Correlación Serial con Wavelets

Alberto Antonio Villa Valencia
Javier Martínez Plazas

Universidad EAFIT
Departamento de Ciencias Básicas
Maestría en matemáticas aplicadas
Medellín
2007

Correlación Serial con Wavelets

Alberto Antonio Villa Valencia
Javier Martínez Plazas

Trabajo de investigación presentado como requisito parcial para optar el título de Magíster en Matemáticas Aplicadas

Directores
Jairo Villegas G
Doctor en Matemáticas Aplicadas
Francisco Zuluaga D
Magíster en Matemáticas Aplicadas

Universidad EAFIT
Departamento de Ciencias Básicas
Maestría en matemáticas aplicadas
Medellín
2007

A Sulbey María, a mis hijas Katheryn y Aslhey Sofía,
y mis padres María Ligia y Fabio Antonio.
Alberto V.

A mi familia personal y profesional.
Javier M.

Agradecimientos

A las universidades, de la Amazonia y EAFIT por brindarnos la oportunidad de cualificarnos; a los profesores Jairo Villegas y Francisco Zuluaga Díaz quienes con su asesoría y su conocimiento facilitaron el desarrollo de este trabajo; a Tomás Olarte, estudiante de Ingeniería Matemática por sus aportes en la programación de los test.

Índice general

Introducción

El objetivo del presente trabajo es aplicar los test $\hat{W}_1$ y $\hat{W}_2$ propuestos por Yongmiao Hong[1] y Chihwa Kao[2], (ver [30]), en pruebas basadas en wavelet para correlación serial de forma no conocida en la estimación residual de un modelo de panel de datos reales obtenidos a partir del estudio de demanda de dinero, trabajo realizado por Francisco I. Zuluaga Díaz[3] cuya base de datos fue la Superintendencia de Sociedades (Supersociedades), en el cual se tomaron 3029 empresas pertenecientes a diversos sectores: construcción, financiero, agropecuario, minero, transporte, energía, salud, educación, entre otros, durante 1998 y 2003 (ver [72]).

A partir del modelo (4.2.1), los métodos de regresión total, efectos fijos y efectos aleatorios y los residuales calculados con el software OX. Para obtener los valores de los test calculados $\hat{W}_1$ y $\hat{W}_2$ programamos la función *wavetest* en Matlab, lo que nos permitió verificar la existencia o no de correlación serial en el componentes de error $\{\varepsilon_{it}\}$.

[1]Dept. of Economics and Dept. of Statistical Science, Cornell University, 424 Uris Hall, Ithaca, NY 14850, USA., and Dept. of Economics, Tsinghua University, Beijing 100084, China; yh20@cornell.edu

[2]Dept. of Economics and Center for Police Research, Syracuse University, 426 Eggers Hall, Syracuse, NY, USA., cdkao@maxwell.syr.edu

[3]Dept. de Ciencias Básicas, Magister en Matemáticas Aplicadas, Universidad EAFIT, Medellín, Colombia, fzuluaga@eafit.edu.co

El trabajo está organizado en cuatro capítulos. En el primer capítulo presentamos en forma general la terminología básica para el fundamento teórico de los capítulos siguientes, tales como el análisis de Fourier, series de tiempo y teoría de probabilidad. En el segundo capítulo se discuten conceptos sobre *wavelet* para el análisis multirresolución de series de tiempo que permiten construir wavelet con mejores propiedades de aproximación. En el tercer capítulo se plantean los conceptos sobre *panel de datos*, sus ventajas y desventajas, los test que determinan los modelos econométricos a utilizar y algunos aspectos metodológicos de la técnica de panel. Finalmente, en el capítulo cuarto se presentan los valores de los test calculados $\hat{W}_1$ y $\hat{W}_2$ con los cuales verificamos las hipótesis y finalmente se presentan las conclusiones del trabajo.

CAPÍTULO 1

Preliminares

1.1. Terminología

En este corto capítulo se presentará alguna terminología necesaria para la lectura de esta monografía. Se hace un corto repaso de temas de análisis, teoría de probabilidad, series de tiempo y teoría asintótica (ver p.e., [6], [14], [17], [28], [46], [52], [66]) .

Recuerde que $L_1(\mathbb{R})$ es el espacio de todas las funciones $f : \mathbb{R} \to \mathbb{C}$, tal que $\int_{\mathbb{R}} |f(t)|dt = \|f\|_{L_1} < \infty$. De igual forma se tiene $L_2(\mathbb{R})$, el espacio las funciones cuadrado-integrables, cuya norma es

$$\|f\|_{L_2} = \left(\int_{\mathbb{R}} |f(t)|^2 dt\right)^{1/2} < \infty.$$

Este espacio se dota con el producto escalar

$$\langle f, g\rangle_{L_2} = \int_{\mathbb{R}} f(t)\overline{g(t)}dt,$$

donde $\overline{g(t)}$ denota el conjugado complejo de $g(t)$. Con este producto interno el espacio $L_2(\mathbb{R})$ es de Hilbert. Las funciones $f, g \in L_2(\mathbb{R})$ son ortogonales si $\langle f, g\rangle_{L_2} = 0$. En general, $L_p(\mathbb{R})$ $(p \geq 1)$, es el espacio de todas las funciones (clases de equivalencia) $f : \mathbb{R} \to \mathbb{C}$, tal que $\int_{\mathbb{R}} |f(t)|^p dt = \|f\|_{L_p}^p < \infty$, acá

$$\|f\|_{L_p} = \left(\int_{\mathbb{R}} |f(t)|^p dt\right)^{1/p}$$

es la norma de f en $L_p(\mathbb{R})$. Otro espacio que se utilizará es $\ell_2(\mathbb{Z})$, el de las sucesiones (x_j), $j \in \mathbb{Z}$, tal que $\sum_j |x_j|^2 < \infty$.

Sea $F = \mathbb{C}$ o $\mathbb{R}$, X y Y espacios normados (espacios vectoriales equipados con una norma). Un operador lineal es una función $T : X \to Y$ tal que $T(a\,u + b\,v) = a\,T(u) + b\,T(v)$, para cada $a, b \in F$ y cada $u, v \in X$. El operador T es continuo en u_0 si para cada $\epsilon > 0$ existe $\delta > 0$ tal que si

$$\|u - u_0\|_X < \delta \quad \text{entonces} \quad \|T\,u - T\,u_0\|_Y < \epsilon. \tag{1.1.1}$$

Si (1.1.1) se cumple para cada $u_0 \in X$ se dice que T es continuo en X. Si δ no depende del punto u_0 se dice que T es uniformemente continuo en X.

El operador T es acotado si y sólo si existe una constante $c > 0$ tal que $\|T\,u\|_Y \leq c\|u\|_X$ para cada $u \in X$.

Si $f, g \in L_1(\mathbb{R})$, entonces la convolución de f y g, denotada $f * g$, se define por

$$(f * g)(t) = \int_{\mathbb{R}} f(t - z)g(z)dz.$$

Un sistema de funciones $\{\phi_j, j \in \mathbb{Z}\}$, $\phi_j \in L_2(\mathbb{R})$, se llama ortonormal si

$$\int_{\mathbb{R}} \phi_j(t)\overline{\phi_k(t)}dt = \delta_{jk},$$

donde δ_{jk} es la delta de Kronecker. Es decir,

$$\delta_{jk} = \begin{cases} 1, & \text{si } j = k; \\ 0, & \text{si } j \neq k. \end{cases}$$

Un sistema ortonormal se llama una base en un subespacio V de $L_2(\mathbb{R})$ si cualquier función $f \in V$ tiene una representación de la forma

$$f(t) = \sum_j c_j \phi_j(t),$$

donde los coeficientes c_j satisfacen $\sum_j |c_j|^2 < \infty$. En lo que sigue se utilizará la notación $\sum_j = \sum_{j=-\infty}^{\infty}$, $\int_{\mathbb{R}} = \int_{-\infty}^{\infty}$, $\|f\|_{L_2} = \|f\|_2$ y $\langle , \rangle_2$.

La función característica del conjunto A, χ_A, se define por

$$\chi_A(t) = \begin{cases} 1, & t \in A; \\ 0, & t \notin A. \end{cases}$$

También se utilizará la notación $I\{A\}$ para denotar esta función y la llaman función indicadora.

1.2. Espacio de probabilidad

A partir de la teoría de la medida, la teoría de probabilidad ha alcanzado un alto grado de formalización. En las siguientes líneas se presentan algunos elementos básicos sobre el tema, para un estudio profundo se puede consultar [6].

Definición 1.2.1. *Sea Ω un conjunto no vacío y $\mathcal{A}$ una colección de subconjuntos de Ω. $\mathcal{A}$ es una $\sigma-$álgebra sobre Ω si y sólo si se satisfacen las siguientes condiciones*

i) $\Omega \in \mathcal{A}$

ii) Si $A_1, A_2, \ldots$ es una sucesión contable de elementos de $\mathcal{A}$, entonces $\bigcup A_n \in \mathcal{A}$

iii) Si $A \in \mathcal{A}$, entonces $A^c \in \mathcal{A}$, donde A^c es el complemento de A en Ω.

La pareja $(\Omega, \mathcal{A})$ se llama espacio medible y a los elementos de $\mathcal{A}$, conjuntos medibles.

Definición 1.2.2. *Sea $\mathcal{C}$ una colección de subconjuntos de Ω. Por $\sigma-$álgebra minimal que contiene a $\mathcal{C}$ o la $\sigma-$álgebra que genera a $\mathcal{C}$, denotada $\sigma(\mathcal{C})$, se entiende una $\sigma-$álgebra de subconjuntos de Ω tal que si $\mathcal{K}$ es otra $\sigma-$álgebra que contiene a $\mathcal{C}$, entonces $\mathcal{C} \subset \sigma(\mathcal{C}) \subset \mathcal{K}$.*

La $\sigma-$álgebra $\mathcal{B}$ generada por todos los conjuntos abiertos de $\mathbb{R}^n$, se llama álgebra de Borel y los elementos en $\mathcal{B}$ se llaman conjuntos de Borel. Esta $\sigma-$álgebra es de gran interés en diversos campos de la matemática, en particular en la teoría de probabilidades.

Definición 1.2.3. *Una probabilidad P es una medida normalizada sobre un espacio medible $(\Omega, \mathcal{A})$; esto es, P es una función de valor real la cual asigna a todo $A \in \mathcal{A}$ el número $P(A)$ tal que* *i)* $P(\Omega) = 1$

ii) *Si* $A_1, A_2, \ldots$ *es una sucesión contable de elementos de* $\mathcal{A}$ *disjuntos dos a dos, entonces*

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n)$$

iii) $P(A) \geq 0$ *para todo* $A \in \mathcal{A}$.

La tripla $(\Omega, \mathcal{A}, P)$ se llama espacio de probabilidad. $P(A)$ se lee como la probabilidad del evento A.

Algunas consecuencias de la definición (1.2.3) son:

1. $P(\emptyset) = 0$.

2. Sean A y B eventos. Si $A \subset B$, entonces $P(A) \leq P(B)$.

3. Si $A_1, A_2, \ldots, A_n$ son eventos disjuntos dos a dos, entonces

$$P\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} P(A_k).$$

4. $P(A^c) = 1 - P(A)$, para todo $A \in \mathcal{A}$.

5. Si $\{A_n\}$ es una sucesión contable de eventos, entonces

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} P(A_n), \quad \text{desigualdad de Boole.}$$

1.2.1. Variable aleatoria

Definición 1.2.4. *Una variable aleatoria* X *es una función de valor real cuyo dominio es* Ω *y la cual es* $\mathcal{A}-$*medible, esto es, para cualquier número real* x, $\{\omega \in \Omega : \ X(\omega) \leq x\} \in \mathcal{A}$.

El conjunto $\{\omega \in \Omega : \ X(\omega) \leq x\}$ se llama conjunto de eventos elementales, se denotará por $[X \leq x]$.

Si X es una variable aleatoria, la función de distribución F_X se define por

$$F_X(x) = P[X \leq x], \quad \text{para todo} \quad x \in \mathbb{R}.$$

Note que diferentes variables aleatorias pueden tener la misma función de distribución. Por ejemplo, sea $\Omega = \{C, S\}$, si $P(C) = P(S) = 1/2$ y si X y Y son variables aleatorias definidas por $X(C) = 1$, $X(S) = 0$, $Y(C) = 0$ y $Y(S) = 1$, entonces

$$F_X(x) = F_Y(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1/2, & \text{si } 0 \leq x < 1 \\ 1, & \text{si } x \geq 1. \end{cases}$$

Si X es una variable aleatoria, entonces la función de distribución F_X tiene las siguientes propiedades:

1. F_X es no decreciente, es decir, si $-\infty < a < b < \infty$, entonces

$$F_X(a) \leq F_X(b$$

2. $\lim_{x\to\infty} F_X(x) = 1$ y $\lim_{x\to-\infty} F_X(x) = 0$

3. F_X es continua por la derecha, esto es,

$$\lim_{h\downarrow 0} F_X(x+h) = F_X(x), \quad \forall x.$$

Una función de distribución F se llama absolutamente continua, si existe una función medible Borel f sobre $\mathbb{R}$ tal que

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad \forall x.$$

La función f se llama la densidad de F.

