BIBLIOTECA VIRTUAL
MIGUEL DE CERVANTES

Universidad de Valladolid

Estimación de identificación de modelos de volatilidad estocástica con memoria larga

Ana Pérez Espartero

Tesis de Doctorado

Facultad: Ciencias Económicas y Empresariales

Directora: Dra. Esther Ruiz Ortega

2000

UNIVERSIDAD DE VALLADOLID

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

BIBLIOTECA VIRTUAL

ESTIMACIÓN E IDENTIFICACIÓN DE MODELOS DE VOLATILIDAD ESTOCÁSTICA CON MEMORIA LARGA

Ana Pérez Espartero

Proyecto de Tesis Doctoral dirigido por la profesora
Esther Ruiz Ortega

Valladolid, 2000

AGRADECIMIENTOS

Quisiera expresar aquí mi agradecimiento a las personas que han hecho posible este trabajo.

En primer lugar, al profesor A.C. Harvey, con quien realicé los cursos de series temporales en la London School of Economics. Bajo su tutela comencé a estudiar los modelos de Volatilidad Estocástica y con él realicé un proyecto de investigación que fue el germen de este proyecto de tesis. Contar con su colaboración supuso para mí un gran aliciente. Por ello, siempre le estaré agradecida.

En segundo lugar, al Departamento de Estadística y Econometría de la Universidad Carlos III de Madrid, por acogerme con tanto cariño durante varios meses de los años 1998 y 1999. En especial a Lorenzo, Angeles, Jose y Sara, por tantos momentos compartidos. El ambiente de trabajo que "respiré" en este departamento hizo muy fructífera mi estancia y contribuyó, en gran medida, a la realización de este trabajo.

En tercer lugar, al Departamento de Estadística y Econometría de la Universidad de Valladolid, por facilitar mi estancia en las universidades antes mencionadas, sin las cuales este trabajo no hubiera sido posible. Mi agradecimiento más sincero a los compañeros del departamento que me apoyaron en los momentos difíciles y que siempre estuvieron a mi lado, en especial a Mercedes, por todos los momentos, buenos y malos, que hemos compartido, aquí y allí.

En cuarto lugar, a mi familia, y en especial a Tula, Andrea y Paula, por hacer mi estancia en Getafe más llevadera, a mis amigos, por su apoyo incondicional en estos años y a Ramón por tanta paciencia.

Y finalmente, a la persona que ha hecho realidad este proyecto, mi directora de tesis, la profesora Esther Ruiz. Quisiera expresarle aquí mi agradecimiento más profundo por haber aceptado dirigir este trabajo en un momento crítico de mi carrera universitaria y por haberlo dirigido con tanta sabiduría y con tanto cariño.

ÍNDICE

INTRODUCCIÓN 1

Capítulo 1. MODELIZACIÓN DE LA VOLATILIDAD Y MODELOS ARFIMA 11

1.1. Modelización de los cambios temporales en la volatilidad 11

1.1.1. Modelos tipo ARCH 13
1.1.1.1. Definición y propiedades 14
1.1.1.2. Estimación 20

1.1.2. Modelos de volatilidad estocástica SV 21
1.1.2.1. El modelo SV autorregresivo 22
1.1.2.2. Extensiones del modelo SV autorregresivo 28
1.1.2.3. Estimación 30

1.2. Modelos con memoria larga para la media: procesos ARFIMA 34

1.2.1. Definición y propiedades 38
1.2.2. Estimación 46
1.2.3. Contrastes de hipótesis 55

1.3. Modelos con memoria larga para la volatilidad 61

1.3.1. Modelos GARCH fraccionalmente integrados 65
1.3.2. El modelo SV con memoria larga. 72
1.3.3. Otros modelos 85

Capítulo 2. ESTIMACIÓN DE LA AUTOCOVARIANZA Y LA AUTOCORRELACIÓN EN MODELOS DE VOLATILIDAD ESTOCÁSTICA 89

2.1. Introducción 89

2.2. Contrastes de incorrelación serial en presencia de volatilidad estocástica 95

2.2.1. Efectos SV sobre las correlaciones muestrales 97
2.2.2. Propiedades en muestras finitas del contraste de incorrelación de las bandas de Barlett en modelos SV 110
2.2.3. Efectos SV sobre el estadístico de Box-Pierce 131
2.2.4. Propiedades en muestras finitas del contraste de incorrelación de Box-Pierce en modelos SV 132

2.3. Autocovarianzas muestrales de la serie de logaritmos de los cuadrados 141

2.3.1. Sesgo y varianza de las autocovarianzas muestrales de la serie de logaritmos de los cuadrados en el modelo ARSV(1) 142
2.3.2. Propiedades en muestras finitas de las covarianzas muestrales de la serie de logaritmos de los cuadrados en el modelo ARSV(1) 158
2.3.3. Sesgo y varianza de las autocovarianzas muestrales de la serie de logaritmos de los cuadrados en el modelo LMSV 164
2.3.4. Propiedades en muestras finitas de las covarianzas muestrales de la serie de logaritmos de los cuadrados en el modelo LMSV 170

2.4. Autocorrelaciones muestrales de la serie de logaritmos de los cuadrados 177

2.4.1. Sesgo y varianza de las autocorrelaciones muestrales de la serie de logaritmos de los cuadrados en el modelo ARSV(1) 179
2.4.2. Propiedades en muestras finitas de las correlaciones muestrales de la serie de los cuadrados y de la serie de logaritmos de los cuadrados en el modelo ARSV(1) 190
2.4.3. Sesgo y varianza de las autocorrelaciones muestrales de la serie de logaritmos de los cuadrados en el modelo LMSV 207
2.4.4. Propiedades en muestras finitas de las correlaciones muestrales de la serie de los cuadrados y de la serie de logaritmos de los cuadrados en el modelo LMSV 211

Capítulo 3. ESTIMACIÓN DEL MODELO LMSV 227

3.1. Introducción 227

3.2. Estimación pseudo-máximo verosímil en el dominio de las frecuencias 233

3.2.1. Estimación del modelo LMSV estacionario 238
3.2.2. Estimación del modelo LMSV no estacionario 241

3.3. Propiedades de los estimadores PMV en muestras finitas. 244

3.3.1. Propiedades de los estimadores en el modelo LMSV estacionario 244
3.3.2. Propiedades de los estimadores en el modelo LMSV no estacionario 257
3.3.3. Propiedades de los estimadores en el modelo ARLMSV 262

3.4. Resultados preliminares de la estimación PMV en el dominio temporal con el Filtro de Kalman 267

3.5. Estimación suavizada de la volatilidad 277

Capítulo 4. APLICACIÓN EMPÍRICA: MODELIZACIÓN DE LA VOLATILIDAD DEL ÍNDICE IBEX-35 287

4.1. Análisis preliminar de los rendimientos del IBEX-35 287

4.2. Estimación de modelos SV para la volatilidad del IBEX-35 301

CONCLUSIONES **317**

APÉNDICES **325**

A. Simulación del proceso ARFIMA(0,d,0) 325

B. Forma reducida del modelo ARSV(1).....329

C. Cumulantes de cuarto orden y momentos de orden superior de la serie $\ln(y_t^2)$ en modelos SV.....332

D. Condiciones de Anderson (1971, p. 471) 337

BIBLIOGRAFÍA **341**

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

INTRODUCCIÓN

Las series temporales de rendimientos de activos financieros observadas con gran frecuencia, suelen caracterizarse, fundamentalmente, por tener exceso de curtosis, escasa correlación en los niveles y correlaciones positivas en la serie de los cuadrados. Estas características se relacionan habitualmente con la presencia de volatilidades cambiantes a lo largo del tiempo, o dicho de otro modo, con la existencia de periodos en los que apenas cambia el valor de la serie alternándose con periodos en los que el margen de variación en los rendimientos es muy amplio.

Entre los distintos modelos propuestos en la literatura econométrica para representar la evolución dinámica de la volatilidad, podríamos distinguir dos grandes grupos: los modelos de heteroscedasticidad condicionada tipo ARCH (AutoRegressive Conditional Heteroscedasticity), basados en la propuesta inicial de Engle (1982), y los modelos de volatilidad estocástica SV (Stochastic Volatility) propuestos por Taylor (1986) y desarrollados posteriormente por Harvey, Ruiz y Shephard (1994).

Ambos modelos reproducen algunas de las propiedades típicas de las series financieras, como exceso de curtosis, agrupamiento de los periodos de la volatilidad, correlación en los cuadrados de la serie, etc, pero difieren sustancialmente en la forma en que se define la ecuación que modeliza la volatilidad. En los modelos tipo ARCH, la volatilidad se considera una función lineal de los cuadrados de las observaciones pasadas de la serie y de su propio pasado, mientras que en los modelos

SV, la volatilidad es una componente no observable cuyo logaritmo suele modelizarse mediante un proceso lineal autorregresivo.

Un resultado muy habitual en aplicaciones empíricas de estos modelos a series reales, es la obtención de una raíz casi unitaria en el proceso que describe la volatilidad. Este resultado concuerda con el comportamiento de las correlaciones muestrales de la serie de los cuadrados, que son significativamente distintas de cero, incluso en retardos muy alejados, y decaen muy lentamente hacia cero, indicando que el efecto de las innovaciones a la volatilidad persiste durante mucho tiempo.

Los modelos ARCH y los modelos SV pueden capturar esta persistencia en la varianza condicional incluyendo de forma explícita raíces unitarias en su componente autorregresiva. Los modelos integrados resultantes se caracterizan, en términos de predicción, porque el efecto de una innovación a la varianza condicional persiste para siempre, en el sentido de que influye en las predicciones futuras de la volatilidad para todos los horizontes de predicción.

Como alternativa a los modelos con raíz unitaria, algunos autores argumentan que sería más apropiado modelizar esa persistencia en los cuadrados de los rendimientos financieros utilizando procesos con memoria larga, ya que el grado extremo de persistencia que implican los modelos integrados no parece ajustarse a las propiedades empíricas de ciertas series financieras. Por ejemplo, Ding, Granger y Engle (1993) y Bollerslev y Mikkelsen (1996) analizan series diarias de rendimientos bursátiles y tipos de cambio, respectivamente, y comprueban que existe un alto grado de dependencia en los cuadrados de dichas series, pero que esta dependencia desaparece finalmente con el horizonte de predicción, eliminando así la posibilidad de un comportamiento integrado. Además, estos autores observan que las correlaciones de los cuadrados de los rendimientos decrecen a un ritmo muy lento, que contrasta con el decrecimiento exponencial impuesto por los modelos ARCH o SV de memoria corta. Crato y de Lima (1994), Breidt, Crato y de Lima (1998), Lobato y Savin (1998) y Olmeda (1998), aplican diferentes contrastes de hipótesis

para detectar la presencia de memoria larga en los cuadrados de diversos índices bursátiles y concluyen que en ninguna de las series analizadas existe evidencia empírica que permita rechazar la hipótesis de memoria larga.

Los resultados de estos trabajos sugieren, en primer lugar, que la varianza condicional de los rendimientos de activos financieros puede tener mayor memoria que la típicamente capturada por los modelos tipo ARCH o SV estacionarios. En segundo lugar, estos trabajos confirman las limitaciones de los modelos con raíces unitarias para reproducir la persistencia observada en los datos financieros, en favor de los procesos con memoria larga.

Actualmente existe ya una literatura muy extensa sobre procesos con memoria larga para la modelización de la media y, en particular, sobre los procesos autorregresivos y de medias móviles fraccionalmente integrados, denominados procesos ARFIMA. Baillie (1996), Beran (1994) y Robinson (1994b), constituyen buenas referencias a este respecto. Sin embargo, la existencia de memoria larga en los momentos condicionados de segundo orden y su correcta modelización, son temas sobre los que todavía existe una literatura escasa. Precisamente, es en este ámbito en el que se enmarca nuestro trabajo de investigación.

Los modelos con memoria larga para la volatilidad propuestos hasta este momento son extensiones de los modelos ARCH y SV que incorporan la integración fraccional con el fin de recoger la memoria larga observada empíricamente en la volatilidad de ciertas series financieras.

En el contexto de los modelos ARCH, Baillie, Bollerslev y Mikkelsen (1996) y Bollerslev y Mikkelsen (1996) proponen el modelo GARCH fraccionalmente integrado (FIGARCH) y el modelo EGARCH fraccionalmente integrado (FIEGARCH), respectivamente. Aunque estos modelos se están utilizando en aplicaciones empíricas con series reales, e incluso están ya implementados en algunos

paquetes informáticos,[1] existen aún muchas cuestiones sin resolver en cuanto a su definición, sus propiedades dinámicas y, fundamentalmente, en cuanto a la estimación de sus parámetros. Para ninguno de ellos existe aún teoría rigurosa sobre las propiedades de los estimadores propuestos.

Como alternativa a estos modelos, Robinson y Zaffaroni (1997) proponen los modelos MA no lineales con memoria larga, y derivan sus propiedades dinámicas para la serie en niveles y la serie de los cuadrados, comprobando que el modelo reproduce muchas de las propiedades empíricas de los rendimientos financieros. Zaffaroni (1997) demuestra que, bajo ciertas condiciones, los estimadores pseudo-máximo verosímiles de Whittle de los parámetros de este modelo, son consistentes y asintóticamente normales.

En el ámbito de los modelos SV, Harvey (1998) y Breidt, Crato y de Lima (1998) han propuesto, simultáneamente, el modelo de Volatilidad Estocástica con Memoria Larga o modelo LMSV (Long Memory Stochastic Volatility), en el que el logaritmo de la volatilidad es un proceso ARFIMA. Las propiedades dinámicas de este modelo son fáciles de obtener e interpretar a partir de las propiedades del proceso ARFIMA subyacente. Desafortunadamente, la estimación del modelo es difícil porque la función de verosimilitud no puede construirse de forma exacta de una manera sencilla y, consecuentemente, hay que buscar métodos de estimación alternativos.

Uno de los objetivos de nuestro trabajo es, precisamente, analizar algunos de los problemas que existen en la estimación de los parámetros del modelo LMSV, y en particular, en el método de estimación pseudo-máximo verosímil (PMV) en el dominio de las frecuencias, propuesto por Harvey (1998) y Breidt, Crato y de Lima (1998). Además, nuestro trabajo pretende abordar otros problemas, como son la detección, la identificación y la validación de modelos SV, que no se habían

[1] Véase por ejemplo E-Views y el módulo GARCH del paquete GAUSS.

planteado hasta este momento en la literatura y que pueden tener una gran relevancia en aplicaciones empíricas con series reales.

En primer lugar, nos proponemos investigar cuáles son las propiedades estadísticas del correlograma de la serie original y del correlograma de la serie de sus cuadrados, en presencia de volatilidad estocástica, y cuál es el efecto que estas propiedades tienen sobre los contrastes de incorrelación que de ellos se derivan.

El estudio de las propiedades del correlograma tiene interés porque el correlograma constituye una de las herramientas básicas en la identificación y validación de modelos de series temporales. Concretamente, en la modelización de la volatilidad, el correlograma de la serie de los cuadrados se utiliza como primer indicio de la dependencia temporal en los segundos momentos condicionales. Por esta razón, es importante conocer las propiedades del correlograma de dicha serie en presencia de volatilidad estocástica.

Uno de nuestros objetivos es derivar expresiones del sesgo y de la varianza asintótica de las covarianzas y de las correlaciones muestrales para la serie de los cuadrados, tanto para los modelos SV autorregresivos como para los modelos SV con memoria larga. Además, nos interesa analizar también la validez de dichos resultados en muestras finitas. Así mismo, queremos estudiar las propiedades de los contrastes de incorrelación habituales en presencia de volatilidad estocástica.

En este sentido, la aportación más importante de nuestro trabajo es haber probado que, en los modelos de volatilidad estocástica con memoria larga, el sesgo de las correlaciones muestrales de la serie de los cuadrados es de tal magnitud que cuestiona la validez del correlograma de dicha serie como herramienta para la identificación y la validación del modelo. De hecho, su utilización puede conllevar la selección de un modelo erróneo, ya que, con los tamaños de muestra habituales, detectar la presencia de memoria larga a través del correlograma es extremadamente difícil, incluso cuando el modelo es muy persistente.

Finalmente, por lo que respecta a la estimación pseudo-máximo verosímil (PMV) en el dominio de las frecuencias, los resultados que se conocen hasta el momento son muy limitados. Breidt, Crato y de Lima (1998) demuestran la consistencia de los estimadores PMV en el dominio de las frecuencias y presentan un estudio de Monte Carlo a partir del cual concluyen que, para ciertos valores de los parámetros, los estimadores tienen buenas propiedades en muestras finitas. Sin embargo, este resultado debe interpretarse con cautela, porque los valores que utilizan estos autores en sus simulaciones no son muy realistas y difieren mucho de los que se obtienen después al estimar el modelo con datos reales.

Nuestro objetivo es extender el estudio de Breidt, Crato y de Lima (1998) en varios sentidos. Por un lado, incorporando en el diseño valores de los parámetros más realistas y modelos no estacionarios y, por otro lado, realizando la estimación de todos los parámetros del modelo. De esta manera, podremos descubrir cuáles son realmente los problemas que pueden surgir al aplicar el método de estimación PMV en la estimación de series financieras reales.

La conclusión más importante que se deriva de nuestros resultados es que existen problemas serios en la estimación PMV de los modelos de volatilidad estocástica con memoria larga cuando los parámetros del modelo toman valores próximos a las fronteras de la no estacionariedad y/o de la homocedasticidad. En estos casos, que corresponden, precisamente, a los modelos más realistas, sería necesario disponer de muestras extremadamente grandes para poder hacer inferencias fiables. Nuestro trabajo supone también una llamada de atención sobre el problema de identificación que existe en el modelo SV autorregresivo y con memoria larga, cuando el proceso que define la volatilidad tiene casi una raíz unitaria. Como veremos más adelante, en estos casos el estimador PMV parece incapaz de distinguir la procedencia de la no estacionariedad y esto puede conllevar la estimación de un modelo erróneo.

El trabajo se organiza de la siguiente manera. En el **capítulo 1** se revisan los principales modelos propuestos hasta este momento en la literatura econométrica para modelizar el comportamiento de la volatilidad de las series financieras y se revisan los modelos con memoria larga para la media. En primer lugar, se presentan los modelos de heteroscedasticidad condicionada tipo ARCH y los modelos SV. Se describen las propiedades estadísticas más relevantes de cada uno de ellos y se exponen los métodos de estimación que con más éxito se han utilizado en aplicaciones empíricas. Se presentan asimismo los nuevos modelos con memoria larga para la varianza condicional, con especial atención al modelo de volatilidad estocástica con memoria larga (LMSV), que constituye el centro de nuestra investigación. Para poder describir correctamente estos modelos, se revisan previamente los procesos autorregresivos y de medias móviles fraccionalmente integrados (ARFIMA). Se describen sus propiedades dinámicas más importantes y se presentan los principales resultados sobre estimación y contrastes de hipótesis para estos modelos.

No se trata en ningún caso de hacer una exposición exhaustiva de cada uno de los modelos existentes, sino más bien de resumir aquellos aspectos de la modelización de la volatilidad y de los modelos fraccionalmente integrados, que consideramos imprescindibles para el desarrollo de nuestro trabajo.

El **capítulo 2** se dedica al análisis de las propiedades estadísticas del correlograma en series con volatilidad estocástica. En este capítulo se presentan las propiedades del correlograma, tanto asintóticas como en muestras finitas, para el modelo de volatilidad estocástica autorregresivo y el modelo de volatilidad estocástica con memoria larga. Se obtienen las expresiones del sesgo y de la varianza asintótica de las covarianzas y de las correlaciones muestrales para la serie del logaritmo de los cuadrados y, en algunos casos, se obtiene además su distribución asintótica. Mediante simulaciones de Monte Carlo, se estudia además la validez de los resultados en muestras finitas. Se analiza asimismo el efecto que tiene la presencia de volatilidad estocástica sobre el tamaño y la potencia de los contrastes habituales

de incorrelación serial para la serie en niveles. En concreto, se estudia el comportamiento asintótico y en muestras finitas del estadístico de Box-Pierce para el modelo de volatilidad estocástica autorregresivo y el modelo de volatilidad estocástica con memoria larga.

En el **capítulo 3** se discute el problema de la estimación del modelo de volatilidad estocástica con memoria larga. Se describe el método de estimación por pseudo-máxima verosimilitud en el dominio de las frecuencias propuesto por Harvey (1998) y se presentan las propiedades asintóticas de los estimadores correspondientes, deducidas por Breidt, Crato y de Lima (1998). Para analizar la validez de estos resultados en muestras finitas se realiza un extenso estudio de Monte Carlo que complementa el realizado por Breidt, Crato y de Lima (1998) en varios sentidos. En primer lugar, se incorporan resultados para el estimador de la varianza de la perturbación del modelo, que es precisamente el que tiene peores propiedades en muestras finitas. En segundo lugar, se estudia el comportamiento de los estimadores para un rango más amplio de valores de los parámetros que incluye modelos más realistas que los que se presentan en Breidt, Crato y de Lima (1998) y modelos no estacionarios. Se discute además el problema de identificación que existe en los modelos SV autorregresivos con memoria larga cuando el proceso que define la volatilidad tiene una raíz unitaria. Como alternativa al método de estimación PMV en el dominio de las frecuencias, se propone un método de estimación PMV en el dominio temporal mediante el filtro de Kalman, basado en la propuesta de Chan y Palma (1998). Se presentan los resultados de un pequeño experimento de Monte Carlo que, en principio, parecen favorables a la estimación en el dominio de las frecuencias. Finalmente, se describe el método propuesto por Harvey (1998) para obtener una estimación suavizada de la volatilidad en modelos LMSV y se resuelve el problema de su aplicación en muestras grandes, donde la propuesta original, que requiere la inversión de matrices TxT, es inviable.

En el **capítulo 4** se presenta una aplicación empírica de los resultados descritos en los capítulos anteriores para una serie diaria del índice bursátil IBEX-35,

en el periodo comprendido entre el 5/1/1987 y el 30/12/1998. Este índice recoge la cotización de los 35 valores con mayor capitalización bursátil en la Bolsa española. Esta aplicación permite ilustrar, por una parte, los problemas que pueden aparecer cuando se utilizan los métodos de estimación pseudo-máximo verosímil para la estimación de la volatilidad de series financieras reales. Por otra parte, esta aplicación confirma los problemas de identificación y selección de modelos que se describieron en los capítulos anteriores.

Finalmente, el **capítulo 5** presenta una recapitulación de las aportaciones más importantes de nuestro trabajo y sugiere algunas de las líneas de investigación en las que pretendemos seguir trabajando en el futuro**.**

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Capítulo 1.

MODELIZACIÓN DE LA VOLATILIDAD Y MODELOS ARFIMA

1.1. Modelización de los cambios temporales en la volatilidad

Muchas series temporales económicas, y especialmente series financieras observadas con gran frecuencia, muestran cambios temporales en los momentos condicionados de segundo orden. Estos cambios tienden a estar correlacionados serialmente, en el sentido de que cambios de gran magnitud en el valor de la serie son seguidos por grandes cambios (periodos de mucha volatilidad), mientras que a cambios pequeños en el valor de la serie les siguen cambios pequeños (periodos de poca volatilidad). En términos estadísticos, este agrupamiento de los periodos de volatilidad se traduce en la presencia de correlaciones positivas en la serie de los cuadrados. Esta propiedad, junto a otras como el exceso de curtosis o la ausencia de correlación en los niveles, son consideradas ya por algunos autores como características típicas de las series financieras.

Algunas de estas características ya fueron observadas por Mandelbrot (1963), pero ha sido el trabajo seminal de Engle (1982) el que ha propiciado el desarrollo de

numerosos modelos que tratan de representar este comportamiento de las series financieras. La formulación básica en todos ellos consiste en modelizar la serie de interés, $\{y_t\}_{t=1}^{T}$, según la ecuación:

$$y_t = \sigma_t \varepsilon_t \qquad [1.1]$$

donde ε_t es una secuencia de variables aleatorias independientes e igualmente distribuidas (I.I.D.) con media cero y varianza unitaria, y σ_t es un factor denominado volatilidad.

Por definición, la serie y_t en [1.1] es un proceso serialmente incorrelacionado con media cero, y σ_t es una función positiva. Los diferentes modelos propuestos para representar la evolución de la volatilidad surgen, precisamente, de las diferentes formas de modelizar la función σ_t^2. Los dos enfoques más utilizados han sido el modelo de heteroscedasticidad condicionada autorregresivo (ARCH) introducido por Engle (1982), y todas sus posibles extensiones, y el modelo de volatilidad estocástica (SV) propuesto por Taylor (1986). Ambos modelos son capaces de reproducir algunas de las propiedades típicas de las series financieras, como exceso de curtosis, agrupamiento temporal de la volatilidad y correlación en los cuadrados, pero difieren sustancialmente en la forma en que se define la ecuación de la volatilidad. Mientras en los modelos ARCH, σ_t es función de las observaciones pasadas de la serie, en los modelos SV, σ_t es función de un proceso estocástico no observable.

