Universidad de la República
Facultad de Ciencias Sociales
DEPARTAMENTO DE ECONOMIA

Notas Docentes

Modelos con Datos Panel

Adriana Cassoni

Nota Docente No. 04

I. INTRODUCCIÓN

Los modelos panel combinan observaciones de corte transversal y de series de tiempo. En ese sentido, incorporan más información que cada uno de los modelos anteriores por separado. Su uso se relaciona a distintas situaciones:

1. Es posible que no exista suficiente información de corte transversal ni de series temporales como para estimar los parámetros de interés utilizando alguno de estos modelos.

2. La teoría puede estar claramente presentada para microdatos (en vez de para datos agregados), pero es de interés analizar el comportamiento temporal de los mismos. Así, el modelo panel permite evitar los problemas de agregación y facilita el seguimiento de los individuos en el tiempo. Como ejemplo se puede pensar en la determinación del salario en función de los retornos a la educación y a la experiencia, así como a las características particulares de los individuos (ecuaciones tipo Mincer). Esta explicación del nivel de salarios se sustenta en la diferenciación entre individuos en un momento del tiempo. Sin embargo, puede resultar del mayor interés analizar si los retornos a la educación y la experiencia o los parámetros que denotan discriminación por sexo, edad, etc., se han modificado a lo largo de un cierto período.

3. No es posible diferenciar en el corte transversal, por ejemplo, lo que una variable está midiendo, con lo cual se hace necesario incorporar variabilidad temporal. Como ilustración es posible pensar en la determinación del salario en distintas empresas, en función de las condiciones tecnológicas ($\mathbf{X}$)y de la existencia o no de un sindicato (S):

$$w_i = \alpha + \beta \mathbf{X}_i + \gamma S_i + u_i$$

En el caso en que γ sea distinta de cero, no es posible afirmar con certeza que si existe un sindicato el salario en la empresa es mayor. Podría ocurrir que esto reflejase el hecho que la empresa contrata trabajadores con más habilidades que el resto, tal vez presionada por el incremento salarial que el sindicato tratará de impulsar, pero no necesariamente es consecuencia de que el sindicato altere las condiciones de movilidad, las compensaciones, etc. Sin embargo, si se tienen observaciones de individuos que se mueven de empresas con sindicatos a empresas sin sindicatos en el tiempo, es posible diferenciar ambos efectos.

4. En ocasiones los modelos, especialmente los dinámicos, presentan una alta colinealidad entre las variables explicativas, que es posible reducir o eliminar con el uso de información de corte transversal.

Desde el punto de vista teórico, los modelos panel pueden verse como un caso particular de los modelos multivariados. Cada variable en el modelo será la misma característica para un individuo distinto. Su forma final dependerá de las restricciones que se vayan imponiendo sobre los parámetros de las distintas ecuaciones.

La notación a usar será la siguiente:

$$y_{it} = \mathbf{X}'_{it}\beta + u_{it} \qquad i=1,2,\ldots,N\ ; \quad t=1,2,\ldots,T; \quad \mathbf{V} = var(\mathbf{y}/\mathbf{X}) \qquad (1)$$

donde i denota el individuo en el corte transversal (CT) y t la unidad de tiempo de la serie temporal (ST).

En general, se pueden diferenciar dos grandes casos:

a) Cuando existe un número reducido de individuos sobre los que se tienen muchas observaciones temporales, consecutivas o no.

b) Cuando existen pocas observaciones temporales para un gran número de individuos. En este caso, los modelos se han dado en llamar *micropaneles*.

En la primera situación, el contexto de análisis se asemeja más al de procesos estocásticos, observándose generalmente complejidades en términos de la matriz de varianzas y covarianzas que reflejan la heterogeneidad de las observaciones de CT.

El segundo caso, por el contrario, se inscribe en el marco de análisis de los modelos de CT, aunque en general deberá incorporar efectos temporales que den cuenta de las variciones en el tiempo.

Dado el modelo *(1)*, el vector de parámetros β puede ser una función de distinto tipo:

- $\beta(t)$: los parámetros cambian con el tiempo aunque son constantes entre individuos.
- $\beta(i)$: los parámetros son constantes en el tiempo pero distintos por individuo.
- β: los parámetros son constantes en el tiempo y por individuo.

Esto determinará la especificación de diferentes restricciones y, por lo tanto, dará lugar a distintos modelos a estimar. Es importante recordar que estas restricciones, si impuestas *a priori*, deben de ser sometidas a prueba de alguna manera, ya que de no ser correctas introducirán sesgo en los estimadores.

Así, si β es distinta por unidades del CT o de las ST, la estimación que se hará, en principio, será análoga a la de un modelo multivariado. Dependiendo de la forma de la matriz de varianzas y covarianzas, se usará el método de estimación correpondiente. La estimación conjunta de las N o T ecuaciones será eficiente respecto a la estimación ecuación por ecuación, cuando $\mathbf{V}$ sea una matriz no diagonal.

Por otra parte, si β es constante $\forall i,t$, se tiene un modelo con NT observaciones. El método de estimación entonces es Mínimos Cuadrados Ordinarios (*MCO*), sujeto a restricciones de igualdad entre los parámetros de las distintas ecuaciones. En este caso, al igual que en el anterior, habrá que analizar la estructura de la matriz de varianzas y covarianzas para definir la conveniencia de usar métodos de estimación alternativos.

En su forma general, la matriz de varianzas y covarianzas condicional a las variables explicativas es del tipo:

$$\mathbf{V}=\begin{bmatrix}\sigma_{11}\Omega_{11}\,.\,.\,.\,.\,.\,.\,.\,\sigma_{1N}\Omega_{1N}\\ \vdots \qquad\qquad\qquad \vdots\\ \sigma_{N1}\Omega_{N1}\,.\,.\,.\,.\,.\,.\quad \sigma_{NN}\Omega_{NN}\end{bmatrix}$$

con $\sigma_{ii}\Omega_{ii}= Var(\mathbf{y}_i/\mathbf{X}_i)$; $\mathbf{y}_i= (y_{i1},...,y_{iT})'$; $\sigma_{ij}\Omega_{ij}= Cov(\mathbf{y}_i\mathbf{y}_j/\mathbf{X}_i\mathbf{X}_j)$

Las matrices Ω_{ii} dan cuenta de la variabilidad de cada individuo en el tiempo, es decir, de los posibles problemas de autocorrelación para cada unidad en el corte transversal, así como de la existencia de heteroscedasticidad o varianza variable en el tiempo. Las matrices Ω_{ij} por su parte, reproducen la estructura de dependencia entre el individuo i y el j en cada momento del tiempo.

Como ejemplo es posible pensar en un panel compuesto por 5 países en el correr de dos décadas. El modelo se inscribe más en el contexto de procesos estocásticos que en el de CT. Se pretende analizar la inversión extranjera directa en función de las características de cada mercado y de factores de oferta comunes a todos los países pero que cambian en el tiempo. El modelo es del tipo:

$$\mathrm{ID}_{it}= \mathbf{X}'_{it}\beta + \mathbf{W}'_t\alpha + u_{it}$$

Según los supuestos que se hagan sobre los distintos coeficientes, es posible derivar formas finales diferentes y, por lo tanto, estimar los parámetros por el método que sea adecuado.

Así, si β y α son constantes para todos los países y a lo largo de ambas décadas, y además la variabilidad de la inversión extranjera directa por país condicional a $\mathbf{X}$ y $\mathbf{W}$ es constante y no existe relación entre los diversos países en cada momento del tiempo, se podrá estimar por MCO restringidos, con 100 observaciones, ya que la matriz de varianzas y covarianzas es diagonal, con elementos idénticos en la diagonal. Si la restricción es adecuada, el modelo estimado por MCO restringidos arrojará estimadores más eficientes de los parámetros que los que se obtendrían de la estimación ecuación por ecuación.

Si, por el contrario, se piensa que en cada país el monto de la inversión dadas las variables explicativas es un proceso autorregresivo o bien que la variabilidad de la inversión cambia en el tiempo, la matriz de varianzas y covarianzas será diagonal en bloques. El método de estimación adecuado es Mínimos Cuadrados Generalizados (MCG).

Es posible que solamente algunos parámetros sean constantes y otros no. Podría pensarse que los factores de oferta tienen un impacto diferenciado por país, de forma que γ es una función de i. En este caso, la inclusión de variables binarias y la estimación por MCO o MCG según corresponda, daría estimadores adecuados.

Otra posibilidad es que, siendo α y β constantes, existan características individuales y/o temporales no observables que deben ser consideradas en la especificación. Esto daría lugar a ciertos modelos especiales denominados de *efectos individuales fijos o aleatorios*.

Finalmente, cabría pensar que en realidad los parámetros α y β son distintos para los diversos países o en cada momento del tiempo. Sin embargo, la existencia de correlación en el CT o entre los distintos años determina que la estimación conjunta del sistema multivariado redunde en ganancia de eficiencia, de forma que lo más adecuado sea obtener estimadores a través del método de Zellner para regresiones aparentemente no relacionadas *(SURE)*.

De lo anterior resulta claro que según los supuestos que se estén manejando en la especificación del modelo, será necesario imponer diferentes restricciones que habrá que analizar siempre que sea posible, para evitar la introducción de sesgo en la estimación. Además, el conjunto de supuestos determinará el método de estimación a utilizar que estará, a su vez, fuertemente relacionado a los distintos problemas que es probable encontrar en cada caso.

Hay dos tipos de situaciones muy frecuentes en los modelos panel, que es importante analizar con precaución:

- La existencia de sesgo en los estimadores por heterogeneidad de las unidades del CT o la ST, que proviene del hecho que **y** no es idénticamente distribuida - $\beta(i)$ o $\beta(t)$.

- La existencia de sesgo de selección en los estimadores debido a que la muestra que se utiliza no es aleatoria, en tanto que se eliminan individuos con determinadas características.

Ambos problemas, por demás comunes, resultan en la obtención de estimadores que son en realidad un promedio de varios, en el primer caso, o una sub o sobre estimación de los verdaderos parámetros en el segundo. Las gráficas que siguen, tomadas de Hsiao (1986), ilustran diversas situaciones de este tipo.

