Efectos Fijos y Aleatorios

Escuela de Métodos ELSOC

Benjamín Muñoz

Massachusetts Institute of Technology

31 de Julio de 2025

Tabla de Contenidos

Motivación

★ Desde Converse (1964), pensamos que los ciudadanos no poseen sistemas de creencias coherentes y estables. *¿Sucede esto con la identificación ideológica?*

Posición	2016	2017	2018	2019	2021	2022
Izquierda	21.4	22.4	23.0	22.5	23.4	20.4
Centro	21.5	19.4	28.9	29.3	37.8	27.1
Derecha	14.7	18.0	19.4	13.8	16.9	17.3
Ninguno	42.5	40.1	28.6	34.4	21.9	35.2

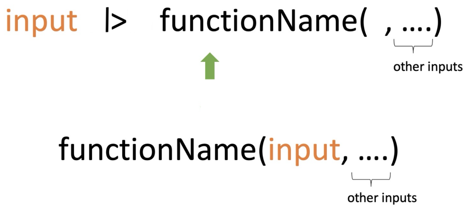
Cuadro 1: Posición Ideológica por Ola, ELSOC 2016-2022

Patrón	Porcentaje
Cambia de Posición	86.41
Centro-Siempre	2.42
Derecha-Siempre	2.84
Izquierda-Siempre	5.25
Ninguno-Siempre	3.09

Cuadro 2: Consistencia Ideológica

Materiales del Workshop

- En la carpeta Drive del curso pueden acceder al archivo comprimido (.zip): `Materiales_Curso__Efectos_Fijos_y_Aleatorios`
- En su interior hay dos tipos de archivos:
  - `Codigo_01__Ejemplos_de_Clases.R`: replicación de los Cuadros y Gráficos de las diapositivas de clases.
  - `Codigo_02__Ejercicios_Practicos.R`: para el módulo práctico.
  - Bases de datos ELSOC (Wide y Long, formato `.dta`).
- El código de R está estructurado en base a **pipes** (Consultar R4DS o Posit):



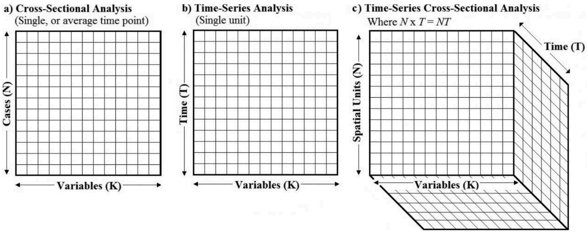
¿Para qué sirven los datos panel?

- Para analizar el **cambio** individual.
  - Más allá del nivel (*Cross-Section*) y la tendencia (*Pooled Cross-Section* o *Time-Series*).
- Para contar con una estrategia más sólida de **causalidad** y de medición.
  - Reducir los riesgos del sesgo por variable omitida.
  - Por medio del modelamiento de la Heterogeneidad Individual o Temporal.
- Sin embargo, su uso implica muchos **desafíos**:
  - Periodicidad y cambio poblacional.
  - Dependencia estadística: Procesos temporales, patrones espaciales, etc.
  - Desafíos Inferenciales (sesgo, parámetros incidentales, etc.).
  - Atrición.
  - Medición Consistente y Efectos de Panel (*panel conditioning*).

Tabla de Contenidos

Datos de Panel

- Observaciones repetidas de “unidades” en el tiempo.
- Tres dimensiones:
  - **Unidades** $i = 1, \ldots, N$.
  - **Variables:** $v = 1, \ldots, V$ (constantes y variables en el tiempo).
  - **Mediciones** (en el tiempo) $t = 1, \ldots, T$.
- ¿Cómo podemos representar la información en matrices de datos de dos dimensiones?



Datos de Panel: Formatos

Formato Largo (Long Format)

- $N \times T$ filas y V columnas
- Fila $\longrightarrow$ unidad-medición (2 *llaves*)

country	year	metric
x	1960	10
x	1970	13
x	2010	15
y	1960	20
y	1970	23
y	2010	25
z	1960	30
z	1970	33
z	2010	35

Formato Ancho (Wide Format)

- N filas y $V \times T$ columnas
- Fila $\longrightarrow$ unidad (1 *llave*)

country	yr1960	yr1970	yr2010
x	10	13	15
y	20	23	25
z	30	33	35

Aplicación en R

country	year	metric
x	1960	10
x	1970	13
x	2010	15
y	1960	20
y	1970	23
y	2010	25
z	1960	30
z	1970	33
z	2010	35

```
pivot_wider(names_from = “year”,
            names_prefix = “yr”,
            values_from = “metric”)
```

country	yr1960	yr1970	yr2010
x	10	13	15
y	20	23	25
z	30	33	35

```
pivot_longer(cols = yr1960:yr2010,
             names_to = “year”,
             names_prefix = “yr”
             values_to = “metric”)
```

Más información sobre las funciones de pivoting de `tidyr`. ▸ Apéndice A: Reshaping

Aplicación en R

Código de R

```
## Transformación de Datos Long a Wide
elsoc_long %>%
  dplyr::select(idencuesta, ola, c15) %>%
  tidyr::pivot_wider(names_from = ola, values_from = c15, names_prefix = "w_") %>% head(3)

#  idencuesta   w_1   w_2   w_3   w_4   w_5   w_6
#       <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#1    1101011    11    12     7     0     2    12
#2    1101012    12    12     5    12    12    12
#3    1101013    12    12    12    12    12    NA

## Transformación de Datos Wide a Long
elsoc_wide %>%
  dplyr::select(idencuesta, starts_with("c15")) %>%
  tidyr::pivot_longer(cols = starts_with("c15"), names_to = "ola", values_to = "c15") %>%
  dplyr::mutate(ola = stringr::str_remove(string = ola, pattern = "^c15_w0")) %>% head(3)

#  idencuesta ola     c15
#       <dbl> <chr> <dbl>
#1    1101011 1        11
#2    1101011 2        12
#3    1101011 3         7
#4    1101011 4         0
#5    1101011 5         2
#6    1101011 6        12
```

Datos de Panel: Aspectos Avanzados

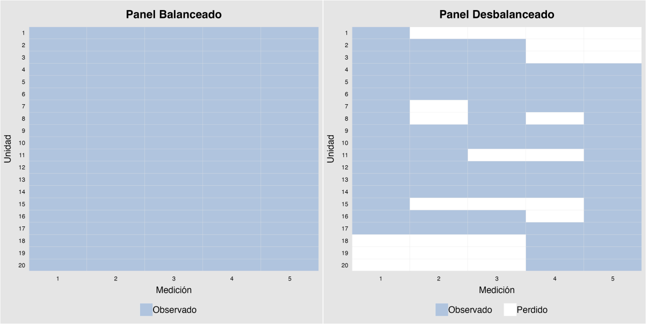
1. **Balance:** presencia de las unidades a lo largo del tiempo.
   - Panel Balanceado: todas las unidades del panel tienen el mismo número de mediciones. N unidades medidas en T ocasiones $\longrightarrow$ No hay Datos Perdidos.
   - Panel Desbalanceado: una o más unidades no son observadas en todas las mediciones. T no es igual para las N unidades (T_i) $\longrightarrow$ **Hay Datos Perdidos**
     - *Atrición* (unidades dejan el panel permanentemente).
     - *Temporary Unit Nonresponse* (salida y entrada del panel)
     - *Ingreso Posterior* al panel.

   ⋆ Los datos perdidos son problemáticos si no podemos asumir **Missingness at Random** (o mejor aún *Missingness Completely at Random*).

Aplicación en R

Código de R

```
## Generar un gráfico de Valores Perdidos
panel_data %>%
  panelView::panelview(data = ., formula = Y ~ X,
                    index = c("Unit","Time"), type = "miss")
```



Más información sobre el paquete `panelView`. Apéndice B: Balance

Datos de Panel: Aspectos Avanzados

2. **«Tamaño» del Panel:** tamaño relativo de las dimensiones del panel.
   - Panel Largo: relativamente pocas unidades medidas en muchas ocasiones.
     - ⋆ $T >> N$
     - ⋆ Se asume N fijo y se examina $T \rightarrow \infty$
     - ⋆ *Time-Series Cross-Sectional* (TSCS) Data o Macro Panel.
     - ⋆ Mayor capacidad para identificar procesos temporales.
     - ⋆ Ejemplo: Países de América Latina (17) observados desde 1990 a 2023 (33).
   - Panel Corto: gran número de unidades medidas en pocas ocasiones.
     - ⋆ $T << N$.
     - ⋆ Se asume T fijo y se examina $N \rightarrow \infty$
     - ⋆ Micro Panel.
     - ⋆ Mayor capacidad para explorar heterogeneidad de unidades.
     - ⋆ Ejemplo: 3000 entrevistados en 7 olas (ELSOC).
3. **Estructura del Panel**: persistencia de unidades en el panel.
   - Panel Fijo: mismas unidades a lo largo del estudio.
   - Panel Rotatorio: proporción fija de unidades es reemplaza en cada medición.

Tabla de Contenidos

Dependencia Estadística

- **Ejemplo:** realizamos una encuesta, recolectando respuestas de 3,000 personas. Repito el proceso de entrevista en 6 ocasiones. ¿Cuántas observaciones tengo?
    1. En un sentido **aritmético** tengo $3,000 \times 6 = 18,000$.
    2. Sin embargo, en un sentido **estadístico** no tengo $18,000$ observaciones.

Ola	N	Media	DE	Cor$(t, t-1)$	Cor$(t, t=1)$
W01: 2016	1274	4.350	1.532	-	-
W02: 2017	1274	4.498	1.572	0.326	0.326
W03: 2018	1274	4.402	1.509	0.398	0.324
W04: 2019	1274	4.230	1.528	0.397	0.322
W05: 2021	1274	4.714	1.551	0.311	0.275
W06: 2022	1274	4.665	1.587	0.312	0.248

Cuadro 3: Estatus Subjetivo Individual (ELSOC)

Apéndice C: Código de R

Dependencia Estadística

- Los valores de Y_{it} a lo largo del tiempo no son independientes.
  - Llamamos **correlación serial** a $\text{Corr}(Y_{it}, Y_{i,t-1})$ (*orden* 1) y $\text{Corr}(Y_{it}, Y_{i,1})$ (*orden* $k = t$).
  - Dependencia disminuye al incrementar distancia entre mediciones.

- ¿Qué explica la dependencia serial observada en Y_{it}?
  - Características **invariantes en el tiempo** (fijas) de las unidades (**Z**, por ejemplo, *sexo* o *color de piel*).
  - Características **variables en el tiempo** de las unidades (**X**, por ejemplo, *nivel de ingresos* y las *redes de conocidos*).
  - La **variable dependiente rezagada** (por ejemplo, Y_{it-1} o Y_{it-2}) $\longrightarrow$ *true state-dependence*.

- ¿Qué explica la disminución de la dependencia serial observada en Y_{it}?
  - Relaciones estocásticas (término de error) y los efectos de **X** se acumulan en el tiempo.

Dependencia Estadística

Correlación Serial	T=1	T=2	T=3	T=4
Sí				
Sí				
No				

Dependencia Estadística

- ¿Qué sucede si ignoro la dependencia/correlación serial?
  1. **Sesgo** en la estimación de la varianza del estimador $\longrightarrow$ Errores estándares sesgados (resultados inferenciales incorrectos).
  2. Como mínimo, el estimador será **ineficiente** y, en ocasiones, también estará sesgado. Apéndice D: Propiedades de un Estimador

- ¿Qué podemos hacer para lidiar con la correlación serial?

Demeaning

La variación entre unidades constante en el tiempo puede ser purgada restando la media específica de la unidad:

$$\begin{aligned}\ddot{Y}_{it} &= Y_{it} - \bar{Y}_i \\ &= Y_{it} - \sum_{t=1}^{T} \frac{Y_{it}}{T}\end{aligned}$$

Dependencia Estadística

Unidad	Tiempo	Valor
A	1	5
A	2	4
A	3	5
A	4	3
A	5	6
B	1	5
B	2	3
B	3	9
B	4	7
B	5	6

$$\frac{\sum_{t=1}^{5} Y_{A,t}}{5} = 4.6$$

$$\frac{\sum_{t=1}^{5} Y_{B,t}}{5} = 6.0$$

Unidad	Tiempo	Valor
A	1	5-4.6= 0.4
A	2	4-4.6=-0.6
A	3	5-4.6=-0.6
A	4	3-4.6=-1.6
A	5	6-4.6= 1.4
B	1	5-6.0=-1.0
B	2	3-6.0=-3.0
B	3	9-6.0= 3.0
B	4	7-6.0= 1.0
B	5	6-6.0= 0.0

Dependencia Estadística

Código de R

```
wagepan %>%
  ### Seleccionar variables relevantes
  dplyr::select(nr, year, lwage) %>%
  ### Transformar a formato wide
  pivot_wider(names_from = year, values_from = lwage, names_prefix = "lw_") %>%
  ### Calcular demeaned variables
  rowwise() %>% dplyr::mutate(Mean = mean(c_across(starts_with("lw")), na.rm = TRUE)) %>%
  dplyr::mutate(across(starts_with("lw"), ~ . - Mean, .names = "dm.col")) %>% ungroup() %>%
  dplyr::select(starts_with("lw"), starts_with("dm")) %>%
  dplyr::summarise(#### Serial Correlation: Original variables respect t=1
        across(starts_with("lw_")[-8], ~ cor(., get(paste0("lw_", as.numeric(sub("lw_", "", cur_column())) + 1)),
                                        use = "complete.obs"), .names = "cor_lw_col"),
                  #### Serial Correlation: Demeaned variables respect t=1
        across(starts_with("dmlw_")[-8], ~ cor(., get(paste0("dmlw_", as.numeric(sub("dmlw_", "", cur_column())) + 1)),
                                        use = "complete.obs"), .names = "cor_dmlw_col")
  ) %>% rename_with(~ gsub("cor_(lw_|cor_next_|dmlw_)", "", .)) %>%
  ### Formato final
  tidyr::pivot_longer(cols = everything(), names_to = c("type", "year"), names_sep = "_") %>%
  tidyr::pivot_wider(names_from = year, values_from = value) %>%
  setNames(nm = c("type", paste0("Y", 1981:1987),"Variable")) %>%
  dplyr::select(Variable, Y1981:Y1987)
```

Variable	Y1981	Y1982	Y1983	Y1984	Y1985	Y1986	Y1987
Raw	0.454	0.611	0.690	0.674	0.664	0.632	0.693
Demeaned	0.025	0.031	0.062	0.028	0.034	0.035	0.251

Cuadro 4: Corrección de Correlación Serial

Lección 1

Siempre Explorar Descriptivamente los Datos

Tabla de Contenidos

Motivación: Medias Condicionales

- Me interesa estudiar los retiros de pensiones. Por lo tanto, he solicitado datos administrativos a la Superintendencia de Pensiones sobre este proceso:

1. **Promedio Global** de los saldos antes 1er Retiro $\frac{\sum_{i=1}^{N} \textbf{Saldo}_i}{N} = \$13,543,045$
2. **Promedio Condicional según Sexo**:
   - $\frac{\sum_{i=1}^{N} \textbf{Saldo}_i \,|\, \textbf{Sexo}_i = Hombre}{N} = \$17,124,200$
   - $\frac{\sum_{i=1}^{N} \textbf{Saldo}_i \,|\, \textbf{Sexo}_i = Mujer}{N} = \$9,558,303$
3. **Promedio Condicional según Sexo, Nacionalidad y Número de Solicitudes:**

Sexo	**Nacionalidad**	**Retiro =1**	**Retiro = 2**	**Retiro =3**
Hombre	Chileno	\$6,819,193	\$16,872,635	\$20,306,031
Hombre	Extranjero	\$1,548,343	\$3,005,798	\$3,902,846
Mujer	Chileno	\$2,412,083	\$7,676,739	\$12,898,882
Mujer	Extranjero	\$947,178	\$1,750,579	\$3,288,924

Cuadro 5: Saldos en Cuentas de Pensiones

Análisis de Regresión con Datos de Panel

- **Model Based Inference:** Uso de modelos estadísticos para hacer inferencias sobre la realidad.

