Economía Aplicada

Datos de Panel

Departmento de Economía
Universidad Carlos III de Madrid

Ver Wooldridge (capítulo 13), y Stock and Watson (capítulo 10)

Datos de Panel vs Secciones Cruzadas Repetidas

- En una sección cruzada cada observación representa a un individuo, o familia, o empresa, o estado, o país, etc.
- En la sección cruzada repetida se tiene más de un periodo, y las observaciones de cada periodo corresponden a distintos individuos o familias, etc.
- La diferencia entre secciones cruzadas repetidas y datos de panel es que el panel sigue a las mismas unidades: individuos, familias, etc., en los distintos periodos.
- Con datos de panel o también denominados datos longitudinales se requiere observar al mismo conjunto de unidades en al menos dos momentos del tiempo diferentes. En general consideraremos aplicaciones con muchas observaciones de sección cruzada pero no

Ejemplo de datos de panel

- Para obtener por ejemplo un panel sobre salarios, empleo, horas trabajadas, se selecciona aleatoriamente a un conjunto de individuos de la población en un momento del tiempo y se recoge esa información. En otro momento (el próximo mes, trimestre, año) se debe realizar la misma entrevista a los mismos individuos. Este ejemplo es el típico procedimiento de las encuestas de hogares y permite obtener la información de interés para el mismo grupo de individuos en diferentes periodos.
- En la primera clase mencionamos algunos ejemplos. Otro es la Encuesta de Presupuestos Familiares del INE. Desde 2006, el tamaño de muestra es de unos 24.000 hogares al año y cada hogar permanece en la muestra dos años consecutivos.

Uso de datos de panel

- Contar con datos de panel nos permite tener en cuenta algunos casos de variables omitidas, que con secciones cruzadas darían estimaciones MCO inconsistentes. Veremos el caso de variables omitidas que difieren entre individuos, pero que son constantes en el tiempo.
- Con datos de panel no es posible suponer que las observaciones son independientes: factores no observados que afectan al salario en un periodo también afectarán al salario el periodo siguiente.
- Por ello es necesario utilizar métodos especiales. Hoy veremos métodos que permiten eliminar los factores no observados que sean constantes en el tiempo.

Notación

Modelo Simple con T=2

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_0 d2_t + a_i + u_{it}, t = 1, 2$$

- El subíndice i hace referencia al individuo, empresa, etc., el subíndice t indica el momento del tiempo. En el caso simple de dos periodos tendremos $t = 1$ (para el primer periodo) y $t = 2$ (para el segundo).
- $d2$ es una variable que vale cero cuando $t = 1$ y uno cuando $t = 2$. Es igual para todos los individuos, por lo tanto no lleva subíndice i.
- La variable a_i captura todos los factores no observados que afectan a Y_{it} y que no cambian en el tiempo: por lo tanto no lleva subíndice t.

Heterogeneidad inobservada

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_0 d2_t + a_i + u_{it}$$

- La variable a_i no es observable, representa la heterogeneidad no observada e invariante (se suele llamar efecto no observado, efecto fijo, heterogeneidad inobservable o heterogeneidad individual).
- A este modelo se le suele llamar modelo de efectos no observados o modelo de efectos fijos.
- El término de error que ya conocemos es u_{it}. Incluye factores no observados que afectan a Y_{it} y que cambian en el tiempo.
- Al término $v_{it} = a_i + u_{it}$ se le suele llamar error compuesto: tiene un componente fijo y otro que cambia en el tiempo.

MCO con heterogeneidad inobservada

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_0 d2_t + v_{it}$$

- Para obtener estimaciones consistentes de β_1 usando MCO debemos suponer que X_{it} no está correlacionada con v_{it}.
- Incluso si suponemos que $C(X_{it}, u_{it}) = 0$, la estimación MCO será sesgada e inconsistente si $C(X_{it}, a_i) \neq 0$.
- El sesgo que surge al omitir a_i al aplicar MCO a datos fusionados se suele llamar sesgo de heterogeneidad, surge de omitir una variable que difiere entre individuos pero no cambia en el tiempo.