Sea el modelo: $y_{it} = \alpha + \mathbf{X}'_{it}\beta + u_{it}$

Las primeras tres gráficas muestran cómo es posible, debido a heterogeneidad en el corte transversal (β constante pero α variable con i), obtener estimadores de α y β que son sesgados, no siendo posible, además, predecir el signo del sesgo.

Las gráficas 4 y 5, por su parte, muestran el caso en que tanto α como β son variables con los individuos, de forma tal que la recta de regresión puede resultar en una aproximación cualquiera (en el primer caso) o bien confundirse el problema con uno de forma funcional (en la segunda ilustración).

Finalmente, la gráfica 6 muestra una posible estimación que adolece de sesgo de selección. Si en la muestra sólo se consideran los individuos cuyos ingresos (variable en el eje vertical) están por debajo de L, la recta de regresión estimada tendrá una ordenada al origen menor (se tendrá un $\hat{\alpha}$ con sesgo positivo) y una pendiente mayor ($\hat{\beta}$ con sesgo negativo).

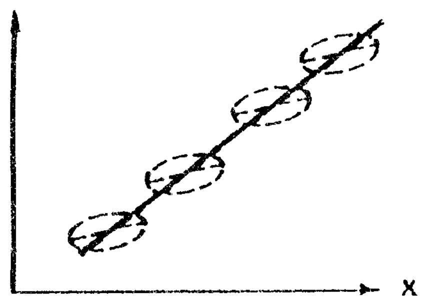

Fig. 1

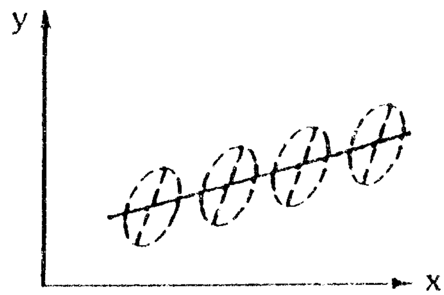

Fig. 2

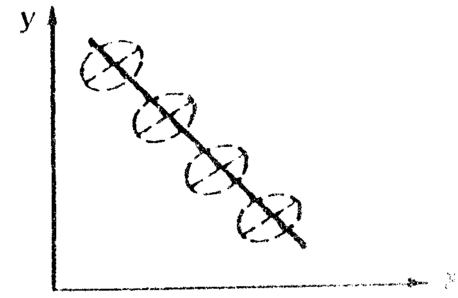

Fig. 3

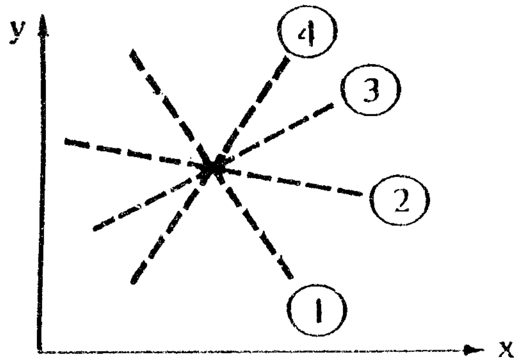

Fig. 4

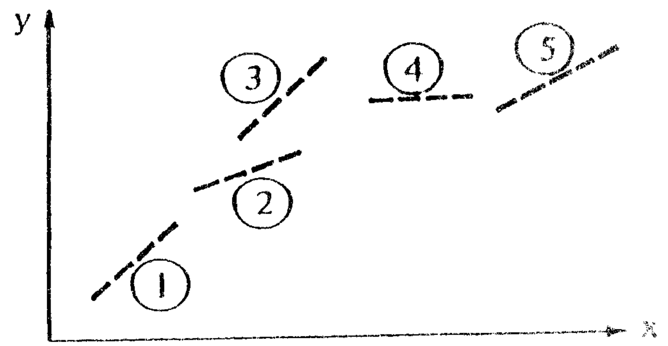

Fig. 5

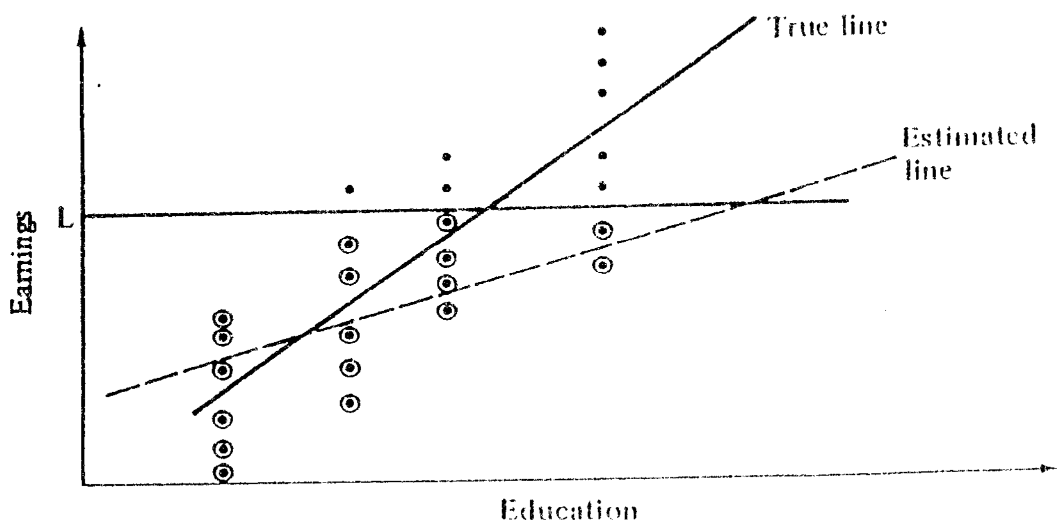

Fig. 6

II. Modelos Estaticos

En el análisis de los modelos estáticos es posible diferenciar dos casos, como se mencionó en la Introducción. El primero es aquel cuya estructura es similar a la de los procesos estocásticos debido a que se tiene un número suficiente de observaciones temporales. Así, es posible analizarlo en el marco de un modelo multivariado con la peculiaridad de que cada variable es la misma característica para un individuo distinto. En este sentido, el tratamiento que se le dé, tanto en términos de estimación como de restricciones, estará asociado estrechamente a la estructura de la matriz de varianzas y covarianzas. Es probable que existan, además, problemas relacionados a la heterogeneidad entre individuos, con lo cual será necesario incorporar efectos individuales. Esto es, a su vez, la mayor complejidad que presentan los modelos del segundo tipo mencionado, es decir, los micropaneles. Su tratamiento, en términos de estimación, implicará algunas modificaciones respecto al primer caso.

Tomando el modelo *(1)* como referencia y la matriz **V** en su forma general, se pueden identificar varios casos de interés:

1. $Var(y_{it}/\mathbf{X}) = \sigma^2 \quad \forall i, t \qquad Cov(y_{it} y_{js}/\mathbf{X}) = 0 \ \forall i \neq j, \ \forall t \neq s$

Con ello, la matriz **V** resulta ser de la forma: $\mathbf{V} = \sigma^2 \begin{bmatrix} I_T & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & I_T \end{bmatrix}$

Así, si los parámetros β son constantes en el CT y en la ST, se tendrá un modelo a estimar por MCO con NT observaciones. En caso que los parámetros varíen por individuos o en el tiempo, se podrá estimar sujeto a la restricción de que la matriz de varianzas y covarianzas tenga la forma descrita arriba, es decir, con una varianza común, agregando así eficiencia a la estimación. Esto equivaldría a una estimación con variables binarias que permitan la variación de los βs.

2. $Var(y_{it}/\mathbf{X}) = \sigma_i^2 \qquad Cov(y_{it} y_{js}/\mathbf{X}) = 0 \ \forall i \neq j, \ \forall t \neq s$

La varianza no es constante en el corte transversal, varía por individuo o bien es heteroscedástica[1]. En este caso la forma que toma la matriz $\mathbf{V}$ es la siguiente:

$$\mathbf{V}=\begin{bmatrix}\sigma_1^2 I_T & & 0\\ & \ddots & \\ 0 & & \sigma_N^2 I_T\end{bmatrix}$$

El tener varianzas variables por individuo puede deberse tanto a que los procesos son intrínsecamente diferentes como a razones relacionadas a las magnitudes de las variables, que pueden diferir de acuerdo a ciertas características como tamaño, o aún a otras no observables.

El método de estimación adecuado en este caso es MCG, para lo cual será necesario estimar las N varianzas en una primera etapa a partir de los residuos mínimo-cuadráticos, y luego transformar el modelo de forma que:

$$\hat{\sigma}_i^2 = \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/(T-k) \qquad \text{con } \hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i'\hat{\beta}_{MCO} \qquad \hat{\beta}_{MCO} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Obtenidas $\hat{\sigma}_i^2$ para $i = 1,\ldots,N$ se construye $\hat{\mathbf{V}}$ y se obtiene $\hat{\beta}_{MCG}$ de acuerdo a:

$$\hat{\beta}_{MCG} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$$

Como siempre, es posible realizar sucesivas iteraciones de estas dos etapas para aumentar la precisión de los estimadores.

Si el número de observaciones es suficiente, resulta del mayor interés someter a prueba la existencia de varianzas variables por individuos (o en el tiempo). Para ello es posible utilizar pruebas bajo distintos principios, siendo la hipótesis nula que las varianzas son iguales para todos los individuos.

Sea s_i^2 el estimador de la varianza para el individuo i: $s_i^2 = \hat{\mathbf{u}}_i'\hat{\mathbf{u}}_i/T$ y sea $s^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/NT$ el estimador bajo la hipótesis de varianza constante.

---

[1] Lo mismo aplica al caso en que la variabilidad se da en la ST.

