



Universidad de Sevilla

Facultad de Matemáticas

Trabajo Fin de Estudios

Grado en Matemáticas

Modelos de Datos Panel

Carlos Pérez Manzano

Tutor

Nombre: Rafael Pino Mejías

Dpto. Estadística e Investigación Operativa

Abstract

The purpose of this project is the development of several panel data models, a type of data with a large presence in fields such as economics, sociology, demography and political science.

We will focus on one-way and two-way error component models. We will build these models and carry out associated hypothesis testing. In addition we will address the problems of poolability and endogeneity. Finally we will make a comparison and interpretation of the results obtained by applying the different alternatives developed in the project.

Resumen

El propósito de este proyecto es el desarrollo de varios modelos de datos panel, un tipo de datos con una gran presencia en ámbitos como la economía, sociología, demografía y ciencias políticas.

Nos centraremos en los modelos de componentes de error unidireccional y bidireccional. Vamos a construir estos modelos y llevar a cabo contrastes de hipótesis asociados. Además abordaremos los problemas de la agrupabilidad y endogeneidad. Finalmente realizaremos una comparación e interpretación de los resultados obtenidos mediante la aplicación de las distintas alternativas desarrolladas en el proyecto.

Índice

Introducción	3
1. Modelo de componente de error unidireccional	5
1.1. Introducción y conceptos previos	5
1.2. Hipótesis sobre los errores	7
1.3. Modelo agrupado	8
1.4. Modelo de efectos fijos	8
1.4.1. Estimador LSDV	8
1.4.2. Estimador intragrupos	9
1.5. Modelos de efectos aleatorios	10
1.5.1. Estimador mediante MCG	10
1.5.2. Estimación de las varianzas de las componentes del error	12
1.6. Comparación de los estimadores	13
1.6.1. Elección entre el modelo de efectos fijos y modelo de efectos aleatorios . .	14
2. Modelo de componente de error bidireccional	15
2.1. Introducción	15
2.2. Hipótesis sobre los errores	16
2.3. Modelo de efectos fijos	16
2.4. Modelo de efectos aleatorios	18
2.4.1. Estimación de la varianza de las componentes del error	19
3. Tests para los modelos de componentes de error	21
3.1. Contrastes de agrupabilidad de los datos	21
3.1.1. Contraste de agrupabilidad con $\varepsilon \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{\mathbf{NT}})$	22
3.1.2. Contraste de agrupabilidad con $\varepsilon \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{\Sigma})$	23
3.2. Contraste sobre los efectos individuales y temporales	23
3.2.1. Caso del modelo de componente de error unidireccional	23
3.2.2. Caso del modelo de componente de error bidireccional	24
3.3. Contraste sobre la correlación de los efectos individuales	25
3.3.1. Test de Hausman	25
4. Coeficientes heterogéneos	27
4.1. Introducción	27
4.2. Coeficientes fijos	28
4.3. Coeficientes aleatorios	28
4.3.1. El estimador de Swamy	28

4.3.2. Estimador de Grupos Medios	30
5. Endogeneidad	31
5.1. Introducción	31
5.2. Método de Variables Instrumentales	32
5.2.1. Estimadores mediante VI	33
5.3. Estimadores mediante VI para el modelo de componente de error unidireccional .	34
5.3.1. Estimador intragrupos mediante VI	34
5.3.2. Estimador para el modelo general	34
6. Caso práctico en R	37
6.1. Carga de los paquetes y del dataset	38
6.2. Análisis descriptivo	39
6.2.1. Visualización de la variable objetivo <code>violent</code>	40
6.2.2. Estudio del factor <code>law</code>	42
6.3. Preparación para los modelos	44
6.4. Modelos de componente de error unidireccional	46
6.4.1. Modelo de efectos fijos	46
6.4.2. Modelos de efectos aleatorios	47
6.4.3. Comparación de los modelos de componente de error unidireccional	48
6.4.4. Verificación y elección del modelo	51
6.5. Modelos de componente de error bidireccional	52
6.5.1. Modelo de efectos fijos	52
6.5.2. Modelos de efectos aleatorios	54
6.5.3. Comparación de los modelos de componente de error bidireccional	56
6.5.4. Verificación y elección del modelo	57
6.6. Comparación de los modelos escogidos e interpretación de los resultados	58
6.7. Algunas consideraciones	61
Bibliografía	63

Introducción

Un diseño de investigación que recoge información de las mismas unidades a lo largo de un cierto periodo de tiempo es denominado un *panel*.

	country	year	exports	imports	resimp	gnp	pgnp	gnpw	pm	px	cpi	pw	exrate	consump	invest	income	pop
1	Indonesia	1964	-7.61279	-8.96084	-14.74513	0.12960	-0.005550435	-2.21541	-0.28160	-1.96169	-2.81341	-0.58577	0.02402	0.02576	-1.31476	0.00851	11.537325
2	Indonesia	1965	-7.65318	-8.96664	-14.96593	0.01653	0.055664506	-2.16165	0.34958	-1.33965	-1.51413	-0.56591	0.04505	-0.03884	-1.36559	-0.11283	11.562382
3	Indonesia	1966	-7.67038	-9.30862	-14.60063	0.15908	0.116879447	-2.11230	1.54867	-0.20691	0.90016	-0.53332	0.14711	0.22324	-1.38655	0.05033	11.588311
4	Indonesia	1967	-7.67121	-9.12042	-16.19614	0.13064	0.178094387	-2.09490	2.04485	0.23114	1.89462	-0.53132	0.23858	0.14438	-1.38932	0.01889	11.613766
5	Indonesia	1968	-7.42591	-8.79655	-13.90201	0.19652	0.239309328	-2.05700	2.72725	0.91458	2.70605	-0.50662	0.47256	0.17004	-1.30994	0.08316	11.639470
6	Indonesia	1969	-7.47061	-8.67138	-13.76654	0.27053	0.300524269	-2.04203	2.85192	1.16096	2.86676	-0.46834	0.51995	0.22435	-1.26739	0.16088	11.665303
7	Indonesia	1970	-7.83957	-8.62897	-13.63308	0.33289	0.361739209	-2.06173	3.00580	1.77602	2.98265	-0.43268	0.57869	0.23743	-1.23001	0.21376	11.690821
8	Korea	1964	-9.46698	-7.69541	-11.22661	0.86738	0.848420326	-2.21541	2.33640	2.44948	2.38601	-0.58577	0.35208	0.77778	-1.09566	0.63825	10.228465
9	Korea	1965	-9.14336	-7.60083	-11.31586	0.83416	0.934504605	-2.16165	2.57471	2.70807	2.51366	-0.56591	0.43860	0.75943	-1.06268	0.58739	10.251677
10	Korea	1966	-8.89550	-7.16218	-11.25729	0.94844	1.020588883	-2.11230	2.56835	2.81323	2.62973	-0.53332	0.44673	0.82426	-0.58408	0.70759	10.273671
11	Korea	1967	-8.73142	-6.88195	-11.25249	1.02022	1.106673162	-2.09490	2.57555	2.85337	2.72981	-0.53132	0.44538	0.90127	-0.49778	0.79354	10.313277
12	Korea	1968	-8.36854	-6.52384	-11.55448	1.14827	1.192757441	-2.05700	2.58773	2.90750	2.83321	-0.50662	0.45547	0.98664	-0.20368	0.93244	10.336568
13	Korea	1969	-8.03694	-6.31016	-11.45235	1.27316	1.278841719	-2.04203	2.61401	2.89721	2.95126	-0.46834	0.47443	1.06636	0.02936	1.04517	10.359012
14	Korea	1970	-7.80942	-6.27771	-11.46572	1.34092	1.364925998	-2.06173	2.72370	3.01558	3.10009	-0.43268	0.51130	1.17145	-0.03206	1.10618	10.380963

Figura 1: Ejemplo base de datos `ForeignTrade` del paquete `pder` de R

En un principio, los estudios sobre datos panel tenían como unidades a individuos de una población, que se seleccionaban generalmente de forma aleatoria. Para estos individuos se recopilaba la información en forma de encuestas. Esto solía provocar el llamado problema del *desgaste o mortalidad del panel*, que hace referencia a la situación en la que un grupo de individuos abandona el panel provocando pérdida de información, sesgo y en consecuencia dificultando la obtención de buenos resultados.

No obstante, con el paso de los años son más los estudios en los que las unidades del diseño no son individuos, sino entidades sociales (por ejemplo empresas, países o regiones), gozando en estos casos de una gran fuente de información de calidad, haciendo disminuir el desgaste del panel. La frecuencia de este tipo de datos en ámbitos como la economía, la sociología y las ciencias políticas ha provocado el desarrollo de modelos más sofisticados que cuentan con las siguientes ventajas respecto de los que han sido empleados tradicionalmente:

1. **Controlar la heterogeneidad individual y temporal no observada:** los modelos de datos panel permiten controlar las diferencias entre las unidades del panel, así como las

variaciones temporales de esas unidades. Este es un problema de los modelos de regresión clásicos pues en general las diferencias entre individuos o periodos de tiempo no son medibles o cuantificables.

2. **Reducción del sesgo por omisión de variables:** habitualmente no conocemos todos los factores que afectan a nuestra variable objetivo, luego los modelos pueden estar incompletos y en consecuencia sesgados. Si poseemos datos transversales, es decir, tomados en el mismo instante de tiempo, no podemos hacer más que ciertas suposiciones que simplifican el efecto de estas variables omitidas. Sin embargo, los datos panel nos van a permitir tener en cuenta ciertas variables omitidas, al menos aquellas que son constantes en el tiempo para cada unidad del panel, o bien variables que son constantes en las unidades para cada instante de tiempo.
3. **Obtención de muestras más grandes en un menor tiempo:** por ejemplo, si una empresa recoge datos internos mensuales desde hace 3 años, el estudio de las series temporales no es muy adecuado debido a la escasez de observaciones. Sin embargo, pueden incluirse en el estudio empresas con características similares, consiguiéndose de esta manera una muestra con un tamaño considerable para la que sí va a ser posible obtener conclusiones sólidas para tomar decisiones en la empresa.

Los modelos que presentaremos a continuación son relevantes en los casos en que los efectos marginales de las covariables sean iguales para todas las unidades del panel, siendo el intercepto a priori diferente para cada individuo o para cada observación. Posteriormente, en el capítulo 4 se dará una solución para los casos en los que lo anterior no se cumpla y los paneles cuenten con ciertas propiedades.

Existen dos tipos de paneles, aquellos para los que se cuentan con el mismo número de observaciones y en los mismos instantes de tiempo para todas las unidades del modelo (*paneles balanceados*) y aquellos en los que no se cumple la condición anterior, por ejemplo causado por un desgaste del panel (*paneles no balanceados*). Para una mayor comodidad en los cálculos consideraremos que los paneles en los modelos son balanceados, aunque todos los resultados que se presentan pueden ser extendidos al caso no balanceado.

Capítulo 1

Modelo de componente de error unidireccional

1.1. Introducción y conceptos previos

El modelo muestral más sencillo y general de datos panel para K variables regresoras se escribe de la siguiente forma:

$$y_{nt} = \alpha + x_{nt}^t \beta + \varepsilon_{nt}, \quad n = 1, \dots, N, \quad t = 1, \dots, T \quad (1.1)$$

siendo este una extensión del modelo clásico de regresión lineal. El subíndice n indica el individuo, mientras que el subíndice t representa el instante de tiempo. Para la muestra completa denotamos por y , X y ε como sigue:

$$y = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ y_{N2} \\ \vdots \\ y_{NT} \end{pmatrix}, \quad X = \begin{pmatrix} x_{11}^1 & x_{11}^2 & \cdots & x_{11}^K \\ x_{12}^1 & x_{12}^2 & \cdots & x_{12}^K \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T}^1 & x_{1T}^2 & \cdots & x_{1T}^K \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1}^1 & x_{N1}^2 & \cdots & x_{N1}^K \\ x_{N2}^1 & x_{N2}^2 & \cdots & x_{N2}^K \\ \vdots & \vdots & \ddots & \vdots \\ x_{NT}^1 & x_{NT}^2 & \cdots & x_{NT}^K \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1T} \\ \vdots \\ \varepsilon_{N1} \\ \varepsilon_{N2} \\ \vdots \\ \varepsilon_{NT} \end{pmatrix}$$

Considerando además j_{NT} un vector formado por unos de tamaño NT obtenemos:

$$y = \alpha j_{NT} + X\beta + \varepsilon \quad (1.2)$$

Suele ser más conveniente expresar (1.2) como:

$$y = Z\gamma + \varepsilon \quad (1.3)$$

donde $Z = (j_{NT}, X)$ y $\gamma = (\alpha, \beta)^t$.

Para el modelo de componente de error unidireccional descomponemos el error en la forma $\varepsilon_{nt} = \eta_n + \nu_{nt}$.

Por un lado η_n representa el efecto individual de n , que se considera constante a lo largo del tiempo y manifiesta la heterogeneidad individual no observada. Por ejemplo, en un estudio acerca del rendimiento académico de un número de estudiantes de diferentes escuelas a lo largo del curso, esta componente representa la inteligencia, habilidad o motivación inherente al sujeto. Si los individuos en el estudio son empresas podríamos pensar que la componente individual manifiesta la reputación de la misma o las habilidades empresariales y de gestión de los líderes de la entidad.

Por otro lado ν_{nt} es la parte idiosincrática o aleatoria del error, que representa la variabilidad incontrolable específica de cada observación.

Sea $\eta = (\eta_i)_{i=1}^N$, y ν el vector de longitud NT conteniendo la parte idiosincrática, ε puede escribirse en forma matricial haciendo uso del producto de Kronecker.

Definición 1.1 (Producto de Kronecker) Sea A una matriz $n \times m$ y B una matriz de tamaño $p \times q$, se define el producto de Kronecker $A \otimes B$ como la matriz bloque $np \times mq$ siguiente:

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1m}B \\ a_{21}B & \cdots & a_{2m}B \\ \vdots & \vdots & \vdots \\ a_{n1}B & \cdots & a_{nm}B \end{pmatrix}$$

De este modo,

$$\varepsilon = Z_\eta \eta + \nu, \quad \text{donde } Z_\eta = (I_N \otimes j_T) \quad (1.4)$$

Empleando esta herramienta, podemos definir las siguientes matrices que utilizaremos a lo largo de este capítulo y los que siguen.

Definición 1.2 (Matriz de transformación interindividual) Definimos la matriz de transformación interindividual como

$$B = I_N \otimes J_T / T, \quad \text{donde } J_T = j_t j_t' \text{ es una matriz de unos cuadrada de tamaño } T$$

Definición 1.3 (Matriz de transformación intraindividual) Definimos la matriz de transformación intraindividual como

$$W = I_{NT} - B$$

Observación 1.4 Es fácil comprobar las siguientes propiedades de las matrices anteriores (ver [19] y [2] para más detalle):

- Se cumple $(Bx)^t = (\overbrace{\bar{x}_1, \bar{x}_1, \dots, \bar{x}_1}^T, \bar{x}_2, \bar{x}_2, \dots, \bar{x}_2, \dots, \bar{x}_N, \bar{x}_N, \dots, \bar{x}_N)$.

- $B = Z_\eta Z_\eta^t / T$.
- Son matrices simétricas, ortogonales entre sí e idempotentes. De esto último se deduce que $\text{rang}(B) = \text{tr}(B) = N$ y $\text{rang}(W) = \text{tr}(W) = N(T - 1)$.
- $B + W = I$, por tanto estas matrices dan lugar a una descomposición de un vector, es decir, $Bz + Wz = z$.

1.2. Hipótesis sobre los errores

Para el desarrollo de los modelos que se describirán a continuación vamos a considerar las siguientes hipótesis:

1. Los valores esperados de las dos componentes del error se suponen 0.
2. Se considera homocedasticidad e incorrelación mutua tanto de la parte idiosincrática como del efecto individual.
3. Las componentes del error están incorreladas entre sí.
4. $\text{Var}(\varepsilon_{nt}) = E(\varepsilon_{nt}^2) = \sigma_\eta^2 + \sigma_\nu^2$
5. La covarianza de los errores para el mismo individuo en distintos instantes de tiempo viene dada por $\text{Cov}(\varepsilon_{nt}, \varepsilon_{ns}) = E(\varepsilon_{nt}\varepsilon_{ns}) = \sigma_\eta^2$
6. La covarianza de dos errores de individuos diferentes es $\text{Cov}(\varepsilon_{nt}, \varepsilon_{mt}) = E(\varepsilon_{nt}\varepsilon_{mt}) = \text{Cov}(\varepsilon_{nt}, \varepsilon_{ms}) = E(\varepsilon_{nt}\varepsilon_{ms}) = 0$
7. Por último vamos a suponer incorrelación de las dos componentes de los errores con las variables predictoras. Esta hipótesis es fundamental, aunque poco frecuente. Se discutirá cómo hacer frente a la situación en la que no se cumple a lo largo del proyecto.

En virtud de lo anteriormente mencionado, la matriz de varianzas-covarianzas de los errores para un individuo viene dada por:

$$\Omega_{nn} = \sigma_\nu^2 I_T + \sigma_\eta^2 J_T \quad (1.5)$$

Debido a la hipótesis (6) la matriz de covarianzas Ω de la muestra total es una matriz cuadrada de dimensión NT diagonal por bloques, viniendo dado cada uno de ellos por (1.5).

$$\Omega = I_N \otimes (\sigma_\nu^2 I_T + \sigma_\eta^2 J_T) = \sigma_\nu^2 I_{NT} + \sigma_\eta^2 (I_N \otimes J_T)$$

o equivalentemente,

$$\Omega = \sigma_\nu^2 (B + W) + T\sigma_\eta^2 B = \sigma_\nu^2 W + \sigma_l^2 B \quad (1.6)$$

donde denotamos $\sigma_l^2 = \sigma_\nu^2 + T\sigma_\eta^2$.