Ejemplo: mortalidad e impuestos al alcohol

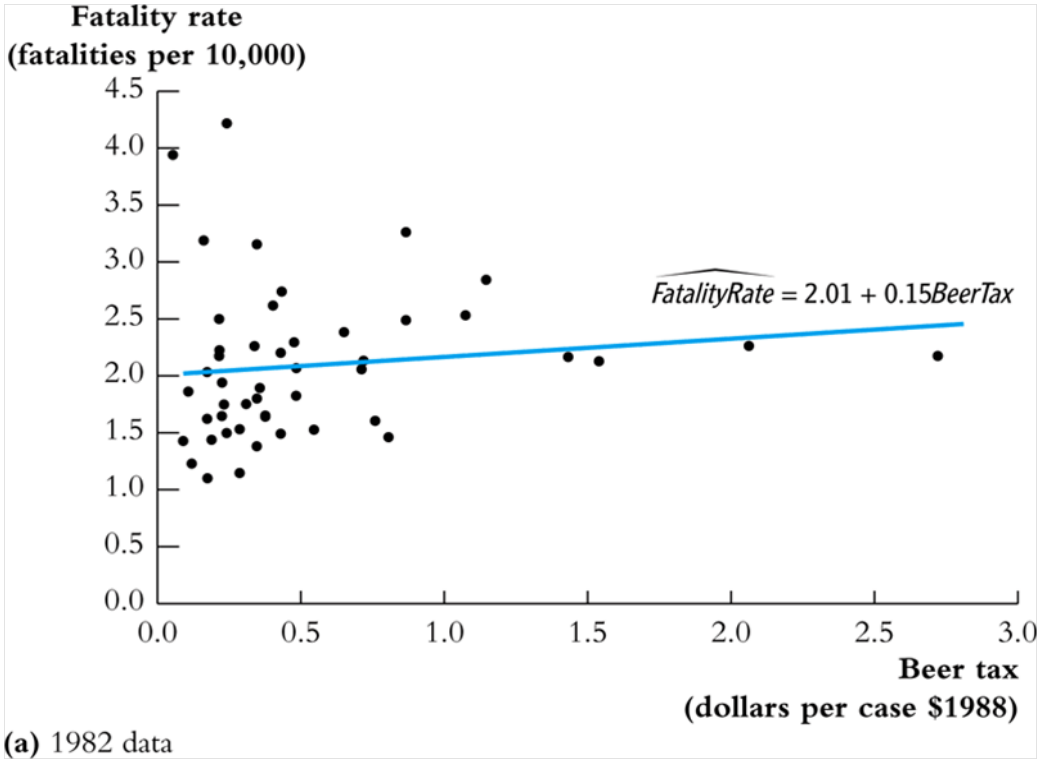
- Estudiar la eficacia de políticas para reducir el número de conductores ebrios.
- Base fatality.gdt en `gretl`. Datos para EEUU para 48 estados. Usaremos los años 1982 y 1988: i es un estado, t un año.
- Variables:

  - tasa de mortalidad (muertes en accidentes de tráfico cada 10.000 habitantes en cada estado, cada año) (variable *TM*)

  - impuesto sobre la cerveza (dólares por caja en dólares de 1988) (variable *ImpCerv*).

Ejemplo: usando una sección cruzada

Modelo simple: $TM_i = \beta_0 + \beta_1 ImpCerv_i + u_i$



(a) 1982 data

El coeficiente no es estadísticamente significativo para 1982. Si usamos los datos fusionados es positivo y significativo (ver en `gretl`). ¿Un impuesto sobre la cerveza provoca más muertes en accidentes de tráfico?

Posibles explicaciones de estos resultados

- Muchos factores afectan a la tasa de mortalidad:
  - Calidad (edad) de los automóviles
  - Calidad de las carreteras
  - Aspectos culturales sobre beber y conducir
  - Densidad de coches
- Si alguno de estos factores está correlacionado con el nivel impositivo sobre la cerveza...
- Podríamos intentar conseguir datos sobre esos factores. Pero no todos son fáciles de medir.
- Con datos de panel podemos obtener estimaciones consistentes si los factores que no observamos, y que están correlacionados con el nivel impositivo sobre la cerveza, se mantienen constantes en el periodo analizado.

