Economía Aplicada

Datos de panel

Universidad Carlos III de Madrid

1. Introducción
2. El modelo lineal básico con datos de panel
   - Exogeneidad en el contexto de datos de panel
   - Heterogeneidad inobservable
   - La falsa dicotomía entre efectos fijos y aleatorios
3. Datos de panel con dos períodos de tiempo
   - El estimador MCO con datos fusionados
   - El estimador de Primeras Diferencias (PD)
   - Ejemplo
   - Limitaciones del estimador de PD
4. Datos de panel con más períodos temporales
   - El estimador de PD con más períodos temporales
5. El estimador IG (Intragrupos)
6. El estimador de "Efectos Aleatorios"
7. Conclusiones finales
   - Ventajas de datos de panel
   - Aspectos prácticos con datos de panel
   - Limitaciones de los datos de panel

Introducción: Datos de panel

- Ver también Wooldridge (capítulo 13) y Stock & Watson (capítulo 10).
- Los datos longitudinales o de panel consisten en una muestra de varias observaciones en períodos de tiempo sucesivos para las mismas unidades individuales.
    - Cada unidad individual se refiere a un individuo, o un hogar, una empresa, una provincia, un país, etc.
- A diferencia de las secciones cruzadas repetidas, donde tenemos varios períodos de tiempo pero las observaciones en cada período corresponden a distintos individuos, con los datos de panel seguimos a los mismos individuos durante varios períodos.
- En general, consideraremos aplicaciones de datos de panel en que el número de individuos (n) es muy grande mientras que el número de períodos de tiempo en que los observamos (T) es mucho más pequeño.

Introducción: Datos de panel (cont.)

- Para caracterizar una población de interés con datos de panel, partiremos de una encuesta a una muestra aleatoria de individuos en un momento del tiempo y volveremos a recoger la misma información para esos mismos individuos en períodos posteriores (mes, trimestre, año...).
- Un ejemplo para España es la "Encuesta de Condiciones de Vida" del INE. La primera encuesta se realizó en 2004 con una muestra de cerca de 15000 hogares, y cada hogar es entrevistado cada año durante cuatro años consecutivos.
- Con datos de panel es posible seguir la evolución temporal de ciertas variables para las mismas unidades individuales.
- Al disponer de observaciones para los mismas unidades individuales durante varios períodos de tiempo, éstas pueden tener dependencia temporal.
  - Por ejemplo, si ciertos factores individuales inobservables afectaron al salario de un individuo en 2014, también pueden afectar a su salario en 2015.

Ventajas de los datos de panel

- Los datos de panel tienen dos ventajas principales con respecto a las secciones cruzadas.
- Estas ventajas surgen porque con datos de panel se dispone de varias observaciones de los mismos individuos a lo largo del tiempo.
- La primera ventaja es la posibilidad de controlar por el sesgo potencial debido a la **heterogeneidad (individual) inobservable**.
- La segunda ventaja es la posibilidad de estimar **modelos dinámicos**.
  - Sin embargo, en este curso nos centraremos en modelos **estáticos** y no discutiremos la estimación de modelos dinámicos.

Ventajas de los datos de panel (cont.)

- La **heterogeneidad (individual) inobservable** consiste en características individuales omitidas que difieren entre individuos pero son constantes en el tiempo.
  - Con varias observaciones de las mismas unidades individuales a lo largo del tiempo, podemos evitar el **sesgo de heterogeneidad inobservable** aplicando métodos y transformaciones específicos.
  - Este problema no se puede resolver con secciones cruzadas y secciones cruzadas repetidas, al observar a cada individuo un solo período.
- Los **modelos dinámicos** (bien conocidos en econometría de series temporales) incluyen la **variable dependiente desfasada como variable explicativa**.
  - Con datos de panel disponemos de una serie temporal (generalmente corta) de cada individuo.
  - Usando i para la unidad individual y t para el período temporal, dada una muestra para dos variables $\{X_{it}, Y_{it}\}$, $(i = 1, \ldots, n;\ t = 1, \ldots, T)$, podemos considerar especificaciones como

$$Y_{it} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 X_{it} + v_{it}.$$

El modelo lineal con datos de panel

- Consideremos una muestra para dos variables X e Y, $\{X_{it}, Y_{it}\}$, $i = 1, \ldots, n$; $t = 1, \ldots, T$.
  - El subíndice i se refiere a la unidad individual (individuo, hogar, empresa, etc.)
  - El subíndice t se refiere al período temporal.
- Sea, para $i = 1, \ldots, n$; $t = 1, \ldots, T$, el modelo lineal

$$Y_{it} = \beta_0 + \sum_{s=2}^{T} \delta_s ds_t + \beta_1 X_{it} + \underbrace{a_i + u_{it}}_{v_{it}}$$

donde

  - ds_t es una variable binaria que vale 1 en el período $t = s$ y 0 en $t \neq s$ (común a todas las unidades, por lo que no tiene subíndice i).
    Captura shocks agregados comunes a todos los individuos.
  - $v_{it} = a_i + u_{it}$ es el error aleatorio inobservable (compuesto por dos términos).

El modelo lineal con datos de panel (cont.)

- El **error aleatorio compuesto** $v_{it} = a_i + u_{it}$ se descompone en:
  - a_i: variable aleatoria inobservable que recoge todos los factores individuales no observados que afectan a Y_{it} y no cambian a lo largo del tiempo (por lo que no tiene subíndice t).
  - u_{it}: error idiosincrático usual que incluye cualquier shock no anticipado factor o factor inobservable que afecta a Y_{it} y difiere tanto a nivel individual como temporal.
- a_i recoge diferencias entre individuos inobservables e invariantes en el tiempo (heterogeneidad inobservable, heterogeneidad individual o efecto fijo individual).
- Las propiedades del modelo dependerán de la relación de la variable explicativa X_{it} con el término de error inobservable v_{it}.
  - Tendremos que analizar dicha relación para cada componente del error, a_i y u_{it}.

