

Ejercicio 2: Modelos de datos de panel

Instituto de Sociología - SOL3063

Matías Deneken

5 de junio, 2023

Preparativos ¹

Pregunta 1

```
# load data
load("mutz_ejercicio.RData")
data <- mutz_ejercicio

# reshape data from wide to long format
data_long <- long_panel(data = data.wide, prefix = '_', begin = 0, end = 1,
                        label_location = 'end', id = 'MNO', wave = 'wave')
```

```
head(data_long)
```

Pregunta 2

2.a) Formular DAG

Efectos fijos. En primer lugar, presentamos un DAG para Efectos Fijos.

En las definiciones entregadas por Wooldridge en su glosario, nos dice que el efecto fijo sirve para controlar la heterogeneidad inobservable. A propósito de esto “[en un modelo panel] una variable inobservable es el término error que no cambia en el tiempo. Para muestras de agrupamientos, una variable inobservable que es común a todas las unidades en el grupo”. Atendiendo a esto, es que se realizó el DAG con la variable (μ_i) como una variable inobservada.

Entendiendo lo anterior, a partir del texto de Mutz identificamos las siguientes variables: 1) Y_{it} refiere a la variable *Republican thermometer advantage*. El subíndice i indica que se mide por individuo y el subíndice t indica periodo de tiempo; donde 0 toma el valor para la elección del 2012 y 1 toma el valor para la elección del 2016. 2) X_{it} refiere a la variable *Social dominance orientation*. El subíndice i indica que se mide por individuo y el subíndice t indica periodo de tiempo; donde 0 toma el valor para la elección del 2012 y 1 toma el valor para la elección del 2016. 3) Como dijimos, el μ_i es una variable inobservada en el tiempo. El DAG nos representa, por un lado, las variables observadas de X_{it} e Y_{it} , mientras que las no observadas e invariantes en el tiempo se representan en el U_i .

¹Los códigos en profundidad pueden verse en el [Repositorio SOL3063](#).

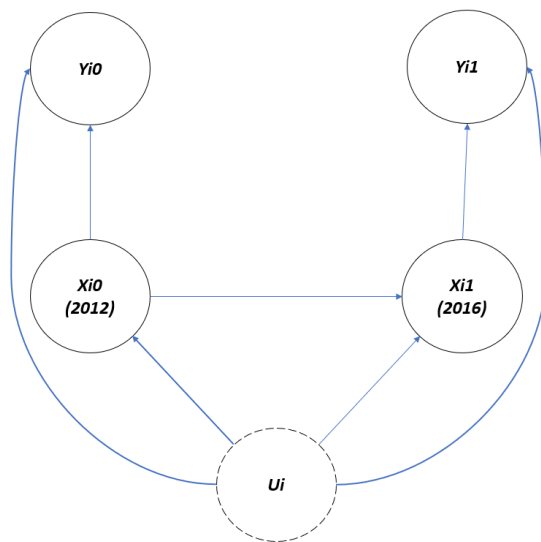


Figure 1: Grafo de efectos fijos

Efectos aleatorios En segundo lugar, dibujamos el de efectos aleatorios.

A diferencia del FE, los FA (Efectos Aleatorios) considera que U_i es independiente de las variables explicativas en los dos períodos de tiempo estudiado. En otras palabras, el FA implica que no existen variables no observadas que afectan en la estimación de las variables X sobre las variables Y .

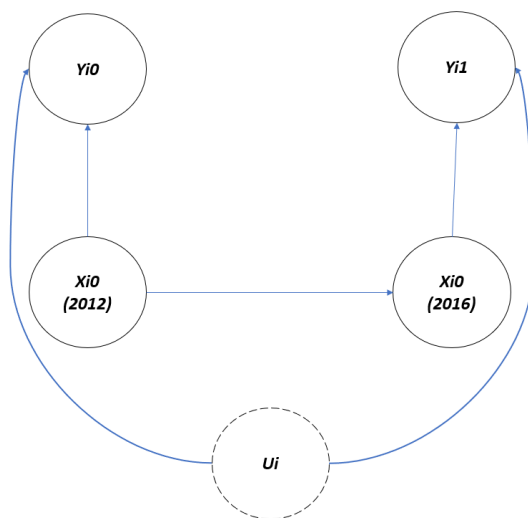


Figure 2: Grafo de efectos aleatorios

2.b) Explicar el supuesto de exogeneidad estricta

La premisa de exogeneidad estricta en modelos de efectos fijos postula que al controlar por U_i , que son variables no observadas constantes en el tiempo, no debería haber ninguna correlación entre las variables independientes que cambian con el tiempo y en el término de error que también experimenta variaciones temporales. En otras palabras, el supuesto de exogeneidad nos plantea que el estimado de FE no se encuentra

sesgado y , por lo tanto, el error no está correlacionado con ninguna variable explicativa dentro de los períodos. Sin embargo, a la luz de la lectura del texto Mutz, es posible identificar la violación de este supuesto.