Si X es una variable aleatoria con función de distribución absolutamente continua y densidad f, entonces el valor esperado de X es dado por

$$E(X) = \int_{\mathbb{R}} x f(x) dx,$$

siempre que la integral sea finita.

1.2.2. Procesos estocásticos

Una variable aleatoria siempre tiene asociada una distribución de probabilidad que mide la probabilidad de ocurrencia de sus distintos resultados. Cuando la variable aleatoria cambia con el tiempo, se le puede asociar una distribución de probabilidad que también varía con el tiempo. En tales ambientes resulta útil definir un proceso estocástico [52].

Definición 1.2.5. *Sea $I \subseteq \mathbb{R}$ un conjunto de índices y $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Una función $X : I \times \Omega \to \mathbb{R}^n$ es un proceso estocástico si para cada $t \in I$ fijo, la función $X_t : \Omega \to \mathbb{R}^n$ es una variable aleatoria, que representa el valor del proceso $X(t, \omega)$, $\omega \in \Omega$. Si $\omega \in \Omega$ es fijo, la aplicación $I \to \mathbb{R}^n$ tal que $t \mapsto X_t(\omega)$ se llama la trayectoria o realización del proceso X.*

Los valores que toma el proceso en $\mathbb{R}^n$ se llaman estados del proceso. Si el conjunto I es contable, el proceso estocástico X se dice que es de tiempo discreto. Por otro lado, si I es un intervalo de los reales no negativos, el proceso estocástico es de tiempo continuo.

Si X es un proceso estocástico continuo, entonces

i) X es independiente si para todo $t, s \in I$ $s \neq t$, las variables aleatorias asociadas X_s y X_t son independientes.

ii) X es independientemente distribuida, si la distribución de probabilidad F_{X_t} es la misma para cada $t \in I$.

iii) X tiene incrementos independientes si para cada $n \geq 1$ y para cualquier partición del intervalo I, $t_0 < t_1 < \cdots < t_n$, las diferencias

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$$

son variables aleatorias independientes.

iv) X tiene incrementos estacionarios si $X_t - X_s \stackrel{d}{=} X_{t+h} - X_{t+s}$ para cada t, s, $t+h$, $s+h$ en I, $s < t$ y $h > 0$. El símbolo $\stackrel{d}{=}$ significa que los términos en comparación tienen la misma distribución de probabilidad. La estacionariedad de un proceso estocástico alude a que la distribución de probabilidad de la diferencia entre dos variables aleatorias permanece invariante bajo cualquier traslación temporal.

Una serie de tiempo es la realización de un proceso estocástico. En otras palabras, una serie de tiempo se puede consider como una colección de variables aleatorias $\{X_t : t \in I\}$ (ver p.e., [28], [46] o [66]).

1.3. Teoría asintótica

Cuando se desea determinar si un estimador es bueno, es decir, consistente, eficiente, suficiente o sesgado, es difícil determinarlo. Sin embargo se puede aproximar al comportamiento a partir de su distribución para tamaños muestrales altos. Esto se puede hacer a partir del límite de la distribución del estimador para inferirlo.

1.3.1. Convergencia en probabilidad.

Definición 1.3.1. ***Convergencia en probabilidad*** *La variable aleatoria x_n converge en probabilidad a una constante c, si*

$$\lim_{n\to\infty} Prob(|x_n - c| > \epsilon) = 0$$

para cualquier $\epsilon > 0$.

La convergencia en probabilidad implica que los valores cercanos a c que toma la variable son cada vez más probables, a medida que n aumenta.

Definición 1.3.2. ***Estimador Consistente.*** *Un estimador $\hat{\theta}$ de un parámetro θ es un estimador consistente de θ si y sólo si $p\lim \hat{\theta} = \theta$.*

1.3.2. Convergencia en distribución y distribución límite

Sea x_n una sucesión de variables aleatorias, cuyo elemento representativo x_n, es una variable aleatoria obtenida de una muestra de tamaño n. Supongamos que x_n tiene una función de distribución $F_n(x)$.

Definición 1.3.3. Convergencia en Distribución. *Decimos que la sucesión de variables aleatorias $\{x_n\}$ converge en distribución a una variable aleatoria x con función de distribución $F(x)$ si*

$$\lim_{n\to\infty} |F_n(x) - F(x)| = 0$$

en todos los puntos en los que $F(x)$ sea continua.

Cabe notar que la convergencia en distribución esta relacionada con la distribución de probabilidad asociada a $\{x_n\}$ a medida que n aumenta; no implica la convergencia de los valores que toma cada variable aleatoria x_n.

Definición 1.3.4. Distribución límite *Si x_n converge en distribución a x, siendo $F(x)$ la función de distribución de x, entonces $F(x)$ es la distribución limite de x. Se representa $x_n \xrightarrow{d} x$.*

1.3.3. Distribuciones asintóticas

Las distribuciones asintóticas obtenidas a partir del teorema central del límite dependen de parámetros desconocidos, ahora bien, lo que haremos es derivar las distribuciones asintóticas de los estimadores que nos interesan. La forma más común de plantear una distribución asintótica es construirla a partir de la distribución límite conocida de la variable aleatoria. Si

$$\sqrt{n}[(\bar{x}_n - \mu/\sigma)] \xrightarrow{d} N[0,1],$$

entonces, aproximadamente, o asintóticamente, $x_n \sim N[\mu, \sigma^2/n]$, lo que escribiremos como

$$x_n \xrightarrow{a} N[\mu, \sigma^2/n].$$

Mediante la afirmación de que $\bar{x}_n$ se distribuye asintóticamente como una normal con media μ y varianza σ^2/n, quiere decirse que esta distribución normal es una aproximación a la verdadera distribución finita, no que la verdadera distribución es exactamente una normal.

Extendiendo la definición, supongamos que $\hat{\theta}$ es un estimador del vector de parámetros θ. La distribución asintótica del vector $\hat{\theta}$ se obtiene de la distribución límite,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N[0, \mathbf{V}] \tag{1.3.1}$$

con $\mathbf{V}$ una matriz de covarianza de 2×2, lo que implica que

$$\hat{\theta} \xrightarrow{d} N[\theta, \frac{1}{n}\mathbf{V}].$$

La matriz de covarianzas de la distribución asintótica es la **matriz de covarianzas asintóticas** y se designa por

$$Asy.Var[\hat{\theta}] = \frac{1}{n}\mathbf{V}.$$

Recordemos que una matriz es semidefinida positiva si $x'Ax \geq 0, \; \forall x \in \mathbb{R}$, con A simétrica y $f(x) = x'Ax$, donde x' es el vector tanspuesto (véase p.e., [4]).

Definición 1.3.5. Normalidad y eficiencia asintótica. $\hat{\theta}$ *es asintóticamente normal si 1.3.1 se cumple y asintóticamente eficiente si la diferencia entre la matriz de covarianzas de cualquier otro estimador consistente que sea asintóticamente normal y* $1/n\mathbf{V}$ *es una matriz semidefinida positiva.*

CAPÍTULO 2

Introducción a las wavelets

2.1. Introducción

El origen de la descomposición de una señal en wavelets está en la necesidad de conocer las características y particularidades de la señal en diferentes instantes de tiempo. La principal virtud de las wavelets es que permite modelar procesos que dependen fuertemente del tiempo y para los cuales su comportamiento no tiene porqué ser suave [1], [13], [15], [16], [22]. Una de las ventajas de las wavelets frente a los métodos clásicos, como la transformada de Fourier, es que en el segundo caso se maneja una base de funciones bien localizada en frecuencia pero no en tiempo, esto es, el análisis en frecuencia obtenido del análisis de Fourier es insensible a perturbaciones que supongan variaciones instantáneas y puntuales de la señal como picos debidos a conmutaciones o variaciones muy lentas como tendencias. En otras palabras, si f es una señal (f es una función definida en todo $\mathbb{R}$ y tiene energía finita $\int_{-\infty}^{\infty} |f(t)|^2 dt$). La transformada de Fourier $\hat{f}(\omega)$ proporciona la información global de la señal en el tiempo localizada en frecuencia. Sin embargo, $\hat{f}(\omega)$ no particulariza la información para intervalos de tiempo específicos, ya que

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

y la integración es sobre todo tiempo (ver [23]). Así, la imagen obtenida no contiene información sobre tiempos específicos, sino que sólo permite calcular el espectro de amplitud total $|\hat{f}(\omega)|$, mientras que la mayoría de las wavelets interesantes presentan una buena localización en tiempo y en frecuencia, disponiendo incluso de bases de wavelets con soporte compacto.

En este capítulo se presenta una introducción a las transformadas de Fourier y wavelets.

2.2. Transformada de Fourier

En esta sección se recordará la definición y algunas propiedades importantes de la transformada de Fourier. En particular, se hará un resumen de resultados básicos de análisis de Fourier omitiendo sus pruebas, las cuales se pueden encontrar en algunos de los siguientes textos [7], [23], [48], [60], [64].

Definición 2.2.1. *Sea $f \in L_1(\mathbb{R})$ y $\omega \in \mathbb{R}$. La transformada de Fourier de f en ω se define por*

$$\hat{f}(\omega) := \int_{\mathbb{R}} f(t) e^{-i\omega t} dt.$$

Como

$$\int_{\mathbb{R}} |f(t)||e^{-it\omega}| dt = \int_{\mathbb{R}} |f(t)| dt = \|f\|_{L_1} < \infty$$

se tiene que la transformada de Fourier está bien definida. La aplicación $f \mapsto \hat{f}$ se llama transformación de Fourier y se denota por $\mathcal{F}$ ($\mathcal{F}(f) = \hat{f}$). La función $\hat{f}$ es continua y tiende a cero cuando $|\omega| \to \infty$ (Lema de Riemann-Lebesgue). Es claro que $\mathcal{F}(a\,f + b\,g) = a\,\mathcal{F}(f) + b\,\mathcal{F}(g)$, para cada $a, b \in \mathbb{R}$.

En general $\hat{f}$ no es una función integrable, por ejemplo, sea

$$f(t) = \begin{cases} 1, & |t| < 1; \\ 0, & |t| > 1. \end{cases}$$

Entonces

$$\begin{aligned} \hat{f}(\omega) &= \int_{-1}^{1} e^{-it\omega} dt = \left[\frac{e^{-i\omega} - e^{i\omega}}{-i\omega} \right] \\ &= \frac{\operatorname{sen} \omega}{\omega} \notin L_1(\mathbb{R}). \end{aligned}$$

Si $\hat{f}(\omega)$ es integrable, entonces existe una versión continua de f y se puede obtener la fórmula de inversión de Fourier

$$f(t) = \mathcal{F}^{-1}\big(\hat{f}(\omega)\big) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\omega)e^{i\omega t}d\omega. \tag{2.2.1}$$

La siguiente proposición recoge algunas propiedades fundamentales de la transformada de Fourier.

Proposición 2.2.2. *Sean f, $g \in L_1(\mathbb{R})$, entonces*

1. $\widehat{(T_x f)}(\omega) = e^{-i\omega x}\hat{f}(\omega)$, *donde* $(T_a f)(t) = f(t-a)$.
2. $(T_x \hat{f})(\omega) = (\widehat{e^{ix(\cdot)}f})(\omega)$
3. $\widehat{f * g} = \hat{f}\hat{g}$
4. *Si* $\epsilon > 0$ *y* $g_\epsilon(t) = g(\epsilon\, t)$ *entonces* $\hat{g}_\epsilon(\omega) = \epsilon^{-1}\hat{g}(\omega/\epsilon)$.

Otro resultado útil es el siguiente: Si $f, g \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$, entonces

$$\|f\|_2^2 = \frac{1}{2\pi}\int_{\mathbb{R}} |\hat{f}(\omega)|^2 d\omega \quad \text{(fórmula de Plancherel)} \tag{2.2.2}$$

$$\langle f, g\rangle_2 = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\omega)\overline{\hat{g}(\omega)}d\omega \quad \text{(fórmula de Parseval).} \tag{2.2.3}$$

Por extensión, la transformada de Fourier se puede definir para cualquier $f \in L_2(\mathbb{R})$. En virtud a que el espacio $L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ es denso en $L_2(\mathbb{R})$. Luego, por isometría (excepto por el factor $1/2\pi$) se define $\hat{f}$ para cualquier $f \in L_2(\mathbb{R})$, y las fórmulas (2.2.2) y (2.2.3) permanecen válidas para todo $f, g \in L_2(\mathbb{R})$.