El estadístico bajo el principio de multiplicadores de Lagrange para realizar la prueba es:

$$T/2\Sigma_i(s_i^2/s^2-1)^2 \sim \chi_N^2$$

o bien, cuando la variabilidad es en el tiempo:

$$N/2\Sigma_i(s_i^2/s^2-1)^2 \sim \chi_T^2$$

Bajo el principio de razón de verosimilitud, el estadístico es:

$$-2ln\lambda = NT ln\hat{s}^2 - T\Sigma_i ln\hat{s}_i^2 \sim \chi_{N-1}^2$$

En el caso en que se desee someter a prueba la existencia de heteroscedasticidad, es posible utilizar la prueba de White (1980), estimando los parámetros de la función que define a la varianza:

$$\hat{\sigma}_i^2 = \alpha_0 + \Sigma_i\alpha_i\mathbf{X}_i + \Sigma_i\Sigma_j\gamma_i\mathbf{X}_i\mathbf{X}_j$$

Este estimador es útil además para el caso en que no es posible suponer normalidad.

Finalmente, la prueba bajo el principio de Wald se realiza utilizando el estadístico:

$$(T/2)\Sigma_i(s_i^2 - s^2)^2/\hat{V}_i \overset{a}{\sim} \chi_N^2 \quad \text{con} \quad \hat{V}_i = [1/[T(T-1)]]\Sigma_t(\hat{u}_{it}^2 - s_i^2)^2$$

3. $Var(y_{it}/\mathbf{X}) = \sigma_i^2 \quad Cov(y_{it}y_{jt}/\mathbf{X}) \neq 0 \ \forall i\neq j, \ Cov(y_{it}y_{is}/\mathbf{X}) = 0 \ \forall t\neq s$

La covarianza no es nula en el corte transversal, pudiendo ser la varianza variable o no por individuo (se supone el caso general en que efectivamente varía con *i*). En el tiempo, sin embargo, se mantiene el supuesto de no autocorrelación. La matriz de varianzas y covarianzas resultante es del tipo:

$$\mathbf{V}=\begin{bmatrix}\sigma_1^2 I_T & \dots\dots\dots & \sigma_{1N} I_T\\ \vdots & & \vdots\\ \sigma_{N1} I_T & \dots\dots\dots & \sigma_N^2 I_T\end{bmatrix}$$

Este caso se aplica a las situaciones en que existen factores no observables que afectan a los individuos de manera diferente, de modo que lo que ocurre con uno de ellos no es independiente de lo que sucede con los demás.

La forma de estimación es, al igual que antes, a través del uso de MCG, sólo que la estructura de la matriz de varianzas y covarianzas ha cambiado, haciendo necesaria la estimación de nuevos parámetros: obtenido $\hat{\beta}_{MCO}$ se calcula $s_{ij} = (1/T)[\mathbf{y}_i'\mathbf{y}_j + \beta\mathbf{X}_i'\mathbf{X}_j\beta' - \beta\mathbf{X}_i'\mathbf{y}_j - \beta\mathbf{X}_j'\mathbf{y}_i]$ y con éste $\hat{\mathbf{V}}$ y el estimador $\hat{\beta}_{MCG}$. Nuevamente, es posible iterar para obtener estimadores más eficientes.

Para someter a prueba la existencia de correlación en el CT se utiliza un estadístico de razón de verosimilitud, siendo la hipótesis nula que las covarianzas son cero:

$$\lambda = T(\Sigma_i ln\hat{\sigma}_i^2 - ln|\hat{\Sigma}|) \sim \chi^2_M$$

con $\hat{\Sigma} = \begin{bmatrix}\hat{\sigma}_1^2 & \dots\dots & \hat{\sigma}_{1N}\\ \vdots & \ddots & \vdots\\ \hat{\sigma}_{N1} & \dots & \hat{\sigma}_N^2\end{bmatrix}$ ya que $\mathbf{V} = \Sigma \otimes I_T$ y $M = N(N-1)/2$

Como alternativa se puede utilizar el estadístico de prueba propuesto por Breusch y Pagan (1979):

$$T\Sigma_{i=2}^{M}\Sigma_{j=1}^{i-1} r_{ij}^2 \sim \chi^2_M \quad \text{con} \quad r_{ij}^2 = \hat{\sigma}_{ij}/(\hat{\sigma}_i^2\hat{\sigma}_j^2)^{1/2}$$

Para realizar la prueba es necesario obtener los estimadores de r_{ij}. Estos pueden calcularse en dos etapas o iterando para mejorar la eficiencia.

4. $Var(y_{it}/\mathbf{X}) = \sigma_i^2 \quad Cov(y_{it}y_{jt}/\mathbf{X}) = 0 \ \forall i \neq j, \ Cov(y_{it}y_{is}/\mathbf{X}) \neq 0 \ \forall t \neq s$

El modelo presenta autocorrelación para las distintas unidades de CT. Para simplificar se puede suponer que no hay correlación entre individuos. De esta forma, la matriz de varianzas y covarianzas es:

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2\Omega_{11} & 0.......0 \\ \vdots & \vdots \\ 0 .... & \sigma_N^2\Omega_{NN} \end{bmatrix}$$

En este caso corresponde estimar por MCG modelando la autocorrelación primeramente. El coeficiente de autocorrelación se estima por alguno de los métodos tradicionales (Durbin, Cochrane-Orcutt, etc.). En el caso de procesos AR(1), la transformación requerida será:

$$\mathbf{y}_i^* = \begin{bmatrix} [1-r_i^2]^{1/2} y_{i1} \\ y_{i2} - y_{i1} r_i \\ \vdots \\ y_{iT} - y_{iT-1} r_i \end{bmatrix}$$ y análogamente para las variables en **X**.

Obtenida la transformación anterior, el modelo es del tipo especificado en el caso *2*, es decir, con una matriz **V** diagonal, con elementos distintos cada *T* filas, aquellos que dan cuenta de las diferentes varianzas individuales. En caso que además existiese correlación entre individuos, el modelo transformado será del tipo *3*, con $\mathbf{V} = \Sigma \otimes I_T$.

Modelos de Efectos Individuales

Cuando la heterogeneidad entre los individuos o entre las observaciones temporales es importante y, además, no es posible modelarlas de forma adecuada, se suelen utilizar los modelos de efectos individuales. Estos darían cuenta de características de los individuos que son, en principio, no observables y, por tanto, no modelables, no medibles a través de variables explicativas. En general el tratamiento del tema en la literatura se concentra en dos tipos de modelos: *modelos de efectos fijos* y *modelos de efectos aleatorios*. La distinción se hizo originalmente pensando en que las características individuales podían tener uno de estos dos comportamientos, es decir, ser constantes distintas por individuo o bien poseer una distribución de probabilidad alrededor de una media común. La interpretación teórica es, sin embargo, poco clara. Desde el punto de vista estadístico, por su parte, resulta más convincente pensar que todos los efectos son aleatorios pero que, si el condicionamiento se realiza sobre la muestra de individuos con que se cuenta es posible considerarlos como fijos mientras que si se condiciona sobre la población sólo es posible considerarlos como efectos aleatorios.

La especificación de los modelos y su estimación se detallan a continuación.

A. *Modelos de Efectos Fijos:* $y_{it} = \alpha_i + \mathbf{X}'_{it}\beta + u_{it}$

Si el número de observaciones totales es tal que es posible estimar los parámetros α_i conjuntamente con los βs (en general cuando el CT posee un número no demasiado grande de unidades), el modelo se especifica con N variables binarias y se estima por MCO, MCG o el método que corresponda de acuerdo a otras especificidades. Los $\hat{\alpha}_i$ determinarán que cada individuo posee una ordenada al origen distinta, debido a características propias, no observables. En los micropaneles, sin embargo, en general no resultará eficiente estimar conjuntamente todos los parámetros, con lo cual es usual transformar el modelo para obtener un buen estimador de β y luego, en caso de ser de interés, calcular los estimadores de los α_is. Esta transformación es la siguiente:

$$y^*_{it} = y_{it} - \bar{y}_{i.} = y_{it} - (1/T)\sum_t y_{it} \qquad \mathbf{X}^*_{it} = \mathbf{X}_{it} - \bar{\mathbf{X}}_{i.} = \mathbf{X}_{it} - (1/T)\sum_t \mathbf{X}_{it}$$

Así, el modelo transformado es:

$$y^*_{it} = \mathbf{X}^{*\prime}_{it}\beta + u^*_{it}$$

Los efectos individuales desaparecen, al ser constantes en el tiempo y β se estima de la forma habitual. Estimado el parámetro β, los $\hat{\alpha}_i$ se calculan de acuerdo a:

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{\mathbf{X}}'_{i.}\hat{\beta}$$

La varianza de los estimadores de β es la usual. En el caso en que la estimación se realice por MCO, ésta es:

$$Var(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1} \qquad \text{con} \qquad \hat{\sigma}^2 = 1/(NT-K-N)\Sigma_i\Sigma_t(y_{it} - \hat{\alpha}_i - \mathbf{X}'_{it}\hat{\beta})^2$$

La varianza de los α_is se calcula como: $Var(\hat{\alpha}_i) = \hat{\sigma}^2/T + \bar{\mathbf{X}}'_{i.}V(\hat{\beta})\bar{\mathbf{X}}_{i.}$

Los estimadores son insesgados y eficientes. Además, $\hat{\beta}$ es consistente cuando N o T o ambos tienden a infinito, mientras que $\hat{\alpha}_i$ son consistentes sólo cuando T tiende a infinito.

