Documentos Técnicos

Curso de Contabilidad Regional Trimestral

Autores: Juan Manuel Rodríguez Poo
Francisco J. Parra Rodríguez
Raquel Bedia Expósito
M. Paz Moral Zuazo

DOC. Nº 2/2010
ISSN 2444 - 1627
Santander,
Cantabria

ICANE
Instituto Cántabro de ESTADÍSTICA

Í N D I C E

I. CUENTAS REGIONALES TRIMESTRALES. INTRODUCCIÓN

I.1.- Las cuentas regionales trimestrales: definición e interés

Para lograr los objetivos de la Unión Monetaria Europea, necesitamos instrumentos estadísticos de corto plazo de alta calidad que suministren a las instituciones de la Comunidad, los gobiernos, el Banco Central Europeo, los bancos centrales nacionales y los operadores económicos y sociales un conjunto de estadísticas de corto plazo comparables y confiables sobre las cuales basar sus decisiones.

Para ello Eurostat en el ámbito de la UE publica un manual de cuentas nacionales trimestrales, solicitado en forma explícita en el Sistema Europeo de Cuentas (SEC 1995) que representa el primer manual de armonización de las cuentas nacionales trimestrales como parte integrante del sistema de cuentas nacionales. Por otro lado, hay que señalar que las cuentas trimestrales son la base analítica para el pronóstico del ciclo económico y en particular para los sistemas de indicadores anticipados.

Las cuentas trimestrales permiten que los economistas estudien los ciclos económicos, midan desde el punto de vista estadístico los efectos de los shocks económicos y efectúen análisis dinámicos con la ayuda de instrumentos estadísticos y econométricos. Permiten la verificación estadística de las hipótesis teóricas y estimaciones econométricas de las ecuaciones económicamente significativas y suministran un insumo para los ejercicios de pronóstico.

Respecto al año corriente, los compiladores de cuentas nacionales trimestrales se ocupan de entregar la estimación más robusta de las variaciones de corto plazo de las principales variables económicas según la información disponible. Estos datos permiten que los analistas del corto plazo detecten los puntos de inflexión y ayuden a los agentes económicos a reconsiderar sus estrategias según sus preferencias y expectativas.

Las cifras trimestrales son importantes para los analistas estadísticos y económicos porque las series trimestrales estimadas contienen la historia íntegra del agregado que interesa. En los institutos de estadística encargados de la compilación de cuentas nacionales trimestrales, el uso de modelos matemáticos y estadísticos asume una importancia que varía según la filosofía de compilación y la información disponible. Así, en distintos países la estimación de los agregados de cuentas trimestrales se realiza según métodos diferentes en que el uso de modelos estadísticos es diferente.

I.2.- Etapas de elaboración de la Contabilidad Trimestral

El proceso de elaboración de la Contabilidad Regional Trimestral (CRT) se resume en los siguientes pasos:

1. Selección de variables o indicadores simples que pueden formar parte del indicador sintético, según sus propiedades cíclicas.

2. Construcción de indicadores sintéticos trimestrales para cada rama de actividad o sector institucional.

3. Uso de los indicadores sintéticos para:

   - predecir las operaciones para cada rama ó sector en el año corriente.
   - desagregar trimestralmente las series de la Contabilidad Anual de referencia.

4. Construcción de los índices volumen trimestrales.

Como los indicadores han de cumplir una serie de requisitos para que sean seleccionados para formar parte del indicador sintético. Esto hace necesario disponer de un número suficiente para decidir entre índices alternativos.

I.3.- Indicadores económicos y Contabilidad Regional Trimestral

En la literatura sobre indicadores se considera como indicador cíclico a aquellos datos o series de datos, que midiendo aspectos significativos de la actividad económica, responden a cambios en el clima económico. La primera lista de indicadores del ciclo económico la realizó en 1938 el "National Bureau of Economic Research" (NBER) de Estados Unidos. Dicha lista se elaboró seleccionando de un gran número de datos trimestrales y mensuales sobre precios, empleo, producción y otros hechos relativos a la economía americana, aquellos más representativos en base a su comportamiento cíclico y relevancia económica.

Dicha selección permitió establecer las características básicas que habrían de cumplir los indicadores del cíclico económico. Estas características que se deben a Mitchell y Burns (1938) son las siguientes:

- Longitud suficiente, es decir, la serie debe ser lo bastante larga para permitir observar varios ciclos.
- La significación económica en su comportamiento respecto al ciclo, que no ha de variar en el futuro.
- Calidad Estadística de la serie, en el sentido de que medirá el proceso económico que representa de una manera similar tanto en el presente como en ejercicios futuros.
- Correspondencia histórica con las fluctuaciones cíclicas observadas en el pasado.
- Consistencia cronológica, esto es, sus adelantos o retrasos con respecto a recuperaciones o caídas de actividad han de ser constantes.
- Perfil suave, debido a un componente irregular de escasa relevancia. Esta propiedad implica que las series candidatas a ser indicadores deben ser previamente filtradas con objeto de eliminar movimientos erráticos (Espasa, 1991 y 1993; Fernández Macho 1991ª y 1991ᵇ; Melis, 1991). En la práctica, se suelen utilizar o bien series corregidas de estacionalidad y efectos calendario o bien el componente de tendencia. (Extracción de señales)
- Prontitud en la disposición de datos

En consecuencia, para la selección de los indicadores coyunturales se han tenido en cuenta las siguientes características:

- longitud
- significación económica
- calidad estadística
- disponibilidad
- inmediatez
- alta frecuencia
- fiabilidad
- eficiencia

- rigor y completitud
- correlación con los agregados objeto de estudio.

El método empleado para elaborar un indicador de cada sector consta de dos etapas:

a) En primer lugar se especifican y estiman modelos ARIMA con Análisis de Intervención para las variables seleccionadas como indicadores; estos modelos se utilizan para prolongar la serie de observaciones del respectivo indicador con predicciones y para corregir esta serie prolongada de anomalías que puedan afectar a la estimación de la tendencia;

b) Se procede a estimar el indicador sintético según el método elegido.

2. NÚMEROS ÍNDICES

2.1.- Introducción

El número índice es un valor expresado como porcentaje de una cifra que se toma como unidad base. Por ejemplo, cuando decimos que el índice de precios de consumo (base media de 1992=100) correspondiente al mes de diciembre de 1997 es 122,9, estamos señalando que los precios en diciembre de 1997 eran un 22,9 más elevados que los que estaban en vigor a lo largo de 1992.

Los números índices no tienen unidades y pueden referirse tanto a precios (índice de precios de consumo, índice de precios percibidos por los agricultores, índice de precios industriales) como a cantidades (índice de producción industrial).

El número índice es un recurso estadístico para medir diferencias entre grupos de datos. Un número índice se puede construir de muchas formas distintas. La forma de cada índice en particular dependerá del uso que se le quiera dar. Los números índices se elaboran tanto con precios (p) como con cantidades (q). El año en que se inicia el cálculo de un número índice se denomina año base y se nombran por p0 o q0 según tratemos de precios o de cantidades, a los precios o las cantidades de los años sucesivos los indicamos por pt o qt.

Las comparaciones pueden ser de una única magnitud, en este caso hablaremos de índices simples, o de varias magnitudes índices complejos o sintéticos. Si trabajamos con diferentes magnitudes o tipos de mercancías utilizamos los subíndices (i) para referirnos a un tipo de mercancía, de modo que utilizamos los símbolos pit o qit para señalar el precio o la cantidad de la mercancía i en el período t.

Dentro de los índices complejos o sintéticos puede que todas las mercancías tengan la misma importancia, índices no ponderados y en caso contrario índices ponderados. Los números índices no ponderados son los más sencillos de calcular, pero deben de utilizarse con especial cuidado. Los números índices ponderados requieren que definamos previamente a su construcción los criterios de ponderación o de peso. Una vez definida una ponderación debe de respetarse en los sucesivos períodos.

Las ventajas de los números índices son:

- Naturaleza adimensional, no tienen unidades y esto nos permite hacer comparaciones.
- Sirven para simplificar la complejidad de ciertos conceptos o fenómenos económicos.

A la hora de elaborar un número índice hay que tener presente una serie de propiedades que el índice debe de cumplir. Dichas propiedades son:

a) Existencia: Todo número índice ha de tener un valor finito distinto de cero.

b) Identidad: Si se hacen coincidir el período base y el período actual el valor del índice tiene que ser igual a la unidad (o 100 si se elabora en porcentajes).

c) Inversión: El valor del índice ha de ser invertible al intercambiar los períodos entre sí. Es decir: $I_t^o = \frac{1}{I_o^t}$ el índice del año o calculado con la base del año *t*, ha de ser igual al inverso del índice del año *t* calculado en base del año *o*.

d) Proporcionalidad: Si en el período actual todas las magnitudes experimentan una variación proporcional, el número índice tiene que experimentar también dicha variación.

e) Homogeneidad: Un número índice no puede estar afectado por los cambios que se realicen en las unidades de medida.

2.2.- Números Índices Simples

Sirven para estudiar la evolución de una sola magnitud en relación a un periodo base y pueden ser:

a) Fijos: el año base es siempre el mismo.

Si $\mathbf{X_0}$ y $\mathbf{X_t}$ representan los valores de la magnitud en los periodos base y actual, respectivamente, el número índice simple se denota por $\mathbf{I}_{t/0}$, y viene dado por:

$$I_0^t = \frac{x_{it}}{x_{i0}} \times 100$$

Que como se indica suele expresarse en porcentajes, aunque también podría expresarse en tanto por uno y nos mide la variación que ha sufrido la magnitud entre los dos periodos considerados.

b) En cadena: cuando el año base varía, es decir cuando el año base es el inmediatamente anterior.

$$I_{t-1}^t = \frac{x_{it}}{x_{it-1}} \times 100$$

Para obtener un índice fijo a partir de un índice en cadena se utiliza la siguiente formula:

$$I_{t-1}^t = \frac{I_0^t}{I_0^{t-1}}$$

Para el caso contrario se utiliza esta fórmula:

$$I_0^t = \prod_{i=1}^{t} I_{i-1}^i$$

Los números índices más utilizados son los siguientes:

- **Precio relativo:** es el cociente entre el precio de un bien en el periodo actual (p_{it}) y el precio del mismo en el periodo base (P_{i0})

$$p_0^t = \frac{p_{it}}{p_{I0}} \times 100$$

- **Cantidad relativa:** es el cociente entre la cantidad de un bien en el periodo actual (q_{it}) y la cantidad del mismo en el periodo base (q_{i0})

$$q_0^t = \frac{q_{it}}{q_{I0}} \times 100$$

- **Valor relativo:** es el cociente entre el valor de un bien de un bien en el periodo actual
($P_{it} * q_{it}$) y la cantidad del mismo en el periodo base ($P_{i0} * q_{i0}$)

$$v_0^t = \frac{p_{it} \cdot q_{it}}{p_{i0} \cdot q_{it}} \times 100 = p_0^t \cdot q_0^t$$

2.3.- Números Índices Complejos o Sintéticos

Son indicadores sintéticos que se elaboran a partir de dos o más series de datos con el objeto de estudiar su evolución conjunta y realizar comparaciones con otras series. Los números índices compuestos se clasifican en:

a) No ponderados: cuando todas las variables tienen asignada la misma importancia.

b) Ponderados: Cuando a cada variable se le asigna un peso o ponderación.

Partimos de una serie de magnitudes simples x1, x2…………..xN, para las que conocemos su valor en el periodo base o de referencia, al que denotaremos por 0, y en el periodo actual t.

A continuación calculamos los índices simples para cada magnitud, de modo que disponemos de la siguiente tabla:

Magnitudes	Valor periodo base	Valor periodo actual	Índices simples
Magnitud 1	X_{10}	X_{1t}	$I_1 = X_{1t}/X_{10}$
Magnitud 2	X_{20}	X_{2t}	$I_2 = X_{2t}/X_{20}$
.	.	.	.
Magnitud N	X_{N0}	X_{Nt}	$I_n = X_{Nt}/X_{N0}$

Con la serie de los N índices simples podemos obtener los siguientes índices compuestos:

2.3.1.- Índices Compuestos Sin Ponderar

2.3.1.1.- Índice media ARITMÉTICA de los índices simples

$$I_0^t = \frac{I_1 + I_2 + ... + I_N}{N} = \frac{\sum_{i=1}^{N} I_i}{N} = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{x_{it}}{x_{i0}}$$

2.3.1.2.- Índice media GEOMÉTRICA de los índices simples

$$I_0^t = \sqrt[N]{I_1 . I_2 .... I_N} = \sqrt[N]{\prod_{i=1}^{N} I_i} = \sqrt[N]{\prod_{i=1}^{N} \frac{x_{it}}{x_{i0}}}$$

2.3.1.3.- Índice media ARMÓNICA de los índices simples

$$I_0^t = \frac{N}{\frac{1}{I_1} + \frac{1}{I_2} + ... + \frac{1}{I_N}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{I_i}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{\frac{x_{it}}{x_{i0}}}}$$

2.3.1.4.- Índice media AGREGATIVA

$$I_0^t = \frac{x_{it} + x_{2t} + ... + x_{Nt}}{x_{1o} + x_{2o} + ... + x_{No}} = \frac{\sum_{i=1}^{N} x_{it}}{\sum_{i=1}^{N} x_{io}}$$

2.3.2.- Índices Compuestos Ponderados

Todos estos mismos índices se pueden calcular con ponderaciones. Una ponderación w_i es un valor de referencia para cada producto que determina su importancia relativa en el índice total. Al ser el ponderador un valor relativo lo normal es que se presente calculado en tanto por uno, por ciento ó por mil, expresando así el porcentaje que representa dicho producto en la cesta de productos que cubre el índice. Por ejemplo,

$$w_i = \frac{p_{i0} \cdot q_{io}}{\sum_{i=1}^{N} p_{io} \cdot q_{io}}$$

Una vez obtenidos los ponderadores (w_i) se calculan el índice media aritmética ponderada de índices simples cuando operamos del siguiente modo:

$$I = \frac{I_1 w_1 + I_2 w_2 + ... + I_N w_N}{w_1 + w_2 + ... + w_N} = \frac{\sum_{i=1}^{N} I_i . w_i}{\sum_{i=1}^{N} w_i}$$

2.3.2.1.- Índices de precios

Los índices de precios se elaboran usualmente utilizando índices complejos ponderados siendo los más utilizados los denominados índices de Laspeyres, Paasche y Fisher.

2.3.2.1.1.- Índice de Laspeyres

El índice de Laspeyres (tanto de precios como cuántico) es el más utilizado en los indicadores generales de precios y producción. Su diseño y posterior cálculo requiere una rigurosa selección de sus componentes y ponderaciones. Ahora bien, a medida que nos alejamos del periodo base, la estructura de coeficientes de ponderación de este índice (y de los demás) es cada vez menos representativa con lo que es necesario fijar un nuevo periodo base y establecer una nueva estructura de ponderaciones.

El índice de Laspeyres es una media aritmética ponderada de índices simples, cuyo criterio de ponderación es $w_i = p_{io}.q_{io}$. La fórmula que define el índice de precios de Laspeyres es la siguiente:

$$Lp = \frac{\sum_{i=1}^{N} \frac{p_{it}}{p_{i0}} \cdot p_{i0} \cdot q_{i0}}{\sum_{i=1}^{N} p_{i0} \cdot q_{i0}} = \frac{\sum_{i=1}^{N} p_{it} \cdot q_{io}}{\sum_{i=1}^{N} p_{io} \cdot q_{io}}$$

Se suele utilizar este índice a la hora de elaborar los índices de precios por cuestiones prácticas ya que únicamente requiere investigar en el año base el valor de los ponderadores, que es la parte mas costosa de la elaboración del índice, (téngase en cuenta que en el IPC se realiza una encuesta de presupuestos familiares en los años base que requiere una muestra de 20.000 hogares). Una vez determinados los ponderadores el índice de Laspeyres únicamente requiere que se investigue en los sucesivos períodos la evolución de los precios.

2.3.2.1.2.- Índice de Paasche

También es una media aritmética ponderada de los índices simples, pero utilizando como coeficiente ponderador $w_i = p_{io}.q_{it}$; por tanto su definición queda como:

$$Pp = \frac{\sum_{i=1}^{N} \frac{p_{it}}{p_{i0}} \cdot p_{i0} \cdot q_{it}}{\sum_{i=1}^{N} p_{i0} \cdot q_{it}} = \frac{\sum_{i=1}^{N} p_{it} \cdot q_{it}}{\sum_{i=1}^{N} p_{io} \cdot q_{it}}$$

La diferencia entre el índice Paasche y el índice Laspeyres es que exige calcular las ponderaciones para cada periodo corriente “t”, haciendo su cálculo estadístico más laborioso, y presentando el inconveniente de que sólo permite comparar la evolución del precio de cada año con el año base, dado que las ponderaciones varían de período en período. Ambas razones han determinado que este índice sea más inusual que el anterior.

2.3.2.1.3.- Índice de Fisher.

El índice de Fisher es la media geométrica de los índices de Laspeyres y Paasche, es decir:

$$Fp = \sqrt{Lp.Pp}$$

Como los índices de precios de consideran un año determinado para calcular el ponderador bien sea a partir de $q_0 .p_0$, o de $q_t .p_0$, utilizan la denominación de año base para referirse al año “0” a partir del que se calcula el ponderador w_i.

2.3.2.2.- Índices cuánticos o de producción

Los índices cuánticos se elaboran de forma similar a los índices de precios. Citaremos los más usuales.

2.3.2.2.1.- *Índice de Laspeyres*

El criterio de ponderación es $w_i=q_{io}.p_{io}$. La fórmula que define el índice cuántico de Laspeyres es la siguiente:

$$Lq = \frac{\sum_{i=1}^{N} \frac{q_{it}}{q_{i0}} \cdot q_{i0} \cdot p_{i0}}{\sum_{i=1}^{N} q_{i0} \cdot p_{i0}} = \frac{\sum_{i=1}^{N} q_{it} \cdot p_{i0}}{\sum_{i=1}^{N} q_{i0} \cdot p_{i0}}$$

2.3.2.2.2.- *Índice de Paasche*

El criterio de ponderación es $w_i= q_{io}.p_{it}$; por tanto su definición queda como:

$$Pq == \frac{\sum_{i=1}^{N} \frac{q_{it}}{q_{i0}} \cdot q_{i0} \cdot p_{it}}{\sum_{i=1}^{N} q_{i0} \cdot p_{it}} = \frac{\sum_{i=1}^{N} q_{it} \cdot p_{it}}{\sum_{i=1}^{N} q_{i0} \cdot p_{it}}$$

2.3.2.2.3.- *Índice de Fisher.*

El índice de Fisher es la media geométrica de los índices de Laspeyres y Paasche, es decir:

$$Fq = \sqrt{Lq.Pq}$$

2.4.- Índices de Volumen encadenados

Tradicionalmente, en los índices compuestos se comparan directamente dos puntos en el tiempo, el periodo actual (t) y el periodo base (0). Las diferencias entre los distintos índices surgen a la hora de agregar los índices simples o elementales. En los índices de tipo Laspeyres se considera la utilización de ponderaciones del periodo base, mientras que los índices de tipo Paasche utilizan las ponderaciones del periodo actual. En ambos casos, si se produce un cambio importante en la composición de las unidades elementales entre los periodos base y actual, la relevancia de ambos índices se ve reducida.

De hecho, Capitulo 10 del SEC trata de la MEDICIÓN DE LAS VARIACIONES DE PRECIO Y VOLUMEN, señalando:

10.61. La elaboración de un sistema integrado de índices de precio y de volumen supone una elección deliberada de los tipos de índices que se deben utilizar.

10.62. La forma más adecuada de medir las variaciones interanuales de volumen es mediante un índice de volumen de Fisher, que se define como la media geométrica de los índices de Laspeyres y de Paasche. Las variaciones de volumen para períodos más largos se obtendrán encadenando, es decir acumulando, los movimientos interanuales de volumen.

10.63. La forma más adecuada de medir las variaciones interanuales de precio es mediante un índice de precios de Fisher. Las variaciones de precio para períodos más largos se obtendrán encadenando los movimientos interanuales de precios.