En teoría de señales, la cantidad $\|f\|_2$ mide la energía de la señal, mientras que $\|\hat{f}\|_2$ representa el espectro de potencia de f.

Si f es tal que $\int_{\mathbb{R}} |t|^k |f(t)|dt < \infty$, para algún entero $k \geq 1$, entonces

$$\frac{d^k}{d\omega^k}\hat{f}(\omega) = \int_{\mathbb{R}} (-it)^k e^{-i\omega t} f(t)dt. \tag{2.2.4}$$

Recíprocamente, si $\int_{\mathbb{R}} |\omega|^k |\hat{f}(\omega)|d\omega < \infty$, entonces

$$(i\omega)^k \hat{f}(\omega) = \mathcal{F}\big(f^{(k)}\big)(\omega). \tag{2.2.5}$$

2.2.1. Serie de Fourier

Sea f una función $2\pi-$periódica en $\mathbb{R}$. Se escribirá $f \in L_p(0, 2\pi)$ si

$$f(t)\chi_{[0,2\pi]}(t) \in L_p(0, 2\pi), \quad p \geq 1.$$

Cualquier función f, $2\pi-$periódica en $\mathbb{R}$, tal que $f \in L_2(0, 2\pi)$, se puede representar por una serie de Fourier convergente en $L_2(0, 2\pi)$

$$f(t) = \sum_n c_n e^{int},$$

donde los coeficientes de Fourier son dados por

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int} dt.$$

Se puede verificar que si $f \in L_1(\mathbb{R})$, entonces la serie

$$S(t) = \sum_k f(t + 2k\pi) \tag{2.2.6}$$

converge casi para todo t y pertenece a $L_1(0, 2\pi)$. Además, los coeficientes de Fourier de $S(t)$ están dados por

$$c_k = \frac{1}{2\pi} \hat{f}(k) = \mathcal{F}^{-1}(f)(-k).$$

En efecto, para ver la expresión (2.2.6), basta probar que

$$\int_0^{2\pi} \sum_k \big| f(t + 2k\pi) \big| dt < \infty.$$

Para la segunda parte se calcula los coeficientes de Fourier

$$\frac{1}{2\pi} \int_0^{2\pi} \Big[\sum_k f(t + 2k\pi) \Big] e^{-ikt} dt.$$

Intercambiando la suma con la integral se obtiene

$$\begin{aligned} \sum_k \frac{1}{2\pi} \int_0^{2\pi} f(t + 2k\pi) e^{-ikt} dt &= \sum_k \frac{1}{2\pi} \int_{2\pi k}^{2\pi(k+1)} f(z) e^{-ikz} dz \\ &= \frac{1}{2\pi} \hat{f}(k). \end{aligned}$$

2.3. Transformadas wavelets

El análisis wavelets es un método de descomposición de una función o señal usando funciones especiales, las wavelets. La descomposición es similar a la de la transformada de Fourier, donde una señal $f(t)$ se descompone en una suma infinita de armónicos $e^{i\omega t}$ de frecuencias $\omega \in \mathbb{R}$, cuyas amplitudes son los valores de la transformada de Fourier de f, $\hat{f}(\omega)$:

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{i\omega t}d\omega, \quad \text{donde} \quad \hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt.$$

El análisis de Fourier tiene el defecto de la no localidad: el comportamiento de una función en un conjunto abierto, no importa cuán pequeño, influye en el comportamiento global de la transformada de Fourier. No se captan los aspectos locales de la señal tales como cambios bruscos, saltos o picos, que se han de determinar a partir de su reconstrucción.

2.3.1. Transformada wavelet continua

La teoría wavelets se basa en la representación de una función en términos de una familia biparamétrica de dilataciones y traslaciones de una función fija ψ, la wavelet madre que, en general, no es senoidal. Por ejemplo,

$$f(t) = \int_{\mathbb{R}^2} \frac{1}{\sqrt{|a|}}\psi\Big(\frac{t-b}{a}\Big)\mathcal{W}_\psi f(a,b)dadb$$

en donde $\mathcal{W}_\psi f$ es una transformada de f definida adecuadamente. También se tiene de modo alterno un desarrollo en serie

$$f(t) = \sum_{j,k} c_{j,k}2^{j/2}\psi(2^j t - k)$$

en donde se suma sobre las dilataciones en progresión geométrica. Para conservar la norma en $L_2(\mathbb{R})$ de la wavelet madre ψ, se insertan los factores $\frac{1}{\sqrt{|a|}}$ y $2^{j/2}$, respectivamente.

Definición 2.3.1. *Una wavelet ψ es una función cuadrado integrable tal que la siguiente condición de admisibilidad se tiene*

$$C_\psi := \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{|\omega|}d\omega < \infty, \tag{2.3.1}$$

donde $\hat{\psi}(\omega)$ es la transformada de Fourier de ψ.

Observación 2.3.1. Si además $\psi \in L_1(\mathbb{R})$, entonces la condición (2.3.1) implica que $\int_{\mathbb{R}} \psi(t)dt = 0$. En efecto, por el Lema de Riemann-Lebesgue (ver p.e., [48]), $\lim_{\omega\to\infty} \hat{\psi}(\omega) = 0$ y la transformada de Fourier es continua, lo cual implica que $0 = \hat{\psi}(0) = \int_{\mathbb{R}} \psi(t)dt$.

Sea $\psi \in L_2(\mathbb{R})$. La función dilatada y trasladada se define por

$$\psi_{a,b}(t) := \frac{1}{\sqrt{|a|}}\psi\Big(\frac{t-b}{a}\Big), \quad a, b \in \mathbb{R}, \quad a \neq 0.$$

Esta función se obtiene a partir de ψ, primero por dilatación en el factor a y, luego, por traslación en b. Es claro que $\|\psi_{a,b}\|_2 = \|\psi\|_2$.

Definición 2.3.2. *Para $f, \psi \in L_2(\mathbb{R})$, la expresión*

$$\mathcal{W}_\psi f(a,b) := \int_{\mathbb{R}} f(t)\overline{\psi_{a,b}(t)}dt \tag{2.3.2}$$

se llama la transformada wavelet de f.

Por la desigualdad de Cauchy, se ve que $\mathcal{W}_\psi f$ es una función acotada con $\big|\mathcal{W}_\psi f(a,b)\big| \leq \|f\|_2\|\psi\|_2$. Note también que

$$\mathcal{W}_\psi f(a,b) = \langle f, \psi_{a,b}\rangle_{L_2(\mathbb{R})} = \langle f, \psi_{a,b}\rangle.$$

La transformada wavelet $\mathcal{W}_\psi f$ de f puede ser descrita en términos del producto de convolución. La convolución de dos funciones $f, g \in L_2(\mathbb{R})$ es dada por

$$(f * g)(t) = \int_{\mathbb{R}} f(t-z)g(z)dz.$$

Observe que esta fórmula está definida para al menos todo $t \in \mathbb{R}$, pero $f * g$ no necesariamente está en $L_2(\mathbb{R})$. Usando la notación $\widetilde{\psi}(t) = \overline{\psi(-t)}$, se tiene $\mathcal{W}_\psi f(a,b) = (f * \widetilde{\psi}_{a,0})(b)$. Note también que $\hat{\tilde{\psi}}_{a,b}(\omega) = \sqrt{|a|}\hat{\tilde{\psi}}(a\,\omega)e^{-i\omega\, b}$. Estos hechos se aplicarán en la prueba de la siguiente proposición, la cual establece la fórmula de Plancherel para la transformada wavelet.

Proposición 2.3.3. *Sea $\psi \in L_2(\mathbb{R})$ y satisface la condición (2.3.1). Entonces para cualquier $f \in L_2(\mathbb{R})$, las siguientes relaciones se tienen*

1. *Isometría*

$$\int_{\mathbb{R}} |f(t)|^2 dt = \frac{1}{C_\psi}\int_{\mathbb{R}^2} \big|\mathcal{W}_\psi f(a,b)\big|^2 db\,\frac{da}{a^2}$$

2. Fórmula de inversión

$$f(t)=\frac{1}{C_\psi}\int_{\mathbb{R}^2}\mathcal{W}_\psi f(a,b)\psi_{a,b}(t)db\,\frac{da}{a^2}$$

Demostración. 1. Es fácil verificar que $(f*\widetilde{\psi}_{a,0})(b)=\sqrt{|a|}\mathcal{F}^{-1}\{\hat{f}(\omega)\tilde{\hat{\psi}}(a\,\omega)\}$. En consecuencia,

$$\begin{aligned}
\int_{\mathbb{R}^2}\left|\mathcal{W}_\psi f(a,b)\right|^2 db\,\frac{da}{a^2} &= \int_{\mathbb{R}}\int_{\mathbb{R}}\left|(f*\widetilde{\psi}_{a,0})(b)\right|^2 db\frac{da}{a^2}\\
&= \int_{\mathbb{R}}\int_{\mathbb{R}}|a|\left|\mathcal{F}^{-1}\Big(\hat{f}(\cdot)\tilde{\hat{\psi}}(a\,\cdot)\Big)(\omega)\right|^2 d\omega\frac{da}{a^2}\\
&= \int_{\mathbb{R}}\int_{\mathbb{R}}\left|\hat{f}(\omega)\right|^2\left|\hat{\psi}(a\,\omega)\right|^2 d\omega\frac{da}{|a|}\\
&= \int_{\mathbb{R}}\left|\hat{f}(\omega)\right|^2\left[\int_{\mathbb{R}}\left|\hat{\psi}(a\,\omega)\right|^2\frac{da}{|a|}\right]d\omega\\
&= C_\psi\int_{\mathbb{R}}\left|\hat{f}(\omega)\right|^2 d\omega=C_\psi\|f\|_2^2.
\end{aligned}$$

Observe que se utilizó el teorema de Fubini y la fórmula de Plancherel para la transformada de Fourier.
2. Para simplificar los cálculos en la fórmula de inversión, suponga que $f,\hat{f}\in L_1(\mathbb{R})$.

$$\begin{aligned}
\int_{\mathbb{R}}\mathcal{W}_\psi f(a,b)\psi_{a,b}(t)db &= \sqrt{|a|}\int_{\mathbb{R}}\mathcal{F}^{-1}\Big(\hat{f}(\cdot)\tilde{\hat{\psi}}(a\,\cdot)\Big)(\omega)\psi_{a,b}(t)d\omega\\
&= \sqrt{|a|}\int_{\mathbb{R}}\hat{f}(\omega)\tilde{\hat{\psi}}(a\,\omega)\mathcal{F}^{-1}(g)(\omega)d\omega,
\end{aligned}$$

donde $g(b):=\psi_{a,b}(t)$. Ahora, la transformada inversa de Fourier de g es

$$\begin{aligned}
\mathcal{F}^{-1}(g)(\omega) &= \frac{1}{2\pi}\int_{\mathbb{R}}g(b)e^{i\omega\,b}db\\
&= \frac{1}{2\pi}\sqrt{|a|}\int_{\mathbb{R}}\psi(z)e^{-ia\omega z}e^{i\omega t}dz\\
&= \frac{1}{2\pi}\sqrt{|a|}\hat{\psi}(a\,\omega)e^{i\omega t}.
\end{aligned}$$

Sustituyendo e integrando respecto a $a^{-2}da$ se obtiene

$$\begin{aligned}\int_{\mathbb{R}^2} \mathcal{W}_\psi f(a,b)\psi_{a,b}(t)db\frac{da}{a^2} &= \frac{1}{2\pi}\int_{\mathbb{R}} |a| \left[\int_{\mathbb{R}} \hat{f}(\omega)\left|\hat{\psi}(a\,\omega)\right|^2 e^{i\omega t}d\omega\right]\frac{da}{a^2}\\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\omega)\left[\int_{\mathbb{R}}\left|\hat{\psi}(a\,\omega)\right|^2\frac{da}{|a|}\right]e^{i\omega t}d\omega\\ &= C_\psi\frac{1}{2\pi}\int_{\mathbb{R}}\hat{f}(\omega)e^{i\omega t}d\omega\\ &= C_\psi f(t).\end{aligned}$$

□

Otro resultado de interés que se presentará en la siguiente proposición, es la fórmula de Parseval para la transformada wavelet.