- De acuerdo con Mutz (2018), no se observa correlación entre la variable independiente en el pasado, la escala de dominancia social (SDO) en 2012, y la variable dependiente en el futuro, la preferencia por Trump en 2016. Sin embargo, en Estados Unidos existe una marcada disparidad racial que podría generar un efecto feedback. Esto se debe a que las personas que tienden a votar por Trump, en su mayoría, son de raza blanca y exhiben niveles más elevados de dominancia social.
- De igual manera, el voto por los Republicanos en 2012 puede influir en la escala de dominancia social para 2016. Debido a que Barack Obama, un presidente de raza negra que defendía la igualdad racial, resultó electo, las personas de raza blanca podrían haber percibido una amenaza a su estatus de supremacía. Esto podría haber conllevado a un aumento en sus niveles de dominancia social para el año 2016 (Año de elección de Trump). En este sentido, existen diferencias entre los votantes que explican la variación en el cambio de apoyo al Partido Republicano, las cuales no pueden ser explicadas por cambios a lo largo del tiempo.

Pregunta 3

Ahora vamos a hacer estimaciones. Para ello tenemos el siguiente modelo:

$$cutdifftherm_{it} = \beta_0 + \beta_1 xparty3_{it} + \beta_2 SDO_{it} + \mu_i + \phi_t + \epsilon_{it} \quad (2)$$

1. La presente Tabla 1 muestra las estimaciones para el *modelo(2)* empleando un modelo de efectos fijos (FE) y efectos aleatorios (RE)

```
fe <- plm(cutdifftherm ~ xparty3 + sdo + factor(wave), data = data_long,
          model = "within")
re <- plm(cutdifftherm ~ xparty3 + sdo + factor(wave), data = data_long,
          model = "random")
```

Table 1: Estimación Efectos Fijos vs Efectos Aleatorios

	Termómetro Republicano	
	FE	RE
xparty3	-1.439*** (0.218)	-3.933*** (0.096)
sdo	0.198*** (0.074)	0.448*** (0.048)
factor(wave)1	-0.418*** (0.128)	-0.634*** (0.127)
Constant		16.344*** (0.328)
Observations	2,111	2,111
R ²	0.057	0.506
Adjusted R ²	-1.160	0.506
F Statistic	18.605*** (df = 3; 921)	2,173.487***

Note: *p<0.1; **p<0.05; ***p<0.01

2. El coeficiente de tiempo (Wave) en el modelo de efectos fijos (Modelo 1) es -0.42, y es estadísticamente significativo al 99,9%. Esto indica que, en promedio, la preferencia por el Partido Republicano (Trump) en el año 2016 disminuye en 0.42 puntos en comparación con el año 2012, considerando todos los demás factores constantes.

Por otro lado, el coeficiente de tiempo en el modelo de efectos aleatorios (Modelo 2) es -0.63, y también es estadísticamente significativo al 99,9%. Esto significa que, en promedio, la preferencia por Trump disminuye en 0.63 puntos para el año 2016 en comparación con el año 2012, manteniendo constantes los demás factores.

Dado que ambos coeficientes de tiempo son estadísticamente significativos al 99,9%, podemos afirmar con un alto grado de confianza que hay una disminución en la preferencia por Trump entre los años 2012 y 2016, sin importar el modelo utilizado. Sin embargo, existe una leve diferencia en la magnitud de la disminución: 0.42 puntos en el Modelo 1 (efectos fijos) y 0.63 puntos en el Modelo 2 (efectos aleatorios). Este resultado difiere del indicado por Mutz (2018), lo que se podría deber a que nuestras estimaciones no están siendo controladas por variables que son relevantes en el modelo de la autora.

3. Se realiza el Test de Hausman para ver cuál estimación es más adecuada.

- H_{0nula} : ambas estimaciones son consistentes, mientras que las estimaciones de efectos aleatorios son eficientes
- $H_{1alternativa}$: las estimaciones de efectos fijos son consistentes mientras que las estimaciones de efectos aleatorios no

```
test <- phtest(fe, re); test
```

Los resultado del test de Hausman nos llevan a rechazar la hipótesis nula, es decir, el modelo de efectos fijos debería ser preferido en vez del modelo de efectos aleatorios. Esto nos dice que el modelo de efectos fijos es consistente. En otras palabras, el Test nos indica que hay presencia de atributos individuales no observados y que devienen relevante en el efecto de la dominancia social sobre la preferencia por los candidatos del partido republicanos.