10.64. Los índices encadenados que utilizan los índices de volumen de Laspeyres para medir variaciones de volumen y los índices de precio de Paasche para medir variaciones interanuales de precios son una alternativa válida a los índices de Fisher.

EUROSTAT ha elaborado un manual sobre esta metodología que se puede consultar en el siguiente enlace:

http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-41-01-543-__-N-EN.pdf

Por su parte, el INE ha publicado un documento sobre las mediciones de volumen mediante índices encadenados accesible en:

http://www.ine.es/daco/daco42/cne00/medic_vol_encad_b2000.pdf

Los índices encadenados consideran que el paso del período 0 al t puede fragmentarse considerando los incrementos parciales, esto es, que el encadenamiento de los índices (i.e. de las variaciones) evaluados con la frecuencia de muestreo máxima posible constituye una valoración más apropiada del cambio realizado desde 0 hasta t. Intuitivamente, se intenta reducir el envejecimiento de la base.
La forma de resolver este problema consiste en efectuar las comparaciones entre períodos que disten lo menos posible (por ejemplo, un período) mediante "eslabones":

$$I^{A}_{s/s-1} = \sum_{j} w_j i_{s/s-1}$$

A partir de los eslabones, la variación entre los periodos 0 y t se encadena:

$$CI^{A}_{t/0} = \prod_{s=1}^{t} I^{A}_{s/s-1}$$

Un índice así construido carece de periodo base o de ponderaciones, ya que van cambiando a lo largo de los distintos periodos. No obstante, se designa un periodo llamado de referencia, al que arbitrariamente se le asigna el valor 100.

En la siguiente tabla se ofrece un ejemplo con datos hipotéticos de dos productos (A y B) y tres años (0, 1 y 2):

PRODUCTO	2000			2001			2002		
	P0*Q0	PRECIO	CANTIDAD	P1*Q1	PRECIO	CANTIDAD	P2*Q2	CANTIDAD	P0*Q0
A	3	5	15	2	9	18	1	9	9
B	4	7	28	5	7	35	6	11	66
TOTAL			43			53			75

Primero, se calculan los eslabones:

PRODUCTO	2000			2001			2002		
	PRECIO	CANTIDAD	P0*Q0	PRECIO	CANTIDAD	P0*Q1	PRECIO	CANTIDAD	P1*Q2
A	3	5	15	2	9	27	1	9	18
B	4	7	28	5	7	28	6	11	55
TOTAL			43			55			73
Eslabón			100			127,9			137,7

$$127{,}9 = \frac{p_0 q_1}{p_0 q_0} = \frac{55}{43} \times 100$$

$$137{,}7 = \frac{p_1 q_2}{p_1 q_1} = \frac{73}{53} \times 100$$

El índice encadenado se obtiene multiplicando cada eslabón anual en forma de índice por la cadena acumulada hasta el año precedente. La cadena así obtenida es un número índice por lo que su conversión en términos monetarios se realiza multiplicándola por el valor a precios corrientes observado en un año particular, llamado "de referencia". En la siguiente tabla se considera el año 0 como periodo de referencia:

PRODUCTO	2000			2001			2002		
	PRECIO	CANTIDAD	P_0*Q_0	PRECIO	CANTIDAD	P_0*Q_1	PRECIO	CANTIDAD	P_1*Q_2
A	3	5	15	2	9	27	1	9	18
B	4	7	28	5	7	28	6	11	55
TOTAL			43			55			73
Eslabón			100			127,9			137,7
Índice encadenado			100			127,9			176,2
Valoración monetaria			43			55			76

$$127{,}9 = 127{,}9 \times 100$$
$$176{,}2 = 137{,}7 \times 127{,}9$$

Debe señalarse que, a diferencia de lo que ocurría con la valoración a precios constantes en la que el año de referencia y base coinciden, en el sistema de valoración a precios del año anterior no son equivalentes. Así, el año de referencia es el que define la escala del índice encadenado (haciéndolo 100), mientras que la base temporal es móvil, existiendo tantas bases como pares de años consecutivos por lo que, en conjunto, la valoración encadenada carece de base fija (base móvil).

La aplicación de esta metodología genera una pérdida de aditividad en las medidas encadenadas de volumen (excepto en los datos correspondientes a los años de referencia y al inmediatamente posterior). La pérdida de aditividad significa, por ejemplo, que la suma de los componentes del Producto Interior Bruto (PIB)no coincide con éste (excepto en los datos correspondientes a los años de referencia y al inmediatamente posterior). De forma general, una variable valorada mediante medidas encadenadas de volumen no coincide con la suma de sus elementos constituyentes igualmente evaluados a través de medidas encadenadas de volumen. La pérdida de aditividad es una consecuencia directa de las propiedades matemáticas del sistema de valoración, por lo que las discrepancias no reflejan deterioro alguno de calidad en el proceso de medida.

2.5.- Elaboración de Índices Compuestos

La fórmula básica para la construcción de los indicadores compuestos IC a partir de N indicadores simples es la siguiente:

$$IC = \sum_{i=1}^{N} w_i \, I(x_i)$$

Donde:

w_i es el ponderador del indicador simple i,
I es el método de normalización,
x_i es el indicador simple.

Cuando se elabora un indicador compuesto es necesario que las series individuales presenten la misma amplitud cíclica relativa, pues de lo contrario, las series con mayor amplitud cíclica dominarían el comportamiento del indicador compuesto, impidiendo así que se revele la información contenida en otras series de menor amplitud. Para lograrlo, se normalizan las series componentes restándoles la media y dividiéndolas por el promedio de las desviaciones de la media en valor absoluto, conforme la siguiente fórmula:

$$I_t = \frac{x_t - \bar{x}}{\frac{1}{n}\sum \left| x_t - \bar{x} \right|}$$

Siendo n el número de observaciones de x. Otros métodos de normalización serían:

— z-score:

$$I_t = \frac{x_t - \bar{x}}{\sigma_x}$$

— Min-max:

$$I_t = \frac{x_t - \min(x_t)}{\max(x_t) - \min(x_t)}$$

Cuando se trabaja con balances de respuestas (encuestas de opiniones empresariales), es conveniente utilizar índices de difusión:

$$ID_t = \frac{x_t + 100}{2}$$

Donde ID_t es el índice de difusión y x_t es el balance de respuestas correspondiente.
La diferencia entre un balance de respuestas y un índice de difusión es que el primero está centrado en cero, con un valor máximo de 100 y un mínimo de -100, mientras que el segundo está centrado en 50, con un valor máximo de 100 y un valor mínimo de cero. El uso de índices de difusión resulta más cómodo que el uso de balances, ya que en tal transformación, las series sólo toman valores positivos, lo que facilita el uso de logaritmos y descomposiciones multiplicativas de las series temporales.

Cuando el índice de difusión es mayor que 50, significa que los entrevistados están optimistas respecto a la evolución de la variable objetivo. Si es menor que 50, los entrevistados se encuentran pesimistas.

La ponderación se puede obtener de datos base de la Contabilidad Nacional Anual, por ejemplo, si se quiere construir un indicador de producción industrial, se puede agregar a partir de los índices subsectoriales y el VAB o empleo de cada subsector.

A continuación se exponen dos metodologías estadísticas de obtención de ponderadores: el método de Granger y Newbold (1986) y los componentes principales.

2.5.1.- El método de Granger y Newbold

Para la construcción del indicador sintético se estima la siguiente ecuación, utilizando la serie anual de la macromagnitud de referencia y el conjunto de variables seleccionadas anualizadas:

$$Y_T = \alpha_1\left(a_1 + b_1 X_T^1\right) + \alpha_2\left(a_2 + b_2 X_T^2\right) + \ldots + \alpha_k\left(a_k + b_k X_T^k\right) + u_T = Z_T + u_T$$

Siendo:

$Y_T \rightarrow$ Valor de la variable a trimestralizar en el año T.

$X_T^j \rightarrow$ Valor del indicador aproximativo, en el año T, proyectado hasta el último trimestre del año actual a través de modelos ARIMA, siendo k el número de indicadores aproximativos utilizados.

$a_j \rightarrow$ Término independiente de la regresión entre Y y X_T^j.

$b_j \rightarrow$ Pendiente de la regresión entre Y y X_T^j.

$\alpha_j \rightarrow$ Peso asignado a la estimación a través de la variable j

$Z_T = \alpha_1\left(a_1 + b_1 X_T^1\right) + \alpha_2\left(a_2 + b_2 X_T^2\right) + \ldots + \alpha_k\left(a_k + b_k X_t^k\right) \rightarrow$ Indicador sintético.

$u_T \rightarrow$ Error del modelo en el año T.

El peso de cada variable en el indicador sintético se establece de forma inversamente proporcional al error típico de su regresión con Y, σ_j, tal que:

$$\alpha_j = \frac{\sigma_j^{-1}}{\sum_{h=1}^{k} \sigma_h^{-1}}$$

Una vez obtenido el indicador Z, se obtiene la serie estimada del valor de la variable en el trimestre *t*:

$$y_t = \alpha_1\left(a_1 + b_1 x_t^1\right) + \alpha_2\left(a_2 + b_2 x_t^2\right) + \ldots + \alpha_k\left(a_k + b_k x_t^k\right)$$

2.5.2.- Estimación del modelo con Componentes Principales

La metodología de componentes principales se realiza en dos fases. En primer lugar se realiza una estimación de los componentes principales de los indicadores estratégicos relacionados con la variable Y, y en segundo lugar se realiza una regresión entre Y y el valor anualizado de los factores resultantes de la fase anterior.

Así pues, siendo $X^1, X^2, \ldots, X^k$, los distintos indicadores que hemos seleccionado como variables relacionadas con Y, este método va a extraer las diferentes funciones lineales que existen entre ellas:

$$\begin{aligned} Z^1 &= a_{11}X^1 + a_{12}X^2 + \ldots . a_{1k}X^k \\ Z^2 &= a_{21}X^1 + a_{22}X^2 + \ldots . a_{2nk}X^k \\ &\ldots . \\ Z^k &= a_{k1}X^1 + a_{k2}X^2 + \ldots . a_{kk}X^k \end{aligned}$$

Este método extrae las funciones lineales seleccionando las así de tal modo que las varianzas de las Z^S sean maximizadas. De este modo, los componentes extraídos son las combinaciones lineales de los indicadores que tienen mayor varianza, siendo Z^1 el componente con mayor varianza explicada, seguido del Z^2 que contiene la segunda mayor varianza explicada pero sin estar correlacionado con Z^1 y así sucesivamente, de modo que la suma de la varianza de todos los componentes explique el total de las variaciones de las X^j y, a su vez, estén incorreladas entre ellas.

Uno de los problemas de esta metodología radica en la determinación del número de componentes principales $m<k$ que deben ser tomados en cuenta para la fase número dos. La práctica más extendida es que sólo serán tomados aquellos componentes cuyos autovalores (raíces características) superen la unidad.

En la segunda fase del modelo de componentes principales, se expresa la relación entre la variable Y y los componentes principales (CP) extraídos del conjunto de indicadores originales.

$$Y_T = \alpha + \beta_1 Z_T^1 + \ldots + \beta_m Z_T^m + u_T$$

Obteniéndose la estimación trimestral de Y a partir de:

$$y_t = \alpha + \beta_1 z_t^1 + \ldots + \beta_m z_t^m + u_t$$

3. INDICADORES DISPONIBLES PARA LA CRT. PROBLEMÁTICA

3.1.- Indicadores de Empleo

- *Mensuales:*

- Ocupados según Seguridad Social

Con el cambio de la CNAE se ha tenido que realizar una transformación a partir de los datos de afiliados a la Seguridad Social según la CNAE 93 para conseguir una serie desde enero de 1995 según la CNAE 2009. Para ello utilizamos la siguiente matriz de conversión.

	Matriz de conversión	Sector CNAE 93					
		Agricultura	**Industria**	**Construcción**	**Servicios de mercado**	**Servicios de no mercado**	**Total general**
Sector CNAE 09	**Agricultura**	91,30	0,00	0,00	0,00	0,00	3,56
	Industria	0,02	98,72	0,00	0,21	2,00	16,59
	Construcción	0,00	0,14	100,00	1,09	0,00	12,29
	Servicios de mercado	8,68	0,94	0,00	97,15	1,50	42,82
	Servicios de no mercado	0,00	0,20	0,00	1,55	96,50	24,74
	Total general	100,00	100,00	100,00	100,00	100,00	100,00

Estos datos se han obtenido a partir de la muestra común de las Cuentas de Cotización y las Afiliaciones comparando la CNAE 93 del 2º semestre de 2008 y la CNAE 2009 del 1er semestre de 2009.

Además los datos disponibles de Servicios hasta el diciembre de 1997 no distinguían entre Servicios de Mercado y no Mercado, por lo que se hizo una estimación de los mismos.

- *Trimestrales:*

- Ocupados según EPA CNAE 93 y Ocupados según EPA CNAE 2009

Las series de ocupados EPA_93 y EPA_09 tienen como periodo común de resultados de datos el año 2008, pero en los microdatos es imposible realizar la transformación llevada a cabo en los datos de la Seguridad Social por ser datos anonimizados.

El indicador de empleo que se ha incorporado para construir cada indicador sintético sectorial en la Contabilidad Trimestral de Cantabria (CTC) ha sido el de Afiliados a la Seguridad Social por tener mejores características que el de la EPA.

3.2.- Otros indicadores sectoriales

3.2.1.- Agricultura

- *Mensuales:*
  - Venta de Ganado
  - Asistencia de Ganado

    No se dispone de las ventas y asistencias de ganado de marzo y abril de 2001 por lo que se ha hecho una estimación.

  - Pesca desembarcada
  - Consumo de fertilizantes
  - Producción agrícola

    En este caso hay un problema y es que ésta no sólo se genera cuando se recoge, por lo que hay que buscar un indicador adecuado para realizar una imputación de esa producción. Si tenemos en cuenta que desde que se siembra hasta que se cosecha se tiene una producción en curso, el consumo de fertilizantes u otros indicadores de coste pueden ser buenos indicadores para realizar la imputación de la producción agrícola.

En agricultura, ganadería, silvicultura y pesca en el caso de la CTC se han considerados series relativas al sector ganadero (asistencias de ganado) y de la pesca (volumen de desembarcos), así como información sobre los ocupados del sector para la construcción del indicador sintético.

3.2.2.- Industria

- *Mensuales:*
  - Series de Importaciones y Exportaciones de Bienes Intermedios, Bienes de Consumo y Bienes de Capital deflactadas

    Para conseguir información de todas estas series de importaciones y exportaciones en términos reales se han transformado deflactándolas por sus respectivos Índices de valor unitario.

  - Índice de Producción Industrial (IPI)

    En los datos del IPI se han producido varios cambios de base además del cambio de la CNAE por lo que se han tenido que realizar estimaciones desde enero de 1995 hasta diciembre de 2001 para obtener una serie homogénea.

  - Índice de Precios Industriales (IPRI)

    Los datos del IPRI solo están disponibles desde enero de 2002, por lo que se han tenido que realizar estimaciones desde enero de 1995 hasta diciembre de 2001 trasladando el crecimiento mensual de España para obtener una serie completa.

  - Nivel y tendencia de Cartera de Pedidos de Encuesta de Coyuntura Industrial
  - Consumo de energía eléctrica industrial

En la industria y energía en el caso de la CTC, se han utilizado las series del Índice de Producción Industrial de Cantabria, las exportaciones de bienes industriales deflactadas por los índices de valores unitarios nacionales y el empleo del sector como indicadores para el cálculo del indicador sectorial.

3.2.3.- Construcción

- *Mensuales:*

- Series de Consumo de Cemento y Consumo de Cemento estimado

Los datos definitivos de las CC.AA. de Consumo de Cemento llegan de OFICEMEN con un desfase temporal importante. Al disponer de información provisional más actualizada de la misma fuente pero por zonas, hemos seleccionado los datos de la zona del Cantábrico, que incluye Galicia, Asturias y Cantabria, para realizar una estimación de los meses donde todavía no tenemos la información desagregada para Cantabria.

- Viviendas a construir periodificadas

- Licitación periodificada en constantes

Tanto en Vivienda a construir como en Licitación hay una diferencia temporal entre el periodo de contabilización y el periodo de ejecución de los mismos, por lo que es necesario distribuir tales datos entre el periodo en que realmente se lleva a cabo la ejecución de los mismos. Además la serie de Licitación, que estaba en términos corrientes, se ha transformado a términos constantes mediante el Índice de Costes del Sector Construcción de la Ingeniería Civil.

La vivienda de nueva construcción debemos periodificarla de la forma que se establece en las publicaciones de la Contabilidad Trimestral de España. Se supone un retraso de un mes desde la concesión de la licencia hasta el comienzo de la obra y la inversión se distribuye en 18 meses según los siguientes porcentajes:

MES	TANTO POR MIL
1	8.08
2	16.16
3	22.22
4	30.30
5	38.38
6	46.46
7	53.23
8	62.63
9	71.72
10	78.79
11	81.82
12	83.94
13	83.94
14	82.93
15	80.81
16	75.66
17	66.67
18	16.16

Estos porcentajes, establecidos según estudio realizado por expertos, se pueden encontrar en el apéndice 2 de la publicación "Boletín Trimestral de Coyuntura de España" de cualquier trimestre.

Para la rehabilitación se considera el periodo de 9 meses, por lo que se agrupa cada periodo de los anteriores en dos. Asimismo, se establece el comienzo de las obras en el periodo t+1.

MES	TANTO POR MIL
1	24,24
2	52,52
3	84,84
4	115,86
5	150,51
6	165,76
7	166,87
8	156,47
9	82,83

Para la edificación no residencial el calendario se establece de forma uniforme en 9 meses, con un lap de 1 mes, por lo que el periodo es de t+1 a t+10.

Para la obra pública el retardo es de 5 meses, tiempo que se estima entre la publicación de la licitación en el boletín y el comienzo de las obras. El plazo de ejecución será el plazo medio por año nacional que suele rondar los 18-19 meses, no siendo fijo, por lo que hay que ir cambiándolo según se modifique. La variabilidad de este plazo depende del tipo de obra que se esté haciendo en cada momento (si predominan las autovías, grandes infraestructuras, el plazo es más amplio).

- Hipotecas en términos constantes

Para tener una serie homogénea es necesario realizar enlaces entre la serie actual con sus respectivas históricas. Además ha tenido que ser transformada a términos constantes mediante el Índice de Costes del Sector Construcción de la Edificación con el fin de tener una serie comparable temporalmente.

En el caso de la Contabilidad Trimestral de Cantabria el indicador se construye en constantes, pero también se hubiera podido construir en corrientes. Existen varias posibilidades para construir el indicador de construcción en corrientes, podríamos utilizar la vivienda hipotecada o el valor de la superficie de vivienda o, incluso, construir nuestro propio indicador a partir de la vivienda residencial y la edificación no residencial. En la construcción en el caso de la CTC para construir el indicador sintético se usan el consumo de cemento y el empleo del sector.

3.2.4.- Servicios de Mercado

- *Mensuales:*

- Pernoctaciones hoteleras de la Encuesta de Ocupación Hotelera (EOH)

Sólo existen datos de la EOH desde enero de 1999, por lo que se completó la serie con la información de la Encuesta de movimientos de viajeros en establecimientos hoteleros.

- Tráfico de mercancías del Puerto de Santander

- Indicador de Actividad del Sector Servicios (IASS)

Se han realizado una serie de agrupaciones: Comercio, Turismo, Transporte, Tecnologías de la Información y Servicios a Empresas. Pero sólo existen datos del IASS desde enero de 2005, por lo que únicamente se puede utilizar para comprobar evoluciones desde esa fecha, pero la longitud de la serie no es suficiente como para utilizarla para construir el indicador sectorial.

- Índice de Comercio al por menor (ICM)

  Como sucede con el IASS, sólo existen datos del ICM de la nueva base desde enero de 2005, por lo que únicamente se puede utilizar para comprobar evoluciones desde esa fecha, pero la longitud de la serie no es suficiente como para utilizarla para construir el indicador sectorial.