Proposición 2.3.4. *Sea $\psi \in L_2(\mathbb{R})$ y satisface la condición (2.3.1). Entonces para cualquier $f, g \in L_2(\mathbb{R})$, se tienen*

$$\langle f, g\rangle_{L_2(\mathbb{R})} = \frac{1}{C_\psi}\int_{\mathbb{R}^2}\mathcal{W}_\psi f(a,b)\overline{\mathcal{W}_\psi g}(a,b)\frac{dadb}{a^2}$$

Demostración. Como $(f * \widetilde{\psi}_{a,0})(b) = \sqrt{|a|}\mathcal{F}^{-1}\{\hat{f}(\omega)\hat{\tilde{\psi}}(a\,\omega)\}$ o de manera equivalente, $\mathcal{F}\big(f * \widetilde{\psi}_{a,0}\big)(\omega) = \sqrt{|a|}\hat{f}(\omega)\hat{\tilde{\psi}}(a\,\omega)$, entonces

$$\int_{\mathbb{R}}\mathcal{W}_\psi f(a,b)\overline{\mathcal{W}_\psi g}(a,b)db = |a|\int_{\mathbb{R}}\hat{f}(\omega)\tilde{\hat{g}}(\omega)|\hat{\psi}(a\,\omega)|^2 d\omega,$$

ahora, integrando respecto a $a^{-2}da$ se sigue

$$\begin{aligned}\int_{\mathbb{R}^2}\mathcal{W}_\psi f(a,b)\overline{\mathcal{W}_\psi g}(a,b)db\frac{da}{a^2} &= \int_{\mathbb{R}}|a|\left[\int_{\mathbb{R}}\hat{f}(\omega)\tilde{\hat{g}}(\omega)\left|\hat{\psi}(a\,\omega)\right|^2 d\omega\right]\frac{da}{a^2}\\ &= \int_{\mathbb{R}}\hat{f}(\omega)\tilde{\hat{g}}(\omega)\left[\int_{\mathbb{R}}\left|\hat{\psi}(a\,\omega)\right|^2\frac{da}{|a|}\right]d\omega\\ &= C_\psi\int_{\mathbb{R}}\hat{f}(\omega)\tilde{\hat{g}}(\omega)d\omega\\ &= C_\psi\langle\hat{f},\hat{g}\rangle_{L_2(\mathbb{R})} = C_\psi\langle f, g\rangle_{L_2(\mathbb{R})}.\end{aligned}$$

Note que se aplicó el teorema de Fubini, y en el último renglón de la expresión anterior, la fórmula de Parseval para la transformada de Fourier. □

En la siguiente proposición se listan algunas propiedades.

Proposición 2.3.5. *Sean ψ y φ wavelets y $f, g \in L_2(\mathbb{R})$. Entonces*

1. $\mathcal{W}_\psi(\alpha f + \beta g)(a,b) = \alpha \mathcal{W}_\psi f(a,b) + \beta \mathcal{W}_\psi g(a,b)$, $\alpha, \beta \in \mathbb{R}$.
2. $\mathcal{W}_{\alpha\psi+\beta\varphi} f(a,b) = \bar{\alpha}\mathcal{W}_\psi f(a,b) + \bar{\beta}\mathcal{W}_\varphi f(a,b)$, $\alpha, \beta \in \mathbb{R}$.
3. $\mathcal{W}_\psi(T_c f)(a,b) = \mathcal{W}_\psi f(a, b-c)$, *donde T_c es el operador traslación definido por $T_c f(t) = f(t-c)$.*
4. $\mathcal{W}_\psi(D_c f)(a,b) = \sqrt{c}\,\mathcal{W}_\psi f(c\,a, c\,b)$, *donde D_c es el operador dilatación definido por $D_c f(t) = \sqrt{c} f(c\,t)$.*

2.3.2. Transformada wavelet discreta

La transformada wavelet continua introduce cierta redundancia, pues la señal original se puede reconstruir completamente calculando $\mathcal{W}_\psi f(a,\cdot)$ para una cantidad numerable de escalas, por ejemplo, potencias enteras de 2. Esto es, si se elige la escala $a = 2^{-j}$ para cada $j \in \mathbb{Z}$, y también se discretiza en el dominio del tiempo en los puntos $b = 2^{-j}k$, $k \in \mathbb{Z}$, la familia de wavelets será ahora dada por

$$\psi_{2^{-j},2^{-j}k}(t) = \frac{1}{\sqrt{2^{-j}}}\psi\Big(\frac{t-2^{-j}k}{2^{-j}}\Big) = 2^{j/2}\,\psi(2^j t - k), \quad \forall j,k \in \mathbb{Z}.$$

Se utilizará la notación ψ_{jk} para denotar la wavelet ψ comprimida 2^j y trasladada el entero k, es decir, $\psi_{jk}(t) = 2^{j/2}\,\psi(2^j t - k)$.

Con la elección de $a = 2^{-j}$ y $b = 2^{-j}k$, observe que el muestreo en el tiempo se ajusta proporcionalmente a la escala, es decir, a mayor escala se toma puntos más distantes, ya que se busca información global, mientras que a menor escala se buscan detalles de la señal, por tal motivo se muestrea en puntos menos distantes entre si. Para otras elecciones de a y b se puede consultar [12].

Definición 2.3.6. *Una función $\psi \in L_2(\mathbb{R})$ es una wavelet si la familia de funciones ψ_{jk} definidas por*

$$\psi_{jk}(t) = 2^{j/2}\,\psi(2^j t - k), \quad \forall j,k \in \mathbb{Z}, \tag{2.3.3}$$

es una base ortonormal en el espacio $L_2(\mathbb{R})$.

Una condición suficiente para la reconstrucción de una señal f es que la familia de dilatadas y trasladadas ψ_{jk} forme una base ortonormal en el espacio $L_2(\mathbb{R})$, ver [19] y [29] para más detalles. Si esto se tiene, cualquier función $f \in L_2(\mathbb{R})$ se puede escribir como

$$f(t) = \sum_{j,k} c_{j,k}\psi_{jk}(t) \tag{2.3.4}$$

o teniendo en cuenta (2.3.3) como

$$f(t) = \sum_{j,k} c_{j,k}2^{j/2}\psi(2^j t - k),$$

donde $c_{j,k} = \langle f, \psi_{2^{-j},2^{-j}k}\rangle = \mathcal{W}_\psi f(2^{-j}, 2^{-j}k)$.

Definición 2.3.7. *Para cada $f \in L_2(\mathbb{R})$ el conjunto bidimensional de coeficientes*

$$c_{j,k} = \langle f, \psi_{jk}\rangle = \int_{\mathbb{R}} 2^{j/2} f(t)\overline{\psi(2^j t - k)}dt$$

se llama la transformada wavelet discreta de f.

En consecuencia, la expresión (2.3.4) se puede escribir en forma alterna como

$$f(t) = \sum_{j,k}\langle f(t), \psi_{jk}(t)\rangle\psi_{jk}(t). \tag{2.3.5}$$

La serie (2.3.5) se llama representación wavelet de f.

Observación 2.3.2. $\psi_{jk}(t)$ es muy apropiada para representar detalles más finos de la señal como oscilaciones rápidas. Los coeficientes wavelet $c_{j,k}$ miden la cantidad de fluctuaciones sobre el punto $t = 2^{-j}k$ con una frecuencia determinada por el índice de dilatación j.
Es interesante notar que $c_{j,k} = \mathcal{W}_\psi f(2^{-j}, 2^{-j}k)$ es la transformada wavelet de f en el punto $(2^{-j}, 2^{-j}k)$. Estos coeficientes analizan la señal mediante la wavelet madre ψ.

CAPÍTULO 3

Panel de Datos

En este capítulo se presentarán los conceptos sobre Panel de datos, sus ventajas y desventajas, los test que determinan el modelo econométrico a utilizar (efectos fijos ó efectos aleatorios) y algunos aspectos metodológicos de la técnica de panel; los cuales dan el fundamento teórico para el análisis del caso de aplicación que se desarrollará en el capítulo 4.

Los modelos usados en el análisis económico se pueden clasificar según los datos utilizados y según las relaciones supuestas entre las variables que intervienen en éstos.

Cuando se realizan estudios económicos, en el análisis de la información pueden existir, entre otras, la dimensión temporal y la dimensión estructural. La primera hace referencia al análisis de series de tiempo, que incorpora información de las variables de estudio en un periodo de tiempo determinado. La segunda representa el análisis de la información para las unidades individuales de estudio restringidas en un momento determinado del tiempo. De las anteriores, interesa obtener conclusiones que se deriven de los modelos estimados y que proporcionen relaciones de causalidad o de comportamiento entre diferentes clases de variables a partir de los datos suministrados.

3.1. Panel de datos

Definición 3.1.1. *Un panel de datos es un conjunto de datos que combina series temporales con unidades de sección cruzada o de corte transversal (países, ciudades, bancos, regiones, empresas, hogares, etc).*

Poder capturar la heterogeneidad no observable, ya sea entre agentes económicos o en el tiempo, es el principal objetivo de la técnica de panel de datos. Cabe destacar que la heterogeneidad en el tiempo no se puede detectar con estudios de series de tiempo y de corte transversal por separado. Además, cuando se realizan estudios en microeconomía se resaltan dos aspectos relevantes: a) los efectos individuales específicos y b) los efectos temporales, los cuales generan heterogeneidad no observable.

Entre las ventajas que presentan los panel de datos se tienen: el control sobre la heterogeneidad individual; más variabilidad, menos colinearidad entre las variables, más grados de libertad y mayor eficiencia; mejor adecuación al estudio de las dinámicas de ajuste; mejor capacidad de identificar y medir efectos que no son detectables en datos puros de sección cruzada o de series temporales y también mejor capacidad de análisis en comportamientos más complicados. Como desventajas, los datos de panel presentan el problema de recolección de datos, distorsiones por errores de medida y la corta dimensión temporal que se tiene generalmente en los conjuntos de datos (veáse [11], [5], [2]).

Teniendo en cuenta sus limitaciones, y a pesar de las ventajas que presentan, cuando nos enfrentamos al análisis de un panel de datos, existe gran cantidad de cuestiones que cabe plantearse a la hora de mantener determinados supuestos y de elegir un método de estimación, para poder así dar un mayor soporte al estudio que se está realizando.

3.2. Modelo general de un panel de datos

La estructura básica para un panel de datos es un modelo de regresión lineal de la forma

$$y_{it} = \mathbf{x}'_{it}\beta + \mathbf{z}'_i\alpha + \varepsilon_{it} \tag{3.2.1}$$

donde $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$. Acá i es la unidad de estudio (corte transversal), t se refire a la dimensión en el tiempo, β es un vector de k parámetros y $\mathbf{x}_{it}$ es la i-ésima observación al momento t para las k

variables explicativas. La heterogeneidad o efecto individual es $\mathbf{z}_i'\alpha$, donde $\mathbf{z}_i$ está conformada por constantes y un conjunto de individuos o un grupo específico de variables, los cuales pueden ser observables (p.e. género, raza, etc.) o no obsevables (p.e. características específicas de familias, destrezas, gustos, etc.) todas invariantes en el tiempo t. Si $\mathbf{z}_i$ es observable para todos los individuos, entonces el modelo se reduce a un modelo de regresión lineal clásico (ver [27]). En caso contrario, se tienen:

1. Regresión Total: Si $\mathbf{z}_i$ contiene solamente términos constantes, el método de mínimos cuadrados ordinarios genera estimadores consistentes y eficientes para α y el vector de pendientes β.

2. Efectos Fijos: Cuando $\mathbf{z}_i$ sea no observable y esté correlacionada con $\mathbf{x}_{it}$, entonces el estimador de mínimos cuadrados para β será inconsistente. Sin embargo el modelo

$$y_{it} = \mathbf{x}_{it}'\beta + \alpha_i + \epsilon_{it} \tag{3.2.2}$$

donde $\alpha_i = \mathbf{z}_i'\alpha$, representa todos los efectos observables. Debe hacerse notar que en el presente se da una pérdida importante de grados de libertad.

3. Efectos Aleatorios: Este modelo considera que los efectos individuales no son independientes entre sí, sino que están distribuidos aleatoriamente alrededor de un valor dado. Una práctica común en el análisis de regresión es asumir que el gran número de factores que afecta el valor de la variable dependiente pero que no han sido excluidas explícitamente como variables independientes del modelo, pueden resumirse apropiadamente en la perturbación aleatoria. El modelo puede ser formulado como:

$$y_{it} = \mathbf{x}_{it}'\beta + E[\mathbf{z}_i'\alpha] + \{\mathbf{z}_i'\alpha - E[\mathbf{z}_i'\alpha]\} + \varepsilon_{it}$$

$$y_{it} = \mathbf{x}_{it}'\beta + \alpha + \mu_i + \varepsilon_{it} \tag{3.2.3}$$

Estos efectos aleatorios se aproxima a especificar que μ_i es un elemento aleatorio de un grupo específico, similar a ε_{it} excepto que para cada grupo hay una gráfica que representa identicamente la regresión para cada periodo. El investigador hace inferencia condicional o marginal respecto a una población.