Heterogeneidad inobservada en datos de panel

- Con datos de panel podemos permitir que los efectos no observados (a_i) estén correlacionados con las variables explicativas.
- Dado que a_i no depende del tiempo, podemos tomar las primeras diferencias de los datos en los dos años y deshacernos de a_i.

$$Y_{i2} - Y_{i1} = (\beta_0 + \beta_1 X_{i2} + \delta_0 + a_i + u_{i2}) - (\beta_0 + \beta_1 X_{i1} + a_i + u_{i1})$$

$$Y_{i2} - Y_{i1} = \delta_0 + \beta_1 (X_{i2} - X_{i1}) + (u_{i2} - u_{i1})$$

Que se puede escribir:

$$\Delta Y_i = \delta_0 + \beta_1 \Delta X_i + \Delta u_i$$

donde Δ hace referencia al cambio de $t = 1$ a $t = 2$.

Estimador de Primeras Diferencias

$$\Delta Y_i = \delta_0 + \beta_1 \Delta X_i + \Delta u_i$$

- Esta ecuación en primeras diferencias es como una ecuación de datos de sección cruzada, donde cada variable aparece en diferencias en la dimensión temporal.
- El intercepto representa el cambio en Y en ausencia de un cambio en X.
- El supuesto clave es que Δu_i no esté correlacionado con ΔX_i. Si se verifica, MCO dará estimaciones consistentes.
- El estimador MCO de β_1 en esta ecuación es el estimador de las primeras diferencias.

Algunas características del estimador de Primeras Diferencias

- Permite correlación entre a_i y los controles X.
- Bajo condiciones generales $plim(\hat{\beta}) = \beta$ cuando $N \to \infty$ y T está fijo.
- No se obtiene ninguna estimación del efecto fijo inobservable a_i.
- Todos los controles que sean invariantes en el tiempo también desaparecen al tomar primeras diferencias.
- Por ello ΔX_i debe variar con i. Esto puede ser un inconveniente si ΔX_i varía muy poco. Nota: a veces puede ser mejor tomar diferencias entre periodos de tiempo más alejados.

Ejemplo mortalidad: datos de panel

Modelo simple

$$TM_{it} = \beta_0 + \beta_1 ImpCerv_{it} + \delta_0 d1988_t + a_i + u_{it}, \text{ con } t = 1982, 1988$$

- a_i es una variable inobservable, que no cambia en el tiempo y que afecta la tasa de mortalidad en el estado i.
- Podría incluir la actitud local hacia beber y conducir (que si cambia lentamente podemos considerarla constante entre 1982 y 1988).
- Estados con una actitud menos favorable al consumo de alcohol tendrán en promedio menos muertes en accidentes de tráfico y probablemente tendrán impuestos más altos sobre el alcohol.
- Sería un caso de sesgo de variable omitida: *ImpCerv* capturaría en parte el efecto de la actitud local hacia beber y conducir.

