/año ####
205 edodat<-peftrad %>% dplyr::select(-gtoorig) %>% pivot_wider(names_
      from = fmi,values_from=gto) %>%
206   mutate(def=replace_na(def,0)) %>%
207   #dplyr::select(-estado) %>% unite("id",c(ISO3,year),sep = ".")
        %>%
208   #column_to_rownames("id") %>%
209   mutate(tot= dplyr::select(.,amb:def) %>% rowSums()) %>%# gasto
        total
210   mutate(across(amb:def, ~./tot)) %>%# convertir a composición
        suma 1
211   #groupparts("resto"=c("seg","def","amb","cul","viv"))
212   mutate(resto=seg+def+amb+cul+viv) %>% dplyr::select(-c(tot,seg,
        def,amb,cul,viv)) %>%# agrupar chicos
213   relocate(c(soc,edu,eco,sal,resto,gob),.after=estado) #orden
214
215 head(edodat)
216 apply(edodat[,4:9],1,sum) #check suma 1
217 # pegar media de estados para cada año
218 yearmed<-function(anio,metodo=TRUE){ #media composicional por año
219   dat<-edodat %>%
220     filter(year==anio) %>% dplyr::select(-c(ISO3:year,estado)) %>%
221     acomp()
222   resultado<-mean(dat,robust=metodo) %>% as.matrix()
223   return(resultado)
224 }
225
226 yearmed(2015)
227 x<-2015:2021
228 promedio<-sapply(x,yearmed) %>% t() %>%  as.data.frame()
229 colnames(promedio)<-names(edodat)[4:9]
230 promedio$year<-x;promedio$estado<-rep("Media",7);promedio$ISO3<-rep
      ("MED",7)
```

```
231 edodat<-bind_rows(edodat,promedio)
232
233 # Media estados ####
234 x<-unique(edodat$ISO3)
235 edomed<-function(edo,metodo=TRUE){ #media composicional por estado
236   dat<-edodat %>%
237     filter(ISO3==edo) %>% dplyr::select(-c(ISO3:estado)) %>%
238     acomp()
239   resultado<-mean(dat,robust=metodo) %>% as.matrix()
240   return(resultado)
241 }
242
243 media<-sapply(x,function(e) edomed(e,F)) %>% t() %>% as.data.frame
      ()
244 names(media)<-names(edodat)[4:9]
245 head(media)
246 apply(media, 1, sum)# check suma 1
247
248 # plot(acomp(media))
249 plot(acomp(media),margin = "gob")
250
251 write.csv(media,"./Data/PEF/mediaedos.csv")
252 media<-read.csv("./Data/PEF/mediaedos.csv") %>% column_to_rownames(
      "X")
253
254
255 # Pruebas de normalidad del gasto por estado
256 library(MVN)
257 library(robCompositions)
258 temp<-alr(media)
259 out<-outCoDa(media)
260 out
261 rownames(media)[out$outlierIndex]
262 mvn(temp,mvnTest = "dh")
263 mvn(temp[!out$outlierIndex,],mvnTest = "dh")
264 mvn(temp[!out$outlierIndex,],mvnTest = "mardia")
265 mvn(temp[!out$outlierIndex,],mvnTest = "royston")
266 mvn(temp[!out$outlierIndex,],mvnTest = "hz")
267
268 # Dendrograma de variables
269 temp<-acomp(edodat %>% filter(!ISO3 %in% c("TOT","MED")) %>% dplyr
      ::select(-c(ISO3:estado)))
270 orden<-mean(temp) %>% sort(decreasing = T) %>% names()
271 head(temp[,orden])
272 varD<-variation(temp[,orden])
273 # Tabla variación composicional
274 varD %>% as.data.frame() %>% bind_rows(mean(temp[,orden])) %>%
275   bind_rows(cumsum(mean(temp[,orden])))
276 #dendrograma funciones con distancia: variación composicional
277 vardend<-varD %>% as.dist() %>% hclust(method = "ward.D") %>%
278   as.dendrogram()
279 plot(vardend)
280 # Dendrograma variables en media
281 temp<-acomp(media)[1:32,c(6,1:5)]
282 varD<-variation(temp)
283 varD %>% as.data.frame() %>% bind_rows(mean(temp)) %>%
284   bind_rows(cumsum(mean(temp)))
```

```
vardend<-varD %>% as.dist() %>% hclust(method = "ward.D") %>%
  as.dendrogram()
plot(vardend)



# Dendrograma ####
# Dendrograma estados
par(mar=c(0,3,2,1))
dendpai<-media[1:32,] %>% alr() %>% dist() %>% hclust(method = "
    ward.D")
pdf("./TeX/Fig/45_1_dendro.pdf",width = 6,height = 4)
plot(dendpai,main = NA,sub=NA,ylab = "Distancia")
dev.off()

# Circular
library(circlize)
library(dendextend)
dend <- dendpai %>% as.dendrogram %>%
  set("branches_lwd", 2) %>%
  set("branches_lty", 1) %>%
  set("labels_cex", 1)

circlize_dendrogram(dend,facing = "outside")#,
                    #labels_track_height = .4,
                    #dend_track_height = 0.5)

# alr PCA ####
n=6
res.pca <- prcomp(alr(media[1:32,]),  scale = TRUE)
#biplot(res.pca)

# Coordinates of individuals
library(factoextra) # para get_pca_ind
ind.coord <- as.data.frame(get_pca_ind(res.pca)$coord)
# Add clusters obtained using the dendrogram cutree

ind.coord$Grupo<-factor(cutree(dendpai,n))
# Add Species groups from the original data sett
ind.coord$code<-rownames(ind.coord)
# Percentage of variance explained by dimensions
eigenvalue <- round(get_eigenvalue(res.pca), 1)
variance.percent <- eigenvalue$variance.percent
head(eigenvalue)

# Coeficientes de PCA ####
# Helper function
#:::::::::::::::::::::::::::::::::::::::
var_coord_func <- function(loadings, comp.sdev){
  loadings*comp.sdev
}
# Compute Coordinates
#:::::::::::::::::::::::::::::::::::::::
loadings <- res.pca$rotation
sdev <- res.pca$sdev
var.coord <- t(apply(loadings, 1, var_coord_func, sdev))
head(var.coord[, 1:4])
```

```
341 print(xtable(var.coord))
342
343 #datos transformados completos vs Grupo
344 edogrupo<-alr(edodat[,4:9]) %>% as.data.frame() %>% mutate(code=
        edodat$ISO3,year=edodat$year) %>%
345   merge(ind.coord %>% dplyr::select(code,Grupo))
346
347 #Estados por grupo
348 temp<-ind.coord %>% group_by(Grupo) %>% summarise(N=n())
349
350 ndato<-temp %>% merge(edogrupo %>% group_by(Grupo) %>% summarise(Ti
        =n())) %>%
351   merge(ind.coord %>% group_by(Grupo) %>% summarise(across(Dim.1:
          Dim.2,mean))) %>%
352   mutate(etiq=paste(N,"estados"))
353
354 #Ajustes para visualizar etiquetas de grupo
355 ndato$Dim.1_adj<-ndato$Dim.1#+c(0.5,-0.25,-0.5,-1,0.5,-0.5)
356 ndato$Dim.2_adj<-ndato$Dim.2#+c(-0.5,0.95,0.75,1,0.5,-0.5)
357
358 library(ggrepel)
359 library(ggpubr)
360 pcagrupo<-ggscatter(
361   ind.coord, x = "Dim.1", y = "Dim.2",
362   color = "Grupo", palette = "npg", ellipse = TRUE, ellipse.type =
          "convex",
363   #shape = "Nivel",
364   size = 2, rug=F, legend = "right", ggtheme = theme_bw(),
365   show.legend.text = TRUE,
366   xlab = paste0("Componente 1 (", variance.percent[1], "%)" ),
367   ylab = paste0("Componente 2 (", variance.percent[2], "%)" )
368 ) +
369   #stat_mean(aes(col=Grupo), size = 3)+
370   coord_fixed() + #xlim = c(-6,3),ylim = c(-4,2))+#default xlim = c
          (-5.1,2.46),ylim = c(-3.88,1.65)
371   geom_text_repel(data = ind.coord,aes(x=Dim.1,y=Dim.2,label=code))
372   #geom_text(data = subset(ind.coord,code=="TOT"),aes(Dim.1,Dim.2,
          label="Promedio"),nudge_x = -0.2,nudge_y = -0.2)#+
373   #geom_text(data=ndato, aes(Dim.1_adj,Dim.2_adj,label=etiq))
374 pcagrupo
375
376
377
378
379
380
381
382
383 # PCA para comparar con paises####
384 #library(factoextra)
385 mediap<-read.csv("./Data/FMI/mediapaises.csv") %>% column_to_
        rownames("X")
386 res.pca <- prcomp(alr(mediap), scale = TRUE)
387 eigenvalue <- round(get_eigenvalue(res.pca), 1)
388 variance.percent <- eigenvalue$variance.percent
389 nivtrans<-function(x,nivel="Nivel"){# para ordenar niveles de
        ingreso
```

```
  x %>% mutate(Nivel=factor(get(nivel),levels = c("Bajo","Medio-
      bajo","Medio-alto","Alto","México")))
}
ind.coordp<-read.csv("./Data/FMI/mediaspaisesPCA.csv",
                     colClasses = c("character",rep("numeric",5),"
                         factor","character","factor")) %>%
  nivtrans()

#library(ggpubr)
pcagrupo<-ggscatter(
  ind.coordp, x = "Dim.1", y = "Dim.2",
  color = "Grupo", palette = "npg", ellipse = TRUE, ellipse.type =
      "convex",
  shape = "Nivel", size = 2, rug=F, legend = "right", ggtheme =
      theme_bw(),
  show.legend.text = TRUE,
  xlab = paste0("Componente 1 (", variance.percent[1], "%)" ),
  ylab = paste0("Componente 2 (", variance.percent[2], "%)" )
) +
  stat_mean(aes(col=Grupo), size = 3)+
  coord_fixed(xlim = c(-6,3),ylim = c(-4,2))#+#default xlim = c
      (-5.1,2.46),ylim = c(-3.88,1.65)
  #geom_text(aes(Dim.1,Dim.2,label=paises$esp))
  #geom_text(data = subset(ind.coord,code=="MEX"),aes(Dim.1,Dim.2,
      label=pais),nudge_x = -0.2,nudge_y = -0.2)+

pcagrupo

mediae<-read.csv("./Data/PEF/mediaedos.csv") %>% column_to_rownames
    ("X")
alr(mediae) %>% head()
dimnames(alr(mediae))
ind.sup.coord <- predict(res.pca, newdata = alr(mediae) %>% as.data
    .frame()) %>%
  as.data.frame()%>%
  rownames_to_column("code")
names(ind.sup.coord)[2:6]<-paste("Dim",1:5,sep = ".")

head(ind.sup.coord)
library(ggrepel)
pcagrupo+
  geom_point(data = ind.sup.coord,aes(x=Dim.1,y=Dim.2))+
  geom_text_repel(data = ind.sup.coord,aes(x=Dim.1,y=Dim.2,label=
      code))

library(gginnards)
pcagrupo
p<-delete_layers(pcagrupo,idx=1L)+
  geom_point(data = ind.sup.coord,aes(x=Dim.1,y=Dim.2))+
  geom_text_repel(data = ind.sup.coord,aes(x=Dim.1,y=Dim.2,label=
      code))
p
ggsave(plot = p,filename = "./TeX/Fig/45_2_pca.pdf",width = 15,
    height = 10,units = "cm")


# composición de gasto por estados ####
```

```
436 lsfun<-list("media"=mean,"dev"=sd,
437               "p25"=function(x,na.rm=T) quantile(x,probs = 0.25,na.rm
                      ),
438               "p75"=function(x,na.rm=T) quantile(x,probs = 0.75,na.rm
                      ),
439               "mín"=min,"máx"=max,
440               "sesgo"=skewness,"kurtosis"=kurtosis)
441 peftrad %>% filter(ISO3!="TOT") %>% mutate(gto=gto/1e9) %>%
442   group_by(year) %>% summarise(across(gto,lsfun))
443 head(peftrad)
444
445 # Total
446 peftrad %>% filter(ISO3!="TOT") %>%#mutate(gto=gto/1e9) %>%
447   ggplot(aes(gto))+
448   geom_density()+
449   scale_x_log10()
450
451 # distribución gtototal por año
452 peftrad %>% filter(ISO3!="TOT") %>% mutate(gto=gto/1e9) %>%
453   ggplot(aes(factor(year),gto))+
454   geom_violin()+
455   scale_y_log10() # no se aprecia gran diferencia
456
457 # Top estados
458 peftrad %>% filter(ISO3!="TOT") %>% group_by(ISO3,year) %>%
      summarise(gto=sum(gto)/1e9) %>%
459   group_by(ISO3) %>% summarise(gto=mean(gto)) %>% arrange(desc(gto)
      ) #miles de millones
460 # en términos del total del gasto, percápita o de PIB
461
462 # composición partes estado
463 temp<-peftrad %>% filter(ISO3!="TOT") %>% group_by(ISO3,year) %>%
      summarise(gto=sum(gto)) %>%#suma de funciones
464   pivot_wider(names_from = ISO3,values_from=gto) %>% column_to_
      rownames("year") %>%
465   acomp() %>% unclass() %>% as.data.frame() %>% rownames_to_column(
      "year") %>%# composición de estados
466   pivot_longer(AGU:ZAC,names_to = "ISO3",values_to="gto") %>%
467   group_by(ISO3) %>% summarise(gto=mean(gto)) %>% arrange(desc(gto)
      )
468 temp
469 temp %>% ggplot(aes(x=reorder(ISO3,gto),label=round(gto*100,1))) +
      geom_bar(aes(weight=gto*100))+
470   geom_text(aes(y=100*gto+2.5),size=3)
471 sum(temp$gto)
472
473 tab<-temp %>% mutate(gtoacu=cumsum(gto))
474 plot(1:32,tab$gtoacu,type = "b")
475 # promedio ponderado con gastos
476
477 # PIB por estado ####
478 temp<-cat_entidad %>% filter(ISO3!="ND",entfed!="Distrito Federal")
      %>% arrange(estado)
479 pibent<-read_excel("./Data/PIB/PIBE/tabulados_pibent/PIBE_2.xlsx",
      skip = 4,n_max = 38) %>%#mdp constantes 2013
480   na.omit() %>% filter(!grepl("roducto|Unidos",Concepto)) %>%
481   arrange(Concepto) %>%
```