4. Test de Especificación de Hausman: Esta prueba permite determinar qué modelo es el más adecuado para el panel de datos que se está analizando, si es el de efectos fijos o de efectos aleatorios. El test de Hausman se utiliza para analizar la posible correlación entre los α_i y los regresores. Se basa en la idea que bajo la hipótesis de no correlación, los modelos OLS, LSDV y GLS [1] son consistentes, pero el OLS es ineficiente, mientras que en la hipotesis alternativa, el OLS es consistente, pero el GLS no lo es. Por lo tanto, bajo la hipótesis nula, los dos estimadores difieren sistematicamente, y el test puede basarse sobre sus diferencias. Otro ingrediente esencial para el test es la matrix de covarianza de el vector diferencia, $[b - \hat{\beta}]$:

$$Var[b - \hat{\beta}] = Var[b] + Var[\hat{\beta}] - 2Cov[b, \hat{\beta}] \tag{3.2.4}$$

El resultado esencial de Hausman es que la covarianza de un estimador eficiente y la diferencia del estimador ineficiente, es cero, lo cual implica que

$$Cov[(b - \hat{\beta}), \hat{\beta}] = Cov[b, \hat{\beta}] - Var[\hat{\beta}] = 0$$

o que

$$Cov[b - \hat{\beta}] = Var[\hat{\beta}]$$

reemplazando este resultado en (3.2.4) la matrix de covarianza requerida para el test,

$$Var[b - \hat{\beta}] = Var[b] - Var[\hat{\beta}] = \Psi. \tag{3.2.5}$$

El test χ^2 se basa en el criterio de Wald:

$$W = \chi^2[K-1] = [b - \hat{\beta}]'\hat{\Psi}^{-1}[b - \hat{\beta}].$$

Para $\hat{\Psi}$, usamos la matrix de covarianza del estimador de pendientes en el modelo LSDV y en el modelo de efecto aleatorio, excluyendo el término constante. Bajo la hipótesis nula, W tiene una distribución límite χ^2 con $K - 1$ grados de libertad.

[1] OLS: Mínimos Cuadrados Ordinarios, LSDV: Mínimo Cuadrados de Variable Dummy, y GLS: Minimos Cuadrados Generalizados

3.3. Criterios para la selección del modelo

Cuando el investigador quiere hacer inferencia debe decidir si va a trabajar respecto a las características de la población o sobre los efectos que se encuentran en la muestra. Si decide trabajar sobre una muestra aleatoria; es decir, hacer inferencias sobre una población, la estructura apropiada para su análisis es de *tipo aleatorio*. Mientras que si toma una muestra seleccionada a conveniencia, el modelo de *efectos fijos* será el apropiado.

Además, si el objetivo del estudio se centra en los coeficientes de las pendientes de los parámetros y no en las diferencias individuales, se debe elegir un modelo que las elimine y que trabaje la heterogeneidad no observable como aleatoria (incorporándolas en el término de error), lo que modifica la varianza del modelo, mientras que en el modelo de efectos fijos la heterogeneidad no observable se incorpora en la ordenada del modelo.

Otro factor que afecta la selección del modelo radica en el tamaño de las dimensiones, tanto temporal como estructural. Cuando t es pequeño y N grande los resultados obtenidos por los dos modelos difieren sustancialmente, además se genera gran cantidad de parámetros de efectos fijos respecto al número de datos disponibles, quienes cuentan con parámetros poco confiables y una estimación ineficiente.

3.4. Modelo de efectos fijos

El modelo (3.2.2) se puede escribir, como

$$y_{it} = \mathbf{i}\alpha_i + \mathbf{X}_i\beta + \epsilon_{it},$$

suponiendo que el término α_i contiene las diferencias entre unidades y debido a ello, dicho parámetro debe ser estimado. En términos matriciales, tenemos:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \mathbf{i} & 0 & \dots & 0 \\ 0 & \mathbf{i} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{i} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

o

$$\mathbf{y} = \begin{bmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \dots & \mathbf{d}_n & \mathbf{X} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \epsilon$$

donde $\mathbf{d}_i$ es una variable dummy[2] que indica la i-ésima unidad. Reuniendo las nT filas se obtiene

$$\mathbf{y} = \mathbf{D}\alpha + \mathbf{X}\beta + \epsilon,$$

con $\mathbf{D} = \begin{bmatrix} d_1 & d_2 & \dots & d_n \end{bmatrix}_{nT\times n}$. Este modelo se denomina *mínimos cuadrados de variables ficticias*, MCVF.

Algunos supuestos necesarios para el modelo de efectos fijos, son: Sea

$$\{(y_{i1}, \dots, y_{iT}, x_{i1}, \dots, x_{iT}, \eta_i), i = 1, \dots, N\}$$

una muestra aleatoria y

$$y_{it} = x'_{it}\beta + \eta_i + \upsilon_{it}$$

el modelo. Además,

1. *Supuesto Uno*:
$$\mathrm{E}(\upsilon_i|x_i, \eta_i) = 0 \quad (t = 1, ..., T),$$
donde $\upsilon_i = (\upsilon_{i1}, ..., \upsilon_{iT})'$ y $x_i = (x_{i1}, ..., x_{iT})'$. Tanto y_{it} como el vector $k \times 1$ de variables explicativas x_{it} son observables, mientras η_i es un regresor no observado invariante en el tiempo.

2. *Supuesto Dos*:
$$Var(\upsilon_i|x_i, \eta_i) = \sigma^2 I_T.$$
Bajo este supuesto los errores son condicionalmente homoscedásticos y no serialmente correlacionados. Bajo el supuesto Uno, tenemos:
$$E(y_i|x_i, \eta_i) = X_i\beta + \eta_i\iota, \tag{3.4.1}$$
donde $y_i = (y_{i1}, ..., y_{iT})'$, ι es un vector $T \times 1$ de unos y $X_i = (x_{i1}, ..., x_{iT})'$ es una matrix $T \times k$. La implicación de (3.4.1) para el valor esperado de y_i dado x_i es
$$E(y_i|x_i) = X_i\beta + E(\eta_i|x_i)\iota. \tag{3.4.2}$$

[2] Una variable dummy, binaria o ficticia es aquella que toma valor de 1 para algunas observaciones indicando la presencia de un efecto sobre miembros de un grupo y 0 para el resto de observaciones.

Sin embargo, bajo el supuesto Dos

$$Var(y_i|x_i, \eta_i) = \sigma^2 I_T \tag{3.4.3}$$

Lo cual implica que

$$Var(y_i|x_i) = \sigma^2 I_T + Var(\eta_i|x_i)\iota\iota' \tag{3.4.4}$$

3. *Supuesto Tres*:

$$E(\upsilon_i|x_i) = 0 \quad (t = 1, 2, ..., T).$$

4. *Supuesto Cuatro*:

$$Var(\upsilon_i|x_i) = \sigma^2 I_T.$$

Frecuentemente se utiliza $E(\upsilon_{it}|x_i) = 0$ a pesar de ser una suposición débil, sin embargo es conveniente hacerlo pues en las aplicaciones se dificultaría imaginar como $E(\upsilon_{it}|x_i) = 0$ tiende hacia $E(\upsilon_{it}|x_i, \eta_i) = 0$.

3.5. Contraste de significatividad de los efectos de grupo

La razón t habitual para a_i implica un contraste de la hipótesis de que α_i es igual a cero. Pero, normalmente, esta hipótesis no es útil en un contexto de regresión. Si estamos interesados en las diferencias entre grupos, podemos contrastar la hipótesis de que los términos constantes son todos iguales, mediante un contraste F. Bajo la hipótesis nula, el estimador eficiente coincide con mínimos cuadrados agrupados. La razón F utilizada para el contraste es

$$F(n-1, nT-n-K) = \frac{(R_u^2 - R_p^2)/(n-1)}{(1-R_u^2)/(nT-n-K)} \tag{3.5.1}$$

donde u indica el modelo no restringido y p indica el modelo agrupado, o restringido, con un único término constante para todos. (Se puede utilizar también la suma de errores al cuadrado, si resulta más conveniente). Si fuese más cómodo, también podría estimarse el modelo con una única constante y $n-1$ variables Dummys. Los demás resultados no cambian, y en vez de estimar α_i , cada coeficiente de las variables Dummys será una estimación de $\alpha_i - \alpha_1$. El contraste F de que los coeficientes de las $n-1$ variables dummys son cero es idéntico al anterior. Es importante tener presente que, aunque los resultados estadísticos sean los mismos, la interpretación de los coeficientes de las variables dummys en las dos formulaciones son diferentes.

3.6. Los estimadores intra y entre grupos

Podemos formular el modelo de regresión de las siguientes tres formas. Primero, la formulación original es

$$y_{it} = \alpha_i + \beta' x_{it} + \epsilon_{it}. \tag{3.6.1}$$

En términos de desviaciones de las medias del grupo,

$$y_{it} - \bar{y}_{i\cdot} = \beta'(x_{it} - \bar{x}_{i\cdot}) + \epsilon_{it} - \bar{\epsilon}_{i\cdot}, \tag{3.6.2}$$

mientras que en términos de las medias de grupo,

$$\bar{y}_{i\cdot} = \alpha_i + \beta' \bar{x}_{i\cdot} + \bar{\epsilon}_{i\cdot}. \tag{3.6.3}$$

Los tres son modelos de regresión clásica y, en principio, los tres podrían ser estimados, al menos consistentemente, aunque no eficientemente, por mínimos cuadrados ordinarios. Consideremos, entonces, las matrices de sumas de cuadrados y productos cruzados que se utilizarían en cada caso, donde nos centraremos solamente en la estimación de β. En (3.6.1), los momentos serían sobre las medias totales, $\bar{\bar{y}}$ y $\bar{\bar{x}}$, y utilizaríamos las sumas totales de cuadrados y productos cruzados,

$$S'_{xx} = \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{\bar{x}})(x_{it} - \bar{\bar{x}})'$$

y

$$S'_{xy} = \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{\bar{x}})(y_{it} - \bar{\bar{y}})$$

Para (3.6.2), como los datos están ya en desviaciones, las medias de $(x_{it} - \bar{x}_{i\cdot})$ y $(y_{it} - \bar{y}_{i\cdot})$ son cero. las matrices de momentos son sumas de cuadrados y productos cruzados intra-grupos (es decir, desviaciones de las medias de los grupos),

$$S_{xx}^{w} = \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})'$$

y

$$S_{xy}^{w} = \sum_{i=1}^{n} \sum_{t=1}^{T} (x_{it} - \bar{x}_{i\cdot})(y_{it} - \bar{y}_{i\cdot})'.$$

Finalmente, para (3.6.3), las medias de las medias de los grupos es la media total. Las matrices de momentos son las sumas de cuadrados y productos cruzados entre-grupos.

$$S_{xx}^{b} = \sum_{i=1}^{n} T(\bar{x}_{i\cdot} - \bar{\bar{x}})(\bar{x}_{i\cdot} - \bar{\bar{x}})'$$

y

$$S_{xy}^{b} = \sum_{i=1}^{n} T(\bar{x}_{i\cdot} - \bar{\bar{x}})(\bar{y}_{i\cdot} - \bar{\bar{y}}).$$

Es fácil comprobar que

$$S_{xx}^{\prime} = S_{xx}^{w} + S_{xx}^{b} \quad \text{y} \quad S_{xy}^{\prime} = S_{xy}^{w} + S_{xy}^{b}.$$

Hay, por lo tanto, tres posibles estimadores de mínimos cuadrados de β, que corresponden a la descomposición analizada. El estimador de mínimos cuadrados es

$$\mathbf{b}' = [S_{xx}^{\prime}]^{-1} S_{xy}^{\prime} = [S_{xx}^{w} + S_{xx}^{b}]^{-1} [S_{xy}^{w} + S_{xy}^{b}]. \qquad (3.6.4)$$

El estimador **intra-grupos** es

$$\mathbf{b}^{w} = [S_{xx}^{w}]^{-1} S_{xy}^{w}.$$

Este es el estimador MCVF. Un estimador alternativo sería el estimador **entre-grupos**,

$$\mathbf{b}^{b} = [S_{xx}^{b}]^{-1} S_{xy}^{b}.$$

Este es el estimador de mínimos cuadrados de (3.6.3) en los n conjuntos de medias de grupos. De la expresión anterior

$$S_{xy}^{w} = S_{xx}^{w} \mathbf{b}^{w} \quad \text{y} \quad S_{xy}^{b} = S_{xx}^{b} \mathbf{b}^{b}.$$

Insertando estos resultados en (3.6.4), vemos que el estimador de MCO es un promedio ponderado matricialmente, de los estimadores intra y entre grupos:

$$\mathbf{b}^t = \mathbf{F}^w b^w + \mathbf{F}^b b^b,$$

donde

$$\mathbf{F}^w = [S_{xx}^w + S_{xx}^b]^{-1} S_{xx}^w = \mathbf{I} - \mathbf{F}^b.$$

3.7. Paneles no balanceados y efectos fijos

Los paneles en que los tamaños de grupos difieren son comunes y se conocen como *paneles no balanceados*. Las modificaciones necesarias para permitir tamaños desiguales, son: el tamaño muestral completo es $\sum_{i=1}^n T_i$ en vez de nT, y las medias de los grupos deben basarse en T_i, que varía entre los grupos. Las medias totales para los regresores son

$$\bar{\bar{x}} = \frac{\sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T_i} x_{it}}{\sum\limits_{i=1}^{n} T_i} = \frac{\sum\limits_{i=1}^{n} T_i \bar{x}_{i\cdot}}{\sum\limits_{i=1}^{n} T_i} = \sum_{i=1}^{n} w_i \bar{x}_{i\cdot},$$

donde $w_i = T_i/(\sum_{i=1}^n T_i)$. Si los grupos son de igual tamaño, $w_i = 1/n$, la matriz de momentos

$$\mathbf{S}_{xx}^w = \mathbf{X}'\mathbf{M_d}\mathbf{X}$$

es una suma de matrices de sumas de cuadrados y productos cruzados,

$$\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{M}_i^o \mathbf{X}_i = \sum_{i=1}^{n} \left(\sum_{t=1}^{T_i} (x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})' \right)$$

sumadas a través de los grupos, denominada la suma de cuadrados **intra-grupos**.