Función de Esperanza Condicional Lineal y Aditiva

$$\text{FEC o F. de Regresión: } \mathbb{E}[Y_i \,|\, X_{i1}, X_{i2}, \dots, X_{ik}] = g(X_{i1}, \dots, X_{ik})$$

$$\text{FEC Lineal: } \mathbb{E}[Y_i \,|\, X_{i1}, X_{i2}, \dots, X_{ik}] = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}$$

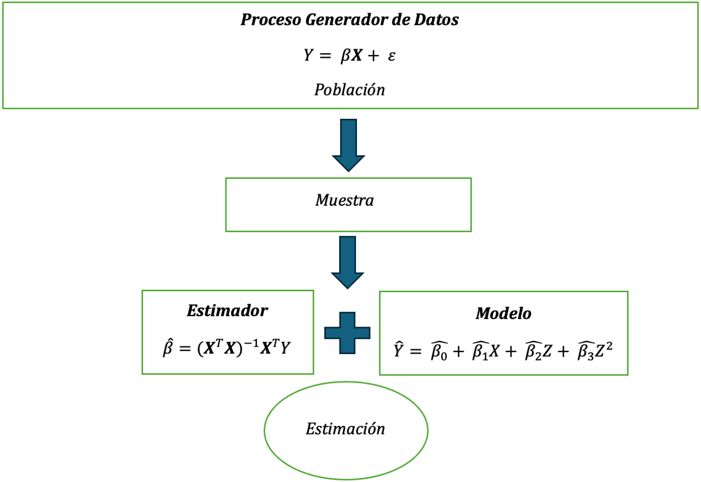
$$\text{Defino el error como: } u_i = Y_i - \mathbb{E}[Y_i \,|\, X_{i1}, X_{i2}, \dots, X_{ik}]$$

$$\text{Resultando en el modelo: } Y_i = \alpha + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i$$

1. La aleatoriedad proviene del **Proceso de Generación de Datos** (DGP) asumido.
2. Modelo gobierna la relación entre variable de resultado (Y) y variables explicativas (X).

⋆ El principal supuesto es que mi modelo está **correctamente especificado**.

Model-Based Inference



Modelo de Componentes del Error

- Comenzamos con el siguiente modelo poblacional:

$$Y_{it} = \alpha + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + u_{it}$$

dónde hay k regresores, $con i = 1, \ldots, N$ y $t = 1, \ldots, T$

  - Este DGP (*Data-Generating Process*) tiene un supuesto de **homogeneidad** en las unidades. Apéndice E: Notación Matricial
    ⋆ Este es un Modelo Lineal, Aditivo y Estático.

- Ahora, le daremos mayor forma al término de error:

$$u_{it} = \delta_i + \epsilon_{it}$$

  - El error es dividido en dos componentes: un efecto individual (δ_i, heterogeneidad individual, invariante en el tiempo) y un error idiosincrático (ϵ_{it}).

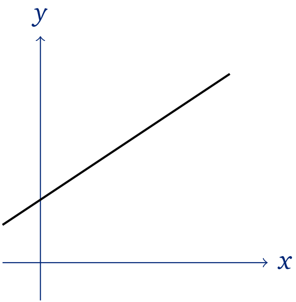
Modelo de Componentes del Error

- Volvamos al modelo poblacional:

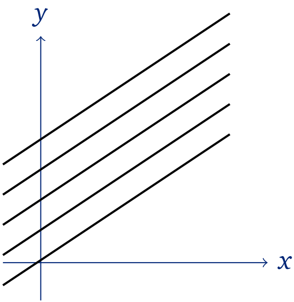
$$
\begin{aligned}
Y_{it} &= \alpha + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + u_{it} \\
&= \alpha + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + (\delta_i + \epsilon_{it}) \\
&= (\alpha + \delta_i) + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + \epsilon_{it} \\
&= \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + \epsilon_{it}
\end{aligned}
$$

- Esta ecuación corresponde al **One-Way Error Component Model**.
  - Comúnmente, α_i es llamado **heterogeneidad** o efectos individuales.
  - Captura todos los factores invariantes en el tiempo (y no incluidos como regresores) que afectan a Y_{it}.

Comparación de Modelos



$$Y_{it} = \alpha + \beta X_{it} + u_{it}$$
$$(\beta > 0)$$



$$Y_{it} = \alpha_i + \beta X_{it} + u_{it}$$
$$(\beta > 0)$$

- Otra manera de llamar al Modelo de Componentes del Error es como el **Modelo de Interceptos Variables** (Varying Intercept Model).
  - ⋆ Las pendientes siguen siendo constantes (efecto común a todas las unidades).

Lección 2

Los datos de panel nos permiten modelar explícitamente la Heterogeneidad Individual

Mantra

Heterogeneidad Individual $\neq$ Causalidad Reversa («Simultaneity Bias»)

Pasos a Seguir

- ¿Qué podemos hacer con nuestro Modelo Poblacional (One-Way Error Component Model)?
- Lo primero es tener claro nuestra Cantidad de Interés ⟶ Efecto Parcial de $\mathbf{X}$.

$$\frac{\partial \mathbb{E}[Y_{it} \mid \mathbf{X}_{it}]}{\partial \mathbf{X}_{it}} = \boldsymbol{\beta}$$

  ⋆ Por el momento, no estamos incluyendo interacciones/efectos condicionales.

  ⋆ No es el interés principal estimar el término α_i.

- En lo que resta del Workshop discutiremos distintos estimadores que nos permiten aproximar el valor de $\boldsymbol{\beta}$ con muestras de datos.

  ⟶ Pooled MCO
  ⟶ Primeras Diferencias
  ⟶ Efectos Fijos
  ⟶ Efectos Aleatorios

Pasos a Seguir

- ¿Cuál es el **Mejor** estimador entre estas 4 opciones?

  ⋆ No hay un estimador óptimo. Depende de ciertas características de los datos/ fenómenos bajo estudio.

- Para elegir el mejor estimador es necesario concebir tres escenarios:
  1. No hay heterogeneidad individual ($\forall \alpha_i = \alpha$).
  2. La heterogeneidad (α_i) es observada para todos los individuos.
  3. La heterogeneidad (α_i) **NO** es observada para todos los individuos.

- Distintos estimadores serán preferibles según que tan creíbles sean dos aspectos adicionales del Modelo:
  - Supuestos sobre la distribución del Error (ϵ_{it}).
  - Supuestos sobre la relación entre el efecto individual (δ_i) y los regresores ($\mathbf{X}_{it}$).

Ejemplo Práctico: Simulación

- Para ver las propiedades de cada estimador, simularemos datos con las siguientes características:
  - **Variable de Resultado:** Aprobación Presidencial (`apr_pres`, V. Continua).
  - **Variable Invariante en el Tiempo** (Z): Sexo (`sexo`, V. Binaria, sólo varía **entre** unidades).
  - **Variable Variante en el Tiempo** (X): Ideología y Nivel de Ingresos (`ideol` e `ingresos`. V. Ordinal y Continua. Ambas varían entre unidades y mediciones (**dentro** de las unidades).
  - NO incluyo **Variables cronológicas** (T) → varían sólo entre mediciones.
- Crearé tres versiones de los datos:
  - `df_pooled`: NO hay heterogeneidad individual.
  - `df_fe`: la heterogeneidad individual está correlacionada con los regresores.
  - `df_re`: la heterogeneidad individual es independiente de los regresores.

Código de R

```
## Parámetros de la simulación
N    <- 5000                   # Número de individuos
T    <- 5                      # Número de períodos

## A) Generar datos: DGP MCO Pooled (No hay heterogeneidad individual)
data.frame(id    = rep(1:N, each = T), # ID de los individuos
           time  = rep(1:T, N))  %>%   # Periodos de tiempo
  dplyr::group_by(id) %>%  #No hay heterogeneidad individual
  dplyr::ungroup() %>%
  dplyr::mutate(sexo     = rep(sample(0:1, N, replace = TRUE), each = T),       # V. constante en el tiempo
                ideol    = round(scales::rescale(rnorm(N * T, 5, 10), to = c(0,10))), # V. cambiante en el tiempo
                ingresos = abs(rnorm(N * T, mean = 50, sd = 25)),               # V. cambiante en el tiempo
                apr_pres = 15 + 1.3 * sexo + 3 * ideol + 0.4 * ingresos + rnorm(n(), 0, 10)
  ) -> df_pooled

## B) Generar datos: DGP Efectos Fijos (Hay heterogeneidad individual correlacionada con X)
data.frame(id    = rep(1:N, each = T), # ID de los individuos
           time  = rep(1:T, N))  %>%   # Periodos de tiempo
  dplyr::group_by(id) %>%
   dplyr::mutate(alpha = unique(rnorm(N, 0, 8)[id]) ) %>% # Heterogeneidad individual
  dplyr::ungroup() %>%
  dplyr::mutate(sexo     = rep(ifelse( alpha > median(alpha), 1, 0)),                   # V. constante en el tiempo
                ideol    = round(scales::rescale( alpha + rnorm(N * T, 5, 10), to = c(0,10))), # V. cambiante en el tiempo
                ingresos = abs( alpha + rnorm(N * T, mean = 50, sd = 25)),              # V. cambiante en el tiempo
                apr_pres = 15 + 1.3 * sexo + 3 * ideol + 0.4 * ingresos + 1 * alpha + rnorm(n(), 0, 10)
  ) -> df_fe

## C) Generar datos: DGP Efectos Aleatorios (Hay heterogeneidad individual no correlacionada con X)
data.frame(id    = rep(1:N, each = T), # ID de los individuos
           time  = rep(1:T, N))  %>%   # Periodos de tiempo
  dplyr::group_by(id) %>%
   dplyr::mutate(alpha = unique(rnorm(N, 0, 8)[id]) ) %>% # Heterogeneidad individual
  dplyr::ungroup() %>%
  dplyr::mutate(sexo     = rep(sample(0:1, N, replace = TRUE), each = T),       # V. constante en el tiempo
                ideol    = round(scales::rescale(rnorm(N * T, 5, 10), to = c(0,10))), # V. cambiante en el tiempo
                ingresos = abs(rnorm(N * T, mean = 50, sd = 25)),               # V. cambiante en el tiempo
                apr_pres = 15 + 1.3 * sexo + 3 * ideol + 0.4 * ingresos + 1 * alpha + rnorm(n(), 0, 10)
  ) -> df_re
```

Simulación

- Simule tres bases de datos:

Datos	**Condición**
`df_pooled`	No hay Heterogeneidad Individual
`df_fe`	Hay Heterogeneidad Individual y está correlacionada con los regresores
`df_re`	Hay Heterogeneidad Individual y **NO** está correlacionada con los regresores

Cuadro 6: Datos Simulados

- En los tres casos, los **verdaderos** valores de los parámetros poblacionales son:
  - $\beta_{Sexo} = 1{,}3$
  - $\beta_{\text{Ideología}} = 3{,}0$
  - $\beta_{Ingresos} = 0{,}4$

Otros Supuestos

One-Way Error Component Model

$$Y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + \epsilon_{it}$$

- También es importante enfatizar los supuestos que trae este modelo:
  - No hay heterogeneidad según tiempo.
  - Los regresores capturan niveles y sus efectos sobre la variable de resultado son **contemporáneos** $\longrightarrow$ No hay efectos rezagados de regresores ni de la variable dependiente.
  - Los efectos de los regresores sobre la variable de resultado es el mismo para todas las unidades (y a lo largo de todo el tiempo) $\longrightarrow$ **Efecto Común**.

⋆ Al final del Workshop discutiremos extensiones para abordar estos aspectos.