- *Trimestrales:*

- Transporte de Mercancías por Carretera

En los servicios destinados a la venta en el caso de la CTC se tienen en cuenta un indicador de actividad del turismo, el número de pernoctaciones en establecimientos hoteleros dado por la Encuesta de Ocupación Hotelera, y el empleo del sector.

3.2.5.- Servicios de no mercado

- *Trimestrales:*

- Gastos del Gobierno de la Comunidad Autónoma en los capítulos I y II

  A partir del primer trimestre de 2000 los datos trimestrales han sido extraídos de la Liquidación del Presupuesto de la Comunidad Autónoma, normalmente publicados en el Boletín Oficial de la Comunidad, y deflactados por los precios de la Contabilidad Nacional Trimestral (CNT). Para completar la serie trimestral desde 1995, a partir de los datos anuales se efectuó una trimestralización según los pesos del periodo 2000-2006.

En los servicios no destinados a la venta en el caso de la CTC se utilizan como indicadores parciales la ejecución del presupuesto de gastos del gobierno de Cantabria y el empleo en el sector.

3.2.6.- Impuestos

- *Mensuales:*

- Series de Importaciones en corrientes de España y de Cantabria

- *Trimestrales:*

- Valor de partidas “IVA”, “impuestos netos a las importaciones”, "resto de impuestos" Brutos y Ajustados de Estacionalidad de Impuestos corrientes de la CNTR

- *Anuales:*

- Series de Recaudación IVA de España y de Cantabria
- Series de Recaudación Capítulo II Impuestos Indirectos de España y de Cantabria
- Series de Recaudación Impuestos Especiales de España y de Cantabria

En los impuestos netos sobre los productos (es decir, impuestos especiales, subvenciones e impuesto sobre el valor añadido) en el caso de la CTC, se tiene un único indicador, que se obtiene regionalizando por indicadores parciales la información de la CNTR española, a partir de la información de la recaudación tributaria y de las importaciones. No obstante, el IVA se hubiera podido repartir por el consumo a partir de la Encuesta de Presupuestos Familiares, pero al desaparecer esta existen otras posibilidades para efectuar el reparto, como utilizar los Valores Añadidos Brutos de la Contabilidad Trimestral o las Ventas del Sector Minorista, etc.

3.3.- Contabilidad Regional Anual (INE)

- *Anuales:*

  - Series de Valor Añadido bruto a precios corrientes de cada uno de los sectores e impuestos

  - Series de Índices de volumen de cada uno de los sectores e impuestos

Estas series han sido incorporadas en cada uno de los cálculos de los indicadores sintéticos sectoriales.

3.4.- Selección de indicadores para la Contabilidad Trimestral

- *Trimestrales:*

  - Valor Añadido bruto a precios corrientes Brutos y Ajustados de Estacionalidad de cada uno de los sectores e impuestos

  - Índice volumen Brutos y Ajustados de Estacionalidad de cada uno de los sectores e impuestos

Una Contabilidad Trimestral se puede elaborar desagregando magnitudes corrientes o constantes, y obteniendo los deflactores como diferencia, o bien obteniendo los valores corrientes a partir de indicadores trimestrales de precios e indicadores de valores constantes (o viceversa). En Cantabria los valores a precios corrientes, se obtiene con los valores a precios constantes, e índices de precios sectoriales procedentes de la Contabilidad Nacional Trimestral (CNT). Únicamente en el caso de la industria se ha considerado el uso como deflactor un indicador regional de precios: el Índice de Precios Industriales (IPRI).

Indicadores seleccionados en la Contabilidad Trimestral de Cantabria (CTC)

Sector	Variables	Periodicidad
Agricultura, ganadería, silvicultura y pesca	Asistencias de ganado en el mercado de Torrelavega (número de cabezas)	Mensual
	Pesca desembarcada (Tm.)	Mensual
	Afiliados a la Seguridad Social	Mensual
	Deflactor del VAB de la CNTR	Trimestral
Industria y energía	Índice de producción Industrial	Mensual
	Exportaciones de bienes industriales deflactadas	Mensual
	Afiliados a la Seguridad Social	Mensual
	Deflactor: Índice de precios industriales	Trimestral
Construcción	Consumo de cemento	Mensual
	Afiliados a la Seguridad Social	Mensual
	Deflactor del VAB de la CNTR	Trimestral
Servicios de mercado	Pernoctaciones en instalaciones hoteleras	Mensual
	Afiliados a la Seguridad Social	Mensual
	Deflactor del VAB de la CNTR	Trimestral
Servicios de No mercado	Gastos del Gobierno de Cantabria en los capítulos I y II	Trimestral
	Afiliados a la Seguridad Social	Mensual
	Deflactor del VAB de la CNTR	Trimestral
Impuestos	Regionalización de los impuestos netos CNTR España	Trimestral
	Deflactor del VAB de la CNTR	Trimestral

3.5.- Recomendaciones de Eurostat

Existe un manual de Eurostat que recoge en su Tabla 4.1 las fuentes que pueden ser utilizadas para obtener indicadores para las cuentas trimestrales.

Table 4.1: Production Approach - Sources

NACE A17- aggregates	Aggregates	Sources
Agriculture, hunting and forestry	• Value added	Difference between output and intermediate input
	• Output	
	– Wheat and barley	– Marketing boards
	– Grains and crops	– Harvesting data
	– Livestock slaughtering	– Quantity of meat produced and prices obtained from abattoirs – Numbers of animals slaughtered
	– Wholemilk and eggs	– Data on deliveries
	– Wool production	– Quantity data • Physical quantity indicators • Other Indicators • Quantitative data multiplied by average producers' prices
	– Animal production	– Sales
	– Crop production	– Allocation of annual estimates
	– Fruits and vegetables	– Quantities and values delivered on auctions
	– Horticultural products	– Trade Association about yearly turnover (allocation to the quarters)
	– Forestry	– State forestry sales – Indicators – Labour force in forestry – Quantity of timber felled – Same movement as the agriculture aggregates – Trend estimation • Official Economic Ministry estimates
	• Intermediate consumption	• Administrative data • Statistics on quantities • Suitable indicators: – Costs of marketing, fodder, fuels,... – Fodder and consumption of fertilisers – Judgmental estimation
Fishing		• Indicators • Value and size of catches • Sales revenue and quantities • Same movement as the agriculture aggregates • Amount of fish slaughtered on fish farms • Fishermen's landings • Trend extrapolation
Mining and quarrying	– Petroleum	• Interpolation/extrapolation according to quarterly movements • Sub-indices of industrial production • Quantity indicators: – Metres drilled

NACE A17- aggregates	Aggregates	Sources
Manufacturing		• Index of industrial production • Interpolation/extrapolation of annual values according to quarterly movements • Sample surveys • Census surveys • Production information • Trend extrapolation
Electricity, gas and water supply		• Physical quantities • Interpolation/extrapolation of annual values according to quarterly movements • Sales • Sub-indices of industrial production • Consumption of inputs
Construction		• Turnover of general trade contractors in general building construction and engineering • Employment figures • Volume index • Interpolation and extrapolation of annual data • Investments
	– Residential constructions – Non-residential constructions	• Indicators: – Estimate of work put in place by type of dwelling – Estimate of work put in place by type of structure – Building and engineering construction surveys
	– Public sector constructions	– Estimates of work done – Employment indicators – Budget data
Wholesale and retail trade, repair of motorvehicles, motorcycles and personal and households goods		• Sales • Gross sales • Surveys of private enterprises • Sales by public market authorities • Interpolation/extrapolation of annual values according to quarterly turnover indicators • Output volume indicators • Activity indicators • Turnover statistics (e.g. from VAT statistics) • Volume trade index • Sum of trade margins

NACE A17- aggregates	Aggregates	Sources
Hotels and restaurants		• Output volume indicators • Balance of payments data • Statistics on nights spent by non-resident • Estimation according to total domestic consumption • VAT statistics on turnover • Sales in restaurants • Nights spent in hotels
Transport, storage and communications	• Transport – Air transport, rail freight, pipeline system, water transport, ferry operations,... – Transit operation – Road haulage – Taxicab services • Communications	• Indicators: – Passengers – Weight/volume-kilometres – Revenues – Real output of industries relying on road haulage – Number of workers – Turnover according to VAT statistics • Indicators: – Audience viewing hours – Sample surveys of radio advertising sales – Number of subscribers to cable services – Data receipts for letters, parcels and telephone calls – Gross revenue of the postal service
Financial intermediation		• Indicators: – Revenue – Stock market volume traded – Issues of stocks and bonds – Mutual fund sales – Extrapolation using hours worked – Volume indices (e.g. number of cheque account transactions) – Employment
Real estate, renting and business activities	– Ownership of dwellings – Accommodation services	– Estimates of end period housing stock – Number of rooms and occupancy rates – Final consumption expenditure of households on dwelling rent – Turnover from VAT statistics
Public administration and defence; compulsory social security		– Number of employees – Extrapolation using hours worked – Wages and salaries

NACE A17- aggregates	Aggregates	Sources
Education		– Number of employees – Extrapolation using hours worked – Wages and salaries – Pupil weeks taught
Health and social work		– Number of employees – Wages and salaries – Extrapolation using hours worked – Government output
Other community, social and personal service activities		– Labour inputs – Extrapolation using hours worked
Private households with employed persons		• Labour force survey
Extra-territorial organisa-tions and bodies		

4. ANÁLISIS DE SERIES TEMPORALES

4.1.- Introducción

El presente epígrafe pretende ser una breve introducción al estudio de las series temporales, las cuales poseen una gran importancia en el campo de la Economía dada la abundancia de este tipo de observaciones; de hecho, las series temporales constituyen la mayor parte del material estadístico con el que trabajan los economistas.

Pero, ¿qué es una serie temporal? Por definición, una *serie temporal es una sucesión de observaciones de una variable realizadas generalmente a intervalos regulares de tiempo*. Según realicemos la medida de la variable considerada podemos distinguir distintos tipos de series temporales:

- Discretas o Continuas, en base al intervalo de tiempo considerado para su medición.
- Flujo o Stock. Cada observación puede medir la acumulación de *algo* durante un periodo de tiempo o su valor en un momento dado. En Economía, se dice que una variable es de tipo flujo si está referida a un período determinado de tiempo (un día, un mes, un año, etc.) como, por ejemplo, las ventas de una empresa (semanales, mensuales, etc.). Por su parte, se dice que una variable es de tipo stock si está referida a una fecha determinada (por ejemplo, el 31 de Diciembre de cada año). Un ejemplo de datos de tipo stock sería la cotización de cierre de las acciones de esa misma empresa, ya que sólo puede ser registrado a una fecha y hora determinadas. Otro ejemplo de serie temporal de tipo stock sería el paro registrado, es decir, el número de personas registradas en las oficinas públicas de empleo el último día de cada mes desde el año 2000 hasta hoy.
- Dependiendo de la unidad de medida, podemos encontrar series temporales en unidades monetarias o en diversas magnitudes físicas (kilogramos, litros, millas, etc.).
- En base a la periodicidad de los datos, podemos distinguir series temporales de datos diarios, semanales, mensuales, trimestrales, anuales, etc.

Antes de profundizar en el análisis de las series temporales es necesario señalar que, para llevarlo a cabo, hay que tener en cuenta los siguientes supuestos:

- Se considera que existe una cierta estabilidad en la estructura del fenómeno estudiado. Para que se cumpla este supuesto será necesario estudiar períodos lo más homogéneos posibles.

- Los datos deben ser homogéneos en el tiempo, o, lo que es lo mismo, se debe mantener la definición y la medición de la magnitud objeto de estudio. Este supuesto no se da en muchas de las series económicas, ya que es frecuente que las estadísticas se perfeccionen con el paso del tiempo, produciéndose saltos en la serie debidos a un cambio en la medición de la magnitud estudiada. Un caso particularmente frecuente es el cambio de base en los índices de precios, de producción, etc. Tales cambios de base implican cambios en los productos y las ponderaciones que entran en la elaboración del índice que repercuten considerablemente en la comparabilidad de la serie en el tiempo.

El objetivo fundamental del estudio de las series temporales es el conocimiento del comportamiento de una variable a través del tiempo para, a partir de dicho conocimiento, y bajo el supuesto de que no van a producirse cambios estructurales, poder realizar predicciones, es decir, determinar qué valor tomará la variable objeto de estudio en uno o más períodos de tiempo situados en el futuro, mediante la aplicación de un determinado modelo previamente seleccionado.

Dado que en la mayor parte de los problemas económicos los agentes se enfrentan a una toma de decisiones bajo un contexto de incertidumbre, la predicción de una variable reviste una importancia notoria pues supone, para el agente que la realiza, una reducción de la incertidumbre y, por ende, una mejora de sus resultados. Indudablemente, la calidad de las previsiones dependerá en buena medida del proceso generador de la serie: así, si la variable observada sigue algún tipo de esquema o patrón de comportamiento fijo (por ejemplo, *una tendencia determinista*) seguramente obtengamos predicciones más fiables, con un grado de error bajo. Por el contrario, si la serie no sigue ningún patrón de comportamiento específico (*serie totalmente aleatoria*), seguramente nuestras predicciones carecerán de validez por completo.

Generalmente, en el caso de las series económicas no existen variables deterministas o aleatorias puras, sino que contienen ambos tipos de elementos. El objeto de los métodos de previsión cuantitativos es conocer los componentes subyacentes de una serie y su forma de integración, con objeto de realizar la predicción de su evolución futura.

Dentro de los métodos de predicción cuantitativos, se pueden distinguir dos grandes enfoques alternativos:

- Por un lado, el análisis univariante de series temporales mediante el cual se intenta realizar previsiones de valores futuros de una variable utilizando como información la contenida en los valores pasados de la propia serie temporal. Dentro de esta metodología se incluyen los métodos de descomposición y la familia de modelos ARIMA univariantes que veremos más adelante.

- El otro gran bloque dentro de los métodos cuantitativos estaría integrado por el análisis multivariante de tipo causal, denominado así porque en la explicación de la variable o variables objeto de estudio intervienen otras adicionales de ella o ellas mismas.

4.2.- Métodos de descomposición de series temporales

Tradicionalmente, en los métodos de descomposición de series temporales se parte de la idea de que la serie temporal se puede descomponer en todos o algunos de los siguientes componentes:

- Tendencia *(T)*, que representa la evolución de la serie en el largo plazo
- Ciclo *(C)*, que refleja las fluctuaciones de carácter periódico, pero no necesariamente regular, a medio plazo en torno a la tendencia. Este componente es frecuente hallarlo en las series económicas, y se debe a los cambios en la actividad económica.

  Para la obtención del ciclo y la tendencia es necesario disponer de una serie larga y de un número de ciclos completo, por lo que, a veces, resulta difícil separar ambos componentes. En estos casos resulta útil englobar ambos componentes en uno solo, denominado ciclo-tendencia.
- Estacionalidad (*S*): recoge aquellos comportamientos de tipo regular y repetitivo que se dan a lo largo de un período de tiempo, generalmente igual o inferior a un año, y que son producidos por factores tales como las variaciones climatológicas, las vacaciones, las fiestas, etc.
- Irregular *(I)*, que recoge los movimientos no sistemáticos, los pequeños efectos accidentales, o erráticos, como resultado de hechos no previsibles.

En este punto, cabe señalar que en una serie concreta no tienen por qué darse los cuatro componentes. Así, por ejemplo, una serie con periodicidad anual carece de estacionalidad.

La asociación de estos cuatro componentes en una serie temporal, Y, puede responder a distintos esquemas; así, puede ser de tipo aditivo:

$$Y=T+C+S+I$$

También puede tener una forma multiplicativa:

$$Y=T\times C\times S\times I$$

O bien ser una combinación de ambos, por ejemplo, el modelo mixto:

$$Y=T\times C\times S+I$$

Si al realizar la representación gráfica se observa que la magnitud de las fluctuaciones son más o menos regulares a lo largo de la serie, sin verse afectadas por la tendencia (véase Fig. 1), se puede emplear el esquema aditivo.

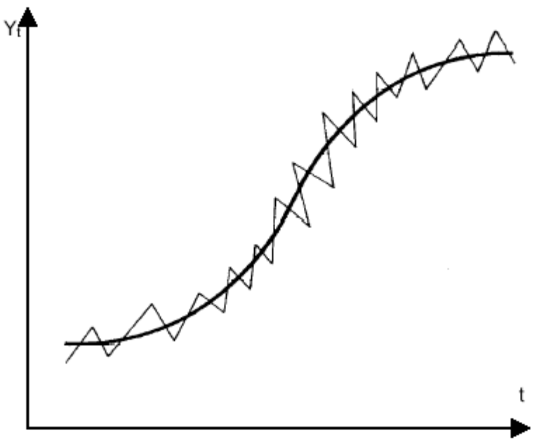

Figura 1. Esquema aditivo

Si, por el contrario, se observa que la magnitud de las fluctuaciones varía con la tendencia, siendo más altas cuando ésta es mayor y más bajas cuando su nivel es menor (véase Fig. 2), se debe adoptar entonces el esquema multiplicativo.

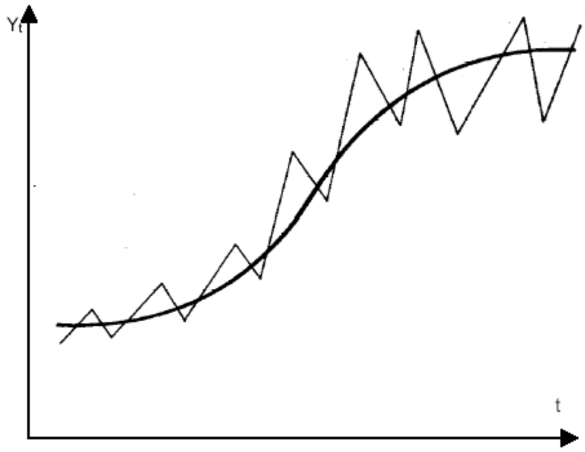

Figura 2. Esquema multiplicativo.

El análisis de series univariantes está basado fundamentalmente en la teoría de los procesos estocásticos. Los modelos univariantes poseen carácter estocástico, pues entre los componentes incluidos en el modelo, al menos uno tiene carácter probabilístico, hecho que confiere a la serie un rasgo aleatorio. En este sentido, los modelos predictivos univariantes de series temporales denominados modelos ARIMA propuestos por Box y Jenkins en 1970 en su libro *Time Series Analysis. Forecasting and Control* serán el punto de atención de este apartado.

4.3.- Procesos Estocásticos

Podemos definir un proceso estocástico como un conjunto de variables aleatorias $\{Y_t\}$ asociadas a distintos instantes de tiempo *t*. Así, en cada período o momento temporal se dispone de una variable que tendrá su correspondiente distribución de probabilidad, tal que si consideramos el proceso para $t = 1$, por ejemplo, Y_1 es una variable aleatoria que tomará diferentes valores con diferentes probabilidades.

Por su parte, diremos que una serie temporal es un conjunto de observaciones o medidas realizadas secuencialmente en intervalos predeterminados y, con carácter general, de igual duración.

Por ejemplo, consideremos el PIB anual a precios de mercado desde 1970 hasta 1990 para un determinado país. En cada año, dicha variable se comportará como una variable aleatoria que puede tomar infinitos valores con distintas probabilidades. La sucesión de observaciones efectuadas cada año desde 1970 hasta 2010 en ese país formaría la serie temporal.

Por tanto, la relación existente entre una serie temporal y el proceso estocástico que la genera será análoga a la que hay entre una muestra y la variable aleatoria de la que procede. De esta forma podemos considerar que una serie temporal es una muestra de de un proceso estocástico, formada por una sola observación de cada una de las variables que componen el proceso durante un intervalo de tiempo determinado. La tarea del investigador será por tanto inferir la forma del proceso estocástico a partir de las series temporales que genera.

Un proceso estocástico $\{Y_t\}$ se suele describir mediante las siguientes características: medias, varianzas, autocovarianzas y coeficientes de autocorrelación.