3.8. Efectos aleatorios

Dentro algún contexto puede ser más apropiado interpretar los términos constantes específicos de la unidad, como distribuidos alaeatoriamente entre las unidades de sección cruzada. Esto es apropiado si creemos que las unidades de sección cruzada de la muestra son extracciones muestrales de una población grande. Retomando (3.2.3), el análisis de familias, se puede interpretar como el conjunto de factores, no incluidos en la regresión, que son específicos en esa familia y además que

$$\begin{aligned} E[\epsilon_{it}] &= E[\mu] = 0, \\ E[\epsilon_{it}^2] &= \sigma_\epsilon^2, \\ E[u_i^2] &= \epsilon_u^2, \\ E[\epsilon_{it} u_j] &= \sigma_\epsilon^2, \\ E[\epsilon_{it}\epsilon_{js}] &= 0; \;\; \text{si } t \neq s \text{ o } i \neq j, \\ E[u_i u_j] &= 0; \;\; \text{si } i \neq j. \end{aligned} \tag{3.8.1}$$

Reescribiendo (3.2.3) en bloques de T observaciones, tenemos:

$$w_{it} = \epsilon_{it} + u_i \quad \text{y} \quad w_i = [w_{i1}, w_{i2}, \ldots, w_{iT}]',$$

el cual se denomina *modelo de componentes del error*. Por consiguiente,

$$E[w_{it}^2] = \sigma_\epsilon^2 + \sigma_u^2,$$

$$E[w_{it} w_{is}] = \sigma_u^2, \quad t \neq s.$$

Para las T observaciones de la unidad i, sea $\mathbf{\Omega} = E[w_i w_i']$. Entonces,

$$\Omega = \begin{bmatrix} \sigma_\epsilon^2 + \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\epsilon^2 + \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_\epsilon^2 + \sigma_u^2 \end{bmatrix} = \sigma_\epsilon^2 \mathbf{I} + \sigma_u^2 \mathbf{ii}', \tag{3.8.2}$$

donde $\mathbf{i}$ es un vector columna $T \times 1$ de unos. Como las observaciones i y j son independientes, la matriz de varianzas y covarianzas de los errores para nT observaciones, es

$$\mathbf{V} = \begin{bmatrix} \mathbf{\Omega} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Omega} \end{bmatrix}$$

3.9. Heterocedasticidad

El problema de heteroscedasticidad se presenta cuando es violado el supuesto de varianza constante de los errores de la función de regresión. La heteroscedasticidad tiene que ver con la relación entre una o más de las variables independientes del modelo y el cuadrado de los errores estimados a partir de la regresión. Este problema se manifiesta en un crecimiento o decrecimiento de la varianza del modelo.

La presencia de heteroscedasticidad es muy común en regresiones estimadas a partir de datos de corte transversal. Por ejemplo, cuando se recolectan datos provenientes de estratos, de regiones, por tamaño de la familia o por tipo de empresa. En general, puede presentasen estudios que incluyen grupos con comportamientos marcados a lo largo de toda la muestra; por ejemplo, la variable ingreso monetario del hogar según el estrato, pues se puede pensar que la varianza del ingreso monetario del grupo de alta riqueza es más alta que la del grupo de escasos recursos.

El problema de heteroscedasticidad repercute directamente sobre la estimación de los parámetros de la regresión. Los estimadores seguirán siendo insesgados y consistentes pero no eficientes. La heteroscedasticidad causa la subestimación o sobre estimación de la varianza del modelo de regresión, por lo tanto el valor del error estándar de los parámetros, el valor de los estadísticos t y los intervalos de confianza cambian con respecto a los resultados que deberían obtenerse en ausencia de heteroscedasticidad. En este sentido, la presencia de heteroscedasticidad en el modelo de regresión hace que las pruebas de hipótesis no tengan validez estadística o que las inferencias sean erróneas.

3.9.1. Detección de la heterocedasticidad

A continuación se presentan los métodos para detectar la existencia de heteroscedaticidad:

1. **Análisis de residuales**: Este método permite evaluar gráficamente si existe heteroscedasticidad causada por una variable independiente en particular o por todo el conjunto de variables independientes. Para el primer caso se elabora un diagrama de dispersión entre x_t y ϵ_t^2 (cuadrado del término de error) donde x_t es el regresor que el investigador supone genera la heteroscedaticidad. En el segundo caso,

se construye el diagrama de dispersión entre y_t estimado y ϵ^2. Si estas gráficas muestran alguna tendencia específica, puede afirmarse que existe heteroscedasticidad en el modelo de regresión. No obstante esta metodología es indicativa y no esta basada en una prueba estadística.

2. **Análisis de regresión**: Es la utilización de una o más regresiones auxiliares. La regresión no se estima entre las variables independientes, sino entre el cuadrado del término de error y el conjunto de regresores del modelo original. Dentro de este método se encuentran las pruebas de Park, White, Glejser, Breusch-Pagan-Godfrey, y Golfeld-Quandt.

3.10. Autocorrelación

El problema de autocorrelación se presenta en una regresión cuando los errores de las diferentes observaciones están relacionados en el tiempo. Esto indica que el efecto de los errores en el tiempo no es instantáneo sino por el contrario es persistente en el tiempo. La autocorrelación es más común en series ordenadas en el tiempo que en información proveniente de encuestas en un tiempo fijo (sección cruzada). La autocorrelación puede estar relacionada con los ciclos económicos; generalmente ésta se presenta en un modelo con variables macroeconómicas donde en el tiempo ocurre un evidente comportamiento tendencial.

Otra causa de la autocorrelación es la presencia de sesgo de especificación en el modelo; principalmente por omisión de variables importantes, las cuales pasan a formar parte del error de la regresión. La autocorrelación puede ser también generada en casos donde se usa una forma funcional incorrecta del modelo, esto hace que los datos se ajusten a una forma funcional que no es la más adecuada. Se argumenta, que la manipulación de información puede llegar a generar también autocorrelación. Un caso típico se presenta en las cuentas nacionales, donde muchos datos son obtenidos a partir de otros, aplicando técnicas de interpolación o extrapolación. Por ejemplo, cuando se convierten datos diarios a semanales. Finalmente, modelos especiales como los de rezagos distribuidos y los autoregresivos pueden originar autocorrelación. Entre las consecuencias de la autocorrelación se tiene la sobreestimación o subestimación de los estadísticos t que juzgan la significancia de las variables independientes en el modelo. Aunque los estimadores siguen siendo insesgados y consistentes son ineficientes. En este sentido se afecta la validez estadística de las pruebas de hipótesis.

3.10.1. Detección de la autocorrelación

Los métodos más comunes para detectar autocorrelación son:

1. **Análisis de residuales**: este método plantea la construcción de diagramas de dispersión para los errores en función de tiempo o en función de un período inmediatamente anterior. El primer paso es estimar el modelo original por MCO. Luego los errores estimados de la regresión son graficados en un eje de coordenadas para identificar si existe alguna tendencia de los mismos en el tiempo, o de estos con su primer rezago.

2. **El estadístico de Durbin-Watson**: Esta prueba es válida para aplicar en errores que se modelan como un proceso autoregresivo de orden 1 "AR(1)", como el mostrado a continuación:

   $$\epsilon_t = \rho\epsilon_{t-1} + \nu_t$$

   El estadístico d oscila entre 0 y 4. Si este se aproxima a 0, se dice que existe autocorrelación positiva (relación directa entre los errores), por el contrario si d se aproxima a 4, existe autocorrelación negativa (relación inversa entre los errores). El Durbin-Watson (d) se estima de la siguiente manera:

   $$d = \frac{\sum_{t=2}^{n}(\epsilon_t - \epsilon_{t-1})}{\sum_{t=1}^{n}\epsilon_t^2} = 2(1-\hat{\rho}), \tag{3.10.1}$$

   donde $\hat{\rho}$ es el coeficiente de autocorrelacion de orden 1, el cual puede despejarse directamente de (3.10.1),

   $$\hat{\rho} = 1 - \frac{d}{2}.$$

La hipótesis planteada es entonces

$$H_o : \rho_{\epsilon_t,\epsilon_{t-1}} = 0, \text{(no existe correlación entre los errores)}$$
$$H_a : \rho_{\epsilon_t,\epsilon_{t-1}} \neq 0, \text{(existe correlación entre los errores).}$$

El estadístico Durbin-Watson puede ser comparado con su respectivo tabulado, teniendo en cuenta el número de observaciones contenidas en la muestra y el numero de regresores. Se debe tener en cuenta que d es utilizado para identificar solo autocorrelación de orden 1 siempre y cuando el modelo tenga intercepto. Además no puede usarse en el caso de modelos autorregresivos.

Prueba de Breusch-Godfrey. Esta es una prueba similar a la prueba de White. Se diferencia de ésta en que la variable dependiente de la regresion auxiliar es el término de error ϵ_t y los regresores sus respectivos rezagos hasta el orden deseado por el investigador. Adicionalmente son incluidos los regresores usados en el modelo original. La hipótesis nula corresponde a que todos los coeficientes de autocorrelación de orden (los coeficientes que acompañan a los residuos rezagados en la regresión auxiliar) son iguales a cero, mientras la hipótesis alterna es que al menos uno de ellos es distinto de cero.

El estadístico de prueba es $(n-s)R^2 \sim \chi^2_s$, donde s es el número de errores rezagados en la regresión auxiliar. Para probar autocorrelación de orden uno, que es la práctica más común, s será igual a uno. La hipótesis nula es rechazada cuando $(n-s)R^2 > \chi^2_s$ a un nivel de significancia α; en este caso se concluye que hay autocorrelacion (veáse [70], [27]).

CAPÍTULO 4

Caso de Aplicación

En este capítulo se aplican los test W_1 y W_2 para datos reales. El trabajo con datos hipotéticos fue realizado por los doctores Yongmiao Hong y Chihwa Kao, publicado en la revista *Econométrica*, (ver [30]). Los residuales (datos reales) los calculamos a partir del modelo obtenido por el magister Francisco I. Zuluaga en su tesis de maestría sobre la demanda de dinero (ver [72]). Posteriormente se presentan los valores calculados de los test $\hat{W}_1$ y $\hat{W}_2$ obtenidos mediante la función *wavetest(resid,N,T,J,W)* programada en Matlab, lo que permitió determinar la existencia o no de correlación serial entre los componentes de error $\{\varepsilon_{it}\}$.