Exogeneidad en el contexto de datos de panel

- Discutiremos la relación de X_{it} con el shock idiosincrático convencional u_{it}.
- Tenemos **exogeneidad contemporánea** cuando $C\left(X_{it}, u_{it}\right) = 0$.
  - Con secciones cruzadas (repetidas o no), basta con la **exogeneidad contemporánea** para obtener estimaciones consistentes de los parámetros de interés.
- Tenemos **exogeneidad estricta** cuando $C\left(X_{it}, u_{is}\right) = 0$, $\forall t, s = 1, \ldots, T$.

Exogeneidad en el contexto de datos de panel (cont.)

- La exogeneidad estricta es mucho más exigente que la exogeneidad contemporánea, porque requiere:
  - (i) X_{it} no correlacionado con shocks futuros: $C\left(X_{it}, u_{is}\right) = 0$, $s > t$.
  - (ii) X_{it} no correlacionado con shocks presentes: $C\left(X_{it}, u_{it}\right) = 0$ (exogeneidad contemporánea).
  - (iii) X_{it} no correlacionado con shocks pasados: $C\left(X_{it}, u_{is}\right) = 0$, $s < t$.
- Es de esperar que siempre se cumpla la condición (i) de no correlación de la(s) variable(s) explicativa(s) con shocks futuros, porque no cabe esperar que shocks no anticipados tengan efecto alguno sobre realizaciones de X anteriores a dichos shocks.
- Pero la condición (iii) es muy exigente en cualquier situación, porque exige que X no se vea afectada por shocks ocurridos en el pasado.

Heterogeneidad inobservable

- ¿Bajo qué condiciones el estimador de MCO con datos fusionados del parámetro de interés β_1 es consistente?
- Comencemos suponiendo exogeneidad contemporánea de X: $C(X_{it}, u_{it}) = 0$.
  - Pero esta condición no es suficiente para que el estimador MCO con datos fusionados sea consistente.
- Si hay exogeneidad contemporánea, la cuestión relevante será si a_i está o no correlacionada con X_{it}.
  - Si además de $C(X_{it}, u_{it}) = 0$ tenemos $C(X_{it}, a_i) = 0$, entonces

$$C(X_{it}, v_{it}) = C(X_{it}, a_i + u_{it}) = \underbrace{C(X_{it}, a)}_{0} + \underbrace{C(X_{it}, u_{it})}_{0} = 0$$

    y en consecuencia el estimador MCO con datos fusionados será consistente.
  - Si por el contrario $C(a_i, X_{it}) \neq 0 \Rightarrow C(X_{it}, v_{it}) \neq 0$ y MCO con datos fusionados será inconsistente.
- El sesgo por omitir a_i cuando $C(a_i, X_{it}) \neq 0$ se denomina **sesgo de heterogeneidad inobservable**.

Heterogeneidad inobservable: Ejemplos

- Función de producción de un campo de cultivo
  - $Y_{it} = \log\left(\text{Producción}/\text{m}^2\right)$, $X_{it} = \log\left(\text{Fertilizante}/\text{m}^2\right)$; a_i = Calidad de la tierra (invariante en el tiempo, conocida por el agricultor pero desconocida por el analista); u_{it} = shocks que afectan a la cosecha producida (pluviometría, sequía, desplazamientos de tierra, etc.)
  - Probablemente, la cantidad de fertilizante por m^2 estará correlacionada con la calidad de la tierra.
- Efecto de la formación en las ganancias salariales de un trabajador
  - $Y_{it} = \log(\text{Ganancias anuales})$, $X_{it} = \log(\text{Horas de formación})$; a_i = Habilidad individual; u_{it} = shocks que afectan a las ganancias (rentabilidad de la empresa, conflictos laborales, etc.)
  - Si cada trabajador decide las horas que dedica a su formación, probablemente la cantidad de dichas horas estará correlacionada con la habilidad individual.

La falsa dicotomía entre efectos fijos y aleatorios

- El enfoque tradicional, ya desfasado, considera dos puntos de vista mutuamente excluyentes sobre los efectos individuales a_i:
  - "**Efectos aleatorios"**: los a_i's $(i = 1, \ldots, n)$ son variables aleatorias, que se suponen no correlacionadas con la(s) variable(s) explicativa(s) (de modo que $C(a_i, X_{it}) = 0$).
  - "**Efectos fijos"**: los a_i's $(i = 1, \ldots, n)$ son n parámetros (fijos) desconocidos asociados a un término constante distinto para cada individuo. Los estimadores que ignoran estos parámetros serán inconsistentes.
- Sin embargo, este enfoque plantea una falsa dicotomía: los efectos individuales a_i's son **variables aleatorias**.
- La **cuestión clave** es si los efectos individuales están o no correlacionados con la(s) variable(s) explicativa(s).
  - Si $C(a_i, X_{it}) = 0$, el estimador MCO con datos fusionados (que ignora a_i) será consistente.
  - Si $C(a_i, X_{it}) \neq 0$, el estimador MCO con datos fusionados será inconsistente. Necesitaremos una transformación del modelo original que elimine los efectos individuales a_i para obtener estimaciones consistentes.

Datos de panel con dos períodos de tiempo

- Consideremos el panel más simple con $T = 2$,

$$\begin{aligned} Y_{it} &= \beta_0 + \delta_2 d2_t + \beta_1 X_{it} + v_{it} \\ v_{it} &= a_i + u_{it} \end{aligned}$$

- Para que el estimador MCO con datos fusionados de β_1 sea consistente, necesitamos $C(X_{it}, v_{it}) = 0$, para lo que la exogeneidad contemporánea de X, $C(X_{it}, u_{it}) = 0$ es condición necesaria pero no suficiente.
  - Aun en ese caso, si $C(X_{it}, a_i) \neq 0$, MCO será inconsistente.
- No controlar por la omisión de a_i hace que el estimador MCO con datos fusionados sea inconsistente a causa del **sesgo de heterogeneidad**.