En las dos secciones siguientes se presenta un resumen de los aspectos teóricos más relevantes de los modelos tipo ARCH y SV, especialmente de aquellos utilizados con mayor frecuencia en aplicaciones empíricas. Nuestro objetivo no es hacer una exposición exhaustiva del tema, sino presentar una visión general que nos permita establecer el marco de referencia en el que se va a desarrollar nuestro trabajo.

La sección 1.1.1 presenta los modelos de heteroscedasticidad condicionada, con especial atención al modelo ARCH de Engle (1982) y a algunas generalizaciones suyas: el modelo ARCH generalizado (GARCH), el modelo GARCH integrado (IGARCH) y el modelo GARCH exponencial (EGARCH). Para cada uno de ellos se analizan sus propiedades estadísticas y se describen los métodos de estimación más utilizados. En la sección 1.1.2 se presentan los modelos de volatilidad estocástica SV. Se estudia con detalle el modelo más sencillo de esta clase, el modelo log-normal de Taylor (1986), y se plantean posibles generalizaciones del mismo. Se describen también los diversos métodos de estimación propuestos en la literatura.

1.1.1. Modelos tipo ARCH

En los modelos de heteroscedasticidad condicionada tipo ARCH, el factor σ_t es una función de las observaciones pasadas de la serie y, por tanto, la volatilidad de hoy depende de la variabilidad de las observaciones anteriores. Desde el punto de vista estadístico, esta clase de modelos tiene una gran ventaja: una vez especificada una distribución para ε_t, la distribución condicionada de y_t/Ω_{t-1}, donde Ω_{t-1} denota el conjunto de información disponible hasta el momento t-1, éste incluido, puede determinarse fácilmente, y su varianza es, precisamente, σ_t^2. A partir de estas distribuciones condicionadas, la función de verosimilitud se construye de forma exacta utilizando la descomposición del error de predicción, lo cual permite obtener estimadores máximo verosímiles y realizar ciertos contrastes de hipótesis. Además, existe cierto paralelismo entre los modelos ARCH y los modelos autorregresivos de medias móviles ARMA, extensamente utilizados para modelizar cambios en la media, lo que va a permitir aplicar algunas técnicas conocidas de la metodología Box Jenkins.

Pero los modelos ARCH también adolecen de ciertos inconvenientes. Por ejemplo, la distribución marginal de y_t suele ser complicada de obtener y tampoco es

fácil establecer las condiciones de estacionariedad ni calcular los momentos de y_t. Además, estos modelos incluyen una única perturbación, ε_t, que afecta conjuntamente al nivel y a la volatilidad de la serie, lo cual resta flexibilidad al modelo y hace que, por ejemplo, el exceso de curtosis y la persistencia de las correlaciones de los cuadrados sean dos características ligadas entre sí, en el sentido de que una no puede cambiar sin que cambie la otra.

Pese a todo, estos modelos han recibido una gran atención en la literatura econométrica de los últimos años. Véanse por ejemplo las excelentes revisiones de Bollerslev, Chou y Kroner (1992), que recopila las principales aplicaciones empíricas de modelos GARCH a series financieras y proporciona cientos de referencias al respecto, Bera y Higgins (1993) y Bollerslev, Engle y Nelson (1994), de carácter más teórico, o más recientemente Palm (1996), que recoge los últimos resultados en modelos GARCH para la volatilidad de series financieras y los compara con otras posibles alternativas. Véase también Engle (1995), que es una recopilación de artículos ya clásicos en la literatura econométrica sobre modelos GARCH, y la revisión sobre modelización ARCH de Novales y Gracia-Díez (1993).

A continuación se describen brevemente las propiedades del modelo ARCH propuesto por Engle (1982) y algunas generalizaciones suyas, así como los métodos de estimación propuestos en la literatura.

1.1.1.1. Definición y propiedades

El modelo ARCH propuesto por Engle (1982) permite modelizar los cambios temporales de la varianza condicional como una función de las observaciones pasadas de la serie. En concreto, en el modelo ARCH(q), la varianza condicional, σ_t^2, es una función lineal de los cuadrados de q observaciones pasadas de y_t, de la forma:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 \qquad [1.2]$$

donde $\alpha_0>0$, $\alpha_i\geq 0$, i=1,…,q. Estas restricciones sobre los parámetros se imponen para asegurar que la varianza sea estrictamente positiva. Además, la condición $\alpha_0>0$ es imprescindible para evitar que el proceso degenere si en algún momento se observara $y_t=0$.

El modelo definido por las ecuaciones [1.1] y [1.2] es una martingala en diferencias[1], ya que $E(y_t/\Omega_{t-1})=0$, y, en consecuencia, la serie y_t tiene media cero y está serialmente incorrelacionada. Engle (1982) demuestra que el modelo ARCH(q) es estacionario[2] si y sólo si $\sum_{i=1}^{q} \alpha_i < 1$, en cuyo caso la varianza marginal de y_t es $Var(y_t)= \alpha_0 /(1-\sum_{i=1}^{q} \alpha_i)$. Además, para el modelo más sencillo de esta clase, el ARCH(1) con perturbación gaussiana, Engle (1982) prueba que si $3\alpha_1^2<1$, el coeficiente de curtosis de y_t es $\frac{E(y_t^4)}{\left[E(y_t^2)\right]^2} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2}$. Si $\alpha_1>0$, este valor es mayor que 3 y, por tanto, el modelo ARCH(1) generará datos con colas más anchas que las de la distribución normal, propiedad de gran interés si se tiene en cuenta que ésta es precisamente una característica típica de muchas series financieras.

Un rasgo común a muchas de las primeras aplicaciones empíricas de los modelos ARCH es que requerían un gran número de parámetros autorregresivos y, para representar adecuadamente el comportamiento dinámico de la varianza, se

---

[1] Un proceso $\{y_t, t\geq 1\}$ es una martingala en diferencias si $E|y_t|<\infty$ y $E(y_t/\Omega_{t-1})=0$, donde Ω_{t-1} denota el conjunto de información disponible hasta el momento t-1, éste incluido.

[2] Salvo que se especifique lo contrario, cuando hablemos de un proceso estacionario nos estaremos refiriendo a la estacionariedad débil o estacionariedad en covarianza, es decir a un proceso y_t tal que $E(y_t)$ es constante y $Cov(y_t,y_s)$ sólo depende de $|t-s|$.

imponía una estructura fija de retardos.[3] Con el fin de flexibilizar estas restricciones Bollerslev (1986) propuso el modelo ARCH generalizado o GARCH.

La generalización del modelo ARCH al modelo GARCH tiene gran similitud con la extensión de los procesos autorregresivos, AR, a los autorregresivos y de medias móviles, ARMA, permitiendo una representación más parsimoniosa de la volatilidad. Bollerslev (1986) considera que la varianza condicional, σ_t^2, además de depender de las observaciones pasadas de y_t, depende también de su propio pasado. Esta dependencia se expresa incluyendo en la ecuación [1.2] cierto número p de retardos de σ_t^2, de forma que la varianza condicional se define entonces como:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad [1.3]$$

donde $\alpha_0>0$, $\alpha_i\geq 0$, i=1,..,q, $\beta_j\geq 0$, j=1,…,p.[4] El nuevo modelo se denomina GARCH(p,q), y se reduce al ya conocido ARCH(q) cuando p=0. Bollerslev (1986) establece las condiciones de estacionariedad, probando que y_t es débilmente estacionario, con $E(y_t)=0$, $Var(y_t)= \alpha_0 / (1- \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j)$ y $Cov(y_t,y_s)=0$ para $s\neq t$, si y sólo si $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$.

Uno de los atractivos de los modelos GARCH es su relación con los modelos ARMA. Definiendo una nueva perturbación, υ_t, como:

$$\upsilon_t = y_t^2 - \sigma_t^2 = \sigma_t^2 (\varepsilon_t^2 - 1) \qquad [1.4]$$

---

[3] Véase Engle (1983) y otras referencias citadas en Engle y Bollerslev (1986) a este respecto.

[4] Aunque estas restricciones se establecen para garantizar que la varianza sea positiva, Nelson y Cao (1992) demuestran posteriormente que la positividad de la varianza está asegurada bajo condiciones más débiles. En concreto demuestran que si el modelo GARCH de la ecuación [1.3] admite una representación ARCH(∞), es suficiente exigir que los coeficientes del polinomio de retardos en dicha representación sean todos positivos.

y reagrupando términos en [1.3], se comprueba fácilmente que el modelo GARCH puede representarse de forma equivalente, según la ecuación:

$$y_t^2 = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) y_{t-i}^2 + \upsilon_t - \sum_{j=1}^{p} \beta_j \upsilon_{t-j} \qquad [1.5]$$

donde m=max(p,q), α_i=0 si i>q y β_i=0 si i>p. Por definición, la perturbación υ_t en [1.4] tiene media cero y está serialmente incorrelacionada, y, por tanto, la expresión [1.5] permite interpretar el modelo GARCH(p,q) como un modelo ARMA(m,p) para la serie y_t^2, propiedad que, como señala Bollerslev (1988), puede utilizarse para identificar los órdenes p y q del modelo.

Muchos trabajos empíricos con series financieras muestran que el más sencillo de los modelos GARCH, el GARCH(1,1), es suficiente para modelizar con éxito los cambios temporales en la varianza condicional, incluso sobre periodos muestrales largos.[5] El modelo GARCH(1,1) se obtiene cuando p=q=1 en la ecuación [1.3] de forma que la varianza condicionada se define como:

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \qquad [1.6]$$

con $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$. Si $\alpha_1+\beta_1 < 1$, la serie y_t tiene varianza finita, y por ser una martingala en diferencias, es ruido blanco, de media 0 y varianza $\alpha_0 / (1-\alpha_1-\beta_1)$. Además, Bollerslev (1986) prueba que si $(\beta_1+\alpha_1)^2+2\alpha_1^2 < 1$, el momento de orden 4 de y_t existe y es finito, y la curtosis de y_t es $\frac{E(y_t^4)}{[E(y_t^2)]^2} = \frac{3[1-(\alpha_1+\beta_1)^2]}{1-2\alpha_1^2-(\alpha_1+\beta_1)^2}$. Cuando

[5] Véase Bollerslev, Chou y Kroner (1992) para referencias al respecto.

$\alpha_1+\beta_1 <1$, este valor es mayor que 3 y, por tanto, el proceso GARCH(1,1) estacionario es leptocúrtico, propiedad que comparte con el modelo ARCH(1).

Cuando p=q=1 en la ecuación [1.5], m=1 y el modelo GARCH(1,1) puede interpretarse como un proceso ARMA(1,1) para la serie y_t^2, cuya función de autocorrelación será:

$$\rho_{y_t^2}(1) = (\alpha_1 + \beta_1) - \beta_1 \frac{\sigma_\upsilon^2}{\gamma_{y_t^2}(0)} \qquad [1.7a]$$

$$\rho_{y_t^2}(k) = (\alpha_1 + \beta_1)^{k-1} \rho_{y_t^2}(1), \quad k \geq 2 \qquad [1.7b]$$

donde σ_υ^2 es la varianza de la perturbación υ_t definida en [1.4] y $\gamma_{y_t^2}(0) = \text{Var}(y_t^2)$.

En las aplicaciones de modelos GARCH(1,1) a series financieras, es casi sistemática la obtención de un valor estimado de $\alpha_1+\beta_1$ prácticamente igual a uno, en especial si la frecuencia de observación es alta. Por ejemplo, los trabajos de Engle y Bollerslev (1986), Bollerslev (1987), Baillie y Bollerslev (1989) y Hsieh (1989) con series de tipos de cambio, Chou (1988), Baillie y DeGennaro (1990) y Poon y Taylor (1992) con índices de bolsa, y otros trabajos citados en Bollerslev, Chou y Kroner (1992) encuentran siempre valores de $\hat{\alpha}_1 + \hat{\beta}_1$ superiores a 0.9. Teniendo en cuenta la forma de la función de autocorrelación de y_t^2 en [1.7], un valor de $\alpha_1+\beta_1$ próximo a uno significa que dicha función apenas decrece, indicando que los cambios en la varianza condicional son relativamente lentos y, por tanto, los *shocks* en la volatilidad persisten. Esta propiedad es interesante porque refleja precisamente una de las características típicas de las series financieras: aunque la serie original está incorrelacionada, existe correlación en la serie de los cuadrados y, además, estas correlaciones decrecen lentamente, mostrando valores significativamente distintos de cero incluso para retardos altos.

Los resultados de los trabajos mencionados anteriormente justifican el interés de un modelo GARCH(1,1) en el que se imponga la condición $\alpha_1+\beta_1=1$. El modelo

resultante, denominado GARCH integrado o IGARCH, fue propuesto por Engle y Bollerslev (1986) y en él la ecuación para la varianza condicionada es:

$$\sigma_t^2 = \alpha_0 + \sigma_{t-1}^2 + \alpha_1(y_{t-1}^2 - \sigma_{t-1}^2) \qquad [1.8]$$

El modelo así definido es una martingala en diferencias, pero ya no es débilmente estacionario, porque su varianza marginal no es finita. Sin embargo, se prueba fácilmente que la ecuación [1.8] admite una representación de la forma:

$$\Delta y_t^2 = \alpha_0 + \upsilon_t - \beta_1 \upsilon_{t-1} \qquad [1.9]$$

donde Δ es el operador de primeras diferencias, tal que $\Delta x_t = x_t - x_{t-1}$, y υ_t es la perturbación definida en [1.4].

La ecuación [1.9] permite interpretar el modelo IGARCH(1,1) como un proceso MA(1) estacionario para las primeras diferencias de y_t^2, lo que indica cierta analogía del modelo IGARCH(1,1) con el proceso ARIMA(0,1,1). Sin embargo, existen diferencias significativas entre ellos. Por ejemplo, Nelson (1990) prueba que el IGARCH(1,1) con $\alpha_0>0$ es estrictamente estacionario y ergódico, a pesar de no ser débilmente estacionario.

Nelson (1991) observó ciertas limitaciones en los modelos GARCH: en primer lugar, las condiciones impuestas sobre los parámetros para asegurar que σ_t^2 no sea negativo son violadas en algunas aplicaciones empíricas; en segundo lugar, el modelo GARCH es incapaz de modelizar una respuesta asimétrica de la volatilidad ante las subidas y bajadas de los precios. Con el fin de solventar estas deficiencias, Nelson (1991) propuso un nuevo modelo que denominó modelo GARCH exponencial o EGARCH.

El modelo EGARCH garantiza la no negatividad de la varianza condicional formulando la ecuación de la volatilidad en términos del logaritmo de σ_t^2, mediante

una representación lineal del tipo:

$$\ln(\sigma_t^2) = \omega + \sum_{j=1}^{p} \varphi_j \ln(\sigma_{t-j}^2) + g(\varepsilon_{t-1}) + \sum_{i=1}^{q} \psi_i g(\varepsilon_{t-i-1}) \qquad [1.10]$$

donde $g(\varepsilon_t) = \lambda_1\varepsilon_t + \lambda_2(|\varepsilon_t| - E|\varepsilon_t|)$. A través de esta función g, que depende del signo y de la magnitud de ε_t, el modelo EGARCH puede capturar una respuesta asimétrica de la volatilidad ante innovaciones de distinto signo, permitiendo así modelizar un efecto contrastado empíricamente en muchas series financieras: las malas noticias (rendimientos negativos) provocan mayor aumento de la volatilidad que las buenas noticias (rendimientos positivos).

Por construcción, las perturbaciones $g(\varepsilon_t)$ son I.I.D. con media cero y varianza constante, y, por tanto, la expresión [1.10] puede considerarse como una representación ARMA para la serie $\ln(\sigma_t^2)$. Aplicando resultados estándar sobre procesos ARMA se obtienen entonces las condiciones de estacionariedad y las propiedades dinámicas de $\ln(\sigma_t^2)$.

1.1.1.2. Estimación

Todos los modelos tipo ARCH que acabamos de describir permiten construir fácilmente la función de verosimilitud utilizando la descomposición del error de predicción. Maximizando esta función se obtienen los estimadores máximo verosímiles. Como es habitual en modelos de series temporales, la función de verosimilitud se construye como el producto de las densidades condicionadas. Asumiendo que las perturbaciones ε_t en [1.1] son variables aleatorias independientes e idénticamente distribuidas con distribución N(0,1), de aquí en adelante NID(0,1), los modelos ARCH son condicionalmente gaussianos y la distribución condicionada de y_t/Ω_{t-1} es $N(0,\sigma_t^2)$. Por tanto, la expresión del logaritmo de la función de verosimilitud resulta ser:

$$\ln L(\theta) = \sum_{t=1}^{T} \ln f(y_t ; \theta / \Omega_{t-1}) = -\frac{T}{2} \ln 2\pi - \frac{1}{2} \sum_{t=1}^{T} \ln \sigma_t^2 - \frac{1}{2} \sum_{t=1}^{T} \frac{y_t^2}{\sigma_t^2}$$

donde θ es el vector de parámetros desconocidos del modelo y $f(y_t;\theta/\Omega_{t-1})$ denota la densidad condicionada de y_t dadas las observaciones previas hasta el instante t-1.

Bajo ciertas condiciones de regularidad se demuestra que, si los momentos de primer y segundo orden están correctamente especificados, los estimadores máximo verosímiles son consistentes y asintóticamente normales. Incluso si el modelo no es condicionalmente gaussiano pero se estima como si lo fuera, Weiss (1986) y Bollerslev y Wooldridge (1992) prueban, para modelos ARCH y GARCH, respectivamente, que los estimadores pseudo-máximo verosímiles (PMV) que se obtienen, son aún consistentes y asintóticamente normales. Lee y Hansen (1994) y Lumsdaine (1996) demuestran la consistencia y la normalidad asintótica del estimador PMV para el modelo GARCH(1,1) bajo condiciones más débiles, y Lumsdaine (1996) establece además resultados similares para el modelo integrado IGARCH(1,1). Por lo que respecta a los modelos EGARCH, aún no existe ningún resultado asintótico satisfactorio que nosotras conozcamos.

1.1.2. Modelos de volatilidad estocástica SV

Una forma alternativa de modelizar los cambios temporales en la volatilidad es a través de los modelos de volatilidad estocástica SV introducidos por Taylor (1986). En estos modelos, σ_t^2 no depende de las observaciones pasadas de la serie, sino de una variable no observable, que habitualmente es un proceso estocástico autorregresivo. Para garantizar la positividad de la varianza, la ecuación de la volatilidad se define para el logaritmo de σ_t^2, al igual que en el modelo EGARCH.

Los modelos SV encajan mejor con la teoría financiera y se generalizan bien al caso multivariante[6]. Sus propiedades dinámicas son fáciles de obtener e interpretar a partir de las del proceso estocástico subyacente, pero desafortunadamente, la estimación es más difícil que en los modelos ARCH, al no poderse construir fácilmente la función de verosimilitud de forma exacta. Esto conlleva la utilización de métodos de estimación como pseudo-máxima verosimilitud, máxima verosimilitud simulada, o el método generalizado de los momentos, entre otros.

Los modelos SV han recibido hasta el momento menos atención que los modelos tipo ARCH, por lo que la literatura al respecto no es tan abundante. Dos buenas revisiones son Taylor (1994) sobre los primeros trabajos con modelos SV, y la más actualizada de Ghysels, Harvey y Renault (1996), que abarca gran variedad de temas e incluye los resultados más recientes y una gran bibliografía. Ruiz (1994a) presenta un estudio comparativo de gran interés entre modelos ARCH y SV, ilustrado con la modelización del Indice Largo de la Bolsa de Madrid. Véase también Shephard (1996) que compara las propiedades de los modelos ARCH y SV e incluye una aplicación empírica de ambos modelos a distintas series financieras.

A continuación se describe con detalle el modelo SV autorregresivo de orden uno y se presentan algunas generalizaciones suyas. Se discuten además los distintos métodos de estimación propuestos en la literatura.

1.1.2.1. El modelo SV autorregresivo

Taylor (1986) propuso el primer modelo de volatilidad estocástica SV, denominado modelo SV log-normal o SV autorregresivo de orden uno. En este modelo, que denotaremos ARSV(1) de aquí en adelante, el logaritmo de σ_t^2 es un proceso AR(1) de la forma:

[6] Véase Harvey, Ruiz y Shephard (1994).

$$\ln(\sigma_t^2) = \gamma + \phi \ln(\sigma_{t-1}^2) + \eta_t \qquad [1.11]$$

donde η_t es NID$(0, \sigma_\eta^2)$ e independiente de la perturbación ε_t en [1.1]. Aunque la hipótesis de normalidad para η_t pueda parecer injustificada a primera vista, Andersen *et al.* (1999) muestran, utilizando datos de tipos de cambio recogidos con una frecuencia elevadísima, que la distribución empírica del logaritmo de la volatilidad se aproxima bien con una distribución normal.

Siguiendo una práctica habitual en la literatura sobre modelos SV, el modelo ARSV(1) dado por las ecuaciones [1.1] y [1.11] se puede reformular como:

$$y_t = \sigma_* \exp(\tfrac{1}{2} h_t)\varepsilon_t \qquad [1.12a]$$

$$h_t = \phi\, h_{t-1} + \eta_t \qquad [1.12b]$$

BIBLIOTECA VIRTUAL

donde σ_* es un factor de escala que evita la inclusión de un término constante en la ecuación [1.12b]. De hecho, $\sigma_*=\exp\{\tfrac{1}{2}\gamma/(1-\phi)\}$ y $h_t=\ln(\sigma_t^2)-[\gamma/(1-\phi)]$.

El modelo SV definido en [1.12] es una martingala en diferencias, ya que $E(y_t/\Omega_{t-1})=0$. En consecuencia, $E(y_t)=0$ y $Cov(y_t, y_s)=0$ para $t\neq s$. Las condiciones para que y_t sea estacionario son las necesarias para asegurar que lo sea el proceso h_t. En este caso, basta que $|\phi|<1$ para que y_t sea estacionario con varianza finita.

El cálculo de los momentos marginales en los modelos SV es relativamente sencillo. Las fórmulas que se presentan a continuación son muy generales y muchas de ellas son válidas para cualquier proceso estacionario h_t, no sólo para el proceso AR(1). Teniendo en cuenta la independencia de ε_t y η_t, de la ecuación [1.12a] se deduce que:

$$E(y_t^r) = \sigma_*^r E(\varepsilon_t^r) E\left\{\exp\left(\frac{r}{2} h_t\right)\right\} \qquad [1.13]$$

Por tanto, si la distribución de ε_t es simétrica, los momentos impares de y_t son cero. Respecto a los momentos de orden par, éstos se calculan utilizando la función generadora de momentos de la distribución logNormal, ya que si η_t es Normal, $\exp(h_t)$ es logNormal y, por tanto, $E[\exp(ah_t)]=\exp(a\mu_h + \frac{1}{2}a^2\sigma_h^2)$, donde μ_h y σ_h^2 denotan la esperanza y la varianza de h_t, respectivamente. En particular, los momentos de orden dos y orden cuatro de y_t son:

$$E(y_t^2) = \sigma_*^2 E(\varepsilon_t^2)E[\exp(h_t)] = \sigma_*^2 \exp(\mu_h + \tfrac{1}{2}\sigma_h^2) \qquad [1.14]$$

$$E(y_t^4) = \sigma_*^4 E(\varepsilon_t^4)E[\exp(2h_t)] = \sigma_*^4 E(\varepsilon_t^4)\exp[2(\mu_h + \sigma_h^2)] \qquad [1.15]$$

De la ecuación [1.15] se deduce que el momento de orden 4 de y_t existe y es finito siempre que lo sea el correspondiente momento de ε_t. Por ejemplo, si se supone que ε_t sigue una distribución N(0,1) se tiene que $E(\varepsilon_t^4) = 3$, y la curtosis de y_t es:

$$\kappa_y = \frac{E(y_t^4)}{[E(y_t^2)]^2} = 3\exp(\sigma_h^2) \qquad [1.16]$$

Para el modelo ARSV(1) en [1.12], h_t es un proceso AR(1) y, por tanto, $\sigma_h^2 = \dfrac{\sigma_\eta^2}{1-\phi^2}$.

En consecuencia, si $\sigma_\eta^2 > 0$, κ_y es mayor que 3, indicando que los modelos SV, al igual que los ARCH, tienen un comportamiento leptocúrtico, pero a diferencia de aquellos, en los modelos SV la estacionariedad de h_t es suficiente para asegurar la existencia del momento de orden cuatro. Además, en los modelos SV, el exceso de curtosis no está ligado necesariamente con la persistencia en la volatilidad, medida por el parámetro ϕ. De hecho, el valor de σ_η^2 puede ser lo suficientemente pequeño como para que un valor de ϕ próximo a uno no provoque exceso de kurtosis.

Dado que en la modelización de la volatilidad, el interés no está tanto en la serie original sino en el comportamiento dinámico de la serie de sus cuadrados,

veremos a continuación cuál es la estructura de correlación de y_t^2. Harvey (1993) establece las propiedades dinámicas de las series y_t^2 y $\ln(y_t^2)$ para el modelo ARSV(1). En esta sección sólo se presenta un resumen de aquellas que son relevantes para nuestro estudio.