1.3. Modelo agrupado

La primera estimación que podemos llevar a cabo es aplicar directamente el *método de mínimos cuadrados ordinarios* (**MCO**) a (1.3), es decir, a los datos sin ninguna modificación previa y sin tener en cuenta la descomposición del error descrita anteriormente. Obtenemos así:

$$\hat{\gamma}_{\text{MCO}} = (Z^t Z)^{-1} Z^t y \quad (1.7)$$

Es un estimador insesgado pues $E(\hat{\gamma}_{\text{MCO}}) = (Z^t Z)^{-1} Z^t E(y) = (Z^t Z)^{-1} Z^t Z \gamma = \gamma$. Sustituyendo la ecuación (1.3) en la expresión del estimador obtenemos $\hat{\gamma}_{\text{MCO}} - \gamma = (Z^t Z)^{-1} Z^t \varepsilon$. Haciendo uso de esta expresión podemos ver que el estimador es consistente si se mantiene la hipótesis de incorrelación de las variables regresoras con la componente del error, y también nos permite calcular la varianza:

$$\text{Var}(\hat{\gamma}_{\text{MCO}}) = (Z^t Z)^{-1} Z^t \Omega Z (Z^t Z)^{-1} \quad (1.8)$$

La expresión anterior es algo complicada y no puede reducirse debido a que la matriz Ω no es diagonal, lo cual es bastante problemático computacionalmente. Tiene además la limitación de que pese a que es insesgado y consistente, hay estimadores con esas mismas características más eficientes.

1.4. Modelo de efectos fijos

En este caso consideramos η_n como parámetros fijos para ser estimados $\forall n = 1, \dots, N$. El modelo de efectos fijos es adecuado cuando se trata con macropaneles, es decir, paneles donde el conjunto de N unidades es fijado previamente e interesa la estimación de los efectos individuales.

1.4.1. Estimador LSDV

Sustituyendo la expresión del error de (1.4) en (1.3) se obtiene:

$$y = Z\gamma + Z_\eta \eta + \nu \quad (1.9)$$

La opción natural es aplicar (**MCO**) al modelo anterior para estimar α , β y η . Identificando el intercepto y los efectos individuales (por ejemplo tomando $\alpha = 0$ o estimando $\hat{\alpha}_n = \alpha + \eta_n$) el modelo cuenta con $N + K$ parámetros a estimar. Las dimensiones de Z y Z_η son $NT \times (K + 1)$ y $NT \times N$ respectivamente, luego la matriz a invertir mediante (**MCO**) es de tamaño $N + K + 1$. Por tanto, en micropaneles de gran tamaño, es decir, con una gran cantidad de individuos, la estimación no es eficiente computacionalmente, además de reducir considerablemente los grados de libertad. Al método descrito anteriormente se le conoce como *mínimos cuadrados de variables ficticias* (en inglés, *least squares dummy variables*) o (**LSDV**).

1.4.2. Estimador intragrupos

Consideremos ahora el modelo multiplicado previamente por la matriz W . Por la definición de esta misma se cumple $W\alpha_j = 0$ y $WZ_\eta\eta = 0$. Obtenemos por tanto:

$$Wy = W(\alpha_j + X\beta + Z_\eta\eta + \nu) = WX\beta + W\nu \quad (1.10)$$

Es decir, esta transformación elimina el intercepto y los efectos individuales. Además también suprime aquellas variables que son constantes en el tiempo como pueden ser la religión, el sexo o la nacionalidad. Aplicando (**MCO**) a (1.10) llegamos al llamado **estimador intragrupos**:

$$\hat{\beta}_w = (X^tWX)^{-1}X^tWy \quad (1.11)$$

Sustituyendo (1.10) en la expresión anterior se llega a:

$$\hat{\beta}_w - \beta = (X^tWX)^{-1}X^tW\nu \quad (1.12)$$

Observando además que $W\varepsilon = W(Z_\eta\eta + \nu) = W\nu$ se obtiene, $\text{Var}(\hat{\beta}_w) = E((\hat{\beta}_w - \beta)(\hat{\beta}_w - \beta)^t) = (X^tWX)^{-1}X^tWE(\varepsilon\varepsilon^t)WX(X^tWX)^{-1} = (X^tWX)^{-1}X^tW\Omega WX(X^tWX)^{-1}$. Finalmente, $W\Omega = W(\sigma_\nu^2W + \sigma_l^2B) = \sigma_\nu^2W$ y por tanto:

$$\text{Var}(\hat{\beta}_w) = \sigma_\nu^2(X^tWX)^{-1} \quad (1.13)$$

El estimador (1.11) es equivalente al que se obtiene mediante (**LSDV**). Puede demostrarse esto último mediante el teorema de Frisch-Waugh (ver [16] y [19]).

Una vez hallada la estimación de β podemos calcular los estimadores de los efectos individuales. Para ello consideremos las siguientes transformaciones del modelo:

$$y_{nt} = \alpha + x_{nt}^t\beta + \eta_n + \nu_{nt} \quad (1.14)$$

Tomando medias muestrales en tiempo:

$$\bar{y}_n = \alpha + \bar{x}_n^t\beta + \eta_n + \bar{\nu}_n \quad (1.15)$$

Restando (1.15) a (1.14) se obtiene el modelo de variables transformadas ya descrito en (1.10):

$$y_{nt} - \bar{y}_n = (x_{nt} - \bar{x}_n)^t\beta + (\nu_{nt} - \bar{\nu}_n) \quad (1.16)$$

En último lugar, tomando la media total de la muestra llegamos a:

$$\bar{\bar{y}} = \alpha + \bar{\bar{x}}^t\beta + \bar{\bar{\nu}} \quad (1.17)$$

Podemos tomar por tanto $\hat{\alpha}_n = \bar{y}_n - \bar{x}_n^t\hat{\beta}_w$ estimador de $\alpha + \eta_n$. Para que la media muestral de los estimadores de los efectos individuales sea 0 (en concordancia con la hipótesis de que el

valor esperado debe ser nulo), se toma como estimador del intercepto $\hat{\alpha} = \bar{y} - \bar{x}^t \hat{\beta}_w$ y entonces los estimadores de los efectos individuales vienen dados por:

$$\hat{\eta}_{n,w} = \hat{\alpha}_n - \hat{\alpha} = (\bar{y}_n - \bar{y}) - (\bar{x}_n - \bar{x})^t \hat{\beta}_w \quad \forall n = 1, \dots, n \quad (1.18)$$

Lo más usual es dar la estimación de $\alpha + \eta_n$ ($\hat{\alpha}_n$) ya que de esta manera se reduce en uno los parámetros a estimar.

1.5. Modelos de efectos aleatorios

Ahora consideramos los efectos individuales como observaciones aleatorias de una distribución específica. En este caso no interesa la estimación de estos efectos individuales, aunque sí va a ser necesaria la estimación de los parámetros de la distribución que siguen con el objetivo de obtener buenos estimadores de las pendientes. El ejemplo acerca del rendimiento académico de unos estudiantes de diferentes escuelas a lo largo del curso es una gran ilustración de la situación anterior y se aplicaría un modelo de efectos aleatorios. En este caso se considera que los alumnos seleccionados forman parte de una población más amplia y que las variaciones individuales de cada uno de ellos son aleatorias.

1.5.1. Estimador mediante MCG

Recordemos que la matriz de covarianzas de los errores venía dada por:

$$\Omega = \sigma_l^2 B + \sigma_v^2 W$$

Es decir, una combinación lineal de las matrices B y W . Debido a que B y W son matrices idempotentes y ortogonales entre sí, la expresión de las potencias de Ω es muy simple:

$$\Omega^k = (\sigma_l^2)^k B + (\sigma_v^2)^k W \quad \forall k \in \mathbb{R} \quad (1.19)$$

Aplicando la fórmula anterior, la matriz $\Omega^{-1} = \frac{1}{\sigma_l^2} B + \frac{1}{\sigma_v^2} W$ es sencilla de obtener. Esto nos permite obtener el estimador por **mínimos cuadrados generalizados (MCG)** de una manera eficiente. Este método consiste en la estimación de los parámetros de (1.3) minimizando la distancia de Mahalanobis de los errores:

$$\hat{\gamma}_{\text{GLS}} = \arg \min_b (y - Zb)^t \Omega^{-1} (y - Zb)$$

Debido a que la función objetivo tiene una forma cuadrática, obtenemos:

$$\hat{\gamma}_{\text{GLS}} = (Z^t \Omega^{-1} Z)^{-1} (Z^t \Omega^{-1} y) \quad (1.20)$$

Para calcular la varianza del estimador, sustituyendo y por $Z\gamma + \varepsilon$:

$$\hat{\gamma}_{\text{GLS}} - \gamma = (Z^t \Omega^{-1} Z)^{-1} (Z^t \Omega^{-1} \varepsilon) \quad (1.21)$$

Por tanto:

$$\begin{aligned}\text{Var}(\hat{\gamma}_{\text{GLS}}) &= (Z^t \Omega^{-1} Z)^{-1} Z^t \Omega^{-1} E(\varepsilon \varepsilon^t) \Omega^{-1} Z (Z^t \Omega^{-1} Z)^{-1} = (Z^t \Omega^{-1} Z)^{-1} \\ &= \left(\frac{1}{\sigma_\nu^2} Z^t W Z + \frac{1}{\sigma_l^2} Z^t B Z \right)^{-1}\end{aligned}\quad (1.22)$$

Para las pendientes, sustituyendo la expresión de Ω^{-1} y usando resultados de inversión de matrices por bloques (ver [18] y [19]) se llega a:

$$\hat{\beta}_{\text{GLS}} = \left(\frac{1}{\sigma_\nu^2} X^t W X + \frac{1}{\sigma_l^2} X^t \bar{B} X \right)^{-1} \left(\frac{1}{\sigma_\nu^2} X^t W y + \frac{1}{\sigma_l^2} X^t \bar{B} y \right) \quad (1.23)$$

$$\text{Var}(\hat{\beta}_{\text{GLS}}) = \left(\frac{1}{\sigma_\nu^2} X^t W X + \frac{1}{\sigma_l^2} X^t \bar{B} X \right)^{-1} \quad (1.24)$$

donde $\bar{B} = B - J/NT$.

Debido a que Ω es simétrica y semidefinida positiva, Ω^{-1} también lo es. Por lo tanto, $\exists C / \Omega^{-1} = CC$. Se puede llegar de manera alternativa al mismo estimador mediante una transformación del modelo multiplicando (1.3) por $C = \Omega^{-\frac{1}{2}} = \frac{1}{\sigma_l} B + \frac{1}{\sigma_\nu} W$ y aplicando (**MCO**) a las variables transformadas. Sea $y^* = Cy$ y $Z^* = CZ$ obtenemos:

$$\hat{\gamma} = [(Z^*)^t Z^*]^{-1} (Z^*)^t y^* = (Z^t C^t C Z)^{-1} Z^t C^t C y = (Z^t \Omega^{-1} Z)^{-1} Z^t \Omega^{-1} y$$

que coincide con $\hat{\gamma}_{\text{GLS}}$. La transformación anterior consiste en una combinación lineal de las matrices de transformación interindividual e intraindividual. Multiplicando por $\sigma_\nu \Omega^{-\frac{1}{2}}$ (que es equivalente) los escalares de la combinación son ahora $\frac{\sigma_\nu}{\sigma_l}$ y 1. Luego las variables transformadas $y^* = \sigma_\nu \Omega^{-\frac{1}{2}} y$ son:

$$y_{nt}^* = \frac{\sigma_\nu}{\sigma_l} \bar{y}_n + (y_{nt} - \bar{y}_n) = y_{nt} - \left(1 - \frac{\sigma_\nu}{\sigma_l}\right) \bar{y}_n = y_{nt} - \theta \bar{y}_n \quad (1.25)$$

donde denotamos $\theta = 1 - \frac{\sigma_\nu}{\sigma_l} = 1 - \phi$, con $\phi = \frac{\sigma_\nu}{\sigma_l}$. Observemos ahora:

$$\theta = 1 - \sqrt{\frac{\sigma_\nu^2}{T\sigma_\eta^2 + \sigma_\nu^2}} = 1 - \frac{1}{\sqrt{1 + T\frac{\sigma_\eta^2}{\sigma_\nu^2}}} \quad (1.26)$$

Como $\text{Var}(\varepsilon_{nt}) = \sigma_\eta^2 + \sigma_\nu^2$, el valor de $\theta \in [0, 1]$ va a venir determinado por la parte de la varianza total del error que es aportada por los efectos individuales. Por un lado, si σ_η^2 domina a σ_ν^2 , θ va a ser próximo a 1 luego la transformación (1.25) va a ser muy similar a (1.16) y por tanto $\hat{\beta}_{\text{GLS}}$ va a ser cercano al estimador intragrupos $\hat{\beta}_w$. Por otro lado si la parte explicada por los efectos individuales es escasa, θ va a ser próximo a 0 luego prácticamente no se llevaría a cabo ninguna modificación del modelo (1.1) y $\hat{\beta}_{\text{GLS}}$ será muy similar al estimador $\hat{\beta}_{\text{MCO}}$ obtenido al aplicar (**MCO**) a los datos sin ningún procesamiento previo.

El estimador $\hat{\beta}_{\text{GLS}}$ depende de parámetros desconocidos. Veamos cómo podemos estimar σ_η^2 y σ_l^2 para hacer factible el estimador.

1.5.2. Estimación de las varianzas de las componentes del error

Vamos a presentar estimadores para σ_l^2 y σ_ν^2 , y teniendo en cuenta que $\sigma_\eta^2 = \frac{\sigma_l^2 - \sigma_\nu^2}{T}$ obtendremos los estimadores para la varianza de los efectos individuales. Fijado un tiempo $t \in \{1, \dots, T\}$, obtengamos la expresión de la varianza para $\bar{\varepsilon}_n$ y $\varepsilon_{nt} - \bar{\varepsilon}_n$. Recordemos que por hipótesis $\text{Var}(\varepsilon_{nt}) = \sigma_\eta^2 + \sigma_\nu^2$. Por un lado:

$$\begin{aligned}\bar{\varepsilon}_n &= \frac{1}{T} \sum_{s=1}^T \varepsilon_{ns} = \frac{1}{T} \sum_{s=1}^T (\eta_n + \nu_{st}) = \eta_n + \frac{1}{T} \sum_{s=1}^T \nu_{ns} \\ \text{Var}(\bar{\varepsilon}_n) &= \sigma_\eta^2 + \frac{\sigma_\nu^2}{T} = \frac{\sigma_l^2}{T}\end{aligned}\quad (1.27)$$

Por otro lado:

$$\begin{aligned}\varepsilon_{nt} - \bar{\varepsilon}_n &= \nu_{nt} - \frac{1}{T} \sum_{s=1}^T \nu_{ns} = \left(1 - \frac{1}{T}\right) \nu_{nt} - \sum_{\substack{s=1 \\ s \neq t}}^T \nu_{ns} \\ \text{Var}(\varepsilon_{nt} - \bar{\varepsilon}_n) &= \left(1 - \frac{1}{T}\right)^2 \sigma_\nu^2 + \frac{T-1}{T^2} \sigma_\nu^2 = \frac{T-1}{T} \sigma_\nu^2\end{aligned}\quad (1.28)$$

Llegamos por tanto a que $\sigma_l^2 = T \text{Var}(\bar{\varepsilon}_n)$ y $\sigma_\nu^2 = \frac{T}{T-1} \text{Var}(\varepsilon_{nt} - \bar{\varepsilon}_n)$. Si ε fuera conocido los estimadores naturales para las varianzas serían:

$$\hat{\sigma}_l^2 = T \frac{\sum_{n=1}^N \bar{\varepsilon}_n^2}{N} = T \frac{\sum_{n=1}^N \sum_{t=1}^T \bar{\varepsilon}_n^2}{NT} = \frac{\varepsilon^t B \varepsilon}{N} \quad (1.29)$$

$$\hat{\sigma}_\nu^2 = \frac{T}{T-1} \frac{\sum_{n=1}^N \sum_{t=1}^T (\varepsilon_{nt} - \bar{\varepsilon}_n)^2}{NT} = \frac{\varepsilon^t W \varepsilon}{N(T-1)} \quad (1.30)$$

Para hacer eficientes estos estimadores, debemos sustituir los errores por estimaciones adecuadas:

1. **Wallace y Hussain** (1969) propusieron el uso de $\hat{\varepsilon}_{\text{MCO}}$ como estimador de ε .
2. **Amemiya** (1971) se basa en el estimador intragrupos. En primer lugar considera la estimación del intercepto $\hat{\alpha} = \bar{y} - \bar{x}^t \hat{\beta}_w$, para posteriormente calcular $\hat{\varepsilon}_w = y - \alpha j_{\text{NT}} - \hat{\beta}_w$. Debemos tener en cuenta que como $\hat{\sigma}_\eta^2 = \frac{\hat{\sigma}_l^2 - \hat{\sigma}_\nu^2}{T}$ y con la transformación intragrupos se eliminan las variables constantes en tiempo, si tenemos presente en el modelo algunas variables de este tipo la varianza de los efectos individuales estará sobreestimada.
3. **Hausman y Taylor** (1981) propusieron un ajuste para el problema anterior, mediante la regresión de $\hat{\varepsilon}_w$ en función de todas las variables constantes en tiempo del modelo. Los residuos $\hat{\varepsilon}_{\text{HT}}$ son los que se computan para la estimación de las varianzas. De esta manera no varía la estimación de σ_ν^2 y decrece la de σ_l^2 . Así también decrecerá la estimación de σ_η^2 .

4. **Swamy y Arora** (1972) emplea para $\hat{\sigma}_t^2$ la estimación de los residuos $\hat{\varepsilon}_B$ del modelo multiplicado previamente por la matriz de transformación interindividual B . Para la segunda forma cuadrática utiliza al igual que Amemiya $\hat{\varepsilon}_w$. Además en ambos se divide por los grados de libertad del modelo, quedando los estimadores como sigue:

$$\hat{\sigma}_t^2 = \frac{\hat{\varepsilon}_B^t W \hat{\varepsilon}_B}{N - K - 1} \quad (1.31)$$

$$\hat{\sigma}_\nu^2 = \frac{\hat{\varepsilon}_w^t W \hat{\varepsilon}_w}{N(T - 1) - K} \quad (1.32)$$

5. **Nerlove** (1971) estima, al contrario que los demás, en primer lugar σ_η^2 mediante la cuasi-varianza muestral de $\hat{\eta}_n$ dado por (1.18) que recordemos cumple $\bar{\hat{\eta}}_n = 0$. Por otro lado $\hat{\sigma}_\nu^2$ es obtenida dividiendo la forma cuadrática por el número de observaciones. De este modo, los estimadores son:

$$\hat{\sigma}_\eta^2 = \sum_{n=1}^N \frac{(\hat{\eta}_n - \bar{\hat{\eta}}_n)^2}{N - 1} \quad (1.33)$$

$$\hat{\sigma}_\nu^2 = \frac{\hat{\varepsilon}_w^t W \hat{\varepsilon}_w}{NT} \quad (1.34)$$

En este caso, también puede estar sobreestimada la varianza de los efectos individuales si el número de unidades no es demasiado grande.