El desarrollo anterior se puede reproducir de manera análoga en el caso en que los efectos no observables se asocien al tiempo en vez de a los individuos. Además, la extensión al caso en que haya efectos individuales y temporales es también inmediata:

$$y_{it} = \alpha_i + \lambda_t + \mathbf{X}'_{it}\beta + u_{it}$$

La transformación adecuada es ahora:

$$y^*_{it} = y_{it} - \bar{y}_{i.} - (\bar{y}_{.t} - \bar{y}) \qquad \mathbf{X}^*_{it} = \mathbf{X}_{it} - \bar{\mathbf{X}}_{i.} - (\bar{\mathbf{X}}_{.t} - \bar{\mathbf{X}})$$

donde: $\bar{y}_{i.} = (1/T)\Sigma_t y_{it}$ $\qquad \bar{y}_{.t} = (1/N)\Sigma_i y_{it}$ $\qquad \bar{y} = (1/NT)\Sigma_t\Sigma_i y_{it}$

$$\mathbf{X}_{i.} = (1/T)\Sigma_t\mathbf{X}_{it} \qquad \bar{\mathbf{X}}_{.t} = (1/N)\Sigma_i\mathbf{X}_{it} \qquad \bar{\mathbf{X}} = (1/NT)\Sigma_t\Sigma_i\mathbf{X}_{it}$$

Los efectos individuales y temporales se calculan como:

$$\hat{\alpha}_i = (\bar{y}_{i.} - \bar{y}) - (\bar{\mathbf{X}}_{i.} - \bar{\mathbf{X}})'\hat{\beta} \qquad \hat{\lambda}_t = (\bar{y}_{.t} - \bar{y}) - (\bar{\mathbf{X}}_{.t} - \bar{\mathbf{X}})'\hat{\beta}$$

Resulta del mayor interés poder someter a prueba la existencia de los efectos individuales. Para ello, es necesario plantear el modelo sin restricciones y aquellos que resultarían de imponer sucesivas condiciones, para luego comparar el desempeño de las diferentes especificaciones a través de estadísticos de Wald. A este tipo de

comparación se le conoce generalmente como *análisis de covarianza*.
Sea el modelo sin restricciones: $y_{it} = \alpha_i + \mathbf{X}'_{it}\beta_i + u_{it}$ *(2)*

con $\mathbf{X}$ una matriz de $T\times K$, $i = 1,\ldots,N$ y $t = 1,\ldots,T$

Se plantean 3 hipótesis[2]:

- β es constante para todos los individuos, α es variable:

$$y_{it} = \alpha_i + \mathbf{X}'_{it}\beta + u_{it} \qquad (3)$$

- β es variable y α es constante entre individuos:

$$y_{it} = \alpha + \mathbf{X}'_{it}\beta_i + u_{it} \qquad (4)$$

- β y α son constantes entre individuos:

$$y_{it} = \alpha + \mathbf{X}'_{it}\beta + u_{it} \qquad (5)$$

Suponiendo una matriz de varianzas y covarianzas del tipo $\sigma^2 I_{NT}$, los estimadores por MCO de β en el modelo *(2)* son:

$$\hat{\beta}_i = (\mathbf{X}^{*\prime}_i\mathbf{X}^*_i)^{-1}\mathbf{X}^{*\prime}_i\mathbf{y}^*_i \qquad \hat{\alpha}_i = \bar{y}_{i.} - \bar{x}'_{i.}\hat{\beta}_i$$

El número de parámetros a estimar es $N(K+1)$ y se supone que $T>K+1$.
En general, a estos estimadores se les conoce como *intra-grupos*.

En el caso *(3)* las fórmulas para $\hat{\alpha}$ y $\hat{\beta}$ son las manejadas en parráfos anteriores, es decir:

$$\hat{\beta} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^* \qquad \hat{\alpha} = \bar{y}_{i.} - \bar{x}'_{i.}\hat{\beta}$$

El número de parámetros a estimar es $N+K$ y se supone que $NT>N+K$.

En el caso *(4)* los estimadores de las pendientes son iguales al caso *(2)* mientras que el estimador de la ordenada al origen es:

[2] El análisis se realiza de forma análoga para variaciones temporales.

$$\hat{\alpha} = \bar{y} - (1/N)\Sigma_i \bar{x}'_{i.} \hat{\beta}_i$$

El número de parámetros a estimar es $NK+1$ y se asume que $NT>NK+1$.

Finalmente, en el modelo *(5)* los estimadores se obtienen a través de:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad \hat{\alpha} = \bar{y} - \bar{x}'\hat{\beta}$$

El total de parámetros a estimar es $K+1$ y el supuesto es $NT>K+1$.

Si u_{it} se distribuye *NIID* sobre i y sobre t, es posible buscar evidencia que sustente o rechace los diferentes modelos a través de pruebas tipo F. Así, se pueden plantear las siguientes hipótesis:

1. H_0: *modelo 5* *versus* H_1: *modelo 2* es decir,

H_0: $\beta_1 = \beta_2 = \ldots \beta_N = \beta$ y $\alpha_1 = \alpha_2 = \ldots \alpha_N = \alpha$ *versus*

H_1: $\alpha_i \neq \alpha$ o $\beta_i \neq \beta$ *para algún* i

El estadístico es: $F = [(\hat{\mathbf{u}}'_5\hat{\mathbf{u}}_5 - \hat{\mathbf{u}}'_2\hat{\mathbf{u}}_2)/(l_5 - l_2)]/[\hat{\mathbf{u}}'_2\hat{\mathbf{u}}_2/l_2]$

con:

$l_2 = NT-N(K+1)$ y $l_5 = NT - (K+1) \Rightarrow l_5 - l_2 = (N-1)(K+1)$

$\hat{\mathbf{u}}_j$: los residuos de la regresión de acuerdo al modelo j.

Si se rechaza H_0 será necesario detectar la fuente principal de variación, con lo cual interesará comparar el modelo *(2)* con los modelos *(3)* y *(4)*.

2. H_0: *modelo 3* *versus* H_1: *modelo 2* es decir,

H_0: $\beta_1 = \beta_2 = \ldots \beta_N = \beta$ *versus* H_1: $\beta_i \neq \beta$ *para algún* i

El estadístico es: $F = [(\hat{\mathbf{u}}'_3\hat{\mathbf{u}}_3 - \hat{\mathbf{u}}'_2\hat{\mathbf{u}}_2)/(l_3 - l_2)]/[\hat{\mathbf{u}}'_2\hat{\mathbf{u}}_2/l_2]$

con $l_3 = NT-(N+K) \Rightarrow l_3 - l_2 = (N-1)K$

En caso de rechazar la hipótesis nula, el proceso a veces se interrumpe, debido a que en general es difícil interpretar el modelo *4* desde un punto de vista teórico. Sin embargo, podría igualmente interesar verificar su validez estadística. En este caso se procede a la prueba siguiente.

3. H_0: modelo 4 versus H_1: modelo 2 es decir,

$H_0: \alpha_1 = \alpha_2 = \ldots \alpha_N = \alpha$ *versus* $H_1: \alpha_i \neq \alpha$ *para algún* i

El estadístico es: $F = [(\hat{u}'_4\hat{u}_4 - \hat{u}'_2\hat{u}_2)/(l_4 - l_2)]/[\hat{u}'_2\hat{u}_2/l_2]$

con $l_4 = NT-(NK+1) \Rightarrow l_4 - l_2 = N-1$

Si también aquí se rechaza H_0, se deberá optar por el modelo *(2)*.

Si, por el contrario, no se rechaza la hipótesis nula en alguno de los dos casos anteriores, se procede a plantear hipótesis condicionadas. Así, la siguiente prueba sería analizar si hay evidencia, dado que no se rechaza la hipótesis de pendientes comunes (modelo 3), para pensar en ordenadas al origen comunes.

4. H_0: modelo 5 versus H_1: modelo 3

$H_0: \alpha_1 = \alpha_2 = \ldots \alpha_N = \alpha$ *dado* $\beta_1 = \beta_2 = \ldots \beta_N = \beta$ *versus*

$H_1: \alpha_i \neq \alpha$ *para algún* i *dado* $\beta_1 = \beta_2 = \ldots \beta_N = \beta$

El estadístico es: $F = [(\hat{u}'_5\hat{u}_5 - \hat{u}'_3\hat{u}_3)/(l_5 - l_3)]/[\hat{u}'_3\hat{u}_3/l_3]$

con $l_5 - l_3 = N-1$

Se procedería análogamente para comparar los modelos *5* y *4*. De igual forma, si la variabilidad a analizar se concentrase en los períodos de tiempo, los estadísticos anteriores serían muy similares, modificándose los grados de libertad.

Es importante señalar que las pruebas anteriores no son independientes, por lo cual es posible obtener resultados inconsistentes. Por ejemplo, podría suceder que en el caso *1* se rechace H_0 y, al mismo tiempo, en el caso *4* no se rechace H_0. Esto es posible porque la hipótesis alternativa es distinta en cada caso (se somete a prueba el modelo *5* contra el *2* y el *3*, respectivamente).

B. Modelos de efectos aleatorios: $y_{it} = \alpha + \mathbf{X}'_{it}\beta + v_{it}$; $v_{it} = u_{it} + w_i + e_t$

Los distintos componentes de la perturbación estarían dando cuenta de la parte no modelable de y_{it} en diferentes dimensiones, incluyendo así un elemento que da cuenta de variación entre unidades pero constante en el tiempo (w_i) y otro elemento que incorpora variación temporal y no individual (e_t). Si además es posible suponer incorrelación entre éstos y las variables *Xs*, entonces la varianza condicional de y_{it} será igual a:

$$\sigma^2_y = \sigma^2_u + \sigma^2_w + \sigma^2_e$$

Por ello, estos modelos se conocen también como *modelos con error o varianza compuestos*. Otros supuestos implícitos en el modelo son:

$E(u_{it}) = E(w_i) = E(e_t) = 0$: la media de los efectos es cero.

$E(w_i u_{it}) = E(w_i e_t) = E(e_t u_{it}) = 0$: los distintos componentes no están correlacionados entre si.

$E(w_i w_j) = E(e_t e_s) = E(u_{it} u_{js}) = 0$ *si* $i \neq j$, $t \neq s$: cada uno de los componentes son incorrelacionados.

De esta forma, la matriz de varianzas y covarianzas ya no será diagonal y tendrá una estructura específica:

$$\mathbf{V}=\begin{bmatrix} \Omega & \sigma_e^2 I_T & \cdots & \sigma_e^2 I_T \\ \vdots & & & \vdots \\ \sigma_e^2 I_T & \cdots & & \Omega \end{bmatrix} \quad \text{con} \quad \Omega=\begin{bmatrix} (\sigma_u^2+\sigma_e^2+\sigma_w^2) & \sigma_w^2 & \cdots & \sigma_w^2 \\ \sigma_w^2 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_w^2 & \cdots & & (\sigma_u^2+\sigma_e^2+\sigma_w^2) \end{bmatrix}$$

Como se observa, la existencia de efectos individuales aleatorios se refleja en autocorrelación (temporal) de los residuos, mientras que los efectos temporales aleatorios generan correlación entre individuos. De esta forma, el método de estimación adecuado, para obtener estimadores eficientes, será MCG. La aplicación de MCO a los desvíos respecto a la media individual (temporal), si bien permitirá obtener estimadores de β insesgados y consistentes cuando N o T tienden a infinito, resultará ineficiente.