La función de medias, μ_t, es la sucesión de las esperanzas matemáticas de las variables que componen el proceso, a lo largo del tiempo:

$$E(Y_t) = \mu_t, \qquad t = 1, 2,3, \ldots$$

Por su parte, la función de varianzas de un proceso estocástico es una sucesión de varianzas, una para cada variable del proceso:

$$var(Y_t) = \sigma_t^2, \qquad t = 1, 2,3, \ldots$$

La función de autocovarianzas está formada por las covarianzas entre cada par de variables del proceso en dos momentos cualquiera t y t+k:

$$cov(Y_t, Y_{t+k}) = \gamma_{t,t+k} \qquad t = 1, 2,3, \ldots \quad k= 0, \pm1, \pm2, \pm3, \ldots$$

La función de autocorrelación es el conjunto de los coeficientes de autocorrelación o coeficientes de correlación entre cada par de variables que componen el proceso:

$$corr(Y_t, Y_{t+k}) = \frac{\gamma_{t,\,t+k}}{\sigma_t \sigma_{t+k}} = \rho_{t,t+k} \qquad t = 1, 2,3, \ldots \quad k= 0, \pm1, \pm2, \pm3, \ldots$$

4.4.- Procesos Estocásticos Estacionarios

Se dice que un proceso es (débilmente) estacionario si sus momentos de primer y segundo orden son estables en el tiempo. Es decir, si todas las variables del proceso tienen la misma media y varianza (finitas):

$$E(Y_t) = \mu, \qquad var(Y_t) = \sigma^2 \qquad t = 1, 2,3, \ldots$$

y el valor de la covarianza entre dos elementos Y_t e Y_{t+k} no depende del momento (t), sino de la distancia o desfase entre ellos (k). Es decir:

$$cov(Y_t, Y_{t+k}) = \gamma_k \qquad t = 1, 2,3, \ldots \quad k= 0, \pm1, \pm2, \pm3, \ldots$$

de modo que $var(Y_t) = \gamma_0$.

En un proceso estacionario, los coeficientes de autocorrelación entre dos elementos sólo depende de su distancia k y $corr(Y_t, Y_{t+k}) = corr(Y_t, Y_{t-k})$. Así, el coeficiente de autocorrelación de orden k de un proceso estacionario, ρ_k, se define:

$$\rho_k = \gamma_k/\sigma^2 = \gamma_k/\gamma_0 \qquad k= 0, \pm1, \pm2, \pm3, \ldots$$

El conjunto de coeficientes ρ_0, ρ_1, ρ_2, ..., forman la función de autocorrelación simple (*fas*) y tiene las siguientes propiedades:

i) $\rho_0 = 1$ ii) $|\rho_k| \leq 1$ iii) $\rho_k = \rho_{-k}$

Otro elemento útil es la función de autocorrelación parcial (*fap*). El coeficiente de autocorrelación parcial de orden k (ρ_k^p) mide la correlación existente entre dos variables del proceso, Y_t e Y_{t+k}, una vez eliminados los efectos sobre las mismas de los elementos intermedios Y_{t+1}, Y_{t+2}, ..., Y_{t+k-1}.

Un tipo de proceso estacionario particular es el denominado ***ruido blanco***, formado por una sucesión de variables aleatorias $\{\varepsilon_t\}$ con media cero, varianza constante (σ^2) e incorrelacionadas entre sí. Se escribe $\varepsilon_t \sim RB(0, \sigma^2)$. Es decir, que para todo momento t y retardo $k \neq 0$ se cumple que:

i) $E(\varepsilon_t) = 0$ ii) $var(\varepsilon_t) = \sigma^2$ iii) $cov(\varepsilon_t, \varepsilon_{t+k}) = 0$

Si además la variable tiene distribución normal se denomina ruido blanco gaussiano y se escribe $\varepsilon_t \sim NID(0, \sigma^2)$. En la Figura 3 se muestra la representación de un ruido blanco.

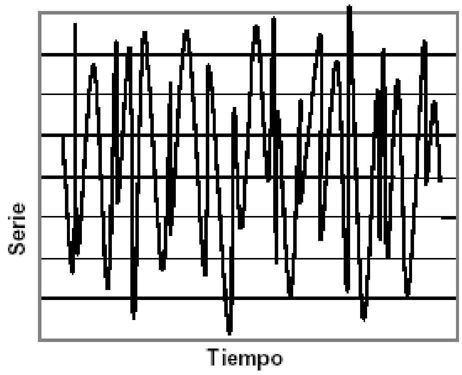

Figura 3. Representación de un ruido blanco.

Estimación de los momentos de un proceso estocástico estacionario. Dado que en la práctica se dispone de una muestra de un proceso estocástico, la serie (Y_1, Y_2, ..., Y_n), se estiman los coeficientes desconocidos del proceso (media, varianza, fas y fap) mediante los momentos muestrales equivalentes. Así:

- Media: $\hat{\mu} = \overline{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$
- Varianza: $\hat{\gamma}_0 = \frac{1}{n}\sum_{t=1}^{n} (Y_t - \overline{Y})^2$
- Autocorrelación simple (fas): $\hat{\rho}_k = r_k = \dfrac{\sum_{t=1}^{n} (Y_t - \overline{Y})(Y_{t-k} - \overline{Y})}{\sum_{t=1}^{n} (Y_t - \overline{Y})^2}$
- Autocorrelación parcial (fap): el coeficiente de autocorrelación parcial muestral de orden k, r_k^p, se define como el coeficiente b_k de la siguiente regresión de mínimos cuadrados ordinarios:

$$\hat{Y}_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \ldots + b_k Y_{t-k}$$

Pero, ¿por qué resulta importante para el investigador que el proceso analizado sea estacionario? La razón fundamental es que los modelos de predicción de series temporales que veremos a continuación están diseñados para ser utilizados con procesos estacionarios. Por ejemplo, es una condición para que los estimadores de momentos que acabamos de ver sean consistentes. Si las características del proceso cambian a lo largo del tiempo, resultará difícil representar la serie para intervalos de tiempo pasados y futuros mediante un modelo lineal sencillo.

Un ejemplo de proceso no estacionario en media es el caso de una serie temporal que presenta una marcada tendencia creciente. Como cada observación de la serie proviene de la distribución de probabilidad del período correspondiente, es razonable pensar que la esperanza matemática de dichas distribuciones también crece en el tiempo. Asimismo, si en el gráfico de la serie se observan oscilaciones desiguales en torno a la tendencia, se puede defender la no estacionariedad de la serie debido a que la varianza no es estable (véase Fig. 2).

Muchas series económicas no proceden de procesos estacionarios, sino que suelen tener una tendencia, ya sea creciente o decreciente, y variabilidad no constante. Dicha limitación en la práctica no es tan importante porque las series no estacionarias se pueden transformar en otras que sí lo son, ya que la mayor parte de las series económicas se convierten en aproximadamente estacionarias después de aplicar diferencias en una ó más etapas.

Por ello, cuando estemos analizando una serie económica que no sea estacionaria en media deberemos trabajar con la serie en diferencias, especificando y estimando un modelo para la misma. Normalmente con una o dos diferencias es suficiente, de tal forma que la serie quedará transformada como sigue:

- Tomando primeras diferencias, tendremos $\Delta Y_t = Y_t - Y_{t-1}$.
- En caso de que el resultado anterior no sea estacionario en media, deberemos tomar una segunda diferencia tal que:

$$\Delta(\Delta Y_t) = \Delta^2 Y_t = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$$

Posteriormente la predicción que realicemos con las series transformadas habrá que trasladarla a una predicción para la serie original, en cuyo análisis está interesado inicialmente el investigador. Si además observamos que la serie presenta no estacionariedad en varianza, deberemos transformarla tomando logaritmos (u otra opción de la transformación de Box-Cox) antes de aplicar diferencias en la serie.

5. MODELOS ARIMA

La representación formal de los procesos aleatorios que generan series reales se puede realizar a través de los modelos lineales de series temporales desarrollados especialmente en Box y Jenkins (1970). Considerando que la serie temporal ha sido generada por un proceso estocástico estacionario, en este epígrafe pasamos a describir los posibles modelos teóricos que permitan explicar el comportamiento de la misma y, por tanto, el de su proceso generador.

5.1.- Procesos estocásticos lineales discretos

Se dice que un proceso estocástico discreto es *lineal* si se puede expresar de la forma:

$$Y_t = \mu + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \dots = \mu + \varepsilon_t + \textstyle\sum_i \Psi_i \varepsilon_{t-i} \qquad (1)$$

con μ, Ψ_1, Ψ_2, ..., parámetros (normalmente desconocidos) y $\{\varepsilon_t\}$ un ruido blanco de media 0 y varianza σ^2. Con frecuencia $\{\varepsilon_t\}$ se denomina innovación porque se corresponde con el error de predicción un periodo hacia delante que cometemos si utilizamos la predicción adecuada. Es decir, es la parte de Y_t no predecible aunque se utilice óptimamente toda la información pasada, Y_{t-1}, Y_{t-2}, ...

¿Por qué el análisis de series temporales se ha centrado en este tipo de procesos? La principal justificación es el **Teorema de la descomposición de Wold (1938):** Un proceso estocástico discreto, estacionario en covarianza, se puede representar unívocamente como la suma de dos procesos mutuamente incorrelacionados, $Y_t = D_t + W_t$, donde D_t es un proceso puramente determinista y W_t es un proceso puramente no determinista, que se puede escribir como un media móvil infinita (1).

Restringiremos el estudio a los procesos lineales discretos que dependan de pocos parámetros conocidos como son los procesos autorregresivos de medias móviles de órdenes *p* y *q*, abreviadamente ARMA (p,q), que se definen

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim RB(0, \sigma^2) \qquad (2)$$

donde δ es un término constante. Estos procesos se pueden escribir de otras dos formas:

- En la forma media móvil, es decir, en función únicamente de la innovación,

$$Y_t = \mu + \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \dots$$

- En la forma autorregresiva, es decir, en función del pasado de la variable y la innovación actual,

$$Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t$$

En principio, consideramos únicamente los procesos ARMA que posean las propiedades de estacionariedad y ergodicidad, que garantizan la resolución del problema de estimación al disponer de estimadores consistentes. Esta última hipótesis implica que los parámetros de la forma media móvil cumplen la condición $\lim_{i\to\infty} \Psi_i = 0$. Además, introducimos la hipótesis de invertibilidad con objeto de que el modelo sea útil para predecir, y que implica la condición $\lim_{i\to\infty} \varphi_i = 0$ en la forma autorregresiva.

Pasamos a mostrar las propiedades que caracterizan a los distintos procesos ARMA. Dentro de esta clase general de modelos encontramos dos casos particulares, los procesos autorregresivos (cuando q=0) y los procesos médias móviles (si p=0).

5.2.- Procesos Autorregresivos (AR(p))

Los procesos autorregresivos son aquellos que representan los valores de una variable durante un instante del tiempo en función de sus valores precedentes. Así, un modelo autorregresivo de orden p, AR(p), tendrá la siguiente forma:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

Veamos a continuación las características particulares de dos procesos autorregresivos, el de orden 1 ó AR(1) y el de orden 2 ó AR(2). Posteriormente, los resultados obtenidos se generalizarán al caso de los autorregresivos de orden p.

5.2.1.- Modelos autorregresivos de primer orden AR(1)

Un proceso autorregresivo de primer orden se define:

$$Y_t = \delta + \phi_1 Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

y se caracteriza por:

- Ser siempre invertible.
- Ser estacionario si $|\phi_1| < 1$.

Si el proceso es estacionario:

1. Su *media* es $\mu = \delta\,(1 - \phi_1)^{-1}$. Por tanto, δ es el parámetro que determina si la media del proceso es cero o no.

2. Su *varianza* es $\gamma_0 .= \sigma^2 \left(1 - (\phi_1)^2\right)$.

3. Su *función de autocorrelación simple* sigue la estructura

$$\rho_k = \phi_1\,\rho_{k-1} = (\phi_1)^k$$

   A la vista del resultado, podemos concluir que los valores de la función de autocorrelación son las sucesivas potencias del parámetro ϕ_1. Por tanto, cualquier elemento del proceso Y_t está correlacionado con cualquier valor pasado o futuro del proceso, y se dice que es un proceso de memoria larga o infinita. La condición de estacionariedad $|\phi_1| < 1$ garantiza que los sucesivos valores de los coeficientes ρ_k convergen a cero, si bien la función puede presentar dos aspectos distintos, dependiendo del signo de ϕ_1 como puede observarse en la Figura 4.

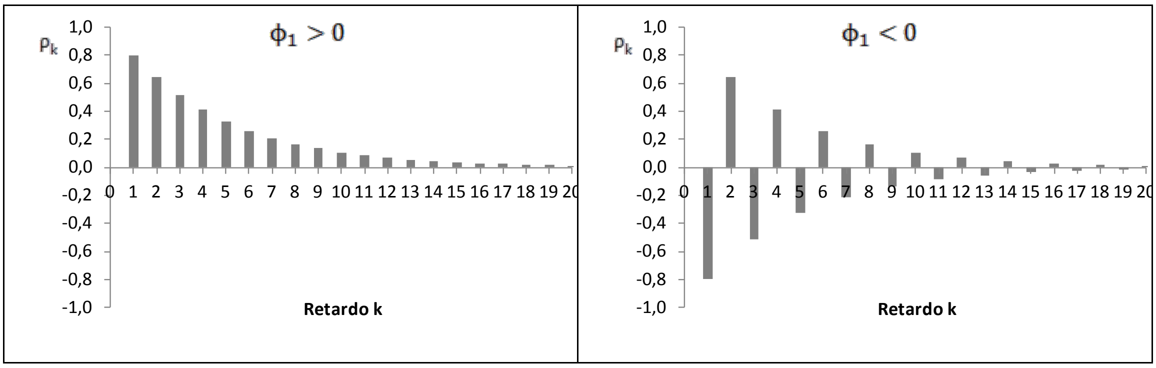

Figura 4. Correlograma simple de un AR(1)

4. Su *función de autocorrelación parcial* (fap) es: $\rho_k^p = \begin{cases} \phi_1 & \text{si } k=\pm 1 \\ 0 & \text{si } |k|>1 \end{cases}$

   Por tanto, su correlograma parcial tiene un único valor distinto de cero en k=1 (véase Fig. 5).

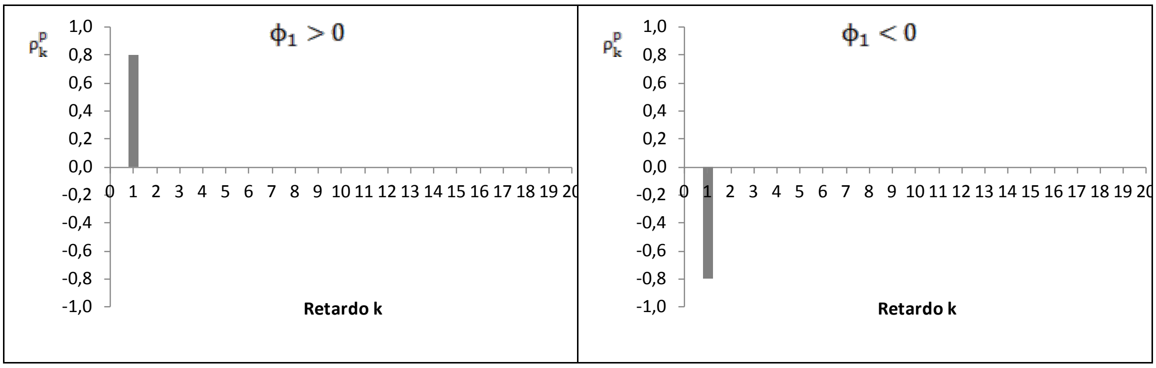

Figura 5. Correlograma parcial de un AR(1)

Pasamos a demostrar los puntos 1, 2 y 3.

1. Un proceso estacionario en media verifica que $\mu = E(Y_t) = E(Y_{t-1})$. Además $E(\varepsilon_t)=0$, de forma que:

$$E(Y_t) = \delta + \phi_1 E(Y_{t-1}) + E(\varepsilon_t) \quad \Rightarrow \quad \mu = \delta + \phi_1 \mu \quad \Rightarrow \quad \mu = \delta(1-\phi_1)^{-1}$$

2. La estacionariedad en varianza implica que $\gamma_0 = \text{var}(Y_t) = \text{var}(Y_{t-1})$, de modo que:

$$\text{var}(Y_t) = (\phi_1)^2 \text{var}(Y_{t-1}) + \text{var}(\varepsilon_t) + 2\,\phi_1 \text{cov}(Y_{t-1},\varepsilon_t)$$

Y_{t-1} es función lineal de elementos pasados del ruido blanco ε_{t-1}, ε_{t-2}, ε_{t-3}, ... que están incorrelacionados con ε_t, por lo que $\text{cov}(Y_{t-1}, \varepsilon_t)=0$ y

$$\gamma_0 = \phi_1^2 \gamma_0 + \sigma^2 \quad \Rightarrow \quad \gamma_0 = \frac{\sigma^2}{1-\phi_1^2}$$

3. Para obtener la fas, partimos de la función de autocovarianzas γ_k. La autocovarianza de orden uno es:

$$\gamma_1 = \text{cov}(Y_t, Y_{t-1}) = \text{cov}(\delta + \phi_1 Y_{t-1} + \varepsilon_t, Y_{t-1}) = \phi_1 \text{var}(Y_{t-}) + \text{cov}(\varepsilon_t, Y_{t-1})$$

Así, $\gamma_1 = \phi_1 \gamma_0$. De forma similar, tenemos que la autocovarianza de orden k cumple:

$\gamma_k = cov(Y_t, Y_{t-k}) = cov(\delta + \phi_1 Y_{t-1} + \varepsilon_t, Y_{t-k}) = \phi_1 var(Y_{t-}) + cov(\varepsilon_t, Y_{t-k}) = \phi_1 \gamma_{k-1} = (\phi_1)^k \gamma_0$

A partir de este resultado podemos obtener la función de autocorrelación simple (*fas*) para un proceso autorregresivo de orden 1:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\phi_1^k \gamma_0}{\gamma_0} = \phi_1^k$$

Podemos expresar de una forma alternativa la condición de estacionariedad. Para ello debemos introducir lo que se conoce como el operador de retardos L^s el cual, aplicado sobre una variable en un periodo del tiempo *t*, la retarda *s* períodos. Así, por ejemplo, $L^3 Y_t = Y_{t-3}$.

Aplicando el operador de retardos en el caso de un proceso AR(1) tenemos que:

$$Y_t = \delta + \phi_1 Y_{t-1} + \varepsilon_t \quad \Rightarrow \quad Y_t = \delta + \phi_1 L Y_t + \varepsilon_{tt} \quad \Rightarrow \quad Y_t - \phi_1 L Y_t = \delta + \varepsilon_t$$

Utilizando el operador polinomial de retardos, $(1 - \phi_1 L) = \Phi(L)$, el proceso queda:

$$\Phi(L)\, Y_t = (1 - \phi_1 L)\, Y_t = \delta + \varepsilon_t$$

La condición de estacionariedad es equivalente a la condición de que la raíz de $\Phi(L) = 0$ caiga fuera del círculo unidad. Resolviendo $(1 - \phi_1 L) = 0$ obtenemos la raíz $L = 1/\phi_1$. Por tanto, si el proceso AR(1) es estacionario:

$$\text{Raíz de } \Phi(L) = 0 \text{ está fuera del círculo unidad} \quad \Leftrightarrow \quad \left|\frac{1}{\phi_1}\right| > 1 \quad \Leftrightarrow \quad |\phi_1| < 1.$$

5.2.2.- Modelos autorregresivos de segundo orden AR (2)

La expresión para un proceso autorregresivo de orden dos es la siguiente:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

y se caracteriza por:

- Ser siempre invertible.
- Ser estacionario si las raíces del polinomio de retardos $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 = 0$ caen fuera del círculo unidad, o bien, si los coeficientes del proceso cumplen las siguientes condiciones:

  (1) $\phi_1 + \phi_2 < 1$ (2) $\phi_2 - \phi_1 < 1$ (3) $|\phi_2| < 1$

Si el proceso AR(2) es estacionario

1. Su *media* es $\mu = \delta (1 - \phi_1 - \phi_2)^{-1}$. Por tanto, δ es el parámetro que determina si la media del proceso es cero o no.
2. Su *función de autocorrelación* sigue la siguiente estructura:

   $$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

3. Los valores iniciales se obtienen de resolver las ecuaciones de Yule-Walker:

   $$\rho_1 = \phi_1 + \phi_2 \rho_1$$
   $$\rho_2 = \varphi_1 \rho_1 + \phi_2$$

La función de autocorrelación no se anula. Por tanto, se dice que es un proceso de memoria larga. Como muestran los correlogramas de la Figura 6, la fas de un AR(2) converge a cero, si bien con distintos patrones dependiendo de las raíces de $(1 - \phi_1 L - \phi_2 L^2) = 0$:

- Si las raíces son reales, presenta un decaimiento exponencial.