1.6. Comparación de los estimadores

En las secciones anteriores hemos llegado a las estimaciones de las pendientes del modelo mediante (**MCO**), el estimador intragrupos y el estimador mediante (**MCG**), cuyas expresiones son:

$$\hat{\beta}_{\text{MCO}} = (X^t W X + X^t \bar{B} X)^{-1} (X^T W y + X^t \bar{B} y)$$

$$\hat{\beta}_w = (X^t W X)^{-1} X^T W y$$

$$\hat{\beta}_{\text{GLS}} = (X^t W X + \phi^2 X^t \bar{B} X)^{-1} (X^T W y + \phi^2 X^t \bar{B} y)$$

Observemos que la expresión anterior para el estimador por (**MCG**) es equivalente a (1.23) y que, como dijimos previamente, $\hat{\beta}_{\text{MCO}}$ y $\hat{\beta}_w$ pueden obtenerse sustituyendo en $\hat{\beta}_{\text{GLS}}$ por $\phi = 1$ y $\phi = 0$ respectivamente. Notemos que con las hipótesis establecidas sobre el modelo, los tres estimadores son insesgados y consistentes, por tanto los resultados de ellos van a ser muy similares si la muestra es suficientemente extensa.

En cuanto a las varianzas, teniendo en cuenta (1.24), $\text{Var}(\hat{\beta}_{\text{GLS}}) = \sigma_\nu^2 (X^t W X + \phi^2 X^t \bar{B} X)^{-1}$ y (1.13) llegamos a que $\text{Var}(\hat{\beta}_w) - \text{Var}(\hat{\beta}_{\text{GLS}})$ es una matriz definida positiva (salvando claramente el caso $\phi = 0$), por tanto $\hat{\beta}_{\text{GLS}}$ es más eficiente que $\hat{\beta}_w$.

1.6.1. Elección entre el modelo de efectos fijos y modelo de efectos aleatorios

Previamente dimos un posible criterio para la elección de un modelo u otro atendiendo a la naturaleza de los datos. Por un lado escogeríamos el modelo de efectos fijos en el caso de que la muestra sea un macropanel, donde se tiene fijado previamente el conjunto de individuos del estudio e interesa la estimación de los efectos individuales. Por otro lado optaremos por un modelo de efectos aleatorios si la muestra es un micropanel, en cuyo caso los individuos forman parte de una población mayor cuya distribución se considera aleatoria y no corresponde la estimación de los efectos individuales.

Sin embargo, el mayor argumento para escoger entre alguna de las dos alternativas consiste en atender a la posibilidad de correlación de algunas variables predictoras y los efectos individuales. Si mantenemos la hipótesis de incorrelación de la parte idiosincrática, tenemos dos posibles situaciones:

1. $E(X^t\eta) \neq 0$, es decir, hay correlación de los efectos individuales con alguna variable. Entonces únicamente el modelo de efectos fijos es consistente, debido a que los efectos individuales desaparecen por la transformación dada por la matriz W . Aun así los resultados no son del todo convincentes para este estimador pues elimina la información acerca de la variación interindividual de la muestra, en especial las variables que son constantes en el tiempo para los individuos.
2. $E(X^t\eta) = 0$. En este caso ambos estimadores son consistentes, pero $\hat{\beta}_w$ es menos eficiente que $\hat{\beta}_{GLS}$ luego sería más conveniente escoger este último.

En el capítulo 3 se planteará el test de Hausman para abordar este problema.

Capítulo 2

Modelo de componente de error bidireccional

2.1. Introducción

Wallace y Hussain (1969), Nerlove (1971) y Amemiya (1971) entre otros consideraron el modelo de regresión (1.1) con la descomposición siguiente del error:

$$\varepsilon_{nt} = \eta_n + \mu_t + \nu_{nt}$$

El elemento η_n denota para cada $n \in \{1, \dots, N\}$ al igual que en el modelo presentado en el capítulo anterior el efecto individual no observado.

La componente μ_t representa para cada $t \in \{1, \dots, T\}$ el efecto temporal no observado. Este parámetro podría representar cambios en las condiciones económicas que afecten a todas las unidades del modelo por igual (por ejemplo una subida de impuestos). También, supongamos que las unidades de nuestro modelo es un conjunto de minoristas de un país en el cual ha habido una recesión económica en un periodo en concreto. Este efecto no observado sería recogido por esta componente. Por último, si las unidades de nuestro modelo son un conjunto de empresas de un mismo sector, μ_t puede representar un cambio de tendencias de los consumidores a lo largo del tiempo que afectan a todas las unidades del modelo.

Al igual que en el modelo de componente de error unidireccional, ν_{nt} denota la parte idiosincrática o aleatoria del error.

El modelo se escribe para cada observación por tanto de la siguiente forma:

$$y_{nt} = \alpha + x_{nt}^t \beta + \eta_n + \mu_t + \nu_{nt} = z_{nt}^t \gamma + \eta_n + \mu_t + \nu_{nt} \quad (2.1)$$

Sea $\mu = (\mu_i)_{i=1}^T$ podemos escribir ε de forma matricial:

$$\varepsilon = Z_\eta \eta + Z_\mu \mu + \nu \quad \text{donde } Z_\mu = j_N \otimes I_T \quad (2.2)$$

Así, matricialmente el modelo se puede expresar como:

$$y = Z\gamma + Z_\eta\eta + Z_\mu\mu + \nu \quad (2.3)$$

2.2. Hipótesis sobre los errores

Vamos a considerar para los efectos temporales las mismas hipótesis que hicimos para los efectos individuales. Estas son:

1. Se considera homocedasticidad de μ con $\text{Var}(\mu) = \sigma_\mu^2$
2. $E(\mu) = 0$
3. Suponemos incorrelación mutua, es decir, $E(\mu_t\mu_s) = 0 \ \forall t \neq s$
4. Los efectos temporales están incorrelados con el resto de las componentes del error.
5. Por último las variables predictoras están incorreladas con μ .

Teniendo en cuenta además las hipótesis sobre los efectos individuales y la parte idiosincrática del modelo de componente de error unidireccional, que también son asumidas para este modelo, la matriz de covarianzas de los errores se puede escribir como:

$$\Omega = \sigma_\nu^2 I_{NT} + \sigma_\eta^2 I_N \otimes J_T + \sigma_\mu^2 J_N \otimes I_T \quad (2.4)$$

2.3. Modelo de efectos fijos

Si η_n y μ_t son considerados parámetros fijos a ser estimados, una primera idea puede ser la estimación por (**LSDV**) en el modelo (2.3), contando con $K + N + T$ parámetros a estimar. Sin embargo, como ya pasaba en el caso del modelo de componente de error unidireccional, si N o T es relativamente grande, la matriz a invertir es de una dimensión para la cual puede no ser computacionalmente factible su inversión, reduce considerablemente los grados de libertad y acentúa el problema de multicolinealidad entre las variables predictoras.

Por ello vamos a considerar análogamente una transformación del modelo que haga desaparecer los efectos individuales y temporales. Observemos que promediando el modelo (2.1) sobre los individuos obtenemos:

$$\bar{y}_{.t} = \alpha + \bar{x}_{.t}\beta + \mu_t + \bar{\nu}_{.t} \quad \text{donde denotamos } \bar{y}_{.t} = \frac{1}{N} \sum_{i=1}^N y_{it} \quad (2.5)$$

Hemos utilizado que $\sum_{i=1}^N \eta_i = 0$, que es una restricción para evitar el problema de la trampa de variables ficticias o multicolinealidad perfecta.

Similarmente, considerando $\sum_{i=1}^T \mu_i = 0$ y promediando (2.1) en tiempo por un lado y en todas las observaciones por otro llegamos a:

$$\bar{y}_{n.} = \alpha + \bar{x}_{n.}^t \beta + \eta_n + \bar{\nu}_{n.} \quad \text{donde denotamos } \bar{y}_{n.} = \frac{1}{T} \sum_{s=1}^T y_{ns} \quad (2.6)$$

$$\bar{y}_{..} = \alpha + \bar{x}_{..}^t \beta + \bar{\nu}_{..} \quad \text{donde denotamos } \bar{y}_{..} = \frac{1}{NT} \sum_{i=1}^N \sum_{s=1}^T y_{is} \quad (2.7)$$

Teniendo en cuenta las expresiones anteriores, la transformación del modelo que buscamos es:

$$(y_{nt} - \bar{y}_{n.} - \bar{y}_{.t} + \bar{y}_{..}) = (x_{nt} - \bar{x}_{n.} - \bar{x}_{.t} + \bar{x}_{..})^t \beta + (\nu_{nt} - \bar{\nu}_{n.} - \bar{\nu}_{.t} + \bar{\nu}_{..}) \quad (2.8)$$

La matriz de transformación Q de manera que $\tilde{y} = Qy = y_{nt} - \bar{y}_{n.} - \bar{y}_{.t} + \bar{y}_{..}$ es:

$$Q = I - B_\eta - B_\mu + \bar{J} \quad (2.9)$$

donde:

- $B_\eta = I_N \otimes J_T/T$ es la matriz de transformación interindividual.
- $B_\mu = J_N \otimes I_T/N$ es la matriz de transformación intertemporal.
- $\bar{J} = J/NT$

Observación 2.1 Las matrices Q , B_η y B_μ son simétricas e idempotentes. Por el contrario no todas son ortogonales entre sí, pues $B_\eta B_\mu = \bar{J} \neq 0$. Esto es porque mediante la transformación del producto de estas dos matrices se obtienen las medias individuales de las medias temporales, que son las medias globales. Gracias a esto sí se tiene $QB_\eta = QB_\mu = 0$

Por tanto, para obtener el estimador del modelo de efectos fijos, basta con aplicar (**MCO**) a la regresión de $\tilde{y} = Qy$ sobre $\tilde{X} = QX$ para obtener el estimador:

$$\hat{\beta}_w = (X^t Q X)^{-1} X^t Q y \quad (2.10)$$

A partir de este estimador podemos obtener estimadores para el intercepto, los efectos individuales y los efectos temporales:

$$\hat{\alpha}_w = \bar{y}_{..} - \bar{x}_{..}^t \hat{\beta}_w \quad (2.11)$$

$$\hat{\eta}_w = (\bar{y}_{n.} - \bar{y}_{..}) - (\bar{x}_{n.} - \bar{x}_{..})^t \hat{\beta}_w \quad (2.12)$$

$$\hat{\mu}_w = (\bar{y}_{.t} - \bar{y}_{..}) - (\bar{x}_{.t} - \bar{x}_{..})^t \hat{\beta}_w \quad (2.13)$$

Observación 2.2 La matriz Q elimina tanto las variables constantes para cada individuo como las variables constantes en cada tiempo para los distintos individuos, luego $\hat{\beta}_w$ no va a tener en cuenta esas covariables en la estimación.

2.4. Modelo de efectos aleatorios

Ahora suponemos que los individuos han sido seleccionados aleatoriamente de una gran población, caso en el que es más adecuado un modelo de efectos aleatorios. Con el objetivo de aplicar (**MCG**), veamos la expresión de Ω^{-1} (en realidad basta con conocer $\Omega^{-1/2}$). Recordemos que la expresión de Ω viene dada por

$$\Omega = \sigma_\nu^2 I_{NT} + \sigma_\eta^2 I_N \otimes J_T + \sigma_\mu^2 J_N \otimes I_T$$

Como en el caso del modelo de componente de error unidireccional, buscamos la expresión de Ω como una combinación lineal de matrices idempotentes y ortogonales entre sí. Teniendo en cuenta que:

1. $I_{NT} = Q + B_\eta + B_\mu - \bar{J}$
2. $I_N \otimes J_T = TB_\eta$
3. $J_N \otimes I_T = NB_\eta$

Obtenemos la expresión $\Omega = \sigma_\nu^2 Q + (T\sigma_\eta^2 + \sigma_\nu^2)B_\eta + (N\sigma_\mu^2 + \sigma_\nu^2)B_\mu - \sigma_\nu^2 \bar{J}$

Atendiendo a lo mencionado en la observación (2.1) estas matrices no son todas ortogonales entre sí, así que considerando las matrices $\bar{B}_\eta = B_\eta - \bar{J}$ y $\bar{B}_\mu = B_\mu - \bar{J}$ obtenemos finalmente:

$$\Omega = \sigma_\nu^2 Q + (T\sigma_\eta^2 + \sigma_\nu^2)\bar{B}_\eta + (N\sigma_\mu^2 + \sigma_\nu^2)\bar{B}_\mu + (T\sigma_\eta^2 + N\sigma_\mu^2 + \sigma_\nu^2)\bar{J} = \sum_{i=1}^4 \lambda_i Q_i \quad (2.14)$$

donde $\lambda_1 = \sigma_\nu^2$, $\lambda_2 = T\sigma_\eta^2 + \sigma_\nu^2$, $\lambda_3 = N\sigma_\mu^2 + \sigma_\nu^2$ y $\lambda_4 = T\sigma_\eta^2 + N\sigma_\mu^2 + \sigma_\nu^2$. Correspondientemente, $Q_1 = Q$, $Q_2 = \bar{B}_\eta$, $Q_3 = \bar{B}_\mu$ y $Q_4 = \bar{J}$.

Ahora sí las matrices de la combinación lineal son ortogonales dos a dos. Además proporcionan una descomposición de un vector, es decir, $\sum_{i=1}^4 Q_i = I$. La ventaja de haber obtenido la expresión (2.14) es que llegamos fácilmente a la expresión de las potencias de Ω . Esta es:

$$\Omega^r = \sum_{i=1}^4 \lambda_i^r Q_i \quad (2.15)$$

El estimador mediante (**MCG**) puede obtenerse aplicando (**MCO**) al modelo transformado multiplicando previamente por $\Omega^{-1/2}$ o equivalentemente $\sigma_\nu \Omega^{-1/2}$. Haciendo uso de la expresión anterior, tenemos:

$$\sigma_\nu \Omega^{-1/2} = \sum_{i=1}^4 \frac{\sigma_\nu}{\sqrt{\lambda_i}} Q_i$$

La regresión que consideramos viene dada por:

$$y^* = Z^* \gamma + \varepsilon^*$$

donde $y^* = \sigma_\nu \Omega^{-1/2} y$, $Z^* = \sigma_\nu \Omega^{-1/2} Z$ y $\varepsilon^* = \sigma_\nu \Omega^{-1/2} \varepsilon$.

Por tanto el estimador del modelo de efectos aleatorios viene dado por:

$$\hat{\gamma}_{\text{GLS}} = [(Z^*)^t Z]^{-1} (Z^*)^t y^* \quad (2.16)$$

La expresión de y_{nt}^* es:

$$y_{nt}^* = y_{nt} - \theta_\eta \bar{y}_{n.} - \theta_\mu \bar{y}_{.t} + (\theta_\eta + \theta_\mu - \theta_2) \bar{y}_{..} \quad \forall n, t \quad (2.17)$$

con:

$$\begin{cases} \theta_\eta = 1 - \sqrt{\frac{\sigma_\nu^2}{\sigma_\nu^2 + T\sigma_\eta^2}} = 1 - \frac{\sigma_\nu}{\sqrt{\lambda_2}} = 1 - \phi_\eta \\ \theta_\mu = 1 - \sqrt{\frac{\sigma_\nu^2}{\sigma_\nu^2 + N\sigma_\mu^2}} = 1 - \frac{\sigma_\nu}{\sqrt{\lambda_3}} = 1 - \phi_\mu \\ \theta_2 = 1 - \sqrt{\frac{\sigma_\nu^2}{\sigma_\nu^2 + N\sigma_\mu^2 + T\sigma_\eta^2}} = 1 - \frac{\sigma_\nu}{\sqrt{\lambda_4}} = 1 - \phi_2 \end{cases}$$

Estos parámetros toman valores entre 0 y 1. Apreciamos que el parámetro θ_η coincide con el parámetro θ del modelo de componente de error unidireccional y su valor va a ser más cercano a 1 cuanto mayor sea el dominio de σ_η^2 frente a σ_ν^2 . Análogamente, el valor de θ_μ va a ser cercano a 1 si los efectos temporales aportan una mayor variabilidad en el modelo que la parte idiosincrática. En cuanto a θ_2 el razonamiento es similar, salvo que ahora se tiene en cuenta la varianza explicada tanto por los efectos individuales como los temporales.

Como el estimador depende de parámetros desconocidos, debemos sustituirlos por estimaciones insesgadas y consistentes de ellos.

2.4.1. Estimación de la varianza de las componentes del error

Teniendo en cuenta que $E(Q_i \varepsilon) = 0$ y $\text{Var}(Q_i \varepsilon) = \lambda_i Q_i$, los mejores estimadores cuadráticos para λ_i son

$$\hat{\lambda}_i = \frac{\varepsilon^t Q_i \varepsilon}{\text{tr}(Q_i)} \quad i = 1, 2, 3$$

De hecho, bajo normalidad de las perturbaciones los estimadores son insesgados de varianza mínima. Sin embargo si se sustituyen los parámetros $\lambda_i \quad i = 1, \dots, 4$ por estos estimadores para estimar las varianzas de las componentes del error se introduce sesgo.

Al igual que para el modelo de componente de error unidireccional, con el objetivo de estimar la varianza de las componentes del error podemos reemplazar ε por estimadores de esta.

Wallace y Hussain (1969) propusieron reemplazar ε por $\hat{\varepsilon}_{\text{MCO}}$. A pesar de que $\hat{\varepsilon}_{\text{MCO}}$ es insesgado y consistente, es ineficiente y resulta en estimadores sesgados de las desviaciones de los errores. Amemiya (1971) alternativamente propone considerar los residuos del modelo de efectos fijos, siendo $\varepsilon_w = y - \hat{\alpha}_w - X \hat{\beta}_w$, con $\hat{\alpha}_w$ dado por (2.11). De hecho, demostró que los estimaciones de las componentes de la varianza de Wallace y Hussain (1969) tienen distintas distribuciones

asintóticas que las verdaderas (suponiéndolas conocidas), mientras que al reemplazar por ε_w sí se tiene la misma distribución asintótica de las estimaciones de la varianza de las componentes del error que las verdaderas perturbaciones. Pueden llevarse a cabo correcciones en los estadísticos $\hat{\lambda}_i$ $i = 1, \dots, 4$ para obtener estimadores insesgados, extendiendo los métodos de estimación de las componentes del error propuestos por Wallace y Hussain (1969), Amemiya (1971), Swamy y Arora (1972) y Nerlove (1971) para el caso unidireccional. Estas correcciones dependen de las trazas de matrices que implica la matriz X (ver [2] para más detalle).