Utilizando las expresiones [1.14] y [1.15] se comprueba de forma inmediata que la varianza de y_t^2 es:

$$\mathrm{Var}(y_t^2) = \sigma_*^4 \exp(2\mu_h + \sigma_h^2)[E(\varepsilon_t^4)\exp(\sigma_h^2) - 1] \qquad [1.17]$$

y teniendo en cuenta que:

$$E(y_t^2 y_{t+k}^2) = \sigma_*^4 E(\varepsilon_t^2 \varepsilon_{t+k}^2) E[\exp(h_t + h_{t+k})] = \sigma_*^4 \exp[2\mu_h + \sigma_h^2 + \gamma_h(k)]$$

se obtiene la siguiente expresión de las autocovarianzas de y_t^2:

$$\mathrm{Cov}(y_t^2, y_{t+k}^2) = \sigma_*^4 \exp(2\mu_h + \sigma_h^2)\{\exp[\gamma_h(k)] - 1\} \qquad [1.18]$$

donde $\gamma_h(k)$ es la función de autocovarianza de h_t. Dividiendo [1.18] entre [1.17] se obtiene la función de autocorrelación de y_t^2:

$$\rho_{Y_t^2}(k) = \rho(y_t^2, y_{t+k}^2) = \frac{\exp[\sigma_h^2 \rho_h(k)] - 1}{[E(\varepsilon_t^4)\exp(\sigma_h^2)] - 1}, \text{ para } k \geq 1 \qquad [1.19]$$

donde $\rho_h(k)$ es la función de autocorrelación de h_t.

Taylor (1986, p. 75) señala que cuando h_t es el proceso AR(1) de la ecuación [1.12b] y σ_h^2 es pequeño y/o ϕ está próximo a uno, la función de autocorrelación de y_t^2 en [1.19] es muy parecida a la de un ARMA(1,1), propiedad similar a la que se

tenía en el modelo GARCH(1,1). Harvey (1998) observa además que cuando σ_h^2 es pequeño y/o $\rho_h(k)$ toma valores próximos a uno, se tiene la siguiente aproximación, válida para cualquier modelo SV estacionario:

$$\rho_{Y_t^2}(k) \cong \rho_h(k)\frac{\exp(\sigma_h^2)-1}{[E(\varepsilon_t^4)\exp(\sigma_h^2)]-1}, \text{ para } k\geq 1$$

Por tanto, la función de autocorrelación de y_t^2 tendrá, aproximadamente, la forma de la función de autocorrelación de h_t, excepto por un factor de proporcionalidad que debe ser menor que uno, ya que $E(\varepsilon_t^4)$ es mayor que uno. Esto significa que en los modelos SV, la función de autocorrelación de la serie de los cuadrados tendrá las mismas características que la función de autocorrelación del proceso h_t pero tomará, en general, valores más pequeños que ésta. Como veremos en el capítulo 2, esto va a dificultar la utilización del correlograma de la serie y_t^2 como herramienta para la identificación del modelo.

Harvey, Ruiz y Shephard (1994) observan que las propiedades dinámicas del modelo SV aparecen de forma más clara en la serie $\ln(y_t^2)$. Aplicando esta transformación a la serie y_t en [1.12a], el modelo ARSV(1) transformado puede escribirse como:

$$x_t = \ln(y_t^2) = \mu + h_t + \xi_t \qquad [1.20a]$$

$$h_t = \phi\, h_{t-1} + \eta_t \qquad [1.20b]$$

donde μ es la media de x_t, $\mu=\ln(\sigma_*^2)+E[\ln(\varepsilon_t^2)]$, y $\xi_t= \ln(\varepsilon_t^2)-E[\ln(\varepsilon_t^2)]$ es un ruido blanco no gaussiano de media 0 y varianza σ_ξ^2, cuyas propiedades dependen de las de ε_t. Por ejemplo, si ε_t se distribuye como una N(0,1), $\ln(\varepsilon_t^2)$ se distribuye como el

logaritmo de una χ_1^2 y se prueba que $E[\ln(\varepsilon_t^2)]=-1.27$ y $\sigma_\xi^2=\pi^2/2\cong 4.93$.[7] Si ε_t tiene una distribución t de Student con ν grados de libertad, $\ln(\varepsilon_t^2)$ puede escribirse como:

$$\ln(\varepsilon_t^2) = \ln(\zeta_t^2) - \ln(\kappa_t)$$

donde ζ_t es una Normal estándar y $\nu\kappa_t$ se distribuye, independientemente de ζ_t, como una χ_ν^2. En este caso se prueba que $E[\ln(\varepsilon_t^2)]=-1.27-\psi(\nu/2)+\ln(\nu/2)$ y $\sigma_\xi^2=(\pi^2/2)+\psi'(\nu/2)$, donde ψ es la función psi de Euler y ψ' es su derivada.[8]

La ecuación [1.20a] representa x_t como la suma de dos procesos mutuamente independientes: un proceso lineal h_t estacionario y un ruido blanco no gaussiano. Esta descomposición permite obtener fácilmente las siguientes expresiones para la varianza, la función de autocovarianza y la función de autocorrelación de x_t:

$$\gamma_X(0) = \mathrm{Var}(x_t) = \sigma_h^2 + \sigma_\xi^2 \qquad [1.21]$$

$$\gamma_X(k) = \mathrm{Cov}(x_t, x_{t+k}) = \mathrm{Cov}(h_t + \xi_t, h_{t+k} + \xi_{t+k}) = \gamma_h(k), \qquad k\geq 1 \qquad [1.22]$$

$$\rho_X(k) = \frac{\gamma_X(k)}{\gamma_X(0)} = \frac{\gamma_h(k)}{\sigma_h^2 + \sigma_\xi^2} = \rho_h(k)\frac{1}{1+(\sigma_\xi^2/\sigma_h^2)}, \quad k\geq 1 \qquad [1.23]$$

donde σ_h^2, $\gamma_h(k)$ y $\rho_h(k)$ son la varianza, la autocovarianza y la autocorrelación de orden k del proceso h_t, respectivamente, y σ_ξ^2 es la varianza de $\ln(\varepsilon_t^2)$.

En la expresión [1.23] vemos que la forma de la función de autocorrelación de x_t es, salvo un factor de proporcionalidad, igual a la de la función de autocorrelación de h_t; por tanto, es el proceso h_t el que gobierna las propiedades dinámicas de x_t. Sin embargo, ese factor de proporcionalidad es típicamente muy

[7] Véase Abramowitz y Stegun (1970, p. 943).

[8] Véase Abramowitz y Stegun (1970, pp. 259-260, 943).

pequeño, debido a que la varianza del ruido, σ_ξ^2, suele ser mucho mayor que la varianza del proceso h_t. En consecuencia, puede ocurrir que las correlaciones de x_t estén muy próximas a cero a pesar de que el proceso h_t tenga una estructura de correlación muy fuerte.

Si h_t es el proceso AR(1) especificado en la ecuación [1.12b], se tiene que $\sigma_h^2 = \sigma_\eta^2/(1-\phi^2)$ y $\rho_h(k)=\phi^k$ para $k\geq 1$. Por tanto, la función de autocorrelación de x_t en [1.23] puede escribirse como:

$$\rho_X(1) = \phi\left\{1 + \frac{\sigma_\xi^2(1-\phi^2)}{\sigma_\eta^2}\right\}^{-1}$$

$$\rho_X(k) = \phi\rho_X(k-1) = \phi^{k-1}\rho_X(1), \quad \text{para } k \geq 2$$

Observamos así que la función de autocorrelación de x_t en el modelo ARSV(1) es equivalente a la de un proceso ARMA(1,1), que toma valores positivos y negativos dependiendo del signo de ϕ, y decrece exponencialmente a partir del segundo retardo.

1.1.2.2. Extensiones del modelo SV autorregresivo

El modelo SV especificado por las ecuaciones [1.12] se puede generalizar de forma inmediata haciendo que h_t sea un proceso ARMA(p,q) estacionario. En este caso, la serie y_t seguiría siendo estacionaria y sus propiedades se deducirían de forma análoga a partir de las propiedades de h_t. Los momentos de y_t^2 y $\ln(y_t^2)$ se podrían calcular utilizando las fórmulas [1.14]-[1.19] y [1.21]-[1.23], respectivamente, ya que éstas fórmulas son válidas para cualquier proceso estacionario h_t.

También es posible extender el modelo SV de forma que la ecuación que modeliza $\ln(\sigma_t^2)$ contenga una raíz unitaria. El principal motivo de interés de este modelo radica en que las aplicaciones empíricas del modelo ARSV(1) a series

financieras producen habitualmente un valor estimado de ϕ muy próximo a uno. Por ejemplo, Harvey, Ruiz y Shephard (1994) estiman un modelo ARSV(1) para cuatro series diarias de tipos de cambio y obtienen en los cuatro casos un valor $\hat{\phi} \geq 0.95$, Ruiz (1994a) obtiene $\hat{\phi} = 0.993$ en una aplicación al Indice Largo de la Bolsa de Madrid, y Sandmann y Koopman (1998) obtienen, con tres métodos de estimación diferentes, $\hat{\phi} \geq 0.96$ para una serie diaria del índice SP500 (*Standard and Poor's 500*) de la bolsa de Nueva York. Estos resultados son similares a los obtenidos en las aplicaciones de modelos GARCH(1,1) que, como vimos en la sección anterior, proporcionan habitualmente casi una raíz unitaria en la ecuación de la varianza condicionada.

El modelo SV con raíz unitaria más sencillo se obtiene suponiendo que $\ln(\sigma_t^2)$ es un paseo aleatorio, y, por tanto, las ecuaciones que definen el modelo son:

$$y_t = \sigma_t \varepsilon_t \qquad [1.24a]$$

$$\ln(\sigma_t^2) = \ln(\sigma_{t-1}^2) + \eta_t \qquad [1.24b]$$

Considerando la misma transformación logarítmica que en el caso estacionario, este modelo admite la siguiente representación:

$$\ln(y_t^2) = \mu + h_t + \xi_t \qquad [1.25a]$$

$$h_t = h_{t-1} + \eta_t \qquad [1.25b]$$

donde $\mu = E[\ln(\varepsilon_t^2)]$, $h_t = \ln(\sigma_t^2)$, y las perturbaciones ξ_t y η_t son como en el caso anterior.

El modelo SV con paseo aleatorio o modelo SV integrado, denotado RWSV (Random Walk SV), no es estacionario. Sin embargo, tomando primeras diferencias en [1.25], se obtiene un modelo estacionario para la serie $\Delta\ln(y_t^2)$ de la forma:

$$\Delta \ln(y_t^2) = \eta_t + \Delta \ln(\varepsilon_t^2)$$

A partir de esta expresión, se comprueba fácilmente que las propiedades dinámicas de $\Delta\ln(y_t^2)$ son las de un MA(1), ya que su función de autocovarianza es:

$$\text{Cov}(\Delta\ln(y_t^2), \Delta\ln(y_{t-1}^2)) = -\sigma_\xi^2$$

$$\text{Cov}(\Delta\ln(y_t^2), \Delta\ln(y_{t-k}^2)) = 0, \text{ si } k \geq 2$$

y, por tanto, $\ln(y_t^2)$ es un proceso ARIMA(0,1,1). Este resultado indica cierta analogía del modelo SV no estacionario con el modelo IGARCH(1,1), con la ventaja de que el modelo SV tiene un parámetro menos. Esto es especialmente importante en predicción. Recordemos que en el modelo IGARCH, la inclusión del término constante en la ecuación de la volatilidad es imprescindible para evitar que el proceso degenere. Sin embargo, como señala Ruiz (1994a), en el modelo SV este requisito no es necesario.

Por último, Harvey y Shephard (1996) prueban que si ε_t y η_t están correlacionados, el modelo SV puede modelizar la respuesta asimétrica de la volatilidad a los cambios de signo en la serie y_t, recogiendo así efectos similares a los del modelo EGARCH de Nelson (1991).

1.1.2.3. Estimación

La mayor dificultad en la estimación de los modelos SV radica en que, a diferencia de lo que ocurre con los modelos tipo ARCH, la distribución condicionada de y_t/Ω_{t-1} no está especificada explícitamente y, por tanto, no es inmediato construir la función de verosimilitud ni obtener los estimadores máximo verosímiles.

Para solventar este problema se han propuesto métodos de estimación alternativos, entre los que cabría mencionar los siguientes: el método de la pseudo-máxima verosimilitud (PMV) basado en el filtro de Kalman, propuesto independientemente por Nelson (1988) y Harvey, Ruiz y Shephard (1994); el método de los momentos generalizados (GMM) utilizado, por ejemplo, por Melino y Turnbull (1990) y Andersen y Sorensen (1996); y, finalmente, los métodos que tratan de evaluar la función de verosimilitud exacta mediante técnicas de simulación intensivas. Entre estos últimos destacan las propuestas de Jacquier, Polson y Rossi (1994) y Kim, Shephard y Chib (1998), que utilizan métodos de muestreo MCMC (Monte Carlo Markov Chain) para construir la verosimilitud, y los métodos de estimación máximo verosímil mediante simulaciones de Monte Carlo propuestos por Danielsson (1994) y Sandmann y Koopman (1998).

El método de estimación pseudo-máximo verosímil (PMV) propuesto por Nelson (1988) y Harvey, Ruiz y Shephard (1994), se basa en lo siguiente. Supongamos el modelo ARSV(1) dado por las ecuaciones [1.12] donde la perturbación ε_t es NID(0,1). El modelo lineal transformado dado por las ecuaciones [1.20] puede considerarse como un modelo estructural en el espacio de los estados donde [1.20a] es la ecuación de medida, [1.20b] es la ecuación de transición, η_t es un ruido blanco NID$(0,\sigma_\eta^2)$ y ξ_t es la perturbación no gaussiana con media 0 y varianza $\sigma_\xi^2=\pi^2/2$, independiente de η_t. Los parámetros de este modelo se pueden estimar por pseudo máxima verosimilitud a través del filtro de Kalman considerando ξ_t como si fuera N$(0,\pi^2/2)$. Las condiciones iniciales para el filtro de Kalman vienen dadas por la media y la varianza de la distribución marginal de h_t.[9]

El modelo SV también puede estimarse por pseudo-máxima verosimilitud sin asumir que ε_t sigue una distribución Normal, hipótesis que no parece muy razonable a juzgar por la excesiva curtosis que se observa empíricamente en muchas series

[9] Para una descripción del filtro de Kalman y de la estimación de modelos estructurales véase Harvey (1989).

financieras. En este caso la varianza de la perturbación ξ_t, σ_ξ^2, se estima como un parámetro más del modelo.[10]

Tanto si se asume para ε_t una distribución Normal como, por ejemplo, una t de Student, Ruiz (1994b) prueba que los estimadores pseudo-máximo verosímiles son consistentes y asintóticamente normales. Sin embargo, el método PMV no es eficiente, y en muestras pequeñas, los estimadores no tienen buenas propiedades, debido a la mala aproximación de la distribución de $\ln(\varepsilon_t^2)$ a una normal. Breidt y Carriquiry (1996) han sugerido una modificación de la transformación $\ln(y_t^2)$ que parece mejorar este comportamiento, pero los estimadores siguen siendo ineficientes.

Frente a este inconveniente, el método PMV tiene la ventaja de poder extenderse, sin ninguna modificación, a casos más generales en los que, por ejemplo, h_t sea un proceso más complicado, o incluso se consideren modelos multivariantes, como en Harvey, Ruiz y Shephard (1994). En particular, el método de estimación PMV puede aplicarse al modelo SV con paseo aleatorio. La única diferencia es que en este caso, en lugar de utilizar la distribución marginal de h_t para inicializar el filtro de Kalman, se utiliza la primera observación disponible. Ruiz (1994b) prueba que el estimador PMV de σ_η^2, el único parámetro del modelo integrado, es asintóticamente normal, y deriva la expresión de su varianza asintótica.

El método de los momentos generalizados (GMM) utilizado, por ejemplo, por Melino y Turnbull (1990) y Andersen y Sorensen (1996), se basa en la convergencia de los momentos muestrales a los momentos poblacionales. En el modelo ARSV(1), esta convergencia está garantizada por las propiedades de ergodicidad y estacionariedad de y_t. Este método es fácil de implementar pero requiere tamaños muestrales grandes y resulta ineficiente cuando el parámetro ϕ de la ecuación [1.12b]

---

[10] Si se supone que ε_t sigue una distribución t_ν de Student, la estimación de ν se obtiene igualando a $(\pi^2/2)+\psi'(\nu/2)$ el valor estimado de σ_ξ^2. Ruiz (1994b) prueba que de esta forma se obtienen estimadores más eficientes que fijando de antemano el valor de ν.

está próximo a uno, algo que, como ya hemos comentado anteriormente, es muy habitual en las aplicaciones empíricas. En muestras pequeñas la eficiencia de los estimadores PMV y GMM es similar, pero PMV domina cuando $\phi \approx 1$.

Jacquier, Polson y Rossi (1994) proponen un enfoque bayesiano utilizando el método MCMC (Monte Carlo Markov Chain). Mediante un intenso ejercicio de simulación, estos autores prueban que, cuando la varianza del proceso subyacente h_t es pequeña en relación con la varianza de la serie original, su método proporciona, en general, estimadores con menor sesgo y error cuadrático medio que los estimadores GMM y PMV. Sin embargo, el método requiere fuertes supuestos en la especificación del modelo y conlleva mucho tiempo en los cálculos computacionales. Además el método necesita modificarse para incorporar generalizaciones como el modelo multivariante o modelos con variables explicativas. Un algoritmo MCMC computacionalmente más eficiente que el de Jacquier, Polson y Rossi (1994) puede encontrarse en Shephard y Pitt (1997).

Kim, Shephard y Chib (1998) proponen aproximar la distribución de la perturbación $\ln(\varepsilon_t^2)$ con una mixtura de normales, y a partir de ésta construir la función de verosimilitud utilizando la descomposición en los errores de predicción. Los estimadores máximo verosímiles se obtienen mediante un algoritmo de simulación que incorpora técnicas MCMC para muestrear el logaritmo de la volatilidad de una manera más eficiente. Las simulaciones presentadas en el trabajo de Kim, Shephard y Chib (1998) muestran que el método propuesto es significativamente más eficiente que otros métodos de estimación de modelos SV propuestos anteriormente. Sin embargo, uno de los inconvenientes de este método es que la tasa de convergencia del algoritmo es muy sensible al número y a los pesos de las componentes de la mixtura. Además, parece que ninguna mixtura de normales puede aproximar correctamente el comportamiento de las colas en la distribución de $\ln(\varepsilon_t^2)$.

Finalmente, Sandmann y Koopman (1998) proponen un método de estimación máximo verosímil mediante simulaciones de Monte Carlo basado en la propuesta de Danielsson (1994) pero aplicada al modelo ARSV(1) transformado en [1.20]. En este caso, la función de verosimilitud se descompone en dos partes: una parte que representa la verosimilitud gaussiana y una función "resto" debida a la no normalidad de $\ln(\varepsilon_t^2)$, cuyo valor se aproxima mediante simulaciones de Monte Carlo. En muestras pequeñas, el método de Sandmann y Koopman (1998) proporciona estimadores con propiedades similares, en términos de sesgo y error cuadrático medio, a los estimadores MCMC, con la ventaja de ser mucho más rápido. Además, este método comparte con el método PMV las ventajas de utilizar la representación en el espacio de los estados del modelo SV, y, por tanto, puede extenderse a modelos más generales sin ser modificado. Sin embargo, aún no existe ningún resultado sobre las propiedades asintóticas de estos estimadores.

1.2. Modelos con memoria larga para la media: procesos ARFIMA

La propiedad de memoria larga suele relacionarse, desde un punto de vista empírico, con la persistencia que muestran las autocorrelaciones muestrales de ciertas series temporales estacionarias, que decrecen a un ritmo muy lento pero finalmente convergen hacia cero. Este comportamiento no es compatible ni con los modelos estacionarios autorregresivos y de medias móviles (ARMA), que imponen un decrecimiento exponencial en las autocorrelaciones, ni con el grado extremo de persistencia de los modelos integrados no estacionarios.

Formalmente, siguiendo la notación de Robinson (1994b) y Baillie (1996), diremos que una serie x_t estacionaria con función de autocorrelación $\rho(k)$ y densidad espectral $f(\lambda)$, tiene memoria larga si: (a) las autocorrelaciones no son absolutamente

sumables, es decir $\Sigma|\rho(k)|=\infty$, o equivalentemente, (b) la función de densidad espectral $f(\lambda)$ no está acotada en las bajas frecuencias, y por tanto $\lim_{\lambda\to 0^+} f(\lambda) = +\infty$.

Además de esta definición, existen otras posibles caracterizaciones de la propiedad de memoria larga. Por ejemplo, se dice que un proceso estacionario x_t tiene memoria larga si los coeficientes de su correspondiente representación de Wold, $x_t=\sum_{k=0}^{\infty}\psi_k\eta_{t-k}$, siendo η_t ruido blanco, verifican la condición $\sum_{k=0}^{\infty}\psi_k^2<\infty$, pero no son absolutamente sumables ($\sum_{k=0}^{\infty}|\psi_k|=\infty$). Otros autores, como Brockwell y Davies (1991) y Beran (1994), por ejemplo, consideran que un proceso estacionario con función de autocorrelación $\rho(k)$ y densidad espectral $f(\lambda)$, tiene memoria larga si existe un número real $0<\alpha<1$ y una constante $C_\rho>0$, tales que $\rho(k)\sim C_\rho|k|^{-\alpha}$ cuando $k\to\infty$, o equivalentemente, si existe un número real $0<\beta<1$ y una constante $C_f>0$ tales que $f(\lambda)\sim C_f|\lambda|^{-\beta}$ cuando $\lambda\to 0^+$, donde el símbolo $\sim$ significa que el cociente de los términos a su derecha e izquierda converge hacia uno. Esta definición es más restrictiva que la de Robinson (1994b) y Baillie (1996), pero para la clase de procesos que estudiaremos en esta sección es suficiente.

Como apunta Beran (1994), es importante señalar que estas definiciones son definiciones asintóticas: sólo nos informan del comportamiento último de las correlaciones cuando $k\to\infty$ (o del espectro cuando $\lambda\to 0^+$). La propiedad de memoria larga sólo determina la tasa de convergencia hacia cero de las correlaciones, pero no la magnitud de cada una de ellas. Cada correlación individual puede ser arbitrariamente pequeña, pero el decrecimiento debe ser lento.

La evidencia empírica sobre datos con memoria larga se remonta a mucho tiempo atrás; quizás el ejemplo más conocido sea el trabajo de Hurst (1951) en el campo de la hidrología. Sin embargo, muchos de los trabajos aplicados a otras ciencias y, sobre todo, muchos de los desarrollos formales sobre estimación y contrastes de hipótesis, son relativamente recientes. En concreto, el interés por los

modelos con memoria larga para la modelización de series económicas surge a partir de los trabajos de Granger (1980) y Granger y Joyeux (1980).

Granger y Joyeux (1980) advierten que la práctica habitual de diferenciar una serie hasta conseguir estacionariedad puede tener consecuencias negativas en la correcta modelización de algunas series. Por ejemplo, muchas series económicas aparentemente no estacionarias con espectro no acotado en el origen suelen diferenciarse para conseguir varianza finita. Sin embargo, la serie diferenciada se convierte a menudo en una serie con espectro nulo en el cero, indicando que se ha eliminado de la serie original la componente de las bajas frecuencias, tan importante en las predicciones a largo plazo. Puede pensarse entonces que para modelizar este tipo de series, la diferenciación es "excesiva" (sobrediferenciación) pero la no diferenciación (no estacionariedad) tampoco es adecuada.

Para cubrir este vacío entre los casos extremos de modelos con raíces unitarias, típicamente utilizados para series no estacionarias, y modelos estacionarios que imponen un decrecimiento exponencial de las autocorrelaciones, y por tanto un espectro acotado en la frecuencia cero, Granger (1980), Granger y Joyeux (1980) y Hosking (1981) proponen una clase de procesos intermedios en los que el orden de integración sea fraccional. Se trata de los procesos autorregresivos y de medias móviles fraccionalmente integrados ARFIMA(p,d,q), donde d es un número real. Al permitir que el orden de integración d sea un número no entero, estos modelos actúan como un "puente" entre los procesos con raíces unitarias ARIMA (d=1) y los procesos ARMA estacionarios que no requieren integración (d=0).

Como veremos más adelante, los procesos ARFIMA producen memoria larga si $0<d<1/2$, en el sentido de que para ese rango de valores se verifican las condiciones (a) y (b) especificadas anteriormente. Además, estos modelos proporcionan una gran flexibilidad en la interpretación de la persistencia, entendida ésta cómo el efecto que tiene un cambio unitario en la perturbación del modelo sobre las predicciones del nivel futuro de la serie. Mientras en los procesos integrados (d=1), el efecto de un *shock* persiste indefinidamente, en un proceso fraccionalmente

integrado con $0<d<1$, el efecto de un *shock* acaba desapareciendo y la serie revierte finalmente a su media, incluso en el intervalo $1/2\leq d<1$, donde el proceso es no estacionario.