Tabla de Contenidos

MCO con Datos Agrupados (Pooled OLS)

- La solución más obvia es usar el estimador de Mínimos Cuadrados Ordinarios para corte transversal con los datos de panel:

$$Y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + \epsilon_{it}$$

$$\widehat{Y_{it}} = \hat{\alpha}^{Pooled} + \sum_{k=1}^{K} \hat{\beta}^{Pooled} \mathbf{X}_{it}$$

- El estimador $\{\hat{\alpha}^{Pooled}, \hat{\beta}^{Pooled}\}$ corresponde a usar `lm(Y ~1 + ., data = df)` Apéndice F: Estimador
- El supuesto central para este estimador es **Exogeneidad Estricta**, que implica:

$$\mathbb{E}[\epsilon_{it} | X_{i1}, \ldots, X_{iT}] = 0, \forall i, t$$
$$\mathbb{E}[\alpha_i | X_{i1}, \ldots, X_{iT}] = 0, \forall i$$

- Que es lo mismo que decir $\mathbb{E}[\epsilon_{it} | X_{i1}, \ldots, X_{iT}, \alpha_i] = 0, \forall i, t$

Exogeneidad Estricta

Apéndice F: Supuestos

- Para cada punto en el tiempo (t), el valor esperado del error idiosincrático condicional a la heterogeneidad individual (α_i)y las covariantes (X) **en todos los puntos en el tiempo** es cero.
- Los regresores y los efectos individuales no tienen correlación con el término de error idiosincrático en **ningún periodo de tiempo**.
- Es decir, es mucho más fuerte que exogeneidad contemporánea $\mathbb{E}[\epsilon_{it} \mid X_{it}, \alpha_i] = 0$ que implica la independencia respecto a los valores contemporáneos de X_{it}, pero no necesariamente de sus valores pasados (lags) o futuros (leads).
- Tampoco permite efectos rezagados de la variable dependiente (Y_{it-1}).

$\star$ Sin embargo, no hemos dicho nada sobre la correlación entre α_i y $X_{i1}, \dots, X_{iT}$.

Lección 3

Exogeneidad Estricta es un Supuesto Clave para TODOS los Estimadores de Panel (y uno Muy Caro)

Aplicación de MCO Pooled

▸ Apéndice F: Insesgamiento

Código de R

```
## Estimación de modelos
m01_a <-        lm(apr_pres ~ sexo + ideol + ingresos,         data = df_pooled)
m01_b <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_pooled, index = c("id", "time"), model = "pooling")
m01_c <- plm::plm(apr_pres ~ sexo + ideol + ingresos + alpha, data = df_fe,     index = c("id", "time"), model = "pooling")
m01_d <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_fe,     index = c("id", "time"), model = "pooling")
m01_e <- plm::plm(apr_pres ~ sexo + ideol + ingresos + alpha, data = df_re,     index = c("id", "time"), model = "pooling")
m01_f <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_re,     index = c("id", "time"), model = "pooling")

## Valores de los Párametros: Intercepto = 15, Sexo = 1.3, Ideología = 3, Ingresos = 0.4

## Extraer coeficientes estimados
data.frame(Parametros = c(15, 1.3, 3, 0.4), LM_Pool  = coef(m01_a), PLM_Pool = coef(m01_b), PLM_FE_1 = coef(m01_c)[1:4],
           PLM_FE_2 = coef(m01_d), PLM_RE_1 = coef(m01_e)[1:4], PLM_RE_2 = coef(m01_f))
```

Parámetro	Valor	LM	PLM	PLM (FE 1)	PLM (FE 2)	PLM (RE 1)	PLM (RE 2)
(Intercepto)	15.0	14.618	14.618	15.272	-0.275	15.215	15.157
Sexo	1.3	1.117	1.117	1.041	11.216	1.231	1.557
Ideología	3.0	3.119	3.119	2.982	4.873	2.974	2.949
Ingresos	0.4	0.399	0.399	0.399	0.429	0.398	0.397

Cuadro 7: Estimación con Pooled MCO según Tipo de Datos

MCO con Datos Agrupados

- MCO con Datos Agrupados es una buena opción como estimador si:
  1. Se cumple Exogeneidad Estricta y **NO** hay heterogeneidad individual.
  2. Se cumple Exogeneidad Estricta y observamos la heterogeneidad individual (la incluimos como regresor en el modelo).
  3. Se cumple Exogeneidad Estricta y la heterogeneidad individual es independiente de los regresores incluidos en el modelo.

- Estas situaciones son poco probables, por lo que MCO con Datos Agrupados **NO** debería ser el estimador por defecto para lidiar con datos panel.

Inferencia Estadística

- Incluso en los escenarios en que el estimador MCO Pooled ($\hat{\beta}^{Pooled}$) es insesgado y consistente, tenemos problemas con el estimador de la varianza $\longrightarrow$ **Correlación Serial** ($\text{Cor}[\epsilon_{it}, \epsilon_{is}] \neq 0$).
- Ignorar la dependencia estadística lleva a sobre-estimar la precisión en la estimación (errores estándares demasiado pequeños).
- Es necesario corregir los errores estándares para realizar inferencia estadística.
- Utilizamos el estimador de la varianza robusto a la clusterización (*cluster-robust variance estimator*).
- Hay una gran variedad de procedimientos alternativos para crear estándares robustos a la clusterización. Relativo consenso que con $N > 50$, los problemas de sesgo por el número de unidades son menores y estos detalles técnicos se vuelven irrelevantes $\longrightarrow$ usar las opciones por default del software.

Tabla de Contenidos

Primeras Diferencias

- ¿Qué podemos hacer si el *DGP* es un modelo con heterogeneidad individual **no observada**?

  ⋆ Usar un estimador capaz de lidiar con $\alpha_i \longrightarrow$ *Transformación de datos*.

- La transformación más simple es la **diferenciación**. Lo ejemplificaremos con un panel con dos mediciones:

$$\begin{aligned} t = 1, \qquad & Y_{i1} = \beta_0 + \beta_1 X_{i1} + \alpha_i + \epsilon_{i1} \\ t = 2, \qquad & Y_{i2} = \beta_0 + \beta_1 X_{i2} + \alpha_i + \epsilon_{i2} \\ (Y_{i2} - Y_{i1}) &= (\beta_0 - \beta_0) + \beta_1 (X_{i2} - X_{i1}) + (\alpha_i - \alpha_i) + (\epsilon_{i2} - \epsilon_{i1}) \\ \Delta Y_i &= 0 + \beta_1 \Delta X_i + 0 + \Delta\epsilon_i \end{aligned}$$

- El estimador de **Primeras Diferencias** (*First Differences* o FD) corresponde a usar MCO con datos diferenciados.

  ⋆ Pasamos de un modelo de niveles (Y_{it}) a un modelo de cambios (ΔY_i).

Primeras Diferencias

- Ahora, pensemos en un caso más general con t mediciones y variables constantes en el tiempo (Z) y cambiantes en el tiempo (X). En este caso, a cada medición le restaré su versión **rezagada** $(t-1)$:

$$Y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + \sum_{l=1}^{L} \gamma_l \mathbf{Z}_i + \epsilon_{it}$$

$$(Y_{it} - Y_{it-1}) = (\alpha_i - \alpha_i) + \sum_{k=1}^{K} \beta_k (\mathbf{X}_{it} - \mathbf{X}_{it-1}) + \sum_{l=1}^{L} \gamma_l (\mathbf{Z}_i - \mathbf{Z}_i) + (\epsilon_{it} - \epsilon_{it-1})$$

$$\Delta Y_{it} = \sum_{k=1}^{K} \beta_k \Delta \mathbf{X}_{it} + \Delta \epsilon_{it}$$

- El Estimador de **Primeras Diferencias** $\hat{\beta}^{FD}$ corresponde a usar la función `plm` con el argumento `model = «fd»`. ▸ Apéndice G: Estimador

- El estimador es insesgado y consistente bajo **Exogeneidad Estricta**. Sin embargo, el uso de MCO no lo hace óptimo en términos de eficiencia. Hay estimadores alternativos que logran lo mismo sin sacrificar tanta precisión. ▸ Apéndice G: Supuestos

Aplicación de Primeras Diferencias

▸ Apéndice G: Insesgamiento

Código de R

```
## Estimación de modelos
m02_a <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_pooled, index = c("id", "time"), model = "fd")
m02_b <- plm::plm(apr_pres ~ sexo + ideol + ingresos + alpha, data = df_fe,     index = c("id", "time"), model = "fd")
m02_c <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_fe,     index = c("id", "time"), model = "fd")
m02_d <- plm::plm(apr_pres ~ sexo + ideol + ingresos + alpha, data = df_re,     index = c("id", "time"), model = "fd")
m02_e <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_re,     index = c("id", "time"), model = "fd")

## Valores de los Párametros: Intercepto = 15, Sexo = 1.3, Ideología = 3, Ingresos = 0.4

## Extraer coeficientes estimados
data.frame(Verdad = c(15, 1.3, 3, 0.4),
           PLM_Pool = c(coef(m02_a)[1],NA,coef(m02_a)[2:3]),
           PLM_FE_1 = c(coef(m02_b)[1],NA,coef(m02_b)[2:3]),
           PLM_FE_2 = c(coef(m02_c)[1],NA,coef(m02_c)[2:3]),
           PLM_RE_1 = c(coef(m02_d)[1],NA,coef(m02_d)[2:3]),
           PLM_RE_2 = c(coef(m02_e)[1],NA,coef(m02_e)[2:3]))
```

Parámetro	Valor	PLM	PLM (FE 1)	PLM (FE 2)	PLM (RE 1)	PLM (RE 2)
Intercepto	15.0	-0.018	-0.017	-0.017	-0.006	-0.006
Sexo	1.3	NA	NA	NA	NA	NA
Ideología	3.0	3.123	2.979	2.979	2.950	2.950
Ingresos	0.4	0.400	0.398	0.398	0.397	0.397

Cuadro 8: Estimación con FD según Tipo de Datos

Primeras Diferencias: Desafíos

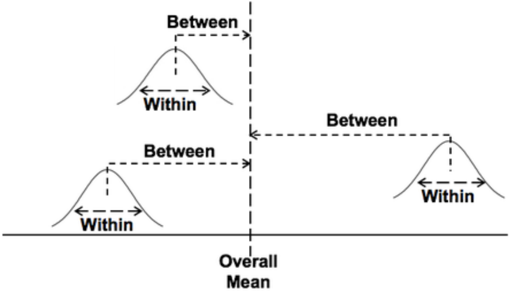
- El uso del estimador de *Primeras Diferencias* (*First Differences* o FD) implica una pérdida de **grados de libertad** equivalente al número de observaciones (eliminamos la primera medición de cada unidad).
- FD remueve las variables explicativas constantes en el tiempo (Z).
  - ¿Tiene sentido argumentar que el cambio (ΔY) depende de un fenómeno constante en el tiempo?
- Otro desafío es el manejo de la **información faltante**:
  - Los modelos de regresión comúnmente usan *listwise deletion*.
  - *FD* no puede usar observaciones con sólo una medición, ni tampoco casos con brechas temporales.
  - Poco útil con *paneles desbalanceados*.

Tabla de Contenidos

Introducción

- De forma general, siempre es posible realizar la siguiente descomposición:

$$\overbrace{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y})^2}^{\text{Total}} = \overbrace{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y}_i)^2}^{\text{Within}} + \overbrace{\sum_{i=1}^{N}T(\bar{y}_i-\bar{y})^2}^{\text{Between}}$$



Introducción

- Con datos panel tenemos dos tipos de variación: a lo largo del tiempo y entre las unidades.
  - **Variación Between**: diferencias constantes entre las unidades.
  - **Variación Within:** cambios a lo largo del tiempo.
- Si sólo estoy interesado en la variación de corte transversal (between), me basta con calcular el promedio por unidad:

$$
\begin{aligned}
\bar{y}_i &= \alpha_i + \bar{\mathbf{x}}_i^t \beta + (\alpha_i - \alpha + \bar{\epsilon}_i) \\
\frac{\sum_t y_{it}}{T} &= \alpha + \frac{\sum_t \mathbf{x}_{it}}{T}\beta + (\alpha_i - \alpha + \frac{\sum_t \epsilon_{it}}{T}
\end{aligned}
$$

Para $i = 1, \ldots, N$.

Introducción

- De forma complementaria, puedo enfocarme únicamente en la variación temporal (within): desviaciones de variables de sus promedios por año (time-averaged values).

$$y_{it} - \bar{y}_i = (\alpha_i - \alpha_i) + (\mathbf{x}_{it} - \mathbf{x}_i)^T \beta + (\epsilon_{it} - \bar{\epsilon}_i)$$

Para $i = 1, \ldots, N, t = 1, \ldots, T$.

- Esta es la lógica básica del estimador de **Efectos Fijos**: eliminar la heterogeneidad individual no observada (α_i) al transformar los datos como desviaciones de las medias por unidad (*demeaning*). Apéndice H: Estimador
- Este estimador también es conocido como el estimador **within**, ya que transforma los datos de modo de remover la variación de corte transversal (o between unidades) y utiliza sólo la información sobre los cambios al interior de las unidades.

Estimador de Efectos Fijos

- El estimador **within** es MCO con datos transformados (sustracción de media grupal $\longrightarrow$ Observaciones son desviaciones respecto al promedio de la unidad).
- El estimador $\hat{\beta}^{FE} = (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}(\tilde{\mathbf{X}}^T\tilde{\mathbf{Y}})$ corresponde a usar la función `plm` con el argumento `model = «within»`.
- El estimador es insesgado y consistente bajo **Exogeneidad Estricta**.
- Para derivar otras propiedades relevantes del modelo necesitamos los otros supuestos de MCO: linealidad en parámetros, no multicolinealidad perfecta, homocedasticidad y ausencia de correlación serial. Apéndice H: Supuestos
- Descartamos la variación between para obtener un modelo más robusto (en el sentido de que funciona para un mayor número de DGPs).
- **NO** necesito ningún supuesto sobre la relación entre α_i y $\mathbf{X}$. La heterogeneidad individual puede estar correlacionada con los regresores.

Estimador de Efectos Fijos

- Es posible estimar el modelo de dos maneras:

  ⋆ **Time Demeaning**: Para cada observación, calcular el promedio grupal. Luego, se calcula la diferencia entre el valor contemporáneo y estev promedio.

  ⋆ **Least Squares Dummy Variables**: utilizar el estimador tradicional y usar la variable que identifica los grupos como *dummy*.

- Ambas opciones son algebraicamente equivalentes y tienen las mismas propiedades. Sin embargo, es necesario tener en cuenta:
  - Calcular manualmente el estimador within requiere una corrección de grados de libertad (De $NT - K$ pasamos a $NT - N - K$.
  - Usar LSDV puede ser computacionalmente intensivo. Si tienes millones de efectos individuales, es recomendable usar otros paquetes como `fixest`.
- Con $T = 2$, el estimador de Efectos Fijos es idéntico al estimador de Primeras Diferencias. ▸ Apéndice H: Equivalencia

¿Qué son los Efectos Fijos?