En términos estadísticos, una razón por la cual se puede considerar que un modelo de efectos fijos es preferible a un modelo de efectos aleatorios es su capacidad para controlar variables no observadas que no varían en el tiempo, evitando problemas de endogeneidad. En contraste, los modelos de efectos aleatorios no controlan por estas variables no observadas (U_i), sino que asumen que estas variables endógenas no afectan ni a las variables independientes ni a las dependientes, lo cual puede generar problemas de exogeneidad.

En el caso específico del texto, tiene sentido pensar que variables como el género o la raza deben ser controladas, dado que en general, los grupos marginalizados como las mujeres o las personas negras tienden a mostrar menor apoyo hacia el Partido Republicano y presentan niveles más bajos en la escala de dominancia social. Por otro lado, son los hombres blancos quienes suelen mostrar un mayor apoyo hacia Trump y presentan niveles más altos de dominancia social.

Una posible solución sería incorporar variables como el sexo y la etnia como variables independientes en el modelo de efectos aleatorios. Sin embargo, en la estimación del modelo solo se consideran las variables independientes que varían en el tiempo, lo cual puede limitar la capacidad de controlar completamente por estas variables.

4. En ambos modelos, el coeficiente estimado para SDO se mantiene constante, ya que los errores estándar agrupados generan estimaciones más eficientes, pero no mejoran la consistencia del estimador. Sin embargo, es importante notar que los errores estándar del modelo de efectos fijos con errores agrupados son ligeramente mayores en comparación con el modelo sin agrupación de errores.

Además, la significancia estadística del coeficiente SDO pasa de un nivel de confianza del 95% en el modelo de efectos fijos con errores agrupados a un nivel de confianza del 99,9% en el modelo de efectos fijos sin agrupación de errores. Esto sugiere que existe una correlación serial entre los clusters/individuos, es decir, la escala de dominancia social reportada por una persona en 2012 es similar en comparación con 2016 para

ese mismo individuo. Este hallazgo tiene sentido, ya que las opiniones políticas de las personas no suelen cambiar mucho entre periodos o cambian muy levemente.

En resumen, el uso de errores estándar agrupados en el modelo de efectos fijos permite estimaciones más eficientes, pero no afecta el coeficiente estimado para SDO. Sin embargo, la significancia estadística y la presencia de correlación serial entre los periodos indican que el modelo de efectos fijos sin agrupación de errores es más apropiado en este contexto.

Pregunta 4

A continuación se arma los códigos correspondientes.

```
data_long <- data_long %>%
  group_by(MNO) %>%
  mutate(msdo = mean(sdo)) #Media de SDO

data_long <- data_long %>%
  group_by(MNO) %>%
  mutate(Wsdo = sdo - msdo) # Resta de SDO con Media de SDO

# Mundlack with RE
mundlack <- plm(cutdifftherm ~ xparty3 + sdo + msdo + factor(wave),
  data = data_long, model = "random")

# Within-between with OLS and Cluster SE
wb_ols <- lm(cutdifftherm ~ xparty3 + Wsdo + msdo + factor(wave),
  data = data_long)

wb_cluster <- lm_robust(cutdifftherm ~ xparty3 + Wsdo + msdo + factor(wave),
  data = data_long, clusters = MNO)
```

- 4.1 Modelo Mundlak

4.1a) El coeficiente de SDO en el modelo de efectos fijos utilizando el enfoque de Mundlak representa la variación “within”, es decir, cómo la escala de dominancia social afecta el apoyo al partido republicano en los dos tiempos (2012 y 2016) en el mismo individuo. El coeficiente es 0.17, y es estadísticamente significativo en un 95% de confianza. En términos empíricos, un aumento de un punto en la escala de dominancia social incrementa en 0.17 puntos la preferencia por Trump en las elecciones presidenciales del 2016.

4.1b) Por otro lado, el coeficiente SDO2 indica la variación “between”, es decir, la diferencia en el apoyo al partido republicano entre las personas. Este coeficiente es 0.32 y es estadísticamente significativo en un 99%. Por lo tanto, un aumento de un punto en la escala de dominancia social, en promedio, incrementa en 0.32 puntos la preferencia por Trump.