```
482   mutate(ISO3=temp$ISO3,estado=temp$estado) %>% dplyr::select(-
          Concepto) %>%
483   #relocate(Concepto,ISO3,estado) %>%
484   pivot_longer(cols = '2003':'2018R',names_to='year',values_to='pib
          ') %>%
485   mutate(year=as.numeric(substr(year,1,4)))
486
487 pibent %>% filter(year==2018) %>%
488   ggplot(aes(reorder(ISO3,-pib),pib/1e6))+geom_bar(stat="identity",
          position = "dodge")+
489   theme_bw()+theme(axis.text.x = element_text(angle=90,vjust = 0.5)
          )
490
491 # población CONAPO (NEW) ####
492 ping<-read_excel("./Data/INEGI/poblacióningresos.xlsx",skip = 4)
      %>%
493   pivot_longer(cols = !c(Entidad,Periodo),names_to="variable",
          values_to="pob") %>%
494   filter(grepl("Cuarto",Periodo),!grepl("cv|est",variable)) %>%
495   pivot_wider(names_from = variable,values_from=pob) %>%
496   mutate(year=as.numeric(substr(Periodo,22,25))) %>%
497   mutate(across(Total:s_ne,as.numeric)) %>% arrange(Entidad,year)
498 cat_ent2<-data.frame(Entidad=sort(unique(ping$Entidad)[-30]), # 30
      es el total
499                      estado=unique(cat_entidad$estado)[-19]) %>%#
                         19 es No Dististribuible
500   merge(cat_entidad %>% filter(entfed!="Distrito Federal") %>%
          dplyr::select(estado,ISO3))
501
502 conapo<-read.csv("./Data/CONAPO/pob_mit_proyecciones.csv") %>%
503   rename(year=AÑO,estado=ENTIDAD) %>% filter(year>=2003,year<=2021,
          estado!="República Mexicana") %>%
504   group_by(estado,year) %>% summarise(pob=sum(POBLACION)) %>%
          ungroup() %>%
505   arrange(estado,year) %>% mutate(ISO3=rep(cat_ent2$ISO3,each=19))
506
507 conapo %>%#filter(year>=2010,estado!="República Mexicana") %>%
508   group_by(year) %>% summarise(pob=sum(pob)/1E6) %>%
509   ggplot(aes(year,pob))+geom_line()
510
511
512 # Poblacion por estado (OLD)###
513 pobent<-read_excel("./Data/PIB/PIBE/POBent.xlsx",skip=4) %>%
514   na.omit() %>% rename(entidad='...1') %>% dplyr::select(entidad,
          T2000,T2005,T2010) %>%
515   mutate(r1=(T2005/T2000)^(1/5)-1,r2=(T2010/T2005)^(1/5)-1,r3=(r1+
          r2)/2) %>%#tasa c.pob. promedio
516   mutate(across(r1:r3,~.*100))
517
518 for (i in 1:8){ #pronóstico de población del 2014 al 2021
519   if (i==1) {
520     proy=pobent$T2010*(1+pobent$r3/100)^(3+i)
521     }
522   else {
523     proy=cbind(proy,pobent$T2010*(1+pobent$r3/100)^(3+i))
524   }
525 }
```

```
526 colnames(proy)<-2014:2021
527 apply(proy,2,sum)/1e6 # total población méxico por año
528 plot(apply(proy,2,sum)/1e6)
529
530 #temp<-apply(proy,2,sum)/1e6;temp[2:8]-temp[1:7]
531
532 pobent<-cbind(pobent[,1:4],proy)
533 apply(pobent[,2:12],2,sum)/1e6
534 pobent<-pobent %>% arrange(entidad) %>%
535   mutate(ISO3=temp$ISO3,estado=temp$estado) %>%
536   relocate(entidad,ISO3,estado) %>% dplyr::select(-c(entidad,T2000,
        T2005,T2010)) %>%
537   pivot_longer(cols = '2014':'2021',names_to='year',values_to='pob'
        )
538
539
540
541
542 # Proyección población 2 ###
543 pobent<-read_excel("./Data/PIB/PIBE/Poblacion_Hist.xlsx",skip=3)
      %>%
544   rename(entidad="...1") %>% select(-2) %>% na.omit() %>%
545   pivot_longer(cols = `1990`:`2010`,names_to="year",values_to="pob"
        ) %>%
546   mutate(year=paste("y",year,sep="")) %>% pivot_wider(names_from =
        year,values_from=pob) %>%
547   mutate(across(y1990:y2010,as.numeric)) %>%
548   mutate(r95=100*((y1995/y1990)^0.2-1),r00=100*((y2000/y1995)
        ^0.2-1),
549          r05=100*((y2005/y2000)^0.2-1),r10=100*((y2010/y2005)
               ^0.2-1))
550
551
552 pobent %>%
553   mutate(across(y1990:y2010,~./y1990)) %>% pivot_longer(cols =
        y1990:y2010) %>%
554   mutate(year=rep(seq(1990,2010,5),33)) %>%
555   ggplot(aes(x=year,y=log(value),col=entidad))+geom_line()+theme(
        legend.position = "none")
556
557 ppob<-pobent %>% select(entidad,y1990:y2010) %>%
558   pivot_longer(cols = y1990:y2010,names_to="time",values_to="pob")
        %>%
559   mutate(year=rep(seq(1990,2010,5),33)-1990) %>%#años desde el
        inicio
560   group_by(entidad) %>%
561   mutate(indice=pob/first(pob),r=-100+100*(pob/dplyr::lag(pob))^(1/
        5)) %>%#indice y tasa
562   ungroup() %>%
563   filter(entidad!="Estados Unidos Mexicanos") %>% arrange(entidad)
        %>%#sin total
564   mutate(ISO3=rep(cat_ent2$ISO3,each=5)) %>%# agregar iso
565   select(-entidad) %>%#filter(year!=0) %>%
566   pdata.frame(index=c("ISO3","time"))
567
568 ppob %>% ggplot(aes(year,r,col=ISO3))+geom_line()+theme(legend.
      position = "none")
```

```
569
570 eq<-log(pob)~year
571 eq2<-r~year
572
573 g_mco<-plm(eq2,ppob,model = "pooling") # ols
574 g_int<-plm(eq2,ppob,model = "within", effect = "individual") #
      within
575 g_mcg<-plm(eq2,ppob,model = "random", effect = "individual") # gls
576
577
578 #presencia de efectos
579 pres<-c(format(pFtest(update(g_int,effect="individual"),g_mco)$p.
      value,digits = 2),"","")
580 #Hausmant test, correlación de efectos
581 gcor<-c("","",format(phtest(g_int,g_mcg)$p.value,digits = 2))
582 #AR(1)
583 ar1<-sapply(list(g_mco,g_int,g_mcg),function(x) format(pbgtest(x)$p
      .value,digits = 2))
584 #AR(2)
585 ar2<-sapply(list(g_mco,g_int,g_mcg),function(x) format(pbgtest(x,
      order=2)$p.value,digits = 2))
586
587
588 stargazer(g_mco,g_int,g_mcg,digits = 2, type = "text",
589           title = "Regresión de panel para crecimiento económico",
590           label = "tab:gmod0_edos",
591           column.labels = c("$\\hat{\\und\\gamma}_{\\text{MCO}}$",
592                             "$\\hat{\\und\\gamma}_{\\text{W}}$",
593                             "$\\hat{\\und\\gamma}_{\\text{MCG}}$"),
594           single.row = T,
595           dep.var.caption = "Variable dependiente: $g_t$ de los
              estados",
596           omit.stat = "F",
597           dep.var.labels.include = F,
598           # covariate.labels=c("periodo","ordenada"),
599           add.lines = list(c("F (efectos) p-val",pres %>% fsignif()
              ),
600                            c("Hausman p-value",gcor %>% fsignif()),
601                            c("AR(1) p-value",ar1 %>% fsignif()),
602                            c("AR(2) p-value",ar2 %>% fsignif()))
603 )
604
605 temp<-data.frame(ISO3=rep(cat_ent2$ISO3,each=10),year=rep(21:30,32)
      ) %>%
606   mutate(pred=predict(g_mcg,newdata = temp),reff=rep(ranef(g_mcg),
        each=10)) %>%
607   mutate(r=pred+reff)
608 ppobest<-ppob %>% merge(temp %>% select(ISO3,year,r),all=TRUE) %>%
609   arrange(ISO3,year) %>% mutate(time=year+1990) %>%
610   group_by(ISO3) %>% mutate(factor=1+r/100)
611
612 for (i in 21:30) {
613   ppobest[ppobest$year==i,"pob"]<-ppobest[ppobest$year==i-1,"pob"]*
        (ppobest[ppobest$year==i,"factor"])
614 }
615
616 ppobest %>% group_by(time) %>% summarise(pob=sum(pob)/1e6)
```

```
617
618
619
620
621 # estimación pib por estado 2019 en adelante ####
622 # PIB total 2019 constantes 2013
623 p19<-(read_excel("./Data/PIB/PIBT_constantes2.xlsx",skip=4,n_max =
        5) %>% t())[190,4] %>% as.numeric()
624 g<-c(-0.1,0.03) # tasa pib real esperada 20-21
625 proy<-data.frame(year=2019:2021,pib=c(p19,p19*(1+g[1]),p19*(1+g[1])
        *(1+g[2])))
626
627 # composición estimada
628 edopib_comp<-pibent %>%dplyr::select(-estado) %>% pivot_wider(names
        _from = ISO3,values_from=pib) %>%
629   column_to_rownames("year") %>% acomp()
630 head(edopib_comp[,1:5])
631
632 nuevos<-data.frame(year=2019:2021)
633 res<-matrix(0,nrow = dim(nuevos)[1],ncol = 32-1)
634 estadistica<-matrix(0,1,31)
635 #proyección composición por estados
636 for(i in 1:31){
637   temp<-data.frame(alr=alr(edopib_comp)[,i],year=2003:2018)
638   mod<-lm(formula = "alr~poly(year,3)",data = temp)
639   res[,i]<-predict(mod,newdata = nuevos)
640   estadistica[[i]]<-summary(mod)$r.squared
641 }
642 summary(t(estadistica))
643 proye<- res %>% alrInv() %>% as.data.frame()
644 colnames(proye)<-names(edopib_comp)
645 proye$year<-2019:2021;proye$pibtot<-proy$pib
646 proye<-proye %>% pivot_longer(cols = AGU:ZAC, names_to="ISO3",
        values_to="comp") %>%
647   mutate(pib=pibtot*comp)
648
649
650 pibent2<-pibent %>%
651   merge(proye %>% dplyr::select(year,ISO3,pib),all = T) %>%
652   group_by(ISO3) %>% fill(estado,.direction = "down") %>% ungroup()
653
654 pibent %>% group_by(year) %>% summarise(pib=sum(pib)) %>%
655   merge(pibent2 %>% group_by(year) %>% summarise(pib2=sum(pib))) #
        check
656 write.csv(pibent,'./Data/temp/pibent.csv',row.names = F)
657 write.csv(pibent2,'./Data/temp/pibent2.csv',row.names = F)
658 # Gráfica datos PIB y Población por estado
659 temp<-pibent2 %>% merge(conapo) %>%
660   mutate(pib=pib/1e6,pob=pob/1e6) #%>%#pib y pob en billones
661 set.seed(1906)
662 temp%>% filter(year<=2018) %>%
663   ggplot(aes(x=pob,y=pib,col=ISO3,label=ISO3))+geom_point()+
664   geom_text_repel(data = subset(temp,year==2018),col="black")+
665   scale_y_log10()+scale_x_log10()+
666   theme_bw()+theme(legend.position = "none")+
667   xlab("Población en millones")+ylab("PIB bdp constantes 2018")
```

```
668 ggsave("./TeX/Fig/45_7_pob_pib.pdf",width = 15,height = 10,units =
        "cm")
669
670
671 # PIB percápita por estado ####
672 pibent <- read.csv('./Data/temp/pibent.csv')
673 pibent2 <- read.csv('./Data/temp/pibent2.csv')
674
675 pibpcent<- pibent2 %>%# PIB más proyecciones '20 y '21
676   merge(conapo,all = T) %>%# proyecciones conapo de población por
          estado
677   mutate(pibpc=pib/pob*1e6, cre=100*(pibpc/dplyr::lag(pibpc)-1)) #
          calculo pibpc y g
678
679 # gráfica PIB pc
680 # pibpcent %>%
681 #   filter(year==2018) %>%
682 #   ggplot(aes(x=reorder(ISO3,pibpc),y=pibpc/1000,label=round(pibpc
        /1000,0)))+
683 #   geom_bar(stat = "identity",position = "dodge",fill=4)+
684 #   geom_text(hjust=-0.2,vjust=0.5,size=2.5,angle=90)+
685 #   #scale_y_log10()+
686 #   coord_cartesian(ylim = c(0,600))+
687 #   theme_bw()+ theme(axis.text.x = element_text(angle=90,vjust
        =0.5))+
688 #   xlab("")+ylab("PIB per cápita mdp")
689 # ggsave("./TeX/Fig/45_3_pibpcedos.pdf",width = 12,height = 7,units
        = "cm")
690
691 # gráfica tasa crecimiento vs PIB pc
692 # pibpcent %>% filter(dplyr::between(year,2015,2019)) %>%# últimos
        5 datos
693 #   ggplot(aes(x=pibpc/1000,y=cre,label=ISO3,col=ISO3))+geom_point
        ()+
694 #   scale_x_log10()+theme_bw()+theme(legend.position = "none")+
695 #   xlab("PIB per cápita mmdp")+ylab("Crecimiento real PIB per
        cápita")
696 # ggsave("./TeX/Fig/45_5_g_pibpc.pdf",width = 12,height = 7,units =
        "cm")
697
698 # Total PIB per cápita
699 temp<-pibpcent %>% group_by(year) %>% summarise(across(pib:pob,sum)
        ) %>%
700   mutate(pibpc=pib/pob*1e6, cre=100*(pibpc/dplyr::lag(pibpc)-1))
          %>%
701   mutate(ISO3="TOT",estado="Total") %>% filter(year>=2015)
702
703 pibpcent<-pibpcent %>% filter(year>=2015) %>% bind_rows(temp) #
        agregar resumen total México
704
705
706 write.csv(edodat,'./Data/PEF/gsubedo.csv',row.names = F)
707 edodat<-read.csv('./Data/PEF/gsubedo.csv')
708 edodat<-edodat %>%
709   merge(pibpcent,all.x = T) #composición del gasto más información
        pib pc
710
```