- Si las raíces son complejas decae siguiendo un comportamiento sinusoidal.

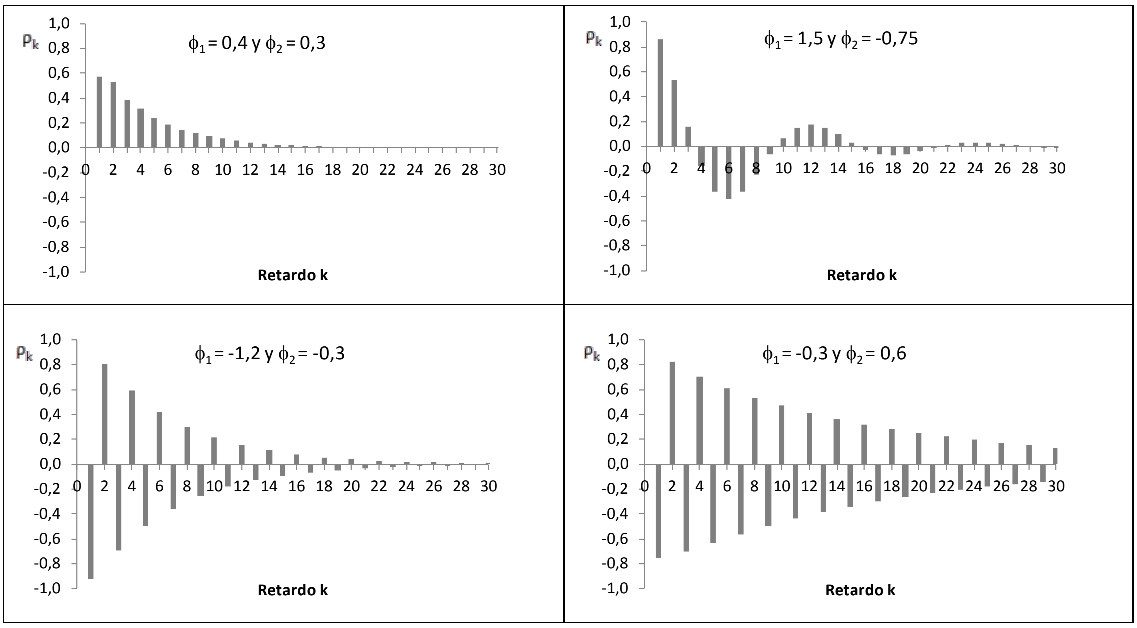

Figura 6. Correlograma simple de un AR(2)

4. Su *función de autocorrelación parcial* se anula para órdenes k>2, como se ve en la Figura 7. Los signos de ρ_1^p y ρ_2^p coinciden, respectivamente con ϕ_1 y ϕ_2.

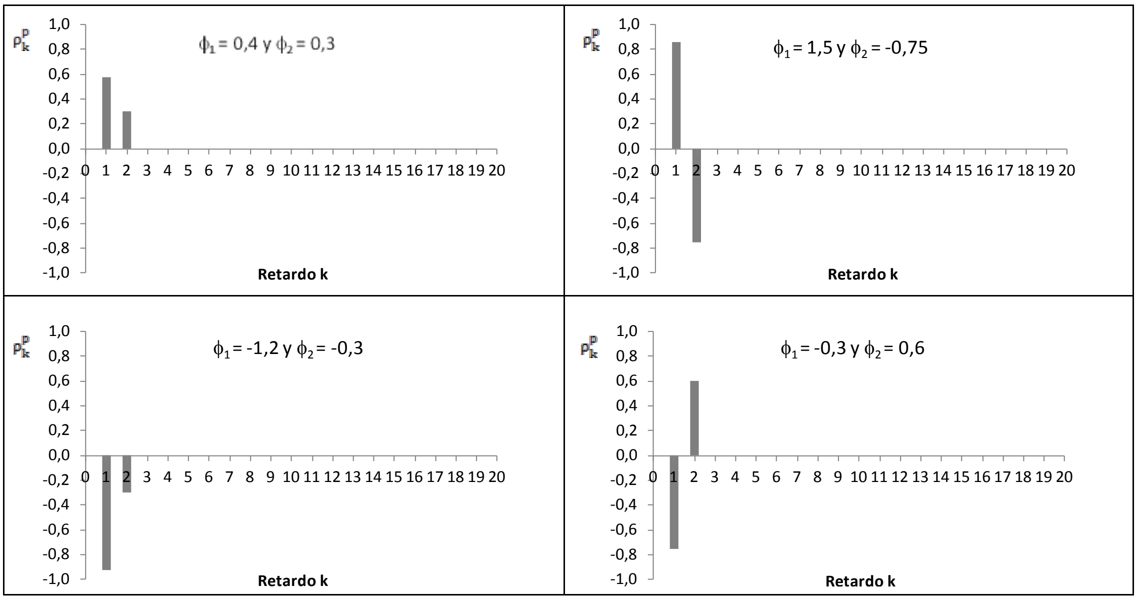

Figura 7. Correlograma parcial de un AR(2)

Pasamos a demostrar los resultados 1 y 2. Si el proceso es estacionario en media y en varianza entonces se tiene que $\mu = E(Y_t) = E(Y_{t-k})$ para todo momento t y retardo k. De aquí se llega a:

$$E(Y_t) = \delta + \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) + E(\varepsilon_t) \quad \Rightarrow \mu = \delta + \phi_1 \mu + \phi_2 \mu \quad \Rightarrow \mu = \delta(1 - \phi_1 - \phi_2)^{-1}$$

Para determinar la estructura de la función de autocorrelación, obtendremos la forma de la función de autocovarianzas. Recordando que γ_0 = var(Y_t) = cov(Y_t, Y_t), tenemos que:

$$\gamma_0 = \mathrm{var}(\delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t,\ Y_t) = \phi_1\gamma_1 + \phi_2\gamma_2 + \mathrm{cov}(\varepsilon_t, Y_t)$$

Como cov(ε_t,Y_{t-1}) = cov(ε_t,Y_{t-2}) = 0, es fácil comprobar que cov(ε_t,Y_t) = σ^2. Por tanto:

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2$$

Para el resto de autocovarianzas se verificará que:

$$\gamma_k = \mathrm{cov}(Y_t, Y_{t-k}) = \mathrm{cov}(\delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t,\ Y_{t-k}) = \phi_1 \mathrm{cov}(Y_{t-1}, Y_{t-k}) + \phi_2 \mathrm{cov}(Y_{t-2}, Y_{t-k})$$

Es decir:

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} \qquad k = \pm 1, \pm 2, \ldots$$

de donde podemos derivar la expresión para la función de autocorrelación simple:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1\frac{\gamma_{k-1}}{\gamma_0} + \phi_2\frac{\gamma_{k-2}}{\gamma_0} \qquad \Rightarrow \qquad \rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} \qquad k = \pm 1, \pm 2, \ldots$$

Por ejemplo, sustituyendo para k=1 y k=2 tenemos las ecuaciones de Yule-Walker.

5.2.3.- Modelos autorregresivos de orden p AR(p)

A la vista de los resultados obtenidos para los procesos AR(1) y AR(2), podemos generalizar las expresiones obtenidas para un proceso de orden p. Sea el proceso autorregresivo de orden p:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

Este proceso se caracteriza por:

- Ser siempre invertible.
- Ser estacionario si las raíces del polinomio en el operador de retardos

  $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p = 0$

  caen fuera del círculo unidad. Una condición necesaria (no suficiente) para que sea estacionario es $\phi_1 + \phi_2 + \ldots + \phi_p < 1$..

Si el proceso es estacionario, sus propiedades son las siguientes:

1. Su media es $\mu = \delta(1 - \phi_1 - \ldots - \phi_p)^{-1}$. Por tanto, δ. es el parámetro que determina si la media del proceso es cero o no.
2. Su *función de autocorrelación simple* sigue la siguiente estructura:

   $$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \qquad k = \pm 1, \pm 2,$$

   Los primeros *p* coeficientes de autocorrelación se pueden calcular mediante las ecuaciones de Yule-Walker (con ρ_0=1):

   $$\begin{aligned} \rho_1 &= \phi_1\rho_0 + \phi_2\rho_1 + \ldots + \phi_p\rho_{p-1} \\ \rho_2 &= \phi_1\rho_1 + \phi_2\rho_0 + \ldots + \phi_p\rho_{p-2} \\ &\ldots \\ \rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots + \phi_p\rho_0 \end{aligned}$$

Un AR(p) es un proceso de memoria larga. La forma en la que su correlograma tiende a cero depende de las raíces del polinomio $\Phi(L)=0$:

- Cuando las raíces son reales presenta un patrón amortiguado hacia cero, tomando siempre valores positivos o alternando de signo.

- Si hay raíces complejas mostrará un decaimiento sinusoidal.

3. Su *función de autocorrelación parcial* (y su correlograma parcial) se anula para retardos superiores al orden p del proceso.

5.3.- Procesos de Media Móvil (MA(q))

En los procesos de media móvil de orden q, cada observación Y_t es generada por una media ponderada de perturbaciones aleatorias, con un retardo de *q* períodos tal que:

$$Y_t = \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim RB(0, \sigma^2)$$

5.3.1.- Modelo media móvil de primer orden MA(1)

Un proceso de media móvil de orden 1 ó MA(1) se define como:

$$Y_t = \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

Se caracteriza por:

- Ser siempre estacionario.

- Ser invertible si $|\theta_1|<1$ o bien si la raíz del polinomio $\Theta(L) = 1 + \theta_1 L = 0$ es mayor que la unidad.

- Su media es $\mu = \delta$ y su varianza es $\gamma_0 = \sigma^2(1+\theta_1^2+\theta_2^2)$.

- Su función de autocovarianzas es:

  $\gamma_0 = \sigma^2(1+\theta_1^2+\theta_2^2)$

  $\gamma_1 = cov(Y_t,Y_{t-1}) = cov(\delta + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \delta + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2}) = \theta_1 var(\varepsilon_{t-1}) = \sigma^2 \theta_1$

  $\gamma_2 = cov(Y_t,Y_{t-2}) = cov(\delta + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \delta + \varepsilon_{t-2} + \theta_1 \varepsilon_{t-3}) = 0$

  De igual forma, se demuestra que $\gamma_k = 0$ si $k > 2$.

- Su *fas* sigue la estructura:

  $$k= 1: \quad \rho_1 = \sigma^2(1+ \theta_1^2)^{-1}, \qquad y \text{ si } |k| > 1 \quad \rho_k = 0$$

  Por tanto, cualquier valor de Y_t está correlacionado con Y_{t-1} e Y_{t+1}, pero no con el resto de elementos del proceso. Es decir, los procesos MA(1) poseen memoria de sólo un período. Su correlograma tiene un único pico distinto de cero, del mismo signo que θ_1 (véase Fig. 8).

- La *función de autocorrelación parcial* de un proceso MA(1) invertible no se anula, aunque las correlaciones decaen a medida que aumenta el orden *k*. Su correlograma parcial presenta un comportamiento amortiguado hacia cero, como muestra la Figura 8.

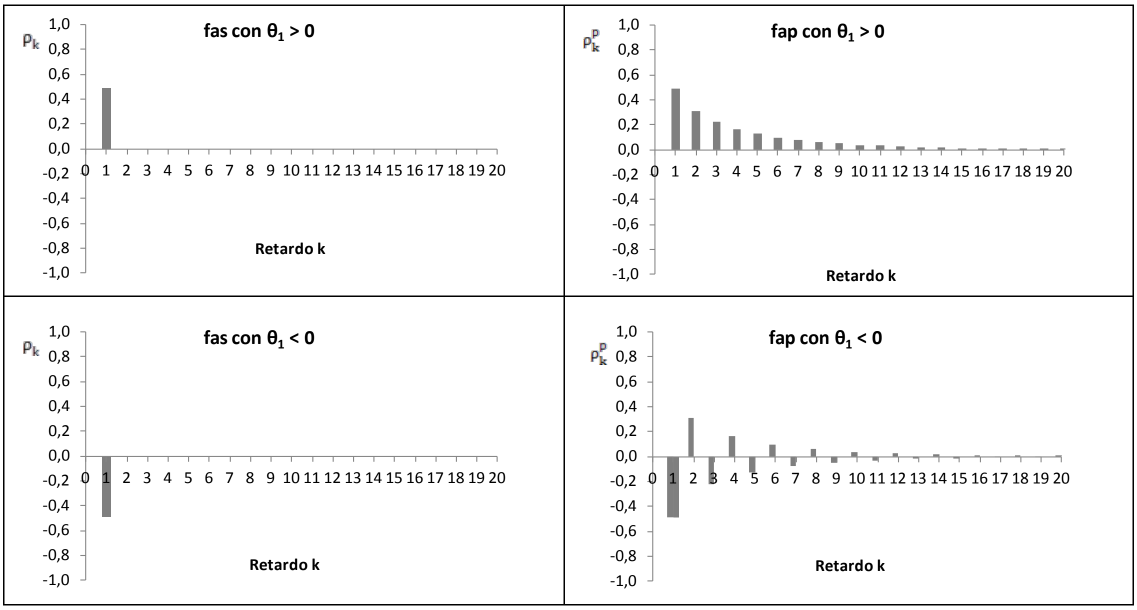

Figura 8. Correlogramas simple y parcial de un MA(1)

5.3.2.- Modelo media móvil de segundo orden MA(2)

Este proceso viene definido por la siguiente ecuación:

$$Y_t = \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

Se caracteriza por:

- Ser siempre estacionario.
- Ser invertible si el polinomio en el operador de retardos $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 = 0$ tiene raíces fuera del círculo unidad.
- Su media es $\mu = \delta$ y su varianza es $\gamma_0 = \sigma^2(1+\theta_1^2+\theta_2^2)$
- Su función de autocovarianzas para $k \neq 0$ es:

  $\gamma_1 = cov(Y_t,Y_{t-1}) = cov(\delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}, \delta + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3}) = \sigma^2 \theta_1(1 + \theta_2)$

  $\gamma_2 = cov(Y_t,Y_{t-2}) = cov(\delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2},, \delta + \varepsilon_{t-2} + \theta_1 \varepsilon_{t-3} + \theta_2 \varepsilon_{t-4}) = \sigma^2 \theta_2$

  $\gamma_3 = cov(Y_t,Y_{t-3}) = cov(\delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2},, \delta + \varepsilon_{t-3} + \theta_1 \varepsilon_{t-4} + \theta_2 \varepsilon_{t-5}) = 0$

  Igualmente, se comprueba que $\gamma_k = 0$ si $k > 3$.

- Su *función de autocorrelación* es $\rho_k = \begin{cases} \theta_1(1+\theta_2)(1+\theta_1^2+\theta_2^2)^{-1} & \text{si } k=\pm 1 \\ \theta_2(1+\theta_1^2+\theta_2^2)^{-1} & \text{si } k=\pm 2 \\ 0 & \text{si } |k|>2 \end{cases}$

  Por tanto, el proceso MA(2) posee memoria de dos períodos. Como se observa en la Figura 9, su correlograma tiene dos valores distinto de cero, con el mismo signo que θ_1 y θ_2, respectivamente.

- La *función de autocorrelación parcial* de un proceso invertible no llega a anularse, de modo que su correlograma parcial presenta un comportamiento amortiguado hacia cero. Su forma depende de las raíces de $\Theta(L)=0$:

  - Si las raíces son reales, decae exponencialmente.
  - Si las raíces son complejas, es una onda amortiguada hacia cero.

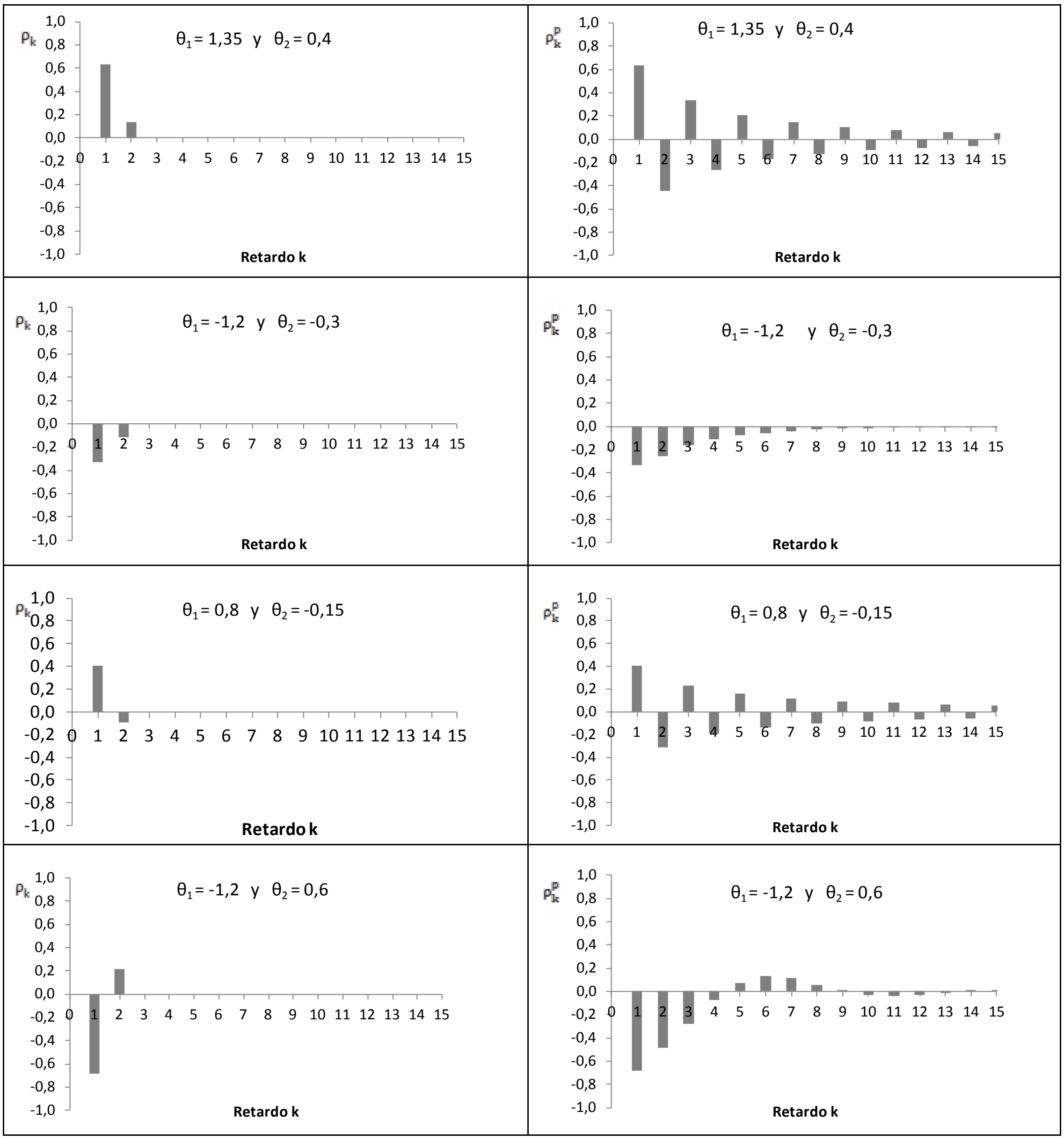

Figura 9. Correlogramas simple y parcial de un MA(2)

Una vez analizados los resultados obtenidos para los procesos de media móvil de orden 1 y 2, ya podemos obtener una generalización de las expresiones anteriores para un proceso de media móvil de orden q cualquiera.