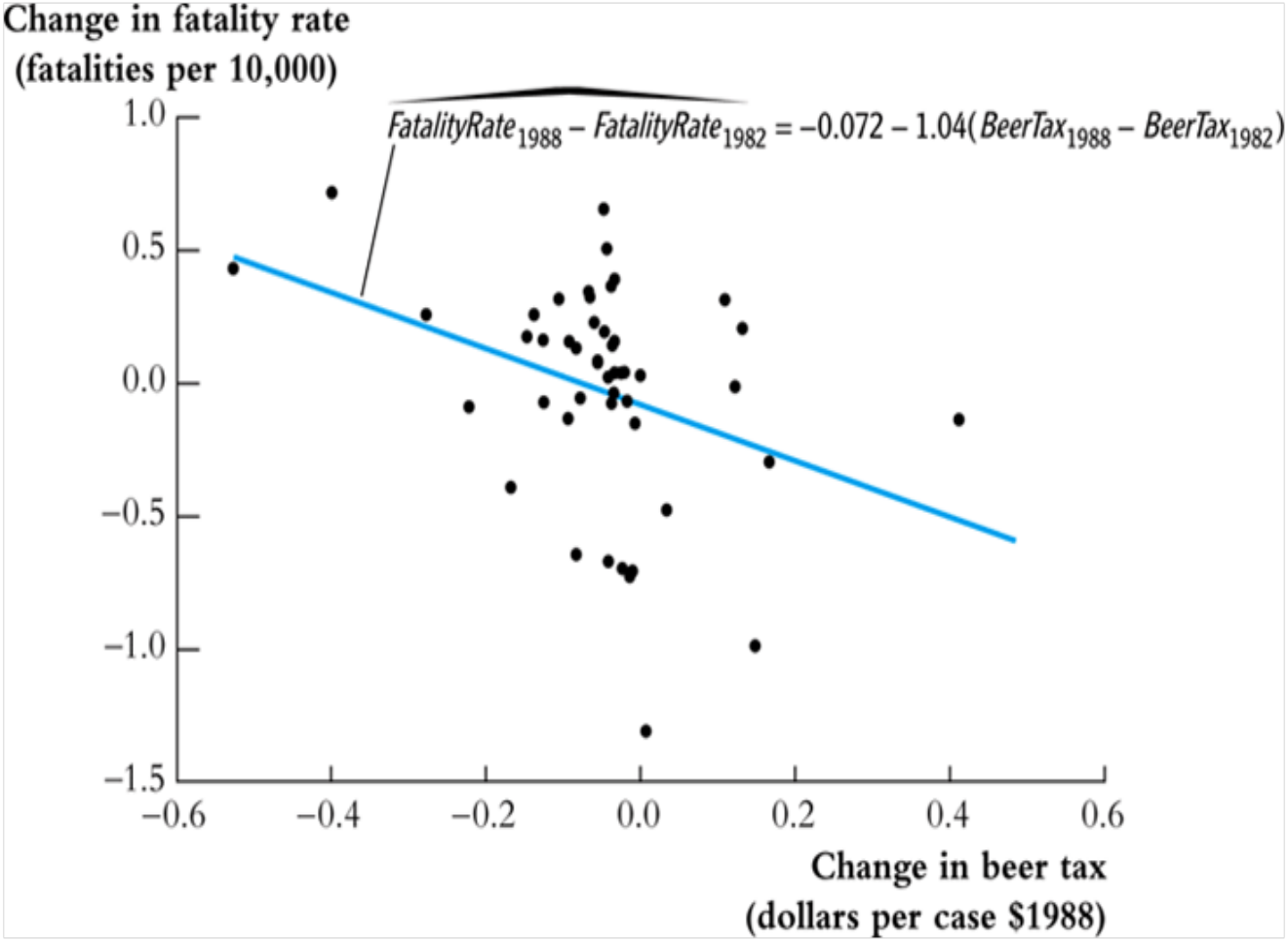
Ejemplo mortalidad: primeras diferencias

- La influencia de a_i puede eliminarse si se toman primeras diferencias.

$$\Delta TM_i = \delta_0 + \beta_1 \Delta ImpCerv_i + \Delta u_i$$

- Intuitivamente:
  - Las actitudes hacia el consumo de alcohol afectan el nivel de conducción en estado de ebriedad y por tanto la tasa de mortalidad en un estado.
  - Sin embargo, si esas actitudes no cambiaron entre 1982 y 1988, no pudieron tener efecto en el cambio en el número de muertes en el estado.
  - Cualquier variación en TM tiene que surgir de otros factores. En la ecuación anterior esos factores son los cambios en el impuesto sobre la cerveza y los cambios en u_i.

Ejemplo: resultados primeras diferencias



Al contrario que cuando usamos secciones cruzadas, un incremento en el impuesto disminuye las muertes en accidentes de tráfico.