4.1. Datos hipotéticos, simulación Monte Carlo.

El artículo *"Wavelet-based testing for correlation of unknown form in panel models"* escrito por los doctores Hong y Kao, (ver [30]), presenta una clase de test robusto y consistente basados en wavelet para detectar correlación serial en modelos de panel estáticos o dinámicos; es decir, busca probar si los procesos de error $\{\epsilon_{it}\}$ son serialmente correlacionados. Las hipótesis de interés son:

$$H_0 : cov(\epsilon_{it}, \epsilon_{it-|h|}) = 0, \text{para todo } h \neq 0 \text{ y todo } i. \tag{4.1.1}$$
$$H_a : cov(\epsilon_{it}, \epsilon_{it-|h|}) \neq 0, \text{al menos para } h \neq 0 \text{ y algún } i. \tag{4.1.2}$$

Los nuevos test presentados para dicho propósito, son:

$$\hat{W}_1 = \left(\sum_{i=1}^{n} 2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \hat{\alpha}_{ijk}^2 - \hat{M} \right) / \hat{V}^{\frac{1}{2}} \tag{4.1.3}$$

y

$$\hat{W}_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(2\pi T_i \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \hat{\alpha}_{ijk}^2 - \hat{M} \right) / V_{i0}^{\frac{1}{2}} \tag{4.1.4}$$

Donde:

$\hat{\alpha}_{ijk} \equiv (2\pi)^{-1/2} \sum_{h=1-T_i}^{T_i-1} \hat{R}_i(h)\hat{\Phi}_{jk}^*(h),$[1]

$\hat{M} \equiv \sum_{i=1}^{n} \hat{R}_i^2(0) M_{i0}, \quad \hat{V} \equiv \sum_{i=1}^{n} \hat{R}_i^2(0) V_{i0},$[2]

$M_{i0} \equiv \sum_{h=1}^{T_i-1} (1 - h/T_i) b_{J_i}(h, h),$

$V_{i0} \equiv 4 \sum_{h=1}^{T_i} \sum_{m=1}^{T_i} (1 - h/T_i)(1 - m/T_i) b_{J_i}^2(h, m),$

$a_J(h, m) \equiv \sum_{j=0}^{J_i} \sum_{k=1}^{2^j} \hat{\Psi}_{jk}(h)\hat{\Psi}_{jk}^*(m), \ b_J(h, m) \equiv 2Re[a_J(h, m) + a_J(h, -m)],$[3]

$\Psi_{ij}(\omega) = (2\pi)^{-1/2} \sum_{h=-\infty}^{\infty} \hat{\Psi}_{jk}(h) e^{ih\omega},$[4]

Intuitivamente, $\hat{W}_2$ puede ser visto como un test de heterocedasticidad corregida mientras que $\hat{W}_1$, como un test de heterocedasticidad consistente, donde la heterocedasticidad surge para diferentes varianzas σ_i^2 y escalas finas J_i. En $\hat{W}_2$, estas dos formas de heterocedasticidad se corrigen para cada i. A partir de lo anterior, $\hat{W}_1$ y $\hat{W}_2$ distribuyen asintóticamente $N(0, 1)$ bajo

[1] $\hat{\alpha}_{ijk}$, coeficiente empírico wavelet.

[2] $\hat{R}_i(h)$ función de autocovarianza.

[3] a_J y b_J son valores reales

[4] $\Psi_{ij}(\omega)$, transformada de Fourier.

la hipóteisis nula, H_0. El test de heterocedasticidad robusta $\hat{W}_1$ puede ser más fuerte que el test de heterocedasticidad consistente $\hat{W}_2$. La distribución límite de estos dos test no es cambiable.

4.2. Los residuales, datos reales.

Para obtener los datos reales; es decir, los *residuales*, se utilizó el modelo obtenido por el magister Francisco I. Zuluaga D. en su trabajo de maestría *"Econometría de Datos de Panel: revisión y una aplicación"* (ver [72]), cuya fuente de información utilizada fue la base de datos de la Superintendencia de Sociedades (Supersociedades) durante los años 1998 al 2003, con un total de 3029 empresas. El objetivo de éste, fue estimar la elasticidad de escala donde la variable dependiente es el dinero y la variable independiente es el ingreso operacional.

El modelo estático obtenido, fue

$$\lg(m_{it}) = \beta(t)\lg(y_{it}) + \gamma_t + \eta_i + \upsilon_{it} \tag{4.2.1}$$

Con *coeficiente de escala* $\beta(t) = \beta_1 + \beta_2 * t + \beta_3 * t^2$.

donde:
γ_t: Captura el efecto temporal generado por cambios en la tasa de interés agregada sobre la demanda de dinero para la firma i en el periodo t.
η_i: Efecto firma.
υ_{it}: Término de error.
$\lg(m_{it})$: Logaritmo de los saldos monetarios reales.
$\lg(y_{it})$: Logaritmo de los ingresos operacionales.

Bajo el modelo (4.2.1) y con ayuda del software Matlab, obtuvios los residuales[5]. Para el cálculo de éstos tomamos una muestra aleatoria de 570 datos.

[5]Al calular el test de Hausman, a partir de éstos, encontramos que el modelo a seguir era el de efectos fijos; dado que $p = 0{,}0075$. Con $\alpha = 0{,}05$.

4.3. Verificación de las hipótesis

En la tabla No.1 se muestran los valores calculados de los test $\hat{W}_1$ y $\hat{W}_2$, a partir de los residuales obtenidos en la sección 4.2. Éstos fueron calculados usando los wavelets, *Franklin o Spline de primer orden (S1)*

$$\hat{\psi}(z) = e^{iz/2}(2\pi)^{-1/2}\frac{sin^4(z/4)}{(z/4)^2}\Big[\frac{P_3(z/4+\pi/4)}{P_3(z/2)P_3(z/4)}\Big]^{1/2} \tag{4.3.1}$$

y *spline de segundo orden (S2)*

$$\hat{\psi}(z) = -ie^{i\omega/2}(2\pi)^{-1/2}\frac{sin^6(z/4)}{(z/4)^3}\Big[\frac{P_5(z/4+\pi/4)}{P_5(z/2)P_5(z/4)}\Big]^{1/2} \tag{4.3.2}$$

donde$P_3(z) \equiv \frac{2}{3} + \frac{1}{3}cos(2z)$, y $P_5(z) \equiv \frac{1}{30}cos^2(2z) + \frac{13}{30}cos(2z) + \frac{8}{15}$, con factores de escala $j = 4, \ldots, 10$.

- Con el método de efectos fijos, éstos caen en zona de rechazo generando la aceptación de la hipótesis alterna, H_a.

- Para efectos aleatorios, se acepta la hipótesis nula con el test $\hat{W}_1$, mientras que con el test $\hat{W}_2$, se rechaza.

- Con efectos totales, para $\hat{W}_1$ con $\alpha = 5\,\%$ se acepta hipótesis nula, mientras que con $\alpha = 10\,\%$ se rechaza para $j = 6$ con el wavelet Franklin y $j = 5$ con el wavelet Splin de segundo orden. Con $\hat{W}_2$ se rechaza H_0 para todo j y $\hat{W}_i$ con $i = 1, 2$.

Efectos Fijos				
	S1		S2	
Factor de escala, j	$\hat{W}_1$	$\hat{W}_2$	$\hat{W}_1$	$\hat{W}_2$
4	60587.0	9734.4	1371.4	239.9
5	7594.9	1017.2	429.3	71.3
6	2478.3	283.2	253.1	38.6
7	1367.0	133.2	237.8	31.1
8	1065.6	87.2	266.5	28.7
9	998.3	67.4	321.0	27.8
10	1039.4	56.9	400.3	27.5
Efectos Aleatorios				
	S1		S2	
Factor de escala, j	$\hat{W}_1$	$\hat{W}_2$	$\hat{W}_1$	$\hat{W}_2$
4	0.72	9565.4	1.01	225.47
5	1.17	973.4	1.39	65.16
6	1.36	263.3	1.17	34.68
7	1.22	120.96	1.08	27.62
8	1.14	77.8	1.04	25.25
9	1.10	59.4	1.03	24.27
10	1.07	49.6	1.02	23.80
Efectos Total				
	S1		S2	
Factor de escala, j	$\hat{W}_1$	$\hat{W}_2$	$\hat{W}_1$	$\hat{W}_2$
4	0.953	9675.4	1.325	233.5
5	1.539	996.7	1.839	68.0
6	1.785	272.3	1.549	36.2
7	1.613	125.9	1.428	28.9
8	1.507	81.4	1.380	26.5
9	1.446	62.3	1.357	25.5
10	1.409	52.2	1.345	25.0

Tabla No.1 Test calculados $\hat{W}_1$ y $\hat{W}_2$ con $S1$ y $S2$

En la tabla 2. se muestran los valores calculados con las wavelets *Daubechies 4, (db4)* y *Daubechies 6, (db6)*(ver [37])

$$\hat{h}(\omega) = \sqrt{2}\left(\frac{1+e^{-i\omega}}{2}\right)^p R(e^{-i\omega}) \tag{4.3.3}$$

para $db4$, con $p = 2$,

$$h_p[n] = [0{,}4829629131445341 \\ 0{,}8365163037378079 \\ 0{,}2241438680420134 \\ -0{,}1294095225512604]$$

y

para $db6$, con $p = 3$,

$$h_p[n] = [0{,}332670552950 \\ 0{,}806891509311 \\ 0{,}459877502118 \\ -0{,}135011020010 \\ -0{,}085441273882 \\ 0{,}035226291882]$$

Efectos Fijos				
	D4		D6	
Factor de escala, j	$\hat{W}_1$	$\hat{W}_2$	$\hat{W}_1$	$\hat{W}_2$
0	1.7960	3.2489	1.7960	3.2489
1	-32.3308	6.8634	129.2648	6.8634
2	-72.2461	8.7703	276.8546	8.7703
3	-142.0513	11.0283	534.2276	11.0283
4	-234.7354	12.9377	875.4749	12.9377
5	-352.9268	14.5224	1310.300	14.5224
6	-505.8021	15.8448	1872.400	15.8448
7	-705.7756	16.9563	2607.400	16.9563
8	-969.6288	17.8976	3577.000	17.8976
9	-1320.100	18.7013	4864.700	18.7013
10	-1787.900	19.3928	6583.600	19.3928

Efectos Aleatorios				
	D4		D6	
Factor de escala, j	$\hat{W}_1$	$\hat{W}_2$	$\hat{W}_1$	$\hat{W}_2$
0	0.4828	3.2489	0.4828	3.2489
1	0.6035	6.4931	0.6035	6.4931
2	0.6584	7.5752	0.6584	7.5752
3	0.7121	9.5628	0.7121	9.5628
4	0.7556	11.0758	0.7556	11.0758
5	0.7901	12.3285	0.7901	12.3285
6	0.8176	13.3717	0.8176	13.3717
7	0.8398	14.2467	0.8398	14.2467
8	0.8578	14.9864	0.8578	14.9864
9	0.8726	15.6169	0.8726	15.6169
10	0.8848	16.1580	0.8848	16.1586

Efectos Total				
	D4		D6	
Factor de escala, j	$\hat{W}_1$	$\hat{W}_2$	$\hat{W}_1$	$\hat{W}_2$
0	0.6358	3.2489	0.6358	3.2489
1	0.7947	4.4999	0.7947	4.4999
2	0.8670	5.5183	0.8670	5.5183
3	0.9376	7.8545	0.9376	7.8545
4	0.9950	9.6488	0.9950	9.6488
5	1.0404	11.1471	1.0404	11.1471
6	1.0766	12.4044	1.0766	12.4044
7	1.1058	13.4663	1.1058	13.4663
8	1.1295	14.3698	1.1295	14.3698
9	1.1490	15.1444	1.1490	15.1444
10	1.1651	15.8133	1.1651	15.8133

Tabla No.2 Test calculados $\hat{W}_1$ y $\hat{W}_2$ con $db4$ y $db6$

A partir de éstos, se encontró:

- Con el método de efectos fijos, cuando $\alpha = 5\,\%$ no se rechaza la hipótesis nula para $\hat{W}_1$ y $\hat{W}_2$ con $j = 0$, tanto en *db*4 como en *db*6, y para el resto de factores de escala se rechaza $\mathbb{H}_0$. Ahora, si $\alpha = 10\,\%$ se rechaza la hipótesis nula para $\hat{W}_1$ y $\hat{W}_2$ con $j = 0, \cdots, 10$.

- Para efectos aleatorios y efectos totales, se acepta la hipótesis nula con $\hat{W}_1$, tanto para *db*4 como para *db*6 para toda j; mientras que en $\hat{W}_2$ se rechaza.

4.4. Conclusiones

Existe correlación serial de forma no conocida entre ε_{it} y $\varepsilon_{it-|h|}$, para los test $\hat{W}_1$ y $\hat{W}_2$ a partir de los residuales obtenidos con el modelo de efectos fijos usando el wavelet Franklin y el wavelet Spline de segundo orden. Para los wavelet *db*4 y *db*6, no existe sólo cuando $\alpha = 5\,\%$ y $j = 0$. Con efectos aleatorios encontramos que no existe correlación serial para el test $\hat{W}_1$ bajo el wavelet Franklin y el wavelet Spline de segundo orden, mientras que para el test $\hat{W}_2$ con ambos wavelet, si existe correlación serial. Con el método de efectos fijos, cuando $\alpha = 5\,\%$ no se rechaza la hipótesis nula para $\hat{W}_1$ y $\hat{W}_2$ con $j = 0$, tanto en *db*4 como en *db*6, y para el resto de factores de escala se rechaza $\mathbb{H}_0$. Ahora, si $\alpha = 10\,\%$ se rechaza la hipótesis nula para $\hat{W}_1$ y $\hat{W}_2$ con $j = 0, \cdots, 10$. Para efectos aleatorios y efectos totales, se acepta la hipótesis nula con $\hat{W}_1$, tanto para *db*4 como para *db*6 para toda j; mientras que en $\hat{W}_2$ se rechaza.