Además de los modelos ARFIMA existen otros modelos paramétricos capaces de producir memoria larga, como son el ruido gaussiano fraccional de Mandelbrot y Van Ness (1968), que se obtiene como los incrementos de procesos autosimilares, o el modelo exponencial fraccional (FEXP) propuesto por Bloomfield (1973). Una descripción de éstos y otros ejemplos de procesos con memoria larga puede encontrarse en el capítulo 2 de Beran (1994).

Entre la extensa literatura existente en estos momentos sobre procesos con memoria larga, tanto a nivel de resultados teóricos como de aplicaciones empíricas, cabría destacar las revisiones de Robinson (1994b) y Baillie (1996), y el libro ya mencionado de Beran (1994). En ellos se revisan las propiedades más importantes de los procesos con memoria larga, en especial de los ARFIMA, y se discuten los métodos de estimación y contrastes de hipótesis que de ellos se derivan. Se incluyen asimismo referencias de trabajos aplicados, en particular de aquellos relativos a la modelización de series macroeconómicas, como Diebold y Rudebush (1989), Sowell (1992b) o Delgado y Robinson (1993), y aquellos aplicados a la modelización de series financieras, como Lo (1991), Cheung (1993a) o Baillie y Bollerslev (1994), entre otros.

En esta sección se describen las propiedades de los modelos ARFIMA, así como los métodos de estimación y los contrastes de hipótesis propuestos en la literatura. En la sección 1.2.1 se presentan estos modelos, poniendo especial énfasis en aquellas propiedades que caracterizan a las series con memoria larga. En la sección 1.2.2 se presentan los métodos de estimación más utilizados en aplicaciones empíricas y aquellos para los que existen resultados teóricos rigurosos. En la sección 1.2.3 se describen brevemente varios métodos para contrastar la hipótesis de memoria corta frente a alternativas de memoria larga así como posibles contrastes de hipótesis para discernir entre la presencia de raíces unitarias y la integración

fraccional.

1.2.1. Definición y propiedades

Siguiendo la definición de Robinson (1994b), diremos que un proceso x_t es integrado de orden d, y lo denotaremos por $x_t \sim I(d)$, si verifica que:

$$(1-L)^d x_t = u_t \qquad [1.26]$$

donde L es el operador retardo, $L^i x_t = x_{t-i}$, y u_t es un proceso estacionario y con densidad espectral acotada en la frecuencia cero. Si d=0, x_t es un proceso estacionario I(0); si d=1, x_t tiene una raíz unitaria y decimos que es un proceso I(1); si d es otro número entero, x_t tiene d raíces unitarias; y si d es un número real no entero, decimos que x_t es un proceso fraccionalmente integrado. Si además el proceso u_t es un ARMA(p,q) de la forma $\phi(L)u_t = \theta(L)\eta_t$, donde η_t es un ruido blanco de media cero y varianza σ_η^2, decimos que x_t es un ARMA fraccionalmente integrado o ARFIMA(p,d,q).

El más sencillo de los procesos fraccionalmente integrados es el ARFIMA(0,d,0) o ruido blanco fraccionalmente integrado, en el que el proceso u_t en [1.26] es ruido blanco ($u_t = \eta_t$). En este caso, la serie x_t se modeliza según la ecuación:

$$(1-L)^d x_t = \eta_t \qquad [1.27]$$

donde d es un número no entero y el polinomio $(1-L)^d$ viene definido, para cualquier d>-1, por la serie binomial:

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k = 1 - dL - \frac{d(1-d)}{2!} L^2 - \frac{d(1-d)(2-d)}{3!} L^3 - \ldots \qquad [1.28]$$

Para los valores de d en el intervalo $-1/2<d<1/2$, el proceso ARFIMA(0,d,0) es estacionario e invertible. En concreto, Hosking (1981) demuestra que si $d>-1/2$, x_t es invertible y tiene una representación AR(∞) de la forma:

$$(1-L)^d x_t = \sum_{k=0}^{\infty} \pi_k x_{t-k} = \eta_t \qquad [1.29]$$

donde π_k son los coeficientes del polinomio $(1-L)^d$ en la expresión [1.28],

$$\pi_k = -\frac{d(1-d)...(k-1-d)}{k!} = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

donde Γ denota la función gamma[11]. Cuando $d<1/2$, Hosking (1981) demuestra que x_t es estacionario y tiene una representación MA(∞) de la forma:

$$x_t = (1-L)^{-d} \eta_t = \sum_{k=0}^{\infty} \psi_k \eta_{t-k}, \text{ con } \psi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} \qquad [1.30]$$

Utilizando la aproximación $[\Gamma(a+x)/\Gamma(b+x)]\sim x^{a-b}$, se comprueba que $\psi_k \sim \frac{k^{d-1}}{\Gamma(d)}$ y, por tanto, los coeficientes de la representación de Wold del proceso ARFIMA(0,d,0) convergen hiperbólicamente hacia cero, en contraposición con el decrecimiento exponencial típico de los procesos ARMA.

La propiedad más característica de los procesos ARFIMA estacionarios ($d<1/2$) es el comportamiento asintótico de su función de autocorrelación, o equivalentemente, el de su densidad espectral en las bajas frecuencias.

Según la ecuación [1.30], la serie x_t puede interpretarse como el proceso resultante de un sistema lineal donde el proceso de entrada, η_t, es un ruido blanco cuya función de densidad espectral es $f_\eta(\lambda)=\sigma_\eta^2/2\pi$. Por tanto, la densidad espectral de x_t puede obtenerse a partir de la de η_t de la siguiente forma:[12]

$$f(\lambda) = f_\eta(\lambda)\,|1-e^{i\lambda}\,|^{-2d} = \frac{\sigma_\eta^2}{2\pi}[2(1-\cos\lambda)]^{-d} = \frac{\sigma_\eta^2}{2\pi}[2\,\mathrm{sen}(\lambda/2)]^{-2d} \qquad [1.31]$$

y cuando $\lambda \to 0+$ se tiene la siguiente aproximación:

$$f(\lambda) \sim \frac{\sigma_\eta^2}{2\pi}\lambda^{-2d} \qquad [1.32]$$

De esta expresión se deduce que para valores $0<d<1/2$, la densidad espectral de x_t es una función decreciente de λ no acotada en el origen, y por tanto el proceso tiene memoria larga, pero si $d<0$, la densidad espectral se anula en el origen y está dominada por las frecuencias altas. En este último caso, se dice que el proceso ARFIMA presenta dependencia negativa o antipersistencia.

Con respecto al comportamiento de las autocorrelaciones, Hosking (1981) deriva las siguientes expresiones para la función de autocovarianza y la varianza del proceso ARFIMA(0,d,0), respectivamente:

$$\gamma(k) = \sigma_\eta^2 \frac{\Gamma(1-2d)\Gamma(k+d)}{\Gamma(d)\Gamma(1-d)\Gamma(k+1-d)}, \text{ si } k\geq 1 \qquad [1.33]$$

$$\gamma(0) = \sigma_\eta^2 \frac{\Gamma(1-2d)}{[\Gamma(1-d)]^2} \qquad [1.34]$$

---

[11] Véase, por ejemplo, Brockwell y Davies (1991, p. 521) para una definición de la función gamma.

[12] Véase Priestley (1981, pp.263-270) para un desarrollo formal sobre análisis espectral de sistemas lineales.

Dividiendo [1.33] entre [1.34] se obtiene la función de autocorrelación:

$$\rho(k) = \frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)} = \prod_{1\le i\le k} \frac{i+d-1}{i-d}, \quad k\ge 1 \qquad [1.35]$$

y aplicando nuevamente la aproximación $[\Gamma(a+x)/\Gamma(b+x)]\sim x^{a-b}$, resulta que:

$$\rho(k) \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}, \text{ cuando } k\to\infty \qquad [1.36]$$

Esta última expresión indica que la función de autocorrelación del proceso ARFIMA(0,d,0) decrece a un ritmo hiperbólico lento, a diferencia de los procesos ARMA, cuya función de autocorrelación decrece exponencialmente.

Esta dependencia a largo plazo, típica de los procesos ARFIMA, se aprecia claramente en los valores de la tabla 1.1, que compara las autocorrelaciones de un proceso ARFIMA(0,1/3,0) con las de un proceso AR(1) de la forma $x_t=\phi x_{t-1}+\eta_t$, con $\phi=0.5$. Aunque en ambos casos la autocorrelación de orden uno es igual a 0.5, existe una gran diferencia en el ritmo al que las correlaciones convergen hacia cero: mientras en el proceso AR(1) la autocorrelación de orden 25 es prácticamente cero, en el proceso fraccionalmente integrado es 0.173 y sólo se reduce a 0.109 en el retardo 100.

Tabla 1.1. Comparación de las funciones de autocorrelación de un proceso ARFIMA, $(1-L)^{1/3}x_t=\eta_t$, y un proceso AR(1), $x_t=0.5x_{t-1}+\eta_t$.

Retardo (k)	ρ(k) [ARFIMA(0,1/3,0)]	ρ(k) [AR(1), con ϕ=0.5]
1	0.500	0.500
2	0.400	0.250
3	0.350	0.125
4	0.318	0.063

5	0.295	0.031
10	0.235	0.001
25	0.173	$2.98x10^{-8}$
50	0.137	$8.88x10^{-16}$
100	0.109	$7.89x10^{-31}$

Resumiendo, podríamos decir que las propiedades más importantes del proceso ARFIMA(0,d,0) son las siguientes. En primer lugar, si $0<d<1/2$, la densidad espectral es una función decreciente de λ, no acotada en el origen, y concentrada en las bajas frecuencias, y las autocorrelaciones son todas positivas y decaen lentamente, a un ritmo hiperbólico k^{2d-1}, y no son sumables. Estas propiedades son precisamente las que caracterizan a los modelos con memoria larga, y por tanto podemos concluir que en el intervalo $0<d<1/2$, el proceso ARFIMA(0,d,0) es un proceso estacionario con memoria larga. Por el contrario, cuando $-1/2<d<0$, x_t es un proceso estacionario con memoria corta ya que la densidad espectral se anula en el origen y está dominada por las frecuencias altas, y las autocorrelaciones son todas negativas y absolutamente sumables. Aunque el rango $-1/2<d<0$ no tiene mucho interés en sí mismo, puede resultar muy útil en la modelización de las primeras diferencias de modelos ARFIMA(0,d,0) no estacionarios con orden de integración $1/2<d<1$. Como veremos más adelante, los procesos fraccionalmente integrados con $1/2<d<1$ permiten modelizar el comportamiento de series no estacionarias pero que finalmente revierten hacia la media, propiedad que no tienen los procesos con raíces unitarias.

El proceso ARFIMA(0,d,0) que acabamos de describir se puede generalizar de forma inmediata haciendo que el proceso u_t en [1.26] sea un proceso autorregresivo y de medias móviles ARMA(p,q) de la forma $\phi(L)u_t=\theta(L)\eta_t$, en lugar de un ruido blanco. En este caso diremos que x_t es un proceso ARMA fraccionalmente integrado o ARFIMA(p,d,q). Este proceso admite una representación de la forma:

$$\phi(L)(1-L)^d x_t=\theta(L)\eta_t \qquad [1.37]$$

donde $\phi(L)$ y $\theta(L)$ son los polinomios de retardos de orden p y q, respectivamente, del proceso ARMA, cuyas raíces están fuera del círculo unidad, y el parámetro d y la perturbación η_t son como en el caso anterior. El proceso ARFIMA(p,d,q) en [1.37] proporciona una gran flexibilidad en la modelización del comportamiento dinámico de una serie temporal, ya que permite describir simultáneamente las propiedades dinámicas en el largo plazo, a través del parámetro d, y la correlación en el corto plazo, a través de los parámetros de la parte ARMA del modelo.

El comportamiento a largo plazo del proceso ARFIMA(p,d,q) es muy similar al del proceso ARFIMA(0,d,0), ya que para observaciones muy distantes, los efectos de los parámetros ARMA son prácticamente despreciables. En concreto, Hosking (1981) demuestra que si $d<1/2$ y todas las raíces de $\phi(z)=0$ están fuera del círculo unidad, x_t es estacionario, y si $d>-1/2$ y todas las raíces de $\theta(z)=0$ están fuera del círculo unidad, x_t es invertible.

En el caso estacionario ($d<1/2$), se prueba que la función de densidad espectral del proceso ARFIMA(p,d,q) definido en [1.37] es de la forma:

$$f(\lambda)=\frac{\sigma_\eta^2}{2\pi}\left|\frac{\theta(e^{i\lambda})}{\phi(e^{i\lambda})}\right|^2\left[2\,\mathrm{sen}(\lambda/2)\right]^{-2d} \qquad [1.38]$$

y por tanto, cuando $\lambda\to 0^+$ resulta que $f(\lambda)\sim\frac{\sigma_\eta^2}{2\pi}\left|\frac{\theta(1)}{\phi(1)}\right|^2\lambda^{-2d}$. En consecuencia, si $0<d<1/2$, el espectro de x_t no está acotado en las frecuencias bajas y decimos que el proceso tiene memoria larga. Esta propiedad también se manifiesta en el comportamiento de las autocorrelaciones. Hosking (1981) demuestra que si $d<1/2$, existe una constante positiva C_ρ tal que $\rho(k)\sim C_\rho k^{2d-1}$ y por tanto, si $0<d<1/2$, las autocorrelaciones decrecen hiperbólicamente y no son sumables. Las fórmulas

exactas de las autocorrelaciones de un proceso ARFIMA(p,d,q) que deriva Hosking (1981) son bastante complicadas, pero su cálculo puede simplificarse considerablemente utilizando la formulación de Sowell (1992a) y Chung (1994).

Una de las propiedades que distingue a los modelos ARFIMA de los modelos ARMA de memoria corta y de los modelos integrados ARIMA es la forma en que la serie responde a un cambio unitario en la perturbación del modelo. Mientras en un modelo ARIMA, el efecto de un *shock* sobre el nivel futuro de la serie persiste para siempre y no hay reversión a la media, en un ARFIMA el efecto finalmente desaparece, aunque a un ritmo más lento que el exponencial de los procesos ARMA.

Para medir el grado de persistencia de los *shocks* sobre la media, Diebold y Rudebusch (1989) y Haubrich y Lo (1989) proponen utilizar la función respuesta al impulso acumulativa[13]. Dada una serie x_t, los valores de esta función son los coeficientes del polinomio $c(L)=1+c_1L+c_2L^2+\ldots$ tal que $x_t=\zeta+c(L)\eta_t$. Así, el coeficiente c_k medirá el impacto que tiene una innovación unitaria en t sobre el nivel de la serie en t+k, y su límite, $c_\infty=\lim_{k\to\infty} c_k$, medirá el impacto a largo plazo de dicha innovación. Cuando la serie x_t en primeras diferencias admite una representación de la forma:

$$(1-L)x_t=a(L)\eta_t$$

el valor límite de la función respuesta al impulso acumulativa coincide con la suma de los coeficientes del polinomio a(L). En concreto, si $a(L)=1+a_1L+a_2L^2+\ldots$, se comprueba fácilmente que $c_k=1+a_1+\cdots+a_k$, y por tanto el impacto a largo plazo que tiene una innovación unitaria sobre el nivel de la serie será:

$$c_\infty = \lim_{k\to\infty} c_k = \lim_{k\to\infty} (1 + a_1 + \ldots + a_k) = a(1)$$

donde a(1) es la suma de los coeficientes del polinomio a(L).

En el modelo ARFIMA(p,d,q), la forma del polinomio a(L) se deduce inmediatamente de la ecuación [1.37], obteniéndose:

$$a(L)=(1-L)^{1-d}\,\phi(L)^{-1}\,\theta(L)$$

donde $(1-L)^{1-d}$ admite una expansión binómica como la de la expresión [1.28]. Utilizando los resultados de Gradshteyn y Ryzhik (1980, sec. 9.1) resulta que si $0<d\leq 1$, $(1-L)^{1-d}=F(d-1,1,1;L)$, donde F es la función hipergeométrica. Esta función toma el valor $F(d-1,1,1;1)=0$, si $d<1$, y $F(d-1,1,1;1)=1$ si $d=1$, y por tanto,

$$a(1) = F(d-1,1,1;1)\phi(1)^{-1}\theta(1) = \begin{cases} 0 & \text{si} \quad 0 < d < 1 \\ \phi(1)^{-1}\theta(1) & \text{si} \quad d = 1 \end{cases}$$

donde $\phi(1)$ y $\theta(1)$ son las sumas de los coeficientes de los polinomios $\phi(L)$ y $\theta(L)$, respectivamente.

Esto significa que en los procesos integrados (d=1), el efecto de un *shock* persiste indefinidamente, ya que $a(1)\neq 0$, y no hay reversión a la media, mientras que en un proceso fraccionalmente integrado I(d) con $0<d<1$, el efecto de un *shock* acaba desapareciendo ($a(1)=0$) y la serie x_t revierte finalmente a su media, ζ, aunque lo hará a un ritmo muy lento caracterizado, precisamente, por el valor del parámetro d. Obsérvese que esto es cierto no sólo en el caso estacionario ($0<d<1/2$), sino también en el intervalo $1/2\leq d<1$, donde el proceso no es estacionario. En este sentido suele decirse que el ARFIMA no estacionario es "menos no estacionario" que los procesos con raíces unitarias.

---

[13] Véase Bloomfield (1976, p.124) para una definición de función respuesta al impulso.

Para terminar esta pequeña revisión de los conceptos básicos sobre modelos fraccionalmente integrados, haremos referencia a algunas de las razones que pueden explicar la potencial relevancia de estos modelos en la modelización de series económicas. En primer lugar cabría recordar que las propiedades del espectro de los procesos ARFIMA estacionarios que acabamos de describir (no acotado en las frecuencias bajas y decreciendo hiperbólicamente hacia cero), coinciden con lo que Granger (1966) denominó la forma "típica" del espectro de las series económicas. A esto hay que añadir los resultados sobre agregación de series de Robinson (1978) y Granger (1980, 1990), que demuestran que la suma de procesos AR(1) independientes en los que el coeficiente autorregresivo es aleatorio, con una distribución, por ejemplo, tipo Beta[14], es un proceso fraccionalmente integrado. Si tenemos en cuenta que muchas variables económicas importantes son agregados de otras variables, este resultado podría explicar la presencia de memoria larga en ciertas series económicas.[15]

1.2.2. Estimación

Los métodos de estimación de modelos con memoria larga propuestos hasta el momento en la literatura pueden clasificarse en dos grandes grupos: métodos paramétricos, en los que se conoce de forma explícita toda la estructura de correlación, tanto en el corto como en el largo plazo, o se conoce la densidad espectral en todas las frecuencias, y métodos semiparamétricos, en los que no se especifica un modelo completo de la estructura de autocovarianzas (o del espectro) sino sólo su comportamiento en los retardos alejados (o en las frecuencias bajas).[16]

[14] Para una descripción completa de la distribución Beta, véase Johnson y Kotz (1970)

[15] Haubrich y Lo (1989) aplican estos resultados a la modelización de series macroeconómicas.

[16] Otros autores distinguen entre métodos de estimación en dos etapas, en los que se estima en primer lugar el parámetro d y en una segunda etapa los parámetros ARMA del modelo, y métodos de estimación en una etapa, que estiman conjuntamente, por máxima verosimilitud, todos los parámetros del modelo.

El enfoque paramétrico tiene la ventaja de que permite caracterizar todas las autocorrelaciones y no sólo su decrecimiento hiperbólico. Además, los métodos propuestos se basan en maximizar la función de verosimilitud gaussiana, y los estimadores máximo verosímiles (MV) resultantes tienen buenas propiedades: son consistentes y asintóticamente normales, y bajo hipótesis de normalidad, asintóticamente eficientes. Sin embargo, como señala Robinson (1995b), si el modelo paramétrico que se supone para $\rho(k)$ o $f(\lambda)$ está mal especificado, los estimadores serán inconsistentes y los métodos semiparamétricos pueden hacer un "papel mejor", ya que son más robustos a una mala especificación del modelo. Además, los estimadores semiparamétricos también van a tener buenas propiedades asintóticas, aunque su tasa de convergencia será inferior a la $\sqrt{T}$ típica de los estimadores paramétricos gaussianos.

El cuadro 1.1 resume los principales métodos de estimación que revisaremos en esta sección. Además de estos métodos existen otros que son menos utilizados en aplicaciones empíricas. Li y McLeod (1986), por ejemplo, analizan las propiedades de los estimadores máximo verosímiles del modelo aproximado que se obtendría truncando el polinomio $(1-L)^d$ en la correspondiente representación AR(∞) del modelo ARFIMA dada en [1.29]. Chung y Baillie (1993) analizan las propiedades en muestras finitas de los estimadores que resultan de minimizar una suma de cuadrados condicional en el dominio del tiempo. Una breve descripción de ambos métodos puede encontrarse en Beran (1994) y Baillie (1996), respectivamente.

Cuadro 1.1. Métodos de estimación para modelos con memoria larga

Métodos paramétricos	• Máxima verosimilitud exacta en el dominio temporal • Aproximación de Whittle a la máxima verosimilitud en el dominio de las frecuencias
Métodos semiparamétricos	• Regresión mínimo-cuadrática en el dominio de las

	frecuencias para el logaritmo del periodograma • Estimador gaussiano semiparamétrico en el dominio de las frecuencias

Los métodos de estimación paramétricos que revisaremos a continuación plantean la estimación conjunta de los parámetros del modelo maximizando la función de verosimilitud gaussiana o alguna aproximación suya.

Yajima (1985) obtuvo los primeros resultados para el más sencillo de los modelos fraccionalmente integrados, el proceso ARFIMA(0,d,0) definido en [1.27] con perturbación η_t gaussiana. Este autor demostró que el estimador del vector paramétrico $\theta=(\sigma_\eta^2,d)$ que se obtiene maximizando la función de verosimilitud exacta,

$$L(\theta;\mathbf{x}) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma(\theta)| - \frac{1}{2}\mathbf{x}'\Sigma(\theta)^{-1}\mathbf{x} \qquad [1.39]$$

donde T es el tamaño muestral, $\mathbf{x}=(x_1,\ldots,x_T)'$ es el vector de observaciones, y $\Sigma(\theta)$ es la matriz de varianzas y covarianzas de $\mathbf{x}$, es consistente y asintóticamente normal. Dahlhaus (1989) extendió estos resultados a una clase más general de modelos gaussianos estacionarios, en la que se incluye el proceso ARFIMA(p,d,q) gaussiano con media μ desconocida. Bajo hipótesis de normalidad y otras condiciones, Dahlhaus (1989) demostró que si $0<d<1/2$, el estimador máximo verosímil de θ, $\hat{\theta}_T$, es aún consistente y asintóticamente normal y además es asintóticamente eficiente. Además, estos resultados siguen siendo válidos si en la función de verosimilitud, $L(\theta,\mu;\mathbf{x})$, se sustituye μ por un estimador suyo consistente, por ejemplo la media muestral, y se maximiza con respecto a θ la función de verosimilitud resultante $L(\theta;\overline{X},\mathbf{x})$. Sowell (1992a) proporciona un algoritmo para calcular la matriz de

varianzas covarianzas $\sum(\theta)$ y obtener posteriormente la función de verosimilitud gaussiana en [1.39].

Aunque el método de estimación MV exacto es muy atractivo por las buenas propiedades de los estimadores que proporciona, el método es computacionalmente muy costoso. Para maximizar la función de verosimilitud hay que emplear métodos numéricos que requieren calcular, en cada iteración, la inversa de la matriz de varianzas y covarianzas. Esto consume mucho tiempo, especialmente si la serie es larga o hay muchos parámetros. Además, si d está próximo a 1/2, la evaluación de la inversa de $\sum(\theta)$ puede ser muy inestable, ya que las covarianzas apenas cambian y la matriz llega a ser casi singular.

Recientemente, Chan y Palma (1998) han propuesto una forma más eficiente de calcular la función de verosimilitud en modelos ARFIMA gaussianos, utilizando la descomposición en los errores de predicción. Esta descomposición permite calcular la verosimilitud de forma recursiva a través del filtro de Kalman en un número finito de pasos, y evita tener que invertir la matriz de varianzas y covarianzas. Además, Chan y Palma (1998) proponen un método de estimación máximo verosímil aproximado que consiste en truncar la representación MA(∞) del modelo ARFIMA en primeras diferencias y maximizar la función de verosimilitud del modelo truncado, utilizando una representación en el espacio de los estados de dicho proceso. Este método simplifica considerablemente los cálculos y proporciona estimadores consistentes y asintóticamente normales y eficientes.

Como alternativa a los métodos de máxima verosimilitud exacta, se han propuesto otros métodos que consisten en maximizar una aproximación a la función de verosimilitud gaussiana en el dominio de las frecuencias.