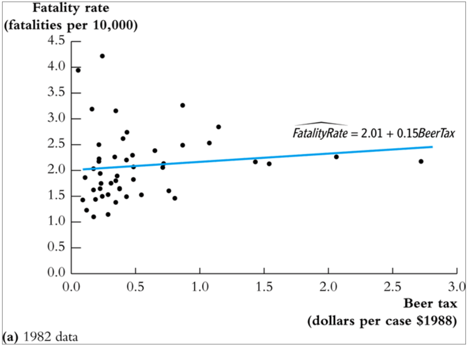
Ejemplo: Impuestos al alcohol y mortalidad en carretera

- Objetivo: Analizar políticas sobre el consumo de alcohol para reducir la mortalidad en carretera.
- Datos `fatality.gdt` en `gretl`: 48 estados de EEUU, 7 años (1982 a 1988).
- Nos concentramos en 1982 y 1988 solamente.
- Variables:
  - Mortalidad en carretera: tasa de muertes en carretera por cada 10,000 residentes (variable TM).
  - Impuesto sobre un paquete de cervezas en dólares de 1988 (variable *beertax*).

Ejemplo: Impuestos al alcohol y mortalidad en carretera

Usando una sección cruzada

- Modelo simple para el año 1982: $TM_i = \beta_0 + \beta_1 beertax_i + u_i$



- La estimación MCO para 1982 del coeficiente es $\widehat{\beta}_1 = 0{,}15$, que es positiva (pero estadísticamente no significativa).

- Usando las secciones cruzadas fusionadas para 1982 y 1988, la estimación MCO del coeficiente es positiva y significativa: ¡¡Comprobar con `gretl`!!
  - Entonces, ¿gravar la cerveza incrementa la mortalidad en carretera?
- Hay muchos factores (omitidos) específicos de cada estado detrás de la siniestralidad vial, como la edad del parque automovilístico, la calidad de las carreteras, la cultura de conducir y beber, la densidad de tráfico...
- Si alguno de estos factores está correlacionado con los impuestos a la cerveza, los estimadores MCO con datos fusionados serán inconsistentes.
- Aunque podemos recoger datos sobre algunos de estos factores, otros (como la cultura de beber y conducir) son difíciles de medir.
- Estos factores omitidos específicos de estado varían poco (o nada en absoluto) en un intervalo temporal de 10 años.
- Si dichos factores permanecen constantes en el período muestral, con datos de panel podremos obtener estimaciones consistentes.

La transformación de PD

- Lo más plausible es que los efectos individuales inobservables a_i estén correlacionados con la(s) variable(s) explicativa(s).
- Podemos explotar el hecho de que Y_{it}, X_{it} varían a lo largo del tiempo, pero a_i no.
- Cuando $T = 2$, podemos escribir para cada período

$$\begin{aligned} Y_{i2} &= \beta_0 + (\delta_2 \times 1) + \beta_1 X_{i2} + a_i + u_{i2} \\ Y_{i1} &= \beta_0 + (\delta_2 \times 0) + \beta_1 X_{i1} + a_i + u_{i1} \end{aligned}$$

  Nótese que $d2_t$ vale 1 cuando $t = 2$ y 0 cuando $t = 1$.
- Si restamos a la ecuación del segundo período la del primer período, obtenemos una ecuación transformada en primeras diferencias:

$$(Y_{i2} - Y_{i1}) = \delta_2 + \beta_1 (X_{i2} - X_{i1}) + (u_{i2} - u_{i1}).$$

El estimador de Primeras Diferencias (PD)

- De forma equivalente, podemos usar el operador Δ para denotar el cambio temporal $\Delta Z_{i2} = Z_{i2} - Z_{i1}$.
- Cuando $T = 2$, la ecuación en primeras diferencias (PD) es una úncia sección cruzada, en la que cada variable original se ha transformado en PD.

$$\Delta Y_{i2} = \delta_2 + \beta_1 \Delta X_{i2} + \Delta u_{i2}$$

- Nótese que la transformación de PD permite eliminar a_i porque $\Delta a_i = (a_i - a_i) = 0$.
- **Condición clave**: $C(\Delta X_{i2}, \Delta u_{i2}) = 0$.
  - Si Δu_{i2} no está correlacionado con ΔX_{i2}, aplicando MCO a la ecuación en PD se obtiene una estimación consistente de β_1.
  - **Pregunta:** Para que se cumpla la condición clave, ¿es suficiente que $C(X_{it}, u_{it}) = 0$?
- El estimador MCO de la ecuación en PD se denomina estimador de Primeras Diferencias (PD).

Sobre el estimador de PD

- Si se cumple la **condición clave**, el estimador de PD es consistente bajo cualquier forma de correlación entre a_i y la(s) variable(s) explicativa(s)..
- Bajo condiciones generales de regularidad, $p \lim_{N \to \infty} \left(\widehat{\beta}_1^{PD} \right) = \beta_1$ para T fijo.
- No es necesario estimar los efectos individuales inobservables a_i.
  - Estos se eliminan al transformar la ecuación original en PD.
- El estimador de PD requiere variabilidad de ΔX_{i2} entre las unidades individuales.
  - Si ΔX_{i2} tiene poca variación, el estimador de PD será muy impreciso (porque la varianza de $\widehat{\beta}_1^{PD}$ será grande).
- A veces, puede tener sentido tomar diferencias entre períodos más separados en el tiempo para asegurar suficiente variabilidad temporal en X_{it}.