Ejemplo: resultados primeras diferencias

- El intercepto permite que el cambio medio en *TM* si no hay cambio en el impuesto sobre la cerveza, sea distinto de cero. En este caso, el valor negativo podría reflejar las mejoras en la seguridad de los vehículos y las carreteras que redujeron el número de muertes.
- Un incremento en el impuesto de 1$ por caja reduce la mortalidad por accidentes de tráfico en promedio en 1,04 muertes por cada 10.000 personas.
- El efecto estimado es grande: la tasa de mortalidad promedio en los datos es de 2 muertos cada 10.000 habitantes por año. La estimación sugiere que las muertes podrían reducirse a la mitad con un incremento del impuesto de 1$ por caja de cerveza.

Ejemplo mortalidad: comentarios

- La regresión en diferencias controla por todos los factores inobservables constantes entre 1982 y 1988, por ejemplo las actitudes hacia la bebida y la conducción.
- Notar que se necesita variación en *ImpCerv* para poder estimar el efecto de los impuestos.
- Pueden existir factores que influyen en la seguridad del tráfico, correlacionados con el nivel impositivo sobre el alcohol, y que cambien en el tiempo. Su omisión dará lugar a un sesgo que no discutiremos aquí.

Primeras Diferencias con más periodos

- Es posible usar la misma estrategia con $T > 2$: podemos tomar diferencias de periodos consecutivos: $t - (t-1)$
- Por simplicidad supongamos $T = 3$:

$$Y_{it} = \delta_1 + \delta_2 d2_t + \delta_3 d3_t + \beta_1 X_{it} + a_i + u_{it}, \text{ para } t = 1, 2, 3$$

- Estamos interesados en β_1. Si a_i correlacionado con X_i, usar MCO dará estimaciones inconsistentes.
- Supuesto clave: $C(X_{it}, u_{is}) = 0$ para todo t, s: una vez eliminado el efecto de a_i, X_i es exógena.

Tomando diferencias eliminamos a_i

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

Primeras Diferencias con más periodos (cont.)

- Supuesto clave para que MCO sea consistente: ΔX_{it} no correlacionado con Δu_{it} para $t = 2, 3$.
- Se puede mostrar que la ecuación anterior es equivalente a incluir una constante y eliminar una variable binaria temporal. Por lo tanto en la práctica se estima:

Primeras diferencias con $T = 3$

$$\Delta Y_{it} = \alpha_0 + \alpha_1 d3_t + \beta_1 \Delta X_{it} + \Delta u_{it}, \text{ para } t = 2, 3$$

- Para $T > 3$ se procede de manera similar: se tendrán $T - 1$ observaciones por unidad y se incluyen $T - 2$ variables binarias temporales.

Otra opción con $T > 2$

- Bajo ciertos supuestos (que no discutimos) es mejor utilizar otro método en lugar de primeras diferencias para eliminar la heterogeneidad individual a_i.
- Partamos del modelo simple:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + a_i + u_{it},\ t = 1, 2, \ldots, T$$

- Para cada i, se escribe la ecuación para los valores medios:

$$\bar{Y}_i = \beta_0 + \beta_1 \bar{X}_i + a_i + \bar{u}_i, \text{ donde } \bar{Y}_i = T^{-1} \sum_{t=1}^{T} Y_{it}.$$

- Y se toma la diferencia de ambas ecuaciones para cada t:

$$Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - \bar{X}_i) + u_{it} - \bar{u}_i,\ t = 1, 2, \ldots, T$$

Desviaciones con respecto a la media

- Que también se suele expresar:

Desviaciones con respecto a la media

$$\ddot{Y}_{it} = \beta_1 \ddot{X}_{it} + \ddot{u}_{it},\ t = 1, 2, \ldots, T$$

- donde $\ddot{Y}_{it} = Y_{it} - \bar{Y}_i$: los datos de Y en desviaciones con respecto a la media.
- El efecto a_i ha desaparecido, con lo que podemos estimar la ecuación usando MCO.
- El estimador MCO de esta ecuación se denomina estimador intragrupos o de desviaciones con respecto a la media. También se conoce como estimador de efectos fijos.