Fox y Taqqu (1986) consideran los estimadores MV aproximados de Whittle (1962), que se obtienen minimizando la función:

$$L_W(\theta;\mathbf{x}) = \frac{1}{4\pi}\int_{-\pi}^{\pi}[\ln f(\lambda;\theta) + \frac{I_T(\lambda)}{f(\lambda;\theta)}]d\lambda \qquad [1.40]$$

donde el subíndice W hace referencia a Whittle, $f(\lambda;\theta)$ es la función de densidad espectral de x_t e $I_T(\lambda)$ es el periodograma de $(x_1,\ldots,x_T)$, definido como:

$$I_T(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T}(x_t - \overline{X})e^{it\lambda}\right|^2, \text{ con } \overline{X} = \frac{1}{T}\sum_{t=1}^{T}x_t$$

En el caso de que x_t sea un proceso estacionario gaussiano cuya densidad espectral admita una parametrización de la forma $f(\lambda;\theta)=\theta_1 g(\lambda;\theta^*)$, con $\int_{-\pi}^{\pi}\ln g(\lambda;\theta^*)d\lambda=0$ y $\theta=(\theta_1,\theta^*)$, Fox y Taqqu (1986) prueban la consistencia de los estimadores MV de Whittle de (θ_1,θ^*), y demuestran que el estimador de θ^*, $\hat{\theta}^*_W$, es asintóticamente normal, pero no obtienen la distribución asintótica del estimador de θ_1. Dahlhaus (1989) extiende este resultado y demuestra que para un proceso estacionario gaussiano con densidad espectral arbitraria $f(\lambda;\theta)$, el estimador de Whittle del vector paramétrico completo, $\hat{\theta}_W$, es consistente y asintóticamente normal, y establece además condiciones para que dicho estimador sea asintóticamente eficiente. El proceso ARFIMA(p,d,q) gaussiano verifica las condiciones anteriores y por tanto el estimador MV de Whittle de d es consistente, asintóticamente normal y asintóticamente eficiente.

Los estimadores aproximados de Whittle son atractivos porque reemplazan la evaluación de la inversa de la matriz de varianzas y covarianzas por transformadas de Fourier del vector de observaciones, y los estimadores tienen, bajo ciertas condiciones, la misma distribución asintótica que los estimadores MV exactos. Sin embargo, el método tampoco es conveniente desde un punto de vista computacional, porque la integral en [1.40] debe evaluarse por métodos numéricos y esto debe

repetirse en cada iteración. La solución adoptada generalmente en estos casos es sustituir esta integral por una suma finita sobre las frecuencias de Fourier y minimizar la función correspondiente:

$$\tilde{L}_W(\theta;\mathbf{x}) = \frac{1}{2T}\sum_{j=1}^{T-1}\left[\ln f(\lambda_j;\theta) + \frac{I_T(\lambda_j)}{f(\lambda_j;\theta)}\right] \qquad [1.41]$$

donde λ_j es la j-ésima frecuencia de Fourier, $\lambda_j = \frac{2\pi j}{T}$. Minimizando $\tilde{L}_W(\theta;\mathbf{x})$ con respecto a θ se obtiene una versión discreta del estimador MV de Whittle en el dominio de las frecuencias, que denotaremos $\tilde{\theta}_W$. Desde un punto de vista computacional, $\tilde{\theta}_W$ es el estimador más conveniente, porque el cálculo de $\tilde{L}_W(\theta;x)$ se puede agilizar utilizando la transformada rápida de Fourier[17] y la simetría de la función de densidad espectral y del periodograma. Además, según Dahlhaus (1989), este estimador tiene las mismas propiedades asintóticas que su análogo anterior, $\hat{\theta}_W$.

Cheung y Diebold (1994) comparan las propiedades en muestras finitas de la versión discreta del estimador aproximado de Whittle, $\tilde{\theta}_W$, y del estimador máximo verosímil exacto en el dominio temporal, $\hat{\theta}_T$, para procesos ARFIMA(0,d,0) gaussianos. Estos autores observan que si la media del proceso es conocida, el estimador $\hat{\theta}_T$ es, por supuesto, más eficiente, pero cuando la media debe estimarse, el estimador $\tilde{\theta}_W$ es preferible, especialmente en muestras grandes donde calcular el estimador MV exacto resulta impracticable. Véase no obstante Smith, Sowell y Zin (1997) que proponen una corrección por el sesgo del estimador $\hat{\theta}_T$ que parece tener buenas propiedades en muestras finitas frente a $\tilde{\theta}_W$.

[17] Véase Bloomfield (1976, cap.4) o Priestley (1981, pp.575-586) para una explicación de la transformada rápida de Fourier (FFT) y sus aplicaciones.

Las propiedades asintóticas del estimador de Whittle que hemos descrito hasta este momento, se derivan, en todos los casos, bajo el supuesto de normalidad. Para procesos lineales con memoria larga no necesariamente gaussianos, existen pocos resultados. Giriatis y Surgalis (1990) derivan un teorema central del límite para el estimador $\hat{\theta}_W^*$, cuando la densidad espectral del proceso admite una parametrización similar a la de Fox y Taqqu (1986). Hosoya (1996, 1997) presenta resultados más generales y deriva las propiedades asintóticas de los estimadores de Whittle para procesos multivariantes no necesariamente gaussianos.

Los estimadores que acabamos de revisar tienen propiedades muy deseables (consistencia, distribución asintótica normal y eficiencia), pero, como señala Robinson (1995b), en caso de que el modelo paramétrico establecido para $f(\lambda)$ esté mal especificado, los estimadores serán sesgados e inconsistentes, y de ahí el interés por estimadores semiparamétricos.

Una primera alternativa a los métodos de estimación paramétricos es el método propuesto por Geweke y Porter-Hudak (1983), GPH de aquí en adelante, basado en un modelo de regresión para el logaritmo del periodograma, o la versión mejorada del mismo, propuesta por Robinson (1995a), que proporciona un estimador más eficiente y establece con rigor la consistencia y su distribución asintótica.

El estimador GPH explota la sencillez de la expresión que tiene la densidad espectral de los procesos fraccionalmente integrados en las frecuencias bajas. Según la expresión [1.38], la densidad espectral de x_t en un modelo ARFIMA(p,d,q) puede escribirse como:

$$f(\lambda)=f_U(\lambda)[4\text{sen}^2(\lambda/2)]^{-d} \qquad [1.42]$$

donde $f_U(\lambda)$ es la densidad espectral del proceso $u_t \sim$ ARMA(p,q) tal que $(1-L)^d x_t = u_t$. Evaluando la expresión [1.42] en las frecuencias de Fourier y tomando logaritmos, se obtiene una expresión aproximada de la forma:

$$\ln[I_T(\lambda_j)] \cong \beta_0 - d\{\ln[4\text{sen}^2(\lambda_j/2)]\} + \varepsilon_j, \text{ con } j=1,2,\ldots,m \qquad [1.43]$$

donde λ_j es la j-ésima frecuencia de Fourier, $I_T(\lambda_j)$ es la j-ésima ordenada del periodograma, $\beta_0=\ln[f_U(0)]-\gamma$, siendo γ la constante de Euler[18], $\varepsilon_j=\ln[I_T(\lambda_j)/f(\lambda_j)]+\gamma$, y $m=g(T)$ es un entero menor que T y tendiendo hacia infinito más despacio que T.

La expresión [1.43] puede interpretarse como un modelo de regresión en el dominio espectral, lo que sugiere estimar d por mínimos cuadrados ordinarios (MCO). GPH (1983) argumentan que las perturbaciones ε_j de la ecuación [1.43] son asintóticamente independientes e idénticamente distribuidas con media cero y varianza $\pi^2/6$, y establecen de forma heurística la consistencia y la normalidad asintótica del estimador $\hat{d}_{MCO}$, cuando $d<0$. Las pruebas rigurosas de las propiedades de este estimador no se establecen hasta el trabajo de Robinson (1995a), que demuestra la consistencia y la normalidad asintótica para todo el rango de valores $-1/2<d<1/2$, y propone una versión modificada y más eficiente del mismo.

El estimador propuesto por Robinson (1995a) es mucho más general que GPH en varios sentidos. En primer lugar, sólo requiere que la densidad espectral en las frecuencias bajas verifique la condición:

$$f(\lambda) \sim C_f|\lambda|^{-2d}, \text{ cuando } \lambda \to 0^+ \qquad [1.44]$$

sin necesidad de explicitar la forma concreta de $f(\lambda)$. En este caso, el modelo de regresión que se estima es:

$$\ln[I_T(\lambda_j)] \cong \ln(C_f) - 2d(\ln|\lambda_j|) + \ln[I_T(\lambda_j)/f(\lambda_j)], \text{ con } j=1,2,\ldots,m \qquad [1.45]$$

[18] Véase Abramovitz y Stegun (1970, p. 255)

donde λ_j, $I_T(\lambda_j)$ y m son como en el caso anterior. Robinson (1995a) sugiere además eliminar las frecuencias más bajas, truncando inferiormente en un punto l que depende de T. Bajo hipótesis de normalidad para x_t, Robinson (1995a) demuestra que el estimador que se obtiene estimando por mínimos cuadrados ordinarios la ecuación [1.45], es consistente y asintóticamente normal, aunque con una tasa de convergencia $\sqrt{m}$ inferior a la $\sqrt{T}$ típica de los estimadores paramétricos gaussianos. Además, todos los resultados de Robinson (1995a) se derivan para modelos multivariantes, y su trabajo se convierte así en el primero en derivar resultados asintóticos para series multivariantes con memoria larga.

Este método de estimación es muy sencillo de aplicar (cualquier ordenador hace una estimación MCO) y no requiere apenas supuestos sobre el modelo. De hecho, el estimador GPH ha sido durante años el más utilizado en aplicaciones empíricas, al menos como punto de comparación para analizar las propiedades de otros estimadores. Sin embargo, la pérdida de eficiencia de este estimador puede llegar a ser extrema en muestras pequeñas. Agiakloglou, Newbold y Wohar (1993) analizan las propiedades de este estimador en muestras finitas y observan que cuando u_t es un AR(1) o un MA(1) con parámetro cercano a la unidad, el estimador tiene un sesgo enorme y es muy ineficiente. Otro problema adicional de este método es la elección del número m de frecuencias que se utilizan. Finalmente, Robinson (1995b) señala que el mayor inconveniente es el carácter restrictivo que supone la hipótesis de normalidad sobre la serie x_t. En dos trabajos recientes, Velasco (1999b, 2000) extiende los resultados de Robinson (1995a) a modelos no estacionarios ($d \geq 1/2$) y modelos no necesariamente gaussianos, respectivamente.

Como alternativa al estimador MCO en el dominio espectral, Robinson (1995b) propone un estimador semiparamétrico para series cuyo espectro verifique la condición [1.44]. El estimador se obtiene maximizando una versión de la aproximación discreta de Whittle a la función de verosimilitud gaussiana dada en [1.41], donde la suma sólo se realiza en un entorno de la frecuencia cero. Este estimador domina a los anteriores en varios aspectos: es más eficiente, no impone

normalidad sobre x_t y evita el recorte de las componentes en las frecuencias bajas así como la elección subjetiva del número m de ordenadas a promediar. Bajo ciertas condiciones, Robinson (1995b) prueba que en el rango $-1/2<d<1/2$, el estimador es consistente y asintóticamente normal, aunque nuevamente la tasa de convergencia es inferior a $\sqrt{T}$. Recientemente Velasco (1999a) extiende estos resultados al caso no estacionario ($d\geq 1/2$).

1.2.3. Contrastes de hipótesis

Existen diversos métodos para contrastar la existencia de memoria larga en una serie. En esta sección se presentan aquellos que han recibido mayor atención en la literatura econométrica, como son el contraste que se deriva del modelo de regresión sobre el periodograma propuesto por Geweke y Porter-Hudak (1983), denotado por GPH en la sección anterior, el contraste basado en el estadístico R/S modificado de Lo (1991), y los contrastes de los multiplicadores de Lagrange de Robinson (1991, 1994a). También se describen brevemente algunos trabajos que analizan el comportamiento de los contrastes de raíces unitarias en el ámbito de los modelos con memoria larga, como Sowell (1990) o Lee y Schmidt (1996).

El contraste GPH se deriva de la estimación mínimo cuadrática del modelo de regresión en el dominio espectral, definido en la ecuación [1.43]. La pendiente de este modelo es precisamente –d, y por tanto, para contrastar la hipótesis H_0:d=0 (memoria corta) frente a la alternativa H_1:d≠0, se propone realizar el contraste habitual de significación sobre el parámetro d con una distribución t-Student. En su propuesta original, la estimación del modelo se realiza sólo sobre las primeras m ordenadas del periodograma, donde m se elige en función del tamaño muestral. Robinson (1995a) sugiere además eliminar las frecuencias más bajas, sin embargo, Hurvich, Deo y Brodsky (1998) demuestran que la reducción de sesgo que se consigue de esta manera no compensa el incremento en la varianza.

Aunque el contraste GPH es muy sencillo de aplicar, Agiakloglou, Newbold y Wohar (1993) prueban que el estimador de d es muy sesgado cuando existen bajo H_0 componentes autorregresivas o de medias móviles con valores grandes de los parámetros, y, por tanto, en estos casos, el contraste rechazará la hipótesis nula, aún siendo cierta, con mucha frecuencia. Esta misma conclusión se deriva de las simulaciones de Cheung (1993b), en las que se observa un sesgo del contraste hacia alternativas del tipo d>0 cuando el proceso generador de los datos es un AR(1) con un valor grande del parámetro autorregresivo.

Haubrich y Lo (1989) y Lo (1991) proponen un contraste de memoria corta basado en una modificación del estadístico R/S (rango reescalado) introducido por Hurst (1951). El estadístico R/S es el rango de las sumas parciales de las desviaciones de una serie a su media, normalizado por su desviación típica muestral. La propuesta de Lo (1991) consiste en utilizar como factor de normalización otro estimador de la varianza que no sea la varianza muestral. En concreto, el estadístico de contraste que se propone es:

$$Q_T(q) = \frac{1}{S_T(q)}\left[\max_{1\le k\le T}\sum_{i=1}^{k}(x_i - \overline{X}) - \min_{1\le k\le T}\sum_{i=1}^{k}(x_i - \overline{X})\right] \qquad [1.46]$$

donde $\overline{X}$ es la media muestral de las T observaciones, $\overline{X} = \frac{1}{T}\sum_{t=1}^{T} x_t$, q es un parámetro de truncamiento, y $S_T(q)$ es la raíz cuadrada de un estimador consistente de la varianza definido como:

$$S_T^2(q) = c(0) + 2\sum_{j=1}^{q} w_j(q)c(j), \text{ con } w_j(q) = 1 - \frac{j}{q+1}, \text{ con } q<T$$

donde c(0) y c(j) denotan la varianza y la autocovarianza muestral de orden j de x_t,

$c(0)=\frac{1}{T}\sum_{t=1}^{T}(x_t-\overline{X})^2$ y $c(j)=\frac{1}{T}\sum_{t=1}^{T-j}(x_t-\overline{X})(x_{t+j}-\overline{X})$, respectivamente.

El estadístico $Q_T(q)$ definido en [1.46] se denomina "rango reescalado modificado", y para q=0 se reduce al estadístico R/S de Hurst. Valores extremos de $Q_T(q)$, en ambos sentidos, se consideran indicativos de la posible existencia de memoria larga en la serie x_t. Bajo ciertas condiciones que incluyen una clase muy general de procesos con memoria corta, Lo (1991) deriva la distribución asintótica del estadístico $V_T(q)=\frac{Q_T(q)}{\sqrt{T}}$, y proporciona los valores críticos del contraste para los niveles de significación habituales. Además, este autor presenta un estudio de Monte Carlo sobre el tamaño y la potencia del contraste en muestras finitas, del cual se deduce que el contraste tiene una potencia razonable, que aumenta con T y disminuye con q, frente a alternativas del tipo ARFIMA(0,d,0). Frente a esta clase de modelos, el contraste parece detectar mejor valores negativos de d que alternativas del tipo d>0. Las simulaciones de Cheung (1993b) sugieren que, al igual que GPH, este contraste es muy sensible a la existencia de correlación bajo la hipótesis nula: por ejemplo, un MA(1) con valores grandes del parámetro sesga el contraste hacia alternativas d<0.

Robinson (1991) desarrolla un estadístico basado en el principio de los multiplicadores de Lagrange (LM) para contrastar la hipótesis de ruido blanco frente a una familia muy general de alternativas que incluye a la clase de los procesos ARFIMA(0,d,0). En este caso, el estadístico admite una expresión muy sencilla de la forma $\Lambda_1=\lambda_1^2$, donde $\lambda_1=\sqrt{\frac{6T}{\pi^2}\sum_{j=1}^{T-1}\frac{r(j)}{j}}$, siendo r(j)=c(j)/c(0) la autocorrelación muestral de orden j.[19] Bajo ciertas condiciones sobre los coeficientes de la representación lineal del modelo y sobre los momentos condicionales de primer y

[19] Además de este contraste, Robinson (1991) propone otra variante que es robusta a la heteroscedasticidad condicionada de las perturbaciones.

segundo orden de las innovaciones de esta, Robinson (1991) demuestra que, bajo la hipótesis de ruido blanco (H_0:d=0), λ_1 tiene una distribución asintótica N(0,1), y por tanto, Λ_1 se distribuye asintóticamente como una χ^2 con un grado de libertad. Estos dos estadísticos permiten contrastar alternativas del tipo d>0, d<0 ó d≠0. Así, el contraste de un lado frente a la alternativa de memoria larga (H_1:d>0), consiste en rechazar H_0 si λ_1 es significativamente positivo respecto al correspondiente valor crítico de una N(0,1), pero si lo se quiere contrastar es H_1:d<0, lo apropiado es un contraste de un lado sobre la cola izquierda de la distribución N(0,1). Finalmente, un contraste frente a la alternativa H_1:d≠0 rechazará la hipótesis de ruido blanco si Λ_1 toma valores significativamente grandes respecto a la distribución χ^2 con un grado de libertad.

El contraste LM de Robinson (1991) presenta ciertas ventajas sobre los anteriores, ya que evita tener que estimar el parámetro d y, además, los estadísticos de contraste tienen una distribución asintótica conocida y sencilla. Cheung (1993b) prueba que, en muestras finitas, este contraste detecta muy bien alternativas fraccionales, aunque es muy sensible a la presencia de correlación serial en los datos: un parámetro AR(1) o MA(1) hace que el contraste rechace H_0 con muchísima más frecuencia que el nivel de significación nominal. Robinson (1994a) sugiere una forma de robustecer el estadístico Λ_1 contra la presencia de autocorrelación débil en la hipótesis nula, y Agiakloglou y Newbold (1994) presentan una generalización del mismo para contrastar H_0:ARMA(p,q) frente a la alternativa H_1:ARFIMA(p,d,q).

Finalmente, Lobato y Robinson (1998) han propuesto un contraste semiparamétrico para la hipótesis I(0) frente a alternativas de integración fraccional, basado en un estadístico LM de los multiplicadores de Lagrange aplicado a la función objetivo del método de estimación de Robinson (1995b). Para procesos lineales multivariantes que verifiquen ciertas condiciones sobre los momentos condicionales de orden cuarto de las innovaciones y ciertas condiciones de diferenciabilidad del espectro, estos autores demuestran que el estadístico LM se distribuye, bajo la

hipótesis nula (H_0:d=0), como una distribución χ^2_1. Por tanto, H_0 se rechaza en favor de la alternativa H_1:d≠0, cuando el valor del estadístico LM es significativamente grande con respecto a la distribución χ^2_1. El contraste parece tener buenas propiedades de eficiencia frente a alternativas fraccionales, comparado con otros contrastes semiparamétricos.

Para el caso univariante, el estadístico del contraste de Lobato y Robinson (1998) admite una expresión de la forma $LM=t^2$, donde t se define como:

$$t = -\sqrt{m}\,\frac{\sum_{j=1}^{m} \upsilon_j I(\lambda_j)}{\sum_{j=1}^{m} I(\lambda_j)}, \text{ con } \upsilon_j = \ln j - \frac{1}{m}\sum_{j=1}^{m} \ln j$$

donde $I(\lambda_j)$ es la j-ésima ordenada del periodograma y m es un parámetro de truncamiento, con $0<m<T/2$. En este caso, la hipótesis H_0:d=0 se rechaza en favor de la alternativa de memoria larga (H_1:d>0), cuando el estadístico t toma un valor significativamente positivo en la cola derecha de la distribución normal N(0,1).

Para terminar esta sección, mencionaremos brevemente algunos contrastes de raíces unitarias que pueden enmarcarse en el ámbito de los procesos fraccionalmente integrados, en cuyo caso la hipótesis de raíz unitaria se convierte en H_0:d=1.

Sowell (1990) deriva la distribución asintótica del estadístico del contraste de Dickey-Fuller para series fraccionalmente integradas I(d), con $1/2<d<3/2$, y demuestra que ésta depende del orden de integración d. Diebold y Rudebush (1991) analizan las propiedades de este contraste en muestras finitas y comprueban que tiene muy poca potencia frente a alternativas fraccionales, y en consecuencia será muy difícil que permita discriminar entre un proceso integrado I(1) y un proceso I(d) con $d<1$. Hassler y Wolters (1994) realizan un estudio similar para el contraste de Dickey-Fuller aumentado y demuestran que éste tiene incluso menos potencia frente a alternativas I(d). Dolado y Marmol (1997) extienden estos resultados al contraste

secuencial de Dickey-Pantula, permitiendo así el análisis de procesos integrados con más de una raíz unitaria.

Lee y Schmidt (1996) consideran el contraste KPSS de Kwiatkowski *et al*. (1992) como un contraste de memoria corta (H_0:d=0) y prueban que es consistente frente a procesos con memoria larga estacionarios I(d), con $-1/2<d<1/2$ y $d\neq 0$. Lee y Amsler (1997) extienden estos resultados a procesos fraccionalmente integrados no estacionarios ($1/2<d<1$) y demuestran que el contraste KPSS no puede distinguir de forma consistente estos procesos de un proceso con raíz unitaria (d=1).

Marmol (1998) revisa los resultados anteriores y prueba que los contrastes unilaterales de Dickey-Fuller (H_0:d=1) y KPSS (H_0:d=0) son consistentes frente a alternativas fraccionales del tipo H_1:d<1 y H_1:d>0, respectivamente, pero la utilización combinada de ambos contrastes sólo conlleva a la decisión correcta, para todo d>-1/2, si éstos se realizan como contrastes de dos lados.

Robinson (1994a) propone un contraste de los multiplicadores de Lagrange para la hipótesis de raíz unitaria y otras formas de no estacionariedad, frente a alternativas muy generales que incluyen a los procesos fraccionalmente integrados. Cuando la perturbación del modelo es un ruido blanco con media cero, este contraste tiene propiedades estándar de eficiencia y tiene una distribución asintótica χ^2 bajo la hipótesis nula. Robinson (1994a) deriva además la distribución asintótica del estadístico de contraste cuando la perturbación del modelo está débilmente correlacionada. En muestras finitas, este contraste compara favorablemente con otros contrastes de raíces unitarias, y permite contrastar no sólo la hipótesis d=1 (raíz unitaria) sino otros valores del parámetro d.

Algunos de los contrastes de memoria larga mencionados en esta sección, se han aplicado a series de rendimientos y series de volatilidades de los mercados financieros, obteniéndose resultados muy diferentes para cada una de ellas. Por ejemplo, los trabajos de Lo (1991), Lobato y Robinson (1998) y Lobato y Savin (1998), indican que el comportamiento de los rendimientos de activos financieros de

la Bolsa de Nueva York es consistente con la hipótesis de memoria corta. Blasco y Santamaría (1994a), Blasco y Santamaría (1994b) y Olmeda (1998), tampoco encuentran evidencia de memoria larga en los rendimientos del mercado de valores español. Sin embargo, Lobato y Robinson (1998) y Lobato y Savin (1998) encuentran una evidencia clara de memoria larga en los cuadrados de los rendimientos para distintas series de tipos de cambio y para una serie diaria del índice bursátil SP500 de la bolsa de Nueva York, respectivamente. Los resultados de estos trabajos confirman los resultados de trabajos anteriores, como Ding, Granger y Engle (1993), Crato y de Lima (1994) y Breidt, Crato y de Lima (1998), que ya habían apuntado la posible existencia de memoria larga en la volatilidad de las series financieras.

Para tratar de modelizar este comportamiento de la volatilidad se han propuesto en la literatura diversos modelos con memoria larga, algunos de los cuales se describen en la sección siguiente.

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

1.3. Modelos con memoria larga para la volatilidad

Como vimos en la sección 1.1, en las aplicaciones empíricas de modelos GARCH o SV a series financieras observadas con gran frecuencia, es muy habitual estimar un fuerte grado de persistencia en los momentos de segundo orden. Esta regularidad empírica motivó la introducción de modelos con raíces unitarias para la volatilidad, como el modelo IGARCH o el modelo SV integrado. Estos modelos se caracterizan, entre otras cosas, porque el efecto de un *shock* en la varianza condicionada es permanente, en el sentido de que influye en las predicciones futuras de la volatilidad para todos los horizontes de predicción.