- Consideremos los años 1982 y 1988:

$$TM_{it} = \beta_0 + \delta_2 d2_t + \beta_1 beertax_{it} + a_i + u_{it}, \quad t = 1982, 1988.$$

- a_i representa variables inobservables específicas de estado e invariantes en el tiempo que afectan a la siniestralidad en el estado i.
  - Puede incluir actitudes locales sobre conducción y consumo de alcohol.
  - También puede recoger otras características inobservables específicas de estado e invariantes en el tiempo que afectan a la siniestralidad. **¿Alguna sugerencia?**
  - Pueden considerarse constantes entre 1982 y 1988 si cambian muy despacio.
- Es muy probable que estados con actitudes sociales poco permisivas sobre conducción y consumo de alcohol tengan en promedio menor tasa de siniestralidad... y que probablemente tengan mayores impuestos al alcohol.
- Se trata de una situación clara de sesgo de variable omitida: *beertax* captura en parte el efecto indirecto de la actitud local dominante respecto a conducir habiendo consumido alcohol.

Ejemplo: Impuestos al alcohol y mortalidad en carretera

Primeras diferencias

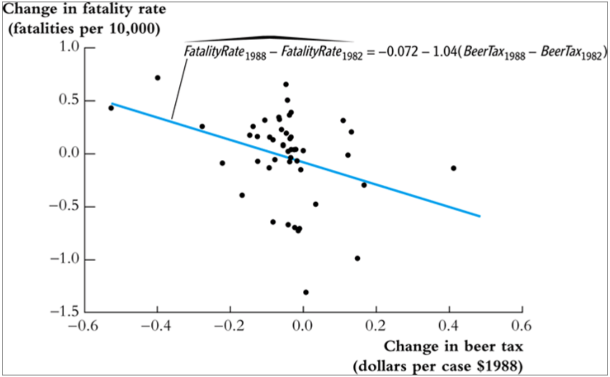
- Los efectos individuales inobservables e invariantes en el tiempo a_i se eliminan tomando PD

$$\Delta TM_{it} = \delta_2 + \beta_1 \Delta beertax_{it} + \Delta u_{it}, \quad t = 1988.$$

  Nótese que la transformación de PD elimina también cualquier término invariante en el tiempo, como el término constante β_0 y el efecto individual a_i.
  Además, $\Delta \delta_2 d2_t = \delta_2 \times (1 - 0) = \delta_2$.

- En resumen:
  - Las actitudes sobre conducir y beber (recogidas en a_i) afectan a la proporción de conductores que beben y por tanto a la tasa de siniestralidad del estado.
  - Si estas actitudes apenas cambian en un intervalo de tiempo relativamente largo (por ej., entre 1982 y 1988), **no deberían afectar al cambio** en la siniestralidad.
  - En consecuencia, el cambio temporal en TM_{it} (entre 1982 y 1988), ΔTM_{it}, se debará a otros factores, como cambios en el impuesto a la cerveza $beertax_{it}$ o en los shocks idiosincráticos u_{it}.

- Modelo para el año 1988 en PD: $\Delta TM_{i2} = \delta_2 + \beta_1 \Delta beertax_{i2} + \Delta u_{i2}$.



- Al contrario que la estimación MCO con datos fusionados, la estimación de PD muestra que aumentar el impuesto a la cerveza reduce la siniestralidad.

- El término constante mide el efecto en la tasa de mortalidad en carretera de efectos agregados comunes a todos los estados.
- Un incremento en el impuesto al paquete de cervezas de 1$ reduce la mortalidad media en carretera en 1,04 muertes por cada 10000 residentes.
- El efecto estimado del impuesto a la cerveza es significativo y relevante:
  - La tasa de mortalidad media es de 2 muertes por cada 10000 residentes.
  - Por tanto, la reducción estimada de 1,04 muertes sugiere que incrementar el impuesto en 1$ puede reducir la mortalidad en carretera a la mitad (50 %).

Extensiones y limitaciones del estimador de PD

- La extensión del estimador de PD a un modelo con varias variables explicativas es inmediata.
- Al igual que con datos de sección cruzada, es muy probable que exista **heterocedasticidad** (varianza del error idiosincrático no constante). Por ello, deben calcularse **errores estándar robustos a heterocedasticidad**.
- El estimador de PD elimina cualquier factor invariante en el tiempo, se observe o no.
  - Por tanto, el estimador de PD no permite estimar el efecto de ninguna variable explicativa que sea invariante en el tiempo.
- El estimador de PD necesita suficiente variación temporal en las variables explicativas.
  - En el ejemplo de mortalidad en carretera, necesitamos que *beertax* cambie a lo largo del tiempo en varios estados para estimar su efecto con suficiente precisión.

Extensiones y limitaciones del estimador de PD (cont.)

- Para que el estimador de PD sea consistente, la **condición clave** es $C\left(\Delta X_{it}, \Delta u_{it}\right) = 0$.
  - Nótese que no basta con que haya exogeneidad contemporánea.
- Podrían existir otros factores individuales con variación temporal que afecten a Y_{it} y que estén correlacionados con la(s) variable(s) explicativa(s)
  Si se omiten dichos factores, entonces $C\left(X_{it}, u_{it}\right) \neq 0$, y por tanto $C\left(\Delta X_{it}, \Delta u_{it}\right) \neq 0$.
  - En ese caso, el estimador de PD sufrirá de **sesgo de endogeneidad** y será inconsistente.
  - Para resolver este problema sería necesario aplicar VI/MC2E, que no discutimos en este curso.