Capítulo 3

Tests para los modelos de componentes de error

La doble dimensión presente en los datos panel tiene la ventaja de poseer una gran fuente de información, pero a su vez en el análisis de un panel de datos debemos plantearnos una gran cantidad de cuestiones para mantener determinados supuestos y tomar la decisión de escoger un modelo u otro.

3.1. Contrastes de agrupabilidad de los datos

En primer lugar nos podemos plantear si agrupar o no los datos, es decir, si plantear un modelo como (1.1) (modelo restringido) donde γ es igual para todos los individuos y en todos los instantes de tiempo o plantear un modelo con parámetros diferentes para cada individuo o para cada instante de tiempo (modelo irrestringido).

Puesto que en el análisis de datos panel es más frecuente encontrar mayor heterogeneidad en el conjunto de individuos, vamos a desarrollar el caso en el que contrastaremos si agrupar o no las series temporales, aunque para el otro caso va a ser muy similar.

Para el modelo irrestringido, tenemos una regresión por cada individuo:

$$y_n = Z_n \gamma_n + \varepsilon_n \quad n = 1, 2, \dots, N \quad (3.1)$$

donde $y_n = (y_{n1}, \dots, y_{nT})^t$, $Z_n = (j_T, X_n)$ con Z_n de dimensión $T \times (K + 1)$. De una manera más compacta se puede escribir el modelo como:

$$y = \begin{pmatrix} Z_1 & 0 & \cdots & 0 \\ 0 & Z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_N \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_N \end{pmatrix} + \varepsilon = Z^* \gamma^* + \varepsilon \quad (3.2)$$

El modelo restringido es $y = Z\gamma + \varepsilon$ donde $Z = (Z_1^t, \dots, Z_N^t)^t$. Es fácil ver que se cumple $Z = Z^*I^*$ con $I^* = j_N \otimes I_{K+1}$.

Nuestro objetivo es realizar el contraste de hipótesis siguiente:

$$\begin{aligned} H_0 : \gamma_n &= \gamma \quad \forall n = 1, 2, \dots, N \\ H_1 : \gamma_n &\neq \gamma \text{ para algún } n \in \{1, \dots, N\} \end{aligned} \quad (3.3)$$

3.1.1. Contraste de agrupabilidad con $\varepsilon \sim N(0, \sigma^2 I_{NT})$

Bajo esta hipótesis el mejor estimador de γ en el modelo restringido es la estimación por *mínimos cuadrados ordinarios*,

$$\hat{\gamma}_{\text{MCO}} = (Z^t Z)^{-1} Z^t y$$

Entonces $y = Z\hat{\gamma}_{\text{MCO}} + \hat{\varepsilon}$ con $\hat{\varepsilon} = y - Z\hat{\gamma}_{\text{MCO}} = y - Z(Z^t Z)^{-1} Z^t y = [I_{NT} - Z(Z^t Z)^{-1} Z^t] y = My = M(Z\gamma + \varepsilon) = M\varepsilon$ pues $MZ = 0$

Análogamente en el modelo irrestringido para cada $n \in \{1, \dots, N\}$ consideramos:

$$\hat{\gamma}_{\text{MCO},n} = (Z_n^t Z_n)^{-1} Z_n^t y_n$$

$$y \hat{\varepsilon}_n = [I_T - Z_n(Z_n^t Z_n)^{-1} Z_n^t] y_n = M_n y_n = M_n \varepsilon_n$$

$$\text{Sea } M^* = I_{NT} - Z^* [(Z^*)^t Z^*]^{-1} (Z^*)^t = \begin{bmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_N \end{bmatrix}$$

Deducimos fácilmente que $y = Z^* \hat{\gamma}_{\text{MCO}}^* + \varepsilon^*$ donde $\varepsilon^* = M^* y = M^* \varepsilon$, con $\hat{\gamma}_{\text{MCO}}^* = [(Z^*)^t Z^*]^{-1} (Z^*)^t y$

Observación 3.1 Las matrices M y M^* son simétricas e idempotentes, luego $M - M^*$ también es idempotente. Además se cumple $MM^* = M^*$. Esto último se deduce de la siguiente forma, teniendo en cuenta que $Z = Z^* I^*$:

$$Z(Z^t Z)^{-1} Z^t Z^* [(Z^*)^t Z^*]^{-1} (Z^*)^t = Z(Z^t Z)^{-1} I^* (Z^*)^t Z^* [(Z^*)^t Z^*]^{-1} (Z^*)^t = Z(Z^t Z)^{-1} Z$$

Observación 3.2 Se tiene $(M - M^*)M^* = MM^* - M^*M^* = M^* - M^* = 0$. De esto deducimos que, debido a la hipótesis distribucional $\varepsilon^t \hat{\varepsilon} - (\hat{\varepsilon}^*)^t \hat{\varepsilon}^* = \varepsilon^t (M - M^*) \varepsilon$ y $(\hat{\varepsilon}^*)^t \hat{\varepsilon}^* = \varepsilon^* M^* \varepsilon^*$ son independientes. Además, una vez dividido por σ^2 las formas cuadráticas siguen una distribución χ^2 .

Estas observaciones nos llevan a considerar el estadístico:

$$F = \frac{[\varepsilon^t \hat{\varepsilon} - (\hat{\varepsilon}^*)^t \hat{\varepsilon}^*] / (N-1)(K+1)}{[(\hat{\varepsilon}^*)^t \hat{\varepsilon}^*] / N(T-K-1)} \stackrel{H_0}{\sim} F_{(N-1)(K+1), N(T-K-1)} \quad (3.4)$$

Este estadístico puede verse como una aplicación del test de Chow extendido al caso de N regresiones lineales.

3.1.2. Contraste de agrupabilidad con $\varepsilon \sim N(0, \sigma^2 \Sigma)$

Consideramos ahora el caso menos restrictivo de que la matriz de covarianzas no tenga por qué ser diagonal, es decir, sin asumir independencia de los errores aunque asumiendo que conocemos un factor escalar y la hipótesis distribucional de normalidad multivariante.

En el caso en que la matriz de covarianzas sea de la forma $\Omega = \sigma^2 \Sigma$ podemos transformar el modelo de manera que la varianza del error de este modelo transformado sea $\sigma^2 I_{NT}$ para así reducirnos al caso anterior.

Multiplicamos pues la ecuación del modelo restringido por $\Sigma^{-1/2}$, obteniendo así el modelo $\tilde{y} = \tilde{Z}\gamma + \tilde{\varepsilon}$ donde $\tilde{y} = \Sigma^{-1/2}y$, $\tilde{Z} = \Sigma^{-1/2}Z$ y $\tilde{\varepsilon} = \Sigma^{-1/2}\varepsilon$. Tal y como queríamos, tenemos $\text{Var}(\tilde{\varepsilon}) = E(\tilde{\varepsilon}\tilde{\varepsilon}^t) = \sigma^2 \Sigma^{-1/2} \Sigma (\Sigma^{-1/2})^t = \sigma^2 I_{NT}$

Así mismo, multiplicamos por $\Sigma^{-1/2}$ el modelo irrestringido (3.2) y de esta manera podemos realizar el test reduciéndonos al caso anterior.

3.2. Contraste sobre los efectos individuales y temporales

La presencia de los efectos individuales o temporales en un modelo de datos panel es fundamental, pues a diferencia de los modelos de regresión tradicionales los modelos de datos panel se basan en el control de estos, luego es trascendental plantearnos si realmente estos efectos están presentes o no.

3.2.1. Caso del modelo de componente de error unidireccional

En un modelo de efectos fijos en el cual nos interesa la estimación de los efectos individuales presentes en el modelo, podemos contrastar la significatividad conjunta de las variables ficticias introducidas en el modelo. Esto es:

$$H_0 : \eta_n = 0 \quad \forall n$$

$$H_1 : \eta_n \neq 0 \text{ para algún } n \in \{1, \dots, N\}$$

Para ello podemos hacer un F-test mediante una aplicación del test de Chow. Consideramos RSS_R la suma de cuadrados de los residuos obtenidos mediante la estimación por (MCO) del modelo agrupado y RSS_ν la suma de cuadrados de los residuos de la estimación mediante (LSDV). El estadístico del contraste es:

$$F_0 = \frac{(RSS_R - RSS_\nu)/(N - 1)}{(RSS_\nu)/(NT - K - N)}$$

Bajo hipótesis de normalidad de los residuos se cumple $F_0 \stackrel{H_0}{\sim} F_{N-1, NT-N-K}$

3.2.2. Caso del modelo de componente de error bidireccional

En este caso podemos contrastar la significatividad conjunta de las variables ficticias introducidas en el modelo:

$$H_0 : \eta_n = 0 \quad \forall n \quad \text{y} \quad \mu_t = 0 \quad \forall t$$

$$H_1 : \eta_n \neq 0 \text{ para algún } n \in \{1, \dots, N\} \quad \text{o} \quad \mu_t \neq 0 \text{ para algún } t \in \{1, \dots, T\}$$

Análogamente al caso anterior, el estadístico del contraste sería::

$$F_1 = \frac{(RSS_R - RSS_\nu)/(N + T - 2)}{(RSS_\nu)/[(N - 1)(T - 1) - K]}$$

donde RSS_R es la suma de cuadrados de los residuos del modelo agrupado y RSS_ν la suma de cuadrados de los residuos de (2.8). Bajo hipótesis de normalidad en los residuos se tiene $F_1 \stackrel{H_0}{\sim} F_{N+T-2, (N-1)(T-1)-K}$

Otra opción es contrastar la presencia conjunta de efectos individuales suponiendo que existen efectos temporales. Luego permitiendo $\mu_t \neq 0$ para algún $t \in \{1, \dots, T\}$ planteamos el contraste de hipótesis siguiente.

$$H_0 : \eta_n = 0 \quad \forall n$$

$$H_1 : \eta_n \neq 0 \text{ para algún } n \in \{1, \dots, N\}$$

Construimos el estadístico del contraste como una aplicación del test de Chow similarmente a los casos anteriores, salvo que en este caso el modelo restringido es la regresión:

$$(y_{it} - \bar{y}_{.t}) = (x_{it} - \bar{x}_{.t})^t \beta + (\nu_{it} - \bar{\nu}_{.t})$$

que es equivalente al que incluye las variables ficticias temporales (y no las individuales). Denotamos la suma de los cuadrados de los residuos $RSS_{\bar{\nu}_{.t}}$ resultando:

$$F_2 = \frac{(RSS_{\bar{\nu}_{.t}} - RSS_\nu)/(N - 1)}{(RSS_\nu)/[(N - 1)(T - 1) - K]}$$

En este caso bajo hipótesis de normalidad de las perturbaciones $F_2 \stackrel{H_0}{\sim} F_{N-1, (N-1)(T-1)-K}$

Por último, de forma análoga podemos contrastar la presencia conjunta de efectos temporales suponiendo que existen efectos individuales, es decir, que $\eta_n \neq 0$ para algún $n \in \{1, \dots, N\}$, planteamos el contraste de hipótesis:

$$H_0 : \mu_t = 0 \quad \forall t$$

$$H_1 : \mu_t \neq 0 \text{ para algún } t \in \{1, \dots, T\}$$

El modelo restringido en este caso viene dado por la regresión (1.16). Denotamos la suma de cuadrados de las perturbaciones como $RSS_{\bar{\nu}_{n.}}$. El estadístico que resulta es:

$$F_3 = \frac{(RSS_{\bar{\nu}_{n.}} - RSS_\nu)/(T - 1)}{(RSS_\nu)/[(N - 1)(T - 1) - K]}$$

que bajo normalidad en las perturbaciones $F_3 \stackrel{H_0}{\sim} F_{T-1, (N-1)(T-1)-K}$

3.3. Contraste sobre la correlación de los efectos individuales

Previamente hemos comentado que a la hora de elegir un método de estimación de los coeficientes es fundamental conocer si existe correlación o no de las variables con los efectos individuales. Vamos a presentar un test que trata esta cuestión.

3.3.1. Test de Hausman

Hausman compara los estimadores $\hat{\gamma}_{\text{GLS}}$ y $\hat{\gamma}_w$, haciendo uso del argumento que explicamos con detalle al final del capítulo 1. Tomamos la diferencia de los dos estimadores $\hat{q} = \hat{\gamma}_w - \hat{\gamma}_{\text{GLS}}$, y consideramos el contraste de hipótesis:

$$H_0 : E(X^t \eta) = 0$$

$$H_1 : E(X^t \eta) \neq 0$$

Es decir, la hipótesis nula es que no existe correlación de las variables explicativas con los efectos individuales. Bajo H_0 ambos estimadores son consistentes, luego se tiene $\text{plim } \hat{q} = 0$, lo cual no se va a cumplir si H_0 no es cierto. En cuanto a la matriz de covarianzas de \hat{q} :

$$\text{Var}(\hat{q}) = \text{Cov}(\hat{\gamma}_w - \hat{\gamma}_{\text{GLS}}, \hat{\gamma}_w - \hat{\gamma}_{\text{GLS}}) = \text{Var}(\hat{\gamma}_w) + \text{Var}(\hat{\gamma}_{\text{GLS}}) - 2\text{Cov}(\hat{\gamma}_w, \hat{\gamma}_{\text{GLS}})$$

Sabemos que $\text{Var}(\hat{\gamma}_w) = \sigma_v^2(Z^t W Z)^{-1}$ y $\text{Var}(\hat{\gamma}_{\text{GLS}}) = (Z^t \Omega^{-1} Z)^{-1}$.

Además $\text{Cov}(\hat{\gamma}_w, \hat{\gamma}_{\text{GLS}}) = \text{Cov}(\hat{q} + \hat{\gamma}_{\text{GLS}}, \hat{\gamma}_{\text{GLS}}) = \text{Cov}(\hat{q}, \hat{\gamma}_{\text{GLS}}) + \text{Var}(\hat{\gamma}_{\text{GLS}})$. Bajo H_0 tenemos $\text{Cov}(\hat{q}, \hat{\gamma}_{\text{GLS}}) = 0$, resultando por tanto la covarianza de \hat{q} de la siguiente forma:

$$\text{Var}(\hat{q}) = \text{Var}(\hat{\gamma}_w) - \text{Var}(\hat{\gamma}_{\text{GLS}}) = \sigma_v^2(Z^t W Z)^{-1} - (Z^t \Omega^{-1} Z)^{-1} \quad (3.5)$$

Teniendo en cuenta que $E(\hat{q}) = 0$ y que bajo H_0 tiende en probabilidad a 0, cuando la muestra es suficientemente grande debe tener un valor muy pequeño. Además, debido a que el estimador de efectos aleatorios es más eficiente que el de efectos fijos, $\text{Var}(\hat{q})$ será un valor considerablemente distinto de 0. Por otro lado, si H_0 no es cierto \hat{q} va a tomar un valor significativamente distinto de 0 pues $\hat{\gamma}_w$ es consistente pero $\hat{\gamma}_{\text{GLS}}$ no lo es. Esto nos lleva a considerar una aplicación del test de Wald tomando como estadístico:

$$m_1 = \hat{q}^T \text{Var}(\hat{q})^{-1} \hat{q} \quad (3.6)$$

Este estadístico sigue bajo H_0 e hipótesis de normalidad asintóticamente una distribución χ_K^2 donde K denota el número de variables predictoras del modelo. Rechazaremos el test si m_1 es significativamente distinto de 0. Para hacer factible el uso del estadístico reemplazaremos Ω por un estimador adecuado $\hat{\Omega}$ (ver [2]).

Nota 3.3 El estimador de Hausman es también válido para el modelo de componente de error bidireccional para contrastar la correlación de los efectos individuales y temporales con las variables.

Capítulo 4

Coeficientes heterogéneos

4.1. Introducción

Hay muchos casos en los que tanto la variable individual como la variable temporal tienen una dimensión bastante grande (paneles largos). Este tipo de paneles son muy comunes en macroeconomía y ciencias políticas. Esto abre un abanico de problemas y posibilidades aunque nosotros nos centraremos en discutir la heterogeneidad de los coeficientes.

En la Sección 3 presentamos un test de agrupabilidad de los datos. A veces para modelar las diferencias entre los distintos individuos no es suficiente con la componente η_n ya que es constante en el tiempo y además la influencia de los regresores en la variable objetivo puede no ser igual para los distintos individuos.

Si la dimensión temporal no es grande (micropaneles) el planteamiento de tantas regresiones separadas como individuos haya en el panel no es una buena opción, ya que en cada serie temporal vamos a tener T observaciones, y por tanto en general será mejor trabajar con un modelo agrupado aunque se rechace la igualdad de los coeficientes para todos los individuos.

En cambio, la presencia de paneles largos nos permiten estimar regresiones separadas en caso de rechazar el test de agrupabilidad (3.3). El modelo de datos panel con heterogeneidad individual en los coeficientes es:

$$y_{nt} = \alpha + \beta_n x_{nt} + \eta_n + \nu_{nt} \quad (4.1)$$

Se ha tomado el modelo de componente de error unidireccional, aunque los cálculos pueden extenderse fácilmente al caso bidireccional.

La decisión de agrupar o no ha sido objeto de muchos estudios. Si bien en el caso de rechazar la agrupabilidad de los datos consideramos un estimador donde las pendientes de las covariables coinciden para todas las unidades del estudio va a introducir sesgo en la estimación, disminuye la varianza del estimador. En general, si la muestra es suficientemente grande va a ser interesante

considerar coeficientes heterogéneos y en caso contrario es más recomendable seleccionar un modelo agrupado.

4.2. Coeficientes fijos

Si la dimensión en tiempo permite una buena estimación, una opción es simplemente estimar β por (MCO) en cada una de las N regresiones, obteniendo así:

$$\hat{\beta}_{\text{MCO},n} = (X_n^t X_n)^{-1} X_n y_n \quad (4.2)$$

Este enfoque reduce considerablemente los grados de libertad ya que los parámetros a estimar son un total de NK contando con NT observaciones. Además puede provocar sobreajuste.

4.3. Coeficientes aleatorios

Consideremos ahora que γ_n es para cada $n \in \{1, \dots, N\}$ un vector aleatorio con media γ a estimar. Este enfoque nos permite aprovechar la riqueza de la doble dimensión de los datos panel, teniendo en cuenta a diferencia del enfoque anterior la variabilidad transversal sin dejar de aprovechar las ventajas de la agrupación.