5.3.3.- Modelo media móvil de orden q MA(q)

El proceso MA(q), $Y_t = \delta + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$, con $\varepsilon_t \sim RB(0, \sigma^2)$ se caracteriza por:

- Ser siempre estacionario.
- Ser invertible si las raíces del polinomio $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q = 0$ caen fuera del círculo unidad. Una restricción necesaria pero no suficiente para que el proceso sea invertible es $-(\theta_1 + \theta_2 + \ldots + \theta_q) < 1$.
- Su media es $\mu = \delta$ y su varianza es $\gamma_0 = \sigma^2(1+\theta_1^2+\theta_2^2+\ldots+\theta_q^2)$.
- Su *función de autocorrelación* se anula para órdenes superiores a *q* (orden del proceso). Se dice que la memoria del MA(q) es igual al orden del proceso. Por tanto, su correlograma tendrá q picos distintos de cero.

- La *función de autocorrelación parcial* de un proceso MA invertible no se anula, de modo que su correlograma parcial muestra una tendencia amortiguada, con decaimiento hacia cero. La forma del decaimiento depende de $\Theta(L) = 0$:
  - Si sus raíces son reales, decae exponencialmente.
  - Si sus raíces son complejas, es una onda amortiguada hacia cero.

5.4.- Procesos ARMA(p, q)

Los procesos ARMA(p,q) son, como su nombre indica, un modelo mixto que posee una parte autorregresiva y otra de media móvil, donde p es el orden de la parte autorregresiva y q, el de la media móvil. La expresión genérica de este tipo de procesos es:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim RB(0, \sigma^2)$$

O bien $\Phi(L)Y_t = \delta + \Theta(L)\varepsilon_t$.

Sus propiedades son las siguientes:

- Es estacionario cuando su parte AR es estacionaria, es decir, si las raíces del polinomio $\Phi(L)=0$ caen fuera del círculo unidad.

- Es invertible cuando su parte MA es invertible, es decir, cuando las raíces del polinomio $\Theta(L)=0$ caen fuera del círculo unidad.

- Además, $\Phi(L)=0$ y $\Theta(L)=0$ no tienen raíces comunes.

Si el proceso es estacionario, invertible y sin raíces comunes:

1. Su media es $\mu = \delta(1 - \phi_1 - \ldots - \phi_p)^{-1}$.
2. Su *función de autocorrelación* tiende a cero a medida que aumenta el desfase temporal *k,* sin llegar a anularse, de forma similar a los procesos AR(p)*.* La presencia del término MA(q) afecta a la determinación de los *q* primeros coeficientes de autocorrelación. A partir del retardo q, su correlograma sigue el patrón de un AR(p): decae en forma de onda seno-coseno, exponencialmente o una mezcla de ambas según las raíces de $\Phi(L)=0$ sean complejas o reales.

3. Su *función de autocorrelación parcial*, como en los procesos MA(q), no se anula, aunque tiene un comportamiento amortiguado hacia cero. A partir del retardo *p*, su correlograma parcial está dominado por un decaimiento exponencial o en forma de onda seno-coseno amortiguada dependiendo de que $\Theta(L)=0$ tenga raíces reales o complejas.

Como ejemplo, nos fijaremos en el modelo mixto más sencillo, ARMA (1,1). La ecuación que define este proceso es:

$$Y_t = \delta + \phi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

O bien $(1-\phi_1 L)Y_t = \delta + (1+\theta_1 L)\,\varepsilon_t$. Presenta las siguientes características:

- Es estacionario cuando su parte AR es estacionaria, es decir, si $|\phi_1| < 1$.
- Es invertible cuando su parte MA es invertible, es decir, si $|\theta_1|<1$.

- No tiene raíces comunes si $\phi_1 \neq -\theta_1$.

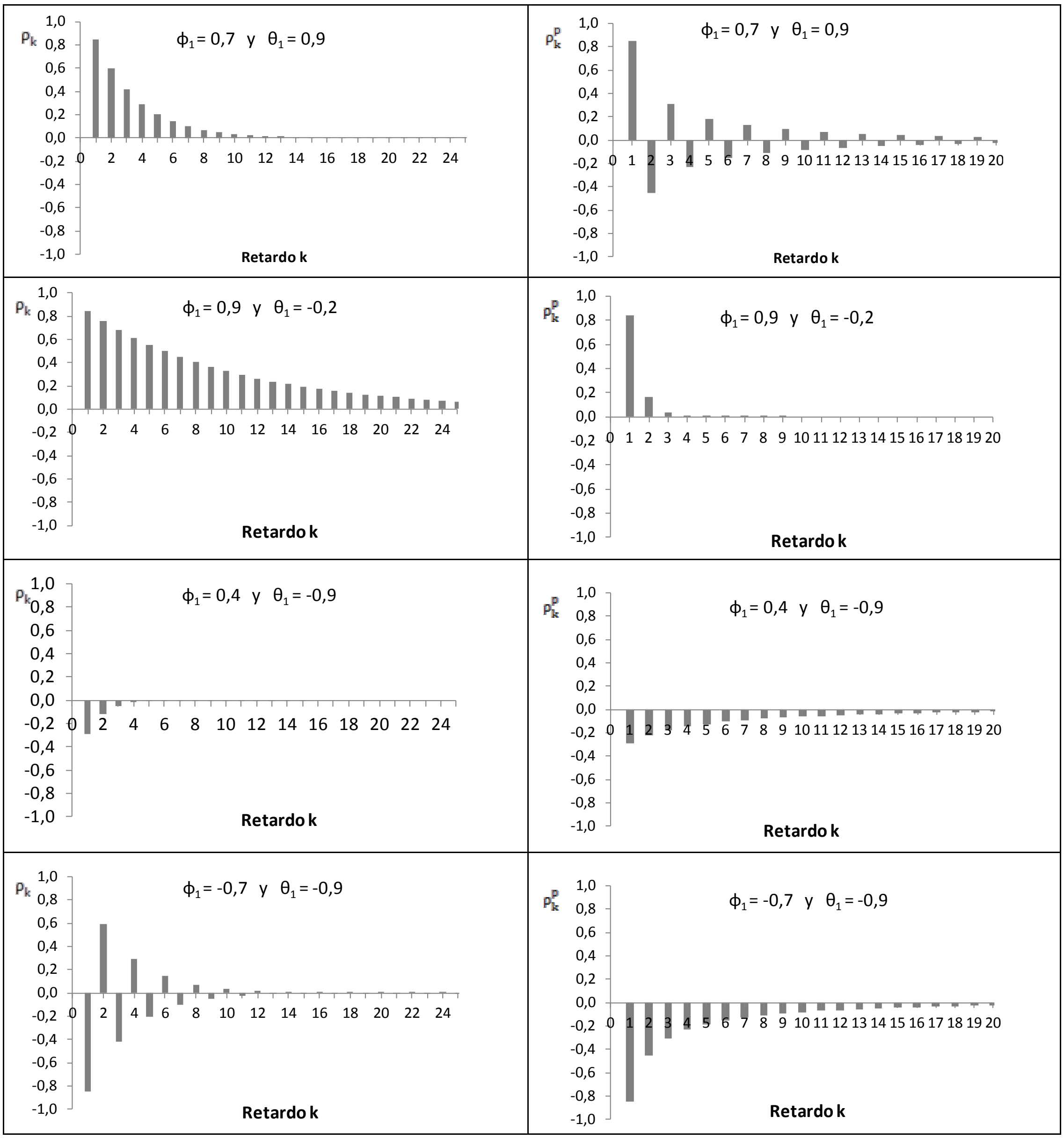

Figura 10. Correlogramas simple y parcial de un ARMA(1,1)

Bajo las condiciones anteriores,

1. Su media es $\mu = \delta\,(1-\phi_1)^{-1}$ y su varianza es igual a $\gamma_0 = \sigma^2\,\dfrac{1+\theta_1^2+2\theta_1\phi_1}{1-\phi_1^2}$.

2. Su *función de autocorrelación* es:

$$\rho_k = \frac{(1+\theta_1\phi_1)(\theta_1+\phi_1)}{1+\theta_1^2+2\theta_1\phi_1}$$

$$\rho_k = \phi_1\,\rho_{k-1} \qquad\qquad k = \pm 2,\ \pm 3,\ \pm 4,\ \ldots$$

   Por tanto, la fas tiende a cero a medida que aumenta el orden k, pero sin llegar a anularse. Es decir, este proceso tiene memoria infinita. Como muestra la Figura 10, su correlograma decae exponencialmente (como un AR(1)), una vez superado el retardo k=1 (orden del componente MA).

3. Como el proceso tiene un componente media móvil, su *función de autocorrelación parcial* no se anula, aunque tiene un comportamiento amortiguado hacia cero. Ejemplos de correlograma parcial se representan en la Figura 10.

5.5.- Procesos estocásticos no estacionarios ARIMA (p, d, q)

El análisis de series temporales se limita al estudio de procesos estacionarios para hacer tratable el problema. En la práctica, muy pocas series económicas son estacionarias. Sin embargo, transformaciones de las series dan lugar a series estacionarias. Si $\{Y_t\}$ es un proceso no estacionario en media y tras tomar consecutivamente *d* diferencias logramos que sea estacionario, decimos que $\{Y_t\}$ es un proceso integrado de orden d.

Generalizando, un proceso estocástico $\{Y_t\}$ se denomina proceso autorregresivo integrado de medias móviles de órdenes *p*, *d*, *q* y se escribe $Y_t \sim \text{ARIMA}(p,d,q)$, si tomando diferencias de orden *d* se obtiene un proceso estacionario Z_t del tipo ARMA(p,q). Es decir:

$$\Delta^d Y_t = Z_t \qquad \text{con} \qquad \Phi(L)Z_t = \delta + \Theta(L)\varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

Un proceso integrado sencillo es el paseo aleatorio o I(1) sin constante, cuyas primeras diferencias es un ruido blanco:

$$\Delta Y_t = Z_t = \varepsilon_t \qquad \text{o bien} \qquad Y_t = Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

Una ampliación del mismo es el paseo aleatorio con deriva o I(1) con constante:

$$\Delta Y_t = Z_t = \delta + \varepsilon_t \text{ o bien } Y_t = \delta + Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2)$$

De forma similar, si la varianza del proceso varía en el tiempo, es decir, $\{Y_t\}$ es un proceso no estacionario en varianza, podemos obtener un proceso estacionario aplicando una transformación de Box-Cox. Se trata de una familia de funciones $Y^{(\lambda)}$ que depende de un parámetro λ:

- para $\lambda = 0$ se calcula $Y^{(\lambda=0)} = \ln(Y)$,
- en otro caso, se define como $Y^{(\lambda)} = (Y^{\lambda} - 1)/\lambda$. El valor $\lambda = 1$ se corresponde con un patrón estacionario, de modo que $Y^{(\lambda=1)} = Y$.

5.6.- Metodología Box-Jenkins de Análisis de Series Temporales

Hasta ahora hemos visto que los procesos ARIMA(p,d,q) son una clase de modelos capaces de generar series temporales, no necesariamente estacionarias y hemos analizado las propiedades de los modelos para d, q, p y δ dados. En la aplicación de la metodología Box-Jenkins se toma el enfoque contrario: se conocen los valores de la serie temporal durante *n* periodos, $(Y_1, \ldots, Y_n)$, y se trata de determinar la estructura ARIMA(p,d,q) que la ha podido generar. Se realiza mediante un proceso iterativo en cuatro etapas (Box, Jenkins & Reinsel, 2008):

- Identificación. – En esta fase se usan los datos o cualquier tipo de información sobre cómo ha sido generada la serie, para sugerir una subclase de modelos que merezca la pena ser investigada. El objetivo es determinar el parámetro λ, los órdenes *d, q, p* y si se incluye o no δ. Es posible que exista más de un modelo que presumiblemente haya generado la serie.
- Estimación. – Una vez seleccionado el modelo, se usa de forma eficiente los datos para hacer inferencia sobre los parámetros condicionado a que el modelo investigado sea apropiado. Dado un determinado proceso, se trata de cuantificar los parámetros del modelo, $\theta_1, \ldots, \theta_q, \phi_1, \ldots, \phi_p, \delta$ y σ^2.
- Validación. – Consiste en la aplicación de una batería de contrastes diagnósticos para comprobar si el modelo se ajusta a los datos; revelar las posibles discrepancias del modelo propuesto y mejorarlo.

- Predicción. – Por último, el modelo seleccionado se utiliza para obtener pronósticos en términos probabilísticos de los valores futuros de la variable. Se distingue predicción puntual o por intervalo. Un modelo es útil si es capaz de predecir, por lo que se trata de evaluar la capacidad predictiva del modelo.

Esta metodología se basa en dos principios:

- Selección de un modelo en forma iterativa: en cada etapa se plantea la posibilidad de rehacer las etapas previas.
- Principio de parametrización escueta: se propone un modelo generador con el mínimo de parámetros posibles y únicamente se procede a ampliarlo en caso de que sea estrictamente necesario para describir el comportamiento de la serie.

5.6.1.- Identificación

La primera etapa, y fundamental, es la identificación del modelo adecuado para los datos. Exige conocer las características y propiedades de los modelos ARIMA. La identificación consiste en:

1. Análisis de estacionariedad, que determina las transformaciones que son necesarias para obtener una serie estacionaria. Incluye dos apartados:
    1.1 Transformación estabilizadora de varianza.
    1.2 Número de diferencias *d* para estacionariedad en media.
2. Elección de los órdenes *p* y *q*, con el fin de determinar el proceso estacionario ARMA (p,q) adecuado.

Los instrumentos que se utilizan en esta etapa son un conjunto de gráficos de la serie original y transformaciones de la misma, junto con sus funciones de autocorrelación simple y parcial muestrales y los contrastes de raíces unitarias.

5.6.1.1.- Análisis de estacionariedad

5.6.1.1.1.- Estacionariedad en varianza

Una serie será *estacionaria en varianza* cuando pueda mantenerse el supuesto de que existe una única varianza para toda la serie temporal. En caso de que no se cumpla esta condición, se trata de obtener una transformación de la variable que sea estacionaria en varianza, es decir, de elegir el parámetro λ adecuado en la trasformación de Box-Cox. Para determinar si la dispersión es estable en el tiempo se utiliza el gráfico de la serie a lo largo del tiempo y el gráfico rango-media de la serie.

- Si se cumple la condición de estacionariedad en varianza, la dispersión de la serie alrededor de su media se mantiene aproximadamente estable a lo largo del tiempo.
- El gráfico rango-media es un gráfico representativo de la relación dispersión-nivel. Se construye dividiendo la muestra en varias submuestras del mismo tamaño y calculando la media y el rango de cada una de las submuestras; posteriormente se representan los resultados en un gráfico de dispersión, con la media en el eje horizontal y el rango en el vertical. Si en el gráfico no se aprecia una relación entre la media y el rango, podemos concluir que se cumple la condición de estacionariedad.

Las series económicas con frecuencia son no estacionarias en varianza, conforme crece la media, el rango (y la varianza) aumenta. En este caso se aplican distintos valores de la transformación de Box-Cox hasta obtener una serie con varianza estable $Y^{(\lambda)}$. En la práctica, muchas series estabilizan su varianza al tomar logaritmos ($\lambda=0$).

5.6.1.1.2.- Estacionariedad en media

En este apartado se trata de determinar el orden de diferenciación d. Si la serie es estacionaria en media, elegimos d=0; en caso contrario, tomaremos d sucesivas diferencias sobre la serie hasta obtener una serie estacionaria $Z_t = \Delta^d Y_t^{(\lambda)}$.

En general, las series económicas son no estacionarias, aunque sí alcanzan la estacionariedad para d<3. Los instrumentos que se utilizan son el gráfico y los correlogramas de la serie y sus transformaciones en diferencias, así como los contrastes de raíces unitarias y estacionariedad.

Las características de un **proceso estacionario** son las siguientes: en lo que respecta al gráfico de la serie, ésta será estacionaria en media cuando pueda mantenerse el supuesto de que existe una única media para toda la serie temporal, es decir, cuando a lo largo del tiempo fluctúa alrededor de un nivel constante. En cuanto al correlograma, tenemos que la función de autocorrelación simple teórica ρ_k de un proceso estacionario en media decae rápidamente. Por tanto, si en el correlograma muestral los coeficientes r_k se aproximan rápidamente a cero podemos considerar que el proceso es estacionario.

Por otro lado, las principales características de los **procesos no estacionarios en media** son: en el gráfico de la serie se observan tendencias o varios tramos con medias diferentes; además, un proceso con alguna raíz unitaria suele presentar un correlograma simple muestral con un decaimiento muy lento, no siendo necesario que se mantenga próximo a la unidad.

Otro tipo de instrumentos son los contraste para determinar si el proceso es estacionario (d=0) o integrado d> 0. Encontramos dos tipos de contrastes:

- Contrastes de raíces unitarias, donde la hipótesis nula es que el proceso tiene raíz unitaria (no es estacionario). El más conocido es el contraste de Dickey-Fuller aumentado o *ADF* (Dickey & Fuller, 1979 y 1981), que en la actualidad se incluye en cualquier paquete econométrico que ofrezca unas herramientas básicas de análisis de series temporales.
- Contrastes de estacionariedad, que plantean la hipótesis nula de estacionariedad del proceso. El más conocido de esta clase es el contraste de Kwiatkowski, Phillips, Schmidt y Shin(1992) o *KPSS*.

5.6.1.2.- Selección de los órdenes del proceso estacionario

Una vez obtenida una serie Z_t estacionaria (t=d+1, ..., n), se determinará el orden de la parte autorregresiva (p) y el de la parte de media móvil (q) del proceso ARMA que se considere que haya podido generar la serie estacionaria. Para tal fin se utilizan los correlogramas muestrales simple y parcial.

En la interpretación de los correlogramas muestrales para tratar de determinar de qué tipo de proceso univariante procede la serie transformada pueden resultarnos muy útiles las siguientes reglas:

- En los modelos AR(p), el correlograma parcial presenta los *p* primeros coeficientes (picos) distintos de cero y el resto nulos. Asimismo, generalmente el correlograma simple presenta un decrecimiento rápido de tipo exponencial, sinusoidal o ambos.
- En los modelos MA(q), sucede el patrón opuesto: el correlograma simple se anula para retardos superiores a *q* y el parcial decrece exponencial o sinusoidalmente.
- En los ARMA los dos correlogramas (simple y parcial) decaen sin llegar a anularse. A diferencia de los procesos MA(q) ó AR(p), en los procesos mixtos ARMA no es fácil identificar los órdenes adecuados con los correlogramas. En la práctica, se puede especificar un ARMA de pocos parámetros (órdenes p y q pequeños) y pasar a la validación para comprobar si es necesario ampliar el modelo.

En cualquier caso, para que una serie sea fácilmente identificable hay que considerar un tamaño muestral elevado. En general, la etapa de identificación suele plantear ciertas dificultades y su objetivo es determinar la especificación tentativa de unos pocos modelos con estructuras sencillas. Las posteriores etapas de estimación y validación de los resultados confirmarán los indicios o, por el contrario, servirán de fundamento para la reformulación de los modelos propuestos.

5.6.2.- Estimación

Una vez identificado el modelo de series temporales apropiado para la serie estacionaria Z_t, se procede a estimar sus parámetros. Evidentemente, si hemos tomado diferencias de orden d se perderán *d* observaciones para la nueva serie, quedando (n–d) datos. Las hipótesis que se adoptan en el proceso de estimación son las siguientes:

a) La innovación ε_t posee estructura de ruido blanco gaussiano, $\varepsilon_t \sim NID(0, \sigma^2)$.

b) La parte AR del proceso es estacionario.

c) La parte MA del proceso es invertible.

d) No hay raíces comunes en la parte MA y la parte AR del proceso.

Bajo estas condiciones existen distintos procedimientos de estimación que son consistentes y asintóticamente normales. Veamos a continuación y de forma somera cómo se realiza la estimación de los distintos modelos univariantes que hemos visto en los epígrafes anteriores.

- **Estimación de procesos autorregresivos.** La estimación en procesos AR(p) es muy sencilla, ya que el modelo a estimar es:

  $Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim RB(0, \sigma^2)$

  con $t=d+p+1, d+p+2, \ldots, n$.