Algunos trabajos recientes cuestionan la validez de los modelos integrados, precisamente porque el grado extremo de persistencia que inducen en la volatilidad no refleja las características observadas empíricamente en ciertas series financieras, que sugieren más bien la existencia de memoria larga en la volatilidad.

El trabajo de Ding, Granger y Engle (1993), pionero en este campo, analiza una serie diaria del índice bursátil SP500, en el periodo comprendido entre el 3/1/1928 y el 30/8/1991, y descubre que las autocorrelaciones de las series de los cuadrados de los rendimientos y de los rendimientos absolutos tienden a persistir, incluso para retardos muy altos, pero finalmente convergen hacia cero. Este comportamiento es consistente con la propiedad de memoria larga que, como vimos en la sección anterior, se caracteriza precisamente por correlaciones que decrecen lentamente hacia cero y no son absolutamente sumables. Bollerslev y Mikkelsen (1996) analizan una serie diaria del índice SP500 sobre un periodo muestral más corto, y observan estas mismas características.

Dacorogna *et al*. (1993) encuentran un comportamiento similar en una serie del tipo de cambio dólar *vs* marco alemán, con datos recogidos con una frecuencia de veinte minutos. Estos autores ajustan dos funciones, una hiperbólica y otra exponencial, al correlograma de los rendimientos absolutos y comprueban que el decrecimiento hiperbólico, típico de los procesos fraccionalmente integrados, se ajusta mejor que el exponencial. Resultados similares aparecen en Andersen y Bollerslev (1997).

Crato y de Lima (1994) contrastan la persistencia en la volatilidad de varias series de índices bursátiles, aplicando los contrastes de memoria larga de Geweke y Porter Hudak (1983) y Lo (1991) a la serie de los cuadrados de los rendimientos y a la serie de sus logaritmos, y concluyen que efectivamente existe dependencia a largo plazo en muchas de ellas. Estos resultados se confirman posteriormente en Breidt, Crato y de Lima (1998). Olmeda (1998) aplica estos mismos contrastes a diferentes series del mercado bursátil español, y concluye que hay cierta evidencia de memoria larga en las volatilidades de los rendimientos.

Lobato y Savin (1998) aplican el contraste semiparamétrico de Lobato y Robinson (1998) a una serie diaria de rendimientos del índice SP500 sobre un periodo comprendido entre julio de 1962 y diciembre de 1994, y concluyen que hay una fuerte evidencia de memoria larga en los cuadrados de la serie, aunque no la haya en la propia serie.

La evidencia empírica presentada en estos trabajos sugiere, en primer lugar, que la varianza condicionada de los rendimientos puede tener mayor memoria que la típicamente capturada por los modelos GARCH o SV, ya que el decrecimiento hiperbólico observado empíricamente en los correlogramas de los cuadrados contrasta con el decrecimiento exponencial impuesto por dichos modelos. En segundo lugar, estos trabajos confirman las limitaciones de los modelos con raíces unitarias para reproducir la persistencia observada en los datos financieros, ya que ciertamente existe un alto grado de dependencia en los momentos de segundo orden, pero éste desaparece finalmente con el horizonte de predicción.

Motivados por estos resultados se han propuesto recientemente varios modelos que tratan de recoger la posible existencia de memoria larga en los momentos condicionados de segundo orden. Todos ellos abordan el problema desde la misma perspectiva: modelizar directamente los cambios en la volatilidad con procesos fraccionalmente integrados capaces de producir memoria larga. Entre los modelos propuestos hasta el momento cabría mencionar los siguientes.

Baillie, Bollerslev y Mikkelsen (1996) y Bollerslev y Mikkelsen (1996) proponen los modelos FIGARCH y FIEGARCH, respectivamente. Estos modelos son versiones con memoria larga de los modelos GARCH y EGARCH, que se obtienen incluyendo en la parte autorregresiva de éstos el operador de diferencias fraccionales, de forma que los modelos resultantes pueden interpretarse como procesos ARFIMA para la serie de los cuadrados. Aunque la extensión es inmediata, las condiciones de estacionariedad y las propiedades dinámicas de los modelos FIGARCH y FIEGARCH no son fáciles de derivar, y tampoco parece fácil resolver el problema de estimación, sobre el que no hay ningún resultado concluyente.

En el contexto de los modelos SV, Harvey (1998) y Breidt, Crato y de Lima (1998) han propuesto el modelo SV con memoria larga o modelo LMSV (Long Memory Stochastic Volatility), en el cual el logaritmo de la volatilidad es un proceso ARFIMA. Las propiedades de este modelo son fáciles de obtener a partir de las del proceso ARFIMA subyacente, pero también existen problemas de estimación derivados de la dificultad de construir la función de verosimilitud exacta. Sin embargo existen ya varios métodos de estimación alternativos propuestos recientemente.

Robinson y Zaffaroni (1997) proponen los modelos MA no lineales con memoria larga, en los que la volatilidad se modeliza mediante un proceso MA(∞) capaz de producir memoria larga en la serie de los cuadrados. Zaffaroni (1997) estudia las propiedades dinámicas de este modelo y propone un método de estimación aproximado de Whittle que es consistente y asintóticamente normal.

Existen además otras propuestas para modelizar la memoria larga en la volatilidad, como los modelos A-PARCH (Asymmetric Power ARCH) de Ding, Granger y Engle (1993) y Ding y Granger (1996), y el modelo con dos perturbaciones propuesto por Robinson y Zaffaroni (1998) como alternativa al modelo LMSV, pero su relevancia en aplicaciones empíricas aún es escasa.

En las secciones siguientes se describirán brevemente algunos de estos modelos, y se presentará con detalle el modelo LMSV, ya que sobre él versarán el resto de los capítulos de este trabajo. En la sección 1.3.1 se presenta el modelo GARCH fraccionalmente integrado o FIGARCH, propuesto por Baillie, Bollerslev y Mikkelsen (1996) y el modelo EGARCH fraccionalmente integrado o FIEGARCH de Bollerslev y Mikkelsen (1996). Se describen las características más importantes de cada uno de ellos y se discuten sus posibles ventajas e inconvenientes. En la sección 1.3.2 se define el modelo de volatilidad estocástica con memoria larga LMSV propuesto por Harvey (1998) y Breidt, Crato y de Lima (1998) y se describen sus propiedades dinámicas. Se discuten las propiedades de la función de autocorrelación de la serie de los cuadrados y de su logaritmo y se plantea un posible problema de

identificación del modelo para ciertos valores de los parámetros. Sin embargo en esta sección no se aborda el problema de estimación del modelo porque éste es precisamente uno de los temas centrales de nuestra investigación, al que dedicamos íntegramente el capítulo 3. Finalmente, en la sección 1.3.3 se describen brevemente otros modelos con memoria larga para la volatilidad.

1.3.1. Modelos GARCH fraccionalmente integrados

Robinson (1991, p. 68) sugirió la primera extensión de los modelos GARCH capaz de producir memoria larga en los cuadrados de la serie, pero son Baillie, Bollerslev y Mikkelsen (1996) quienes posteriormente desarrollan el modelo y lo denominan GARCH fraccionalmente integrado o FIGARCH. Este modelo combina muchas de las características de los procesos fraccionalmente integrados para la media con las de los procesos GARCH, permitiendo así modelizar las dependencias temporales observadas empíricamente en la volatilidad de las series financieras.

Para comprender mejor las propiedades del modelo FIGARCH conviene situar el nuevo modelo en el marco general de la clase de los modelos GARCH descritos en la sección 1.1. Recordemos que el modelo GARCH(p,q) puede interpretarse, según la ecuación [1.5], como un proceso ARMA en la serie de los cuadrados de la forma:

$$[1-\alpha(L)-\beta(L)]y_t^2 = \alpha_0 + [1-\beta(L)]\upsilon_t \qquad [1.47]$$

donde $\alpha(L)$ y $\beta(L)$ son polinomios de la forma $\alpha(L) = \sum_{i=1}^{q}\alpha_i L^i$, $\beta(L) = \sum_{j=1}^{p}\beta_j L^j$, y υ_t es el ruido blanco definido en [1.4]. Cuando el polinomio autorregresivo de la ecuación [1.47] posee una raíz unitaria, tenemos el modelo GARCH integrado o IGARCH, que puede interpretarse como un proceso ARIMA en la serie de los

cuadrados de la forma:

$$\phi(L)(1-L)y_t^2 = \alpha_0 + [1-\beta(L)]\upsilon_t$$

donde $\phi(L)(1-L) = 1-\alpha(L)-\beta(L)$, y $\phi(z)=0$ tiene todas las raíces fuera del círculo unidad.

El modelo GARCH fraccionalmente integrado o FIGARCH(p,d,q) es un modelo intermedio entre los dos modelos anteriores, en el cual la propiedad de memoria larga se recoge permitiendo que el polinomio autorregresivo de la ecuación [1.47] tenga raíces unitarias fraccionales. La ecuación de la varianza condicionada se define entonces como:

$$\phi(L)(1-L)^d y_t^2 = \alpha_0 + [1-\beta(L)]\upsilon_t \qquad [1.48]$$

donde d es un número real $0 \le d \le 1$, $(1-L)^d$ es el operador de diferencias fraccionales definido en [1.28], υ_t es el ruido blanco definido en la ecuación [1.4] y todas las raíces de los polinomios $\phi(L)$ y $[1-\beta(L)]$ están fuera del círculo unidad.

Al permitir órdenes de integración fraccionales, el modelo FIGARCH proporciona una gran flexibilidad en la modelización de la dependencia temporal de la varianza condicionada, e incluye como casos particulares los modelos GARCH (d=0) e IGARCH (d=1). El modelo es, por tanto, el análogo para la varianza condicionada de los modelos ARFIMA para la media que hemos estudiado en la sección 1.2. Obsérvese que en el modelo FIGARCH(1,d,0), definido por la ecuación $(1-\phi L)(1-L)^d y_t^2 = \alpha_0+\upsilon_t$, existe un problema de identificación, ya que este modelo puede convertirse en un modelo IGARCH tomando dos tipos de restricciones sobre los parámetros: $\{\phi=1,d=0\}$ o $\{\phi=0,d=1\}$.

Si sustituimos en la ecuación [1.48] la perturbación υ_t por su valor en [1.4], $\upsilon_t = y_t^2 - \sigma_t^2$, obtenemos una expresión equivalente del modelo FIGARCH de la forma:

$$[1-\beta(L)]\sigma_t^2 = \alpha_0 + [1-\beta(L)-\phi(L)(1-L)^d]y_t^2$$

y por tanto, la ecuación que representa la varianza condicionada, σ_t^2, en función de las observaciones pasadas de la serie, admite una representación de la forma:

$$\sigma_t^2 = \alpha_0[1-\beta(1)]^{-1} + \lambda(L)y_t^2 \qquad [1.49]$$

donde $\lambda(L)=1-[1-\beta(L)]^{-1}\phi(L)(1-L)^d$. Para que el modelo FIGARCH esté bien definido y se garantice la no negatividad de la varianza condicionada, todos los coeficientes del polinomio $\lambda(L)$ tienen que ser positivos. Al igual que ocurría con los modelos GARCH clásicos, esta condición implica restricciones sobre los parámetros, lo cual supone una gran desventaja y complica la estimación. Además, no existe hasta el momento ningún resultado que permita establecer de forma general cuáles son esas restricciones, salvo en algunos modelos muy concretos. Por ejemplo, Baillie, Bollerslev y Mikkelsen (1996) y Bollerslev y Mikkelsen (1996) establecen las condiciones para los modelos FIGARCH(1,d,0) y FIGARCH(1,d,1), respectivamente.

Respecto a las condiciones de estacionariedad, Baillie, Bollerslev y Mikkelsen (1996) prueban que si d>0, el proceso no es débilmente estacionario, ya que la suma de los coeficientes del polinomio $\lambda(L)$ en la representación [1.49] no es menor que uno[20], y por tanto la varianza de y_t no es finita. De hecho, el modelo FIGARCH sólo es débilmente estacionario cuando d=0, en cuyo caso se reduce al GARCH estándar. Sin embargo, Baillie, Bollerslev y Mikkelsen (1996) argumentan que el modelo es

[20] Se comprueba fácilmente que la suma de los coeficientes de $\lambda(L)$ es $\lambda(1)=1-[1-\beta(1)]^{-1}\phi(1)F(-d,1,1;1)$, y dado

estrictamente estacionario y ergódico, de la misma forma que lo es el modelo IGARCH.

Una de las ventajas del modelo FIGARCH es la flexibilidad que proporciona a la hora de estudiar el grado de persistencia de los *shocks* sobre la volatilidad. Al igual que en los modelos para la media Baillie, Bollerslev y Mikkelsen (1996) proponen cuantificar este grado de persistencia utilizando la función de respuesta al impulso acumulativa.

En los modelos FIGARCH, los valores de esta función son los coeficientes del polinomio $c(L)=1+c_1L+c_2L^2+\ldots$ tal que $y_t^2=\zeta+c(L)\upsilon_t$. Así, el coeficiente c_k medirá el impacto que tiene una innovación unitaria en t sobre el nivel de la serie de los cuadrados en t+k, y el coeficiente $c_\infty=\lim_{k\to\infty} c_k$ medirá el impacto a largo plazo. Si la serie y_t^2 en primeras diferencias puede escribirse como $(1-L)\,y_t^2=a(L)\upsilon_t$, con $a(L)=1+a_1L+a_2L^2+\ldots$, se comprueba que $c_k=1+a_1+\cdots+a_k$, y por tanto su límite será:

$$c_\infty=\lim_{k\to\infty} c_k=\lim_{k\to\infty}(1+a_1+\ldots+a_k)=a(1)$$

donde a(1) es la suma de los coeficientes del polinomio a(L). Según la ecuación [1.48], la serie y_t^2 en primeras diferencias para el modelo FIGARCH puede escribirse como:

$$(1-L)y_t^2=\zeta+(1-L)^{1-d}\phi^{-1}(L)[1-\beta(L)]\,\upsilon_t$$

y por tanto $a(L)=(1-L)^{1-d}\phi^{-1}(L)[1-\beta(L)]$. Para este modelo, Baillie, Bollerslev y Mikkelsen (1996) prueban que:

---

que $F(-d,1,1;1)=0$ si $0<d\leq 1$, se tiene que $\lambda(1)=1$.

$$c_\infty = a(1) = F(d-1,1,1;1)\phi(1)^{-1}[1-\beta(1)]$$

donde F es la función hipergeométrica, $\phi(1)$ y $\beta(1)$ son las sumas de los coeficientes de los polinomios $\phi(L)$ y $\beta(L)$, respectivamente, y $0 \leq d \leq 1$. Teniendo en cuenta que la función F toma el valor $F(d-1,1,1;1)=0$ si $0 \leq d < 1$, y $F(d-1,1,1;1)=1$ si $d=1$,[21] resulta que para el modelo IGARCH (d=1), la función respuesta al impulso converge hacia una constante distinta de cero, $c_\infty = \phi(1)^{-1}[1-\beta(1)] \neq 0$, y para los modelos GARCH (d=0) y para los FIGARCH con $0<d<1$, dicha función converge hacia cero ($c_\infty=0$). En términos predictivos, esto significa que en el modelo IGARCH, el efecto de los *shocks* sobre la volatilidad persiste indefinidamente, mientras que en los modelos GARCH y FIEGARCH, el efecto del *shock* desaparece finalmente con el horizonte de predicción.

Sin embargo, existe una diferencia sustancial entre los modelos GARCH y FIGARCH, que radica precisamente en la velocidad con la que c_k converge hacia cero. Por ejemplo, en un modelo GARCH(1,1) estacionario ($\alpha_1+\beta_1<1$), el polinomio c(L) puede escribirse como $c(L)=[1-(\alpha_1+\beta_1)L]^{-1}(1-\beta_1 L)$, y se demuestra que los valores de la función respuesta al impulso acumulativa, que son los coeficientes de este polinomio, son de la forma:

$$c_k = \alpha_1 (\alpha_1 + \beta_1)^{k-1} \qquad [1.50]$$

para $k \geq 1$ y $c_0=1$. Sin embargo, en un modelo FIGARCH(1,d,0), el polinomio c(L) es de la forma $c(L)=(1-L)^{-d}(1-\beta_1 L)$. Utilizando la expresión de los coeficientes del polinomio $(1-L)^{-d}$ en [1.30], se obtienen los siguientes valores de la función respuesta al impulso:

[21] Ver Gradshteyn y Ryzhik (1980, p. 1040)

$$c_k = \left[(1-\beta_1) - \frac{(1-d)}{k}\right]\frac{\Gamma(k-1+d)}{\Gamma(k)\Gamma(d)} \cong \frac{(1-\beta_1)}{\Gamma(d)} k^{d-1}, \text{ si } k\to\infty \qquad [1.51]$$

para $k\geq 1$ y $c_0=1$.

De las expresiones [1.50] y [1.51] se deduce que, mientras en un modelo GARCH(1,1) los coeficientes c_k decrecen hacia cero exponencialmente, en un modelo FIGARCH(1,d,0), el efecto de un *shock* se desvanece muchísimo más despacio, a un ritmo hiperbólico determinado por el valor del parámetro d. En consecuencia, el parámetro d se identifica con el ritmo al que se propagan los *shocks* en la volatilidad y no con el impacto último que éstos tienen sobre la predicción de la varianza condicionada, que será cero tanto en los modelos GARCH (d=0) como en los FIGARCH (0<d<1).

El gráfico 1.1 representa los valores de la función respuesta al impulso, c_k, para los modelos GARCH(1,1) y FIGARCH(1,d,0) que Baillie, Bollerslev y Mikkelsen (1996) obtienen al modelizar la volatilidad de una serie diaria de rendimientos del tipo de cambio Dólar/Marco alemán. Los valores estimados de los parámetros para ambos modelos son $\{\hat{\alpha}_1=0.113,\ \hat{\beta}_1=0.867\}$ y $\{\hat{\beta}_1=0.762,\ \hat{d}=0.823\}$, respectivamente. En este gráfico se aprecia claramente la diferencia entre la velocidad con la que el impacto de un *shock* desaparece en cada uno de los modelos, aunque en ambos casos el impacto finalmente desaparece y converge hacia cero.

Gráfico 1.1. Función de respuesta al impulso para un modelo GARCH(1,1) con parámetros $\{\alpha_1=0.113, \beta_1=0.867\}$ y un modelo FIGARCH(1,d,0) con parámetros $\{\beta_1=0.762, d=0.823\}$.

BIBLIOTECA VIRTUAL

Con respecto a la estimación del modelo FIGARCH, Baillie, Bollerslev y Mikkelsen (1996) proponen el método de máxima verosimilitud en el dominio del tiempo, bajo la hipótesis de normalidad. A juzgar por los resultados de las simulaciones que presentan estos autores, los estimadores parecen tener buenas propiedades en muestras pequeñas, pero la aplicación de este método tiene varios inconvenientes. En primer lugar, para calcular la función de verosimilitud se requiere truncar el polinomio $(1-L)^d$ en un retardo suficientemente grande que permita capturar las propiedades de memoria larga de la serie. Además, no existe teoría asintótica rigurosa sobre las propiedades de los estimadores, y tampoco parece fácil desarrollar esa teoría. Pensemos que incluso en los modelos GARCH (d=0) e IGARCH (d=1) sólo existen resultados para los casos más sencillos.

Como alternativa al modelo FIGARCH, Bollerslev y Mikkelsen (1996) consideran una generalización similar a la anterior para el modelo EGARCH de Nelson (1991). Esta propuesta, que ya había sugerido el propio Nelson (1991),

consiste en modelizar la serie $\ln(\sigma_t^2)$ mediante una estructura de retardos en ε_t y en su propio pasado, que incluya el operador en diferencias fraccionales, $(1-L)^d$, en la parte autorregresiva de la representación [1.12] del modelo EGARCH. El modelo así definido se denomina EGARCH fraccionalmente integrado o FIEGARCH(p,d,q), e incluye como casos particulares el EGARCH (d=0) y el EGARCH con raíz unitaria (d=1).

El modelo FIEGARCH tiene las mismas ventajas que tenía el modelo EGARCH sobre los modelos GARCH: garantiza la no negatividad de la varianza condicionada modelizando su logaritmo, no impone restricciones en los parámetros, y puede modelizar una respuesta asimétrica de la volatilidad a cambios de signo en los valores de la serie. Además, si $0<d<1/2$, precisamente el rango de valores en el que un ARFIMA presenta memoria larga, Bollerslev y Mikkelsen (1996) demuestran que $\ln(\sigma_t^2)$ es un proceso ergódico y estrictamente estacionario, y es también débilmente estacionario e invertible.

Sin embargo, el modelo FIEGARCH adolece de los mismos inconvenientes que el modelo FIGARCH con respecto a la estimación. El método propuesto por Bollerslev y Mikkelsen (1996) es nuevamente la máxima verosimilitud, pero no existe ningún resultado sobre las propiedades de los estimadores. Nótese que ni siquiera para el caso EGARCH estándar (d=0) existen resultados al respecto y de hecho, son escasísimos los trabajos empíricos que han utilizado este modelo. Previsiblemente lo mismo ocurrirá con el modelo EGARCH fraccionalmente integrado, a pesar de que éste presente ciertas ventajas sobre el FIGARCH.

Una alternativa a estos modelos que conserva la idea de modelizar $\ln(\sigma_t^2)$ para asegurar la no negatividad de σ_t^2, es el modelo SV con memoria larga, que se presenta en la siguiente sección.

1.3.2. El modelo SV con memoria larga.

Harvey (1998) y Breidt, Crato y de Lima (1998) sugieren una forma alternativa de modelizar la dependencia temporal en los momentos de segundo orden, basada en los modelos de volatilidad estocástica SV. Como vimos en la sección 1.1.2, los modelos SV consideran la volatilidad, σ_t, como una componente no observable, cuyo logaritmo se modeliza mediante un proceso estocástico lineal que habitualmente es un ARMA. Por ejemplo, en el modelo ARSV(1) definido en [1.12] ese proceso es un AR(1).

El modelo de volatilidad estocástica con memoria larga LMSV (Long Memory Stochastic Volatility) no es más que una extensión del modelo anterior, en el que el proceso h_t que define la volatilidad es un proceso ARFIMA. En su formulación más sencilla, h_t será un ARFIMA(0,d,0) y el modelo LMSV básico vendrá definido por las siguientes ecuaciones:

$$y_t = \sigma_* \exp(\tfrac{1}{2} h_t)\, \varepsilon_t \qquad [1.52a]$$

$$(1-L)^d h_t = \eta_t, \qquad [1.52b]$$

donde σ_* es un factor de escala, ε_t es IID(0,1), η_t es NID$(0, \sigma_\eta^2)$ e independiente de la perturbación ε_t, y d puede tomar cualquier valor en el intervalo $0 \le d \le 1$.[22] Este modelo incluye como caso particular el modelo SV con paseo aleatorio (d=1), que se describió en la sección 1.1.2. En esta sección nuestro interés se centrará en los valores no enteros del intervalo $0<d<1$.

Harvey (1998) establece las propiedades dinámicas del modelo LMSV en [1.52]. Muchas de estas propiedades pueden derivarse de las fórmulas del modelo SV presentadas en la sección 1.1.2, siempre que el proceso h_t definido en [1.52b] sea estacionario. A continuación se resumen algunas de ellas.

---

[22] Conservamos aquí la notación original de Harvey (1998) que considera el modelo LMSV definido en $0 \le d \le 1$, a diferencia de Breidt, Crato y de Lima (1998) que consideran el rango $-1/2<d<1/2$, donde, como veremos a continuación, el modelo es estacionario e invertible.

En primer lugar, se demuestra que la serie y_t es una martingala en diferencias, porque $E(y_t/\Omega_{t-1})=0$, y, por tanto, $E(y_t)=0$ y $Cov(y_t, y_s)=0$ para $t\neq s$. Además, la serie y_t es estacionaria siempre que lo sea el proceso h_t. En este caso, h_t es un proceso ARFIMA(0,d,0) y, como vimos en la sección 1.2, este proceso es estacionario si $d<1/2$ e invertible si $d>-1/2$, y presenta memoria larga si $0<d<1/2$. Si el modelo LMSV se define sobre el intervalo $0\leq d\leq 1$, como en Harvey (1998), la serie y_t es estacionaria con varianza finita si $0\leq d<1/2$. Además, Breidt, Crato y de Lima (1998) demuestran que en este intervalo la serie y_t^2 es débil y estrictamente estacionaria. De aquí en adelante supondremos que $0\leq d<1/2$ y, por tanto, trabajaremos con un modelo LMSV estacionario.[23] Finalmente analizaremos el caso no estacionario.