4.3.1. El estimador de Swamy

Swamy, desde un punto de vista paramétrico, propuso el modelo siguiente:

$$y_{nt} = \gamma_n^t z_{nt} + \nu_{nt} \quad (4.3)$$

donde suponemos γ_n i.i.d. según una $N(\gamma, \Delta)$ o equivalentemente $\delta_n = \gamma_n - \gamma \sim N(0, \Delta)$. En forma matricial:

$$y_n = Z_n \gamma_n + \nu_n \quad (4.4)$$

Para la parte idiosincrática ν se admite heterocedasticidad, aunque sí suponemos homocedasticidad en cada individuo, es decir, $\text{Var}(\nu_{nt}) = \sigma_n^2 \quad \forall t$. También $\text{Cov}(\nu_{nt}, \nu_{ms}) = 0 \quad \forall n \neq m$, $\text{Cov}(\nu_{nt}, \nu_{ns}) = 0 \quad \forall t \neq s$ y $\text{Cov}(\nu_{nt}, \delta_n) = 0 \quad \forall n, t$. De esta forma el modelo anterior puede ser reescrito como:

$$y_{nt} = \gamma^t z_{nt} + \varepsilon_{nt} \quad (4.5)$$

con $\varepsilon_{nt} = \nu_{nt} + \delta_n^t z_{nt}$. En forma matricial:

$$y_n = Z_n \gamma + \varepsilon_n \quad (4.6)$$

con $\varepsilon_n = \nu_n + Z_n \delta_n$.

Observemos que al no suponer homocedasticidad en ν_{nt} , los errores en el modelo (4.5) no son homocedásticos. Además, los errores correspondientes a los mismos individuos van a estar

correlados debido a que contienen el mismo vector de parámetros δ_n . Para un individuo en concreto, la matriz de covarianzas de los errores es:

$$\Omega_n = \text{Var}(\nu_n + Z_n \delta_n) = \text{Var}(\nu_n) + \text{Var}(Z_n \delta_n) = \sigma_n^2 I_T + Z_n \Delta Z_n^t \quad (4.7)$$

De esta forma, la matriz Ω para la muestra completa es una matriz diagonal por bloques, siendo cada uno de ellos igual a Ω_n .

La estimación del modelo mediante (**MCO**) no es la más indicada debido a que tenemos presente correlación y heterocedasticidad en los errores. En su lugar aplicaremos (**MCG**), que como hemos visto en varias ocasiones en las secciones anteriores es equivalente a multiplicar por $\Omega^{-1/2}$ y luego aplicar (**MCO**) al modelo transformado. Además, como la matriz es diagonal por bloques, obtenemos el mismo resultado multiplicando por $\Omega_n^{-1/2}$ en los datos de los individuos correspondientes. Claramente, este método es infactible pues Ω_n es desconocida, aunque podemos hacer funcional la expresión (4.7) sustituyendo los parámetros desconocidos por estimadores adecuados. Tenemos por tanto que estimar $\sigma_1^2, \dots, \sigma_N^2$ (N en total) y los elementos Δ que es una matriz $(K+1) \times (K+1)$ simétrica. Para ello, comenzamos estimando el modelo (4.4) mediante (**MCO**):

$$\hat{\gamma}_n = (Z_n^t Z_n)^{-1} Z_n^t y_n = \gamma_n + (Z_n^t Z_n)^{-1} Z_n^t \nu_n$$

El estimador natural de σ_n^2 es por tanto:

$$\hat{\sigma}_n^2 = \sum_{t=1}^T \frac{\hat{\varepsilon}_{nt}^2}{T - (K+1)} \quad (4.8)$$

Para la estimación de Δ nos basamos en la expresión de $\hat{\gamma}_n - \hat{\gamma}$. Desarrollando:

$$\begin{aligned} \hat{\gamma}_n - \hat{\gamma} &= \gamma_n + (Z_n^t Z_n)^{-1} Z_n^t \nu_n - \frac{1}{N} \sum_{n=1}^N \hat{\gamma}_n \\ &= \gamma_n + (Z_n^t Z_n)^{-1} Z_n^t \nu_n - \frac{1}{N} \sum_{m=1}^N [\gamma_m + (Z_m^t Z_m)^{-1} Z_m^t \nu_m] \\ &= \frac{N-1}{N} \gamma_n + \frac{N-1}{N} (Z_n^t Z_n)^{-1} Z_n^t \nu_n - \frac{1}{N} \sum_{m \neq n} \gamma_m - \frac{1}{N} \sum_{m \neq n} (Z_m^t Z_m)^{-1} Z_m^t \nu_m \end{aligned}$$

De este modo, hemos expresado $\gamma_n - \hat{\gamma}$ como combinación lineal de variables aleatorias independientes, pudiendo desarrollar fácilmente:

$$\begin{aligned} E((\hat{\gamma}_n - \hat{\gamma})(\hat{\gamma}_n - \hat{\gamma})^t) &= \left(\frac{N-1}{N}\right)^2 \Delta + \left(\frac{N-1}{N}\right)^2 \sigma_n^2 (Z_n^t Z_n)^{-1} - \frac{N-1}{N^2} \Delta \\ &\quad + \frac{1}{N^2} \sum_{m \neq n} \sigma_m^2 (Z_m^t Z_m)^{-1} \\ &= \frac{N-1}{N} \Delta + \frac{N-2}{N} \sigma_n^2 (Z_n^t Z_n)^{-1} + \frac{1}{N^2} \sum_{m=1}^N \sigma_m^2 (Z_m^t Z_m)^{-1} \end{aligned}$$

Por tanto:

$$\begin{aligned}
 E \left(\sum_{m=1}^N (\hat{\gamma}_m - \hat{\gamma})(\hat{\gamma}_m - \hat{\gamma})^t \right) &= (N-1)\Delta + \frac{N-2}{N} \sum_{m=1}^N \sigma_m^2 (Z_m^t Z_m)^{-1} + \frac{N}{N^2} \sum_{m=1}^N \sigma_m^2 (Z_m^t Z_m)^{-1} \\
 &= (N-1)\Delta + \frac{N-1}{N} \sum_{m=1}^N \sigma_m^2 (Z_m^t Z_m)^{-1} \\
 E \left(\frac{1}{N-1} \sum_{m=1}^N (\hat{\gamma}_m - \hat{\gamma})(\hat{\gamma}_m - \hat{\gamma})^t \right) &= \Delta + \frac{1}{N} \sum_{m=1}^N \sigma_m^2 (Z_m^t Z_m)^{-1}
 \end{aligned}$$

Teniendo en cuenta la expresión anterior, llegamos finalmente al estimador de Δ :

$$\hat{\Delta} = \frac{1}{N-1} \sum_{m=1}^N (\hat{\gamma}_m - \hat{\gamma})(\hat{\gamma}_m - \hat{\gamma})^t - \frac{1}{N} \sum_{m=1}^N \sigma_m^2 (Z_m^t Z_m)^{-1} \quad (4.9)$$

Este estimador es insesgado y T-consistente. Sin embargo, la matriz $\hat{\Delta}$ no es siempre semi-definida positiva. Claramente, para hacer factible el estimador sustituiremos las varianzas σ_m^2 por $\hat{\sigma}_m^2 \quad \forall m = 1, \dots, N$.

Una vez obtenidas las expresiones de los estimadores de σ_n^2 y Δ , podemos entonces realizar la estimación por (**MCG**).

4.3.2. Estimador de Grupos Medios

Si los supuestos distribucionales del modelo de Swamy no se cumplen, suponiendo únicamente exogeneidad de las variables predictoras y los errores e incorrelación de los errores, el parámetro γ puede ser estimado por el *Método de Grupos Medios* (**MG**), siendo este un enfoque no pramétrico que sencillamente es:

$$\hat{\gamma}_{\text{MG}} = \frac{1}{N} \sum_{n=1}^N \hat{\gamma}_{\text{MCO},n} \quad (4.10)$$

La varianza puede expresarse a través de la covarianza empírica del individuo n:

$$\text{Var}(\hat{\gamma}_{\text{MG}}) = \frac{1}{N(N-1)} \sum_{n=1}^N (\hat{\gamma}_{\text{MCO},n} - \hat{\gamma}_{\text{MG}})(\hat{\gamma}_{\text{MCO},n} - \hat{\gamma}_{\text{MG}})^t \quad (4.11)$$

Podemos observar la similitud con la varianza obtenida por el modelo de Swamy, siendo la varianza dada por el estimador de Grupos Medios una versión simplificada. Esta es, si se cumplen las hipótesis del modelo de Swamy sesgado y T-consistente, aunque es siempre semidefinida postiva a diferencia del anteriormente mencionado. De hecho, Swamy (1970) sugiere este estimador como una alternativa para los casos en los que $\hat{\Delta}$ no es semidefinida positiva.

Capítulo 5

Endogeneidad

5.1. Introducción

Consideremos el estimador por (MCO) dado por (1.7) :

$$\hat{\gamma}_{\text{MCO}} = (Z^t Z)^{-1} Z^t y$$

Sustituyendo y por $Z\gamma + \varepsilon$ y suponiendo que el tamaño de la muestra es N :

$$\hat{\gamma}_{\text{MCO}} = \gamma + (Z^t Z)^{-1} Z^t \varepsilon = \gamma + \left(\frac{1}{N} Z^t Z\right)^{-1} \frac{1}{N} Z^t \varepsilon$$

Este estimador es consistente si $\lim_{N \rightarrow \infty} \frac{1}{N} Z^t \varepsilon = 0$, es decir, si los regresores y el error son incorrelados. Análogamente, para los estimadores $\hat{\beta}_w$ y $\hat{\beta}_{\text{GLS}}$ la consistencia es cierta si admitimos la hipótesis anterior.

Definición 5.1 (Endogeneidad) Decimos que existe endogeneidad en un modelo si al menos una variable predictora está correlada con el término de error.

Este problema se puede deber a diversos motivos. Destacamos los tres más comunes:

- **Fallo en la medición:** si el verdadero modelo es $y = \alpha + \beta x + \varepsilon$ y observamos $\tilde{x} = x + \omega$ el modelo a estimar sería $y = \alpha + \beta(\tilde{x} - \omega) + \varepsilon$ o equivalentemente $y = \alpha + \beta\tilde{x} + \nu$ con $\nu = \varepsilon - \beta\omega$. Por consiguiente, ν está correlado con \tilde{x}
- **Omisión de variables:** si el verdadero modelo es $y = \alpha + \beta_1 x + \beta_2 z + \varepsilon$ y z no es contemplada o es inobservable, el modelo a estimar sería $y = \alpha + \beta_1 x + \nu$ donde $\nu = \beta_2 z + \varepsilon$. Por tanto, si la variable x y z fueran correladas tendríamos un problema de endogeneidad en el modelo.
- **Simultaneidad:** hace referencia al caso en el que alguna variable explicativa es determinada al mismo tiempo que la variable respuesta. Un ejemplo clásico de este hecho se

encuentra en el análisis de la oferta y la demanda de un bien o servicio. Consideremos un mercado en el que tanto la demanda como la oferta están determinadas por los precios de los mismos. Supongamos ahora que los precios incrementan debido a un factor externo como puede ser una subida de impuestos. Esto puede provocar una disminución de la demanda debido al alza de los precios y una mayor oferta debido a un menor margen de ganancias por producto. Sin embargo, debemos tener en cuenta que el precio también puede ser influenciado por la cantidad de demanda y oferta. La endogeneidad surge debido a la simultaneidad en la determinación de las variables.

Para hacer frente a este problema vamos a presentar el método de *Variables Instrumentales* (VI).

5.2. Método de Variables Instrumentales

Consideramos el modelo $y = Z\gamma + \varepsilon$ con $\text{Var}(\varepsilon) = \sigma^2 I$, con K variables explicativas y al menos una de ellas correlada con los errores.

Definición 5.2 (Variable Instrumental (VI)) Sea X una variable regresora del modelo con $\text{Cov}(X, \varepsilon) \neq 0$ (variable endógena) y Z una variable cumpliendo:

1. Z no está correlada con el error, es decir, $\text{Cov}(Z, \varepsilon) = 0$ (exogeneidad del instrumento).
2. Z está correlada con X , es decir, $\text{Cov}(Z, X) \neq 0$ (relevancia del instrumento).

Entonces se dice que Z es una variable intrumental o instrumento.

Observación 5.3 Para la elección de una variable instrumental cumpliendo $\text{Cov}(Z, \varepsilon) = 0$ debemos atender a argumentos basados en la naturaleza de los datos. Por ejemplo, si consideramos un modelo lineal simple en el que estudiamos el salario de una persona en función de su experiencia laboral, podríamos pensar que la variable que indica la última letra de la matrícula de su coche es incorrelada con el error del modelo, pues en el error estaría compuesto por variables omitidas como puede ser la educación. No es posible verificar con exactitud la condición de exogeneidad del instrumento debido a que ε es desconocido.

Observación 5.4 La relevancia del instrumento sí puede ser comprobada. La manera más sencilla sería considerar la regresión $X = \pi_0 + \pi_1 Z + \nu$ y plantear el contraste de hipótesis con $H_0 : \pi_1 = 0$. Si rechazamos la hipótesis nula podríamos afirmar que $\text{Cov}(Z, X) \neq 0$.

De forma análoga a Z vamos a considerar L la matriz formada por las variables instrumentales (con la primera columna también formada por unos) de tamaño $N \times (M + 1)$ donde M es el número de variables instrumentales. Por la construcción de la matriz de variables instrumentales se cumple $\lim_{N \rightarrow \infty} \frac{1}{N} L^t \varepsilon = 0$. Vamos a suponer sin pérdida de generalidad que $M \geq K$, de

forma que tomamos como instrumento todas las variables exógenas y al menos un instrumento por cada variable endógena.

5.2.1. Estimadores mediante VI

En el caso particular en el que contemos con el mismo número de instrumentos que de covariables se puede obtener el estimador mediante el planteamiento del sistema $L^t \varepsilon = 0$ o equivalentemente $L^t(y - Z\gamma) = 0$. Este sistema puede ser resuelto directamente, obteniendo:

$$\hat{\gamma} = (L^t Z)^{-1} L^t y \quad (5.1)$$

Observación 5.5 Pese a que el sistema anterior tiene solución única (exceptuando el caso de la presencia de multicolinealidad perfecta en las covariables o las variables instrumentales), el estimador (5.1) no es único, pues depende de la matriz L seleccionada.

Si contamos con un mayor número de variables instrumentales ($M > K$) el sistema anterior está sobredeterminado, luego en general no va a tener solución. Vamos a presentar dos procedimientos para llegar al estimador de γ .

El primero de ellos consiste en multiplicar el modelo por L^t , teniendo así:

$$L^t y = L^t Z \gamma + L^t \varepsilon$$

La matriz de covarianzas de los errores es $\text{Var}(L^t \varepsilon) = \sigma^2 L^t L$. Como carecemos de incorrelación en los errores apliquemos (**MCG**), obteniendo de este modo:

$$\hat{\gamma}_{\text{VI}} = [Z^t L (L^t L)^{-1} L^t Z]^{-1} Z^t L (L^t L)^{-1} L^t y = (Z^t P_L Z)^{-1} Z^t P_L y \quad (5.2)$$

donde $P_L = L (L^t L)^{-1} L^t$.

El segundo procedimiento consiste en el denominado estimador de *Mínimos Cuadrados en 2 Etapas* (**MC2E**), que consiste en lo siguiente:

Etapas 1: Consideremos la regresión de cada variable z sobre todas las variables instrumentales. Esta es denominada la forma reducida de z . Aplicando (**MCO**) a la regresión obtenemos el estimador $\hat{\gamma} = (L^t L)^{-1} L^t z$ y los valores ajustados $\hat{z}_L = L \hat{\gamma} = L (L^t L)^{-1} L^t z = P_L z$. En general denotamos $\hat{Z}_L = P_L Z$ los valores ajustados de las variables predictoras para la regresión anterior.

Etapas 2: Planteamos ahora la regresión de la variable respuesta sobre \hat{Z}_L . Esto es $y = \hat{Z}_L \gamma + \varepsilon$. Mediante (**MCO**) llegamos a:

$$\hat{\gamma}_{\text{MC2E}} = (\hat{Z}_L^t \hat{Z}_L)^{-1} \hat{Z}_L^t y \quad (5.3)$$

Observación 5.6 P_L es la matriz de proyección en el subespacio generado por las variables instrumentales. Se comprueba fácilmente que es una matriz simétrica e idempotente.

Observación 5.7 Desarrollando (5.3) teniendo en cuenta la observación anterior:

$$\hat{\gamma}_{\text{MC2E}} = (\hat{Z}_L^t \hat{Z}_L)^{-1} \hat{Z}_L^t y = (Z^t P_L^t P_L Z)^{-1} Z^t P_L^t y = (Z^t P_L Z)^{-1} Z^t P_L y$$

Quedando demostrada de esta manera la equivalencia de (5.2) y (5.3)

5.3. Estimadores mediante VI para el modelo de componente de error unidireccional

Volvamos ahora al modelo de datos panel $y = Z\gamma + \varepsilon$ con el error compuesto por el efecto individual η y la parte idiosincrática ν . Debemos tener en cuenta la correlación que puedan tener las variables con ambas componentes del error. En el modelo de efectos fijos, la matriz de transformación intraindividual elimina los efectos individuales del modelo, luego no tenemos por qué tener en cuenta la posible correlación de las variables con los efectos individuales. Presentaremos el estimador de γ para ese caso y posteriormente para el modelo general, es decir, para el cual no se ha llevado a cabo ninguna transformación.

5.3.1. Estimador intragrupos mediante VI

Consideremos el modelo multiplicado previamente por W , $Wy = WZ\gamma + W\nu$. Aplicamos (VI) multiplicando por L^t obteniendo así:

$$L^t Wy = L^t WZ\gamma + L^t W\nu$$

La matriz de covarianzas del modelo es $\text{Var}(L^t W\nu) = \sigma^2 L^t W L$. Aplicamos (MCG) para llegar al estimador:

$$\hat{\gamma}_{\text{VIw}} = (Z^t P_{\text{WL}} Z)^{-1} Z^t P_{\text{WL}} y \quad (5.4)$$

donde $P_{\text{WL}} = WL(L^t WL)^{-1} L^t W$.

Este estimador es consistente incluso en el caso de haber correlación de las variables con los efectos individuales. Sin embargo, si se cumple lo anterior no es muy eficiente ya que solo tiene en cuenta una fuente de variabilidad de la muestra.

5.3.2. Estimador para el modelo general

En el marco de las posibles situaciones que podemos discernir según la correlación de las variables con las componentes de error, vimos que si no existe correlación entre las variables y ambas componentes del error, el mejor estimador es el dado por (MCG), y que si existe correlación de algunas variables con los efectos individuales el estimador intragrupos es el que mantiene la consistencia, aunque no es muy eficiente debido a que no explota ambas fuentes de variabilidad del panel. También existe la posibilidad de correlación de algunas variables con la parte idiosincrática del error, en cuyo caso ninguno de los estimadores dados posee consistencia.