  Bajo las hipótesis establecidas, se demuestra que el estimador MCO es consistente, asintóticamente normal y los estadísticos *t* habituales pueden aplicarse para estudiar la significatividad de los coeficientes.

- **Estimación de procesos de medias móviles y ARMA.** El procedimiento de obtención de las estimaciones es más complicado cuando el modelo tiene un componente MA. Por ejemplo, en el modelo más sencillo, el MA(1)

  $$Z_t = \delta + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

  La variable que acompaña a θ_1 es el error no observable ε_{t-1}, que también hay que estimar. Como consecuencia, la función que se maximiza ó minimiza al estimar δ y θ_1 es no lineal en parámetros, por lo que se aplican algoritmos de optimización numérica. Por ello generalmente la estimación de estos procesos resulta complicada, debiéndose utilizar algún programa informático especializado[1].

5.6.3.- Validación

En esta etapa se comprobará la capacidad de ajuste del modelo estimado a los datos por medio de un conjunto de contrastes sobre los residuos y los parámetros del modelo. Si éste no supera satisfactoriamente esta fase, será necesario reformularlo. En este sentido, cabe señalar que los resultados de la comprobación de la validez del modelo suelen dar pistas para proceder a la re-especificación de uno más adecuado.

[1] Ejemplos de software estadístico-econométricos son Eviews, R, SPSS, Demetra, Gauss, RATS, SAS o TSP.

5.6.3.1.- Análisis de los residuos

Como sabemos, una de las hipótesis de los modelos univariantes es que el término de error ε_t es ruido blanco. Por ello, los residuos del modelo estimado:

$$e_t = Z_t - \hat{Z}_t = Z_t - (\hat{\delta} + \hat{\phi}_1 Z_{t-1} + \ldots + \hat{\phi}_p Z_{t-p} + \hat{\theta}_1 e_{t-1} + \ldots + \hat{\theta}_1 e_{t-q})$$

deben ser un proceso puramente aleatorio. En otro caso, contendrían información relevante para la predicción. Con objeto de estudiar si los residuos se aproximan al comportamiento de un proceso ruido blanco, se disponen de las siguientes herramientas:

- **Representación gráfica de los residuos**. La representación de los residuos en el tiempo permite observar si la varianza es constante y si la media está próxima a cero. También es útil para detectar la existencia de residuos atípicos, esto es, residuos que en valor absoluto exceden en tres o cuatro veces su desviación típica $S_e = \hat{\sigma}_e$. Además, se puede comprobar si se desvían del patrón esperado en una distribución normal: un número alto de residuos, por encima del 5%, que caigan fuera del intervalo $\pm 1{,}96 S_e$ puede indicar ausencia de normalidad.

- **Correlogramas de los residuos y contrastes de significatividad individual de los coeficientes de autocorrelación simple y parcial.** Si el proceso propuesto es correcto, las innovaciones ε_t deben estar incorrelacionadas y, por lo tanto, también su equivalente muestral, los residuos e_t. La serie de residuos es aleatoria si los coeficientes de las funciones de autocorrelación simple y parcial son iguales a cero. Anderson (1942) demostró que los coeficientes de autocorrelación muestral r_k de un ruido blanco siguen asintóticamente una distribución normal de media 0 y varianza 1/n, siendo n el tamaño de la serie temporal.

  En consecuencia, se considera que el coeficiente de autocorrelación de orden k de los errores es significativamente distinto de cero si se sale de las bandas de confianza del 95% ($\pm 2n^{-1/2}$). La mayor parte de los paquetes econométricos dibujan estas bandas de confianza en los correlogramas simple y parcial. Si más del 5% de estos coeficientes de autocorrelación simple $r_k(e)$ o parcial $r_k^p(e)$ caen fuera de este intervalo, entonces se rechazará la hipótesis nula de no autocorrelación.

- **Contrastes de significatividad conjunta sobre los coeficientes de autocorrelación simple del error.** En un segundo tipo de contraste se plantea la hipótesis nula de que los primeros *m* coeficientes de autocorrelación simple de los residuos son iguales a cero. El estadístico de contraste más utilizado es el propuesto por Ljung y Box(1978):

$$Q^*(m) \quad \text{ó} \quad LB(m) = n(n+2)\sum_{k=1}^{m}\frac{r_k(e)^2}{n-k}$$

  donde n es el tamaño de la serie de residuos y $r_k(e)$ es su coeficiente de autocorrelación muestral de orden k. Bajo la hipótesis nula de que los residuos del modelo son ruido blanco, este estadístico sigue asintóticamente una distribución chi-cuadrado con m-p-q grados de libertad si el cociente m/n es pequeño y m moderadamente grande. Por tanto, se rechazará la hipótesis nula si el valor de observado de LB es superior que el tabulado de la distribución a un nivel de significación dado. Cuanto mayor sea m el test se extenderá a desfases mayores, pero disminuye la potencia del contraste, es decir, la probabilidad de rechazar la hipótesis nula cuando es falsa.

En caso de que los residuos de la estimación no se comporten como un proceso ruido blanco, su *fas* y *fap* son instrumentos valiosos a la hora de reformular el modelo. Si se detecta algún tipo de estructura univariante que no se encontraba originalmente en el modelo, podremos incorporarla para estimar un nuevo modelo más amplio. Así, por ejemplo, si estimamos un AR(1) y los residuos resultantes de dicha estimación presentan una estructura MA(2), deberemos incorporar a la especificación original dicho componente de media móvil, debiendo estimar ahora un modelo ARMA(1,2). Además de los anteriores contrastes, también es posible aplicar a los residuos contrastes de heterocedasticidad, normalidad, estabilidad estructural o valores atípicos.

5.6.3.2.- Análisis de los coeficientes estimados

- **Condiciones de estacionariedad e invertibilidad.** Si alguna de las raíces de los polinomios de retardos del componente autorregresivo o del componente media móvil estimados no cayeran fuera del círculo unidad, se rechazaría el modelo.

  Por ejemplo, si el valor absoluto de alguna de las raíces del polinomio AR es menor de la unidad, el modelo obtenido no es estacionario. En el límite, si alguna de las raíces del polinomio de retardos del componente autorregresivo estuviera cercana a uno es posible que la serie original esté subdiferenciada, por lo que puede que precise alguna diferencia adicional.

  Del mismo modo, si las raíces del polinomio de retardos del componente media móvil estuvieran cercanas a uno, posiblemente el modelo esté sobrediferenciado.

- **Significatividad de los coeficientes.** Se contrasta si los coeficientes son individualmente significativos mediante el estadístico t. Dicho estadístico está construido bajo la hipótesis nula de que el coeficiente es cero y sigue una distribución asintótica normal N(0,1). Si concluimos que alguno no es significativo (estadístico t menor, en valor absoluto, que su valor en tablas) se puede eliminar del modelo.

5.6.4.- Criterios de selección de modelos

Es posible que tengamos varios modelos que superen la fase de validación y, por tanto, puede servir para predecir. En estos casos podemos utilizar criterios adicionales de selección de modelos para optar por un modelo. Otra opción es comparar la capacidad predictiva de los modelos, como veremos en la siguiente sección.

Los criterios de selección de modelos más utilizados tienen en cuenta dos factores: los errores de ajuste y la complejidad o número de coeficientes del modelo. Así, el criterio de información de Akaike, o AIC, formulado por Akaike (1974) es:

$$AIC = n^* \ln(S_e^2) + 2\, k^*$$

donde n* es el tamaño muestral usado para la estimación del modelo, S_e^2. es la varianza de los residuos y k* el número de parámetros estimados.

El criterio de información bayesiano propuesto por Schwarz(1978) o BIC es:

$$BIC = n^* \ln(S_e^2) + k^* \ln(n^*)$$

Un modelo con mejor ajuste tiene una menor varianza de los errores y, por lo tanto, un menor AIC y BIC. Por otro lado, en los dos criterios se penaliza la inclusión de coeficientes adicionales en el modelo. En general, se elige aquél modelo que tiene un menor AIC o BIC. El modelo elegido puede variar dependiendo del criterio utilizado, ya que la penalización por aumentar el número de parámetros del modelo es mayor en el BIC.

5.6.5.- Predicción

Una vez que el modelo ha superado la fase de diagnosis, se convierte en un instrumento útil para la predicción. Dado el modelo y la serie temporal **Y** = (Y_1, ..., Y_T), se pueden obtener predicciones por punto y por intervalo para periodos futuros Y_{T+1}, ..., Y_{T+L}.

Consideramos que una predicción por punto ℓ periodos hacia delante $\hat{Y}_{T+\ell|T}$ es óptima si minimiza el error cuádratico medio $ECM(\hat{Y}_{T+\ell|T}) = E(\hat{Y}_{T+\ell|T} - Y_{T+\ell}|\mathbf{Y})^2$. La función de predicción por punto óptima en modelos ARIMA se obtiene calculando la media condicionada $E(Y_{T+\ell}|\mathbf{Y})$. Por ejemplo, en un modelo AR(1) la predicción óptima es la siguiente:

- Para $\ell = 1$: $\hat{Y}_{T+1|T} = \delta + \phi_1 Y_T$
- Para $\ell > 1$: $\hat{Y}_{T+\ell|T} = \delta + \phi_1 \hat{Y}_{T+\ell-1|T} = \delta\left(1 + \phi_1 + \ldots + (\phi_1)^{\ell-1}\right) + \ldots + (\phi_1)^{\ell} Y_T$

También se puede obtener la predicción en los procesos no estacionarios. Por ejemplo, en un paseo aleatorio $Y_t = Y_{t-1} + \varepsilon_t$, la predicción óptima en T es el último valor observado, $\hat{Y}_{T+\ell|T} = Y_T$, mientras que en el paseo aleatorio con deriva es $\hat{Y}_{T+\ell|T} = \ell\,\delta + Y_T$. A partir de las predicciones por punto y sus ECM se calculan intervalos de confianza para la predicción, que proporcionan muchos programas econométricos.

Para evaluar la capacidad predictiva de un modelo se suele dividir la muestra de tamaño n en dos partes: los primeros T=n-L datos se usan para estimar el modelo, mientras que las últimas L observaciones se reservan para calcular diversos errores de predicción. Por ejemplo, se puede aplicar un contraste de estabilidad similar al contraste predictivo de Chow o comparar la calidad de la predicción de varios modelos mediante criterios como, por ejemplo, la raíz del error cuadrático medio de la predicción:

$$RECM(L) = \frac{1}{\sqrt{L}}\sqrt{\sum_{\ell=1}^{L}\left(Y_{T+\ell} - \hat{Y}_{T+\ell|T}\right)^2}$$

6. METODOLOGÍA DE LA CONTABILIDAD TRIMESTRAL

6.1.- Introducción

La Contabilidad Trimestral es una estadística de síntesis de carácter coyuntural, cuyo bjetivo primordial es proporcionar una descripción cuantitativa coherente del conjunto de la actividad económica, mediante un cuadro macroeconómico trimestral elaborado bien desde la óptica de la oferta, la demanda y/o las rentas. Estas estimaciones se ajustan a los mismos principios de coherencia y equilibrio contable que la Contabilidad Nacional de frecuencia anual y, por consiguiente, al marco del Sistema Europeo de Cuentas de 1995 (SEC-95).

El capítulo 12 del SEC trata los aspectos básicos de la Contabilidad Trimestral, existiendo un manual específico dedicado a los métodos de las Cuentas Trimestrales publicado por EUROSTAT (*Handbook on Quarterly National Accounts, 1999 edition*)

http://ec.europa.eu/eurostat/documents/3859598/5854053/CA-22-99-781-EN.PDF

Según el SEC, las cuentas económicas trimestrales forman parte integrante del sistema de cuentas nacionales y, entre sus diversos usos, cabe citar la gran importancia que tienen para el análisis del año corriente y el cálculo de las estimaciones provisionales del año precedente. Las cuentas económicas trimestrales forman un conjunto coherente de operaciones, cuentas y saldos contables, definido en el ámbito financiero y no financiero y registrado trimestralmente. Adoptan los mismos principios, definiciones y estructura que las cuentas anuales, con algunas modificaciones debidas al período de tiempo que abarcan.

Los métodos estadísticos utilizados para la elaboración de las cuentas trimestrales difieren de los empleados en las cuentas anuales. Los citados métodos pueden clasificarse en dos grandes categorías: Los procedimientos directos y los procedimientos indirectos.

- Los procedimientos directos se basan en la disponibilidad, a intervalos trimestrales y con las simplificaciones apropiadas, de fuentes similares a las utilizadas para elaborar las cuentas anuales.
- Los procedimientos indirectos se basan en la desagregación temporal de los datos de las cuentas anuales, de acuerdo con métodos matemáticos o estadísticos y utilizando indicadores de aproximación que permiten la extrapolación para el año corriente.

Los procedimientos indirectos son los métodos de extrapolación o de trimestralización, a la hora de elegir entre las diferentes metodología el SEC, señala que es preciso, ante todo, procurar que éstos minimicen el error de las previsiones para el año corriente, con la finalidad de que las estimaciones anuales a que dan lugar se acerquen lo más posible a las cifras estimadas posteriormente por las cuentas nacionales. Dicha elección dependerá, entre otras cosas, de la información trimestral disponible.

Por otro lado hay que tener presente que aunque el carácter estacional forma parte integrante de los datos trimestrales, supone a menudo un obstáculo para la identificación y el análisis correctos del componente de ciclo-tendencia. Por este motivo, se plantea la necesidad de elaborar cuentas brutas y cuentas ajustadas estacionalmente, debiéndose garantizar la coherencia contable de las cifras de estas últimas cuentas.

Dado que para las cuentas trimestrales se utiliza el mismo marco que para las cuentas anuales, es preciso que exista una coherencia en el tiempo entre ambos tipos de cuentas. Esto supone, en el caso de las variables flujo, que para cada año la suma de los datos trimestrales ha de ser igual a las cifras estimadas por las cuentas anuales. En principio, tal condición puede cumplirse sin problema alguno para las cuentas de años ya pasados. No obstante, para el año corriente se plantea un problema de prioridad temporal entre los datos trimestrales y los anuales, ya que las cifras trimestrales suelen estar disponibles antes que las anuales. El SEC señala que este problema puede solventarse si se llega al acuerdo de obtener una primera estimación provisional de las cifras anuales por medio de la agregación de los datos trimestrales. Cuando se disponga de nuevas estimaciones anuales que supongan una revisión de las cifras provisionales, los datos trimestrales tendrán que modificarse con arreglo a ellas.

Desde un punto de vista teórico no existe obstáculo alguno para que en las cuentas trimestrales se use el mismo esquema que el utilizado en las cuentas anuales. No obstante, en la práctica, resultará útil simplificarlo y agregarlo para obtener cifras trimestrales fiables lo más rápidamente posible. El SEC establece un programa de tablas y cuadros del que recogemos el apartado correspondiente a la oferta:

Tabla 1: **Principales agregados** (1) **(ejercicios trimestral y anual)**

Código	Lista de variables	Desglose (2)	Precios corrientes	Precios del año anterior y volúmenes encadenados
Valor añadido y producto interior bruto				
B.1g	1. Valor añadido bruto a precios básicos	A6†	x	x
D.21	2. a) Impuestos sobre los productos (3)		x	x
D.31	b) Subvenciones a los productos (3)		x	x
B.1*g	3. Producto interior bruto a precios de mercado		x	x

Señalar por último que uno de los aspectos a resaltar de la Contabilidad Trimestral es el de las frecuentes y, en ocasiones, notables revisiones que, necesariamente se producen. Cabe citar las siguientes:

- modificaciones en los datos de la contabilidad Anual,
- revisión de los indicadores,
- sustitución de predicciones en los indicadores por datos reales disponibles,
- modificaciones inducidas por los filtros de ajuste estacional
- variaciones en el proceso de equilibrio general entre recursos y empleos (cuando se estiman cuadros de oferta y demanda)

6.2.- Métodos de desagregación temporal

La desagregación temporal expresa la idea de que el procedimiento realizado corresponde a una desagregación de los datos anuales de baja frecuencia en datos trimestrales de alta frecuencia.

Los métodos matemáticos y estadísticos de elaboración de contabilidades trimestrales se suelen diferencian en:

- métodos que no utilizan indicadores relacionados;
- métodos que utilizan series relacionadas;
- métodos de extrapolación.

Los métodos que no precisan de indicadores relacionados con la serie anual obtienen las estimaciones trimestrales mediante una división ponderada conforme a un criterio puramente matemático, y entregan una trayectoria trimestral suficientemente ajustada y coherente con las restricciones de desagregación temporal o utilizando modelos de series cronológicas. Son métodos que pueden emplearse cuando los únicos datos disponibles son los relativos a las series anuales.

Los métodos que utilizan series relacionadas estiman la trayectoria trimestral en función de la información trimestral que proporcionan la serie o variables trimestrales que deben de estar relacionadas lógica y/o económicamente con la magnitud a anualizar.
Los métodos de extrapolación utilizan la información proveniente de las series de indicadores para obtener una estimación de los agregados deseados. La idea básica es que las series de indicadores y el agregado tienen el mismo perfil temporal y, en consecuencia, tienen la misma tasa de crecimiento, de manera que el perfil del agregado para las cifras desconocidas se construye según el perfil conocido de las series de indicadores.

El problema de la desagregación temporal se expresarse así:

Sea $Y = \{Y_T : T = 1..N\}$ la serie anual observada y $x = \{x_{i,t,T} : i = 1..p, t = 1..4, T = 1,.., N\}$ una matriz *n*x*p* cuyas filas recogen las n observaciones disponibles sobre *p* indicadores de frecuencia trimestral, siendo $p \geq 1$ y $n=4N$.

El problema de la desagregación temporal consiste en estimar una serie $y = \{y_{t,T} : t = 1..4, T = 1,.., N\}$ que satisfaga la restricción temporal asociada a que la suma de los cuatro trimestres pertenecientes a un mismo año coincida con el total anual correspondiente:

$$\sum_{t=1}^{4} y_{t,T} = Y_T \qquad \forall T \tag{3}$$

Esta restricción longitudinal se puede expresar en forma matricial como:

$$By = Y \tag{4}$$

Donde *B* es una matriz Nxn de agregación temporal definida como:

$$B = I_N \otimes f = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \dots & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 & 1 \end{bmatrix} \tag{5}$$

$\otimes$ denota el producto tensorial de Kronecker y *f* es el vector de agregación, *f*= [1,1,1,1]. Esta expresión permite considerar otros casos: si *f*=[¼,¼,¼,¼] se trata de la distribución temporal de un índice y si *f*=[0,0,0,1], se obtiene un problema de interpolación.

Los métodos de desagregación temporal se idearon originalmente para ofrecer un desglose de las cifras de baja frecuencia en cifras de alta frecuencia (por ejemplo, cifras anuales en cifras trimestrales). Reconstruyen la trayectoria de alta frecuencia de las series dando la posibilidad de la extrapolación. Hay diferentes métodos de desagregación temporal que exigen diferentes cantidades de información básica.

De hecho, la trimestralización de la serie anual, Y, puede realizarse sin disponer de indicadores de aproximación trimestral, a partir de los “métodos de desagregación temporal sin indicadores”, que sólo tienen en cuenta la información contenida en la serie anual Y. Dentro de éstos se encuentran los de Lisman y Sandee (1964) y Boot, Feibes y Lisman (1967).

El procedimiento de Boot, Feibes y Lisman es el más utilizado entre los métodos que no utilizan indicadores. Este procedimiento minimiza la suma de los cuadrados de las primeras ó segundas diferencias entre trimestres consecutivos, es decir:

$$\min \sum\nolimits_{t,T} \left(\Delta\, y_{t,T}\right)^2 \qquad \text{o} \qquad \sum\nolimits_{t,T} \left(\Delta^2\, y_{t,T}\right)^2$$

Con las N restricciones dadas por la condición de agregación (3).