PD con más períodos temporales

$$Y_{it} = \beta_0 + \sum_{s=2}^{T} \delta_s ds_t + \beta_1 X_{it} + a_i + u_{it}, \quad i = 1, \ldots, n; \;\; t = 1, \ldots, T.$$

- La estrategia de PD discutida anteriormente se puede aplicar igualmente cuando $T > 2$, tomando diferencias entre períodos consecutivos (entre t y $t-1$).
- Al igual que con $T = 2$, la **condición clave** es $C\left(\Delta X_{it}, \Delta u_{it}\right) = 0 \; \forall t \geq 2$.
  - Nótese que $C\left(\Delta X_{it}, \Delta u_{it}\right) = C\left(X_{it} - X_{i,t-1}, u_{it} - u_{i,t-1}\right) = C\left(X_{it}, u_{it}\right) - C\left(X_{it}, u_{i,t-1}\right) - C\left(X_{i,t-1}, u_{it}\right) + C\left(X_{i,t-1}, u_{i,t-1}\right)$.
  - Entonces, la **condición clave** requiere $C\left(X_{it}, u_{it}\right) = 0$, $C\left(X_{it}, u_{i,t-1}\right) = 0$, $C\left(X_{it}, u_{i,t+1}\right) = 0$,
    o de forma equivalente $C\left(X_{it}, u_{i,t+s}\right) = 0, \; \forall t, s = -1, 0, 1$.

PD con más períodos temporales (cont.)

- Por simplicidad, consideremos $T = 3$,

$$Y_{it} = \beta_0 + \delta_2 d2_t + \delta_3 d3_t + \beta_1 X_{it} + a_i + u_{it}, \quad t = 1, 2, 3.$$

- Nuestro parámetro de interés es β_1.
- Si $C(X_{it}, a_i) \neq 0$, MCO con datos fusionados será inconsistente, incluso aunque $C(X_{it}, u_{it}) = 0$.
- **Condición clave** para la consistencia del estimador de PD: $C(\Delta X_{it}, \Delta u_{it}) = 0 \Leftrightarrow C(X_{it}, u_{i,t+s}) = 0$, $\forall t, s = -1, 0, 1$.

PD con más períodos temporales (cont.)

- Transformando en PD para eliminar a_i,

$$\Delta Y_{it} = \delta_2 \Delta d2_t + \delta_3 \Delta d3_t + \beta_1 \Delta X_{it} + \Delta u_{it}, \quad t = 2, 3.$$

lo que equivale a incluir un término constante y una dummy temporal:

$$\Delta Y_{it} = \alpha_0 + \alpha_3 d3_t + \beta_1 \Delta X_{it} + \Delta u_{it}, \quad t = 2, 3.$$

- En general, cuando $T > 3$ tendremos $T - 1$ observaciones por unidad individual, e incluiremos $T - 2$ dummys temporales:

$$\Delta Y_{it} = \alpha_0 + \sum_{s=3}^{T} \alpha_s ds_t + \beta_1 \Delta X_{it} + \Delta u_{it}, \quad t = 2, \ldots, T.$$

Otra transformación para eliminar los efectos individuales

- Existen transformaciones alternativas para eliminar el término de heterogeneidad individual inobservable.
- Consideremos el modelo lineal simple, haciendo abstracción de efectos temporales agregados:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + a_i + u_{it}, \quad i = 1, \ldots, n; \ t = 1, \ldots, T.$$

- Consideremos, para cada individuo, la media temporal de la expresión anterior:

$$\overline{Y}_i = \beta_0 + \beta_1 \overline{X}_i + a_i + \overline{u}_i, \quad i = 1, \ldots, n,$$

donde $\overline{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, $\overline{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, $\overline{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$.

- Si restamos la ecuación de la media a la ecuación original en cada período t:

$$\left(Y_{it} - \overline{Y}_i\right) = \beta_1 \left(X_{it} - \overline{X}_i\right) + \left(u_{it} - \overline{u}_i\right), \quad i = 1, \ldots, n; \ t = 1, \ldots, T.$$

El estimador IG

- Las variables originales para cada unidad individual i quedan transformadas en desviaciones con respecto a sus medias individuales $\widetilde{Y}_{it} = Y_{it} - \overline{Y}_i$, $\widetilde{X}_{it} = X_{it} - \overline{X}_i$, etc.

$$\widetilde{Y}_{it} = \beta_1 \widetilde{X}_{it} + \widetilde{u}_{it}, \quad i = 1, \ldots, n; \;\; t = 1, \ldots, T.$$

- Esta transformación se denomina **intragrupos** (IG).
- El efecto individual a_i se elimina, porque su transformación IG es igual a 0.
- El estimador MCO con datos fusionados aplicado a la transformación IG del modelo se conoce como estimador IG.
- La **condición clave** para la consistencia del estimador IG es $C\left(\widetilde{X}_{it}, \widetilde{u}_{it}\right) = 0$, ¡¡que requiere **exogeneidad estricta**!!

Estimación IG o de Efectos Fijos (EF)

- El estimador IG se conoce también como estimador de efectos fijos (EF).
- La razón es que estimar por MCO la transformación IG del modelo original es equivalente a estimar por MCO el modelo original ampliado con un conjunto de variables binarias adicionales para cada una de las unidades individuales i:

$$Y_{it} = \sum_{i=1}^{n} \gamma_i D_i + \beta_1 X_{it} + u_{it}, \quad i = 1, \ldots, n; \;\; t = 1, \ldots, T,$$

  donde D_i es una variable binaria de *efecto fijo*, que vale 1 si la observación corresponde al individuo i y 0 en caso contrario.
- De hecho, aunque cada efecto fijo $\gamma_i D_i$ no es en realidad un parámetro fijo, permanece constante en el período muestral.
- El término de efecto fijo se corresponde con un enfoque desfasado que considera los efectos individuales inobservables a_i como parámetros "molestos" desconocidos, en vez de como variables aleatorias.
- Los efectos individuales inobservables deben verse como variables aleatorias espcíficas de individuo que apenas varían a lo largo del período muestral.