Nuestro objetivo es ahora llegar a un estimador mediante el uso de (VI) que sea consistente y que además aproveche las dos fuentes de variabilidad de la muestra.

Podemos distinguir tres tipos de variables regresoras según la correlación con las componentes del error.

1. Las variables endógenas, que son aquellas que están correladas con ambas componentes del error.
2. Las variables simplemente exógenas, que son incorreladas con la parte idiosincrática y correladas con la parte correspondiente a los efectos individuales.
3. Las variables doblemente exógenas, que están incorreladas con ambas componentes del error.

El mayor problema para encontrar un buen estimador radica en la captación de los instrumentos adcaudos. Esto en general es dificultoso como dijimos en la sección anterior, pero en este caso la riqueza de los datos panel sí nos va a permitir hacer frente a este problema empleando las matrices de transformación intraindividual e interindividual.

Las variables del primer tipo no proporcionan ningún instrumento. Las correspondientes a la segunda categoría proporcionan un instrumento dado por la transformación intraindividual debido a la ortogonalidad con los efectos individuales que se da por construcción. En cuanto al tercer grupo, van a proporcionar dos instrumentos mediante el uso de ambas transformaciones.

Observación 5.8 Sea x una variable predictora invariante en tiempo, se tiene $Wx = 0$ y $Bx = x$, luego mediante las transformaciones o bien proporcionan un instrumento (que sería la variable en sí misma en el caso en el que no exista correlación con los efectos individuales) o bien, en caso contrario, no aportan ninguno.

Consideramos el modelo $\Omega^{-\frac{1}{2}}y = \Omega^{-\frac{1}{2}}Z\gamma + \Omega^{-\frac{1}{2}}\varepsilon$ que ha sido multiplicado previamente por $\Omega^{-\frac{1}{2}} = \frac{1}{\sigma_v}W + \frac{1}{\sigma_l}B$ y nos planteamos aplicar el método de (VI). Sea L_1 y L_2 las matrices cuyas columnas están formadas por las variables doblemente exógenas y simplemente exógenas respectivamente y $L = (L_1, L_2)$. Estas variables aportan los instrumentos $A = (WL, B\tilde{L})$ donde \tilde{L} es el conjunto de variables que proporciona un instrumento al aplicarse la transformación interindividual, es decir, las variables doblemente exógenas.

Denotamos la matriz de proyección $P_A = A(A^tA)^{-1}A^t$ Esta matriz puede ser expresada mediante la suma de dos matrices de proyecciones definidas por los instrumentos dados por L transformados por las matrices B y W :

$$P_A = WL(L^tWL)^{-1}L^t + B\tilde{L}(\tilde{L}^tB\tilde{L})^{-1}\tilde{L}^tB = P_{WL} + P_{B\tilde{L}}$$

El estimador es por tanto:

$$\hat{\gamma}_{VI} = (Z^t\Omega^{-\frac{1}{2}}P_A\Omega^{-\frac{1}{2}}Z)^{-1}Z^t\Omega^{-\frac{1}{2}}P_A\Omega^{-\frac{1}{2}}y$$

$$\hat{\gamma}_{VI} = \left(\frac{1}{\sigma_v^2} Z^t P_{WL} Z + \frac{1}{\sigma_l^2} Z^t P_{BL} Z \right)^{-1} \left(\frac{1}{\sigma_v^2} Z^t P_{WL} y + \frac{1}{\sigma_l^2} Z^t P_{BL} y \right) \quad (5.5)$$

Considerando de nuevo la notación $\phi^2 = \frac{\sigma_v^2}{\sigma_l^2}$:

$$\hat{\gamma}_{VI} = (Z^t P_{WL} Z + \phi^2 Z^t P_{BL} Z)^{-1} (Z^t P_{WL} y + \phi^2 Z^t P_{BL} y) \quad (5.6)$$

Vamos a presentar ahora algunos casos especiales en función del tipo de variables predictoras que estén presentes en el modelo.

5.3.2.1. Caso $L = Z$ y $L = 0$

Si todas las covariables son simplemente exógenas se tiene $L = Z$ y $\tilde{L} = 0$, luego (5.6) coincide con $\hat{\gamma}_{VIW}$. La matriz A va a estar formada por todas las variables del modelo transformadas por W . En el caso en el que estén presentes variables endógenas y simplemente exógenas, debemos recurrir a al menos tantos instrumentos externos como variables endógenas haya. Si estos son a su vez simplemente exógenos se cumple también $\tilde{L} = 0$ y por tanto el estimador también coincidiría con (5.4).

5.3.2.2. Caso $L = \tilde{L}$

Este caso se da cuando todas las covariables son doblemente exógenas, luego cada una de ellas se utiliza dos veces, aprovechando ambas transformaciones. Consideramos los modelos:

$$\begin{aligned} L^t W y &= L^t W Z + L^t W \varepsilon \\ L^t B y &= L^t B Z + L^t B \varepsilon \end{aligned}$$

Agrupando ambas ecuaciones, pues el parámetro γ a calcular es el mismo en ambos modelos obtenemos

$$\begin{pmatrix} L^t W y \\ L^t B y \end{pmatrix} = \begin{pmatrix} L^t W Z \\ L^t B Z \end{pmatrix} \gamma + \begin{pmatrix} L^t W \varepsilon \\ L^t B \varepsilon \end{pmatrix}$$

La matriz de covarianzas de los errores del modelo conjunto es:

$$Var \begin{pmatrix} L^t W \varepsilon \\ L^t B \varepsilon \end{pmatrix} = E \begin{pmatrix} L^t W \varepsilon \varepsilon^t W L & L^t W \varepsilon \varepsilon^t B L \\ L^t B \varepsilon \varepsilon^t W L & L^t B \varepsilon \varepsilon^t B L \end{pmatrix} = \sigma_v^2 \begin{pmatrix} L^t W L & 0 \\ 0 & \phi^2 L^t B L \end{pmatrix}$$

Podemos aplicar ahora (**MCG**) y mediante un cálculo sencillo llegamos al estimador (**EC2SLS**):

$$\hat{\gamma}_{EC2SLS} = (Z^t P_{WL} Z + \phi^2 Z^t P_{BL} Z)^{-1} (Z^t P_{WL} y + \phi^2 Z^t P_{BL} y) \quad (5.7)$$

que como vemos coincide con (5.6).

Capítulo 6

Caso práctico en R

Vamos a destinar este capítulo a la explicación e interpretación de los resultados del código desarrollado en RStudio¹, complementada principalmente con la versión 2.6-3 del paquete **plm**, que contiene funciones y métodos de ajuste relevantes para este tipo de datos. Se van a usar adicionalmente los siguientes paquetes:

- **AER**: para la carga de datos.
- **lattice**: incluye funciones gráficas avanzadas.
- **ggplot2**: incluye funciones gráficas avanzadas.
- **cramer**: para la realización de un contraste no paramétrico de diferencia de medias.

Utilizaremos el dataset **Guns** del paquete **AER**. Esta base de datos contiene variables relacionadas con la criminalidad en los estados de EEUU además del distrito de Columbia desde el año 1977 hasta 1999. Se trata de un panel de datos balanceado con 1173 observaciones y 13 variables que se describen a continuación:

state: factor indicando el estado.

year: factor indicando el año.

violent: tasa de delitos violentos (incidentes por 100.000 personas).

murder: tasa de asesinatos (incidentes por 100.000 personas).

robbery: tasa de robo (incidentes por 100.000 personas).

prisoners: tasa de encarcelamientos del año anterior (por cada 100.000 personas).

afam: porcentaje de la población con edad entre 10 y 64 años que es afroamericana.

¹Se puede acceder al código completo en <https://github.com/carlosperman/AnexoGuns>

cauc: porcentaje de la población con edad entre 10 y 64 años que es caucásica.

male: porcentaje de la población que es hombre, con edad entre 10 y 29 años.

population: población en millones de personas.

income: ingreso personal per cápita en dólares.

density: densidad de población por milla cuadrada de superficie terrestre dividida por 1.000.

law: variable de tipo factor que indica si el estado cuenta con una ley de portabilidad de armas vigente (1) o no (0).

La variable a predecir en nuestro estudio va a ser **violent**.

6.1. Carga de los paquetes y del dataset

Cargamos los paquetes que vamos a necesitar a lo largo del script:

```
library(AER)
library(plm)
library(lattice)
library(ggplot2)
```

Importamos también el conjunto de datos. Vamos a reordenar las variables de manera que la primera y segunda columna sean la variable individual y temporal respectivamente. Esto permite el almacenamiento de los datos en un tipo de objeto `pdata.frame`, siendo este una extensión del clásico `data.frame` aunque incluyendo un doble índice para cada observación (primero individual y luego temporal).

```
data("Guns")
Guns = cbind(Guns$state , Guns[,-12])
colnames(Guns)[1] = "state"
Guns = pdata.frame(Guns)
```

```
pdim(Guns)

## Balanced Panel: n = 51, T = 23, N = 1173
```

6.2. Análisis descriptivo

Realicemos un análisis descriptivo de los datos previo a la preparación y aplicación de los modelos.

```
GunsCont = Guns[ , -c(1,2,13)]
(matcor = round(cor(as.matrix(GunsCont)),2))

##          violent murder robbery prisoners  afam  cauc  male population income
## violent      1.00   0.83   0.91    0.70  0.57 -0.57 -0.17      0.32   0.41
## murder       0.83   1.00   0.80    0.71  0.60 -0.62  0.01      0.10   0.22
## robbery      0.91   0.80   1.00    0.57  0.58 -0.58 -0.09      0.32   0.41
## prisoners    0.70   0.71   0.57    1.00  0.53 -0.53 -0.45      0.10   0.46
## afam         0.57   0.60   0.58    0.53  1.00 -0.98  0.02      0.06   0.26
## cauc        -0.57  -0.62  -0.58   -0.53 -0.98  1.00 -0.01     -0.07  -0.19
## male        -0.17   0.01  -0.09   -0.45  0.02 -0.01  1.00     -0.10  -0.53
## population   0.32   0.10   0.32    0.10  0.06 -0.07 -0.10      1.00   0.22
## income       0.41   0.22   0.41    0.46  0.26 -0.19 -0.53      0.22   1.00
## density      0.66   0.75   0.78    0.56  0.54 -0.56 -0.06     -0.08   0.34
##          density
## violent      0.66
## murder       0.75
## robbery      0.78
## prisoners    0.56
## afam         0.54
## cauc        -0.56
## male        -0.06
## population   -0.08
## income       0.34
## density      1.00

det(round(cor(as.matrix(GunsCont[, -1])),2))

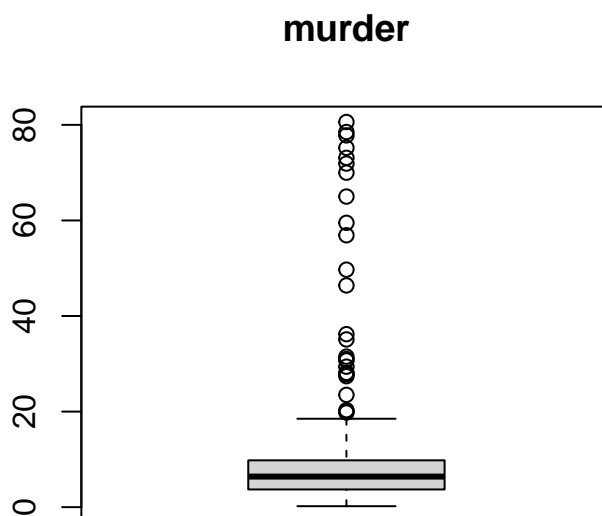
## [1] 0.0001581329
```

Contamos con la presencia de multicolinealidad en las covariables.

Destacamos una alta correlación lineal de la variable **violent** con **robbery**, **murder** y **prisoners** principalmente.

Las variables cuentan con una gran cantidad de casos atípicos. Mostramos un ejemplo:

```
boxplot(GunsCont[,2] , main = colnames(GunsCont)[2])
```



Vamos a llevar a cabo en lo que resta la transformación logarítmica en las variables continuas para reducir la influencia de estos valores extremos. Es razonable la gran presencia de casos atípicos debido a que EE.UU es un país muy extenso y en el que hay mucha diferencia de estilo de vida entre unos estados y otros.

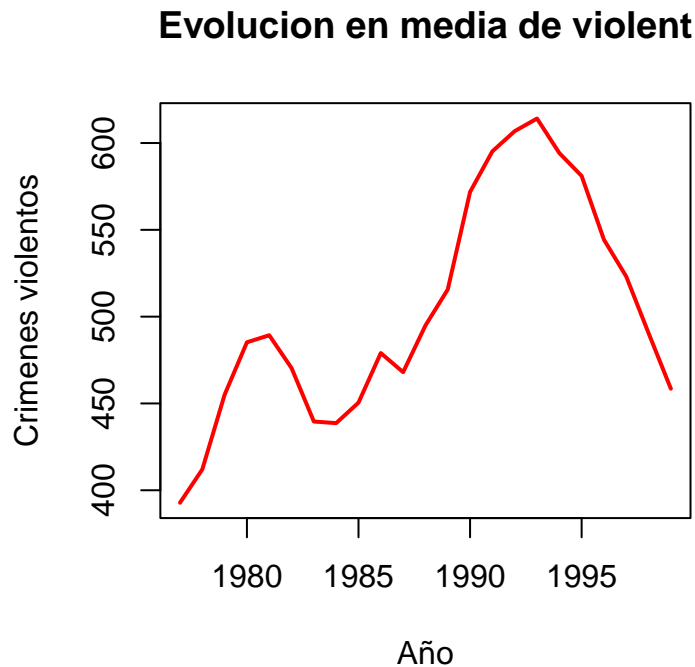
6.2.1. Visualización de la variable objetivo violent

6.2.1.1. Comportamiento medio global a lo largo del tiempo

Veamos como evoluciona en tiempo la variable `violent` en media globalmente, es decir, en todos los estados.

```
T = pdim(Guns)$nT$T
time = paste(1977:(1977+(T-1)))
medviol = rep(NA,T)
for (i in 0:(T-1)){
  medviol[i+1] = mean(Guns[Guns[,2] == (1977 + i), 3])
}
medviol = cbind(medviol)
rownames(medviol) = time
matplot(time , medviol , type = "l" , col = "red" , lwd = 2 ,
```

```
xlab = "Año" , ylab = "Crímenes violentos" ,  
main = "Evolucion en media de violent" )
```



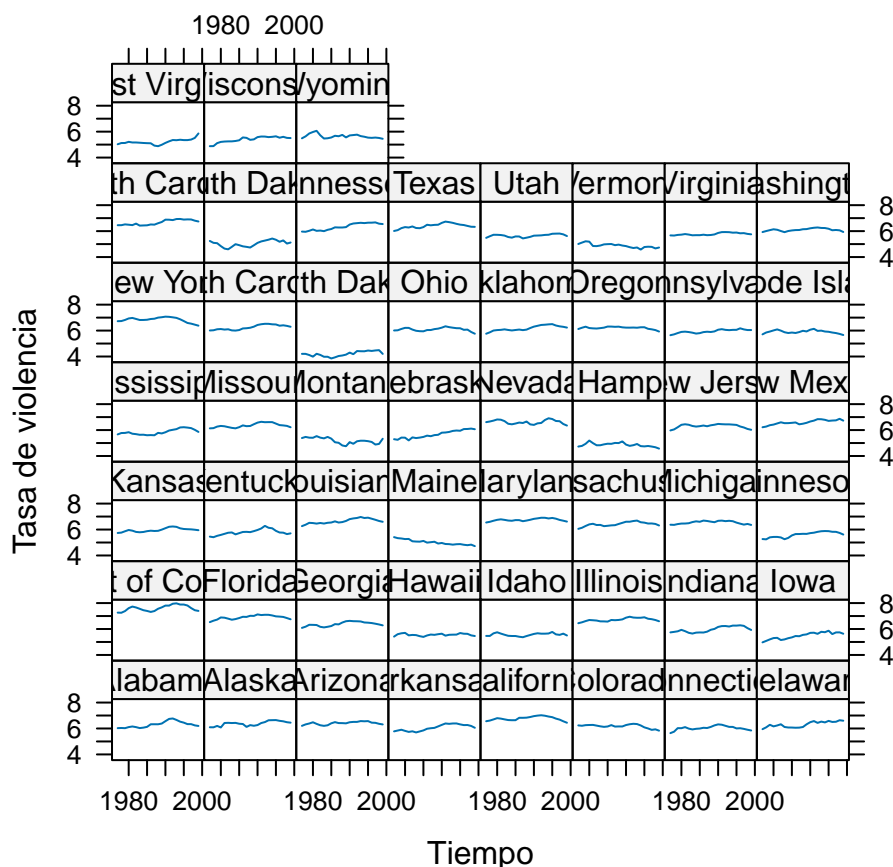
```
which.max(medviol)  
  
## [1] 17  
  
medviol[which.max(medviol)]  
  
## [1] 614.1098  
  
which.min(medviol)  
  
## [1] 1  
  
medviol[which.min(medviol)]  
  
## [1] 392.8255
```

En principio, podemos observar una clara subida de la tasa de crímenes violentos a partir del año 1987, con un maximo de 614.1098 en 1994. A partir de ese momento notamos una clara bajada de los casos. En un principio la tasa es mucho menor al resto de años. De hecho la tasa minima se encuentra en el primer año registrado (1977).

6.2.1.2. Comportamiento a lo largo del tiempo en los distintos estados

Mostramos a continuación la siguiente gráfica:

```
xyplot(log(violent) ~ as.numeric(as.character(year)) | state, data = Guns,
       type = "l", xlab = "Tiempo" , ylab = "Tasa de violencia")
```



Podemos observar que en general `violent` fluctúa bastante en el tiempo. Destaca por su alta criminalidad el Distrito de Columbia claramente sobre el resto. Le siguen Florida y New York.

6.2.2. Estudio del factor `law`

Esta es la única covariable que es de tipo factor.

```
table(Guns$law)
```

```
##
## no yes
## 888 285
```

En 888 observaciones no hay una ley vigente mientras que en 285 sí. Calculemos la media de la variable objetivo distinguiendo las observaciones según el factor `law`.

```
meanviollaw = cbind(t(sapply(split(Guns$violent, Guns$law) ,
                                function(m) mean(m))))
rownames(meanviollaw) = names(Guns)[3]
meanviollaw

##              no              yes
## violent 542.2377 381.0509
```

Aparentemente la tasa de crímenes violentos es mayor en los casos en los que no hay vigente una ley de portabilidad de armas.