En el caso en que sólo haya efectos individuales aleatorios ($e_t = 0\ \forall t$) la transformación adecuada de las variables (equivalente a la aplicación de MCG) que permitirá obtener estimadores de β eficientes es como sigue:

$$y_i^*=\Omega^{-1/2}y_i=\begin{bmatrix} y_{i1}-\theta\bar{y}_{i.} \\ \vdots \\ y_{iT}-\theta\bar{y}_{i.} \end{bmatrix} \qquad \text{con } \theta = 1 - \sigma_u/[T\sigma_w^2+\sigma_u^2]^{1/2}$$

$$\Omega^{-1/2}= I-(1/T)\theta\iota\iota'$$

La misma transformación se realiza para cada variable X. Cuando θ se aproxima a 1, el estimador de MCG se aproxima al obtenido con el modelo de efectos fijos. Este caso se obtiene cuando T tiende a infinito. Por otro lado, si θ se acerca a cero, la varianza de los efectos individuales es muy pequeña y el estimador es equivalente al que resulta de aplicar MCO, es decir, el problema de eficiencia es menor.

Un uso de este modelo que resulta de particular interés es la diferenciación de la variabilidad total entre aquella originada en la variación al interior de cada grupo de la que proviene de la variación entre los grupos, suponiendo nuevamente que no hay efectos temporales. Se considera como un "grupo" al conjunto de observaciones temporales sobre un individuo. Así, la variabilidad al interior del grupo da cuenta de las desviaciones de los y_{it} con respecto a la media temporal $\bar{y}_{i.}$. Por su parte, la variabilidad entre los grupos se mide eliminando la variación temporal para cada individuo y comparando entonces las medias temporales entre sí. La idea anterior resulta más clara si se piensa en un conjunto de datos panel en el que, en lugar de individuos en el tiempo, se tiene, por ejemplo, N familias gastando en T bienes. La variabilidad intra-grupos se mediría aquí como los desvíos del gasto de las familias en los diferentes bienes respecto a su gasto medio en un bien tipo, mientras que los desvíos del gasto medio de cada familia respecto a la media de la población como un todo daría cuenta de la variabilidad entre-grupos.

Para realizar la distinción anterior, conviene expresar el modelo de tres formas distintas:

$$(i)\ y_{it} = \alpha + \mathbf{X}'_{it}\beta + (u_{it} + w_i)$$

$$(ii)\ y_{it} - \bar{y}_{i.} = (\mathbf{X}_{it} - \bar{\mathbf{X}}_{i.})'\beta + (u_{it} - \bar{u}_{i.})$$

$$(iii)\ \bar{y}_{i.} = \alpha + \bar{\mathbf{X}}'_{i.}\beta + (\bar{u}_{i.} + w_i)$$

El modelo (i) permite estimar β alrededor de las medias totales, considerando variación temporal e individual ($\hat{\beta}$). En el caso (ii), por su parte, el estimador se calcula alrededor de los desvíos respecto a las medias grupales ($\hat{\beta}^{i}$), mientras que en (iii) se obtiene a partir de la variación de las medias individuales respecto a las totales ($\hat{\beta}^{e}$). El segundo estimador, equivalente al estimador del modelo de efectos fijos, es el estimador *intra-grupos*, mientras que el último es el estimador *entre-grupos*. De esta forma, $\hat{\beta}$ es un promedio ponderado de los restantes dos:

$$\hat{\beta} = F^{i}\hat{\beta}^{i} + F^{e}\hat{\beta}^{e} \quad \text{con } F^{i} = [S^{i} + S^{e}]^{-1}S^{i} \quad \text{y} \quad F^{i} = I - F^{e}$$

siendo S la suma del cuadrado de los residuos de la regresión correspondiente, de forma que el ponderador es el porcentaje de la variación total que se debe a la variabilidad intra-grupos.

Finalmente, los estimadores de los distintos componentes de la varianza se obtienen a partir de:

$$\hat{\sigma}_u^2 = (1/[N(T-1)-K])\Sigma_i\Sigma_t[(y_{it} - \bar{y}_{i.})(X_{it} - \bar{X}_{i.})'\hat{\beta}^i]^2$$

$$\hat{\sigma}_w^2 = (1/[N-(K+1)])\Sigma_i(\bar{y}_{i.} - \bar{X}'_{i.}\hat{\beta}^e)^2 - (1/T)\hat{\sigma}_u^2$$

$$\hat{\sigma}_e^2 = (1/[T-(K+1)])\Sigma_t(\bar{y}_{.t} - \bar{X}'_{.t}\hat{\beta}^m)^2 - (1/N)\hat{\sigma}_u^2$$

donde $\hat{\beta}^m = \left[\Sigma_i(X_{.t}-\bar{X})(X_{.t}-\bar{X})'\right]^{-1}\left[\Sigma_i(X_{.t}-\bar{X})(y_{.t}-\bar{y})'\right]$

Para someter a consideración la existencia de efectos aleatorios, Breusch y Pagan (1979) propusieron la siguiente prueba:

$H_0 : \sigma_w^2 = 0$ *versus* $H_1 : \sigma_w^2 \neq 0$

es decir, $\sigma_y^2 = \sigma_u^2$ bajo la hipótesis nula y, en todo caso, hay ordenadas al origen distintas por individuos.

El estadístico de prueba es:

$$NT/[2(T-1)]\{[\Sigma_i(\Sigma_t\hat{u}_{it})^2/\Sigma_i\Sigma_t\hat{u}_{it}^2] - 1\}^2 \sim \chi_1^2$$

El desarrollo del modelo de efectos aleatorios se ha hecho, hasta aquí, suponiendo que los efectos individuales *no están correlacionados* con las variables explicativas. La única diferencia en términos de estimación entre realizar el condicionamiento sobre la muestra o la población se refiere a la eficiencia de los estimadores. Esto, sin embargo, no parece explicar la observación muy frecuente de diferencias importantes entre los estimadores de β, según se considere que los efectos individuales son fijos o aleatorios. La explicación brindada por Mundlak (1978) es que en realidad, cuando se encuentran dichas diferencias, éstas se deben a que existe correlación entre efectos y variables explicativas. Así, el modelo *(2)* está incorrectamente especificado ya que en él se está considerando la distribución marginal de los efectos en vez de la distribución conjunta de éstos y las variables *Xs*. Sin embargo, al ser los efectos no observables, no es posible obtener dicha distribución. La propuesta de Mundlak es aproximar la esperanza condicional de α_i a $\mathbf{X}_{it}$ a través de una función lineal[3]. De esta forma, el modelo transformado es:

$$y_{it} = \alpha + \mathbf{X}'_{it}\beta + \bar{\mathbf{X}}'_{i.}\gamma + v_{it} \qquad (6)$$

siendo $v_{it} = u_{it} + \eta_i$ y con α_i en *(2)* tal que:

$$\alpha_i = \bar{\mathbf{X}}'_{i.}\gamma + \eta_i \qquad \eta_i \sim N(0,\sigma^2_\eta)$$

de forma que: $E(\alpha_i/\mathbf{X}_{it}) = \bar{\mathbf{X}}'_{i.}\gamma$

Así, cuando los α_is no están correlacionados con las $\mathbf{X}_{it}$, $\gamma=0$, y el modelo *(2)* está anidado en el *(6)*.

Los estimadores adecuados (insesgados y de mínima varianza), se calculan por MCG, de acuerdo a lo siguiente:

$$\hat{\alpha} = \bar{y} - \bar{\mathbf{X}}'\hat{\beta}^e \qquad \hat{\beta}_{MCG} = \hat{\beta}^i \qquad \hat{\gamma} = \hat{\beta}^e - \hat{\beta}^i$$

Es importante señalar que los resultados obtenidos por Mundlak son válidos siempre que la correlación se presente entre los efectos individuales y *todas* las variables explicativas. Además, los mismos son válidos en el contexto de modelos estáticos, no siendo posible mantener los mismos resultados cuando el modelo es dinámico.

Para definir si es apropiado o no el uso de efectos aleatorios, *vis à vis* efectos fijos, resulta importante someter a prueba la existencia de correlación entre las α_is y las $\mathbf{X}_{it}$. Una opción es realizar la prueba tipo *F* de $\gamma=0$ usando el modelo *(6)*. El estadístico de prueba es:

$$\frac{[NT-(2K+1)]}{K} * \frac{\Sigma_i \hat{\mathbf{u}}_i^{M2\prime} V^{*-1}\mathbf{u}_i^{M2} - \Sigma_i \hat{\mathbf{u}}_i^{M6\prime} V^{*-1}\mathbf{u}_i^{M6}}{\Sigma_i \hat{\mathbf{u}}_i^{M6\prime} V^{*-1}\mathbf{u}_i^{M6}} \sim F_{K,[NT-(2K+1)]}$$

[3] En caso de no ser posible suponer linealidad, se puede usar un predictor lineal óptimo, obteniéndose iguales resultados.

con:

$$V^{*-1} = (1/\sigma_u^2)[I_T - (1/T)\iota\iota' + \psi^*(1/T)\iota\iota']$$

$$\psi^* = \sigma_u^2/(\sigma_u^2 + T\sigma_\eta^2)$$

Una propuesta alternativa realizada por Hausman (1978) es comparar el desempeño de los estimadores de MCG y MCO bajo ambas hipótesis y en base a ello decidir. El modelo a estimar es el *(2)*. Esto equivale a estimar el modelo *(6)* bajo la hipótesis: $H_0: \gamma = 0$. Bajo este supuesto, los estimadores de MCG y MCO son consistentes aunque sólo el de MCG es eficiente. Bajo la hipótesis alternativa, por su parte, el estimador de MCO continúa siendo consistente, pero el de MCG ya no lo es: es un estimador sesgado debido a la omisión de variables relevantes - las $\bar{\mathbf{X}}_{i.}$.