- En el modelo de componentes del error, los efectos fijos (estimaciones de α_i) fueron derivadas del término de error compuesto.
- Capturan toda la heterogeneidad individual (constante en el tiempo). Por lo tanto, no es posible incluir en el modelo variables constantes en el tiempo (Z), ya que serán **colineales** con los efectos fijos.
- Los efectos fijos son **interceptos variables** (uno por cada unidad en el panel). Por lo tanto, el modelo resultante no tiene un intercepto global.
- En el estimador **within**, el estimador α_i es inconsistente si $N \to \infty$ y T fijo (en paneles cortos).
  - ⋆ Problema de Parámetros Incidentales
- Esto no es un problema serio, los efectos fijos (interceptos por unidad) comúnmente **NO** son la cantidad de interés científico.

Tabla de Contenidos

Efectos Fijos y Errores Estándares

- Incluso tras controlar por α_i la correlación serial puede persistir.
- A su vez, la transformación **within** no resuelve los potenciales problemas de heterocedasticidad.
- Por lo tanto, siempre es recomendable explorar las propiedades de los errores estándares:
  1. Usar errores estándares robustos a la heterocedasticidad o robustos y clusterizados.
  2. Utilizar block bootstraping u otra alternativa para derivar errores estándares.
  3. Usar Mínimos Cuadrados Generalizados Factibles.

Aplicación de Efectos Fijos

▸ Apéndice H: Insesgamiento

Código de R

```
## Estimación de modelos
m03_a <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_pooled, index = c("id", "time"), model = "within", effect
m03_b <- plm::plm(apr_pres ~ sexo + ideol + ingresos + alpha, data = df_fe,     index = c("id", "time"), model = "within", effect
m03_c <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_fe,     index = c("id", "time"), model = "within", effect
m03_d <- plm::plm(apr_pres ~ sexo + ideol + ingresos + alpha, data = df_re,     index = c("id", "time"), model = "within", effect
m03_e <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_re,     index = c("id", "time"), model = "within", effect

## Valores de los Párametros: Intercepto = 15, Sexo = 1.3, Ideología = 3, Ingresos = 0.4

## Extraer coeficientes estimados
data.frame(Verdad = c(15, 1.3, 3, 0.4),
           PLM_Pool = c(NA,NA,coef(m03_a)),
           PLM_FE_1 = c(NA,NA,coef(m03_b)),
           PLM_FE_2 = c(NA,NA,coef(m03_c)),
           PLM_RE_1 = c(NA,NA,coef(m03_d)),
           PLM_RE_2 = c(NA,NA,coef(m03_e)))
```

Parámetro	Valor	PLM	PLM (FE 1)	PLM (FE 2)	PLM (RE 1)	PLM (RE 2)
Intercepto	15.0	NA	NA	NA	NA	NA
Sexo	1.3	NA	NA	NA	NA	NA
Ideología	3.0	3.103	2.962	2.962	2.961	2.961
Ingresos	0.4	0.399	0.400	0.400	0.399	0.399

Cuadro 9: Estimación con Within Estimator según Tipo de Datos

Tabla de Contenidos

Efectos Aleatorios

- Efectos Aleatorios es otra forma de concebir los interceptos variantes:

$$
\begin{aligned}
y_{it} &= \alpha_i + \mathbf{x}_{it}^T \beta + \epsilon_{it} \\
\alpha_i &\sim \mathcal{N}(\alpha, \sigma_\alpha^2) \\
\epsilon_{it} &\sim \mathcal{N}(0, \sigma_\epsilon^2)
\end{aligned}
$$

- No sólo asumo una distribución para la heterogeneidad individual, también asumo que es **independiente** de **X**.
- El estimador $\hat{\beta}^{RE}$ corresponde a usar la función `plm` con el argumento `model = «random»`. ▸ Apéndice I: Estimador
- Al igual que en los casos anteriores, necesito **Exogeneidad Estricta** junto a los supuestos de distribución de α_i. ▸ Apéndice I: Supuestos
- Al incluir más supuestos paramétricos logro explotar la variación between y within de forma simultánea ⟶ Ganancias en **Eficiencia**.

Estimador de Efectos Aleatorios

- La estimación eficiente de $\hat{\beta}^{RE}$ usa **Mínimos Cuadrados Generalizados**:

1. **Estimar un Modelo Pooled OLS para Obtener los Residuales**
   - Estimar el modelo agrupado (pooled OLS) ignorando la estructura de panel:

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it}$$

   - Obtener los residuales $\hat{u}_{it}$ del modelo estimado.

2. **Estimar los Componentes de Varianza**
   - Calcular las varianzas dentro y entre grupos usando los residuales:

$$\hat{\sigma_\epsilon}^2 = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2$$

$$\hat{\sigma_\alpha}^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} \hat{u}_{it} \right)^2$$

Pasos para Estimar un Modelo de Interceptos Aleatorios

3. **Usar $\hat{\sigma_\epsilon}^2$ y $\hat{\sigma_\alpha}^2$ para Estimar Ω_τ**
   - Construir la matriz de varianzas-covarianzas Ω_τ:

$$\Omega_\tau = \hat{\sigma_\epsilon}^2 I_{NT} + \hat{\sigma_\alpha}^2 J_N \otimes I_T$$

   donde I_{NT} es la matriz identidad de tamaño NT y J_N es una matriz de unos de tamaño $N \times N$.

4. **Actualizar los Coeficientes por FGLS**
   - Reestimar los coeficientes usando FGLS:

$$\widehat{\beta}^{RE} = \left(X'\Omega_\tau^{-1}X\right)^{-1} X'\Omega_\tau^{-1}y$$

$$\hat{\alpha} = \bar{y} - \widehat{\beta}^{RE}\bar{\mathbf{X}}$$

   - α_i son tratados como un «error» y no son estimados directamente.

Aplicación de Efectos Aleatorios

Apéndice I: Insesgamiento

Código de R

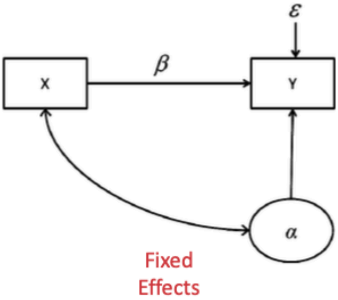
```
## Estimación de modelos
m04_a <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_pooled, index = c("id", "time"), model = "random", effect
m04_b <- plm::plm(apr_pres ~ sexo + ideol + ingresos + alpha, data = df_fe,     index = c("id", "time"), model = "random", effect
m04_c <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_fe,     index = c("id", "time"), model = "random", effect
m04_d <- plm::plm(apr_pres ~ sexo + ideol + ingresos + alpha, data = df_re,     index = c("id", "time"), model = "random", effect
m04_e <- plm::plm(apr_pres ~ sexo + ideol + ingresos,         data = df_re,     index = c("id", "time"), model = "random", effect

## Valores de los Párametros: Intercepto = 15, Sexo = 1.3, Ideología = 3, Ingresos = 0.4

## Extraer coeficientes estimados
data.frame(Verdad = c(15, 1.3, 3, 0.4),
           PLM_Pool = coef(m04_a)[1:4],
           PLM_FE_1 = coef(m04_b)[1:4],
           PLM_FE_2 = coef(m04_c)[1:4],
           PLM_RE_1 = coef(m04_d)[1:4],
           PLM_RE_2 = coef(m04_e)[1:4] )
```

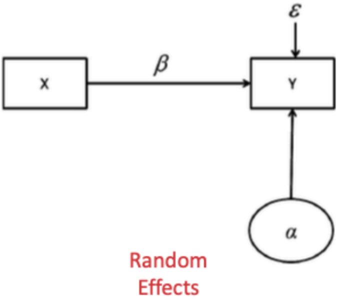
Parámetro	Valor	PLM	PLM (FE 1)	PLM (FE 2)	PLM (RE 1)	PLM (RE 2)
Intercepto	15.0	14.618	15.274	2.140	15.216	15.064
Sexo	1.3	1.117	1.041	11.881	1.231	1.557
Ideología	3.0	3.119	2.981	4.369	2.974	2.957
Ingresos	0.4	0.399	0.399	0.421	0.398	0.398

Cuadro 10: Estimación con FGLS (RE) Estimator según Tipo de Datos

Tabla de Contenidos

Efectos Fijos y Efectos Aleatorios





Efectos Aleatorios

- El estimador de efectos aleatorios añade el supuesto de independencia de los regresores con la heterogeneidad individual.
- Es un supuesto poco realista, pero permite no limitarnos a la variación within (*quasi-demeaned data* o *partial pooling*).
- En la práctica, esto nos permite aprovechar ambas fuentes de variación y estimar los efectos de variables invariantes en el tiempo.
- Hay mucha discusión académica sobre cómo elegir entre uno y otro estimador:
  1. Test de Hausman.
  2. Justificación teórica sobre tipo de panel: en paneles largos, tiene más sentido Efectos Fijos ya que las unidades son únicas. En paneles cortos, es más plausible una distribución normal de efectos aleatorios.

  ⋆ **Mi recomendación** es seguir pautas disciplinares, pero siempre comenzar con **Efectos Fijos** para compararlo con los **Efectos Aleatorios**.

Efectos Aleatorios: Balance Between y Within

- El modelo de efectos aleatorios puede verse como una combinación de los estimadores de efectos entre (between) y efectos dentro (within).
- El estimador de efectos aleatorios combina los estimadores entre (between) y dentro (within) ponderando por la varianza de los componentes del error:

$$\hat{\beta}_{RE} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y$$

- La matriz Ω incorpora la varianza de los errores idiosincráticos y específicos del individuo.
- Es una media ponderada del estimador entre y dentro:

$$\hat{\beta}^{RE} = \lambda\hat{\beta}^{FE} + (1-\lambda)\hat{\beta}^{Between}$$

- Donde λ es una ponderación basada en la varianza de los componentes de error, y toma valores entre cero y uno: $\lambda = 1 - \sqrt{\frac{\sigma_\epsilon^2}{(\sigma_\epsilon^2 + T\sigma_\alpha^2}})$

Efectos Aleatorios: Balance Between y Within

Apéndice I: Comparación

- Esto deriva en un modelo de la forma:

$$y_{it} - \lambda \bar{y}_i = \alpha(1 - \lambda) + (\mathbf{x}_{it} - \lambda \mathbf{x}_i)^T \beta + (\epsilon_{it} - \lambda \bar{\epsilon}_i)$$

- λ se mueve entre 0 y 1 y puede ser entendido como un modo de ponderar la información (qué proporción de la variación between voy a preservar en mi análisis):
  - Si $\lambda = 1$ es el Estimador de Efectos Fijos (Within).
  - Si $\lambda = 0$ es el Estimador MCO con Datos Agrupados.

Tabla de Contenidos

Test de Hausman para Efectos Aleatorios

- El test de Hausman compara los estimadores de efectos fijos y efectos aleatorios.
- Hipótesis:
  - H_0: Los efectos aleatorios son consistentes y eficientes.
- Estadístico del test:

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})^T \left(\text{Var}(\hat{\beta}_{FE}) - \text{Var}(\hat{\beta}_{RE})\right)^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE})$$

- Si el valor de H es grande, rechazamos H_0 y preferimos otro estimador.
- Interpretación:
  - Un valor significativo del test indica que los efectos específicos están correlacionados con las variables explicativas, invalidando el uso de efectos aleatorios.

⋆ La hipótesis nula es que *Efectos Aleatorios* es correcto. Pero fallar en rechazar la nula no implica que sea el estimador adecuado para el DGP (ambos FE y RE podrían ser incorrectos).

Test de Hausman para Efectos Aleatorios

Código de R

```
# Test de Comparación de Modelos: Efectos Fijos vs Efectos Aleatorios

## Test de Hausman
plm::phtest(x = m03_a, x2 = m04_a) # Datos simulados SIN heterogeneidad individual
plm::phtest(x = m03_c, x2 = m04_c) # Datos simulados CON heterogeneidad individual Correlacionada con X
plm::phtest(x = m03_e, x2 = m04_e) # Datos simulados CON heterogeneidad individual Independiente de X
```

- Hay otros tests estadísticos para evaluar la pertinencia del estimador de efectos aleatorios.
  - Test de Chamberlain.
  - Test de Angrist y Newey.

- Sin importar el test utilizado, ninguna prueba estadística es suficiente para justificar qué estimador usar. Estas pruebas sólo te permiten concluir (si se rechaza la hipótesis nula del test) que, en una determinada aplicación **no es adecuado usar efectos aleatorios**.

  ⋆ Para profundizar sobre pruebas estadísticas, consultar el Capítulo 3 de (Croissant y Millo, 2018).

Lección 4

No hay una forma mecánica para elegir entre Efectos Fijos (Within) y Efectos Aleatorios (Within + Between)

Tabla de Contenidos

Corrección de Errores Estándares

- Los estimadores que hemos discutido no han sido diseñados para impedir que los errores en el modelo estén correlacionados (ni tampoco para resolver otros problemas como la heterocedasticidad).

- Es probable, que al usar un modelo que controla por la heterogeneidad individual, la correlación serial disminuya.
  - ⋆ Sin embargo los estimadores (Efectos Fijos, Efectos Aleatorios, etc.) asumen que es cero ($Cov[\epsilon_{it}, \epsilon_{is} | \mathbf{X}_{it}, \mathbf{Z}_i, \alpha_i] = 0, t \neq s$).

- Por lo tanto, es recomendable corregir los errores estándares resultantes, reemplazándolos por **errores estándares robustos**.
  - ⋆ En el contexto panel, las observaciones están clusterizadas a nivel de unidad.

- Resolveremos el problema de la correlación serial al tener errores estándares robustos a la dependencia al interior de un cluster.

Corrección de Errores Estándares

- La discusión sobre cómo calcular errores estándares robustos es extensa (Ver (Abadie et al., 2023) y (MacKinnon et al., 2023) para resúmenes recientes; Zeileis et al., 2020 es una introducción accesible sobre los errores estándares en R).
- Desde una perspectiva práctica:
  - Necesitas al menos 50 unidades para que funcionen (algo frecuente en paneles cortos).
  - Usa las opciones por defecto de R.
  - No asumas que resuelven todos los problemas de un modelo.