```
711 rm(temp)
712 #promedios
713 temp<-edodat %>% filter(!grepl('TOT|MED',ISO3)) %>%
714   group_by(year) %>% summarise(across(pib:cre,mean))
715 edodat[edodat$ISO3=='MED',c('year','pib','pob','pibpc','cre')]<-
      temp
716 write.csv(edodat,'./Data/PEF/edodat.csv',row.names = F)
717 edodat<-read.csv('./Data/PEF/edodat.csv')
718
719 edodat %>% filter(ISO3=="CAM") %>% select(ISO3,year,pibpc)
720
721 # rm(pibent,pobent,proy,promedio,tab,temp,total)
722
723 # Mapas ####
724 # library("devtools")
725 # install.packages('digest')
726 # install.packages('rlang')
727 # install.packages('backports')
728 # install.packages('ps')
729 #
730 # devtools::install_github("diegovalle/mxmaps")
731 library(mxmaps)
732 temp<-df_mxstate[,c('region','state_name')]
733 cat_entidad<-cat_entidad %>% merge(temp,all.x = T,by.x = 'estado',
      by.y = 'state_name')
734
735 edodat<-edodat %>%
736   merge(cat_entidad %>% dplyr::select(ISO3,region) %>% unique,all.x
        = T) %>%
737   relocate(ISO3,region)
738
739 edodat %>% filter(year==2015,!is.na(region)) %>%
740   #mutate(value=round(pob/1e6,1)) %>%
741   #mutate(value=round(pib/1e6,1)) %>%
742   #mutate(value=round(eco*100,1)) %>%
743   #mutate(value=round(cre,1)) %>%
744   mutate(value=round(pibpc,1)) %>%
745   mxstate_choropleth()
746 rm(mxstate.map)
747
748 write.csv(edodat,"./Data/PEF/edodat.csv",row.names = F)
749 write.csv(cat_entidad,"./Data/PEF/cat_entidad.csv",row.names = F)
750 # PCA sin atípicos
751
752 # Evolución compo estados
753
754 # Evolución gastoxfunción algunos estados
755
756 # Cruce contra variables por estado: g, esperanza de vida,
      educación, desempleo, ingreso promedio etc
757
758
759
760
761
762
763
```

```
764
765 # Completar gasto en precios constantes
766
767 #PIB total constante
768 temp<-read_excel("./Data/PIB/PIBE/tabulados_pibent/PIBE_2.xlsx",
        skip = 4,n_max = 38) %>%#mdp constantes 2013
769   na.omit() %>% filter(grepl("Producto",Concepto)) %>%dplyr::select
          (-Concepto) %>% t() %>%
770   as.data.frame() %>% rename(pib=V1) %>%
771   rownames_to_column("year") %>% mutate(year=as.numeric(substr(year
          ,1,4))) %>%
772   bind_rows(proy) %>% filter(year>=2015) # agregar proyección
          '20:-10%, '21:2%
773
774 #pib total corriente + deflactor
775 deflac<-pib %>% merge(temp) %>% mutate(factor=pib/PIBcorr) #%>%#
        factor precios 2013
776 deflac$f18=deflac$factor/deflac$factor[4] # factor precios 2018
777
778 #gasto total en precios constantes (usando deflactor del PIB)
779 gtoconst<-peftrad %>% group_by(ISO3,year) %>% summarise(gto=sum(gto
        )) %>%#gasto total precios corrientes
780   merge(deflac %>% dplyr::select(year,factor,f18)) %>%#deflactor
781   mutate(gtoc=gto*f18) #gasto precios constantes 2018
782
783 # Gráficas preliminares y exploratorias ####
784
785
786
787 # datos
788 g_dat<-edodat %>% filter(!ISO3%in%c("TOT","MED")) %>%#excluir
        total y promedio
789   merge(gtoconst %>% dplyr::select(ISO3,year,gto,gtoc,factor,f18))
          %>%# agregar gasto en precios constantes
790   mutate(pib=pib/factor*f18,pibpc=pibpc/factor*f18) %>%# convertir
          PIB a precios constantes 2018
791   mutate(gto2=gtoc/pib/1e6*100) %>%#gto en porcentaje de PIB
792   mutate(gtopc=gtoc/pob/1e3) %>%#gasto per cápita
793   mutate(across(soc:resto,~log(./gob))) %>% dplyr::select(-gob) %>%
          #logcocientes
794   filter(year<=2019) %>% pdata.frame() #quitar proyección '20 y '21
795
796 write.csv(g_dat,'./Data/g_dat_edos.csv',row.names = F)
797 g_dat=read.csv('./Data/g_dat_edos.csv') %>% pdata.frame()
798
799 g_dat %>% filter(ISO3=="CAM") %>% select(ISO3,year,pibpc)
800 g_dat %>% group_by(year) %>% summarise(pib=sum(pib))
801
802 # pib per cápita constantes 2013 caja y brazos (no usado)
803 # g_dat %>%
804 #   ggplot(aes(x=reorder(ISO3,pibpc),y=pibpc/1000,label=round(pibpc
        /12,0)))+geom_boxplot()+
805 #   scale_y_log10()+
806 #   theme_bw()+ theme(axis.text.x = element_text(angle=90,vjust
        =0.5))+
807 #   xlab("")+ylab("PIB per cápita mdp")
```

```
# ggsave("./TeX/Fig/45_3_pibpcedos.pdf",width = 15,height = 10,
    units = "cm")

# pib per cápita 2018 constantes 2013
g_dat %>% as.data.frame() %>%
  filter(year %in% c('2018')) %>%
  ggplot(aes(x=reorder(ISO3,-pibpc),y=pibpc/1000,label=round(pibpc/
      1000,0)))+
  geom_bar(stat = "identity",position = "dodge",fill=4)+
  geom_text(size=2.5,angle=90,nudge_y = c(rep(-30,3),-250,rep
      (-30,28)),col="white")+
  #scale_y_log10()+
  coord_cartesian(ylim = c(0,450))+
  theme_bw()+ theme(axis.text.x = element_text(angle=90,vjust=0.5))
      +
  xlab("")+ylab("PIB per cápita (miles)")
ggsave("./TeX/Fig/45_3_pibpcedos.pdf",width = 12,height = 7,units =
    "cm")

# g en caja y brazos
g_dat %>%
  group_by(ISO3) %>% mutate(m=median(cre)) %>% ungroup() %>%
  ggplot(aes(x=reorder(ISO3,m),y=cre,label=round(cre,1)))+geom_
      boxplot()+
  theme_bw()+xlab("")+ylab("Crecimiento real PIB per cápita")+
  theme(axis.text.x = element_text(angle=90,vjust=0.5))
ggsave("./TeX/Fig/45_4_g_edos.pdf",width = 12,height = 7,units = "
    cm")


# Cuadro medias gasto por grupo
edoagrupado<-edodat %>% filter(!ISO3%in%c("TOT","MED")) %>%#
    excluir total y promedio
  merge(gtoconst %>% dplyr::select(ISO3,year,gto,gtoc,factor,f18))
      %>%# agregar gasto en precios constantes
  mutate(pib=pib/factor*f18,pibpc=pibpc/factor*f18) %>%# convertir
      PIB a precios constantes 2018
  mutate(gto2=gtoc/pib/1e6*100) %>%#gto en porcentaje de PIB
  mutate(gtopc=gtoc/pob/1e3,pob=pob/1e6) %>%#gto per cápita
  #mutate(across(soc:resto,~log(./gob))) %>% dplyr::select(-gob)
      %>%#logcocientes
  filter(year<=2019) %>% #quitar proyección '20 y '21
  merge(ind.coord %>% dplyr::select(code,Grupo), by.x = "ISO3", by.
      y = "code")
edoagrupado %>%
  group_by(Grupo) %>% summarise(across(c(soc:gob,pibpc,cre,gtopc,
      gto2,pob),mean),n=n_distinct(ISO3)) %>%
  mutate(across(soc:gob,~.*100),pibpc=pibpc/1000) %>%
  relocate(gtopc,gto2,.after=gob) %>%
  xtable(caption="Resumen de los estados por grupos",label="tab:
      edos_resumen",
         digits=c(rep(0,8),1,0,0,1,1,0)) %>%
  print(include.rownames=F)

# Gasto % de PIB del estado
g_dat %>% as.data.frame() %>%
  filter(year==2018) %>%
```

```
851    ggplot(aes(x=reorder(ISO3,-gto2),y=gto2,label=format(gto2,digits
           =2)))+
852    geom_bar(stat = "identity",position = "dodge",fill=4)+
853    geom_text(size=2.5,angle=90,nudge_y = -3,col="white")+
854    #coord_cartesian(ylim = c(0,600))+
855    theme_bw()+ theme(axis.text.x = element_text(angle=90,vjust=0.5))
           +
856    xlab("")+ylab("Gasto público (% del PIB)")
857 ggsave("./TeX/Fig/45_8_gtoedos.pdf",width = 12,height = 7,units = "
       cm")
858
859 # gasto per cápita
860 g_dat %>% as.data.frame() %>%
861    filter(year==2018) %>%
862    ggplot(aes(x=reorder(ISO3,-gtopc),y=gtopc,label=round(gtopc,1)))+
863    geom_bar(stat = "identity",position = "dodge",fill=4)+
864    geom_text(size=2.5,angle=90,nudge_y = c(rep(-5,3)
           ,-95,-5,-5,-140.2,rep(-5,25)),col="white")+
865    coord_cartesian(ylim = c(0,60))+
866    theme_bw()+ theme(axis.text.x = element_text(angle=90,vjust=0.5))
           +
867    xlab("")+ylab("Gasto per cápita (miles)")
868 ggsave("./TeX/Fig/45_9_gtopcedos.pdf",width = 12,height = 7,units =
       "cm")
869
870
871 # pib per cápita vs comp ingreso población
872 # library(MASS)
873
874 rm(temp)
875 temp<-g_dat %>%
876    merge(ping2 %>% mutate(s_0a3=s_0+s_0a1+s_1a2+s_2a3,
877                           year=factor(year)) %>%
878             dplyr::select(ISO3,year,s_0a3,s_3a5,s_5mas)) %>%   #
                 agregar población en ingreso bajo
879    group_by(ISO3) %>% summarise(across(c(pibpc,s_0a3:s_5mas),mean))
           %>%
880    pivot_longer(cols = s_0a3:s_5mas,names_to="sm",values_to="pob")
881
882 sm.labs <- c("menos de 3SM", "de 3SM a 5SM", "mas de 5SM")
883 names(sm.labs) <- c("s_0a3", "s_3a5", "s_5mas")
884
885 sm.labs
886
887 temp   %>%
888    ggplot(aes(x=pibpc/1000,y=pob*100,label=ISO3))+geom_point()+
889    geom_smooth(data=subset(temp,ISO3!="CAM"),method = "lm",formula =
           'y~x',se=F)+
890    facet_wrap(~sm,scales = "free",nrow = 1,
891               labeller = labeller(sm = sm.labs))+
892    theme_bw()+theme(legend.position = "none")+
893    scale_x_log10()+xlab("PIB per cápita (miles)")+ylab("Población %
           )
894 ggsave("./TeX/Fig/45_10_pibpc_ing.pdf",width = 15,height = 10,units
       = "cm")
895
896
```

```
897
898 # Ternarias # Gasto
899 edoagrupado %>% dplyr::select(ISO3,year,soc:gob) %>%
900   unite("id",ISO3:year,sep = ".") %>% column_to_rownames("id") %>%
901   acomp() %>%
902   plot(margin="gob")
903
904 # Variación composicional
905 edoagrupado %>% dplyr::select(ISO3,year,soc:gob) %>%
906   unite("id",ISO3:year,sep = ".") %>% column_to_rownames("id") %>%
907   acomp() %>% variation()
908
909 # barras por año
910 edoagrupado %>%
911   group_by(Grupo,year) %>%
912   #summarise(across(c(pob,pib),sum),pibpc=pib/pob) %>%
913   summarise(g=mean(cre),lg=quantile(cre,0.05),ug=quantile(cre,0.95)
        ) %>%
914   ggplot(aes(x=year,y=g)) + geom_bar(stat = "identity",position = "
        dodge")+
915   geom_errorbar(aes(ymin=lg, ymax=ug), width=.2,
916                 position=position_dodge(.9)) +
917   facet_wrap(~Grupo)+theme_bw()
918
919 # boxplot por año
920 edoagrupado %>%
921   ggplot(aes(x=year,y=pibpc/1e3)) + geom_boxplot()+
922   facet_wrap(~Grupo)+
923   scale_y_log10()
924
925 # Composición media, acumulada (gasto, ingresos, pib)
926 # PIB per cápita, gasto percápita, Ingresos población
927 # Cuadro medias gfun por clusters entidades
928 # Normalidad datos estados
929
930 # Modelos x grupos
931
932
933
934 # modelos panel ####
935 g_mco<-plm(cre~gto2+soc+edu+eco+sal+resto,g_dat,model = "pooling")
      # ols
936 g_int<-plm(cre~gto2+soc+edu+eco+sal+resto,g_dat,model = "within",
      effect = "individual") # within
937 g_mcg<-plm(cre~gto2+soc+edu+eco+sal+resto,g_dat,model = "random",
      effect = "individual") # gls
938
939
940 # summary(g_mco)
941 # summary(g_int)
942 # summary(g_mcg)
943
944 fsignif=function(x,lvl=c(0.1,0.05,0.01)){
945   y<-as.numeric(x)
946   codigos = (y<lvl[1]) + (y<lvl[2]) + (y<lvl[3])
947   codigos[codigos==3]<-"$^{***}$"
948   codigos[codigos==2]<-"$^{**}$"
```

```
949   codigos[codigos==1]<-"$^{*}$"
950   res=paste(x,replace_na(codigos,""),sep="")
951   return(res)}
952
953 #presencia de efectos
954 pres<-c(format(pFtest(update(g_int,effect="individual"),g_mco)$p.
      value,digits = 2),"","")
955 #Hausmant test, correlación de efectos
956 gcor<-c("","",format(phtest(g_int,g_mcg)$p.value,digits = 2))
957 #AR(1)
958 ar1<-sapply(list(g_mco,g_int,g_mcg),function(x) format(pbgtest(x)$p
      .value,digits = 2))
959 #AR(2)
960 ar2<-sapply(list(g_mco,g_int,g_mcg),function(x) format(pbgtest(x,
      order=2)$p.value,digits = 2))
961
962
963 stargazer(g_mco,g_int,g_mcg,digits = 2, #type = "text",
964           title = "Regresión de panel para crecimiento económico",
965           label = "tab:gmod0_edos",
966           column.labels = c("$\\hat{\\und\\gamma}_{\\text{MCO}}$",
967                             "$\\hat{\\und\\gamma}_{\\text{W}}$",
968                             "$\\hat{\\und\\gamma}_{\\text{MCG}}$"),
969           single.row = T,
970           dep.var.caption = "Variable dependiente: $g_t$ de los
                estados",
971           omit.stat = "F",
972           dep.var.labels.include = F,
973            covariate.labels=c("gto","$\\log$(soc/gob)","$\\log$(edu
                /gob)",
974                               "$\\log$(eco/gob)","$\\log$(sal/gob)"
                                  ,
975                               "$\\log$(resto/gob)","ordenada"),
976           add.lines = list(c("F (efectos) p-val",pres %>% fsignif()
                ),
977                            c("Hausman p-value",gcor %>% fsignif()),
978                            c("AR(1) p-value",ar1 %>% fsignif()),
979                            c("AR(2) p-value",ar2 %>% fsignif()))
980 )
981
982
983
984 # # g
985 # g_dat %>%
986 #   group_by(ISO3) %>% mutate(m=median(cre)) %>% ungroup() %>%
987 #   ggplot(aes(x=reorder(ISO3,m),y=cre,label=round(cre,1)))+geom_
      boxplot()+
988 #   theme_bw()+xlab("")+ylab("Crecimiento real PIB per cápita")+
989 #   theme(axis.text.x = element_text(angle=90,vjust=0.5))
990 # ggsave("./TeX/Fig/45_4_g_edos.pdf",width = 12,height = 7,units =
      "cm")
991 #
992 # g_dat %>%
993 #   #group_by(ISO3) %>% summarise(pibpc=mean(pibpc),cre=mean(cre))
      %>%
994 #   #filter(year==2018) %>%
```