Mostremos ahora la evolución de la variable `law` a lo largo del tiempo.

```
table(Guns$law, Guns$year)

##
##      1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990
## no      47  47  47  47  47  46  46  46  46  44  43  42  42  39
## yes     4   4   4   4   4   5   5   5   5   7   8   9   9  12
##
##      1991 1992 1993 1994 1995 1996 1997 1998 1999
## no      36  34  34  34  30  25  22  22  22
## yes     15  17  17  17  21  26  29  29  29
```

Observamos un claro aumento a lo largo del tiempo del número de estados con una ley de portabilidad de armas vigente, lo que nos conduce a reflexionar lo siguiente:

Hemos visto que la media de crímenes violentos del grupo de observaciones donde `law = 1` es menor que en el caso `law = 0`, luego podríamos pensar que el factor `law = 1` hace disminuir la tasa de criminalidad, sin embargo, si observamos la tabla anterior y el gráfico representado anteriormente sobre la evolución de la media de `violent`, en el periodo de 1987 a 1994 aumentó considerablemente el número de estados con una ley vigente, no obstante, la media de la tasa de criminalidad crece significativamente en ese periodo. Posteriormente sigue en aumento el número de estados con `law = 1` mientras que la media de la tasa de criminalidad baja. Nos surge por tanto la pregunta de si realmente la variable `law` es significativa. Razonando con una regresión lineal simple:

```
lm(log(violent) ~ law, data = Guns)$coefficients[2]

##      lawyes
```

```
## -0.4429646
```

El coeficiente es negativo, concluimos por tanto que el factor $\text{law} = 1$ disminuye la tasa de crímenes violentos.

6.3. Preparación para los modelos

Debido a que vimos anteriormente que contamos con un alto grado de multicolinealidad en las variables regresoras, debemos decidir qué variables escoger para los modelos. Destacamos las siguientes correlaciones lineales entre variables predictoras:

- `cauc` y `afam`: -0.98
- `murder` y `robbery`: 0.8
- `robbery` y `density`: 0.78
- `murder` y `density`: 0.75

Procederemos eliminando del modelo una variable de cada uno de los siguientes pares: (`cauc`, `afam`), (`murder`, `robbery`). Con la idea de eliminar las variables con mayores casos atípicos, optamos por suprimir de los modelos las variables `cauc` y `robbery`. Con el propósito de tener una mayor seguridad, podemos hacer uso del *factor de inflación de la varianza* (**FIV**) o más bien conocido por sus siglas en inglés (**VIF**). Consideraremos variables que generan una alta multicolinealidad aquellas con un **VIF** mayor que 5

Cálculo del (**VIF**) con todas las covariables:

```
(vifvalues <- car::vif(lm(log(violent) ~ log(murder) + log(robbery)
+ log(prisoners) + log(afam) + log(cauc) + log(male)
+ log(population) + log(income) + log(density),
data = Guns)))
```

##	log(murder)	log(robbery)	log(prisoners)	log(afam)	log(cauc)
##	4.864927	6.431694	4.069917	5.661684	4.032101
##	log(male)	log(population)	log(income)	log(density)	
##	2.953319	2.464337	2.223232	2.430003	

Eliminando la variable `cauc`:

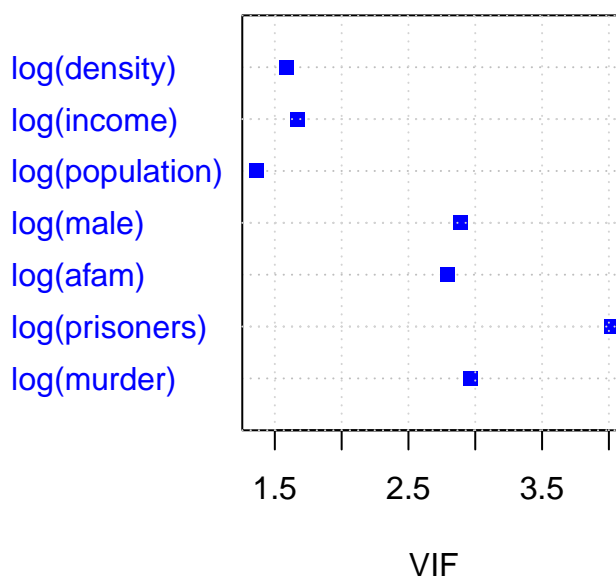
```
(vifvalues <- car::vif(lm(log(violent) ~ log(murder) + log(robbery)
+ log(prisoners) + log(afam) + log(male)
+ log(population) + log(income) + log(density),
data = Guns)))
```


##	log(murder)	log(robbery)	log(prisoners)	log(afam)	log(male)
##	4.836668	6.376678	4.035344	2.865385	2.952948
##	log(population)	log(income)	log(density)		
##	1.767345	2.220092	1.859972		

Eliminando además la variable **robbery**:

```
vifvalues <- car::vif(lm(log(violent) ~ log(murder) + log(prisoners)
                        + log(afam) + log(male) + log(population)
                        + log(income) + log(density), data = Guns))
dotchart(vifvalues , main = c("VIF de las covariables continuas") ,
         pch = 15 , col = "blue" , xlab = "VIF")
grid()
```

VIF de las covariables continuas



Todos los **VIF** son menores o iguales que 5. De esta manera hemos reducido la multicolinealidad considerablemente.

Definimos la formula que describe la variables predictora y las covariables de los modelos que se plantean a continuación.

```
form <- log(violent) ~ log(murder) + log(prisoners) + log(afam) +
      log(male) + log(population) + log(income) + log(density) + law
```

6.4. Modelos de componente de error unidireccional

Pasamos a la aplicación de los modelos de componente de error unidireccional.

6.4.1. Modelo de efectos fijos

Consideramos el modelo de efectos fijos con las variables escogidas en el capítulo anterior. Podríamos aplicar la función `summary` para obtener los outputs más importantes del modelo.

```
Guns.within1 <- plm(form , data = Guns ,
                    effect = "individual" , model = "within")
summary(Guns.within1)
```

Las estimaciones de las pendientes del modelo son las siguientes:

```
coef(Guns.within1)
```

##	log(murder)	log(prisoners)	log(afam)	log(male)	log(population)
##	0.26778135	0.06154899	-0.23112739	-0.75615261	3.93523600
##	log(income)	log(density)	lawyes		
##	0.18613963	-3.94469574	0.02503313		

Para los coeficientes correspondientes a los efectos individuales tenemos distintas opciones,

- `"level"` (valor por defecto): devuelve los interceptos individuales, es decir $\hat{\alpha} + \hat{\eta}_n$. Esta es la opción más usual.
- `"dfirst"`: toma las desviaciones de los efectos individuales respecto del primer individuo, el cual se toma como estimación del intercepto $\hat{\alpha}$.
- `"dmean"`: devuelve las desviaciones de los efectos individuales respecto de las medias. En este caso $\hat{\alpha}$ es la media de los interceptos individuales.

Como el comportamiento predeterminado de `lm` con una variable de tipo factor es eliminar el primer nivel del factor, podemos ver numéricamente la equivalencia del estimador **LSDV** con el modelo de efectos fijos estableciendo la opción `"dfirst"`.

```
coeflsdv <- head(coef(lm(log(violent) ~ log(murder) + log(prisoners)
                      + log(afam) + log(male) + log(population) + log(income)
                      + log(density) + law + factor(state), data = Guns))[-1])
coefwithin1 <- head(coef(Guns.within1 , type = "dfirst"))
cbind(coeflsdv , coefwithin1)
```

```
##               coeflsdv coefwithin1
## log(murder)    0.26778135  0.26778135
## log(prisoners) 0.06154899  0.06154899
## log(afam)     -0.23112739 -0.23112739
## log(male)     -0.75615261 -0.75615261
## log(population) 3.93523600  3.93523600
## log(income)   0.18613963  0.18613963
```

6.4.2. Modelos de efectos aleatorios

Para el modelo de efectos aleatorios, los resultados van a depender del estimador que escogamos de ε para las estimaciones de las componentes de la varianza, aunque como veremos, los resultados son muy similares. Presentamos aquellas dadas por Wallace-Hussain (1969), Amemiya (1971), Swamy-Arora (1972) y Nerlove (1971).

```
Guns.rwalhus1 <- plm(form , data = Guns , effect = "individual" ,
                     model = "random", random.method = "walhus")
Guns.ramemiya1 <- update(Guns.rwalhus1 , random.method = "amemiya")
Guns.rswar1 <- update(Guns.rwalhus1 , random.method = "swar")
Guns.rnerlove1 <- update(Guns.rwalhus1 , random.method = "nerlove")
Guns.rmodels1 <- list( walhus = Guns.rwalhus1 , amemiya = Guns.ramemiya1,
                      swar = Guns.rswar1 , nerlove = Guns.rnerlove1)
```

Los coeficientes para los distintos modelos son:

```
sapply(Guns.rmodels1 , function(m) coef(summary(m))[, "Estimate"])
```

##	walhus	amemiya	swar	nerlove
## (Intercept)	4.0918169958	4.09755146	4.646234546	3.98492048
## log(murder)	0.3286685254	0.26139135	0.315665764	0.26143264
## log(prisoners)	0.1049872738	0.05625341	0.092476132	0.05629371
## log(afam)	0.0618372109	-0.21121950	0.030804968	-0.21146764
## log(male)	-0.2178106955	-0.70962662	-0.304327281	-0.71016703
## log(population)	0.1162834599	0.51378520	0.113694429	0.54453012
## log(income)	0.1414692550	0.16996480	0.123185155	0.17013665
## log(density)	0.0569070855	-0.48712608	0.060828590	-0.51829699
## lawyes	-0.0007825797	0.02253809	0.004225241	0.02256707

Analicemos la información proporcionada por la función `ercomp`:

```
lapply(Guns.rmodels1, function(m) ercomp(m))

## $walhus
##               var std.dev share
## idiosyncratic 0.03282 0.18116 0.385
## individual    0.05235 0.22880 0.615
## theta: 0.8371
##
## $amemiya
##               var std.dev share
## idiosyncratic 0.02331 0.15268 0.001
## individual    34.48791 5.87264 0.999
## theta: 0.9946
##
## $swar
##               var std.dev share
## idiosyncratic 0.02348 0.15323 0.281
## individual    0.05998 0.24490 0.719
## theta: 0.8706
##
## $nerlove
##               var std.dev share
## idiosyncratic 0.0223 0.1493 0.001
## individual    35.1787 5.9312 0.999
## theta: 0.9948
```

Observamos que la varianza de los efectos individuales en los métodos de Amemiya y Nerlove son muy superiores al resto, explicando prácticamente la totalidad de la varianza, mientras que por el método de Walhus y Swamy la variabilidad individual explican un 61.5 % y un 72 % de la totalidad respectivamente. En el siguiente apartado justificaremos por qué se da esta situación.

6.4.3. Comparación de los modelos de componente de error unidireccional

Veamos en primer lugar el parámetro θ para los distintos modelos de efectos aleatorios.

```
sapply(Guns.rmodels1 , function(m) ercomp(m)$theta)

## walhus.id amemiya.id swar.id nerlove.id
## 0.8371088 0.9945789 0.8706309 0.9947504
```



```
## idiosyncratic 0.02348 0.15322 0.065
## individual    0.33862 0.58191 0.935
## theta: 0.9452
```

```
ercomp(Guns.rnerlove1)
```

```
##                var std.dev share
## idiosyncratic 0.02246 0.14986 0.061
## individual    0.34643 0.58858 0.939
## theta: 0.947
```

Comparamos las desviaciones típicas de los estimadores de las pendientes. Como los resultados son muy similares entre los distintos métodos de efectos aleatorios, seleccionamos únicamente el dado por Swamy-Arora (1972)

```
difstd <- rbind(coef(summary(Guns.within1))[ , "Std. Error"],
               coef(summary(Guns.rswar1))[ , "Std. Error"][-c(1, 6 , 8)])

rownames(difstd) = c("within" , "random")
difstd

##          log(murder) log(prisoners)  log(afam) log(male) log(income)      lawyes
## within  0.02049414      0.02599681 0.05186876 0.1245495  0.08462531 0.01670420
## random  0.02000589      0.02589970 0.03642558 0.1221259  0.08213188 0.01714188
```

Las desviaciones correspondientes al estimador intragrupos son algo mayores tal y como dijimos en la sección 1.6.

6.4.4. Verificación y elección del modelo

En primer lugar veamos si verdaderamente es razonable el uso de un modelo de componente de error unidireccional. Para ello vamos a testear la presencia conjunta de efectos individuales en el panel de datos.

Esto se realizará mediante el F-test explicado en el capítulo 3. Este test está implementado en el paquete **plm** mediante la función **pFtest**.

```
Guns.pooling <- plm(form , data = Guns, effect = "individual", model = "pooling")
pFtest(Guns.within1, Guns.pooling)

##
## F test for individual effects
```

```
##
## data:  log(violent) ~ log(murder) + log(prisoners) + log(afam) + log(male) + ...
## F = 64.926, df1 = 48, df2 = 1116, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Se rechaza la hipótesis nula y por tanto hay presencia de efectos individuales en los datos. Por último, suponiendo que no existe endogeneidad en las variables, debemos determinar si seleccionamos un modelo de efectos fijos o un modelo de efectos aleatorios.

Para ello contrastamos la incorrelación de los efectos individuales con los regresores. Hagamos entonces el test de Hausman. Escogeremos al igual que para el cálculo de las desviaciones el modelo de Swamy-Arora (1972) como representante de los modelos de efectos aleatorios.

```
phtest(Guns.within1, Guns.rswar1)

##
##  Hausman Test
##
## data:  log(violent) ~ log(murder) + log(prisoners) + log(afam) + log(male) + ...
## chisq = 616.69, df = 6, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

Se rechaza la hipótesis nula, por tanto hay correlación de los efectos individuales y los regresores. En este caso, el estimador intragrupos es el único que proporciona estimaciones consistentes.

6.5. Modelos de componente de error bidireccional

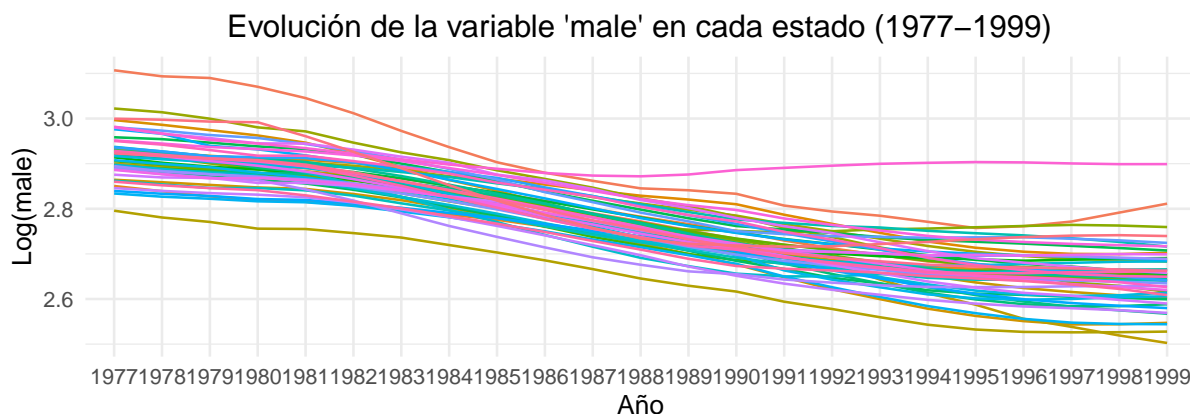
Pasamos ahora a la construcción de los modelos de componente de error bidireccional.

6.5.1. Modelo de efectos fijos

La transformación establecida por la matriz Q elimina las variables constantes para cada individuo, luego tomamos en principio para el modelo las variables consideradas finalmente para el modelo de componente de error unidireccional. Además Q elimina aquellas variables que sean constantes para los estados en cada instante de tiempo. Podemos ver gráficos de las covariables continuas escogidas para `Guns.within1` que nos permitan analizar ese aspecto. Ya vimos en el análisis descriptivo que la variable `law` no tenía este comportamiento. Destacamos en ese sentido la variable `male`.²

²Se pueden ver el resto de gráficas en el script completo.


```
ggplot(Guns, aes(x = year, y = log(male), group = state, color = factor(state))) +
  geom_line() +
  labs(title = "Evolución de la variable 'male' en cada estado (1977-1999)",
       x = "Año",
       y = "Log(male)") +
  theme_minimal() + theme(legend.position = "none",
plot.title = element_text(hjust = 0.5))
```



Vamos a eliminar esta variable del modelo para así tener estimaciones más favorables.

```
Guns.within2 <- plm(log(violent) ~ log(murder) + log(prisoners) + log(afam) +
  + log(income) + law, data = Guns ,
  effect = "twoways" , model = "within")
```

Para la estimación de los efectos temporales, contamos con las mismas opciones descritas para la componente individual. Destacamos la más usual:

```
fixef(Guns.within2 , effect = "time" , type = "level")
```

##	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
##	4.8928	4.9399	5.0409	5.1042	5.1028	5.0969	5.0618	5.0887	5.1146	5.1642	5.1535
##	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
##	5.2002	5.2443	5.3512	5.3953	5.4266	5.4334	5.4332	5.4263	5.3846	5.3821	5.3351
##	1999										
##	5.2907										

que estima los interceptos temporales, es decir, $\hat{\alpha} + \hat{\mu}_t$, $t \in \{1, \dots, T\}$.