Código de R

```
# B) Corrección de Errores Estándares
lmtest::coeftest(m01_b)

## Lo más sencillo: Robustez a Heterocedasticidad
lmtest::coeftest(m01_b, vcov = plm::vcovHC(m01_b, type = "HC1", method = "arellano"))

## Sin embargo, no es suficiente. Dependencia estadística es clusterizada
lmtest::coeftest(m01_b, vcov = plm::vcovHC(m01_b, type = "HC1", method = "arellano", cluster = "group"))

## Esto funciona con cualquier tipo de estimador
lmtest::coeftest(m02_c, vcov = plm::vcovHC(m02_c, type = "HC1", method = "arellano", cluster = "group"))
```

Lección 5

Es importante examinar y corregir la dependencia en los Errores Estándares

Tabla de Contenidos

Otras Pruebas Estadísticas

- Todos los estimadores que discutimos son modelos de regresión (y la mayoría utiliza MCO). Por lo tanto, puedes usar pruebas t, pruebas F, ANOVA, etc. de forma tradicional.
  ⋆ Es conveniente usar errores estándares robustos en el análisis.
- También hay tests estadísticos que son específicos para la investigación aplicada con datos panel.

1. **Presencia de Heterogeneidad Individual**
   - **Prueba F:** comparar un modelo estimado con Efectos Fijos con un modelo estimado con MCO y Datos Agrupados.
   - **Prueba Multiplicador Lagrange** usa los residuos de un modelo estimado con MCO Agrupado.
     ⋆ En ambos casos, **la hipótesis nula es la ausencia de heterogeneidad individual**. Rechazar H_0 implica que es preferible Efectos Fijos/ Primeras Diferencias/ Efectos Aleatorios en vez de MCO con Datos Agrupados.

Otras Pruebas Estadísticas

Código de R
```
# C1) Test de Comparación de Modelos: Incluir Efectos Fijos? (versus Pooled MCO)

## Test F para Heterogeneidad Individual o de Tiempo
plm::pFtest(x = m03_a, z = m01_b) # Datos simulados SIN heterogeneidad individual
plm::pFtest(x = m03_c, z = m01_d) # Datos simulados CON heterogeneidad individual

## Test de Multiplicador de Lagrange para Heterogeneidad Individual o de Tiempo
plm::plmtest(x = m01_b) # Datos simulados SIN heterogeneidad individual
plm::plmtest(x = m01_d) # Datos simulados CON heterogeneidad individual

## Test de Combinalidad/Poolability de los Datos
plm::pooltest(x = m01_b, z = m03_a) # Datos simulados SIN heterogeneidad individual
plm::pooltest(x = m01_d, z = m03_c) # Datos simulados CON heterogeneidad individual
```

2. **Presencia de Correlación Serial:** Los estimadores analizados no garantizan la eliminación de la correlación serial.
   - Si hay correlación serial, nuestras estimaciones de los errores estándares (y por lo tanto, las pruebas de significancia estadística) estarán **sesgadas**.
   - Dado lo anterior, hay muchas pruebas diseñadas específicamente para identificar correlación serial. Las diferencias se deben al tipo de estimador:
     - `pwtest` para MCO Pooled.
     - `pwartest` para Efectos Fijos.
     - `pbgtest` y `pdwtest` para todo tipo de estimadores.

Otras Pruebas Estadísticas

- Para todas estas pruebas, la **hipótesis nula es la ausencia de correlación serial**. Rechazar H_0 implica que es necesario corregir los errores estándares y hacerlos robustos a la correlación serial (*errores estándares clusterizados*).

Código de R

```
# C2) Test de Correlación Serial

## Para MCO Pooled
plm::pwtest(x = m01_b, effect = "individual") # Datos simulados SIN heterogeneidad individual
plm::pwtest(x = m01_b, effect = "time")       # Datos simulados SIN heterogeneidad individual

plm::pwtest(x = m01_d, effect = "individual") # Datos simulados CON heterogeneidad individual
plm::pwtest(x = m01_d, effect = "time")       # Datos simulados CON heterogeneidad individual

### Para Efectos Fijos
plm::pwartest(x = m03_a) # Datos simulados SIN heterogeneidad individual
plm::pwartest(x = m03_c) # Datos simulados CON heterogeneidad individual
plm::pwartest(x = m03_e) # Datos simulados CON heterogeneidad individual

## Para Primeras Diferencias
plm::pwfdtest(x = m02_a, h0  = "fd") # Datos simulados SIN heterogeneidad individual
plm::pwfdtest(x = m02_c, h0  = "fd") # Datos simulados CON heterogeneidad individual
plm::pwfdtest(x = m02_e, h0  = "fd") # Datos simulados CON heterogeneidad individual
```

3. **Presencia de Heterocedasticidad** se pueden usar las pruebas tradicionales como el test de Breusch-Pagan (`lmtest::bptest`).

Tabla de Contenidos

Heterogeneidad Temporal

- Hasta el momento, sólo hemos discutido sobre heterogeneidad individual (α_i). Sin embargo, también es posible utilizar las técnicas de panel para modelar **heterogeneidad temporal**.

- Para esto, volvamos a la expresión del modelo de componentes del error:

$$Y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + \epsilon_{it}$$

- Sin ningún problema podemos incorporar el tiempo como un regresor en este momento. Por ejemplo, podemos definir t $= 1, \dots, T$ como un contandor de tiempo continuo:

$$Y_{it} = \alpha_i + \gamma t + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + \epsilon_{it}$$

Heterogeneidad Temporal

- Este modelo asume un efecto lineal de la **tendencia temporal** (*time trend*). En otras palabras, captura el efecto promedio (y común a todas las unidades) del aumento de 1 unidad en el tiempo sobre la variable de resultando, manteniendo todo lo demás constante.

- Es posible concebir otras formas funcionales para el efecto del tiempo sobre la variable de resultado:
  - **Efecto Cuadrático:** $\gamma_1 t + \gamma_2 t^2$.
  - **Efecto Exponencial:** $\exp(t)$. Esto sólo es conveniente si el outcome es un logaritmo.
  - **Efecto Discontinuo del tiempo:** $\gamma_1 d_1 + \cdots + \gamma_l d_{T-1}$.

- Este último caso, corresponde a incorporar $T-1$ variables dummies de tiempo (se excluye una para evitar colinealidad perfecta). En este caso, estamos asumiendo que el efecto del tiempo es distinto para cada ola (no hay una tendencia como hay en el caso lineal).

Heterogeneidad Temporal

- El modelo con efectos discontinuos del tiempo es una buena manera de motivar un *DGP* más complejo.
- En el contexto de Efectos Fijos demostramos que incluir dummies de las unidades es lo mismo que la transformación *within*. Por lo tanto, podemos concebir una forma más general del **modelo de componentes del error**:

$$Y_{it} = \alpha + \sum_{k=1}^{K} \beta_k \mathbf{X}_{it} + u_{it}$$

- Volvemos a descomponer el error compuesto:

$$u_{it} = \delta_i + \gamma_t + \epsilon_{it}$$

- Dónde tenemos heterogeneidad individual (α_i), heterogeneidad temporal (γ_t) y un error idiosincrático (ϵ_{it}).

Heterogeneidad Temporal

- Es decir, ahora reconocemos que hay una fuente de heterogeneidad que son atributos fijos de las unidades (α_i) y otra fuente de heterogeneidad que es común a todas las unidades en cada punto del tiempo (γ_t).

- Este DGP es conocido como **Two-Ways Error Component Model**. Todos los estimadores que hemos discutido en este Workshop pueden ser adaptados para incluir ambas fuentes de heterogeneidad.

- Para hacerlo, sólo es necesario cambiar el argumento `effect`:
  - Sólo Heterogeneidad Individual ⟶ `effect = «individual»`.
  - Heterogeneidad Individual y Temporal ⟶ `effect = «twoways»`.
- El controlar por heterogeneidad individual y temporal es bastante frecuente entre investigadores que prefieren el estimador de **efectos fijos**.

Código de R
```
## Estimación de modelos
m03_a <- plm::plm(apr_pres ~ sexo + ideol + ingresos, data = df_pooled, index = c("id", "time"),
                  model = "within", effect = "twoways")
```

Tabla de Contenidos

Inclusión de Otro Tipo de Regresores

- En este workshop discutimos tres tipos de regresores:
  - Variables constantes en el tiempo ($\mathbf{Z}_i$).
  - Variables variables en el tiempo ($\mathbf{X}_{it}$).
  - Variables que capturan tendencias temporales (t o un conjunto de dummies $\sum d_l$).
- Hasta ahora, hemos asumido que el efecto de las variables es contemporáneo.
- Sin embargo, no hay ninguna razón para sostener este supuesto. Es posible modelar efectos **rezagados** de variables cambiantes en el tiempo.
  - En vez de incluir el efecto de $\mathbf{X}_{it}$, puedes incluir el efecto rezagado en un tiempo $\mathbf{X}_{it-1}$.
  - También es posible incluir ambos de forma simultánea ($\mathbf{X}_{it}$ y $\mathbf{X}_{it-1}$). Esto implica asumir que hay un efecto contemporáneo (el nivel de la variable $\mathbf{X}_{it}$ t sobre Y_{it}) pero que dicho nivel sigue operando en el siguiente período (sin importar el valor que tome $\mathbf{X}_{it}$ en este nuevo período).

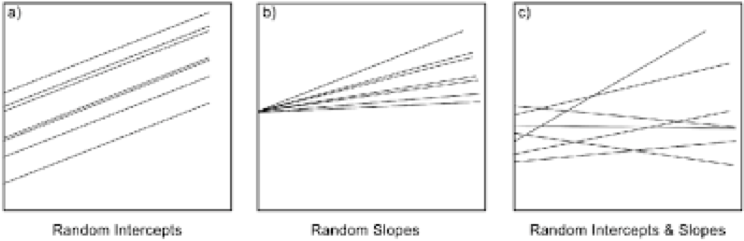
Inclusión de Otro Tipo de Regresores

- Puedes crear variables rezagadas con la función `plm::lag`.
- Es raro encontrar casos aplicados de efectos rezagados mayores a $t-1$ (una medición de rezago). De todos modos, es posible incluirlos si crees que tu efecto es muy persistente.
- En la mayoría de aplicaciones, no tiene mucho sentido incluir efectos futuros (leads).
- Los estimadores discutidos en esta sesión **NO** permiten la inclusión de variables dependientes rezagadas (Y_{it-1}). Su inclusión redunda en que el estimador es sesgado e inconsistente.
  ⋆ Hay estimadores específicos para lidiar con este tipo de especificaciones con **efectos dinámicos**.
  · El Capítulo 7 de (Croissant y Millo, 2018) trata sobre Modelos Dinámicos.

Tabla de Contenidos

Modelamiento Multinivel

- El Workshop se enfocó en presentar estimadores que modelan la heterogeneidad individual y/o temporal.
- Sin embargo, en todos los casos asumimos que los efectos de las variables (*slopes*) son constantes.
- Existe una estrategia de modelamiento llamada **«Modelamiento Multinivel/Mixto»** que permite estimar efectos de variables que varían según unidad y/o tiempo.



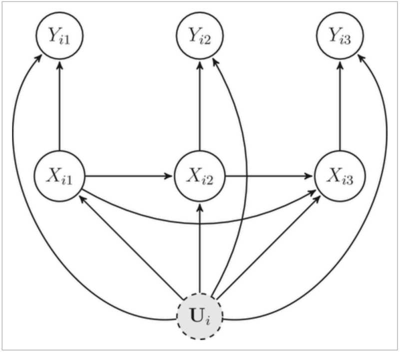
Causalidad con Modelamiento Panel

- Los datos panel son muy populares dentro de la *«design-based inference»* inspirada en el modelo de **Resultados Potenciales** (Imbens y Rubin, 2014).
- Imai y Kim (2019) proponen cómo formalizar con un DAG («directed acyclical graph») causal el análisis con **Efectos Fijos** (FE de Unidad).

- 3 Mediciones: $t = 1$, $t = 2$ y $t = 3$.

- Variable de resultado (Y_{it}), tratamiento (X_{it}) y confusor (no observado) invariante en el tiempo (U_i).

- En un DAG, la **ausencia** de variables o fechas son los supuestos.



Causalidad con Modelamiento Panel

- Por lo tanto, para identificar un efecto causal usando **Efectos Fijos** de Unidad, asumo:

1. No hay un confusor variable en el tiempo.
2. Valores pasados del outcome no afectan directamente el valor presente del outcome.
3. Valores pasados del outcome no afectan directamente el valor presente del tratamiento.
4. Valores pasados del tratamiento no afectan directamente el valor presente del outcome.

⋆ La lección es que la **interpretación causal** de los resultados obtenidos con el estimador Within depende de **MUCHOS** supuestos.

⇝ El artículo reseña qué se puede hacer para relajar estos cuatro supuestos.

Causalidad con Modelamiento Panel

- De forma similar, Imai y Kim (2020) demuestran que el uso de **Two-Way Fixed Effects** (TWFE, Efectos Fijos de Unidad y Tiempo) también realiza muchos supuestos (incluyendo linealidad) para aislar un efecto causal.
- Por último, también es posible darle una interpretación causal al estimador de **Efectos Aleatorios** (RE). Sin embargo, es importante tener en cuenta:
  1. Al igual que en el caso de FE, los supuestos necesarios para identificar un efecto causal no son triviales. Esto incluye la ausencia de **Confusores no observados variantes en el tiempo**.
  2. Incluso si se cumple el supuesto de independencia de **X** y la heterogeneidad individual (α_i), el estimador RE no aísla un efecto causal (*confounding*).
     ⋆ Para resolver estos problemas es necesario aplicar un *«bias-correction procedure»* y calcular errores estándares clusterizados (detalles en (Hazlett y Wainstein, 2020).

Tabla de Contenidos

Resumen: General

- Los datos de panel proveen la oportunidad de estudiar el cambio individual.
- Hay características de los datos panel que merecen nuestra atención:
  - Correlación Serial.
  - Heterogeneidad Individual.
- En este workshop discutimos distintos estimadores para manejar la heterogeneidad individual:
  - MCO con Datos Agrupados.
  - Primeras Diferencias.
  - Efectos Fijos (Within).
  - Efectos Aleatorios
- En todos los casos necesitamos importantes supuestos sobre los datos. El más importante de ellos es **Exogeneidad Estricta**.
- Los cuatro estimadores asumen Exogeneidad Estricta, a lo que se suman otros detalles según el caso.