```
#    ggplot(aes(x=pibpc/1000,y=cre,label=ISO3))+geom_point()+#geom_
     text_repel()+
#    scale_x_log10()+
#    theme_bw()+xlab("PIB per cápita mdp")+ylab("Crecimiento real
     PIB per cápita")
# ggsave("./TeX/Fig/45_5_g_pibpc.pdf",width = 12,height = 7,units =
     "cm")


# Efectos fijos
temp<-data.frame(ef=fixef(g_int,type="dmean")) %>% rownames_to_
    column("ISO3")
q<-sapply(temp$ef,function(x) sign(x)*min(abs(x),6))
elcolor<-ifelse(abs(temp$ef)>6,"white","black")
temp %>%
  ggplot(aes(x=reorder(ISO3,-ef),y=ef))+
  geom_bar(stat = "identity",position = "dodge",fill=4)+
  geom_text(aes(y=q+2*sign(ef),label=format(ef,digits = 1)),angle
      =90,size=2.5,col=elcolor)+
  theme_bw()+theme(axis.text.x=element_text(angle=90,vjust = 0.5))+
      #coord_flip()+
  coord_cartesian(ylim = c(-10,10))+
  xlab("")+ylab("")
ggsave("./TeX/Fig/45_11_g_ef.pdf",width = 12,height = 7,units = "cm
    " )

# Correlación efectos fijos
fnames=c("soc","edu","eco","sal","resto")
sapply(c(fnames,"gto2","cre"), function (x) cor(fixef(g_int),
    between(g_dat[,x]))) %>%
  t() %>% as.data.frame() %>%
  xtable(caption = "Correlación vs los efectos individuales",label=
      "tab:cor_ef_edo") %>%
  print(includel.rownames=F)

# Autocorrelaciones
temp<-pibent2 %>% mutate(cre=100*(pib/dplyr::lag(pib)-1)) %>%
    filter(!is.na(cre),year<=2019)
temp<-g_int$residuals %>% as.data.frame() %>% rownames_to_column("
    id") %>%
  separate(id,into = c("ISO3","year")) %>% rename(nus=3)

auto<-by(temp$nus, temp$ISO3, function(i) { acf(i,  plot = FALSE)$
    acf })

temp<-matrix(NA,nrow = 32,ncol = 6)
for (i in 1:32) {
  temp[i,1]<-names(auto)[i]
  m<-length(auto[[i]])
  temp[i,2:(m+1)]<-auto[[i]]
}
colnames(temp)<-c("ISO3",0:4)
auto<-temp %>% as.data.frame() %>%
  pivot_longer(cols = `0`:`4`,names_to="lag",values_to="cor") %>%
  na.omit() %>% mutate(cor=as.numeric(cor),lag=as.numeric(lag))

auto %>% filter(lag <=15) %>%
```

```
  ggplot(aes(x=lag,y=cor,group=lag)) +
  geom_boxplot()+theme_bw()+
  ylab("")
ggsave("./TeX/Fig/45_12_acfbox.pdf",width = 10,height = 8,units = "
    cm")
```

```
# sys GMM ####
eqsgmm<-cre ~ lag(cre) + gto2+
  soc+edu+eco+sal+resto |
  lag(cre, 2:99) | # exógenas en ef. ind. e idios.
  lag(gto2,2)+lag(soc, 2)+lag(edu,2)+lag(eco,2)+lag(sal,2)+lag(
      resto,2)

4*5/2 #system
6 # normales
temp<-g_dat %>% filter(!ISO3 %in% c("CAM","TAB")) %>% as.data.frame()
    %>% droplevels() %>% pdata.frame()
class(g_dat)
class(temp)
pdim(temp)
pdim(g_dat)

base <- pgmm(formula=eqsgmm,data=g_dat,
             index=c("ISO3", "year"),model="twosteps",
             effect="twoways",transformation = "ld")
base.ind<-pgmm(formula=eqsgmm,g_dat,
               index=c("ISO3", "year"),model="twosteps",# subset =
                   !ISO3 %in% c("CAM","TAB"),
               effect="individual",transformation = "ld")
# basefil<-pgmm(formula=eqsgmm,data=temp,
#               index=c("ISO3", "year"),model="twosteps",# subset =
    !ISO3 %in% c("CAM","TAB"),
#               effect="twoways",transformation = "ld")
# coeftest(basefil,vcov. = vcovHC)

sbase<-summary(base)
sbasei<-summary(base.ind)
# sbasef<-summary(basefil)

lmodel<-list(sbasei,sbase)
vsarg<-format(sapply(lmodel, function(x) x$sargan$p.value),digits =
    2)
ar1<-format(sapply(lmodel, function(x) x$m1$p.value),digits = 1)
ar2<-format(sapply(lmodel, function(x) x$m2$p.value),digits = 2)
waldmu<-c("",format(sbase$wald.td$p.value[[1]],digits = 1))#,format
    (sbasef$wald.td$p.value[[1]],digits = 1))

stargazer(base.ind,base, type="text",
          digits = 2,
          title = "sys-GMM para el crecimiento económico por
              estados",
          label = "tab:gsGMM_edos",
```

```
1088            column.labels = c("Sin $\\mu$", "Con $\\mu$", "Con $\\mu$
                    ex.pet."),
1089            dep.var.caption = "Variable dependiente: $g_t$ de los
                    estados",
1090            single.row = T,
1091            dep.var.labels.include = F,
1092            covariate.labels=c("$g_{t-1}$","gto","$\\log$(soc/gob)","
                    $\\log$(edu/gob)",
1093                               "$\\log$(eco/gob)","$\\log$(sal/gob)"
                                       ,
1094                               "$\\log$(resto/gob)"),
1095            omit.stat = "n",
1096            add.lines = list(#c("Observaciones",rep("1,296",3)),
1097              c("Estados",c(32,32)),
1098              c("Instrumentos",16,16),
1099              c("Sargan p-value",vsarg %>% fsignif()),
1100              c("AR(1) p-value",ar1 %>% fsignif()),
1101              c("AR(2) p-value",ar2 %>% fsignif()),
1102              c("Wald $\\mu_t$ p-value",waldmu %>% fsignif())
1103            ))
1104
1105
1106
1107
1108 # errores ####
1109
1110 lsfun<-list("media"=mean,#"dev"=sd,
1111             "plo"=function(x,na.rm=T) quantile(x,probs = 0.25,na.rm
                     ),
1112             "pup"=function(x,na.rm=T) quantile(x,probs = 0.75,na.rm
                     ))
1113
1114
1115 modelos<-list("(2) Temp"=base,"(1) Indiv"=base.ind)#,"(3) Temp ex.
        pet."=basefil)
1116
1117 for (i in 1:length(modelos)) {
1118   errores<-modelos[[i]]$residuals %>% as.data.frame() %>%
1119     mutate(year=names(base$residuals[[1]]),
1120            model=c(rep('d',3),rep('l',4))) %>%
1121     pivot_longer(!c(year,model),names_to='code',values_to='resid')
           %>%
1122     mutate(resid=na_if(resid,0)) %>% na.omit()
1123   temp<-errores %>% group_by(year) %>% summarise(across(resid,lsfun
          )) %>% mutate(modelo=names(modelos)[i])
1124   if (i==1) todo=temp else todo=rbind(todo,temp)
1125 }
1126
1127 todo %>% ggplot(aes(x=as.numeric(year),y=resid_media))+
1128   geom_ribbon(aes(ymin=resid_plo,ymax=resid_pup,fill=modelo),alpha
          =0.2)+
1129   geom_line(aes(col=modelo),size=1)+theme_bw()+
1130   xlab("Año")+ylab("")+labs(col="",fill="")+
1131   theme(legend.position = c(.45,.95),
1132         legend.background = element_rect(fill = "transparent"))+
1133   guides(fill=guide_legend(nrow=1))
```

```
ggsave("./TeX/Fig/45_12_errores.pdf",width = 12,height = 5,units =
    "cm",bg="transparent")

# Efectos temporales
base$coefficients[[2]][9:11]
coeftest(base,vcov=vcovHC)
sqrt(diag(vcovHC(base)))
sqrt(diag(vcovHC(base))[9:11])
sqrt(diag(vcov(base))[9:11])
base$coefficients
base$vcov[c("2017","2018","2019"),c("2017","2018","2019")]

# Cuadros efectos sustitución
pibent2[131,'pib']/deflac[5,'factor']/100 #1% del PIB CMX 2019
    precios corrientes en millones

pibent2 %>% filter(year==2019) %>%
  summarise(pib=mean(pib)/deflac[5,"factor"]/100) #1% del PIB del
      edo prom. 2019 precios corrientes en millones

fnames<-c("soc","edu","eco","sal","resto")
gm<-function(iso,año){
  edoagrupado %>% mutate(across(soc:gob,~.*gto2),lg=dplyr::lag(cre)
      ) %>%
    filter(ISO3==iso,year==año) %>%
    select(ISO3,year,lg,gto2,soc:gob,cre) %>%
    unite("id",ISO3:year,sep=".") %>% column_to_rownames("id")
}
logob<-function(x) x %>% mutate(across(soc:resto,~log(./gob))) %>%
  dplyr::select(-gob)
gm("CMX",2019)
base.ind$coefficients[[2]]

cambios<-c(0.1,0.5,1)
# inicial<-gm("CMX",2019)
# rebal<-inicial
# rebal[fnames[1]]<-gm("CMX",2019)[fnames[1]]+cambios[1]
# rebal[fnames[2]]<-gm("CMX",2019)[fnames[2]]-cambios[1]
# inicial
# rebal
# sum(logob(inicial)[1:7]*base.ind$coefficients[[2]])-sum(logob(
    rebal)[1:7]*base.ind$coefficients[[2]])

n<-length(fnames)
reasignaciones<-function(modelo){
  temp<-lapply(cambios, function(x){
    res=matrix(0,n,n)
    for(j in 1:n){
      for(k in 1:n){
        inicial<-gm("MEX",2019)
        rebal<-inicial
        rebal[fnames[j]]<-rebal[fnames[j]]+x
        rebal[fnames[k]]<-rebal[fnames[k]]-x
        res[j,k]<-sum(logob(rebal)[1:7]*modelo$coefficients[[2]])-
            sum(logob(inicial)[1:7]*modelo$coefficients[[2]])
      }
    }
```

```
1185      return(res)
1186    })
1187    matriz<-rbind(temp[[1]],temp[[2]],temp[[3]])
1188    return(matriz)
1189  }
1190
1191  tab<-cbind(reasignaciones(base.ind),reasignaciones(base)) %>% as.
        data.frame()
1192  tab
1193  colnames(tab)<-paste(rep(c("Ind","Temp."),each=n),rep(fnames,2))
1194  tab$fun<-rep(fnames,3)
1195  # tab$delta<-rep(cambios,each=n)
1196
1197  tab<-tab[11:15,] %>% relocate(fun,`Ind soc`)
1198
1199  tab %>%
1200    xtable(caption = "Efectos de la reasignación del gasto",
1201           label="tab:gsGMM_delta_edos",digits=1) %>%
1202    print(include.rownames=F)
1203
1204
1205  # Forecast modelos # Predict 2020 - 2021
1206  logob<-function(x) x %>% mutate(across(soc:resto,~log(./gob))) %>%
1207    dplyr::select(-gob)
1208  g_datp<-edodat %>% filter(!ISO3%in%c("TOT","MED")) %>%#excluir
        total y promedio
1209    merge(gtoconst %>% dplyr::select(ISO3,year,gto,gtoc,factor,f18))
          %>%# agregar gasto en precios constantes
1210    mutate(pib=pib/factor*f18,pibpc=pibpc/factor*f18) %>%# convertir
        PIB a precios constantes 2018
1211    mutate(gto2=gtoc/pib/1e6*100) %>%#gto en porcentaje de PIB
1212    mutate(gtopc=gtoc/pob/1e3) %>%#gasto per cápita
1213    mutate(across(soc:resto,~log(./gob))) %>% dplyr::select(-gob) #
          %>%#logcocientes
1214    #filter(year>=2019)  # proyección '20 y '21
1215
1216
1217  gm<-function(iso,año){
1218    g_datp %>%
1219      group_by(ISO3) %>%
1220      mutate(gob=0,across(c(soc:resto,gob),exp),lg=dplyr::lag(cre))
            %>%
1221      ungroup() %>%
1222      mutate(tot= select(.,c(soc:resto,gob)) %>% rowSums(),
1223             across(c(soc:resto,gob),~./tot)) %>%
1224      mutate(across(c(soc:resto,gob),~.*gto2)) %>%
1225      filter(ISO3==iso,year==año) %>%
1226      select(ISO3,year,lg,gto2,soc:resto,gob,cre) %>%
1227      unite("id",ISO3:year,sep=".") %>% column_to_rownames("id")
1228  }
1229
1230  gm("CMX",2020)
1231
1232  iso="MEX";año=2020
1233
1234  temp<-g_datp %>%#filter(year>2019) %>%
1235    mutate(year=factor(year))
```