Propiedades del estimador IG

- El estimador IG o de EF es consistente si $C(X_{it}, u_{is}) = 0 \; \forall t, s$ (**exogeneidad estricta**).
  - Esta condición clave es muy restrictiva, pudiendo incumplirse en muchas situaciones.
- Tanto el estimador PD como el IG permiten cualquier tipo de correlación entre los efectos individuales a_i y las variables explicativas.
  - Además, las transformaciones de PD e IG eliminan cualquier variable invariante en el tiempo que esté incluida en el modelo. En consecuencia, los estimadores de PD e IG no permiten identificar el coeficiente de ninguna variable observada que no tenga variación temporal.
- En el caso más habitual de n grande y T pequeño, la elección entre el estimador de PD o el IG depende de la eficiencia relativa de cada uno (puesto que ambos serán consistentes bajo exogeneidad estricta).
  - Cuando $T > 2$ el estimador IG es en general más eficiente que el de PD.
  - Cuando $T = 2$, ambos estimadores son numéricamente idénticos.

- Transformación IG del modelo para el año 1988: $\widetilde{TM}_{it} = \beta_1 \widetilde{beertax}_{it} + \widetilde{u}_{it}$ con $t = 1982, 1988$.
- Debemos decirle a `gretl` que tenemos datos de panel, declarando las variables que identifican la unidad individual (en este caso, `state`) y el período temporal (en este caso, `year`):
  `setobs state year --panel-vars`
- Después utilizamos el comando `panel`. Podemos usar la opción `--time-dummies` para incluirlas en la estimación del modelo.
- Obtenemos $\widehat{\beta}_1^{IG} = -1{,}04097$, con un error estándar de 0,3475.
  - La estimación IG es idéntica a la estimación MCO con datos fusionados del modelo ampliado con dummys de estado.
  - **Pregunta**: ¿Cómo se diferencian los resultados de las estimaciones IG y de PD en este ejemplo?

“Efectos Aleatorios” (EA)

- Lo habitual es que los efectos individuales a_i estén correlacionados con las variables explicativas.
- El estimador IG es consistente se cumpla o no $C(a_i, X_{it}) = 0$, siempre que se verifique la **condición clave** $C(X_{it}, u_{is}) = 0 \forall t, s = 1, \ldots, T$.
- Sin embargo, conviene insistir en que $C(a_i, X_{it}) = 0$ es muy improbable.
- Cuando $C(a_i, X_{it}) = 0$ y se cumple la condición clave anterior:
  - El estimador IG es consistente pero ineficiente.
  - El estimador de MCO con datos fusionados (del modelo original) es también consistente, pero ineficiente, porque ignora la **estructura del término de error compuesto** $a_i + u_{it}$.
- Tradicionalmente, este marco de análisis se ha denominado incorrectamente “efectos aleatorios”, pero deberia denominarse **efectos aleatorios no correlacionados**.

Efectos Aleatorios (EA) no correlacionados

- Consideremos el modelo más simple

$$Y_{it} = \beta_0 + \beta_1 X_{it} + a_i + u_{it}, i = 1, \ldots, n; t = 1, \ldots, T.$$

- Si $C(X_{it}, u_{it}) = 0$ y $C(X_{it}, a_i) = 0$, el estimador MCO con datos fusionados del modelo proporcionará una estimación consistente de β_1.
  - Incluso con $T = 1$ (una única sección cruzada) obtendríamos una estimación consistente.
- Sin embargo, el término de error compuesto $v_{it} = a_i + u_{it}$ presenta autocorrelación debida al efecto individual.
- Bajo homocedasticidad y ausencia de autocorrelación en u_{it} ($C(u_{it}, u_{is}) = 0$, $\forall t \neq s$), definiendo $\sigma_a^2 = V(a_i)$, $\sigma_u^2 = V(u_{it})$:
  - $V(v_{it}) \equiv V(a_i + u_{it}) = \sigma_a^2 + \sigma_u^2$.
  - $C(v_{it}, v_{is}) = C(a_i + u_{it}, a_i + u_{is}) = \sigma_a^2$, $\forall t \neq s$.
- MCO con datos fusionados será ineficiente al ignorar esta autocorrelación.

El estimador de "Efectos Aleatorios" (EA)

- En lugar de MCO, podemos aplicar Mínimos Cuadrados Generalizados (MCG) Factibles, que tiene en cuenta la autocorrelación inducida por el efecto individual (aunque omitimos los detalles en este curso).
- El denominado estimador de EA (Efectos Aleatorios) es un estimador de MCG que aplica MCO a la siguiente transformación del modelo:

$$Y_{it} - \lambda \overline{Y}_i = \beta_0 (1 - \lambda) + \beta_1 \left(X_{it} - \lambda \overline{X}_i \right) + \left(v_{it} - \lambda \overline{v}_i \right)$$

donde $\overline{Z}_i = \frac{1}{T} \sum_{t=1}^{T} Z_{it}$ denota la media temporal de Z_{it}.

- λ es un parámetro desconocido, entre 0 y 1, que verifica $(1 - \lambda)^2 = 1/ \left(1 + T\sigma_a^2/\sigma_u^2\right)$. Por tanto, λ es creciente con la varianza relativa σ_a^2/σ_u^2.
  - A menor σ_a^2/σ_u^2 más se acercará λ a 0, y el estimador EA será similar al estimador MCO con datos fusionados.
  - A mayor σ_a^2/σ_u^2 más se acercará λ a 1, y el estimador EA será similar al estimador IG o de EF.

El estimador de "Efectos Aleatorios" (EA) (cont.)

- Para obtener el estimador de EA hay que estimar λ, lo que requiere estimaciones de σ_a^2 y σ_u^2, que pueden obtenerse a partir del estimador IG.
- El cálculo del estimador de EA es inmediato en `gretl`.
- Sin embargo, este marco de análisis es poco realista, porque ¡¡requiere que el error idiosincrático u_{it} no tenga autocorrelación y sea homocedástico!!
- Si hay heterocedasticidad y/o autocorrelación, la transformación de EA no es óptima. En ese caso, el estimador de EA no sería el estimador Factible de MCG y ya no sería eficiente.
- En la práctica, es muy improbable que los supuestos anteriores necesarios para la consistencia y eficiencia del estimador de EA se cumplan.
- En particular, los efectos individuales están habitualmente correlacionados con las variables explicativas, en cuyo caso el estimador de EA no es consistente (difiriendo sustancialmente del estimador IG).