Resumen: MCO Pooled y Primeras Diferencias

- MCO con Datos Agrupados (Pooled) corresponde a utilizar el estimador MCO ignorando la naturaleza panel de los datos.
  - ⋆ Se asume que las unidades son homogéneas, por lo que podemos combinarlas (*pooling*) en una única muestra.
  - ⋆ Esto implica un intercepto común a todas las unidades.
- El modelo MCO Pooled sólo es adecuado si el DGP no incluye heterogeneidad individual. Incluso si no hay heterogeneidad individual, necesitamos corregir los errores estándares (correlación serial) $\longrightarrow$ Errores Estándares clusterizados.

Resumen: Efectos Fijos

- El estimador within resuelve la heterogeneidad individual (α_i) eliminando toda la variación de corte transversal (between).

  ⋆ Hemos reducido el total de variación que vamos a estudiar: sólo modelamos la variación a lo largo del tiempo (al interior de las unidades, within). Los $\hat{\beta}^{FE}$ son estimados usando **únicamente las desviaciones** del valor respecto al promedio de cada unidad.

  ⋆ Purgamos la influencia de los factores constantes en el tiempo (*between-unit time-invariant factors*).

- Es posible también remover heterogeneidad temporal al incluir Efectos Fijos de Tiempo: sólo modelamos la variación *within-time*.

- Podemos purgar ambos tipos de heterogeneidad (individual y temporal) $\longrightarrow$ **Two-Way Fixed Effects** (TWFE). Este estimador:
  - Elimina los factores constantes dentro de las unidades.
  - Elimina los factores constantes a todos los individuos.

  ⋆ Sólo usa la variación que no puede explicarse ni por características fijas de las unidades ni por características fijas de los períodos de tiempo.

Resumen: Efectos Aleatorios

- El estimador de efectos aleatorios resuelve la heterogeneidad individual (α_i) haciendo supuestos parámetricos:
  - Supuesto de independencia de los regresores respecto a la heterogeneidad individual también es conocido como **exogeneidad de los covariantes**.
  - Normalidad de los residuos.
- Esto implica en la práctica:
  - $\mathbb{E}[\alpha_i | \mathbf{X}_{it}, \mathbf{Z}_i] = \mathbb{E}[\epsilon_{it} | \mathbf{X}_{it}, \mathbf{Z}_i] = 0$
  - $Cov(\mathbf{X}_{it}, \alpha_i) = 0, Cov(\mathbf{X}_{it}, \epsilon_{it}) = 0$
- Es poco probable que estos supuestos se cumplan. Pero la verdadera pregunta, es **¿cuánto sesgan los resultados?** y **¿qué estoy ganando al usar efectos aleatorios?**
- **Partial Pooling:** las unidades no son homogéneas, pero vienen de una misma distribución/población.
  - ⋆ Uso la variación between y within y puedo explorar los efectos de variables constantes en el tiempo.

Resumen de Resultados

- En base a toda la sesión, tenemos los siguientes resultados:

	Supuestos sobre α_i		
Estimador	Homogeneidad	Heterogeneidad Correlacionada	Heterogeneidad Independendiente
MCO Agrupado	Consistente	Inconsistente	Consistente
Primeras Diferencias	Consistente	Consistente	Consistente
Efectos Fijos	Consistente	Consistente	Consistente
Efectos Aleatorios	Consistente	Inconsistente	Consistente

Recomendaciones Prácticas

- La decisión más importante de un investigador aplicado es la especificación del modelo de regresión (¿cómo medir las variables? ¿qué variables incluir? ¿incluir interacciones?).
- Lo central es identificar cuál variable explicativa es más importante para mi investigación (lógica de *efectos de causas*). La decisión del estimador se ve influenciado si esta variable central es constante en el tiempo (Z) o cambiante en el tiempo (X).
- Luego de esto, la decisión del estimador se vuelve más sencilla.
  - Siempre es bueno seguir las pautas disciplinares (qué estimador es más popular en tu campo).
  - Siempre vale la pena comenzar explorando con un estimador más versátil (Efectos Fijos).
  - Sin embargo, esto no debe impedir usar otros estimadores, especialmente si son más relevantes para tus objetivos de investigación (por ejemplo, quiero estudiar el efecto total (no within) de una variable constante en el tiempo).

Recomendaciones Prácticas

- Comúnmente, los resultados de los modelos de regresión serán similares con los distintos estimadores (Efectos Fijos vs Aleatorios) ¿Qué hacer cuando difieren?
  - Explorar con mayor detalle los datos. Crear gráficos descriptivos que mapeeen diferencias promedio entre unidades (between) y diferencias a lo largo del tiempo (within).
  - Estimar un modelo between también puede resultar útil.
- El estimador de Efectos Fijos **únicamente utiliza varianza within**, por lo que es necesario entender de forma adecuada este tipo de variación. Mummolo y Peterson (2015) proveen varias recomendaciones prácticas para mejorar la interpretación de los resultados obtenidos con FE.

Recomendaciones Prácticas

- Tal como afirman Clark y Linzer (2015), lo verdaderamente importante no es si el estimador de efectos aleatorios tiene o no tiene sesgo. La verdadera pregunta es cuál es la magnitud de dicho sesgo.
- Por lo tanto, incluso si tenemos buenas razones para pensar que los regresores están correlacionados con la heterogeneidad individual, podemos usar un estimador de efectos aleatorios.
- Sin embargo, es recomendable incluir modelos between y within como pruebas de robustez para demostrar que tus resultados no se basan únicamente en un supuesto poco realista.
    - ⋆ Incluir en el Apéndice resultados comparativos para determinar si el efecto es similar con otros estimadores.
- Sin importar el estimador utilizado, siempre es necesario revisar si la correlación serial persiste. En caso que así sea, es necesario corregir los errores estándares.

Tabla de Contenidos

Literatura Sugerida

- Si desean seguir aprendiendo, recomiendo los siguientes libros para una introducción al tema:
  - Capítulo 17 *«Panel Data»* del manual Econometrics de Bruce Hansen (2022).
  - Capítulo 21 *«Linear Panel Models: Basics»* del manual Microeconometrics: Methods and Applications de Colin Cameron y Pravin Trivedi (2005).
- A nivel intermedio:
  - Applied Panel Data Analysis for Economic and Social Surveys de Hans-Jürgen Andreβ, Katrin Golsch y Alexander Schmidt (2013).
- A nivel avanzado:
  - Capítulo 10 *«Basic Linear Unobserved Effects Panel Data Models»* del manual Econometric Analysis of Cross Section and Panel Data de Jeffrey Wooldridge (2010).
  - Panel Data Econometrics with R de Yves Croissant y Giovanni Millo (2018). Este libro tiene un **tratamiento matemático avanzado pero contiene código de R** $\longrightarrow$ Altamente Recomendado.

Apéndice

A: Reshape
B: Balance
C: Dependencia
D: Estimadores
E: Matrices
F: Pooled MCO
G: Primeras Diferencias
H: Efectos Fijos
I: Efectos Aleatorios

Apéndice A: Reshaping Data

Volver

- En R, hay muchas maneras de realizar la transformación Wide/Long (y viceversa). Algunos puntos a considerar:
  1. Siempre chequea los valores de las variables de identificación (unidad y tiempo) con tablas de frecuencias.
  2. R no permite columnas con más de un type. Es necesario chequear esto de antemano. También es importante revisar los nombres de las columnas y el manejo de missing values (NA).
  3. Funciones disponibles en R:
     - `base::reshape`
     - `reshape2::dcast` (y `reshape2::melt` para wide to long).
     - `data.table::dcast` (y `data.table::melt` para wide to long)
     - `panelr::widen_panel` (y `panelr::long_panel` para wide to long)
  4. Las funciones de `tidyr` son robustas a la **ausencia de variables en ciertas mediciones en el formato wide** (completará con `NA`) y su flexibilidad para manejar los nombres de las variables.

Apéndice B: Balance

Volver

- Reporte sobre como usar panelView.
- Hay muchos paquetes útiles para el análisis de datos de panel en `R`. La mayor parte de las tareas exploratorias pueden hacerse con `base R` o con `tidyverse`. Aquí hay un buen recurso sobre gráficos descriptivos con datos panel.
- Recomiendo las funciones del paquete `plm`, `sjmisc` y `panelr` para explorar datos de panel:
  1. `plm`: `pdata.frame`, `is.pbalanced`, `is.pconsecutive`.
  2. `sjmisc`: `is_crossed` y `has_na`.
  3. `panelr`: `are_varying`.
- El objetivo debe ser identificar las causas del desbalance del panel para poder evaluar qué podemos hacer.

Apéndice B: Balance

Volver

- Hay causas del desbalance que pueden ser resueltas:
  - Muestras de refresco o diseños rotatorios (ingreso tardío de unidades).
  - Preguntas que no son incluidas en todas las mediciones.
  - Cambio en el nombre de las variables.
  - Inconsistencias en el manejo de valores perdidos para subconjuntos de variables o individuos.
- Es importante tener en cuenta que:
  - Hay varios estimadores capaces de lidiar con paneles desbalanceados.
  - Siempre es posible transformar un panel desbalanceado en un panel balanceado. Esto se puede hacer **imputando** observaciones faltantes o **descartando** observaciones que no cumplen con los estándares de balance (valores perdidos, atrición, etc.).
  - El remover observaciones o imputar filas tiene efectos sobre el tipo de inferencias que podemos extraer.

Apéndice C: Correlación Serial

Volver

Código de R

```
elsoc_wide %>%
  ### Recode missing values
  dplyr::mutate(across(where(is.numeric),   ~case_when(.x %in% c(-999, -888, -777, -666) ~ NA_real_, TRUE ~ .x))) %>%
  ### Keep only the original sample
  dplyr::filter(muestra == 1) %>%
  ### Select relevant variables
  dplyr::select(idencuesta,contains("d01_01_")) %>%
  ### Drop missing values (attrition)
  dplyr::mutate(filter = rowSums(across(d01_01_w01:d01_01_w06, is.na))) %>% dplyr::filter(filter == 0)  %>%
  ### Summarise distribution
  dplyr::summarize(across(d01_01_w01:d01_01_w06, list(n = ~ sum(!is.na(.)),
                                                    mean = ~ mean(., na.rm = TRUE),
                                                    sd = ~ sd(., na.rm = TRUE)), .names = "fn_col"),
                   cor_A_w01_w02 = cor(d01_01_w01, d01_01_w02, use = "complete.obs"),
                   cor_A_w02_w03 = cor(d01_01_w02, d01_01_w03, use = "complete.obs"),
                   cor_A_w03_w04 = cor(d01_01_w03, d01_01_w04, use = "complete.obs"),
                   cor_A_w04_w05 = cor(d01_01_w04, d01_01_w05, use = "complete.obs"),
                   cor_A_w05_w06 = cor(d01_01_w05, d01_01_w06, use = "complete.obs"),
                   cor_B_w01_w02 = cor(d01_01_w01, d01_01_w02, use = "complete.obs"),
                   cor_B_w01_w03 = cor(d01_01_w01, d01_01_w03, use = "complete.obs"),
                   cor_B_w01_w04 = cor(d01_01_w01, d01_01_w04, use = "complete.obs"),
                   cor_B_w01_w05 = cor(d01_01_w01, d01_01_w05, use = "complete.obs"),
                   cor_B_w01_w06 = cor(d01_01_w01, d01_01_w06, use = "complete.obs") ) %>%
  dplyr::mutate(cor_A_w01_w01 = NA_real_, cor_B_w01_w01 = NA_real_) %>%
  ### Reshape the output to long format (1 row per wave)
  tidyr::pivot_longer(cols = everything(), names_to = c(".value", "wave"), names_pattern = "(.*)_(.*)") %>%
  ### Recode, rename and reorder variables
  dplyr::mutate(cor_A = coalesce(cor_A_w01, cor_A_w02, cor_A_w03, cor_A_w04, cor_A_w05)) %>%
  dplyr::mutate(wave = paste0(stringr::str_to_title(wave),": ",c(2016, 2017, 2018, 2019, 2021, 2022)) ) %>%
  dplyr::rename(n = n_d01_01, mean = mean_d01_01, sd = sd_d01_01, cor_B = cor_B_w01) %>%
  dplyr::select(wave, n, mean, sd, cor_A, cor_B) %>%
  dplyr::mutate(across(where(is.numeric), ~ round(.x, 3)))
```

Apéndice D: Propiedades de Estimadores

Volver

- **Estimando** o Parámetro/Cantidad de Interés que buscamos estimar (atributo fijo -único valor- e inobservable de la población, simbolizado por θ).
- **Estimador** es una función de los datos (muestra) que usamos para aprender sobre los parámetros [es una estadística relacionada a la cantidad de interés poblacional. La representamos con $\hat{\theta}$].
  - ⋆ Es una variable aleatoria cuya variación proviene del **muestreo repetido** de la población.
  - ⋆ Con una nueva muestra, obtendré un nuevo valor de $\hat{\theta}$.
- **Estimación** es un valor particular que toma el estimador dada una muestra de datos.

Apéndice D: Propiedades de Estimadores

Volver

- Las **propiedades de un estimador** son las características de su distribución muestral.

1. **Propiedades de Muestra Finita**: aplican a cualquier tamaño muestral.
   - ***Insesgamiento:*** la distribución muestral del estimador está centrada en el valor del parámetro $\mathbb{E}[\hat{\theta}] - \theta = 0$.
   - ***Eficiencia:*** la varianza de la distribución muestral del estimador es razonablemente pequeña $\mathbb{V}[\widehat{\theta_1}] < \mathbb{V}[\widehat{\theta_2}]$
2. **Propiedades Asintóticas**: se manifiestan cuando el tamaño muestral crece.
   - ***Consistencia:*** al aumentar el N hacia el infinito, la distribución muestral del estimador converge (*en probabilidad*) al valor del parámetro $\theta \xrightarrow{p} \theta$. Otra forma de escribirlo es como $\lim_{n\to\infty} Pr(|\hat{\theta} - \theta| > \epsilon) = 0$, si para todo $\epsilon > 0$.
   - ***Normalidad Asintótica:*** al aumentar el N, la distribución muestral del estimador se aproxima a la distribución muestral $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_\theta^2)$. Un modo similar de expresarlo es como $\frac{\hat{\theta}-\theta}{se} \xrightarrow{d} \mathcal{N}(0, 1)$.