```
estima<-rep(0,nrow(temp))
for(i in 1:nrow(temp)){
  estima[i]<-sum(base.ind$coefficients[[2]]*
        logob(gm(temp$ISO3[i],temp$year[i]))[1:7])
}
temp$g_est<-estima


fit<-base.ind$fitted.values %>%
  as.data.frame() %>%
  mutate(year=c(2017:2019,2016:2019),modelo=c(rep("d",3),rep("l",4)
      )) %>%
  pivot_longer(AGU:ZAC,names_to="ISO3",values_to="gfit") %>%
  filter(modelo=="l") %>% arrange(ISO3,year)

res<- base.ind$residuals %>%
  as.data.frame() %>%
  mutate(year=c(2017:2019,2016:2019),modelo=c(rep("d",3),rep("l",4)
      )) %>%
  pivot_longer(AGU:ZAC,names_to="ISO3",values_to="resid") %>%
  filter(modelo=="l") %>% arrange(ISO3,year)

g_dat2<-g_dat %>% merge(temp,all = T) %>%
  dplyr::select(ISO3,year,estado,cre,pibpc,g_est) %>%
  merge(ind.coord %>% select(code,Grupo),by.x = "ISO3",by.y = "code
      ",all.x = T) %>%
  #merge(fit %>% dplyr::select(-modelo),all=T) %>%
  merge(res %>% dplyr::select(-modelo),all=T) %>%
  mutate(year=as.numeric(as.character(year))) %>%
  group_by(ISO3) %>%
  mutate(pibfit=case_when(year<=2019 ~ pibpc,
                          year==2020 ~ dplyr::lag(pibpc)*(1+g_est/
                              100),
                          year==2021 ~ dplyr::lag(pibpc,2)*(1+dplyr
                              ::lag(g_est,1)/100)*(1+g_est/100)))
                              %>%
  mutate(check=100*(pibfit/dplyr::lag(pibfit)-1))

# Gráfica obs vs fitted g

sm.labs <- paste("Grupo",1:6,
                 c('(centro)','(norte)','(golfo)','(petrolero)','(
                     bajo ing.)','(cmdx)'))
names(sm.labs) <- as.character(1:6)

sm.labs
g_dat2 %>% group_by(Grupo,year) %>%
  summarise(se=sd(resid,na.rm = T)/n(),across(c(cre,g_est,resid),
      mean,na.rm=T)) %>%
  mutate(cre=ifelse(year<=2019,cre,NA)) %>%#quitar proyecciones
      iniciales
  ggplot(aes(x=year))+
  #geom_bar(aes(y=cre),stat="identity",position = "dodge")+
  #geom_ribbon(aes(ymin=g_est-se,ymax=g_est+se),fill=4)+
  geom_point(aes(y=cre),col=1)+
  geom_line(aes(y=g_est),col=2)+
```

```
1283    facet_wrap(~Grupo,scales = "free",labeller = labeller(Grupo=sm.
            labs))+
1284    theme_bw()+xlab(NULL)+ylab('Crecimiento económico real %)
1285  ggsave("./TeX/Fig/45_13_g20y21.pdf",width = 15,height = 10,units =
        "cm")
1286
1287
1288  # Gráfica obs vs fitted pibpc
1289  g_dat2 %>% group_by(Grupo,year) %>%
1290    summarise(across(c(cre,g_est,resid,pibpc,pibfit),mean,na.rm=T))
            %>%
1291    mutate(pibpc=ifelse(year<=2019,pibpc,NA)) %>%#quitar
            proyecciones iniciales
1292    ggplot(aes(x=year))+
1293    #geom_bar(aes(y=cre),stat="identity",position = "dodge")+
1294    #geom_ribbon(aes(ymin=g_est-se,ymax=g_est+se),fill=4)+
1295    geom_point(aes(y=pibpc),col=1)+
1296    geom_line(aes(y=pibfit),col=2)+
1297    facet_wrap(~Grupo,scales = "free")+
1298    theme_bw()
1299
1300
1301  # población por ingresos ####
1302
1303  ping<-read_excel("./Data/INEGI/poblacióningresos.xlsx",skip = 4)
            %>%
1304    pivot_longer(cols = !c(Entidad,Periodo),names_to="variable",
            values_to="pob") %>%
1305    filter(grepl("Cuarto",Periodo),!grepl("cv|est",variable)) %>%
1306    pivot_wider(names_from = variable,values_from=pob) %>%
1307    mutate(year=as.numeric(substr(Periodo,22,25))) %>%
1308    mutate(across(Total:s_ne,as.numeric)) %>% arrange(Entidad,year)
1309  cat_ent2<-data.frame(Entidad=sort(unique(ping$Entidad)[-30]), # 30
            es el total
1310                estado=unique(cat_entidad$estado)[-19]) %>%# 19 es No
                    Dististribuible
1311    merge(cat_entidad %>% filter(entfed!="Distrito Federal") %>%
            dplyr::select(estado,ISO3))
1312
1313  # con catalogo ISO3 para estado
1314  ping2<-ping %>% merge(cat_ent2 %>% dplyr::select(Entidad,ISO3)) %>%
1315    mutate(across(s_0a1:s_0,~./Total)) %>%#en porcentaje del total
1316    relocate(ISO3,year) %>% relocate(s_0,.before=s_0a1) %>%
1317    dplyr::select(-Entidad,-Periodo) %>%
1318    arrange(ISO3,year) #%>%
1319    # group_by(ISO3) %>%
1320    # mutate(across(Total:s_0,~100*(./dplyr::lag(.)-1),.names ="g_{.
            col}" ))
1321
1322  pnames <- c('SM=0','SM in (0,1]','SM in (1,2]','SM in (2,3]',
1323               'SM in (3,5]','SM in (5,inf]')
1324  pnames <- c('0','(0,1]','(1,2]','(2,3]','(3,5]',
1325               paste('(5,',expression(infinity),')',sep=''))
1326  pnames <- c('0','De 0 a 1','De 1 a 2','De 2 a 3','De 3 a 5','5+')
1327  onames <- names(ping2)[4:9]
1328  names(ping2)[4:9] <- pnames
1329
```

```
# PEA composición salario detalle total
peacomp <- ping2 %>%
  dplyr::select(-c(Total,s_ne)) %>%
  unite("id",ISO3:year,sep='.') %>% column_to_rownames("id") %>%
  acomp()

# Variation matrix

rbind(mean(peacomp,robust = T),variation(peacomp,robust = T)) %>%
  xtable(caption = 'Media y Varianza Composicional',
         label='tab:pea_medvarcomp',digits=2) %>%
  print(include.rownames=T)

# Dendrograma
pdf('./TeX/Fig/45p_1_pea_dendro.pdf',width = 6,height = 4)
variation(peacomp,robust = T) %>% as.dist() %>% hclust(method = "
    ward.D") %>%
  plot(main=NA,sub=NA,ylab='Distancia')
dev.off()


# categorías reducidas ###
names(ping2)[4:9] <- onames
ging_dat <- ping2 %>% dplyr::select(ISO3,year,Total:s_ne) %>%
  mutate(s_0a3=s_0+s_0a1+s_1a2,s_3a5=s_2a3+s_2a3) %>% dplyr::select
      (-c(s_0a1:s_3a5,s_0)) %>%
  relocate(s_0a2,s_2a5,.before=s_5mas)


# boxplots tasas categorías reducidas
pnames<-c('De 0 a 2 SM','De 2 a 5 SM','Más de 5 SM')
names(pnames) <- c('s_0a2','s_2a5','s_5mas')
ging_dat %>% dplyr::select(ISO3,year,Total:s_5mas) %>%
  group_by(ISO3) %>% mutate(m1=median(s_0a2),m2=median(s_2a5)) %>%
      ungroup() %>%
  pivot_longer(cols = Total:s_5mas,names_to="grupo",values_to="p")
      %>% na.omit() %>%
  filter(grupo!="Total") %>% mutate(p=p*100) %>%
  ggplot(aes(x=reorder(ISO3,-m1),y=p))+geom_boxplot()+
  facet_wrap(~grupo,scales = "free",nrow = 3,labeller = labeller(
      grupo = pnames))+
  theme_bw()+xlab("")+ylab("Población %")+
  theme(axis.text.x = element_text(angle=90,vjust=0.5))
ggsave("./TeX/Fig/45_6_pobingreso.pdf",width = 12,height = 15,units
    = "cm")

# scatter bottom vs gasto
g_dat %>% dplyr::select(ISO3,year,soc:resto,gto2) %>%
  pivot_longer(cols = c(soc:resto,gto2),names_to="explic",values_to
      ="valor") %>%
  merge(ging_dat %>% dplyr::select(ISO3,year,Total:s_5mas) %>%
          mutate(year=factor(year))) %>%
  ggplot(aes(x=valor,y=s_0a2))+
  facet_wrap(~explic,scales = "free")+
  geom_point()+xlab('Gasto público')+
  ylab('% de la población < 2 SM')+
  theme(legend.position = "none")
```

```
ggsave('./TeX/Fig/45p_2_gto_vs_bajo.pdf',height = 10,width = 12,
    units = 'cm')

# scatter logc bottom vs gasto
g_dat %>% dplyr::select(ISO3,year,soc:resto,gto2) %>%
  pivot_longer(cols = c(soc:resto,gto2),names_to="explic",values_to
      ="valor") %>%
  merge(ging_dat %>% dplyr::select(ISO3,year,Total:s_5mas) %>%
          mutate(year=factor(year),s_0a2=log(s_0a2/s_5mas))) %>%
  ggplot(aes(x=valor,y=s_0a2))+
  facet_wrap(~explic,scales = "free")+
  geom_point()+xlab('Gasto público')+
  ylab('% de la población < 2 SM')+
  theme(legend.position = "none")


# Datos
ging_dat <- ging_dat %>% dplyr::select(ISO3,year,Total:s_5mas) %>%
  merge(g_dat %>% dplyr::select(ISO3,year,soc:resto,gto2)) %>%
      pdata.frame()

pdim(ging_dat)
# modelos panel población| ingresos ####
g_mco<-plm(s_0a2~gto2+soc+edu+eco+sal+resto,ging_dat,model = "
    pooling") # ols
g_int<-plm(s_0a2~gto2+soc+edu+eco+sal+resto,ging_dat,model = "
    within", effect = "individual") # within
g_mcg<-plm(s_0a2~gto2+soc+edu+eco+sal+resto,ging_dat,model = "
    random", effect = "individual") # gls


# summary(g_mco)
# summary(g_int)
# summary(g_mcg)


#presencia de efectos
pres<-c(format(pFtest(update(g_int,effect="individual"),g_mco)$p.
    value,digits = 2),"","")
#Hausmant test, correlación de efectos
gcor<-c("","",format(phtest(g_int,g_mcg)$p.value,digits = 2))
#AR(1)
ar1<-sapply(list(g_mco,g_int,g_mcg),function(x) format(pbgtest(x)$p
    .value,digits = 2))
#AR(2)
ar2<-sapply(list(g_mco,g_int,g_mcg),function(x) format(pbgtest(x,
    order=2)$p.value,digits = 2))


stargazer(g_mco,g_int,g_mcg,digits = 2, type = "text",
          title = "Regresión de panel para población de ingreso
              bajo",
          label = "tab:pmod0_edos",
          column.labels = c("$\\hat{\\und\\gamma}_{\\text{MCO}}$",
                            "$\\hat{\\und\\gamma}_{\\text{W}}$",
                            "$\\hat{\\und\\gamma}_{\\text{MCG}}$"),
          single.row = T,
```

```
1427            dep.var.caption = "Variable dependiente: Población|
                    Ingreso < 2SM",
1428            omit.stat = "F",
1429            dep.var.labels.include = F,
1430            # covariate.labels=c("gto","$\\log$(soc/gob)","$\\log$(
                    edu/gob)",
1431            #                       "$\\log$(eco/gob)","$\\log$(sal/gob)
                    ",
1432            #                       "$\\log$(resto/gob)","ordenada"),
1433            add.lines = list(c("F (efectos) p-val",pres %% fsignif()
                    ),
1434                             c("Hausman p-value",gcor %% fsignif()),
1435                             c("AR(1) p-value",ar1 %% fsignif()),
1436                             c("AR(2) p-value",ar2 %% fsignif()))
1437 )
1438
1439 # sys GMM población| ingresos ####
1440
1441 eqsgmm<-s_0a3 ~ lag(s_0a3) + gto2+
1442   soc+edu+eco+sal+resto |
1443   lag(s_0a3, 2:99) | # exógenas en ef. ind. e idios.
1444   lag(gto2,2)+lag(soc, 2)+lag(edu,2)+lag(eco,2)+lag(sal,2)+lag(
          resto,2)
1445
1446 # 4*5/2 #system
1447 # 6 # normales
1448
1449
1450 base <- pgmm(formula=eqsgmm,data=ging_dat,
1451               index=c("ISO3", "year"),model="twosteps",# subset = !
                      ISO3 %in% c("CAM","TAB"),
1452               effect="twoways",transformation = "ld")
1453 base.ind<-pgmm(formula=eqsgmm,ging_dat,
1454                 index=c("ISO3", "year"),model="twosteps",# subset =
                        !ISO3 %in% c("CAM","TAB"),
1455                 effect="individual",transformation = "ld")
1456
1457 # coeftest(base,vcov. = vcovHC)
1458 # sbase<-summary(base)
1459 # sbasei<-summary(base.ind)
1460
1461
1462 lmodel<-list(sbasei,sbase)
1463 vsarg<-format(sapply(lmodel, function(x) x$sargan$p.value),digits =
        2)
1464 ar1<-format(sapply(lmodel, function(x) x$m1$p.value),digits = 1)
1465 ar2<-format(sapply(lmodel, function(x) x$m2$p.value),digits = 2)
1466 waldmu<-c("",format(sbase$wald.td$p.value[[1]],digits = 1))#,format
      (sbasef$wald.td$p.value[[1]],digits = 1))
1467
1468 stargazer(base.ind,base,# type="text",
1469           digits = 2,
1470           title = "sys-GMM para para población de ingreso bajo",
1471           label = "tab:psGMM_edos",
1472           column.labels = c("Sin $\\mu$", "Con $\\mu$", "Con $\\mu$
                  ex.pet."),
```