- 4.2 Modelo within-between

4.2a) En el modelo “within-between” utilizando regresión OLS, el coeficiente wsdo identifica el efecto “within” al restar el valor individual en la escala de dominancia social al efecto promedio ($X_{ij} - X_j$). Esto representa cómo la dominancia social afecta el apoyo al partido republicano a lo largo del tiempo en un mismo individuo. En este caso, el coeficiente es 0.14 y no es estadísticamente significativo, ya que el intervalo de confianza incluye el valor 0. Por lo tanto, la escala de dominancia social no puede explicar la variación individual en el apoyo al candidato republicano en el año 2016.

4.2b) Por otro lado, el coeficiente sdo2 en este modelo representa el efecto contextual, es decir, la variación “between” menos la varianza de los individuos a lo largo del tiempo. Este coeficiente es 0.48 y es estadísticamente significativo al 90%. Esto indica que un aumento de un punto en la escala de dominancia social, en promedio, incrementa en 0.48 el apoyo al partido republicano. Este efecto contextual puede deberse a políticas públicas, shocks económicos u otras variables exógenas que influyen en el cambio del apoyo al partido republicano.

Table 2: Estimación modelos híbridos

	Mundlack RE	W-B Ols Clusters
(Intercept)	15.354*** (0.417)	16.458*** (0.441)
xparty3	-3.815*** (0.105)	-4.186*** (0.114)
sdo	0.177* (0.076)	
msdo	0.463*** (0.103)	0.553*** (0.066)
factor(wave)1	-0.549*** (0.131)	-0.557*** (0.138)
Wsdo		0.156 (0.087)
s_idios	2.733	
s_id	2.391	
R ²	0.500	0.599
Adj. R ²	0.499	0.598
Num. obs.	1921	1921
RMSE		3.729
N Clusters		997

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

A modo de conclusión, el modelo within-between ofrece la posibilidad de descomponer los efectos fijos y aleatorios en un único modelo, lo cual brinda una tolerancia significativa en su configuración al combinar las ventajas de los modelos de efectos fijos y aleatorios. Al utilizar este enfoque, se puede estimar el efecto de la variable sdo a nivel individual, al mismo tiempo que se proporcionan estimaciones del efecto a nivel temporal que no están sesgadas por una posible correlación con el error de los grupos o clusters (individuos). Además, al comparar los efectos dentro (within-effects) y entre (between-effects), o su diferencia, se puede evaluar hasta qué punto la heterogeneidad no observada en las características individuales U_i es responsable de la relación observable entre el resultado y una variable que varía en el tiempo y no se explica en el estudio. Esta comparación permite analizar el grado de influencia de dicha heterogeneidad no observada en la relación estudiada.

Pregunta 5

A partir de una acuciosa lectura del texto de Morgan se desprenden las siguientes aseveraciones sobre la crítica que hace éste a Mutz.

1. Morgan cuestiona que la amenaza de estatus por sí sola no es suficiente para explicar la victoria de Trump en 2016, coincidiendo con Mutz. Morgan encuentra que los predictores relacionados con intereses económicos tienen un mayor poder explicativo en el cambio en el apoyo a Trump, mientras que el coeficiente de amenaza de estatus es relativamente pequeño.
2. Morgan destaca que los cambios en el apoyo a Trump a lo largo del tiempo parecen ser impulsados por diferencias estables entre las personas, en lugar de variaciones temporales dentro de un mismo individuo. Esto tiene implicaciones en la elección de los modelos estadísticos, ya que Morgan argumenta que los modelos de efectos fijos son más apropiados para identificar la causalidad al controlar por variables no observadas constantes en el tiempo.
3. La inclusión de variables adicionales en el análisis de Mutz, como la identificación partidaria y otros predictores relacionados con temas específicos, genera preocupación en Morgan debido a la endogeneidad que introducen, al estar correlacionadas con variables de raza y alineación de posiciones individuales con las del candidato preferido.
4. Morgan critica la inclusión de efectos de interacción en el estudio de Mutz, ya que no se presenta una tabla bivariada para observar el efecto de cada predictor individualmente, dificultando una comparación adecuada entre los efectos en 2012 y 2016. Además, el uso generalizado de efectos de interacción no aclara las interpretaciones sobre la causalidad directa entre la amenaza de estatus y la diferencia en el apoyo al partido republicano.

En conclusión, Morgan plantea que no se deben descartar los factores de intereses económicos al explicar la victoria de Trump, a pesar de reconocer la importancia potencial de la escala de dominancia social mencionada por Mutz. Estas críticas destacan la importancia de utilizar análisis estadísticos rigurosos y considerar la endogeneidad y otras variables relevantes para comprender fenómenos políticos y sociales complejos.