Al igual que en el modelo ARSV(1) estacionario, los momentos de y_t se calculan utilizando la expresión [1.13]. Por tanto, si la distribución de ε_t es simétrica, todos los momentos impares de y_t son cero. Los momentos de orden par se deducen fácilmente utilizando las propiedades de la distribución logNormal. Las fórmulas [1.14]-[1.15] para los momentos de segundo y cuarto orden de y_t, respectivamente, que derivamos en la sección 1.1, aún son válidas para el modelo LMSV estacionario. Aplicando dichas fórmulas, y teniendo en cuenta que en este caso el proceso h_t tiene media cero, se prueba que la varianza y la curtosis de y_t en el modelo LMSV estacionario son:

$$Var(y_t) = E(y_t^2) = \sigma_*^2 \exp(\tfrac{1}{2}\sigma_h^2) \qquad [1.53]$$

$$\kappa_y = \frac{E(y_t^4)}{[E(y_t^2)]^2} = \kappa_\varepsilon \exp(\sigma_h^2) \qquad [1.54]$$

respectivamente, donde σ_h^2 es la varianza de h_t, cuya expresión aparece en la ecuación [1.34], y κ_ε es la curtosis de ε_t. Si por ejemplo ε_t sigue una distribución

[23] Los resultados que obtengamos para este modelo serían igualmente válidos sobre el intervalo $-1/2<d<1/2$, donde el modelo sigue siendo estacionario e invertible.

Normal (κ_ε=3) y σ_η^2>0, el modelo LMSV produce una serie con exceso de curtosis (κ_y>3), recogiendo así una de las características típicas de las series financieras.

Recordemos que otra característica típica de los rendimientos de activos financieros, que fue precisamente la que motivó los modelos con memoria larga para la volatilidad, es que las series de los cuadrados de los rendimientos y los rendimientos absolutos muestran correlaciones que decaen muy despacio y tienden a persistir para retardos muy altos. Harvey (1998) demuestra que el modelo LMSV es capaz de reproducir este comportamiento de las correlaciones. En concreto, este autor deriva la función de autocorrelación de la serie $|y_t|^c$ para cualquier valor de c, incluidos valores no enteros, y comprueba que esta función refleja fielmente las características observadas empíricamente en los correlogramas de series financieras, incluso en los descritos por Ding, Granger y Engle (1993) para potencias fraccionales. Utilizando las fórmulas de Harvey (1998) para c=1 y c=2, se tienen sendas expresiones para las funciones de autocorrelación de las series $|y_t|$ e y_t^2, respectivamente:

$$\rho_{|y_t|}(k) = \rho(|y_t|,|y_{t+k}|) = \frac{\exp[\frac{1}{4}\sigma_h^2\rho_h(k)]-1}{\kappa_1\exp(\frac{1}{4}\sigma_h^2)-1}, \text{ para } k\geq 1 \qquad [1.55]$$

$$\rho_{y_t^2}(k) = \rho(y_t^2, y_{t+k}^2) = \frac{\exp[\sigma_h^2\rho_h(k)]-1}{\kappa_2\exp(\sigma_h^2)-1}, \text{ para } k\geq 1 \qquad [1.56]$$

donde σ_h^2 y $\rho_h(k)$ denotan la varianza y la función de autocorrelación de h_t, respectivamente, y κ_c es la constante:

$$\kappa_c = E(|\varepsilon_t|^{2c}) / [E(|\varepsilon_t|^c)]^2$$

Para la serie de los cuadrados (c=2), esta constante es la curtosis de ε_t, que toma el valor κ_2=3 si ε_t sigue una distribución N(0,1), y $\kappa_2 = 3\Gamma(\frac{\nu}{2}-2)\Gamma(\frac{\nu}{2})[\Gamma(\frac{\nu}{2}-1)]^{-2}$ si ε_t

tiene una distribución t_ν de Student. Para la serie de los valores absolutos (c=1), resulta que $\kappa_1=\pi/2$ cuando ε_t es N(0,1), y $\kappa_1=\frac{\pi}{2}\Gamma(\frac{\nu}{2}-1)\Gamma(\frac{\nu}{2})[\Gamma(\frac{\nu-1}{2})]^{-2}$ si ε_t tiene una distribución t_ν de Student.[24]

Al igual que ocurría en el modelo ARSV(1), cuando σ_h^2 es pequeño y/o $\rho_h(k)$ toma valores próximos a uno, se tiene la siguiente aproximación:

$$\rho_{y_t^2}(k) \cong \rho_h(k)\frac{\exp(\sigma_h^2)-1}{\kappa_2\exp(\sigma_h^2)-1}, \text{ para } k\geq 1$$

En el modelo LMSV, esto significa que la función de autocorrelación de y_t^2 decrecerá hiperbólicamente hacia cero, al igual que la función de autocorrelación de h_t, pero tomará, en general, valores más pequeños que ésta. Como veremos en el capítulo 2, esto va a dificultar la detección de la memoria larga en la volatilidad a través del correlograma de la serie y_t^2.

Al igual que ocurría con el modelo ARSV(1), el modelo LMSV en [1.52] resulta en general más sencillo de analizar transformándolo en un modelo lineal de la forma:

$$x_t = \ln(y_t^2) = \mu + h_t + \xi_t \qquad [1.57a]$$

$$(1-L)^d h_t = \eta_t, \qquad [1.57b]$$

donde $\mu=\ln(\sigma_*^2)+E[\ln(\varepsilon_t^2)]$ y $\xi_t=\ln(\varepsilon_t^2)-E[\ln(\varepsilon_t^2)]$ es un ruido blanco no gaussiano de media 0 y varianza σ_ξ^2, cuyas propiedades dependen de las de ε_t.[25] Si $d<1/2$, la serie x_t es estacionaria y, como veremos más adelante, si $0<d<1/2$ la serie x_t además tiene memoria larga.

---

[24] Véase Harvey (1998)

La ecuación [1.57a] representa x_t como la suma de dos procesos mutuamente independientes, un proceso lineal ARFIMA(0,d,0) y un ruido blanco ξ_t no gaussiano. Si d<1/2, el proceso ARFIMA es estacionario y, por tanto, las fórmulas [1.21]-[1.23] que derivamos en la sección 1.1.2 para calcular la varianza, la función de autocovarianza y la función de autocorrelación, respectivamente, de $x_t=\ln(y_t^2)$ en modelos SV estacionarios, aún son válidas para el modelo LMSV con d<1/2.

Según la ecuación [1.22], la función de autocovarianza de x_t coincide con la del proceso h_t, y por tanto, en el modelo LMSV estacionario la función de autocovarianza de x_t es la de un proceso ARFIMA(0,d,0), cuya expresión aparece en la ecuación [1.33], y se tiene:

$$\gamma_X(k) = \gamma_h(k) = \sigma_\eta^2 \frac{\Gamma(1-2d)\Gamma(k+d)}{\Gamma(d)\Gamma(1-d)\Gamma(k+1-d)}, k \geq 1 \qquad [1.58]$$

BIBLIOTECA VIRTUAL

donde Γ denota la función gamma. Por lo que respecta a la varianza de x_t, la ecuación [1.21] nos dice que ésta es igual a la suma de las varianzas del proceso h_t y de la perturbación ξ_t. Teniendo en cuenta la expresión de la varianza de h_t en [1.34] resulta que:

$$\gamma_X(0) = \mathrm{Var}(x_t) = \sigma_h^2 + \sigma_\xi^2 = \sigma_\eta^2 \frac{\Gamma(1-2d)}{[\Gamma(1-d)]^2} + \sigma_\xi^2 \qquad [1.59]$$

De la ecuación [1.58] se deduce que la serie x_t tiene la misma memoria que h_t, en el sentido de que las propiedades de memoria que se manifiestan en la función de autocovarianza de x_t son las mismas que las de h_t, ya que $\gamma_x(k)=\gamma_h(k)$ para $k\geq 1$. Por tanto, las autocovarianzas de x_t convergen hacia cero hiperbólicamente y no son absolutamente sumables, y en consecuencia, si 0<d<1/2, la serie x_t, al igual que h_t,

[25] Véase la sección 1.1.2.1 para las propiedades de ξ_t.

tiene memoria larga. Sin embargo existe una diferencia sustancial en el valor de la varianza de ambos procesos, ya que, según la ecuación [1.59], la varianza de x_t es mucho mayor que la de h_t debido al efecto adicional de la varianza del ruido ξ_t.

Las ecuaciones [1.58]-[1.59] se pueden resumir en la siguiente expresión:

$$\gamma_X(k) = \gamma_h(k) + \sigma_\xi^2 \, I_{\{k=0\}}$$

donde $I_{\{k=0\}}$ es la función indicador, que toma el valor 1 si k=0, y cero en el resto, y $\gamma_h(0)=\sigma_h^2$. Utilizando esta expresión, es inmediato probar que la función de autocovarianzas de la serie $x_t=\ln(y_t^2)$ en el modelo LMSV es la misma que en un modelo FIEGARCH con $g(\varepsilon_t)=\lambda_1\varepsilon_t+\lambda_2(|\varepsilon_t|-E|\varepsilon_t|)$ y $\lambda_2=0$; véase Breidt, Crato y de Lima (1998).

Dividiendo [1.58] entre [1.59], se obtiene la función de autocorrelación de x_t:

$$\rho_X(k) = \frac{\gamma_X(k)}{\gamma_X(0)} = \frac{\gamma_h(k)}{\sigma_h^2+\sigma_\xi^2} = \rho_h(k)\left\{1+\frac{\sigma_\xi^2}{\sigma_\eta^2}\frac{[\Gamma(1-d)]^2}{\Gamma(1-2d)}\right\}^{-1}, \text{ para } k\geq 1 \qquad [1.60]$$

donde $\rho_h(k)$ denota la función de autocorrelación de h_t y $\sigma_\xi^2=\mathrm{Var}(\xi_t)$. De la ecuación [1.60] se derivan varias conclusiones importantes. En primer lugar observamos que la función de autocorrelación de $x_t=\ln(y_t^2)$ es proporcional a la función de autocorrelación de h_t y, por tanto, heredará algunas de las propiedades de $\rho_h(k)$ sobre el intervalo $0<d<1/2$. En concreto, tomará siempre valores positivos, aunque más pequeños debido a ese factor de proporcionalidad, y decrecerá a un ritmo hiperbólico muy lento. Sin embargo, la función de autocorrelación de x_t, a diferencia de la de h_t, depende conjuntamente de los dos parámetros del modelo (d,σ_η^2) y esto va a dificultar, en muchas ocasiones, la estimación e identificación del modelo. Por ejemplo, para un valor d fijo, valores pequeños de σ_η^2 hacen que el cociente $\sigma_\xi^2/\sigma_\eta^2$

sea muy grande y en consecuencia, los valores de la función de autocorrelación de x_t pueden ser tan pequeños que sea realmente difícil distinguir la serie x_t de un ruido blanco, incluso para valores grandes de d (modelos muy persistentes).

Una expresión equivalente a [1.60] puede obtenerse sustituyendo $\rho_h(k)$ por su expresión en [1.35], en cuyo caso se tiene que:

$$\rho_X(1) = \frac{d}{1-d}\left\{1 + \frac{\sigma_\xi^2}{\sigma_\eta^2}\frac{[\Gamma(1-d)]^2}{\Gamma(1-2d)}\right\}^{-1}$$
$$\rho_X(k) = \frac{k-1+d}{k-d}\rho_X(k-1), \quad \text{para } k \geq 2 \qquad [1.61]$$

En esta expresión se aprecia claramente que en el intervalo 0<d<1/2, la función de autocorrelación de x_t toma valores siempre positivos y decrece hiperbólicamente hacia cero a partir del segundo retardo.

La representación lineal del modelo LMSV en [1.57] facilita asimismo el cálculo de la densidad espectral de la serie x_t, cuyo comportamiento también permite caracterizar la propiedad de memoria larga del modelo. En concreto, por la independencia de η_t y ε_t, de la ecuación [1.57a] se deduce que $f_X(\lambda) = f_h(\lambda) + f_\xi(\lambda)$, donde $f_X(\lambda)$, $f_h(\lambda)$ y $f_\xi(\lambda)$ representan las densidades espectrales de x_t, h_t y ξ_t respectivamente. Como la perturbación ξ_t es un ruido blanco con media cero y varianza σ_ξ^2, su densidad espectral es $f_\xi(\lambda) = \sigma_\xi^2/2\pi$. La expresión de la densidad espectral de h_t aparece en la ecuación [1.31]. La suma de ambas proporciona la densidad espectral de x_t, dada por la siguiente ecuación:

$$f_X(\lambda) = \frac{1}{2\pi}\{\sigma_\eta^2[2(1-\cos\lambda)]^{-d} + \sigma_\xi^2\} \qquad [1.62]$$

Para valores positivos de d esta función no está acotada en las bajas frecuencias, lo

que confirma que en el intervalo $0<d<1/2$ la serie x_t tiene la propiedad de memoria larga.

Cuando en el modelo LMSV, el parámetro d toma valores en el intervalo $1/2\leq d\leq 1$, el proceso h_t en [1.57b] ya no es estacionario, y por tanto tampoco lo es la serie x_t. Sin embargo, tomando primeras diferencias en [1.57a] se consigue un modelo estacionario para Δx_t de la forma:

$$\Delta x_t = \Delta h_t + \Delta \xi_t \qquad [1.63a]$$

$$\Delta h_t = (1-L)^{-d^*} \eta_t, \quad -1/2 \leq d^* \leq 0 \qquad [1.63b]$$

donde Δh_t es un ARFIMA(0,d*,0) sobre el intervalo $-1/2\leq d^*\leq 0$, con $d^*=d-1$, y por tanto estacionario.[26] La perturbación del modelo [1.63], $\Delta\xi_t$, tiene media cero y varianza $\{2\sigma_\xi^2\}$, pero ya no es ruido blanco, porque la autocovarianza de orden uno es $Cov(\Delta\xi_t, \Delta\xi_{t-1}) = -\sigma_\xi^2 \neq 0$.

El modelo LMSV básico definido en [1.52] puede generalizarse de forma que el proceso h_t en [1.52b] sea un proceso ARFIMA(p,d,q), de la forma:

$$(1-L)^d \phi(L) h_t = \theta(L)\eta_t \qquad [1.64]$$

donde los polinomios $\phi(L) = \sum_{j=1}^{p} \phi_j L^j$ y $\theta(L) = \sum_{i=1}^{q} \theta_i L^i$ tienen todas las raíces fuera del círculo unidad, y el parámetro d y la variable η_t son como en el caso anterior.

Las condiciones de estacionariedad y la estructura dinámica del nuevo modelo se obtienen de forma similar a las del modelo LMSV en [1.52], ya que muchas de las fórmulas descritas anteriormente aún son válidas para el modelo más

[26] Obsérvese que el mismo resultado se obtendría si $-1/2\leq d<3/2$, en cuyo caso Δh_t sería un ARFIMA(0,d*,0) con d* en el intervalo $-1/2<d^*<1/2$, donde el proceso sigue siendo estacionario.

general. Sin embargo, las expresiones de la varianza y la función de autocorrelación de las series y_t^2 y $\ln(y_t^2)$ son más complicadas. Por ejemplo, si consideramos p=1 y q=0 en la ecuación [1.64], el proceso h_t es un ARFIMA(1,d,0) y se obtiene el siguiente modelo:

$$y_t = \sigma_* \exp(\tfrac{1}{2} h_t)\, \varepsilon_t \qquad [1.65a]$$

$$(1-\phi L)(1-L)^d h_t = \eta_t \qquad [1.65b]$$

que denotaremos ARLMSV (AutoRegressive Long Memory Stochastic Volatility). Este modelo ha sido utilizado por Henry y Payne (1997) para modelizar la volatilidad de series de tipos de cambio observadas transacción a transacción, y por Breidt, Crato y de Lima (1998) y Chan y Petris (2000) para modelizar una serie diaria del índice SP500 de la bolsa de Nueva York. En ambos casos se obtiene evidencia de memoria larga en la volatilidad.

El modelo ARLMSV en [1.65] es estacionario si d<1/2 y $|\phi|<1$ y sus propiedades dinámicas se derivan fácilmente a partir de las fórmulas [1.53]-[1.56], que aún son válidas para este modelo. Sin embargo, en este caso, las fórmulas para calcular la varianza y la función de autocorrelación de h_t ya no son tan sencillas, y por ende tampoco serán sencillas las fórmulas correspondientes a las series y_t^2 y $\ln(y_t^2)$. Siguiendo la notación de Hosking (1981), la varianza y la función de autocorrelación del proceso ARFIMA(1,d,0) en [1.65b] admiten una expresión de la forma:

$$\sigma_h^2 = \sigma_{h'}^2 \frac{F(1,1+d;1-d;\phi)}{(1+\phi)}$$

$$\rho_h(k) = \rho_{h'}(k) \frac{\{F(1,d+k;1-d+k;\phi) + F(1,d-k;1-d-k;\phi) - 1\}}{(1-\phi)F(1,1+d;1-d;\phi)}$$

donde $\sigma_{h'}^2$ y $\rho_{h'}(k)$ denotan la varianza y la función de autocorrelación de orden k del proceso ARFIMA(0,d,0), y F es la función hipergeométrica definida como:

$$F(a,b;c;\phi)=1+\frac{a\,b}{c.1}\phi+\frac{a(a+1)\,b(b+1)}{c(c+1).1.2}\phi^2+\frac{a(a+1)(a+2)\,b(b+1)(b+2)}{c(c+1)(c+2).1.2.3}\phi^3+...$$

A pesar de lo complicadas que puedan parecer estas fórmulas, Sowell (1992a) y Chung (1994) proporcionan expresiones alternativas de $\rho_h(k)$ que pueden calcularse fácilmente y con gran precisión en un ordenador.

Uno de los aspectos interesantes del modelo ARLMSV en [1.65] es que incluye como casos particulares muchos de los modelos SV analizados anteriormente. En particular, si d=0 el modelo ARLMSV se convierte en el modelo ARSV(1) de la ecuación [1.14] y si ϕ=0 se obtiene el modelo LMSV básico definido en [1.52]. Sin embargo, el modelo SV con raíz unitaria puede obtenerse a partir del modelo ARLMSV mediante dos combinaciones diferentes de los parámetros: {ϕ=1,d=0} y {ϕ=0,d=1}. Esto puede tener consecuencias importantes para la identificación del modelo [1.65] cuando el proceso que define la volatilidad esté próximo a la frontera de la no estacionariedad.

La tabla 1.2 recoge los valores de la función de autocorrelación de la serie y_t^2 para los modelos ARSV(1), LMSV y ARLMSV, con diferentes valores de los parámetros. Éstos se han elegido de forma que la autocorrelación de orden uno sea igual en todos los modelos y que el cuadrado del coeficiente de variación de σ_t^2, $CV=Var(\sigma_t^2)/[E(\sigma_t^2)]^2=\{\exp(\sigma_h^2)-1\}$, sea igual a uno en todos ellos, con el fin de hacerlos comparables. Las autocorrelaciones de la tabla 1.2 muestran que la dinámica que generan en la serie de los cuadrados los distintos modelos analizados es muy similar cuando dichos modelos se acercan a las fronteras de la no estacionariedad. En esta tabla se puede observar además que existe un compromiso entre los parámetros ϕ y d, ya que ambos están midiendo la persistencia del modelo. Por ejemplo, la

funciones de autocorrelación de y_t^2 correspondientes a los modelos ARSV(1) con $\{\phi=0.98, \sigma_\eta^2=0.02745\}$ y ARLMSV con $\{\phi=0.8821, d=0.3, \sigma_\eta^2=0.02589\}$ sólo empiezan a distinguirse a partir de retardos muy alejados. Aunque no se presentan aquí, para la función de autocorrelación de la serie $\ln(y_t^2)$se obtienen resultados similares a éstos.

El gráfico 1.2 representa las funciones de autocorrelación de las series y_t^2 y $\ln(y_t^2)$ para los dos modelos seleccionados. En el gráfico 1.2(a) observamos que la función de autocorrelación de la serie y_t^2 tiene distinta forma en los modelos ARSV(1) y ARLMSV, especialmente en el largo plazo. Características similares se observan en el gráfico 1.2(b) para la serie $\ln(y_t^2)$. La función de autocorrelación de y_t^2 en el modelo ARSV(1) comienza tomando valores más grandes pero luego decrece rápidamente hacia cero. Sin embargo, en el modelo ARLMSV, aunque la función de autocorrelación decae en los primeros retardos, después apenas decrece y toma valores superiores a los correspondientes al modelo ARSV(1).

A pesar de estas diferencias, en la práctica no será fácil discernir entre ambos modelos debido, en primer lugar, a que las autocorrelaciones muestrales de las series y_t^2 y $\ln(y_t^2)$ en modelos con memoria larga tienen un sesgo considerable, como veremos en el capítulo 2, y en segundo lugar, a que la magnitud de dichas correlaciones será, en general, muy pequeña.

Tabla 1.2. Función de autocorrelación de y_t^2 para modelos ARSV(1), LMSV y ARLMSV con coeficiente de variación igual a uno y la misma autocorrelación de orden uno.

Modelo	ARSV(1)	ARLMSV	ARLMSV	ARLMSV	ARLMSV	LMSV
Parámetros	ϕ=0.9 d=0 σ_η^2=0.1317	ϕ=0.77094 d=0.2 σ_η^2=0.13038	ϕ=0.664 d=0.3 σ_η^2=0.12937	ϕ=0.46515 d=0.4 σ_η^2=0.12693	ϕ=0.23672 d=0.45 σ_η^2=0.11988	ϕ=0 d=0.473684 σ_η^2=0.10666
$\rho(1)$	0.17321	0.17321	0.17321	0.17321	0.17321	0.17321
$\rho(2)$	0.15064	0.14683	0.14688	0.15069	0.15744	0.16527
$\rho(3)$	0.13150	0.12441	0.12550	0.13478	0.14829	0.16072
$\rho(4)$	0.11516	0.10606	0.10889	0.12367	0.14235	0.15756
$\rho(5)$	0.10115	0.09121	0.09608	0.11569	0.13808	0.15515
$\rho(10)$	0.05468	0.04969	0.06325	0.09566	0.12616	0.14793
$\rho(50)$	0.00072	0.01378	0.02938	0.06517	0.10323	0.13263
$\rho(100)$	0.000004	0.00890	0.01636	0.05566	0.09490	0.12662
Parámetros	ϕ=0.95 d=0 σ_η^2=0.06758	ϕ=0.86243 d=0.2 σ_η^2=0.06635	ϕ=0.78565 d=0.3 σ_η^2=0.06556	ϕ=0.63785 d=0.4 σ_η^2=0.06527	ϕ=0.45902 d=0.45 σ_η^2=0.06443	ϕ=0 d=0.48718 σ_η^2=0.0539
$\rho(1)$	0.18637	0.18637	0.18637	0.18637	0.18637	0.18637
$\rho(2)$	0.17386	0.17073	0.17003	0.17104	0.17378	0.18209
$\rho(3)$	0.16235	0.15552	0.15436	0.15756	0.16429	0.17957
$\rho(4)$	0.15174	0.14146	0.14031	0.14644	0.15730	0.17780
$\rho(5)$	0.14195	0.12876	0.12804	0.13742	0.15205	0.17644
$\rho(10)$	0.10288	0.08336	0.08806	0.11164	0.13777	0.17227
$\rho(50)$	0.01096	0.02009	0.03733	0.07445	0.11219	0.16304
$\rho(100)$	0.00082	0.01278	0.02761	0.06341	0.10302	0.15925
Parámetros	ϕ=0.98 d=0 σ_η^2=0.02745	ϕ=0.92974 d=0.2 σ_η^2=0.02658	ϕ=0.8821 d=0.3 σ_η^2=0.02589	ϕ=0.7873 d=0.4 σ_η^2=0.02570	ϕ=0.6681 d=0.45 σ_η^2=0.02605	ϕ=0 d=0.49495 σ_η^2=0.02169
$\rho(1)$	0.19449	0.19449	0.19449	0.19449	0.19449	0.19449
$\rho(2)$	0.18917	0.18741	0.18679	0.18673	0.18744	0.19270
$\rho(3)$	0.18402	0.17980	0.17838	0.17854	0.18063	0.19163
$\rho(4)$	0.17904	0.17208	0.16988	0.17064	0.17458	0.19087
$\rho(5)$	0.17423	0.16445	0.16159	0.16329	0.16934	0.19029
$\rho(10)$	0.15237	0.13040	0.12669	0.13604	0.15247	0.18847
$\rho(50)$	0.05743	0.03620	0.05097	0.08713	0.12270	0.18434
$\rho(100)$	0.01926	0.02040	0.03689	0.07389	0.11250	0.18260
Parámetros	ϕ=0.9999 d=0 σ_η^2=1.38623e^{-4}	ϕ=0.99848 d=0.2 σ_η^2=1.2711e^{-4}	ϕ=0.99612 d=0.3 σ_η^2=1.1294e^{-4}	ϕ=0.9906 d=0.4 σ_η^2=0.9617e^{-4}	ϕ=0.98342 d=0.45 σ_η^2=0.8975e^{-4}	ϕ=0 d=0.499975 σ_η^2=1.08877e^{-4}
$\rho(1)$	0.19997	0.19997	0.19997	0.19997	0.19997	0.19997
$\rho(2)$	0.19994	0.19993	0.19992	0.19992	0.19991	0.19996
$\rho(3)$	0.19992	0.19988	0.19986	0.19984	0.19983	0.19996
$\rho(4)$	0.19989	0.19983	0.19978	0.19974	0.19973	0.19995
$\rho(5)$	0.19986	0.19977	0.19969	0.19963	0.19960	0.19995
$\rho(10)$	0.19972	0.19939	0.19913	0.19887	0.19880	0.19994
$\rho(50)$	0.19862	0.19477	0.19118	0.18792	0.18805	0.19992
$\rho(100)$	0.19725	0.18738	0.17835	0.17202	0.17486	0.19991

Gráfico 1.2. Función de autocorrelación teórica de las series y_t^2 y $\ln(y_t^2)$ en un modelo ARSV(1) con $\{\phi=0.98,\ \sigma_\eta^2=0.02745\}$ y en un modelo ARLMSV con $\{\phi=0.8821,\ d=0.3,\ \sigma_\eta^2=0.02589\}$.