```
1473            dep.var.caption = "Variable dependiente: Población|
                    Ingreso < 3SM",
1474            single.row = T,
1475            dep.var.labels.include = F,
1476            # covariate.labels=c("$g_{t-1}$","gto","$\\log$(soc/gob)
                    ","$\\log$(edu/gob)",
1477            #                   "$\\log$(eco/gob)","$\\log$(sal/gob)
                    ",
1478            #                   "$\\log$(resto/gob)"),
1479            omit.stat = "n",
1480            add.lines = list(#c("Observaciones",rep("1,296",3)),
1481              c("Estados",c(32,32)),
1482              c("Instrumentos",16,16),
1483              c("Sargan p-value",vsarg %>% fsignif()),
1484              c("AR(1) p-value",ar1 %>% fsignif()),
1485              c("AR(2) p-value",ar2 %>% fsignif()),
1486              c("Wald $\\mu_t$ p-value",waldmu %>% fsignif())
1487            ))
1488
1489
1490 # categorías reducidas ####
1491 ging_dat <- ping2 %>% dplyr::select(ISO3,year,Total:s_ne) %>%
1492   mutate(s_0a3=s_0+s_0a1+s_1a2+s_2a3) %>% dplyr::select(-c(s_0a1:s_
          2a3,s_0)) %>%
1493   relocate(s_0a3,.before=s_3a5) %>% select(-s_ne)
1494
1495 ging_dat %>% head()
1496 # boxplots tasas categorías reducidas
1497 pnames<-c('De 0 a 3 SM','De 3 a 5 SM','Más de 5 SM')
1498 #pnames<-c(expression('p'[3]),expression('p'[5]),expression('p
        '[5+]))
1499 names(pnames) <- c('s_0a3','s_3a5','s_5mas')
1500 ging_dat %>% dplyr::select(ISO3,year,Total:s_5mas) %>%
1501   group_by(ISO3) %>% mutate(m1=median(s_0a3)) %>% ungroup() %>%
1502   pivot_longer(cols = Total:s_5mas,names_to="grupo",values_to="p")
          %>% na.omit() %>%
1503   filter(grupo!="Total") %>% mutate(p=p*100) %>%
1504   ggplot(aes(x=reorder(ISO3,-m1),y=p))+geom_boxplot()+
1505   facet_wrap(~grupo,scales = "free",nrow = 3,labeller = labeller(
          grupo = pnames))+
1506   theme_bw()+xlab("")+ylab("Población %")+
1507   theme(axis.text.x = element_text(angle=90,vjust=0.5))
1508 ggsave("./TeX/Fig/45_6_pobingreso.pdf",width = 12,height = 15,units
        = "cm")
1509
1510 # scatter bottom vs gasto
1511 g_dat %>% dplyr::select(ISO3,year,soc:resto,gto2) %>%
1512   pivot_longer(cols = c(soc:resto,gto2),names_to="explic",values_to
          ="valor") %>%
1513   merge(ging_dat %>% dplyr::select(ISO3,year,Total:s_5mas) %>%
1514           mutate(year=factor(year))) %>%
1515   ggplot(aes(x=valor,y=s_0a3))+
1516   facet_wrap(~explic,scales = "free")+
1517   geom_point()+xlab('Gasto público')+
1518   ylab('% de la población < 3 SM')+
1519   theme(legend.position = "none")
```

```
#ggsave('./TeX/Fig/45p_2_gto_vs_bajo.pdf',height = 10,width = 12,
    units = 'cm')

# scatter logc bottom vs gasto
g_dat %>% dplyr::select(ISO3,year,soc:resto,gto2) %>%
  pivot_longer(cols = c(soc:resto,gto2),names_to="explic",values_to
      ="valor") %>%
  merge(ging_dat %>% dplyr::select(ISO3,year,Total:s_5mas) %>%
          mutate(year=factor(year),s_0a3=log(s_0a3/s_5mas))) %>%
  ggplot(aes(x=valor,y=s_0a3))+
  facet_wrap(~explic,scales = "free")+
  geom_point()+xlab('Gasto público')+
  ylab('% de la población < 3 SM')+
  theme(legend.position = "none")


# Datos
ging_dat <- ging_dat %>% dplyr::select(ISO3,year,Total:s_5mas) %>%
  merge(g_dat %>% dplyr::select(ISO3,year,soc:resto,gto2)) %>%
      pdata.frame()

pdim(ging_dat)
# modelos panel población| ingresos ####
g_mco<-plm(s_0a3~gto2+soc+edu+eco+sal+resto,ging_dat,model = "
    pooling") # ols
g_int<-plm(s_0a3~gto2+soc+edu+eco+sal+resto,ging_dat,model = "
    within", effect = "individual") # within
g_mcg<-plm(s_0a3~gto2+soc+edu+eco+sal+resto,ging_dat,model = "
    random", effect = "individual") # gls


# summary(g_mco)
# summary(g_int)
# summary(g_mcg)


#presencia de efectos
pres<-c(format(pFtest(update(g_int,effect="individual"),g_mco)$p.
    value,digits = 2),"","")
#Hausmant test, correlación de efectos
gcor<-c("","",format(phtest(g_int,g_mcg)$p.value,digits = 2))
#AR(1)
ar1<-sapply(list(g_mco,g_int,g_mcg),function(x) format(pbgtest(x)$p
    .value,digits = 2))
#AR(2)
ar2<-sapply(list(g_mco,g_int,g_mcg),function(x) format(pbgtest(x,
    order=2)$p.value,digits = 2))


stargazer(g_mco,g_int,g_mcg,digits = 2, type = "text",
          title = "Regresión de panel para población de ingreso
              bajo",
          label = "tab:pmod0_edos",
          column.labels = c("$\\hat{\\und\\gamma}_{\\text{MCO}}$",
                            "$\\hat{\\und\\gamma}_{\\text{W}}$",
                            "$\\hat{\\und\\gamma}_{\\text{MCG}}$"),
          single.row = T,
```

```
1567            dep.var.caption = "Variable dependiente: Población|
                    Ingreso < 2SM",
1568            omit.stat = "F",
1569            dep.var.labels.include = F,
1570            # covariate.labels=c("gto","$\\log$(soc/gob)","$\\log$(
                    edu/gob)",
1571            #                   "$\\log$(eco/gob)","$\\log$(sal/gob)
                    ",
1572            #                   "$\\log$(resto/gob)","ordenada"),
1573            add.lines = list(c("F (efectos) p-val",pres %% fsignif()
                    ),
1574                             c("Hausman p-value",gcor %% fsignif()),
1575                             c("AR(1) p-value",ar1 %% fsignif()),
1576                             c("AR(2) p-value",ar2 %% fsignif()))
1577 )
1578
1579 # Graficas pendientes ####
1580 # Ternarias
1581 # install.packages('crop')
1582 # library(crop)
1583
1584 pnames <- c(expression('p'[3]),expression('p'[5]),expression('p'['
        5+']))
1585 par(mar=rep(0,4))
1586 pdf('./TeX/Fig/45p_3_pea_tern.pdf',width = 4,height = 4)
1587 ging_dat %% select(s_0a3:s_5mas) %% acomp() %% plot(axes=T,
        labels=pnames)
1588 dev.off()
1589
1590 pdf('./TeX/Fig/45p_3b_pea_tern_centro.pdf',width = 4,height = 4)
1591 ging_dat %% select(s_0a3:s_5mas) %% acomp() %% plot(center=T,
        axes = T,labels=pnames)
1592 dev.off()
1593
1594 # Zoom ultimo año
1595 ping2 %% dplyr::select(-c(Total,s_ne)) %%
1596   filter(year==2018|year==2019) %%
1597   group_by(ISO3) %%
1598   mutate(across(s_0:s_5mas,~.-dplyr::lag(.))) %% ungroup() %%
1599   na.omit() %%
1600   pivot_longer(cols = s_0:s_5mas,names_to='salario',values_to='
          porcentaje') %%
1601   ggplot(aes(x=ISO3,y=porcentaje,col=salario,group=salario))+geom_
          line(size=1)+
1602     theme_bw()+theme(axis.text.x = element_text(angle = 90))+
1603     xlab(NULL)+ylab('Cambio 2018 a 2019')+scale_color_locuszoom()
1604
1605 pnames <- c('Hasta 1','Hasta 2','Hasta 3','Hasta 5','Más de 5')
1606 ping2 %% dplyr::select(-c(Total,s_ne)) %%
1607   filter(year %in% c(2016:2019)) %%
1608   group_by(ISO3) %%
1609   mutate(across(s_0:s_5mas,~.-dplyr::lag(.))) %%
1610   na.omit() %% ungroup() %%
1611   group_by(year) %% summarise(across(s_0:s_5mas,mean)) %%
1612   pivot_longer(cols = s_0:s_5mas,names_to='salario',values_to='
          porcentaje') %%
1613   filter(salario!='s_0') %%
```

```
1614    mutate(salario=rep(pnames,3)) %>%
1615    ggplot(aes(x=salario,y=porcentaje,fill=factor(year)))+
1616    geom_bar(stat='identity',position='dodge')+
1617    theme_bw()+xlab('Ingreso en SM')+ylab('Cambio en la composición')
            +
1618    labs(fill='Año')
1619  ggsave('./TeX/Fig/45p_4_cambiopea.pdf',height = 7,width = 12,units
        = 'cm')
1620
1621
1622  ging_dat %>% filter(year==2018|year==2019) %>%
1623    ggplot(aes(x=ISO3,y=s_0a3,col=year))+
1624    geom_point()+
1625    theme_bw()+theme(axis.text.x = element_text(angle = 90))+
1626    xlab(NULL)+ylab('Población con Ingreso < 3SM')
1627
1628  # Efectos fijos
1629  temp<-data.frame(ef=fixef(g_int,type="dmean")*100) %>% rownames_to_
        column("ISO3")
1630  q<-sapply(temp$ef,function(x) sign(x)*min(abs(x),13))
1631  elcolor<-ifelse(abs(temp$ef)>15,"white","black")
1632  temp %>%
1633    ggplot(aes(x=reorder(ISO3,-ef),y=ef))+
1634    geom_bar(stat = "identity",position = "dodge",fill=4)+
1635    geom_text(aes(y=q+sign(ef)*2.5,label=format(ef,digits = 1)),angle
            =90,size=2.5,col=elcolor)+
1636    theme_bw()+theme(axis.text.x=element_text(angle=90,vjust = 0.5))+
          #coord_flip()+
1637    coord_cartesian(ylim = c(-20,20))+
1638    xlab("")+ylab("")
1639  ggsave("./TeX/Fig/45p_5_p_ef.pdf",width = 12,height = 7,units = "cm
        " )
1640
1641  # Correlacion con efectos
1642  fnames=c("soc","edu","eco","sal","resto")
1643  sapply(c(fnames,"gto2","cre"), function (x) cor(fixef(g_int),
        between(g_dat[,x]))) %>%
1644    t() %>% as.data.frame() %>%
1645    xtable(caption = "Correlación vs los efectos individuales",
1646           label="tab:cor_p_ef_edo") %>%
1647    print(includel.rownames=F)
1648  # AR residuales
1649
1650  #temp<-pibent2 %>% mutate(cre=100*(pib/dplyr::lag(pib)-1)) %>%
        filter(!is.na(cre),year<=2019)
1651  temp<-g_mcg$residuals %>% as.data.frame() %>% rownames_to_column("
        id") %>%
1652    separate(id,into = c("ISO3","year")) %>% rename(nus=3)
1653
1654  auto<-by(temp$nus, temp$ISO3, function(i) { acf(i,  plot = FALSE)$
        acf })
1655
1656  temp<-matrix(NA,nrow = 32,ncol = 6)
1657  for (i in 1:32) {
1658    temp[i,1]<-names(auto)[i]
1659    m<-length(auto[[i]])
1660    temp[i,2:(m+1)]<-auto[[i]]
```

```
}
colnames(temp)<-c("ISO3",0:4)
auto<-temp %>% as.data.frame() %>%
  pivot_longer(cols = '0':'4',names_to="lag",values_to="cor") %>%
  na.omit() %>% mutate(cor=as.numeric(cor),lag=as.numeric(lag))

auto %>% filter(lag <=15) %>%
  ggplot(aes(x=lag,y=cor,group=lag)) +
  geom_boxplot()+theme_bw()+
  ylab("")
ggsave("./TeX/Fig/45p_6_acfbox.pdf",width = 10,height = 8,units = "
    cm")



# errores ####

lsfun<-list("media"=mean,#"dev"=sd,
            "plo"=function(x,na.rm=T) quantile(x,probs = 0.25,na.rm
                ),
            "pup"=function(x,na.rm=T) quantile(x,probs = 0.75,na.rm
                ))


modelos<-list("(1) MCO"=g_mco,"(2) E. fijos"=g_int,"(3) E.
    aleatorios"= g_mcg)

for (i in 1:length(modelos)) {
  errores<-modelos[[i]]$residuals %>% as.data.frame() %>% rename('
      resid'=1) %>%
    #mutate(year=names(g_mco$residuals))
    rownames_to_column('id') %>% separate('id',into = c('ISO3','
        year'))
    #mutate(resid=na_if(resid,0)) %>% na.omit()
  temp<-errores %>% group_by(year) %>% summarise(across(resid,lsfun
      )) %>%
    mutate(modelo=names(modelos)[i])
  if (i==1) todo=temp else todo=rbind(todo,temp)
}

todo %>% ggplot(aes(x=as.numeric(year),y=resid_media))+
  geom_ribbon(aes(ymin=resid_plo,ymax=resid_pup,fill=modelo),alpha
      =0.2)+
  geom_line(aes(col=modelo),size=1)+theme_bw()+
  xlab("Año")+ylab("")+labs(col="",fill="")+
  theme(legend.position = c(.45,.95),
        legend.background = element_rect(fill = "transparent"))+
  guides(fill=guide_legend(nrow=1))
ggsave("./TeX/Fig/45p_7_errores.pdf",width = 12,height = 5,units =
    "cm",bg="transparent")
rm(temp,todo,modelos,lsfun)

# Cuadros efectos sustitución


fnames<-c("soc","edu","eco","sal","resto")
gm<-function(iso,año){
```

```
  edoagrupado %>% mutate(across(soc:gob,~.*gto2),lg=1) %>%
    merge(ging_dat %>% dplyr::select(ISO3,year,s_0a3)) %>%
    mutate(re=rep(ranef(g_mcg),each=5)) %>%
    filter(ISO3==iso,year==año) %>%
    select(ISO3,year,lg,gto2,soc:gob,s_0a3,re) %>%
    unite("id",ISO3:year,sep="-") %>% column_to_rownames("id")
}
logob<-function(x) x %>% mutate(across(soc:resto,~log(./gob))) %>%
  dplyr::select(-gob)
gm("CMX",2019) %>% logob
g_mcg$coefficients

cambios<-c(0.1,0.5,1)


n<-length(fnames)
reasignaciones<-function(modelo){
  temp<-lapply(cambios, function(x){
    res=matrix(0,n,n)
    for(j in 1:n){
      for(k in 1:n){
        inicial<-gm("MEX",2019)
        rebal<-inicial
        rebal[fnames[j]]<-rebal[fnames[j]]+x
        rebal[fnames[k]]<-rebal[fnames[k]]-x
        res[j,k]<-predict(modelo,logob(rebal))-predict(modelo,logob
            (inicial))
      }
    }
    return(res)
  })
  matriz<-rbind(temp[[1]],temp[[2]],temp[[3]])
  return(matriz)
}

tab<-cbind(reasignaciones(g_mcg)) %>% as.data.frame()
tab
colnames(tab)<-fnames
tab$fun<-rep(fnames,3)
# tab$delta<-rep(cambios,each=n)

tab<-tab[11:15,] %>% relocate(fun,.before=1) %>% mutate(across(soc:
    resto,~.*100))
tab
tab %>%
  xtable(caption = "Efectos de la reasignación del gasto en la PEA
      con ingreso < 3SM",
         label="tab:mcg_pdelta_edos",digits=1) %>%
  print(include.rownames=F)