Sea $\hat{\beta}^i$ el estimador de MCO en el modelo *(2)* (bajo H_0) y sea $\hat{\beta}_{MCG}$ el correspondiente estimador por MCG aplicados al mismo modelo. Entonces, es posible definir lo siguiente:

$$\hat{q} = \hat{\beta}_{MCG} - \hat{\beta}^i \qquad Var(\hat{q}) = Var(\hat{\beta}^i) - Var(\hat{\beta}_{MCG})$$

con lo cual el estadístico propuesto es:

$$H = \hat{q}' Var(\hat{q})^{-1} \hat{q} \sim \chi^2_K$$

Si H_0 no se rechaza, entonces $\gamma=0$ y $\hat{\beta}^i$ es equivalente a $\hat{\beta}_{MCG}$. No existe correlación entre efectos individuales y variables explicativas y es posible usar el modelo de efectos aleatorios tal cual se especificó en *(2)*. Por el contrario, si se rechaza la hipótesis nula, convendrá utilizar el modelo de efectos fijos, ya que el estimador de β resultante es equivalente al que se obtiene de acuerdo al modelo *(6)*.

Observación: cuando T tiende a infinito, la prueba no tiene sentido, ya que $\hat{\beta}_{MCG}$ bajo la hipótesis alternativa será asintóticamente insesgado.

III. Modelos Dinamicos

En los modelos estáticos, cuando no existe correlación entre los efectos individuales y las variables explicativas, la diferencia en el uso de modelos de efectos fijos o aleatorios se relaciona a la eficiencia de los estimadores. Sin embargo, si existe correlación entre $\mathbf{X}_{it}$ y α_i, la aplicación de MCG al modelo *(2)*, en el que no se considera explícitamente dicha correlación, resulta en estimadores sesgados por omisión de variables. De esta forma, la sugerencia es estimar el modelo *(6)* por MCG.

Al considerar modelos dinámicos, la diferenciación entre modelos de efectos fijos y aleatorios posee nuevas implicancias para la estimación. La inclusión de rezagos de la variable dependiente determina que la aplicación de MCO a un modelo de efectos fijos no resulte en estimadores de β consistentes cuando N tiende a infinito y T está fijo. Por otra parte, la consistencia de los estimadores por MCG en el modelo de efectos aleatorios dependerá de la forma en que T y N tiendan a infinito y, además, de los valores iniciales de y.

A. Modelos de Efectos Fijos

Sean: $y_{it} = \alpha_i + \lambda y_{it-1} + u_{it}$ con $|\lambda| < 1$ *(7)*

$$\bar{y}_i = (1/T)\sum_{t=1}^{T} y_{it} \quad \bar{y}_{i,-1} = (1/T)\sum_{t=1}^{T} y_{it-1}$$

La estimación de *(7)* por máxima verosimilitud o el método de covarianzas dará estimaciones consistentes de λ y α_i cuando N y T tienden a infinito.

Sin embargo, si T está fijo, los estimadores serán inconsistentes. Dicha inconsistencia surge de la eliminación de los efectos fijos que, a su vez, crea correlación entre la nueva variable explicativa y los residuos transformados. Sea:

$$y_{it} - \bar{y}_i = \lambda(y_{it-1} - \bar{y}_{i,-1}) + (u_{it} - \bar{u}_i) \qquad (8)$$

Entonces, los estimadores son:

$$\hat{\alpha}_i = \bar{y}_i - \hat{\lambda}\bar{y}_{i,-1} \qquad para\ i=1,2,\ldots,N$$

$$\hat{\lambda} = \frac{\Sigma_i \Sigma_t (y_{it} - \bar{y}_i)(y_{it-1} - \bar{y}_{i,-1})}{\Sigma_i \Sigma_t (y_{it-1} - \bar{y}_{i,-1})^2} =$$

$$= \lambda + \frac{\Sigma_i \Sigma_t (u_{it} - \bar{u}_i)(y_{it-1} - \bar{y}_{i,-1})}{\Sigma_i \Sigma_t (y_{it-1} - \bar{y}_{i,-1})^2}$$

El estimador de λ es consistente sólo si el numerador tiende a cero cuando NT tiende a infinito, siendo el denominador no nulo.

$$N = \underset{N->\infty}{\text{plim}} \frac{1}{NT} \Sigma_i \Sigma_t (u_{it} - \bar{u}_i)(y_{it-1} - \bar{y}_{i,-1}) = \frac{T(1-\lambda)+\lambda^T-1}{(1-\lambda)^2} * (\sigma_u^2/T^2)$$

$$D = \underset{N->\infty}{\text{plim}} \frac{1}{NT} \Sigma_i \Sigma_t (y_{it-1} - \bar{y}_{i,-1})^2 = \left[1-(1/T) - \frac{2\lambda[T(1-\lambda)+\lambda^T-1]}{(1-\lambda)^2 T^2}\right] * \sigma_u^2/(1-\lambda^2)$$

Así, el estimador existe, ya que el denominador es no nulo, y es sesgado para T fijo, dado que el numerador no es cero. Sin embargo, cuando T tiende a infinito, el numerador tiende a anularse mientras que el denominador tiende a una constante:

$$\lim_{T->\infty} N = -\lim_{T->\infty} \sigma_u^2/T + \lim_{T->\infty} \sigma_u^2/(1-\lambda^2)T^2 = 0$$

$$\lim_{T->\infty} D = \sigma_u^2/(1-\lambda^2).$$

El sesgo, conocido como *sesgo de Nickell*, es:

$$-(1+\lambda)(1-\lambda)[\lambda^T-1+T(1-\lambda)]/[(1-\lambda)T(T-1)-2\lambda(T-1+\lambda^T-T\lambda)]$$

A pesar de que asintóticamente el sesgo se anula, seguirá siendo de un orden importante para valores relativamente grandes de T. Por ejemplo, para un número de observaciones temporales de 15 su valor será del orden del 20% (Arellano y Bond, 1991).

En los casos en que T sea "chico", para obtener estimadores consistentes se deberá usar Variables Instrumentales. La propuesta es, entonces, estimar el modelo *(7)* en primeras diferencias:

$$\Delta y_{it} = \lambda \Delta y_{it-1} + \Delta u_{it} \qquad \text{con:} \qquad \Delta y_{it} = y_{it} - y_{it-1} \qquad \Delta u_{it} = u_{it} - u_{it-1}$$

Así, Δy_{it-1} está correlacionado con Δu_{it}. Un instrumento adecuado es Δy_{it-2}, que estando correlacionado con Δy_{it-1}, no lo está con Δu_{it}. Alternativamente se puede usar y_{it-2}.

Los estimadores resultantes son:

$$\hat{\lambda}_{VI} = \frac{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{it} - y_{it-1})(y_{it-2} - y_{i,t-3})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{it-1} - y_{it-2})(y_{it-2} - y_{it-3})}$$

o bien:

$$\hat{\lambda}_{VI} = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it} - y_{it-1})y_{it-2}}{\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{it-1} - y_{it-2})y_{it-2}}$$

y:

$$\hat{\alpha}_i = \bar{y}_i - \hat{\lambda}_{VI}\bar{y}_{i,-1}$$

B. Modelo de Efectos Aleatorios

Al igual que en el modelo anterior, las variables explicativas siempre estarán correlacionadas con el error en el modelo dinámico transformado en diferencias. Así, MCO no porporcionará estimadores insesgados de los parámetros.

Sea el siguiente modelo:

$$\mathbf{y}_i = \lambda \mathbf{y}_{i,-1} + \mathbf{X}'_i\beta + \mathbf{z}'_i\rho + \iota\alpha_i + \mathbf{u}_i \qquad (9)$$

donde las variables z_i son exógenas y no varían en el tiempo. Esto es un caso más general que el planteado en el modelo de Mundlak *(6)*, de forma que el término $(\bar{\mathbf{X}}'_{i.}\gamma+\iota\eta_i)$ queda sustituido por $(\iota z'_i\rho+\iota\alpha_i)$.

Además, y al igual que en *(6)*, este modelo anida a aquel en el que no existe correlación entre las variables exógenas y los efectos individuales ($\rho=0$).

Para ver que, nuevamente, la aplicación de MCO al modelo transformado que elimina los efectos individuales resulta en estimadores sesgados e inconsistentes cuando T tiende a infinito, se considera el caso más sencillo en que no hay variables exógenas:

$$y_{it} = \alpha_i + \lambda y_{it-1} + u_{it} \qquad \text{con } |\lambda|<1 \qquad (10)$$

$$\hat{\lambda}_{MCO} = \frac{\Sigma_i \Sigma_t (y_{it} y_{it-1})}{\Sigma_i \Sigma_t y^2_{it-1}} = \lambda + \frac{\Sigma_i \Sigma_t (u_{it} + \alpha_i) y_{it-1}}{\Sigma_i \Sigma_t y^2_{it-1}}$$

$$\underset{N\to\infty}{\text{plim}} \frac{1}{NT} \Sigma_i \Sigma_t (u_{it}+\alpha_i) y_{it-1} = (1/T)[(1-\lambda^T)/(1-\lambda)*cov(y_{i0}\alpha_i)+$$

$$+ (1/T)\sigma^2_\alpha (T-1-T\lambda+\lambda^T)/(1-\lambda)^2$$

Esta expresión no tiende a cero cuando T tiende a infinito, sino a una constante igual a: $\sigma^2_\alpha/(1-\lambda)$.

Por su parte, el límite probabilístico del denominador es también una expresión no nula. Cuando T tiende a infinito, dicha expresión tiende a una constante igual a: $\sigma^2_\alpha/(1-\lambda)^2 + \sigma^2_u/(1-\lambda^2)$.

De esta forma, el sesgo es positivo y aumentará con la varianza de los efectos (σ^2_α).