"Efectos Fijos" vs "Efectos Aleatorios"

- Supongamos que se cumple la **condición clave** $C(X_{it}, u_{is}) = 0 \forall t, s = 1, \ldots, T$ y que el error idiosincrático es homocedástico y no presenta autocorrelación.
- Consideremos la hipótesis nula $H_0 : C(X_{it}, a_i) = 0$ frente a la alternativa $H_1 : C(X_{it}, a_i) \neq 0$.

Método de estimación	Hipótesis sobre $C(X_{it}, a_i)$	
	$H_0 : C(X_{it}, a_i) = 0$	$H_1 : C(X_{it}, a_i) \neq 0.$
IG o EF	consistente	consistente
EA	consistente eficiente	inconsistente

- El estimador IG/EF es consistente tanto bajo la nula como bajo la alternativa. Pero bajo la nula, el estimador de EA es consistente y además más eficiente que el estimador IG/EF.
- En ausencia de correlación entre efectos individuales y variables explicativas, el estimador de EA es preferible al IG/EF.

Contraste de "Efectos Fijos" vs "Efectos Aleatorios"

- Si la condición clave se cumple, podemos realizar un **contraste de Hausman** para evaluar si los efectos individuales están o no correlacionados con las variables explicativas. La hipótesis nula es $H_0 : C(X_{it,} a_i) = 0$.
- El contraste de Hausman valora dicha hipótesis evaluando si la diferencia entre los estimadores IG/EF y de EA es o no estadísticamente significativa.
- Este contraste es inmediato de obtener con `gretl` (aunque no vamos a derivarlo).

Contraste de "Efectos Fijos" vs "Efectos Aleatorios" (cont.)

- Nótese, sin embargo, que la validez del contraste de Hausman se basa en supuestos muy fuertes para garantizar la consistencia del estimador IG/EF.
  - En concreto, debe cumplirse la **condición clave** $C(X_{it}, u_{is}) = 0 \forall t, s = 1, \ldots, T$.
  - Si dicha condición no se cumple, el estimador IG/EF será inconsistente tanto bajo H_0 como bajo H_1 (y el estimador de EA también).
- Aunque se cumpla la condición clave, necesitaremos que el error indiosincrático sea homocedástico y no presente autocorrelación.
  - De lo contrario, el contraste de Hausman será inconsistente.
- Por tanto, la estrategia de confrontar el estimador de EF y el de EA basada en el contraste de Hausman no es recomendable.
- Esta estrategia está actualmente en desuso en la práctica.

Ejemplo EF vs EA: GMR

- Consideremos el modelo siguiente para gasto en innovación (inn) con datos de panel de empresas:

$$\texttt{inn}_{it} = \beta_{1j}\texttt{GMR}_{it}\left(j\right) + \beta_{2k}\texttt{Dsize}_{it}\left(k\right) + \beta_{3l}\texttt{Dsector}_{it}\left(l\right) + a_i + u_{it}$$

  donde $\texttt{GMR}\left(j\right)$ es el ámbito de mercado en que la empresa opera, con j entre 1 y 4 que corresponden a local (1), nacional (2), europeo (3) or internacional (4) y $\texttt{Dsize}\left(k\right)$, $\texttt{Dsector}(l)$ corresponden al tamaño y el sector de actividad de la empresa.
- Si GMR afecta positivamente a la innovación, β_{1j} será mayor cuanto mayor sea j.
- Podemos usar el "Panel de Innovación Tecnológica" (PITEC) para España, un panel incompleto o no equilibrado con información detallada para cada empresas sobre innovación desde 2003.
  - Nuestra muestra abarca de 2003 a 2008.

```
ols inn const dummify(gmr) dummify(year) dummify(sector) --robust
```

```
Model 1: Pooled
Included 5004 cross-sectional units
Time-series length: minimum 1, maximum 5
Dependent variable: inn
Robust (H
```

	coefficient	std. error	t-ratio	p-value	
const	0.504561	0.0472255	10.68	1.40e-26	***
Dgmr_2	0.0928605	0.0174111	5.333	9.73e-08	***
Dgmr_3	0.190534	0.0191665	9.941	3.08e-23	***
Dgmr_4	0.275561	0.0179527	15.35	6.64e-53	***
Dyear_2	-0.0347065	0.00787474	-4.407	1.05e-05	***
Dyear_3	0.0141995	0.00744819	1.906	0.0566	*
Dyear_4	0.0175798	0.00649652	2.706	0.0068	***
Dyear_5	-0.00381661	0.00481397	-0.7928	0.4279	
Dsector_2	0.0528567	0.0528875	0.9994	0.3176	
Dsector_3	0.105232	0.0467047	2.253	0.0243	**
Dsector_4	0.0568898	0.0481531	1.181	0.2374	
Dsector_5	0.0699164	0.0468809	1.491	0.1359	
Dsector_6	0.125310	0.0470796	2.662	0.0078	***
Dsector_7	0.0890264	0.0469188	1.897	0.0578	*
Dsector_8	0.0244243	0.0536698	0.4551	0.6491	

```
panel inn const dummify(gmr) dummify(year) dummify(sector)
--robust
```

```
# Fixed-effects
? panel inn const dummify(gmr) dummify(year) dummify(sector) --robust

Model 2: Fixed-effects, using 22548 observations
Included 5004 cross-sectional units
Time-series length: minimum 1, maximum 5
Dependent variable: inn
Robust (H


                 coefficient   std. error   t-ratio   p-value
  -----------------------------------------------------------
  const          0.411051     0.110877      3.707    0.0002    ***
  Dgmr_2         0.0189910    0.0163840     1.159    0.2464
  Dgmr_3         0.0565919    0.0188047     3.009    0.0026    ***
  Dgmr_4         0.0848563    0.0194780     4.357    1.33e-05  ***
  Dyear_2       -0.00946958   0.00767044   -1.235    0.2170
  Dyear_3        0.0220273    0.00718667    3.065    0.0022    ***
  Dyear_4        0.0223016    0.00629260    3.544    0.0004    ***
  Dyear_5       -0.00125518   0.00466273   -0.2692   0.7878
  Dsector_2      0.140703     0.130886      1.075    0.2824
  Dsector_3      0.213858     0.117661      1.818    0.0691    *
  Dsector_4      0.246948     0.121435      2.034    0.0420    **
  Dsector_5      0.136018     0.116954      1.163    0.2448
  Dsector_6      0.138575     0.120310      1.152    0.2494
  Dsector_7      0.201261     0.118028      1.705    0.0882    *
  Dsector_8      0.208629     0.134680      1.549    0.1214
  Dsector_9      0.291881     0.148914      1.960    0.0500    *
```