Apéndice E: Notación Matricial ▸ Volver

- La gran mayoría de los manuales sobre modelamiento con datos de panel está escrito en notación matricial. Siempre es útil invertir tiempo en revisar los conceptos básicos.
  - Ver el Capítulo 2 de las Notas de Damián Clarke.
  - Ver las Notas de Álgebra Matricial de Álvaro Parra.
- En las diapositivas centrales del Workshop utilizo una notación ligera que combina matrices con ecuaciones simples. Aquí, formalizo las ideas en notación matricial de forma más rigurosa:
  - $i \in \{1, \ldots, N\} \longrightarrow$ Individuos
  - $t \in \{1, \ldots, T_i\} \longrightarrow$ Mediciones en el tiempo.
- La notación matricial es útil para demostrar que los datos de panel son un tipo de datos **jerárquicos** o multinivel. Los datos multinivel es una forma más general de estructura dónde tengo individuos anidados en grupos (estudiantes en salones, votantes en comunas, etc.). Los estimadores panel también pueden ser usados con otros tipos de datos multinivel.

Apéndice E: Notación Matricial ▸ Volver

- Las matrices para el individuo i son:

$$\underbrace{Y_i}_{T_i \times 1} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT_i} \end{pmatrix} \text{ y } \underbrace{X_i}_{T_i \times K} = \begin{pmatrix} x_{i11} & x_{i12} & \cdots & x_{i1K} \\ x_{i21} & x_{i22} & \cdots & x_{i2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{iT_i1} & x_{iT_i2} & \cdots & x_{iT_iK} \end{pmatrix}$$

- Combinamos las matrices para todos los N individuos de forma **apilada** (*stacked*):

$$\underbrace{\mathbf{Y}}_{\Sigma_1^N T_i \times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix} \quad \text{y} \quad \underbrace{\mathbf{X}}_{\Sigma_i^N T_i \times K} = \begin{pmatrix} x_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}$$

Apéndice E: Notación Matricial

Volver

- Comenzamos con el modelo de forma **apilada**:

$$\underbrace{\mathbf{Y}}_{\sum_{i=1}^{N} T_i \times 1} = \underbrace{\mathbf{X}}_{\sum_{i=1}^{N} T_i \times K} \underbrace{\boldsymbol{\beta}}_{K \times 1} + \underbrace{\boldsymbol{u}}_{\sum_{i=1}^{N} T_i \times 1}$$

- Si reexpresamos el error compuesto con los subíndices:

$$\underbrace{u_{it}}_{\text{Error Compuesto}} = \underbrace{\alpha_i}_{\text{Heterogeneidad individual}} + \underbrace{\epsilon_{it}}_{\text{Error idiosincrático}}$$

$$\underbrace{\mathbf{u}}_{\sum_{i=1}^{N} T_i \times 1} = \underbrace{\mathbf{D}}_{\sum_{i=1}^{N} T_i \times N} \underbrace{\boldsymbol{\alpha}}_{N \times 1} + \underbrace{\boldsymbol{\epsilon}}_{\sum_{i=1}^{N} T_i \times 1}$$

Apéndice E: Notación Matricial

Volver

- Esto equivale a incorporar una **matriz de diseño** (**D**) que incluye 1s para indicar la correspondencia de los αs para cada unidad:

$$\begin{pmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1T_1} \\ u_{21} \\ \vdots \\ u_{NT_N} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_1 \\ \vdots \\ \alpha_2 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{1T_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{NT_N} \end{pmatrix}$$

- Dado lo anterior, tenemos el DGP del **One-Way Error Component Model** en notación matricial:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$$

Apéndice F: Estimador Pooled MCO

Volver

- El estimador MCO con Datos Agrupados es:

$$\widehat{\beta}^{Pooled} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y})$$

- El supuesto de **Exogeneidad Estricta** (independencia estricta en media):

$$\mathbb{E}[\epsilon_{it}|\mathbf{X}_{it}, \alpha_i] = 0$$

- Lo que implica independencia de todos los regresores X_{it} para todos los períodos de tiempo $t = 1, \ldots, T$ (pasados, presentes y futuros). Es decir, es mucho más exigente que sólo asumir $\mathbb{E}[\epsilon_{it}|X_{it}, \alpha_i] = 0$ (**Exogeneidad Contemporánea**).

Apéndice F: Supuestos Pooled MCO

Volver

1. **Linealidad:** El modelo es lineal en los parámetros $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.
2. **Exogeneidad Estricta:** Los errores no están correlacionados con los regresores en todos los períodos $\mathbb{E}[\epsilon_{it} | \mathbf{X}_{it}, \alpha_i] = 0$, para todo i y t.
   - ⋆ Con $\alpha_i \neq 0$ es necesario que $Cor(\mathbf{X}_{it}, \alpha_i) = 0$.
3. **No Multicolinealidad Perfecta:** Los regresores $\mathbf{X}$ son linealmente independientes: $\text{Rank}(\mathbf{X}) = K$.
4. **Homoscedasticidad:** La varianza de los errores es constante e independiente de i y t: $\text{Var}(\boldsymbol{\epsilon}_i | \mathbf{X}_i) = \sigma^2 \mathbf{I}_{T_i}$.
5. **No Correlación Serial:** Los errores no están correlacionados entre diferentes periodos y unidades: $\text{Cov}(\epsilon_{it}, \epsilon_{js} | \mathbf{X}) = 0 \quad \text{para} \quad (i \neq j)$ o $(t \neq s)$.
6. **Normalidad (opcional):** Los errores son normalmente distribuidos, lo que es necesario para inferencia exacta: $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$.

Apéndice F: Insesgamiento Pooled MCO

Volver

- Esta es la prueba formal del insesgamiento del estimador $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$.

$$
\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{\beta}}] &= \mathbb{E}[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}] \\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\mathbb{E}[\boldsymbol{y}] \\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\mathbb{E}[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{\epsilon}] \\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\alpha} + \underbrace{\mathbb{E}[\boldsymbol{\epsilon}]}_{\text{Exogeneidad estricta}}) \\
&= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{D}\boldsymbol{\alpha} \\
&= \boldsymbol{\beta} + (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\underbrace{\boldsymbol{X}^\top\boldsymbol{D}\boldsymbol{\alpha}}_{0} \\
&= \boldsymbol{\beta}
\end{aligned}
$$

- $\hat{\boldsymbol{\beta}}$ es un estimador insesgado de $\boldsymbol{\beta}$ si se cumple heterogeneidad estricta y $\boldsymbol{X}^\top\boldsymbol{D}\boldsymbol{\alpha} = \mathbf{0}$. Esta última condición equivale a que $\boldsymbol{\alpha} = 0$ (no hay heterogeneidad individual) o $Cor(\mathbf{X}, \boldsymbol{\alpha}) = 0$ (los regresores son independientes de la heterogeneidad individual).

Apéndice G: Estimador de Primeras Diferencias

Volver

- El estimador de Primeras diferencias comienza por transformar los datos desde una ecuación de niveles a una de cambios:

$$\Delta \boldsymbol{y} = \Delta \boldsymbol{X} \boldsymbol{\beta} + \Delta \boldsymbol{\epsilon}$$

- En la práctica, esto implica transformar los datos. El estimador de Primeras Diferencias es:

$$\hat{\boldsymbol{\beta}}^{FD} = \left(\Delta \boldsymbol{X}^{\top} \Delta \boldsymbol{X}\right)^{-1} \Delta \boldsymbol{X}^{\top} \Delta \boldsymbol{y}$$

- Esta fórmula evidencia que es simplemente el estimador MCO aplicado a datos transformados (diferenciación corresponde a restar el valor $t-1$ al valor t).
- El estimador de Primeras Diferencias también puede ser estimado con Mínimos Cuadrados Generalizados (en dicho caso, la fórmula es marginalmente distinta).

Apéndice G: Supuestos de Primeras Diferencias

▸ Volver

1. **Linealidad:** El modelo es lineal en los parámetros $\Delta\boldsymbol{y} = \Delta\boldsymbol{X}\boldsymbol{\beta} + \Delta\boldsymbol{\epsilon}$.
2. **Exogeneidad Estricta en Primeras Diferencias:** Loa cambios en los errores no están correlacionado con los regresores en todos los períodos $\mathbb{E}[\Delta\epsilon_{it} \,|\, \Delta\mathbf{X}_{it}] = 0$ para todo i y t.
   - ⋆ En Primeras Diferencias $\Delta\boldsymbol{\alpha} = 0$.
3. **No Multicolinealidad Perfecta:** $\text{Rank}(\Delta\boldsymbol{X}^\top\Delta\boldsymbol{X}) = K$.
4. **Homoscedasticidad:** La varianza de los errores en diferencias es constante: $\text{Var}(\Delta\boldsymbol{\epsilon} \,|\, \Delta\boldsymbol{X}) = \sigma^2\boldsymbol{I}$.
5. **No Correlación Serial:** No hay correlación serial en los errores diferenciales: $\text{Cov}(\Delta\epsilon_{it}, \Delta\epsilon_{is} \,|\, \Delta\boldsymbol{X}) = 0 \quad \text{para} \quad t \neq s$.
6. **Normalidad (opcional):** Los errores diferenciales siguen una distribución normal $\Delta\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

Apéndice G: Insesgamiento de Primeras Diferencias

▸ Volver

- Esta es la prueba del insesgamiento del estimador $\hat{\boldsymbol{\beta}}^{FD} = (\Delta \boldsymbol{X}^\top \Delta \boldsymbol{X})^{-1} \Delta \boldsymbol{X}^\top \Delta \boldsymbol{y}$.

$$
\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{\beta}}^{FD}] &= \mathbb{E}[(\Delta \boldsymbol{X}^\top \Delta \boldsymbol{X})^{-1} \Delta \boldsymbol{X}^\top \Delta \boldsymbol{y}] \\
&= (\Delta \boldsymbol{X}^\top \Delta \boldsymbol{X})^{-1} \Delta \boldsymbol{X}^\top \mathbb{E}[\Delta \boldsymbol{y}] \\
&= (\Delta \boldsymbol{X}^\top \Delta \boldsymbol{X})^{-1} \Delta \boldsymbol{X}^\top \mathbb{E}[\Delta(\boldsymbol{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\epsilon})] \\
&= (\Delta \boldsymbol{X}^\top \Delta \boldsymbol{X})^{-1} \Delta \boldsymbol{X}^\top \mathbb{E}[\Delta \boldsymbol{X}\boldsymbol{\beta} + \Delta \boldsymbol{\epsilon}] \\
&= (\Delta \boldsymbol{X}^\top \Delta \boldsymbol{X})^{-1} \Delta \boldsymbol{X}^\top (\Delta \boldsymbol{X}\boldsymbol{\beta} + \underbrace{\mathbb{E}[\Delta \boldsymbol{\epsilon}]}_{\text{Exogeneidad estricta}}) \\
&= (\Delta \boldsymbol{X}^\top \Delta \boldsymbol{X})^{-1} \Delta \boldsymbol{X}^\top \Delta \boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}
\end{aligned}
$$

Apéndice H: Estimador de Efectos Fijos

Volver

- Comenzamos con el *one-way error component model*: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\alpha} + \boldsymbol{\epsilon}$.

1. Promediamos el modelo para cada individuo i: $\overline{\boldsymbol{y}}_i = \overline{\boldsymbol{X}}_i\boldsymbol{\beta} + \alpha_i \mathbf{1}_T + \overline{\boldsymbol{\epsilon}}_i$
   - dónde $\overline{\boldsymbol{y}}_i = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_{it}$, $\overline{\boldsymbol{X}}_i = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{it}$, y $\overline{\boldsymbol{\epsilon}}_i = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\epsilon}_{it}$.
2. Restamos las medias individuales del modelo original:

$$\boldsymbol{y}_i - \overline{\boldsymbol{y}}_i = (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)\boldsymbol{\beta} + (\alpha_i\mathbf{1}_T - \alpha_i\mathbf{1}_T) + (\boldsymbol{\epsilon}_i - \overline{\boldsymbol{\epsilon}}_i)$$

3. Apilamos las diferencias individuales para todas las unidades:

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}$$

dónde $\tilde{\boldsymbol{y}}$, $\tilde{\boldsymbol{X}}$, y $\tilde{\boldsymbol{\epsilon}}$ son los vectores y matrices apilados de las diferencias individuales:

$$\tilde{\boldsymbol{y}} = \begin{pmatrix} \boldsymbol{y}_1 - \overline{\boldsymbol{y}}_1 \\ \boldsymbol{y}_2 - \overline{\boldsymbol{y}}_2 \\ \vdots \\ \boldsymbol{y}_N - \overline{\boldsymbol{y}}_N \end{pmatrix}, \quad \tilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_1 - \overline{\boldsymbol{X}}_1 \\ \boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2 \\ \vdots \\ \boldsymbol{X}_N - \overline{\boldsymbol{X}}_N \end{pmatrix}, \quad \tilde{\boldsymbol{\epsilon}} = \begin{pmatrix} \boldsymbol{\epsilon}_1 - \overline{\boldsymbol{\epsilon}}_1 \\ \boldsymbol{\epsilon}_2 - \overline{\boldsymbol{\epsilon}}_2 \\ \vdots \\ \boldsymbol{\epsilon}_N - \overline{\boldsymbol{\epsilon}}_N \end{pmatrix}$$

Apéndice H: Estimador de Efectos Fijos

Volver

- Lo anterior se conoce como transformación **within** y equivale a transformar los datos desde niveles a desviaciones en relación a la media individual.
- La operación de restar las medias individuales se puede representar sintéticamente por medio de una matriz de transformación: $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$, dónde
  - $\boldsymbol{I}$ es la matriz identidad de tamaño $NT \times NT$.
  - $\boldsymbol{P}$ es la matriz de proyección, que calcula las medias individuales: $\boldsymbol{P} = \boldsymbol{D}(\boldsymbol{D}^\top \boldsymbol{D})^{-1}\boldsymbol{D}^\top$
- La matriz $\boldsymbol{P}$ promedia las observaciones de $\boldsymbol{y}$ y $\boldsymbol{X}$ para cada individuo, y $\boldsymbol{M}$ se encarga de restar estos promedios de las observaciones originales.
- Por lo tanto, si aplicamos **M** al *one-way error component model*:

$$\boldsymbol{My} = \boldsymbol{MX\beta} + \underbrace{\boldsymbol{MD\alpha}}_{0} + \boldsymbol{M\epsilon}$$

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}$$

- El estimador de Efectos Fijos se obtiene aplicando MCO a los datos transformados: $\hat{\boldsymbol{\beta}}^{FE} = (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{y}}$

Apéndice H: Supuestos de Efectos Fijos

Volver

1. **Linealidad:** El modelo es lineal en los parámetros $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$.
2. **Exogeneidad Estricta:** Los errores idiosincráticos no están correlacionados con los regresores en todos los períodos $\mathbb{E}[\epsilon_{it} | \boldsymbol{X}_{it}, \alpha_i] = 0$, para todo i y t.
3. **No Multicolinealidad Perfecta:** Las columnas de $\boldsymbol{X}$ son linealmente independientes: $\text{Rank}(\boldsymbol{X}) = K$.
4. **Homoscedasticidad:** La varianza de los errores idiosincráticos es constante. $\text{Var}(\boldsymbol{\epsilon}_i | \boldsymbol{X}_i) = \sigma^2 \boldsymbol{I}_T$.
5. **No Ccorrelación Serial:** No hay correlación serial en los errores idiosincráticos. $\text{Cov}(\epsilon_{it}, \epsilon_{is} | \boldsymbol{X}_i) = 0, \quad \text{para} \quad t \neq s$
6. **Normalidad (opcional):** Los errores son normalmente distribuidos, lo que es necesario para inferencia exacta: $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$.