Cuando el modelo posee variables exógenas el sesgo se reduce pero manteniendo su dirección. Los coeficientes de las variables exógenas, por su parte, están sesgados hacia cero. En el caso de procesos autorregresivos de orden mayor que 1 no es posible prever la dirección del sesgo.

Para determinar los estimadores es necesario analizar, además, el comportamiento de los valores iniciales. Partiendo del modelo *(9)*, los supuestos implícitos son:

$$E(\alpha_i) = E(u_{it}) = 0 \quad \forall i,t \qquad\qquad E(\alpha_i u_{jt}) = 0 \quad \forall i,j,t$$

$$E(\alpha_i \mathbf{z}'_i) = E(\alpha_i \mathbf{x}'_i) = \mathbf{0} \quad \forall i$$

$$E(\alpha_i \alpha'_j) = \begin{cases} \sigma^2_\alpha & i=j \\ 0 & i \neq j \end{cases} \qquad\qquad E(u_{it} u_{js}) = \begin{cases} \sigma^2_u & i=j \text{ si } i=j,\ t=s \\ 0 & i\neq j \text{ otro caso} \end{cases}$$

La forma final del modelo es:

$$y_{it} = \lambda^t y_{i0} + \beta' \sum_{j=0}^{t-1} \lambda^j X_{i,t-j} + (1-\lambda^{t-1})/(1-\lambda)(\rho' \mathbf{z}_i + \alpha_i) + \sum_{j=0}^{t-1} \lambda^j u_{i,t-j}$$

Cuando t tiende a infinito, la forma final converge a:

$$y_{it} = \beta' \sum_j \lambda^j X_{i,t-j} + (\rho' \mathbf{z}_i + \alpha_i)/(1-\lambda) + \sum_j \lambda^j u_{i,t-j}$$

Se distinguen dos casos de acuerdo al carácter de los valores iniciales:

1. y_{i0} es exógeno, está fijo.
En este sentido, y_{i0} no se vería afectado por los efectos individuales y, con ello: $Cov(\alpha_i y_{i0}) = 0$. El proceso se movería, a partir de determinado valor inicial y dados α_i, $\mathbf{x}_i$, $\mathbf{z}_i$, hacia los valores dados por la forma final cuando t tiende a infinito. El supuesto es bastante cuestionable si se piensa que en general la elección del momento $t=0$ es arbitraria, dependiendo fuertemente de la disponibilidad de información.

2. y_{i0} es endógeno, es una variable aleatoria con media y varianza comunes para los individuos - μy_0 y σy_0^2 - de forma que: $y_{i0} = \mu y_0 + \varepsilon_i$

Esto equivaldría a suponer que ε_i da cuenta de las dotaciones iniciales, afectando la desviación del proceso respecto a su media. Alternativamente, es posible plantear que resulta irrelevante saber cómo se llegó a un estado inicial dado, siendo sólo de interés conocer su distribución. Bajo esta hipótesis se deben considerar dos subcasos:

2.1. y_{i0} independiente de α_i: $Cov(y_{i0}\alpha_i) = 0$

El comportamiento del proceso en esta situación es muy similar al del caso *1*, en cuanto a que el efecto de las condiciones iniciales va desapareciendo con el paso del tiempo hasta desaparecer cuando t tiende a infinito: $Cov(y_{i0}y_{it}) = \lambda^t\sigma y_0^2$

2.2 y_{i0} correlacionado con α_i: $Cov(y_{i0}\alpha_i) = \phi\sigma y_0^2$

En este caso, la covarianza entre el proceso en un momento dado del tiempo y su valor inicial no se anula, aún cuando t sea suficientemente grande, ya que:

$$Cov(y_{i0}y_{it}) = \phi\sigma y_0^2/(1-\lambda) + \sigma y_0^2[\lambda^t - \phi\lambda^{t-1}/(1-\lambda)]$$

Esta expresión converge a $\phi\sigma y_0^2/(1-\lambda)$ cuando t tiende a infinito, con lo cual el efecto de las condiciones iniciales es permanente.

La estimación de los parámetros se puede realizar, en cualquiera de los casos, por Máxima Verosimilitud (MV), utilizando algún método iterativo como el de Newton-Raphson. Sin embargo, cada una de las opciones planteadas determinará una forma distinta de la función de verosimilitud. Por su parte, las propiedades de los estimadores son iguales: cuando N y/o T tienden a infinito $\hat{\beta}$, $\hat{\lambda}$, $\hat{\sigma}_u^2$ son consistentes en todos los casos, mientras que $\hat{\rho}$ y $\hat{\sigma}_\alpha^2$ sólo son consistentes cuando N tiende a infinito. En el caso *2*, además, ocurre esto mismo con los estimadores de μy_0 y σy_0^2. La razón es que la variabilidad entre individuos es insuficiente para poder estimar correctamente dichos parámetros.

Es posible utilizar un procedimiento de estimación alternativo, si se considera al panel como un sistema simultáneo de ecuaciones con N observaciones cada una. El número de ecuaciones será T o $T+1$, dependiendo de que y_{i0} sea fijo o aleatorio. La estimación por Mínimos Cuadrados en 3 Etapas restringidos (MC3Er) -ya que existirán restricciones lineales entre los coeficientes de las distintas ecuaciones- proporciona estimadores con propiedades idénticas a los de MV en los casos *1* y *2.1* mientras que en el caso *2.2* sólo se garantiza consistencia de $\hat{\beta}$, $\hat{\lambda}$, $\hat{\sigma}^2_u$ cuando T tiende a infinito.

Las complejidades de estos métodos aunadas al hecho que es necesario decidir sobre las características que poseen los valores iniciales, vuelve atractivo un método que sea independiente de éstos. Aún más, los resultados obtenidos a partir de un procedimiento así podrían, a su vez, utilizarse como condiciones iniciales para la aplicación de MV. Un método que reúne parcialmente estas condiciones (Anderson y Hsiao, 1981) es el de Variables Instrumentales (VI) de forma que los estimadores resultantes se comportan como los de MV, independientemente de qué características posean las condiciones iniciales ($\hat{\lambda}$, $\hat{\beta}$, $\hat{\sigma}^2_u$ consistentes cuando N y/o T tienden a infinito; $\hat{\rho}$ y $\hat{\sigma}^2_\alpha$ y, si corresponde μy_0 y $\sigma^2 y_0$, sólo cuando N tiende a infinito). El procedimiento de estimación es como sigue:

1. Se transforma el modelo *(9)* de acuerdo a:

$$y_{it} - y_{it-1} = \lambda(y_{it-1} - y_{it-2}) + (\mathbf{X}'_{it} - \mathbf{X}'_{it-1})\beta + (u_{it} - u_{it-1}) \qquad (11)$$

y se estiman λ y β usando y_{it-2} o $(y_{it-2} - y_{it-3})$ como instrumentos.

2. Se estima ρ por MCO aplicados a la ecuación:

$$\bar{\mathbf{y}}_i - \hat{\lambda}\bar{\mathbf{y}}_{i,-1} - \bar{\mathbf{X}}'_i\hat{\beta} = \mathbf{z}'_i\rho + \alpha_i + \bar{\mathbf{u}}_i \qquad \text{para } i=1,2,\ldots,N$$

3. Se estiman las varianzas de acuerdo:

$$\hat{\sigma}^2_u = \Sigma_i\Sigma_t[(y_{it} - y_{it-1}) - \hat{\lambda}(y_{it-1} - y_{it-2}) - (\mathbf{X}'_{it} - \mathbf{X}'_{it-1})\hat{\beta}]^2/2N(T-1)$$

$$\hat{\sigma}^2_\alpha = \Sigma_i[\bar{\mathbf{y}}_i - \hat{\lambda}\bar{\mathbf{y}}_{i,-1} - \bar{\mathbf{X}}'_i\beta - \mathbf{z}'_i\hat{\rho}]^2/N - \hat{\sigma}^2_u/T$$

El problema que presenta esta propuesta es que la equivalencia en términos de las propiedades de los estimadores entre MV y VI se puede garantizar únicamente para $T \leq 3$. Para valores mayores, sólo se dará bajo restricciones sobre la media y la varianza del proceso. Arellano y Bover (1990) proponen una generalización al procedimiento, en la que los instrumentos que se utilizan para cada ecuación son distintos. Este método resulta equivalente al Método Generalizado de Momentos (MGM).

Sea el siguiente modelo: $$y_{it} = \alpha_i + \lambda y_{it-1} + u_{it} \qquad (12)$$

Este se puede transformar tomando primeras diferencias en:

$$y_{it} - y_{it-1} = \lambda(y_{it-1} - y_{it-2}) + (u_{it} - u_{it-1})$$

O, lo que es lo mismo: $$\Delta y_{it} = \lambda \Delta y_{it-1} + \Delta u_{it}$$

Al igual que en el caso anterior, dado que Δy_{it-1} está correlacionado con Δu_{it}, la estimación por MCO proporciona estimadores sesgados, mientras que VI soluciona el problema. En esta alternativa, sin embargo, se plantea la definición de una matriz de instrumentos, de forma que éstos vayan variando con t (hay un instrumento para $t=3$ y hasta $T-2$ para $t=T$):

$$\mathbf{Z}_i = \begin{pmatrix} y_{i1} & 0 & \ldots\ldots\ldots\ldots & 0 & \ldots\ldots\ldots & 0 \\ 0 & y_{i1} & y_{i2} & 0 \ldots\ldots & 0 \ldots\ldots\ldots & 0 \\ \vdots & & & & & \vdots \\ 0 & \ldots\ldots\ldots\ldots & & y_{i1} & \ldots\ldots & y_{iT-2} \end{pmatrix}_{(T-2)\times[(T-1)(T-2)/2]} \quad \text{para } i=1,2,\ldots,N$$

Sean: $$\mathbf{u}^*_i = \begin{pmatrix} \Delta u_{i3} \\ \vdots \\ \Delta u_{iT} \end{pmatrix} \qquad \mathbf{y}^*_{i,-1} = \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{iT-1} \end{pmatrix}$$

Entonces, se cumple que: $$E(\mathbf{Z}'_i \mathbf{u}^*_i) = 0 \quad y \quad E(\mathbf{u}^*_i \mathbf{u}^{*\prime}_i) = \sigma^2_u \mathbf{H}$$