Con efectos aleatorios encontramos que no existe correlación serial para el test $\hat{W}_1$ bajo el wavelet Franklin, el wavelet Spline de segundo orden y el wavelet *db*4 y *db*6, mientras que para el test $\hat{W}_2$ con los anteriores wavelet, si existe.

Ahora, con el modelo de efectos totales verificamos que no existe correlación serial para el test $\hat{W}_1$ bajo los wavelet Franklin y Spline de segundo orden con $\alpha = 5\,\%$, sin embargo cuando utilizamos $\alpha = 10\,\%$ encontramos que existe correlación serial cuando el factor de escala j es 6 y 5, respectivamente. El test $\hat{W}_2$ presenta correlación serial, independientemente de los valores j y α.

Como se mencionó en párrafos anteriores, se programó en el software Matlab la función *wavetest(resid,N,T,J,W)*, la cual es una contribución de nuestro trabajo al realizado por Hong y Kao. Dicha función tiene como argumentos los residuales (*resid*), el número de categorías o empresas (N), número de datos por categoría (T), factor de escala (j) y la función wavelet (W). Dicha función arroja [$w1$ $w2$], donde $w1$ y $w2$ representan los valores calculados de los test $\hat{W}_1$ y $\hat{W}_2$, respectivamente. La anterior función queda disponible para las personas interesadas[6].

Comparando nuestros resultados con los obtenidos en la tesis *Econometría de datos de panel: Revisión y una aplicación*, observamos grandes ventajas, primero (ver [72]) se utilizó un test para detectar correlación serial de orden uno el cual no es robusto ni consistente para heterocedasticidad mientras que lo test $\hat{W}_1$ y $\hat{W}_2$ son robustos respectivamente; además que son test para correlación serial de forma no conocida, en ([72]) el modelo es en esencia es un modelo de efectos fijos y se encontró que existe correlación serial de orden uno, mientras que los test $\hat{W}_1$ y $\hat{W}_2$ detectan correlación serial de orden n.

[6]Contactar a los e-mail: albervi32@latinmail.com y javiermartinezplazas@yahoo.com

Bibliografía

[1] Aldroubi, A. The wavelet transform: A surfing guide, pp 3-36 in *Wavelets in Medicine and Biology*, A. Audroubi, M. Unser (eds.), CRC Press, New York, 1996.

[2] Arellano, M. and Bover O., *La Econometría Datos de Panel*, Investigaciones Económicas(Segunda época). Vol XIV, no. 1, 1990, páginas 3-45.

[3] Arellano, M. *Panel Data Econometrics*. Oxford University Press, New York, 2003

[4] Asmar, Ch. Abraham. *Tópicos en teoría de matrices*, Primera edición, Medellín, 1995.

[5] Baltagi, B., *Econometrics Analsis of Panel Data*, 2^{nd} edition. Wiley, 2001.

[6] Billingsley, P. *Probability and Measure*, 3^{nd} edition. John Wiley & Sons, Inc. New York, 1995.

[7] Boggess, A., Narcowich, F. J.*A First Course in Wavelets with Fourier Analysis*. Prentice Hall, New Jersey, 2001.

[8] Bonani, A., Durand, S. and Weiss, G. *Wavelets obtained by continuos deformations of the Haar wavelet*, *Revista Mat. Iberoamericana*, Vol. 12, No 1, 1996.

[9] Burdisso, T., *Estimación de una función de costos para los bancos privados Argentinos utilzando datos de panel*. Banco Central de la Republica de Argentina, documento de trabajo N° 3, 1997.

[10] Burrus, C. S., Gopinath, R. A., Guo, H. *Introduction to Wavelets and Wavalets Transforms A Primer*. Prentice Hall, New Jersey, 1998.

[11] Cobacho, T. Ma. Belén and Bosch M. Mariano. *Contrastes de Hipótesis de Datos en Panel*, XIII Jornadas de ASEPUMA.

[12] Chui, C. K. *An Introduction to Wavelets*. Academic Press, Boston, 1992.

[13] Chui, C. K. *Wavelets: A Mathematical Tool for Signal Analysis*. SIAM Monographs on Mathematical Modeling and Computation, Philadelphia, 1997.

[14] Chung, K. L. *Elementary Probability Theory with Stochastic Processes*. Springer-Verlag. New York, 1979.

[15] Clarke, R. J. *Transform Coding of Images*. Academic Press, San Diego, CA, 1985.

[16] Clarke, R. J. *Digital Compression of Still Images and Video*. Academic Press, San Diego, CA, 1995.

[17] Cryer, J. D. *Time Series Analysis*, Duxbury Press. Boston, 1986.

[18] Daubechies, I. Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.*, 41 1988, 909-996.

[19] Daubechies, I. *Ten Lectures on Wavelets*, CBMS Series 61, SIAM, Philadelphia 1992.

[20] Daubechies, I. The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inform. Theory*, vol. 36 1990, 961-1005.

[21] Delyon, B. and Juditsky, A. Estimating wavelets coefficients, pp 151-168 in *Wavelets and Statistics*, Springer-Verlag, Tokyo, 1995.

[22] Donoho, D. *Nonlinear Wavelet Methods for Recovery of Signals, Densities and Spectra from Indirect and Noyse Data*. "Different Perspectives on Wavelets", Proceedings of Symposia in Pure Math., AMS., I Daubechies, Edt., 47, 173-205, 1993.

[23] Folland, G. B. *Real Analysis*, 2^{nd} edition. John Wiley & Sons, Inc. New York, 1999.

[24] Gabor, D. Theory of communications, *J. Inst. Elect. Eng. London.* 93(III) 1946, 429-457.

[25] Gersho, A. and Gray R. M. *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.

[26] Gonzalez, R. C. and Woods, R. E. *Digital Image Processing.* Addison-Wesley, New York, 1992.

[27] Greene, W.H. *Análisis Econométrico.* Quinta Edición, Prentice Hall, 1998.

[28] Hamilton, J. D. *Time Series Analysis.* Princeton University Press, Princeton, New Jersey, 1994.

[29] Hernández, E. Weiss, G. *A First Course on Wavelets.* CRC Press, Boca Raton, FL, 1996.

[30] Hong, Y. and Kao, Ch. *Wavelet-based testing for serial correlation of unknown form in panel models.* Econometrica, Vol. 72, N° 5 (Septiembre, 2004), Pág. 1519-1563.

[31] Hsiao, Ch. *Anilysis of Panel Data.* Econometric Society Monographs. Cambridge University Press, 1986.

[32] Kikut, O. *Uso de datos de tiempo y sección cruzada.* Monografía, Febrero, 1999.

[33] Irino, T., Kawahara, H. Signal reconstruction from modified auditory wavelet transform, *IEEE Trans. Signal Process.*, vol 41, 1993, 3549-3553.

[34] Lee, J. and Hong. *Testing for Serial Correlation of Unknown form using Wavelets Methods.* Econometrics Theory, 2001, Pág. 17, 386-423.

[35] Mallat, S. Multiresolution approximations and wavelet orthonormal bases for $L^2(\mathbb{R}^d)$, *Trans. of Amer. Math. Soc.* 315, 1989, 69-87.

[36] Mallat, S. A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, 1989, 674-693.

[37] Mallat, S., 'textita wavelet tour of signal processing, second edition, Academic Press, 1999.

[38] Mulligan, C.B. Scale Economies, the Value of Time,and the Demand for Money:Longitudinal Evidence from Firm. Journal of Political Economy,105,1061-1079,1997

[39] Mulligan, C.B. The Demand for Money by Firms:Some Aditional Empirical Results, Discussion Paper 125,Federal Reserve Bank of Minneapolis,1-52,1997.

[40] Martin, G. *Introducción a la econometría*. Prentice Hall, 1997.

[41] Meyer, Y. *Ondelettes et opérateurs, I: Ondelettes*. Herman, Paris, 1990.

[42] Mulligan, C.B. *Scale Economies, the Value of Time,and the Demand for Money:Longitudinal Evidence from Firms*, Journal of Political Economy,105,1061-1079,1997.

[43] Mulligan, C.B. *The Demand for Money by Firms: Some Aditional Empirical Results*, Discussion Paper 125,Federal Reserve Bank of Minneapolis,1-52,1997.

[44] Newey, W. y D.McFadden. Large Sample Estimation and Hypothesis Testing, en Engle y McFadden(eds) Handbook of Econometrics, IV, North-Holland, 1994.

[45] Oppenheim, A., Shafer, R. *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[46] Percival, D. B., Walde, A. T. *Wavelet Methods for Time Series Analysis*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2000.

[47] Pindyck, R. and Rubinfeld, D. *Econometric models and Economics forecasts*. McGraw-Hill International Edition, Fourth Edition,1993.

[48] Pinsky, M. A. *Introduction to Fourier Analysis and wavelet*, Brooks/Cole, NJ, 2001.

[49] Proakis, J. G., Manolakis, D. G. *Signal Processing: Principles, Algorithms and Applications*, Macmilan, New York, 1992.

[50] Quak, E. and Weyrich, N. Decomposition and reconstruction algorithms for spline wavelet on a bounded interval, *Appl. and Comp. Harmonic Anal.* (ACHA), 1 1994, 217-231.

[51] Rioul, O., Duhamel, P. Fast algorithms for discrete and continuous wavelet transform, *IEEE Trans. Inform. Theory*, vol. IT-38 1992, 569-586.

[52] Ross, S. M. *Introduction to Probability Models*, 8^{nd} edition. Academic Press. New York, 2003.

[53] Shapiro, J. M. Embedded image coding using zerotrees of wavelet coefficients, *IEEE Trans. Signal Proc.* 41 1993, 3445-3462.

[54] Shen, X. A Galerkin-wavelet method for a singular convolution equation on the real line, *J. Int. Equa. Appl.* 12 2000, 157-176.

[55] Strang, G. and Nguyen, T. *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Cambridge, MA, 1996.

[56] Stremler, F. G. *Introduction to Comunication Systems*, 2^{nd} edition. Addison-Wesley Publishing Company, Inc. Massachusetts, 1982.

[57] Sun, W. and Zhou, X. Sampling theorem for wavelet subspaces: error estimate and irregular sampling, *IEEE Trans. Signal Process.*, vol. 48 2000, 223-226.

[58] Unser, M. and Aldroubi, A. B-spline processing I: Theory and II: Efficient design and application, *IEEE Trans. Signal Process.*, vol 41 1993, 821-848.

[59] Unser, M. A practical guide to implementation of the wavelet transform, pp 37-76 in *Wavelets in Medicine and Biology*, A. Audroubi, M. Unser (eds.), CRC Press, New York 1996.

[60] Walnut, D. *An Introduction to Wavelets Analysis.* Birkhäuser, Boston, 2002.

[61] Walter, G. G. Approximation with impulse trains, *Results Math.* 34 1998, 185-196.

[62] Walter, G. G. and Shen, X. A substitute for summability in wavelet expansions, *Appl.Numer. Harmon. Anal.*, Birkhäuser Boston, Boston, MA 1999, 51-63.

[63] Walter, G. G. and Shen, X. Deconvolution using Meyer wavelets, *J. Integral Equations Appl.* 11 1999, 515-534.

[64] Walter, G. G., Shen, X. *Wavelets and Other Orthogonal Systems*, 2^{nd} edition. Chapman & Hall/CRC, Boca Raton, FL, 2001.

[65] Wayne, A. F. *Introduction to Statistical Time Series*, Second Edition, A Wiley-Interscience Publication, John Wiley and Sons, INC. 1996.

[66] Wei, W. W. S. *Time Series Analysis*, 2^{nd} edition. Addison-Wesley Publishing Company, Inc. Boston, 2006.

[67] Yamada, K., Ohkitani, K. Orthonormal wavelet expansion and its application to turbulence, *Progr. Theoret. Phys.*, vol. 83-5, 1990, 819-823.

[68] Yamada, K., Ohkitani, K. An identification of energy cascade in turbulence by orthonormal wavelet analysis, *Progr. Theoret. Phys.*, vol. 84-4, 1991, 799-815.

[69] Yamada, K., Ohkitani, K. Orthonormal wavelet analysis of turbulence, *Fluid Dynam. Res.*, vol. 8, 1991, 101-115.

[70] Rosales, R. y Bonilla, J. A. *Introducción a la econometría.* Series apuntes de clase Cede, octubre de 2006.

[71] Wooldridge, J. Econometric Analysis of Cross Section and Panel Data, The MIT Press, 2001.

[72] Zuluaga, F. I. *Econometría de Datos de Panel: Revisión y una Aplicación.* Universidad Eafit, Medellín, julio de 2005.

Índice alfabético