Apéndice H: Equivalencia FD y FE con T = 2 ▸ Volver

- Sea $\bar{\mathbf{x}}_i = \frac{1}{2}(x_{i1} + x_{i2}), \bar{y}_i = \frac{1}{2}(y_{i1} + y_{i2}), \ddot{\mathbf{x}}_{i1} = \mathbf{x}_{i1} - \bar{\mathbf{x}}_i, \ddot{\mathbf{x}}_{i2} = \mathbf{x}_{i2} - \bar{\mathbf{x}}_i,$ $\ddot{y}_{i1} = y_{i1} - \bar{y}_i$, y $\ddot{y}_{i2} = y_{i2} - \bar{y}_i$. Para $T = 2$

$$\hat{\beta}^{FE} = \left[\sum_{i=1}^{N}(\ddot{\mathbf{x}}_{i1}^T\ddot{\mathbf{x}}_{i1} + \ddot{\mathbf{x}}_{i2}^T\ddot{\mathbf{x}}_{i2})\right]^{-1}\left[\sum_{i=1}^{N}(\ddot{\mathbf{x}}_{i1}^T\ddot{y}_{i1} + \ddot{\mathbf{x}}_{i2}^T\ddot{y}_{i2})\right]$$

Algebraicamente, tenemos:

$$\ddot{\mathbf{x}}_{i1} = \mathbf{x}_{i1} - \bar{\mathbf{x}}_i = \mathbf{x}_{i1} - \frac{1}{2}(\mathbf{x}_{i1} + \mathbf{x}_{i2}) = \frac{1}{2}(\mathbf{x}_{i1} - \mathbf{x}_{i2}) = \frac{1}{2}\Delta\mathbf{x}_i$$

$$\ddot{\mathbf{x}}_{i2} = \mathbf{x}_{i2} - \bar{\mathbf{x}}_i = \mathbf{x}_{i2} - \frac{1}{2}(\mathbf{x}_{i1} + \mathbf{x}_{i2}) = \frac{1}{2}(\mathbf{x}_{i2} - \mathbf{x}_{i1}) = -\frac{1}{2}\Delta\mathbf{x}_i$$

$$\ddot{y}_{i1} = y_{i1} - \bar{y}_i = y_{i1} - \frac{1}{2}(y_{i1} + y_{i2}) = \frac{1}{2}(y_{i1} - y_{i2}) = \frac{1}{2}\Delta y_i$$

$$\ddot{y}_{i2} = y_{i2} - \bar{y}_i = y_{i2} - \frac{1}{2}(y_{i1} + y_{i2}) = \frac{1}{2}(y_{i2} - y_{i1}) = -\frac{1}{2}\Delta y_i$$

Apéndice H: Equivalencia FD y FE con T = 2

Volver

- Entonces:

$$\ddot{\mathbf{x}}_{i1}^T \ddot{\mathbf{x}}_{i1} + \ddot{\mathbf{x}}_{i2}^T \ddot{\mathbf{x}}_{i2} = \Delta \mathbf{x}_i^T \Delta \mathbf{x}_i^T / 4 + \Delta \mathbf{x}_i^T \Delta \mathbf{x}_i^T / 4 = \Delta \mathbf{x}_i^T \Delta \mathbf{x}_i^T / 2$$
$$\ddot{\mathbf{x}}_{i1}^T \ddot{y}_{i1} + \ddot{\mathbf{x}}_{i2}^T \ddot{y}_{i2} = \Delta \mathbf{x}_i^T \Delta y_i^T / 4 + \Delta \mathbf{x}_i^T \Delta y_i^T / 4 = \Delta \mathbf{x}_i^T \Delta y_i^T / 2$$

- Por lo tanto:

$$\begin{aligned}
\widehat{\beta}^{FE} &= \left[\sum_{i=1}^{N} \frac{1}{2} \Delta \mathbf{x}_i^T \Delta \mathbf{x}_i \right]^{-1} \left[\sum_{i=1}^{N} \frac{1}{2} \Delta \mathbf{x}_i^T \Delta y_i \right] \\
&= \left[\frac{1}{2} \sum_{i=1}^{N} \Delta \mathbf{x}_i^T \Delta \mathbf{x}_i \right]^{-1} \left[\frac{1}{2} \sum_{i=1}^{N} \Delta \mathbf{x}_i^T \Delta y_i \right] \\
&= \left[\sum_{i=1}^{N} \Delta \mathbf{x}_i^T \Delta \mathbf{x}_i \right]^{-1} \left[\sum_{i=1}^{N} \Delta \mathbf{x}_i^T \Delta y_i \right] \\
&= \widehat{\beta}^{FD}
\end{aligned}$$

Apéndice H: Insesgamiento de Efectos Fijos

Volver

- Esta es la prueba del insesgamiento del estimador $\hat{\boldsymbol{\beta}}^{FE} = (\tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^\top \tilde{\boldsymbol{y}}$
- Al *one-way error component model* $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\alpha} + \boldsymbol{\epsilon}$ le aplicamos la matriz de transformación $\boldsymbol{My} = \boldsymbol{MX\beta} + \boldsymbol{M\epsilon}$

$$\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{\beta}}^{FE}] &= \mathbb{E}[(\boldsymbol{MX})^\top (\boldsymbol{MX})]^{-1} (\boldsymbol{MX})^\top \boldsymbol{My}] \\
&= (\boldsymbol{MX})^\top (\boldsymbol{MX})]^{-1} (\boldsymbol{MX})^\top \mathbb{E}[\boldsymbol{My}] \\
&= (\boldsymbol{MX})^\top (\boldsymbol{MX})]^{-1} (\boldsymbol{MX})^\top \mathbb{E}[\boldsymbol{M}(\boldsymbol{X\beta} + \boldsymbol{D\alpha} + \boldsymbol{\epsilon})] \\
&= (\boldsymbol{MX})^\top (\boldsymbol{MX})]^{-1} (\boldsymbol{MX})^\top \mathbb{E}[\boldsymbol{MX\beta} + \boldsymbol{M\epsilon}] \\
&= (\boldsymbol{MX})^\top (\boldsymbol{MX})]^{-1} (\boldsymbol{MX})^\top (\boldsymbol{MX\beta} + \underbrace{\mathbb{E}[\boldsymbol{M\epsilon}]}_{\text{Exogeneidad estricta}}) \\
&= (\boldsymbol{MX})^\top (\boldsymbol{MX})]^{-1} (\boldsymbol{MX})^\top \boldsymbol{MX\beta} \\
&= \boldsymbol{\beta}
\end{aligned}$$

Apéndice I: Estimador de Efectos Aleatorios

Volver

- Comenzamos con el *one-way error component model*: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\alpha} + \boldsymbol{\epsilon}$.
- En este caso, no transformaremos los datos. En cambio, realizaremos supuestos paramétricos sobre la heterogeneidad individual $\boldsymbol{\alpha}$ y $\boldsymbol{\epsilon}$:

1. Distribución Normal de la heterogeneidad individual (con media α y varianza σ_α^2): $\boldsymbol{\alpha} \sim \mathcal{N}(\alpha, \sigma_\alpha^2 \boldsymbol{I}_N)$.
2. Exogeneidad de heterogeneidad individual: $\mathbb{E}[\boldsymbol{\alpha} \,|\, \mathbf{X}] = 0$. Es decir, la heterogeneidad individual no está correlacionada con los regresores.
3. Distribución Normal de los errores idiosincráticos (con media cero y varianza σ_ϵ^2): $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I}_{NT})$

Apéndice I: Estimador de Efectos Aleatorios

▸ Volver

- Al asumir que los efectos individuales ($\boldsymbol{\alpha}$) están distribuidos normalmente, y son independientes de los regresores (y también de los errores idiosincráticos, al igual que en los casos anteriores, asumiremos exogeneidad estricta), podemos construir una matriz de varianzas y covarianzas $\boldsymbol{\Omega}$ que refleja la correlación dentro de las unidades debido a la presencia de heterogeneidad individual.

- Estos supuestos nos permiten aprovechar otro estimador llamado **Mínimos Cuadrados Generalizados**.

- Por lo tanto, el **estimador de Efectos Aleatorios** es:

$$\hat{\boldsymbol{\beta}}_{RE} = \left(\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y}$$

donde:

- $\boldsymbol{\Omega} = \sigma_\alpha^2 \boldsymbol{D}\boldsymbol{D}^\top + \sigma_\epsilon^2 \boldsymbol{I}$ es la matriz de varianzas y covarianzas de los errores.

Apéndice I: Supuestos de Efectos Aleatorios

Volver

1. **Linealidad:** El modelo es lineal en los parámetros: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\alpha} + \boldsymbol{\epsilon}$.
2. **Exogeneidad Estricta:** Los errores idiosincráticos no están correlacionados con los regresores en todos los períodos $\mathbb{E}[\boldsymbol{\epsilon_{it}} | \mathbf{X}_{it}] = 0$ para todo i y t.
   - ⋆ También lo asumimos para la heterogeneidad individual: $\mathbb{E}[\boldsymbol{\alpha i} | \mathbf{X}_{it}] = 0$ para todo i y t.
3. **No Multicolinealidad Perfecta:** Las columnas de $\boldsymbol{X}$ son linealmente independientes. Rank$(\boldsymbol{X}) = K$
4. **Homoscedasticidad:** La varianza de los errores idiosincráticos es constante: Var$(\boldsymbol{\epsilon} | \boldsymbol{X}) = \sigma_\epsilon^2 \boldsymbol{I}_{NT}$
   - ⋆ También lo asumimos para la heterogeneidad individual: Var$(\boldsymbol{\alpha} | \mathbf{X}) = \sigma_\alpha^2 \boldsymbol{I}_N$
5. **No Correlación Serial:** No hay correlación serial en los errores idiosincráticos: Cov$(\epsilon_{it}, \epsilon_{is} | \boldsymbol{X}) = 0, \quad$ para $\quad t \neq s$
6. **Normalidad:** Los efectos individuales y los errores idiosincráticos son normalmente distribuidos: $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \sigma_\alpha^2 \boldsymbol{I}_N), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I}_{NT})$

Apéndice I: Insesgamiento de Efectos Aleatorios

▸ Volver

- Esta es la prueba del insesgamiento del estimador $\left(X^{\top}\Omega^{-1}X\right)^{-1}X^{\top}\Omega^{-1}y$

$$
\begin{aligned}
\mathbb{E}[\hat{\beta}_{RE}] &= \mathbb{E}\left[\left(X^{\top}\Omega^{-1}X\right)^{-1}X^{\top}\Omega^{-1}y\right] \\
&= \left(X^{\top}\Omega^{-1}X\right)^{-1}X^{\top}\Omega^{-1}\mathbb{E}[y] \\
&= \left(X^{\top}\Omega^{-1}X\right)^{-1}X^{\top}\Omega^{-1}\mathbb{E}[X\beta + D\alpha + \epsilon] \\
&= \left(X^{\top}\Omega^{-1}X\right)^{-1}X^{\top}\Omega^{-1}\left(X\beta + \underbrace{\mathbb{E}[D\alpha]}_{\mathbb{E}[\alpha|X]=0} + \underbrace{\mathbb{E}[\epsilon]}_{\mathbb{E}[\epsilon|X]=0}\right) \\
&= \left(X^{\top}\Omega^{-1}X\right)^{-1}X^{\top}\Omega^{-1}X\beta \\
&= \beta
\end{aligned}
$$

Apéndice I: Comparación de Modelos

▸ Volver

- El estimador de efectos aleatorios ($\hat{\boldsymbol{\beta}}^{RE}$) se puede expresar como un compromiso entre el estimador Within ($\hat{\boldsymbol{\beta}}^{FE}$) y el estimador Between ($\hat{\boldsymbol{\beta}}^{Between}$):

$$\hat{\boldsymbol{\beta}}^{RE} = \lambda\hat{\boldsymbol{\beta}}^{FE} + (1-\lambda)\hat{\boldsymbol{\beta}}^{Between}$$

  donde λ es un factor de ponderación que depende de las varianzas σ_α^2 y σ_ϵ^2. El estimador de efectos aleatorios usa ambas fuentes de variación y pondera las estimaciones para capturar tanto la variabilidad entre unidades como la variabilidad dentro de las unidades.

  - Si la varianza de los efectos individuales (σ_α^2) es grande, λ se aproxima a 1, y $\hat{\boldsymbol{\beta}}_{RE}$ se aproxima a $\hat{\boldsymbol{\beta}}_{Within}$.
  - Si la varianza de los efectos individuales (σ_α^2) es pequeña, λ se aproxima a 0, y $\hat{\boldsymbol{\beta}}_{RE}$ se aproxima a $\hat{\boldsymbol{\beta}}_{Pooled}$.

  ⋆ Ver páginas 33-47 de (Croissant y Millo, 2018).