Con $\mathbf{H}$ tal que:

$$\mathbf{H} = \begin{pmatrix} 2 & -1 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & 0 \dots & 0 \\ \vdots & & & & & \vdots \\ 0 & \dots & \dots & 0 & -1 & 2 \end{pmatrix}$$

Así, la matriz de varianzas y covarianzas de $\mathbf{Z}'_i\mathbf{u}^*_i$ es:

$$E(\mathbf{Z}'_i\mathbf{u}^*_i\mathbf{u}^{*\prime}_i\mathbf{Z}_i) = \sigma^2_u E(\mathbf{Z}'_i\mathbf{H}\mathbf{Z}_i)$$

Con lo que el estimador óptimo de λ es:

$$\hat{\lambda} = \frac{\Sigma_i \mathbf{y}^*_{i,-1}\mathbf{Z}_i(\Sigma_i \mathbf{Z}'_i\mathbf{H}\mathbf{Z}_i)^{-1}\mathbf{Z}'_i\mathbf{y}^*_i}{\Sigma_i \mathbf{y}^*_{i,-1}\mathbf{Z}_i(\Sigma_i \mathbf{Z}'_i\mathbf{H}\mathbf{Z}_i)^{-1}\mathbf{Z}'_i\mathbf{y}^*_{i-1}}$$

Si las perturbaciones u_{it} estuviesen autocorrelacionadas o fuesen heteroscedásticas, lo que corresponde es usar el MGM en dos etapas. En la primera fase, se obtiene $\hat{\lambda}$ tal cual se definió más arriba y con éste $\hat{\mathbf{u}}^*$ y $\hat{\sigma}^2_u$. Dados estos últimos, se calcula un estimador más general de la varianza de $\mathbf{Z}'_i\mathbf{u}^*_i$ de acuerdo a: $\hat{\sigma}^2_u\Sigma_i\mathbf{Z}'_i\mathbf{u}^*_i\mathbf{u}^{*\prime}_i\mathbf{Z}_i$. Con él, se re-estima λ en una segunda etapa.

Finalmente, dado que al aplicar primeras diferencias al modelo *(12)* se introduce autocorrelación en los residuos, Arellano (1988) sugiere una transformación alternativa, de acuerdo a:

$$c_t y^0_{it} = \lambda c_t y^0_{it-1} + c_t u^0_{it}$$

donde:

c_t es tal que: $c^2_t = (T-t)/(T-t-1)$

$y^0_{it} = y_{it} - [1/(T-t)](y_{it+1} + \dots + y_{iT})$

$u^0_{it} = u_{it} - [1/(T-t)](u_{it+1} + \dots + u_{iT})$

Así, $c_t u^0_{it}$ es no autocorrelacionada y homoscedástica, y es posible aplicar MCO al modelo transformado para obtener buenos estimadores.

Si el modelo tiene variables exógenas, el marco de análisis es el mismo, sólo que se dispone de un mayor número de instrumentos. Es interesante diferenciar tres casos.

Sea el modelo:

$$y_{it} = \alpha_i + \lambda y_{it-1} + \mathbf{X}'_{it}\beta + u_{it} \qquad \text{con } |\lambda|<1 \qquad (13)$$

1. Las X_{it} son predeterminadas y correlacionadas con α_i. Las perturbaciones son no autocorrelacionadas.

Esto implica:

$$E(X_{it}u_{is}) = 0 \quad t \leq s \qquad\qquad E(\alpha_i u_{it}) = 0 \quad \forall i$$

Entonces:

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\beta} \end{pmatrix} = \left[\mathbf{W}^{*\prime}\mathbf{Z}A_N\mathbf{Z}'\mathbf{W}^{*}\right]\left[\mathbf{W}^{*\prime}\mathbf{Z}A_N\mathbf{Z}'\mathbf{Y}^{*}\right]$$

donde:

$$\mathbf{W}_{it} = \begin{pmatrix} y^{*}_{i,t-1} \\ \cdots\cdots \\ \mathbf{X}^{*}_{i,t-1} \end{pmatrix} \qquad \mathbf{Z}_i = diag(\mathbf{Z}'_{it}) \qquad A_N = \left[\sum_{i=1}^{N}\mathbf{Z}'_i\mathbf{u}^{*}_i\mathbf{u}^{*\prime}_i\mathbf{Z}_i\right]^{-1}$$

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{12} \\ \vdots \\ \mathbf{W}_{1T} \\ \vdots \\ \mathbf{W}_{NT} \end{pmatrix} ; \qquad \mathbf{y} = \begin{pmatrix} y_{12} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{NT} \end{pmatrix}$$

$\mathbf{W}^{*}$ y $\mathbf{y}^{*}$ son las desviaciones ortogonales o primeras diferencias de $\mathbf{W}$ y $\mathbf{y}$, respectivamente.

2. Las X_{it} son predeterminadas y correlacionadas con α_i; u_{it} autocorrelacionado, pero su proceso posee pocos parámetros.

En este caso habrá que estimar el modelo para las perturbaciones conjuntamente con el sistema de ecuaciones y utilizar MGM en dos etapas. La existencia de componentes de medias móviles complica bastante la especificación debiéndose imponer restricciones adicionales. También es posible usar MC3E y lograr estimadores eficientes.

3. Las X_{it} son estrictamente exógenas: $E(X_{it}u_{is}) = 0 \quad \forall t,s$

Entonces, es posible usar todas las variables Xs como instrumentos válidos en cada una de las ecuaciones, en todos los períodos.

Finalmente, al utilizar MGM se está suponiendo que las perturbaciones son no autocorrelacionadas. Pero si esto no ocurre, se obtendrán estimadores inconsistentes de los parámetros. De esta forma, resulta imprescindible someter a prueba dicha hipótesis. Para ello es importante recordar que cuando u_{it} no está autocorrelacionada, su primer diferencia Δu_{it} lo estará por definición. Sin embargo, las segundas diferencias $\Delta\Delta u_{it}$ serán no autocorrelacionadas. En base a ello se propone el siguiente estadístico:

$$m_2 = (1/\hat{\omega})\sum_{i=1}^{N}\sum_{t=4}^{T}\hat{u}^*_{it}\hat{u}^*_{it-2} \sim N(0,1)$$

Con:

$$\hat{u}^*_{it} = \Delta\hat{u}_{it}$$

$$\hat{\omega}^2 = \sum_{i=1}^{N}\hat{\mathbf{u}}^{*\prime}_{i,-2}\hat{\mathbf{u}}^0_i\hat{\mathbf{u}}^{0\prime}_i\hat{\mathbf{u}}^*_{i,-2} - 2\hat{\mathbf{u}}^*_{i,-2}\mathbf{M}^0(\mathbf{M}'\mathbf{Z}A_N\mathbf{Z}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{Z}A_N(\sum_{i=1}^{N}\mathbf{Z}'_i\hat{\mathbf{u}}^*_i\hat{\mathbf{u}}^{0\prime}_i\hat{\mathbf{u}}^*_{i,-2}) +$$

$$+ \hat{\mathbf{u}}^{*\prime}_{-2}\mathbf{M}^0 avar(\hat{\delta})\mathbf{M}^{0\prime}\hat{\mathbf{u}}^{*}_{-2}$$

siendo $\mathbf{M}^0$ y $\hat{\mathbf{u}}^0_i$ la matriz y los vectores creados eliminando observaciones para que sean de igual orden que $\mathbf{M}$ y $\hat{\mathbf{u}}^*_i$ respectivamente, mientras que $\mathbf{M} = (\mathbf{y}_{-1}, \mathbf{X})'$ y $\delta = (\lambda, \beta)'$.

$\hat{\omega}^2$ es el estimador de la varianza de $(1/N)^{1/2}\sum_i\sum_t\hat{u}^*_{it}\hat{u}^*_{it-2}$ que se puede interpretar como la contrapartida muestral de la esperanza de las autocovarianzas, que bajo la hipótesis nula serán cero. Así, si m_2 es cero, no hay autocorrelación de segundo orden y se puede afirmar u_{it} es no autocorrelacionada.

Una segunda prueba de interés se refiere a si la elección de los instrumentos es óptima, dado que MGM implica la existencia de restricciones de sobre-identificación. Para ello se propone el siguiente estadístico:

$$s = \left[\Sigma_i \hat{\mathbf{u}}_i^{*\prime} \mathbf{Z}_i\right] A_N \left[\Sigma_i \mathbf{Z}_i' \hat{\mathbf{u}}_i^*\right] \sim \chi_r^2$$

con r igual a la diferencia entre el número de columnas en $\mathbf{Z}$ y el número de parámetros a estimar, es decir, las relaciones sobre-identificadoras. Si el conjunto de instrumentos es adecuado, habrá una buena aproximación muestral a los momentos $E(\mathbf{Z}_i' \mathbf{u}_i^*) = 0$

REFERENCIAS BIBLIOGRAFICAS

Anderson, T.W. y C. Hsiao (1981) 'Estimation of dynamic models with error components', *Journal of the American Statistical Association 76*: 598-606.

Arellano, M (1988) 'An alternative transformation for fixed effects models with predetermined variables', Applied Economic Discussion Paper 57, Oxford.

__________ y S. Bond (1991) 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations', *Review of Economic Studies 58*: 277-297.

__________ y Bover (1990) 'La econometría de datos panel', Investigaciones Económicas (Segunda época), Vol.XIV, ≠1.

Breusch, T.S. y A.R. Pagan (1979) 'A simple test for heteroscedasticity and random coefficient variation', *Econometrica 47*: 1287-1294.

Green, W.H. (199?) *Econometric analysis*, MacMillan.

Hausman, J.A. (1978) 'Specification tests in econometrics', *Econometrica 46*: 1251-1272.

Hsiao C. (1986) *Analysis of panel data,* Cambridge Univ. Press.

Mundlak, Y. (1978) 'On the pooling of time series and cross section data', *Econometrica 46*: 69-85.

White H.(1980) 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity'*Econometrica 48*: 817-838.