En notación matricial la función a minimizar sería

$$F(y,\Lambda) = y' A\, y + 2\Lambda' \left(Y - By\right)$$

donde Λ es el vector de N multiplicadores de Lagrange y la matriz A cambia según el procedimiento elegido:

- $A = A_{fd} = D' D'$ si se emplean primeras diferencias, y
- $A = A_{sd} = D' D' DD$ si se emplean segundas diferencias, siendo D la matriz nxn:

$$D = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \tag{6}$$

Boot, Feibes y Lisman obtienen las soluciones siguientes:

$$y_{fd} = (D'D)^{-1}B'\left[B(D'D)^{-1}B'\right]^{-1}Y$$

si se emplean las primeras diferencias, y

$$y_{sd} = (D'D'DD)^{-1}B'\left[B(D'D'DD)^{-1}B'\right]^{-1}Y$$

si se emplean las segundas diferencias.

La relación entre los valores trimestral y y anual Y puede estar condicionada por la información contenida en los indicadores trimestrales x. En este caso se tienen los llamados "métodos de desagregación temporal basados en indicadores". Dentro de éstos existen dos enfoques principales: métodos de ajuste y métodos basados en modelos. Los primeros consideran la estimación de y como la solución de un programa de optimización restringida mientras que los segundos plantean dicha estimación como un problema inferencial: dada la estructura del modelo, derivar estimadores lineales, insesgados y de varianza mínima (ELIO), que permitan obtener y en función de Y y de x, verificando al mismo tiempo la restricción longitudinal (3) ó (4).

Es un método de ajuste el propuesto por Denton (1971), y basados en modelos los propuestos por Chow y Lin (1971), Fernández (1981) y di Fonzo (1990). En muchos casos la estimación de los valores trimestrales se realiza en dos etapas. En primer lugar se utilizan los indicadores para obtener una primera estimación de las series trimestrales, y se aplica después algún criterio de optimización que corrige esta estimación inicial hasta conseguir que la agregación de los valores trimestres estimados coincida con el valor anual observado.

El procedimiento de Denton (1971) se deriva del procedimiento Boot, Feibes y Lisman antes comentado. Su principio es que la variable trimestral y debe preservar en la medida de lo posible el movimiento de un único indicador x, manteniendo la condición de agregación (3). Matemáticamente, se elige una función que penaliza las discrepancias en los valores trimestrales x e y que tiene la forma $(y-x)'A(y-x)$, siendo A una matriz nxn simétrica no singular. Denton utiliza dos funciones de penalización basadas en las primeras diferencias de la variable de interés y y el indicador x. Siendo D la matriz (6) de diferencias, considera:

- $A = A_{fd} = D'D'$, que da lugar a $(y-x)'A_{fd}\,(y-x) = \sum_{t,T}\left(\Delta\, y_{t,T} - \Delta\, x_{t,T}\right)^2$
- $A = A_{sd} = D'D'DD$, de modo que $(y-x)'A_{sd}\,(y-x) = \sum_{t,T}\left(\Delta^2\, y_{t,T} - \Delta^2\, x_{t,T}\right)^2$.

En forma matricial, la función a minimizar queda:

$$F(y,\Lambda) = (y-x)'A(y-x) - 2\Lambda'(Y - By)$$

en donde, nuevamente, Λ es el vector de N multiplicadores de Lagrange. Las soluciones en los dos casos mencionados son las siguientes series trimestrales:

$$y_{fd} = x + (D'D)^{-1}B'\left[B(D'D)^{-1}B'\right]^{-1}(Y - Bx);$$
$$y_{sd} = x + (D'D'DD)^{-1}B'\left[B(D'D'DD)^{-1}B'\right]^{-1}(Y - Bx)$$

El método propuesto por Chow y Lin resuelve el problema de la estimación trimestral de manera elegante bajo un enfoque estadístico. Concretamente, este método obtiene un estimador lineal, insesgado y de varianza mínima (estimador ELIO) de los valores trimestrales a partir de un modelo de regresión lineal múltiple entre la magnitud a trimestralizar y un conjunto de indicadores representativos de su evolución. Es por ello que, de los diferentes métodos de trimestralización con indicadores propuestos, sea probablemente el más utilizado.

El procedimiento de Chow y Lin, presupone que existe un modelo trimestral lineal que se deduce a partir del modelo anual estimado.

Así, si partimos de la existencia de un modelo lineal que relaciona una variable trimestral inobservada (y), con un vector x de $p \geq 1$ variables que si son observadas (indicadores):

$$y_{t,T} = x_{t,T}\beta + u_{t,T} \qquad \forall t,T$$

donde el vector de perturbaciones u se distribuye como una normal multivariante de media cero y matriz de varianzas y covarianzas V, $u \sim N(0,V)$. Además, la variable $y_{t,T}$ no es observable, pero sí su agregación anual $\sum_{t,}\, y_{t,T} = Y_T$. El objetivo es estimar $y_{t,T}$ de forma óptima, dadas la relación entre variable e indicador y la condición de agregación (4).

El procedimiento es el siguiente: en primer lugar, con los datos anuales se estima el vector β de forma óptima. Para ello se obtiene la relación sobre las magnitudes anuales observadas premultiplicando la relación trimestral por la matriz de agregación B dada en (5). Matricialmente:

$$By = Bx\beta + Bu \quad \Rightarrow Y = X\beta + U, \qquad U \sim N(0,V_*), \quad \text{con} \quad V_* = BVB$$

En este modelo se estima el vector β por Mínimos cuadrados generalizados:

$$\hat{\beta}_G = \left(X'V_*^{-1}X_o\right)^{-1} X'V_*^{-1}Y$$

La estimación final de la serie trimestral es:

$$y = x\beta_G + VB'V_*^{-1}(Y - X\beta_G)$$

Quilis(2000) analiza los diferentes métodos de trimestralización y concluye que:

1. En todos los métodos considerados se utilizan filtros lineales cuyos coeficientes varían con el tiempo, de forma que el filtro que se utiliza en los extremos de la serie no es el mismo que el que se aplica en el tramo central. Esta dependencia temporal del filtro genera inhomogeneidades y revisiones en la serie trimestral.

2. La distinción entre los métodos de ajuste y los basados en modelos no es una separación infranqueable. Ambos enfoques han de realizar hipótesis relativamente fuertes sobre la serie trimestral inobservable. Los primeros lo hacen indirectamente al plantear qué medida de volatilidad se desea minimizar y, los segundos, al definir qué estructura gobierna las propiedades estocásticas de dicha serie.

3. En todos los métodos basados en indicadores, la estimación de la serie trimestral se genera a partir de la suma algebraica de dos elementos, uno vinculado con el indicador y otro asociado con la distribución temporal de un residuo. En consecuencia, las propiedades dinámicas de la serie trimestralizada son una combinación de las de los indicadores y de las del residuo anual distribuido. Así, sus propiedades de carácter infraanual están determinadas por las del indicador. En particular, la estacionalidad de la serie trimestralizada es la del indicador igual que otros elementos de alta frecuencia como efectos de calendario, valores atípicos, etc.

4. El procedimiento de trimestralización propuesto por Chow y Lin (1971) ha adquirido una extraordinaria difusión debido a su generalidad, recurso a métodos de regresión sobradamente conocidos y de gran utilidad, consistencia con la práctica usual del

análisis de la coyuntura, empleo de un modelo estadístico explícito y facilidad de generalización al caso multivariante.

5. Frecuentemente, el método de Fernández, que es un caso límite del de Chow y Lin, resulta un procedimiento computacionalmente conveniente y compatible con la relación estimada entre agregado e indicadores en la frecuencia anual.

No obstante, siempre debe recordarse que el uso de métodos más complejos o de modelos más sofisticados y generales no conduce, de forma necesaria, a mejores resultados. En las situaciones prácticas de trimestralización, la longitud de las series, la disponibilidad de buenos indicadores y la calidad de la información disponible juegan un papel muy importante, de manera que técnicas muy sofisticadas pueden resultar poco adecuadas.

Señalar por último, que pueden utilizarse la macro SOLVER de Excel para programar estos métodos. Dicha macro utiliza el algoritmo del Gradiente Reducido Generalizado (GRG), en la versión GRG2.

6.3.- Índices trimestrales con encadenamiento anual

La aplicación de los índices encadenados a las series de alta frecuencia (mensuales o trimestrales) de tipo económico para la elaboración de las cuentas trimestrales, plantea una serie de inconvenientes a considerar. En primer lugar, las oscilaciones de la componente estacional e irregular pueden distorsionar y complicar las comparaciones entre dos periodos consecutivos. En segundo lugar, es necesario que las estimaciones de alta y baja frecuencia sean cuantitativamente consistentes. Además, el uso de un encadenamiento trimestral concatenando las valoraciones a precios del trimestre anterior, puede dar lugar a desviaciones sistemáticas o derivas que provocan un alejamiento del índice de su agregado anual.

Para subsanar estos problemas se utilizan métodos de encadenamiento anual descritos en el apartado 2.4. A continuación, se va a detallar la técnica de solapamiento anual, de acuerdo con el método utilizado para el cálculo de la Contabilidad Trimestral de España.

Antes de describir los distintos esquemas de encadenamiento anual de índices trimestrales, se exponen tres conceptos que se van a emplear de forma continua en todos ellos:

1. Cantidad media anual:

$$\bar{q}_{jT} = \frac{\sum_{t=1}^{4} q_{jtT}}{4}$$

2. Valor medio anual:

$$\bar{v}_{jT} = \frac{\sum_{t=1}^{4} v_{jtT}}{4} = \frac{\sum_{t=1}^{4} p_{jtT} \times q_{jtT}}{4}$$

3. Precio medio anual, es un concepto del tipo “valor unitario” que se deduce de los dos anteriores:

$$\bar{p}_{jT} = \frac{\bar{v}_{jT}}{\bar{q}_{jT}}$$

Los diversos procedimientos de encadenamiento tratan de resolver los problemas que plantea el encadenamiento cuando se aplica la metodología de índices encadenados al caso trimestral. En todos los casos la fórmula general se expresa como sigue:

$$CQ^L_{t/0} = \prod_{s=1}^{t} Q^L_{s/s-1}$$

Siendo los eslabones que se consideran, una aplicación directa de la fórmula de Laspeyres, información del trimestre anterior, tanto para efectuar la comparación como para calcular las ponderaciones. Estos dos aspectos van a ser modificados por los diversos esquemas de encadenamiento:

- solapamiento anual ("Annual Overlap Technique")
- solapamiento trimestral ("One-quarter Overlap Technique")

$$Q^L_{s/s-1[s-1]} = \sum_j w_{js-1} \frac{q_{js}}{q_{js-1}} = \frac{\sum_j p_{js-1} q_{js}}{\sum_j p_{js-1} q_{js-1}}$$

6.3.1.- Encadenamiento mediante solapamiento anual (annual overlap technique)

En el esquema de solapamiento anual el planteamiento es diferente, es este caso las ponderaciones van a ser las correspondientes a los valores medios del año anterior (T-1) y serán las mismas para todo el año T. De esta forma, la expresión del eslabón trimestral según esta técnica sería:

$$Q^L_{(t,T)/(T-1)[T-1]} = \sum_j w_{jT-1} \frac{q_{jtT}}{\overline{q}_{jT-1}} = \frac{\sum_j \overline{p}_{jT-1} q_{jtT}}{\sum_j \overline{p}_{jT-1} \overline{q}_{jT-1}}$$

donde $w_{jT-1} = \dfrac{\overline{p}_{jT-1}\overline{q}_{jT-1}}{\sum_j \overline{p}_{jT-1}\overline{q}_{jT-1}}$, $\overline{q}_{jT-1} = \dfrac{\sum_{t=1}^{4} q_{jtT-1}}{4}$, $\overline{p}_{jT-1} = \dfrac{\sum_{t=1}^{4} p_{jtT-1} q_{jtT-1}}{\sum_{t=1}^{4} q_{jtT-1}}$

el período actual es el trimestre *t* del año *T*, y la referencia y la base coinciden pero son anuales (*T-1*).

En la expresión anterior, q_{jtT} es el único elemento de alta frecuencia.

De esta forma, la cadena trimestral se construye de acuerdo con la expresión:

$$CQ^L_{(t,T)/0} = CQ^L_{T-1/0} Q^L_{(t,T)/T-1[T-1]} = \left(\prod_{S=1}^{T-1} Q^L_{S/S-1[S-1]} \right) Q^L_{(t,T)/T-1[T-1]}$$

Donde el primer término es el índice anual encadenado desde 0 hasta el periodo T-1 y el segundo término es el eslabón de Laspeyres trimestral calculado anteriormente.

Una interesante propiedad de estos índices es que su estructura de ponderaciones es igual que la de su homólogo anual.

En el caso de los índices de precio trimestrales de Paasche encadenados anualmente, la expresión del eslabón de la cadena, de acuerdo con el método de solapamiento anual sería:

$$P^P_{(t,T)/T-1[T-1]} = \sum_j w_{jtT} \frac{p_{jtT}}{\overline{p}_{jT-1}} = \frac{\sum_j p_{jtT} q_{jtT}}{\sum_j \overline{p}_{jT-1} q_{jtT}}$$

donde $w_{jtT} = \dfrac{\overline{p}_{jT-1} q_{jtT}}{\sum_j \overline{p}_{jT-1} q_{jtT}}$

Así, el índice de precios trimestral de Paasche encadenado anualmente quedaría:

$$CP^P_{(t,T)/0} = CP^P_{T/0} P^P_{(t,T)/T-1[T-1]} = \left(\prod_{S=1}^{T} P^P_{S/S-1[S-1]} \right) P^P_{(t,T)/T-1[T-1]}$$

donde $CP^P_{T/0}$ es la cadena anual, y $P^P_{S/S-1[S-1]} = \dfrac{\sum_j \overline{p}_{jT} \overline{q}_{jT}}{\sum_j \overline{p}_{jT-1} q_{jT}}$

La serie de volumen encadenada carece de unidades y puede quedarse como tal, como un número índice. No obstante, puede resultar conveniente expresar dichas series en términos monetarios, esto es, utilizando como numerario una unidad de cuenta específica (por ejemplo, euros o dólares). Existen dos maneras de conseguirlo.

En la primera se aplica un término o factor de valoración al índice de cantidad encadenado:

SERIE MONETARIA(t) = INDICE ENCADENADO(t) * FACTOR DE VALORACIÓN(0)

La segunda implica deflactar las cantidades trimestrales valoradas a precios medios del año anterior mediante el índice de precios anual de Paasche encadenado:

SERIE MONETARIA(t) = CANTIDAD(t) * PRECIO(0)

Ambas posibilidades son equivalentes.

Esta valoración se denomina "medida de volumen encadenado referida a su nivel nominal del año 0" y no refleja una valoración según los precios de un período específico.

6.3.2.- Encadenamiento mediante solapamiento en un trimestre (one-quarter overlap)

En este caso se utilizan pesos del año anterior (concretamente, los precios medios del año anterior valoran las cantidades del cuarto trimestre). Las comparaciones se efectúan con respecto al último trimestre del año anterior.

$$Q^L_{(t,T)/(4,T-1)[T-1]} = \sum_j w_{j,4,T-1} \frac{q_{jtT}}{\overline{q}_{j,4,T-1}} = \frac{\sum_j \overline{p}_{jT-1} q_{jtT}}{\sum_j \overline{p}_{jT-1} \overline{q}_{j,4,T-1}}$$

donde $w_{jT-1} = \dfrac{\overline{p}_{jT-1} q_{j,4,T-1}}{\sum_j \overline{p}_{jT-1} q_{j,4,T-1}}$, $\qquad \overline{p}_{jT-1} = \dfrac{\sum_{t=1}^{4} p_{jtT-1} q_{jtT-1}}{\sum_{t=1}^{4} q_{jtT-1}}$

Las ponderaciones son de naturaleza trimestral (es la valoración del cuarto trimestre), pero se mantienen fijas a lo largo de todo el año. Dado que ya no coinciden con las correspondientes anuales, se pierde la consistencia temporal.

Referencias bibliográficas

Akaike, H.: "A New Look at the Statistical Model Identification". *IEEE Transactions on Automatic Control*, vol. 19, 1974, pp. 716-723.

Anderson, T. W.: "Distribution of the serial correlation coefficient". *Annals of Mathematical Statistics*, vol. 13, 1942, pp. 1-13.

Boot, J. C. G., Feibes, W. y Lisman, J. H. C.: "Further methods of derivation of quaterly figures from annual data".Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 16, 1967, pp. 65-75.

Box, G. E. P., Jenkins, G. M. y Reinsel, G. C.: "Time Series Analysis: Forecasting and Control", 4ª edición. Wiley Series in Probability and Statistics. John Wiley. 2008.

Burns, A. F. and C. Mitchell W.C.: "Measuring Business Cycles". New York. National Bureau of economic Research. 1946.

Chow, G. and Lin, A.L.: "Best linear unbiased distribution and extrapolation of economic time series by related series", *Review of Economic and Statistics*, vol. 53. 1971. pp. 372-375.

Denton, F. T.: "Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization". *Journal of the American Statistical Association*, vol. 66, 1971, pp. 99-102.

Di Fonzo, T.: "The estimation of M disaggregate time series when contemporaneous and temporal aggregates are known", Review of Economic and Statistics, vol. 72, 1990, pp. 178-182.

Dickey, D. A. y Fuller, W. A.: "Distribution of the Estimators for Autorregressive Time Series with a unit root". *Journal of the American Statistical Association,* vol. 74, 1979, pp. 427-431.

Dickey, D. A. y Fuller, W. A.: "Likelihood ratio for Autorregressive Time Series with a unit root". *Econometrica,* vol. 49, 1981, pp. 1057-1072.

Espasa, A.: "Métodos Cuantitativos y Análisis de la Coyuntura Económica". Banco de España, Documento de Trabajo 8805. 1988.

Espasa, A.: "El perfil de crecimiento de un fenómeno económico". *Estadística Española,* vol. 33, nº 126, 1991, pp. 57-71.

Espasa, A. y Cancelo, J.R.: "Métodos Cuantitativos para el análisis de la coyuntura económica". Alianza Economía. 1993.

Fernández, R.B.: "Methodological note on the estimation of time series", *Review of Economic and Statistics*, vol. 63, 1981, pp. 471-478.

Fernández Macho, F. J.: "Indicadores Sintéticos de Aceleraciones y Desaceleraciones en la Actividad Económica". *Revista Española de Economía,* vol. 8, nº 1, 1991, pp. 125- 156.

Fernández Macho, F. J.: "El crecimiento subyacente en variables económicas". *Estadística Española,* vol. 33, nº 126, 1991, pp. 73-98.

Granger C. W. J. and Newbold P.: "Forecasting Economic Time Series", 2ª edición. Academic Press. 1986.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. y Shin, Y.: "Testing the null hypothesis of stationarity against the alternative of a unit root". *Journal of Econometrics*, vol. 54, 1992, pp. 159-178.

Lisman, J. H. C. y Sandee, J. "Derivation of quarterly figures from annual data". *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 13, nº 2, 1964, pp. 87-90.

Ljung, G.M. y Box, G.E.P. "On a measure of lacke of fit in time series models". *Biometrika*, vol. 65, 1978, pp. 297-303.

Maravall, A.: "On Structural Time Series Models and the Characterization of Components". *Journal of Business and Economic Statistics*, vol. 3, 1985, pp. 350-355.

Melis, F.:"La estimación del ritmo de variación en series económicas". *Estadística Española*, vol. 33, nº 126, 1991, pp. 7-56.

Mitchell, W.C. y Burns, A.F.: "Statistical Indicators of Cyclical Residuals". Bulletin 69. NBER. New York. 1938.

Quilis, E.: "Notas sobre desagregación temporal de variables económicas". Instituro de Estudios Fiscales, Papeles de trabajo PT. Nº 1/01. Disponible en http://www.ief.es/documentos/recursos/publicaciones/papeles_trabajo/2001_01.pdf.

Rodríguez, J.: "Un ciclo de referencia para la economía española: primeras aproximaciones". Banco de España, Servicio de Estudios. 1976.

Rodríguez, J.: "Una aproximación al ciclo de referencia de la economía española: 1965-1975". Banco de España, Servicio de Estudios. 1977.

Schwartz, G. "Estimating the dimension of a model". *The Annals of Statistics*, vol. 6, 1978, pp. 461-464.