Ejemplo EF vs EA: GMR

"Efectos Aleatorios" (EA)

`panel inn const dummify(gmr) dummify(year) dummify(sector) --random-effects`

```
# Random-effects
? panel inn const dummify(gmr) dummify(year) dummify(sector) --random-effects

Model 4: Random-effects (GLS), using 22548 observations
Included 5004 cross-sectional units
Time-series length: minimum 1, maximum 5
Dependent variable: inn


               coefficient   std. error   t-ratio    p-value
  ----------------------------------------------------------
  const         0.529804     0.0402741    13.15      2.23e-39 ***
  Dgmr_2        0.0625103    0.0108095     5.783     7.44e-09 ***
  Dgmr_3        0.137182     0.0125538    10.93      1.00e-27 ***
  Dgmr_4        0.192742     0.0122588    15.72      2.08e-55 ***
  Dyear_2      -0.0142156    0.00652366   -2.179     0.0293   **
  Dyear_3       0.0192957    0.00619107    3.117     0.0018   ***
  Dyear_4       0.0214864    0.00612074    3.510     0.0004   ***
  Dyear_5      -0.00224029   0.00616410   -0.3634    0.7163
  Dsector_2     0.0563393    0.0483025     1.166     0.2435
  Dsector_3     0.122983     0.0430382     2.858     0.0043   ***
```

Ejemplo EF vs EA: GMR

El parámetro λ y el contraste de Hausman

- 'Within-variance' = 0.0837569 'Between-variance' = 0.130407
- Hausman test - Null hypothesis: GLS estimates are consistent
  - Asymptotic test statistic: Chi-square(29) = 374.727 with p-value = 9.52973e-62
- Nótese que el contraste de Hausman es válido sólo si el estimador IG es consistente, lo que requiere la **condición clave** $C(X_{it}, u_{is}) = 0 \forall t, s = 1, \ldots, T$, que es muy restrictiva.

Ventajas de datos de panel

- A diferencia de las secciones cruzadas o secciones cruzadas repetidas, los datos de panel permiten controlar la heterogeneidad individual inobservable (efectos individuales inobservables).
- Los efectos individuales a_i $(i = 1, \ldots, n)$ son variables aleatorias inobservables específicas de individuo e invariantes en el tiempo que recogen diferencias inobservables entre individuos que se mantienen constantes en el tiempo (al menos, durante el período muestral).
- Dichos efectos individuales están potencialmente correlacionados con las variables explicativas, por lo que el estimador MCO con datos fusionados será inconsistente.
- Para obtener estimaciones consistentes, debemos considerar transformaciones del modelo original que eliminen los efectos individuales, como PD o IG.
- Bajo **exogeneidad estricta** $C(X_{it}, u_{is}) = 0 \forall t, s$, el estimador MCO aplicado a las transformaciones PD o IG del modelo de interés permite obtener estimaciones consistentes de los parámetros de interés.

Aspectos prácticos con datos de panel

- Los efectos individuales inobservables son variables aleatorias, generalmente correlacionadas con las variables explicativas.
- Esta situación habitual desaconseja la estrategia desfasada y en desuso de confrontar los estimadores de EF/IG y de EA.
  - El estimador de EA se apoya en supuestos muy poco realistas.
  - Si dichos supuestos no se cumplen, el contraste de Hausman es inapropiado y puede llevar a conclusiones erróneas.
- Además, al igual que en el caso de secciones cruzadas, cabe esperar heterocedasticidad, debiendo utilizar errores estándar robustos a heterocedasticidad.
- También deben incluirse dummys temporales para controlar por shocks agregados comunes.
- En general, algunos individuos dejarán de responder en algún período y saldrán de la muestra, debiendo refrescar la muestra con nuevos individuos para mantener la representatividad. Por tanto, en general no tendremos observaciones para todos los individuos de la muestra en todo el período muestral. Además, el panel puede ser incompleto o no equilibrado, observando a los diferentes individuos en un número de períodos diferente (T_i).

Limitaciones de los datos de panel

- Frecuentemente, la **condición clave** $C\left(X_{it}, u_{is}\right) = 0 \forall t, s = 1, \ldots, T$ (**exogeneidad estricta**) no se cumple.
- Si alguna variable explicativa no es estrictamente exógena, los estimadores de PD e IG serán inconsistentes.
  - Obviamente, esta condición no se cumple en modelos dinámicos con la variable endógena desfasada entre las variables explicativas (no estudiados en este curso).
- Nótese que el estimador IG exige condiciones mucho más estrictas que el de PD, de manera que el estimador IG es menos robusto que el de PD.
- Si la condición clave no se cumple, debemos considerar métodos de variables instrumentales aplicados a la transformación de PD (que no vamos a estudiar aquí).