Estableciendo **effect = twoways** se obtienen los interceptos en cada observación del panel, es decir, $\hat{\alpha} + \eta_n + \mu_t$ con $n \in \{1, \dots, N\}$, $t \in \{1, \dots, T\}$.

```
fixef(Guns.within2 , effect = "twoways" , type = "level")
```

Las pendientes de las covariables son las siguientes:

```
coef(Guns.within2)
```

##	log(murder)	log(prisoners)	log(afam)	log(income)	lawyes
##	0.17182147	-0.08599552	-0.37923624	0.21144407	-0.01159806

6.5.2. Modelos de efectos aleatorios

Los resultados que se obtienen en un modelo de efectos aleatorios van a depender del método escogido para la estimación de los parámetros desconocidos correspondientes a la transformación realizada sobre el modelo original.

```
Guns.rwalhus2 <- plm(form , data = Guns , effect = "twoways" ,
                    model = "random", random.method = "walhus")
Guns.ramemiya2 <- update(Guns.rwalhus2 , random.method = "amemiya")
Guns.rswar2 <- update(Guns.rwalhus2 , random.method = "swar")
Guns.rnerlove2 <- update(Guns.rwalhus2 , random.method = "nerlove")
Guns.rmodels2 <- list( walhus = Guns.rwalhus2 , amemiya = Guns.ramemiya2,
                    swar = Guns.rswar2 , nerlove = Guns.rnerlove2)
```

Veamos los parámetros θ .

```
sapply(Guns.rmodels2, function(x) ercomp(x)$theta)
```

##	walhus	amemiya	swar	nerlove
## id	0.844225	0.9944452	0.8913011	0.9946739
## time	0.5741935	0.9420086	0.28459	0.9451591
## total	0.5663023	0.9419833	0.2839826	0.9451342

Los parámetros θ_η para todos los modelos son bastante elevados, principalmente en los casos de `Guns.ramemiya2` y `Guns.rnerlove2` que son prácticamente 1. Como sucede en el caso unidireccional, la componente individual de la varianza es sobreestimada si se emplea alguno de estos métodos, pues tenemos presente en el modelo variables prácticamente constantes en el tiempo para los distintos estados (`population` y `density`). Además, para estos dos casos el parámetro θ_μ toma también un valor cercano a 1, lo cual va a suponer que la componente temporal del error sea también sobreestimada, ya que la transformación va a eliminar casi completamente las variables constantes en cada instante de tiempo para los distintos estados (`male`). Para los

modelos `Guns.rswar` y `Guns.rwalhus` el parámetro θ_η toma también un valor muy alto, siendo además en el modelo de Wallace y Hussain el parámetro θ_μ un valor considerable. Analizaremos luego si para estos modelos debemos eliminar alguna variable.

```
sapply(Guns.rmodels2, function(x) ercomp(x)$sigma2)

##           walhus      amemiya      swar      nerlove
## idios 0.03001371 0.01645544 0.0165759897 0.01543136
## id    0.05247204 23.18656599 0.0602753357 23.65102707
## time 0.00265732 0.09561995 0.0003100179 0.10030363
```

Vemos la sobreestimación clara de la componente individual y también aunque bastante más leve de la componente temporal en los modelos de Amemiya y Nerlove.

Reescribimos por tanto estos modelos eliminando las variables ya mencionadas.

```
Guns.ramemiya2 <- plm(log(violent) ~ log(murder) + log(prisoners) + log(afam) +
  + log(income) + law, data = Guns ,
  effect = "twoways" , model = "random" ,
  random.method = "amemiya")
Guns.rnerlove2 <- plm(log(violent) ~ log(murder) + log(prisoners) + log(afam) +
  + log(income) + law, data = Guns ,
  effect = "twoways" , model = "random" ,
  random.method = "nerlove")
Guns.rmodels2 <- list( walhus = Guns.rwalhus2 , amemiya = Guns.ramemiya2,
  swar = Guns.rswar2 , nerlove = Guns.rnerlove2)

sapply(Guns.rmodels2, function(x) ercomp(x)$theta)

##           walhus      amemiya      swar      nerlove
## id    0.844225 0.9655877 0.8913011 0.9670219
## time 0.5741935 0.8859875 0.28459 0.8926431
## total 0.5663023 0.8845375 0.2839826 0.891205

sapply(Guns.rmodels2, function(x) ercomp(x)$sigma2)

##           walhus      amemiya      swar      nerlove
## idios 0.03001371 0.01779966 0.0165759897 0.01669193
## id    0.05247204 0.65274134 0.0602753357 0.66658555
## time 0.00265732 0.02650052 0.0003100179 0.02806997
```

Los valores de los parámetros θ han experimentado unos cambios mínimos, mientras que las estimaciones de las componentes de la varianza sí se han reducido considerablemente.

Los coeficientes para los distintos modelos son:

```
sapply(Guns.rmodels2 , function(m) coef(summary(m))[, "Estimate"])
```

```
## $walhus
##      (Intercept)      log(murder)  log(prisoners)      log(afam)      log(male)
##      0.373340950      0.256336420      0.118647770      0.058815809      0.346206541
## log(population)      log(income)    log(density)      lawyes
##      0.127630857      0.373907485      0.059680540      0.004129197
##
## $amemiya
##      (Intercept)      log(murder)  log(prisoners)      log(afam)      log(income)
##      3.6021602604      0.1801485723     -0.0405158329     -0.2670015897      0.2804101656
##      lawyes
##      -0.0009257358
##
## $swar
##      (Intercept)      log(murder)  log(prisoners)      log(afam)      log(male)
##      2.7254136962      0.2657300227      0.0943080737      0.0003186885     -0.0715141480
## log(population)      log(income)    log(density)      lawyes
##      0.1155330220      0.2702078673      0.0633985069      0.0102484761
##
## $nerlove
##      (Intercept)      log(murder)  log(prisoners)      log(afam)      log(income)
##      3.689964120      0.179582786      -0.044199236     -0.275089699      0.274441170
##      lawyes
##      -0.001784265
```

6.5.3. Comparación de los modelos de componente de error bidireccional

Para los modelos de Amemiya y Nerlove, todos los parámetros θ son muy cercanos a uno, luego los coeficientes de efectos aleatorios estimados mediante esos métodos van a ser muy similares al de efectos fijos. Para el resto sí notaremos algunas variaciones respecto del modelo de efectos fijos, aunque no van a ser muy significativas si las variables con las que contamos en el modelo son adecuadas.

```
cbind(as.matrix(sapply(list(amemiya = Guns.ramemiya2 , nerlove = Guns.ramemiya2)
                        ,function(m) coef(m)))[-1,], within = coef(Guns.within2))
```

```
##              amemiya      nerlove      within
## log(murder)    0.1801485723  0.1801485723  0.17182147
## log(prisoners) -0.0405158329 -0.0405158329 -0.08599552
## log(afam)      -0.2670015897 -0.2670015897 -0.37923624
## log(income)    0.2804101656  0.2804101656  0.21144407
## lawyes        -0.0009257358 -0.0009257358 -0.01159806
```

```
cbind(as.matrix(sapply(Guns.rmodels2[c("walhus" , "swar")],
                        function(m) coef(m)))[-1,])
```

```
##              walhus      swar
## log(murder)    0.256336420  0.2657300227
## log(prisoners) 0.118647770  0.0943080737
## log(afam)      0.058815809  0.0003186885
## log(male)      0.346206541 -0.0715141480
## log(population) 0.127630857  0.1155330220
## log(income)    0.373907485  0.2702078673
## log(density)   0.059680540  0.0633985069
## lawyes        0.004129197  0.0102484761
```

El coeficiente de la variable `male` para el método de Wallace-Hussain es positivo y considerablemente alto, lo cual no es razonable atendiendo al coeficiente obtenido por el modelo de Swamy-Arora y los coeficientes obtenidos en los modelos anteriores para esa variable. Esto es debido al valor del parámetro θ_μ . Consideraremos en adelante este modelo eliminando esa variable.

```
Guns.rwalhus2 <- plm(log(violent) ~ log(murder) + log(prisoners) + log(afam) +
                    log(population) + log(income)+ log(density) + law,
                    data = Guns , effect = "twoways" , model = "random" ,
                    random.method = "walhus")
```

6.5.4. Verificación y elección del modelo

Veamos si realmente es oportuno el empleo de un modelo de componente de error bidireccional. Debemos verificar si existen efectos temporales e individuales.

```
Guns.pooling2 <- plm(form , data = Guns, effect = "twoways", model = "pooling")
pFtest(Guns.within2, Guns.pooling2)

##
## F test for twoways effects
##
## data: log(violent) ~ log(murder) + log(prisoners) + log(afam) + +log(income) + ...
## F = 65.103, df1 = 69, df2 = 1095, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Se rechaza la hipótesis nula y por tanto existen efectos individuales y temporales. Por último, debemos determinar si seleccionar un modelo de efectos fijos o aleatorios. Tomamos el modelo de Swamy como representante de efectos aleatorios y realizamos el test de Hausman

```
phtest(Guns.within2,Guns.rswar2)

##
## Hausman Test
##
## data: log(violent) ~ log(murder) + log(prisoners) + log(afam) + +log(income) + ...
## chisq = 132.29, df = 5, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

Seleccionamos por tanto el modelo de efectos fijos.

6.6. Comparación de los modelos escogidos e interpretación de los resultados

Los modelos que hemos seleccionado finalmente son los de efectos fijos para ambos tipos de descomposición del error. Para ello nos hemos basado en el test de Hausman en ambas ocasiones, aunque esta elección también coincide basándonos en la naturaleza de los datos. El conjunto de individuos del panel en el estudio son todos los estados de EEUU, luego esos individuos son fijados de antemano y no son seleccionados aleatoriamente de una población mayor. En esos casos ya hemos mencionado que es más conveniente en principio un modelo de efectos fijos. Recordemos las variables y los coeficientes de los modelos.

```
coef(Guns.within2 , type = "level")

## log(murder) log(prisoners) log(afam) log(income) lawyes
## 0.17182147 -0.08599552 -0.37923624 0.21144407 -0.01159806
```

```
coef(Guns.within1 , type = "level")

##      log(murder) log(prisoners)      log(afam)      log(male)      log(income)
##      0.25872595      0.05495490     -0.20806877     -0.71895063      0.16899181
##           lawyes
##      0.02373469
```

En el modelo bidireccional contamos con las mismas variables que en el caso unidireccional, exceptuando la variable `male` pues recordemos que fue eliminada debido a la transformación dada por la matriz Q . Sin embargo, no debemos pensar que esto supone una pérdida de información con respecto al modelo unidireccional, pues el efecto de esta variable estaría recogido por la componente temporal del error, así como las variables que puedan haber sido omitidas en el estudio que cuenten también con un comportamiento constante para los estados en cada unidad de tiempo. Comparemos los ajustes de ambos modelos mediante la suma de los valores absolutos de los errores.

```
err <- rbind( uni = sum(abs(exp(predict(Guns.within1)) - Guns$violent)),
              bi = sum(abs(exp(predict(Guns.within2)) - Guns$violent)))
colnames(err) = "sum.errabs"
err

##      sum.errabs
## uni    63561.34
## bi     50468.28
```

El modelo de componente de error bidireccional se ajusta en general mejor a los datos, aunque podemos localizar estados para los cuales se obtienen resultados muy similares de la medida del error empleada, o incluso mejores para el caso unidireccional.

Pese a que hemos dicho que la variable `male` es constante para los distintos individuos del panel, el estado de Alaska se diferencia algo del resto con un mayor ratio de hombres en general. Analicemos las diferencias de los valores ajustados en ambos modelos para este estado.

```
time <- seq(1977, 1999)
vector_alaska <- paste("Alaska", time, sep="-")

errAlaska <- rbind(uni = sum(abs(exp(predict(Guns.within1)[vector_alaska]) -
                                Guns[Guns$state=="Alaska",3])),
                  bi = sum(abs(exp(predict(Guns.within2)[vector_alaska]) -
                                Guns[Guns$state=="Alaska",3])))
colnames(errAlaska) <- "sum.error"
```

```
errAlaska

##      sum.error
## uni  1424.204
## bi   1459.405
```

Vemos como la medida del error en el estado de Alaska es menor para el modelo unidireccional. Esto es debido a que el ratio de hombres en Alaska es considerablemente distinto en el instante inicial y en el instante final (tiempos en los que los valores ajustados van a tener menor error para el modelo unidireccional) y en el resto de las unidades de tiempo toma valores cada vez más cercanos al resto de estados.

Errores en el ajuste para los instantes iniciales y finales de Alaska:

```
errAlaskainit<-rbind(uni = abs(exp(predict(Guns.within1)[vector_alaska[1]])) -
                      Guns[Guns$state=="Alaska",3][1]),
                    bi = abs(exp(predict(Guns.within2)[vector_alaska[1]])) -
                      Guns[Guns$state=="Alaska",3][1]))

n = length(vector_alaska)
errAlaskaend <- rbind(uni = abs(exp(predict(Guns.within1)[vector_alaska[n]])) -
                      Guns[Guns$state=="Alaska",3][n]),
                    bi = abs(exp(predict(Guns.within2)[vector_alaska[n]])) -
                      Guns[Guns$state=="Alaska",3][n]))

errAlaska <- cbind(errAlaskainit , errAlaskaend)
colnames(errAlaska) <- c("sum.error.in" , "sum.error.end")
errAlaska

##      sum.error.in sum.error.end
## uni      57.43359      5.549054
## bi      73.25285      62.138995
```

Sin embargo, en la modelización de los datos panel nos interesa la extracción de conclusiones globales para el conjunto de individuos. Por tanto seleccionaremos finalmente el modelo de efectos fijos de componente de error bidireccional `Guns.within2`.

Los coeficientes del modelo son:

```
coef(Guns.within2)

##      log(murder) log(prisoners)      log(afam)      log(income)      lawyes
##      0.17182147  -0.08599552  -0.37923624    0.21144407   -0.01159806
```


Atendiendo a los coeficientes del modelo podemos extraer las siguientes conclusiones:

- **murder**: un aumento de un 1 % en esta variable supone según el modelo un incremento de un 0.17 % aproximadamente de la variable objetivo. Es decir, un aumento en la tasa de asesinatos causa un aumento de la tasa de delitos violentos.
- **prisoners**: un aumento de un 1 % en esta variable supone un descenso de un 0.09 % aproximadamente de **violent**. Por tanto, un aumento en la tasa de encarcelamientos provocaría (aunque no notablemente) un descenso de la tasa de delitos violentos en el año posterior.
- **afam**: un aumento de un 1 % de esta variable supone un descenso de un 0.38 % aproximadamente de **violent**. Luego según el modelo un aumento de la población afroamericana provoca un considerable descenso en la tasa de crímenes violentos. Este es un porcentaje muy alto, pero debemos tener en cuenta que es fiable en el contexto y periodo de estudio, pudiendo haber factores subyacentes que favorecen la reducción de la violencia en zonas con una mayor proporción de población afroamericana.
- **income**: un aumento de un 1 % en esta variable supondría un aumento de un 0.21 % aproximadamente en la variable objetivo. Es decir, las zonas con mayor ingreso tienen una mayor tasa de delitos violentos. Es razonable dado que EEUU es un país con una importante desigualdad económica, estableciéndose en los alrededores de las zonas más ricas suburbios donde la tasa de crímenes violentos es alta.
- **law**: según los resultados del modelo interpretamos que el gozar de una ley vigente sobre portabilidad de armas disminuye, aunque no excesivamente (un 0.01 % aproximadamente) la tasa de delitos violentos. Este dato es un argumento a favor de los que apoyan el control de las armas en EEUU, aunque no es del todo convincente pues el valor del coeficiente es muy cercano a 0, luego deberíamos hacer un estudio más profundo para ese debate.

6.7. Algunas consideraciones

En primer lugar, los p-valores para los contrastes de hipótesis se han obtenido suponiendo normalidad multivariante en los datos, caso para el cual los estadísticos de los tests tienen distribuciones conocidas. Sin suponer esa hipótesis se podrían obtener mediante el método bootstrap.

En segundo lugar, podríamos sospechar que en los modelos descritos en las secciones anteriores contamos con endogeneidad debido a la omisión de alguna variable. Por ejemplo, el nivel de educación, que está correlada con la variable **income**. Así podríamos aplicar (VI), buscando una variable que sea incorrelada con el error (entendamos que es la variable educación y otras similares), y sea correlada con **income**, como puede ser la media de herencias familiares.

Por último, también podríamos pensar en considerar un modelo con coeficientes heterogéneos, contando con coeficientes diferentes para cada estado, ya que el contraste de agrupabilidad de los datos se rechaza.

```
notpoolingmodel <- pvcm(formula = form ,
                        data = Guns, effect = "individual",
                        model = "within")
pooltest(Guns.within1 , notpoolingmodel )

##
## F statistic
##
## data:  log(violent) ~ log(murder) + log(prisoners) + log(afam) + log(male) + ...
## F = 9.9661, df1 = 402, df2 = 714, p-value < 2.2e-16
## alternative hypothesis: unstability
```

Sin embargo, la dimensión temporal no es excesivamente grande y en esos casos ya hemos mencionado que es más recomendable el uso del modelo agrupado, evitando el sobreajuste.

Bibliografía

- [1] Arlette Beltrán y Juan Francisco Castro, *Modelos de datos de panel y variables dependientes limitadas: teoría y práctica*. Universidad del Pacífico, 2020.
- [2] Badi H. Baltagi, *Econometric Analysis of Panel Data*. John Wiley & Sons Ltd, 2005.
- [3] Carsten Franz, *Multivariate Nonparametric Cramer-Test for the Two-Sample-Problem*. R package version 0.9-4, 2024. <https://CRAN.R-project.org/package=cramer>
- [4] Cheng Hsiao, *Analysis of Panel Data*. Cambridge University Press, 2014.
- [5] Christopher Dougherty, *Introduction to Econometrics*. Oxford University Press, 2007.
- [6] Croissant Y, Millo G, *Panel Data Econometrics in R: The plm Package*. Journal of Statistical Software, 2008. <https://doi.org/10.18637/jss.v027.i02>
- [7] Graybill, Franklin A, *An introduction to linear statistical models*. McGraw-Hill, 1961.
- [8] Hadley Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. <https://ggplot2.tidyverse.org>
- [9] Hans-Jürgen Andreß, Katrin Golsch y Alexander W. Schmidt, *Applied Panel Data Analysis for Economic and Social Surveys*. Springer, 2013.
- [10] Jeffrey M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*. MIT press, 2010.
- [11] Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*. South-Western College Publishing, 2012.
- [12] John R. Lott, Jr., *More Guns, Less Crime*. University of Chicago Press, 2000.
- [13] Kleiber C, Zeileis A, *Applied Econometrics with R*. Springer-Verlag, New York, 2008. <https://CRAN.R-project.org/package=AER>
- [14] Marno Verbeek, *Panel Methods for Finance*. De Gruyter, 2022.
- [15] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2023. <https://www.R-project.org/>

-
- [16] Russel Davidson y James G. MacKinnon, *Estimation and Inference in Econometrics*. John Wiley Sons Ltd, 1993.
 - [17] Sarkar D, *Multivariate Data Visualization with R*. Springer, New York, 2008.
<http://lmdvr.r-forge.r-project.org>
 - [18] Strang, Gilbert, *Lecture 3: Multiplication and inverse matrices*. MIT Open Course ware, 1999.
 - [19] Yves Croissant y Giovanni Millo, *Panel Data Econometrics with R*. John Wiley Sons Ltd, 2019.