Efectos fijos

- En economía aplicada se suele pensar a_i como un parámetro a estimar para cada i. Con esa idea, a_i sería el intercepto para la unidad i y se podría estimar junto con β_1.
- La manera de hacerlo es incluir una variable binaria para cada unidad i.
- Cada variable binaria se denomina efecto fijo. Notar que no se trata de parámetros fijos, sino que simplemente para cada individuo son constantes en el periodo analizado.
- Se puede mostrar que el estimador intragrupos es equivalente al estimador que surge del modelo con variables binarias individuales.

Características del estimador intragrupos o de efectos fijos

- El estimador intragrupos o de efectos fijos para ser consistente requiere que u_{it} esté incorrelacionado con X_{it} en todos los periodos.
- Al igual que el estimador de primeras diferencias, el estimador intragrupos permite cualquier tipo de correlación entre a_i y las variables explicativas. Y también desaparecerán las variables explicativas que sean constantes en el tiempo.
- Para el caso más común de N grande y T pequeño, la elección entre primeras diferencias o intragrupos se basa en la eficiencia relativa de los estimadores ya que ambos serán consistentes.
- En la mayoría de los casos y con más de dos periodos se verifica que el estimador intragrupos es más eficiente.

Ejemplo mortalidad: estimador intragrupos

$$\ddot{TM}_{it} = \beta_1 \ddot{ImpCerv}_{it} + \ddot{u}_{it}, \text{ con } t = 1982, 1988$$

- Lo veremos en los ordenadores en detalle, aquí lo vemos resumidamente.
- Primero debemos establecer que nuestros datos son un panel, declarando la variable de unidad y la de tiempo (`setobs` state year −−panel−vars).
- Luego usamos el comando `panel` (con la opción −−time-dummies si queremos agregar variables temporales).
- Obtenemos: $\hat{\beta}_1 = -1.04097$ con un error estándar de 0.3475.
- Notar que es el mismo resultado que al incluir variables binarias para cada estado.

Efectos Aleatorios (RE)

- El método de efectos fijos (FE) transforma el modelo para eliminar la heterogeneidad inobservada dando un estimador consistente sin imponer ningún supuesto adicional sobre a_i.
- El estimador de efectos aleatorios (RE: Random Effects) es útil si asumimos que la heterogeneidad no observada no está correlacionada con los controles.
- Si incluimos las variables adecuadas en nuestro modelo, podemos pensar que la heterogeneidad no observada produce autocorrelación en los términos de error pero no entre el término de error y los controles.
- Si este supuesto adicional es correcto, el estimador de efectos aleatorios será consistente y más eficiente que el estimador de efectos fijos, dado que utiliza más información.

Modelo de Efectos Aleatorios

- Modelo con heterogeneidad inobservada

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + \beta_2 X_{it2} + ... + \beta_k X_{itk} + a_i + u_{it},\ t = 1, 2, ..., T$$

- Notar que X puede incluir dummies temporales.
- Supongamos que a_i no está correlacionada con X de cada periodo. En ese caso, si transformamos el modelo para eliminar a_i obtendremos estimadores infeficientes.
- Este modelo se llama modelo de efectos aleatorios cuando se asume que la heterogeneidad no observada a_i está incorrelacionada con cada una de los controles:
  $Cov(X_{itj}, a_i) = 0, t = 1, 2, ..., T; j = 1, 2, ..., k$

MCO con Efectos Aleatorios

- Si creemos que la heterogeneidad individual está incorrelacionada con las variables explicativas, los coeficientes β_j pueden estimarse consistentemente usando una simple regresión con datos de sección cruzada: no necesitamos datos de panel.
- Pero usar una única regresión con datos de sección cruzada implica no considerar información muy útil. Podemos usar entonces una estimación por MCO fusionados y obtener estimadores consistentes.
- Pero seguimos sin tener en cuenta que el término de error compuesto $v_{i,t} = a_i + u_{it}$, presenta autocorrelación a lo largo del tiempo.
- Como los errores estándar de MCO fusionados no tienen en cuenta esta autocorrelación, estarán mal calculados, al igual que los estadísticos de contraste.