# Forecast modelos # Predict 2020 - 2021
g_datp<-edodat %>% filter(!ISO3%in%c("TOT","MED")) %>%#excluir
    total y promedio
  merge(gtoconst %>% dplyr::select(ISO3,year,gto,gtoc,factor,f18))
      %>%# agregar gasto en precios constantes
```

```
1760    mutate(pib=pib/factor*f18,pibpc=pibpc/factor*f18) %>%# convertir
            PIB a precios constantes 2018
1761    mutate(gto2=gtoc/pib/1e6*100) %>%#gto en porcentaje de PIB
1762    mutate(gtopc=gtoc/pob/1e3) %>%#gasto per cápita
1763    mutate(across(soc:resto,~log(./gob))) %>% dplyr::select(-gob) %>%
            #logcocientes
1764    merge(ging_dat %>% dplyr::select(ISO3,year,s_0a3),all.x = T)
1765 #filter(year>=2019)  # proyección '20 y '21
1766
1767 g_datp %>% head()
1768 gm<-function(iso,año){
1769    g_datp %>%
1770      group_by(ISO3) %>%
1771      mutate(gob=0,across(c(soc:resto,gob),exp),lg=1) %>%
1772      ungroup() %>%
1773      mutate(tot= select(.,c(soc:resto,gob)) %>% rowSums(),# total
1774             across(c(soc:resto,gob),~./tot)) %>%# cerradura
                   composicional
1775      mutate(across(c(soc:resto,gob),~.*gto2)) %>%# base porc de PIB
1776      filter(ISO3==iso,year==año) %>%
1777      select(ISO3,year,lg,gto2,soc:resto,gob,s_0a3) %>%
1778      unite("id",ISO3:year,sep=".") %>% column_to_rownames("id")
1779 }
1780
1781 gm("CMX",2020)
1782
1783 iso="CMX";año=2020
1784
1785 temp<-g_datp %>%#filter(year>2019) %>%
1786    mutate(year=factor(year))
1787 estima<-rep(0,nrow(temp))
1788 for(i in 1:nrow(temp)){
1789    estima[i]<-sum(base.ind$coefficients[[2]]*
1790                     logob(gm(temp$ISO3[i],temp$year[i]))[1:7])
1791 }
1792 temp$g_est<-estima
1793
1794
1795 res <- g_mcg$residuals %>% as.data.frame() %>% rownames_to_column('
        id') %>%
1796    separate(id,into = c('ISO3','year')) %>% rename(resid=3)
1797
1798 g_dat2<-g_datp %>%dplyr::select(-c(pib:f18)) %>%
1799    mutate(s_0a3est=predict(g_mcg,g_datp) + rep(ranef(g_mcg),each=7))
           %>%
1800    dplyr::select(ISO3,year,estado,s_0a3,s_0a3est) %>%
1801    merge(ind.coord %>% select(code,Grupo),by.x = "ISO3",by.y = "code
           ",all.x = T) %>%
1802    #merge(fit %>% dplyr::select(-modelo),all=T) %>%
1803    merge(res ,all=T) %>%
1804    mutate(year=as.numeric(as.character(year))) #%>%
1805    # group_by(ISO3) %>%
1806    # mutate(pibfit=case_when(year<=2019 ~ pibpc,
1807    #                         year==2020 ~ dplyr::lag(pibpc)*(1+g_est
           /100),
1808    #                         year==2021 ~ dplyr::lag(pibpc,2)*(1+
           dplyr::lag(g_est,1)/100)*(1+g_est/100))) %>%
```

```
  # mutate(check=100*(pibfit/dplyr::lag(pibfit)-1))



# Gráfica obs vs fitted g
sm.labs <- paste("Grupo",1:6)
names(sm.labs) <- as.character(1:6)

sm.labs
g_dat2 %>% group_by(Grupo,year) %>%
  summarise(se=sd(resid,na.rm = T)/n(),
            across(c(s_0a3,s_0a3est,resid),~mean(.)*100,na.rm=T)) %>%
  ggplot(aes(x=year))+
  #geom_bar(aes(y=cre),stat="identity",position = "dodge")+
  #geom_ribbon(aes(ymin=g_est-se,ymax=g_est+se),fill=4)+
  geom_point(aes(y=s_0a3),col=1)+
  geom_line(aes(y=s_0a3est),col=2)+
  facet_wrap(~Grupo,scales = "free",labeller = labeller(Grupo=sm.labs))+
  theme_bw()
ggsave("./TeX/Fig/45p_8_p20y21.pdf",width = 15,height = 10,units = "cm")

# Mapas
rm(temp)
temp <- ind.coord %>% group_by(Grupo,code) %>% summarise(n=n())
paste(temp$code[temp$Grupo==1],collapse = T)

library(mxmaps)
temp<-df_mxstate[,c('region','state_name')]
cat_entidad<-cat_entidad %>% merge(temp,all.x = T,by.x = 'estado',by.y = 'state_name')

edomap<-edodat %>%
  merge(cat_entidad %>% dplyr::select(ISO3,region) %>% unique,all.x = T) %>%
  relocate(ISO3,region) %>%
  merge(ind.coord %>% dplyr::select(code,Grupo),by.x = 'ISO3',by.y = 'code')

edo %>% filter(year==2015,!is.na(region)) %>%
  #mutate(value=round(pob/1e6,1)) %>%
  #mutate(value=round(pib/1e6,1)) %>%
  #mutate(value=round(eco*100,1)) %>%
  #mutate(value=round(cre,1)) %>%
  mutate(value=round(pibpc,1)) %>%
  mxstate_choropleth()
rm(mxstate.map)


edomap<-ind.coord %>%
  merge(cat_entidad %>%
          dplyr::select(ISO3,region) %>%
          unique,all.x = T,by.x = 'code',by.y = 'ISO3') %>%
  rename(ISO3=code) %>%
  relocate(ISO3,region)
```

```
1860
1861 edomap %>% mutate(value=Grupo) %>%
1862   mxstate_choropleth(legend = "Grupo") + scale_fill_npg()
1863 ggsave('./TeX/Fig/45p_9_gtogrupos.pdf',height = 7,width = 12,units
       = 'cm')
1864
1865 edomap %>% filter(Grupo==5)
1866
1867 edodat %>% filter(!ISO3%in%c('MED','TOT')) %>%
1868   #mutate(across(soc:gob,~.*gto2)) %>%
1869   dplyr::select(ISO3,year,soc:gob) %>%
1870   filter(year%in%c('2019','2020','2021')) %>%
1871   group_by(ISO3) %>%
1872   summarise(across(soc:gob,~100*(.-dplyr::lag(.)))) %>%
1873   na.omit() %>% ungroup() %>%
1874   mutate(year=rep(2020:2021,32)) %>%
1875   pivot_longer(cols = soc:gob,names_to='funcion',values_to='delta19
         ') %>%
1876   ggplot(aes(x=funcion,y=delta19))+geom_boxplot()+
1877   theme_bw()+xlab(NULL)+ylab('Cambio en %del gasto total')+
1878   #coord_cartesian(ylim=c(-5,5))+
1879   facet_wrap(~year)
1880 ggsave('./TeX/Fig/45p_PEF2021.pdf',width = 12,height = 7,units = '
       cm')
1881
1882 temp <- edodat %>% filter(!ISO3%in%c('MED','TOT')) %>%
1883   #mutate(across(soc:gob,~.*gto2)) %>%
1884   dplyr::select(ISO3,year,soc:gob) %>%
1885   filter(year%in%c('2019','2020','2021')) %>%
1886   group_by(ISO3) %>%
1887   summarise(across(soc:gob,~100*(.-dplyr::lag(.)))) %>%
1888   na.omit() %>% ungroup() %>%
1889   mutate(year=rep(2020:2021,32))
1890
1891 quantile(temp$sal[temp$year==2020],.25)
1892 # No usado
1893 # sys GMM población| ingresos
1894
1895 eqsgmm<-s_0a3 ~ lag(s_0a3) + gto2+
1896   soc+edu+eco+sal+resto |
1897   lag(s_0a3, 2:99) | # exógenas en ef. ind. e idios.
1898   lag(gto2,2)+lag(soc, 2)+lag(edu,2)+lag(eco,2)+lag(sal,2)+lag(
         resto,2)
1899
1900 # 4*5/2 #system
1901 # 6 # normales
1902
1903
1904 base <- pgmm(formula=eqsgmm,data=ging_dat,
1905               index=c("ISO3", "year"),model="twosteps",# subset = !
                      ISO3%in%c("CAM","TAB"),
1906               effect="twoways",transformation = "ld")
1907 base.ind<-pgmm(formula=eqsgmm,ging_dat,
1908                index=c("ISO3", "year"),model="twosteps",# subset =
                       !ISO3%in%c("CAM","TAB"),
1909                effect="individual",transformation = "ld")
1910
```

```
# coeftest(base,vcov. = vcovHC)
# sbase<-summary(base)
# sbasei<-summary(base.ind)


lmodel<-list(sbasei,sbase)
vsarg<-format(sapply(lmodel, function(x) x$sargan$p.value),digits =
    2)
ar1<-format(sapply(lmodel, function(x) x$m1$p.value),digits = 1)
ar2<-format(sapply(lmodel, function(x) x$m2$p.value),digits = 2)
waldmu<-c("",format(sbase$wald.td$p.value[[1]],digits = 1))#,format
    (sbasef$wald.td$p.value[[1]],digits = 1))

stargazer(base.ind,base,# type="text",
          digits = 2,
          title = "sys-GMM para para población de ingreso bajo",
          label = "tab:psGMM_edos",
          column.labels = c("Sin $\\mu$", "Con $\\mu$
              ex.pet."),
          dep.var.caption = "Variable dependiente: Población|
              Ingreso < 3SM",
          single.row = T,
          dep.var.labels.include = F,
          # covariate.labels=c("$g_{t-1}$","gto","$\\log$(soc/gob)
              ","$\\log$(edu/gob)",
          #                    "$\\log$(eco/gob)","$\\log$(sal/gob)
              ",
          #                    "$\\log$(resto/gob)"),
          omit.stat = "n",
          add.lines = list(#c("Observaciones",rep("1,296",3)),
            c("Estados",c(32,32)),
            c("Instrumentos",16,16),
            c("Sargan p-value",vsarg %>% fsignif()),
            c("AR(1) p-value",ar1 %>% fsignif()),
            c("AR(2) p-value",ar2 %>% fsignif()),
            c("Wald $\\mu_t$ p-value",waldmu %>% fsignif())
          ))
```

../Code/7_estados.R

A.3. Miscelánea

Clasificación de países del Banco Mundial.

```
# Obtiene la clasificación de países por nivel de Ingreso según el
    Banco Mundial
# consolida la clasificación con variables del Banco Mundial:
#     gdp PPP, gdp pc PPP, pob, densidad pob, recursos, #negocios/
    1000hab, income share.
# lee: Country_classif, WDIData.csv
# crea: wbclasif.csv
#

```

```
# Paquetes
library(tidyverse)
library(countrycode) # para convertir nombres de paises a códigos
suppressMessages(library(magrittr)) #para operadores pipeline como %<>%
suppressMessages(library(readxl)) # para leer archivos xls
library(xtable) #para imprimir tablas en formato latex

# Directorio

setwd("C:/Users/sgome/Dropbox/#tesis")

# Lectura archivos

#Clasificación
clasif<-read_excel(path = "./Data/World Bank/Country_classif.xls",
                   sheet="Country Analytical History",
                   range = "A6:AI229",
                   na="..", .name_repair="unique")

names(clasif)[1:2]=c("code","country")
clasif17<- clasif %>% filter(!is.na(code)) %>% select(code, country,33) %>% rename(clas="2017")

#Population
wbdata<-read.csv(file="./Data/World Bank/WDIData.csv", encoding = "UTF-8")

#wbdata %>% group_by(Indicator.Name) %>% summarise(n=n_distinct(Country.Code))
pobla<- wbdata %>% filter(Indicator.Name=="Population, total") %>%
  select(Country.Code,c(5:64)) %>%
  pivot_longer(cols = starts_with("X"), names_to = "year",values_to = "pob") %>%
  mutate(year=substr(year,2,5)) %>%
  rename(code=Country.Code)

#gdp per cápita GDP per capita, PPP (current international $)
dens<-wbdata %>% filter(Indicator.Name=="Population density (people per sq. km of land area)") %>%
  select(Country.Code,c(5:64)) %>%
  pivot_longer(cols = starts_with("X"), names_to = "year",values_to = "surf") %>%
  mutate(year=substr(year,2,5)) %>%
  rename(code=Country.Code)

reso<-wbdata %>% filter(Indicator.Name=="Total natural resources rents (% of GDP)") %>%
  select(Country.Code,c(5:64)) %>%
  pivot_longer(cols = starts_with("X"), names_to = "year",values_to = "reso") %>%
  mutate(year=substr(year,2,5)) %>%
  rename(code=Country.Code)

gdp<-wbdata %>% filter(Indicator.Name=="GDP, PPP (constant 2011 international $)") %>%
  select(Country.Code,c(5:64)) %>%
```

```
55    pivot_longer(cols = starts_with("X"), names_to = "year",values_to
          = "gdp") %>%
56    mutate(year=substr(year,2,5))%>%
57    rename(code=Country.Code)
58
59  gdpperc<-wbdata %>% filter(Indicator.Name=="GDP per capita, PPP (
        constant 2011 international $)") %>%
60    select(Country.Code,c(5:64)) %>%
61    pivot_longer(cols = starts_with("X"), names_to = "year",values_to
          = "gdppc") %>%
62    mutate(year=substr(year,2,5))%>%
63    rename(code=Country.Code)
64
65  bussi<-wbdata %>% filter(Indicator.Name=="New business density (new
        registrations per 1,000 people ages 15-64)") %>%
66    select(Country.Code,c(5:64)) %>%
67    pivot_longer(cols = starts_with("X"), names_to = "year",values_to
          = "bussi") %>%
68    mutate(year=substr(year,2,5))%>%
69    rename(code=Country.Code)
70
71  ishare<-wbdata %>% filter(str_detect(Indicator.Name,"Income share")
        ) %>%
72    select(Country.Code,Indicator.Name,c(5:64)) %>%
73    pivot_longer(cols = starts_with("X"), names_to = "year",values_to
          = "ishare") %>%
74    mutate(year=substr(year,2,5))%>%
75    rename(code=Country.Code) %>%
76    pivot_wider(names_from = Indicator.Name, values_from = ishare)
77  names(ishare)[3:9]<-c("4th20","top10","top20","bot10","bot20","2
        nd20","3rd20")
78
79  clasif17 %>% group_by(clas)%>%summarise(n=n())
80  # Otras a agregar Income share held by , New business, Surface area
        (sq. km), Total natural resources rents (% of GDP)
81
82
83
84  # Guardar base para cruzar con otros ejercicios
85  # todos los años, datos de población, clasificación y gdppcPPP
86
87  lclasif<- clasif %>% filter(!is.na(code)) %>%
88    pivot_longer(cols = 3:35, names_to = "year", values_to = "clas")
89
90
91  wbclasif<- merge(x=lclasif,y=pobla) %>% merge(y=gdpperc) %>%
92    merge(y=gdp) %>% merge(y=dens) %>% merge(reso) %>% merge(bussi)
          %>% merge(ishare)
93  tradclase<- data.frame(clas=c("L","LM","UM","H"),Nivel=c("Bajo","
        Medio-bajo","Medio-alto","Alto")) #Traducción categorías
94  wbclasif<-merge(wbclasif,tradclase)
95  write.csv(x=wbclasif,file = "./Data/wbclasif.csv",row.names = FALSE
        )
96
97  # Resumen categorías, ingreso y población
98  # filtrar los datos de 2017
99  pobla %<>% filter(year==2017) %>% select(-year)
```

```
gdpperc %<>% filter(year==2017)%>% select(-year)

clasif2<- merge(x=clasif17,pobla) %>% merge(y=gdpperc)

clasif2 %>% filter(is.na(gdppc))
clasif2 %>% mutate(nodato=is.na(gdppc)+is.na(pob)) %>%
  group_by(nodato,clas) %>%
  summarise(n=n_distinct(code),totpob=sum(pob)/1e9,medgdppc=median(
      gdppc)/1e3,pgdppc=mean(gdppc)/1e3)
rclasif<-clasif2 %>% na.omit() %>%
  group_by(clas) %>%
  summarise(n=n_distinct(code),totpob=sum(pob)/1e9,medgdppc=median(
      gdppc)/1e3,pgdppc=mean(gdppc)/1e3) %>%
  arrange(pgdppc)

tot<-rclasif %>% summarise_if(is.numeric,sum)
resumen<-rclasif %>% bind_rows(tot) %>% mutate(Nivel=c("Bajo","
    Medio-bajo","Medio-alto","Alto","Total"))%>%
  select(Nivel,n,totpob,medgdppc,pgdppc) %>%
  rename( "Población total"=totpob,
          "No. de países"=n,
          "Mediana PIB per cápita"=medgdppc,
          "Promedio PIB per cápita"=pgdppc)

print(xtable(resumen ,digits=1),include.rownames=FALSE)

# Anexo con la clasificación para todos los países
tradclase<- data.frame(clas=c("L","LM","UM","H"),Nivel=c("Bajo","
    Medio-bajo","Medio-alto","Alto"))
apenclasif<- clasif2 %>% na.omit() %>%
  merge(y=tradclase) %>%
  select(country,code,Nivel,pob,gdppc) %>%
  arrange(desc(gdppc)) %>%
  mutate(pob=pob/1e6,gdppc=gdppc/1e3) %>%
  rename(País=country,"Código"=code, "Población millones"=pob,"PIB
      per cápita miles $USD"=gdppc)

print(xtable(apenclasif ,digits=1),include.rownames=FALSE)
```

../Code/0_wb_income_classif.R

Gráfica del simplex en tres dimensiones.

```

x=y=linspace(0,1,40)';
[xx,yy]=meshgrid(x,y);
z=1-xx-yy;

mesh(xx,yy,z);

##vectorx=[0 1]
##vectory=[1 0]
##vectorz=[0 0]
##plot3(vectorx,vectory,vectorz)

axis ([0,1,0,1,0,1]);
```

```
15
16 xlabel("x");
17 ylabel("y");
18 zlabel("z");
19 ##title("Simplex en 3 dimensiones");
```

../Code/planeTest.m

Código para diagramas ternarios.

```
1
2
3  suppressMessages(library(ggtern))
4  set.seed(1)
5  plot <- ggtern(data = data.frame(x = runif(100),
6                                   y = runif(100),
7                                   z = runif(100)),
8                 aes(x, y, z))
9  plot + stat_density_tern(geom='polygon',
10                           n         = 200,
11                           aes(fill  = ..level..,
12                               alpha = ..level..)) +
13   geom_point() +
14   theme_rgbw() +
15   labs(title = "Example Density/Contour Plot")    +
16   scale_fill_gradient(low = "blue",high = "red")  +
17   guides(color = "none", fill = "none", alpha = "none")
18
19
20 #Datos simplex aleatorios
21 set.seed(1906)
22 x<-runif(100)
23 temp<-runif(100)
24 y<-(1-x)*temp
25 z<-(1-x)*(1-temp)
26
27 plot <- ggtern(data = data.frame(x = x,
28                                  y = y,
29                                  z = z),
30                aes(x, y, z))
31 plot + stat_density_tern(geom='polygon',
32                          n         = 200,
33                          aes(fill  = ..level..,
34                              alpha = ..level..)) +
35   geom_point() +
36   theme_rgbw() +
37   labs(title = "Example Density/Contour Plot")    +
38   scale_fill_gradient(low = "blue",high = "red")  +
39   guides(color = "none", fill = "none", alpha = "none")
40
41 #####Datos simple puntos#####
42
43 nombre<-c("A")
44
45 x<-c(1/5)
46 y<-c(2/5)
47 z<-c(2/5)
```

```
plot <- ggtern(data = data.frame(x = x,
                                 y = y,
                                 z = z),
               aes(x, y, z))
plot+geom_point()+
  geom_text(aes(label=nombre),hjust=1, vjust=0) +
  theme_light()

#####Datos simple sumas#####

suppressMessages(library(compositions))
x<-c(0.5,0.1,0.4)
y<-c(.1,.3,.6)#c(1/3,1/3,1/3)
z<-c(.4,.6,0)#c(0.125,0.125,0.75)
nombre<-c("A","B","C")

x<-c(5/8,1/9,1/10)
y<-c(1/8,6/9,1/10)#c(1/3,1/3,1/3)
z<-c(2/8,2/9,8/10)#c(0.125,0.125,0.75)
plot <- ggtern(data = data.frame(x = x,
                                 y = y,
                                 z = z),
               aes(x, y, z))
plot+geom_point()+
  geom_text(aes(label=nombre),hjust=1, vjust=0) +
  theme_light()

datos<-data.frame(x = x, y = y, z = z,row.names = nombre)

AmasB<-perturbe(as.matrix(datos[1,]),as.matrix(datos[2,]))
BmasC<-perturbe(as.matrix(datos[2,]),as.matrix(datos[3,]))
AmasC<-perturbe(as.matrix(datos[1,]),as.matrix(datos[3,]))

oper<-rbind(AmasB,BmasC,AmasC)

nombre2<-c(nombre,"A+B","B+C","A+C")
datos2<-data.frame(rbind(as.matrix(datos), oper),row.names =
    nombre2)
plot <- ggtern(data = datos2,
               aes(x, y, z))
plot+geom_point()+
  geom_text(aes(label=nombre2),hjust=1, vjust=0) +
  theme_gray()
#####Datos simple potencias#####

A<-c(36/100,33/100,31/100)
pot<-seq(1,30)
n<-length(pot)
predat<-matrix(0,n,3)
#predat[1,]<-A
for (i in pot) {
  predat[i,] <-power.acomp(A,pot[i])
}

datpot<-data.frame(predat)
plot<-ggtern(data=datpot, aes(X1,X2,X3))
plot+geom_point()#+geom_text(aes(label=row.names(datpot)),hjust=1,
```

```
    vjust=-2,size=2)

#####Datos simple perturbacion sucesiva#####

A<-c(54/100,22/100,24/100)
B<-c(23/100,23/100,54/100)
pot<-seq(1,10)
n<-length(pot)
predat<-matrix(0,n+2,3)
predat[1,]<-A
predat[n+2,]<-B
for (i in pot+1) {
  if (i %%2==0) {
    predat[i,] <-perturbe(predat[i-1,],B)
  }else{
    predat[i,] <-perturbe(predat[i-1,],A)
  }

}
datnper<-data.frame(predat)
plot<-ggtern(data=datnper, aes(X1,X2,X3))
plot+geom_point()+geom_text(aes(label=c("A",pot,"B")),hjust=1,vjust
    =-2,size=2)
```

../Code/ternaryTest.R

Código para diagramas de distancias en el simplex.

```
####Paquetes a usar#####


if (!require("ggtern")) install.packages("ggtern")
suppressMessages(library(ggtern))
if (!require("compositions")) install.packages("compositions")
suppressMessages(library(compositions))
if (!require("Ternary")) install.packages("Ternary")
suppressMessages(library(Ternary))


n=3
#definir centro
cero<-acomp(c(1,1,1))
#probar distancias
u<-acomp(c(.4,.3,.3))
plot(cero)
TernaryPlot()
TernaryPoints(cero)
TernaryPoints(u)

disait<-function(a,b){
  n<-length(a)
  suma=0
  dena<-geometricmean(a)
  denb<-geometricmean(b)
  for (i in 1:n) {
    suma=suma+ (log(a[i]/dena)-log(b[i]/denb))^2
  }
```

```
  return(sqrt(suma))
}

disait(cero,u)

cerot<-clr(cero)
ut<-clr(u)

norm(cerot-ut) # validación norma euclidiana de trans. vs norma
    aitchison

u<-acomp(c(.9999998,.0000001,.0000001))
disait(cero,u)


###### Centrado en cero ###########

#definir distancia
d=.5
# intervalos de x
a=-sqrt(2/3)*d
b=+sqrt(2/3)*d

#partición para x
n=50
x=seq(from=a, to=b, length.out = n)
#calculo y, z
y1=-x/2+0.5*sqrt(2*d^2-3*x^2)
y2=-x/2-0.5*sqrt(2*d^2-3*x^2)
z1=y2
z2=y1
#concentrar en una tabla
datU=data.frame(x=c(x,x),y=c(y1,y2),z=c(z1,z2))
plot(datU)
matU=as.matrix(datU)

round(x+y1+z1,2)
dtest=rep(0,2*n)
for (i in 1:(2*n)){dtest[i]=norm(matU[i,])}
dtest

#graficar en U
#install.packages("plot3D")
#library(plot3D)
scatter3D(datU$x,datU$y,datU$z)

#transformar a simplex
matS=clrInv(datU)
datS=data.frame(matS)

#corroborar distancia aitchison
dtest=rep(0,2*n)
for (i in 1:(2*n)){dtest[i]=disait(cero,matS[i,])}
dtest


#graficar en simplex
```

```
TernaryPlot()
TernaryPoints(cero)
TernaryLines(datS[c(1:50),])
TernaryLines(datS[c(51:100),])

bol0<-function(n,d){
  # intervalos de x
  a=-sqrt(2/3)*d
  b=+sqrt(2/3)*d
  #partición para x
  x=seq(from=a, to=b, length.out = n/2)
  #calculo y, z
  y1=-x/2+0.5*sqrt(2*d^2-3*x^2)
  y2=-x/2-0.5*sqrt(2*d^2-3*x^2)
  z1=y2
  z2=y1
  datU=data.frame(x=c(x,x),y=c(y1,y2),z=c(z1,z2))
  matS=clrInv(datU)
  datS=data.frame(matS)
  TernaryPlot()
  TernaryPoints(cero)
  TernaryLines(datS[c(1:(n/2)),])
  TernaryLines(datS[c((n/2+1):n),])
  return(datS)
}

temp=bol0(100,.8)



###### Centrado en otro punto ############

#definir centro
x0=c(.1,.2,.7)
u=clr(x0)
#definir distancia
d=.5
# intervalos de x
a=-sqrt(2/3)*d+u[1]
b=+sqrt(2/3)*d+u[1]

#partición para x
n=50
x=seq(from=a, to=b, length.out = n)
#calculo y, z
y1=-(x-u[1])/2+0.5*sqrt(2*d^2-3*(x-u[1])^2)+u[2]
y2=-(x-u[1])/2-0.5*sqrt(2*d^2-3*(x-u[1])^2)+u[2]
z1=-(x-u[1])/2-0.5*sqrt(2*d^2-3*(x-u[1])^2)+u[3]
z2=-(x-u[1])/2+0.5*sqrt(2*d^2-3*(x-u[1])^2)+u[3]
#concentrar en una tabla
datU=data.frame(x=c(x,x),y=c(y1,y2),z=c(z1,z2))
plot(datU)
matU=as.matrix(datU)

round(x+y1+z1,2)
dtest=rep(0,2*n)
for (i in 1:(2*n)){dtest[i]=norm(matU[i,]-u)}
```

```
143 dtest
144
145 #graficar en U
146 #install.packages("plot3D")
147 #library(plot3D)
148 scatter3D(datU$x,datU$y,datU$z)
149
150 #transformar a simplex
151 matS=clrInv(datU)
152 datS=data.frame(matS)
153
154 #corroborar distancia aitchison
155 dtest=rep(0,2*n)
156 for (i in 1:(2*n)){dtest[i]=disait(x0,matS[i,])}
157 dtest
158
159
160 #graficar en simplex
161 TernaryPlot()
162 TernaryPoints(x0)
163 TernaryLines(datS[c(1:50),])
164 TernaryLines(datS[c(51:100),])
165
166
167
168 bolx0<-function(x0,n,d){
169   #transformar centro
170   u=clr(x0)
171   # intervalos de x
172   a=-sqrt(2/3)*d+u[1]
173   b=+sqrt(2/3)*d+u[1]
174   #partición para x
175   x=seq(from=a, to=b, length.out = n/2)
176   #calculo y, z
177   y1=-(x-u[1])/2+0.5*sqrt(2*d^2-3*(x-u[1])^2)+u[2]
178   y2=-(x-u[1])/2-0.5*sqrt(2*d^2-3*(x-u[1])^2)+u[2]
179   z1=-(x-u[1])/2-0.5*sqrt(2*d^2-3*(x-u[1])^2)+u[3]
180   z2=-(x-u[1])/2+0.5*sqrt(2*d^2-3*(x-u[1])^2)+u[3]
181   datU=data.frame(x=c(x,x),y=c(y1,y2),z=c(z1,z2))
182   matS=clrInv(datU)
183   datS=data.frame(matS)
184   TernaryPlot()
185   TernaryPoints(x0)
186   TernaryLines(datS[c(1:(n/2)),])
187   TernaryLines(datS[c((n/2+1):n),])
188   return(datS)
189 }
190
191 temp=bolx0(c(.2,.4,.4),100,.5)
```

../Code/distancias.R