El estimador de Efectos Aleatorios

- Un método llamado mínimos cuadrados generalizados (MCG) arroja estimadores asintóticamente eficientes (no vemos los detalles aquí).
- La transformación por *MCG* es la siguiente:

$$y_{it} - \lambda \overline{y}_i = \beta_0(1-\lambda) + \beta_1 \left(X_{it1} - \lambda \overline{X}_{i1}\right) + (v_{it} - \lambda \overline{v}_i)$$

  donde $\overline{y}_i$ representa valores medios temporales, y λ está entre 0 y 1 y depende de las varianzas de a_i y u_{it}, y de T:
  - Si a_i es relativamente poco importante, λ estará cerca de 0 y el estimador de efectos aleatorios será similar a *MCO*
  - Si a_i es importante, λ estará cerca de 1 y el estimador de efectos aleatorios será similar al de efectos fijos
- MCG es el estimador MCO de la ecuación transformada (se computa de manera directa en `gretl`).

Efectos Fijos vs. Efectos Aleatorios

- Si estamos interesados en el efecto de una variable que no varía en el tiempo en un estudio de datos de panel, usar efectos fijos no es útil.
- Sin contar con una VI, lo cual no es fácil, Efectos Aleatorios es probablemente la única opción.
- Es necesario agregar variables que capturen la parte de a_i correlacionada con X, por ejemplo incluir variables dummy para grupos (si contamos con suficientes observaciones por grupo). Por ejemplo, si tenemos datos de panel a nivel de estudiante, podemos incluir variables dummy por escuela.
- Incluir variables dummy para grupos controla hasta cierto punto la heterogeneidad que pueda estar correlacionada con los elementos invariantes de X.

Contraste Efectos Fijos vs. Efectos Aleatorios

- Si la heterogeneidad no observada invariante está correlacionada con los controles, el estimador de efectos fijos es consistente y el de efectos aleatorios no
- Si los supuestos del modelo de efectos aleatorios son correctos, ambos estimadores serán consistentes, pero efectos aleatorios será más eficiente
- Hausman: contrastar si la heterogeneidad no observada está correlacionada con los controles comparando ambos estimadores
  - H_0 : supuestos del modelo de efectos aleatorios son correctos $\Rightarrow$ diferencias entre estimador de efectos aleatorios y efectos fijos no significativas
  - H_1 : supuestos del modelo de efectos aleatorios son falsos $\Rightarrow$ diferencias entre estimador de efectos aleatorios y efectos fijos significativas

El contraste de Hausman

- El contraste de Hausman se basa en la diferencia entre ambos estimadores: efectos fijos vs. efectos aleatorios
- Una diferencia estadísticamente significativa sugiere que la hipótesis nula es falsa, y es por tanto evidencia en contra del modelo de efectos aleatorios
- No derivamos el contraste, pero sabemos que bajo la hipótesis nula, $H \xrightarrow{a} \chi^2_K$ donde K es el número de regresores
- Se implementa muy fácilmente en `gretl`.

Innovación en las empresas

- hay consenso sobre los efectos de la innovación en el crecimiento
- los determinantes de la innovación han sido menos estudiados: rango del mercado geográfico, *GMR*
  - Las empresas innovadoras con éxito están en condiciones de conquistar nuevos mercados para sus productos
  - Las empresas con una gran presencia en mercados distantes tienen más probabilidades de incurrir en costes de innovación para mantener las cuotas de mercado
  - Factores no observables: gestión de alta calidad

GMR

- Consideramos el modelo de efectos no observables

$$innovation_{it} = \beta_{1j} GMR_{it}(j) + \beta_{2k} Dsize_{it}(k) + \beta_{3l} Dsector(l) + a_i + u_{it}$$

- $GMR(j)$: rango geográfico 1: local, 2: nacional, 3: europeo, 4: internacional
- Si GMR afecta la innovación positivamente, β_{1j} aumentaría con j
- Podemos usar el "Panel de Innovación Tecnológica" o PITEC. Un panel no balanceado de empresas con información detallada sobre la innovación desde 2003 a 2008

MCO Fusionados

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
```

Efectos Fijos

```
# 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	*

Efectos Aleatorios

```
# 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   ***
```

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