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

1.3.3. Otros modelos

En esta sección se describen brevemente otros modelos con memoria larga para la volatilidad propuestos en la literatura econométrica, pero cuya relevancia en aplicaciones empíricas aún es escasa.

Ding, Granger y Engle (1993) analizan una serie diaria de rendimientos del índice SP500 y encuentran que las correlaciones muestrales de los rendimientos absolutos decrecen lentamente y son significativamente positivas incluso en retardos muy alejados. Un comportamiento similar se observa en transformaciones del tipo $|y_t|^d$, para distintos valores de d, siendo y_t la serie de los rendimientos. Para modelizar este comportamiento, los autores proponen una generalización del modelo GARCH(p,q) que denominan A-PARCH (Asymetric Power ARCH). El nuevo modelo viene definido por las siguientes ecuaciones:

$$y_t = \sigma_t \varepsilon_t$$

$$\sigma_t^{\delta} = \alpha_0 + \sum_{i=1}^{q} \alpha_1 (| y_{t-i} | - \gamma_i y_{t-i})^{\delta} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^{\delta}$$

donde ε_t es IID(0,1), $\alpha_0>0$, $\delta\geq 0$, $\alpha_i\geq 0$, i=1,…,q, $\beta_j\geq 0$, j=1,…,p y $|\gamma_i|<1$, i=1,…,q. Cuando $\gamma_i=0$ $\forall$i, y $\delta=2$, el modelo se reduce al GARCH(p,q) estándar. Valores de γ_i distintos de cero permiten modelizar la respuesta asimétrica de la volatilidad ante cambios de distinto signo en el valor de los rendimientos, al igual que ocurría en el modelo EGARCH de Nelson (1991). Ding, Granger y Engle (1993) establecen las condiciones de estacionariedad de y_t y proponen estimar los parámetros de este modelo por máxima verosimilitud, pero nada se sabe sobre las propiedades teóricas de estos estimadores. En realidad se desconocen también las propiedades teóricas del modelo en cuanto a estructura de correlación, curtosis, etc.

Ding y Granger (1996) extienden el trabajo anterior a otras series de rendimientos de activos financieros y encuentran propiedades similares en todas ellas. Estos autores proponen un modelo en el que σ_t^2 es una suma ponderada de N componentes, σ_{it}^2, i=1,2,…,N, cada una de las cuales es la varianza condicionada de un modelo GARCH(1,1) con parámetros $\{\alpha_i, \beta_i\}$. Suponiendo para los coeficientes β_i una distribución Beta(p,q), con p≠0, se demuestra que cuando N tiende a infinito, el modelo propuesto admite una representación ARCH(∞) cuyos coeficientes decrecen hiperbólicamente hacia cero; por tanto la varianza condicionada, σ_t^2, tiene memoria larga. Esta representación es muy similar a la del modelo FIGARCH de Baillie, Bollerslev y Mikkelsen (1996).

Robinson y Zaffaroni (1997) proponen una nueva clase de modelos, denominados modelos MA no lineales con memoria larga, en los que la volatilidad, σ_t, se modeliza mediante una estructura infinita de retardos de la propia perturbación ε_t. El modelo viene definido por las siguientes ecuaciones:

$$y_t = \sigma_t \varepsilon_t$$

$$\sigma_t = \rho + \sum_{j=1}^{\infty} \alpha_j \varepsilon_{t-j}$$

donde ρ es un escalar, la perturbación ε_t es IID$(0,\sigma^2)$, con momento de cuarto orden finito, y los coeficientes α_j verifican la condición $\alpha_j \sim Cj^{d-1}$, cuando $j \to \infty$, siendo C una constante y $0<d<1/2$. Estos modelos son capaces de producir incorrelación en los niveles de la serie y_t y memoria larga en la serie de sus cuadrados. En concreto, se prueba que la función de autocovarianza de y_t^2 admite una expresión de la forma:

$$\gamma_{y_t^2}(k) = \mathrm{Cov}(y_t^2, y_{t+k}^2) = \begin{cases} Ck^{4d-2} & \text{si} \rho = 0 \\ Ck^{2d-1} & \text{si} \rho \neq 0 \end{cases}$$

y por tanto, si $\rho=0$, la serie y_t^2 tiene memoria larga si $1/4<d<1/2$, y si $\rho\neq0$, la serie y_t^2 tiene memoria larga en todo el rango $0<d<1/2$.

En un trabajo posterior, Zaffaroni (1997) propone una generalización del modelo anterior y demuestra que, bajo ciertas condiciones, este modelo reproduce también otras propiedades típicas de las series de rendimientos de activos financieros, como el exceso de curtosis y la asimetría en la respuesta de la volatilidad a cambios de signo en los rendimientos. Además, este autor propone un método de estimación pseudo-máximo verosímil en el dominio de las frecuencias basado en la aproximación discreta de Whittle a la función de verosimilitud gaussiana de y_t, y demuestra que los estimadores que se obtienen son consistentes y asintóticamente normales.

Finalmente, Robinson y Zaffaroni (1998) proponen un modelo no lineal de volatilidad estocástica con dos perturbaciones, alternativo al modelo LMSV de Harvey (1998) y Breidt, Crato y de Lima (1998). En este modelo, la ecuación de la volatilidad de y_t se define como:

$$\sigma_t = \rho + \sum_{j=1}^{\infty} \alpha_j \eta_{t-j} \qquad [1.60]$$

donde η_t es $IID(0, \sigma_\eta^2)$ e independiente de ε_t, y con momento de cuarto orden finito. Robinson y Zaffaroni (1998) derivan las principales propiedades estadísticas de y_t en el dominio temporal y en el dominio de las frecuencias, y demuestran que, bajo ciertas condiciones sobre los momentos de ε_t, la serie y_t está incorrelacionada y tiene exceso de kurtosis, y la serie de sus cuadrados puede tener memoria larga. Para estimar los parámetros de este modelo, los autores proponen un estimador pseudo-máximo verosímil en el dominio de las frecuencias, y argumentan que no sería difícil establecer la consistencia y la distribución asintótica de dicho estimador, pero no presentan ninguna demostración formal de tales resultados.

Capítulo 2.

ESTIMACIÓN DE LA AUTOCOVARIANZA Y LA AUTOCORRELACIÓN EN MODELOS DE VOLATILIDAD ESTOCÁSTICA

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

2.1. Introducción

El correlograma o función de autocorrelación muestral es una de las herramientas más utilizadas en la identificación y validación de modelos de series temporales. Concretamente, en la modelización de la volatilidad, el comportamiento del correlograma de la serie de los cuadrados suele utilizarse como un primer indicio de la dependencia temporal en los momentos de segundo orden y de la necesidad de modelos no lineales. Es más, en los modelos tipo GARCH(p,q), el correlograma de la serie de los cuadrados puede utilizarse, en teoría, para identificar los valores (p,q) y el correlograma de los cuadrados de los residuos se utiliza para validar el modelo correspondiente.[1] Lo mismo podría decirse del modelo SV autorregresivo de orden uno, donde la función de autocorrelación de la serie del logaritmo de los cuadrados es la de un ARMA(1,1), según vimos en el capítulo 1.

En modelos lineales, es bien sabido que tanto la covarianza como la correlación muestral son estimadores sesgados de sus correspondientes funciones poblacionales. Fuller (1996) propone fórmulas aproximadas para calcular el sesgo y la varianza de ambos estadísticos en modelos lineales con coeficientes absolutamente sumables y perturbaciones independientes, y con ciertas condiciones sobre los momentos de dichas perturbaciones. Las expresiones de la varianza asintótica pueden encontrarse también en Brockwell y Davies (1991). Kendall, Stuart y Ord (1968) calculan el sesgo de las correlaciones muestrales en procesos AR(1) y MA(1) gaussianos. De Gooijer (1980) obtiene el valor exacto de la esperanza y la varianza de las autocorrelaciones muestrales en modelos ARMA y ARIMA gaussianos y, para determinados modelos, compara estos valores con los que se obtendrían utilizando las aproximaciones de Marriot y Pope (1954) y Kendall (1954).

Bajo condiciones mucho más generales, Anderson (1971) deriva expresiones del sesgo y de la varianza de la covarianza muestral para procesos débilmente estacionarios, no necesariamente lineales, y demuestra que, bajo ciertas condiciones sobre la función de autocovarianza poblacional, el sesgo es de orden $O(T^{-1})$.[2]

Por lo que respecta a la distribución asintótica de las covarianzas muestrales, Anderson (1971), Brockwell y Davies (1991) y Fuller (1996) presentan varios resultados para modelos lineales, bajo diferentes supuestos sobre los coeficientes y las perturbaciones del modelo. Hannan (1976) extiende estos resultados al caso multivariante y obtiene la distribución asintótica de las covarianzas muestrales bajo condiciones muy generales sobre las perturbaciones.

La distribución asintótica de las correlaciones muestrales para procesos lineales univariantes con perturbaciones I.I.D. se deriva en Anderson y Walker (1964). Hannan y Heyde (1972) y Anderson (1992) generalizan este resultado a

---

[1] Ver Bollerslev (1986).

[2] Se dice que una sucesión de números reales $\{a_n\}$ es a lo sumo de orden $g_n>0$, y se denota por $a_n=O(g_n)$, si existe un número real M tal que $\{|a_n|/g_n\}\leq M$ para todo n.

modelos lineales en los que las perturbaciones son una martingala en diferencias, y derivan la distribución asintótica bajo diferentes supuestos sobre los momentos de dichas perturbaciones. Roy (1989) deriva la distribución asintótica de las correlaciones muestrales en modelos lineales multivariantes, extendiendo así los resultados de Hannan (1976) para las covarianzas muestrales.

En el ámbito de los modelos ARFIMA estudiados en la sección 1.2, no existen muchos trabajos que analicen el comportamiento de las autocorrelaciones muestrales. Newbold y Agiakloglou (1993) utilizan la fórmula aproximada del sesgo de las correlaciones muestrales propuesta por Marriot y Pope (1954), y estudian su validez en muestras finitas bajo hipótesis de normalidad. Hosking (1996) analiza el comportamiento de la media, la covarianza y la correlación muestral en modelos estacionarios ($0<d<1/2$) y obtiene expresiones del sesgo asintótico de la covarianza y la correlación muestral, así como sus distribuciones asintóticas. Sus fórmulas confirman los resultados de las simulaciones de Newbold y Agiakloglou (1993), en el sentido de que existe siempre un sesgo negativo en las correlaciones muestrales, nada despreciable, y mayor cuanto más persistente es el modelo y cuanto mayor es el orden de la correlación que se estima. Esto significa que en modelos ARFIMA el correlograma subestima la función de autocorrelación teórica y, por tanto, será difícil reconocer la memoria larga a través del correlograma. He (1996) deriva, bajo condiciones diferentes a las de Hosking (1996), la distribución asintótica de las correlaciones muestrales para modelos ARFIMA(0,d,0) con $d<1/4$.

Por lo que respecta a los modelos para series temporales heteroscedásticas estudiados en la sección 1.1, tampoco existe mucha literatura en la que se estudie el comportamiento de las correlaciones muestrales. Milhoj (1985) y Bollerslev (1988) derivan la función de autocorrelación de la serie de los cuadrados y establecen las condiciones de existencia de los momentos de cuarto orden en modelos ARCH(q) y GARCH(1,1), respectivamente. He y Terasvirta (1999b) y Karanasos (1999) extienden estos resultados al modelo GARCH(p,q). Milhoj (1985) analiza además el efecto que tiene la heteroscedasticidad condicionada sobre las propiedades

asintóticas de las autocorrelaciones muestrales de la serie en niveles, y demuestra que, en presencia de efectos ARCH, el contraste de incorrelación de las bandas de Barlett puede llevar a conclusiones erróneas. Diebold (1988) extiende estos resultados al contraste de incorrelación de Box-Pierce y propone una corrección por heteroscedasticidad. Sin embargo, ninguno de estos autores presenta resultados teóricos sobre el comportamiento de los momentos muestrales de la serie de los cuadrados. Bollerslev (1988) observa, a través de experimentos de Monte Carlo, la aparición de un sesgo negativo sistemático en las correlaciones muestrales de dicha serie para ciertos modelos ARCH(2) y GARCH(1,1). He y Terasvirta (1999a) obtienen el mismo resultado en distintos modelos GARCH(1,1), incluso con series de 100000 observaciones. Li y Mak (1994) derivan la distribución asintótica de las correlaciones muestrales de los cuadrados de los residuos en modelos ARCH.

Por lo que respecta a los modelos de volatilidad estocástica, Wright (1999) demuestra, a partir de los resultados de Hannan (1976), que las autocovarianzas muestrales de la serie $\ln(y_t^2)$ en el modelo LMSV con $d<1/4$, son asintóticamente normales, y proporciona la expresión de la varianza asintótica de dicha distribución. Sin embargo, el comportamiento de las correlaciones muestrales de series generadas por otros modelos SV, y de las series de sus cuadrados y del logaritmo de sus cuadrados, no ha sido analizado anteriormente en la literatura y éste es el objetivo central de este capítulo. En primer lugar, presentaremos las propiedades de las correlaciones muestrales de la propia serie y_t y las de los contrastes que de ellas se derivan. En segundo lugar, analizaremos el comportamiento de las correlaciones muestrales de la serie de sus cuadrados. En este caso, obtendremos previamente las propiedades de las covarianzas muestrales, que resultan imprescindibles para poder derivar después resultados sobre las correlaciones muestrales. Dada la dificultad que supone obtener resultados teóricos para la serie y_t^2, debido a la no linealidad de los modelos SV, nos hemos centrado en las correlaciones de la serie $\ln(y_t^2)$, que es además la serie que se obtiene al aplicar a los modelos SV la transformación logarítmica que permite obtener su representación lineal en el espacio de los estados.

El capítulo se organiza como sigue. En la sección 2.2 se consideran las propiedades de la función de autocorrelación muestral de la serie y_t y del estadístico de Box-Pierce para contrastar incorrelación en dicha serie en modelos SV. Siguiendo el desarrollo de Diebold (1988), probamos que, en presencia de volatilidad estocástica, estos estadísticos no tienen la distribución habitual y proponemos una corrección por volatilidad. En la sección 2.3 se derivan el sesgo y la varianza asintóticos de las covarianzas muestrales de la serie $\ln(y_t^2)$ en el modelo ARSV(1), y se obtiene su distribución asintótica. Por lo que respecta al modelo LMSV, a partir de los resultados de Hosking (1996), se obtiene una expresión del sesgo de la covarianza muestral de la serie $\ln(y_t^2)$, y se presenta el resultado de Wright (1999) sobre la distribución asintótica de las covarianzas muestrales de dicha serie en modelos con $d<1/4$. Se comprueba además que las fórmulas de Hosking (1996) para la varianza asintótica de las covarianzas muestrales no pueden aplicarse directamente en modelos LMSV. En la sección 2.4 se presentan las fórmulas del sesgo asintótico y la desviación estándar de las correlaciones muestrales de la serie $\ln(y_t^2)$ en el modelo ARSV(1), y se deriva su distribución asintótica. Se prueba además que los resultados de Hosking (1996) sobre correlaciones muestrales tampoco son directamente aplicables al modelo LMSV y se deriva la distribución asintótica de las correlaciones muestrales en el modelo LMSV con $d<1/4$. En cada sección, los resultados asintóticos se complementan con los resultados de un estudio de Monte Carlo para investigar su validez en muestras finitas.

Antes de entrar en cada una de estas secciones, describimos a continuación la notación que se va emplear en este capítulo.

Dado un proceso x_t estacionario con media μ y función de autocovarianza $\gamma(k)$, un posible estimador de $\gamma(k)$ es la autocovarianza muestral de orden k dada por:

$$c^*(k)=\frac{1}{T}\sum_{t=1}^{T-k}(x_t-\mu)(x_{t+k}-\mu) \qquad [2.1]$$

En general, la media, μ, es desconocida y debe sustituirse por algún estimador. Se obtiene así otro posible estimador de $\gamma(k)$:

$$c(k) = \frac{1}{T}\sum_{t=1}^{T-k}(x_t - \overline{X})(x_{t+k} - \overline{X}) \qquad [2.2]$$

donde $\overline{X}$ es la media muestral de todas las observaciones.[3] También suele utilizarse como estimador de $\gamma(k)$ otra versión del estadístico anterior de la forma:

$$\tilde{c}(k) = \frac{1}{T-k}\sum_{t=1}^{T-k}(x_t - \overline{X})(x_{t+k} - \overline{X}) \qquad [2.3]$$

Para k=0, se obtiene, en ambos casos, la siguiente expresión de la varianza muestral:

$$c(0) = \tilde{c}(0) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \overline{X})^2$$

A partir de estos estimadores de la covarianza muestral se definen de forma inmediata los estimadores de la función de autocorrelación de x_t, definida como $\rho(k)=\gamma(k)/\gamma(0)$. Cuando la media del proceso es conocida, un posible estimador de $\rho(k)$ es $r^*(k)=c^*(k)/c^*(0)$. En caso contrario, $\rho(k)$ puede estimarse utilizando la correlación muestral de orden k, r(k), definida como:

$$r(k)=\frac{c(k)}{c(0)} = \frac{\sum_{t=1}^{T-k}(x_t - \overline{X})(x_{t+k} - \overline{X})}{\sum_{t=1}^{T}(x_t - \overline{X})^2} \qquad [2.4]$$

[3] Existen otras formas de definir la covarianza muestral, por ejemplo, utilizando como estimador de μ la media de T-k observaciones. Véase Anderson (1971, p.440).

Salvo que se indique lo contrario, siempre que hablemos de la correlación muestral nos estaremos refiriendo a esta última definición.

2.2. Contrastes de incorrelación serial en presencia de volatilidad estocástica

Dada una serie temporal $\{y_t\}_{t=1,\ldots,T}$, los contrastes habituales de incorrelación serial se basan en las propiedades de las autocorrelaciones muestrales de dicha serie. Si las variables y_t son independientes e igualmente distribuidas (IID) con media cero y varianza constante, se demuestra[4] que la correlación muestral de orden k de la serie y_t, que denotaremos $r_Y(k)$, se distribuye asintóticamente como una variable normal con media cero y varianza:

$$\mathrm{Var}(r_Y(k)) = \frac{T-k}{T(T+2)} \sim \frac{1}{T}$$

Utilizando esta última aproximación se obtienen las habituales bandas de confianza de Barlett al 95%, $\pm 1.96/\sqrt{T}$, y se construye el estadístico de Box-Pierce,

$$Q(K)=T\sum_{j=1}^{K}[r_Y(j)]^2 \qquad [2.5]$$

que permite contrastar la significación conjunta de las K primeras autocorrelaciones. Bajo la hipótesis nula de incorrelación, el estadístico Q(K) se distribuye asintóticamente como una variable χ^2 con K grados de libertad.

[4] Véase por ejemplo Brockwell y Davies (1991, pp. 222-223).

Sin embargo, la presencia de heteroscedasticidad condicionada en la serie y_t invalida estos resultados, porque la hipótesis de independencia deja de ser cierta. Una posible solución a este problema es parametrizar la forma de la heteroscedasticidad y desarrollar contrastes de incorrelación serial para esa parametrización. Diebold (1988) adopta esta solución modelizando la heteroscedasticidad con modelos ARCH. Utilizando los resultados de Milhoj (1985) sobre la varianza asintótica de las correlaciones muestrales, Diebold (1988) demuestra que las bandas de confianza para contrastar incorrelación en modelos ARCH son más anchas que las habituales $\pm 1.96/\sqrt{T}$ y la distribución del estadístico de Box-Pierce tampoco es la habitual χ^2. Este autor propone correcciones de ambos estadísticos que tengan en cuenta los efectos ARCH.[5]

Como vimos en el capítulo 1, la familia de modelos tipo ARCH no es la única parametrización posible en la modelización de series heteroscedásticas, y los modelos SV constituyen una alternativa interesante, cuyos efectos sobre los contrastes de incorrelación serial también merecen ser estudiados. En esta sección analizamos, en primer lugar, las propiedades de las correlaciones muestrales de y_t en modelos SV y demostramos que su varianza asintótica es mayor que 1/T. En consecuencia, proponemos unas bandas de confianza corregidas por volatilidad para contrastar incorrelación serial en este tipo de modelos, y comprobamos que estas correcciones producen resultados satisfactorios. En segundo lugar, estudiamos las propiedades del estadístico de Box-Pierce y probamos que en modelos SV su distribución asintótica no es la habitual χ^2, y proponemos una corrección por volatilidad que sí tiene tal distribución.

[5] Las correcciones de Diebold (1988) mantienen el "espíritu" del estadístico de Box-Ljung, en el sentido de que para mejorar el contraste de Box-Pierce, se corrige $r_Y(k)$ por una aproximación a su varianza asintótica mejor que 1/T. Recordemos que el estadístico de Box-Ljung mejora los resultados de Box-Pierce en procesos IID porque corrige las correlaciones muestrales por (T-k)/[T(T+2)] que es una aproximación a la varianza asintótica de $r_Y(k)$ más precisa que 1/T.

2.2.1. Efectos SV sobre las correlaciones muestrales

En esta sección demostramos que, en presencia de volatilidad estocástica, la distribución asintótica de las correlaciones muestrales de la serie y_t no es la habitual N(0,1/T). La siguiente proposición establece el resultado principal de esta sección, a partir del cual se derivarán los demás resultados.

Proposición 2.1. Sea el modelo de volatilidad estocástica ARSV(1) estacionario definido por las siguientes ecuaciones:

$$y_t = \sigma_* \exp(\tfrac{1}{2} h_t)\varepsilon_t \qquad [2.6a]$$

$$h_t = \phi\, h_{t-1} + \eta_t \qquad [2.6b]$$

donde σ_* es un factor de escala, ε_t es IID con media cero y varianza unitaria, η_t es NID$(0, \sigma_\eta^2)$ e independiente de ε_t y $|\phi|<1$. Si ε_t tiene distribución simétrica y momento de cuarto orden finito, la correlación muestral de orden k de la serie y_t es asintóticamente normal con media cero y varianza:

$$C_1(T,k) = \frac{1}{T}\exp\{\gamma_h(k)\} \qquad [2.7]$$

donde $\gamma_h(k)$ es la autocovarianza de orden k del proceso h_t.

Demostración:

En la demostración seguiremos los pasos de Milhoj (1985) y Kramer y Runde (1994) que demuestran este mismo resultado para modelos ARCH.

En el capítulo 1 vimos que la serie y_t del modelo ARSV(1) definido en las ecuaciones [2.6], es una martingala en diferencias y, por tanto, es una serie con

media cero ($E(y_t)=0$) e incorrelacionada ($Cov(y_t,y_{t+k})=0$, para $k>0$). Por ser la media de y_t conocida e igual a cero, utilizaremos la correlación muestral definida como:

$$r_Y^*(k) = c_Y^*(k)/c_Y^*(0)$$

donde $c_Y^*(k)$ es la covarianza muestral con media conocida, definida en [2.1]:

$$c_Y^*(k) = \frac{1}{T}\sum_{t=k+1}^{T} y_t y_{t-k}$$

Para obtener la distribución asintótica de $r_Y^*(k) = c_Y^*(k)/c_Y^*(0)$ utilizaremos un resultado clásico de teoría asintótica, en virtud del cual, si $X_n \xrightarrow{d} X$ y $a_n \xrightarrow{p} a$, entonces $a_n X_n \xrightarrow{d} aX$, donde $\xrightarrow{d}$ denota convergencia en distribución y $\xrightarrow{p}$ denota convergencia en probabilidad. En nuestro caso, obtendremos en primer lugar la distribución asintótica de $c_Y^*(k)$ y después probaremos que $c_Y^*(0) \xrightarrow{p} \gamma_Y(0)$, donde $\gamma_Y(0)$ es la varianza marginal de y_t.

Para obtener la distribución asintótica de $c_Y^*(k)$, consideremos la variable:

$$\sqrt{T}c_Y^*(k) = \frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} y_t y_{t-k} = \frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} U_t$$

donde $U_t=y_t y_{t-k}$, con $k>0$. En el modelo ARSV(1), se comprueba fácilmente que el proceso U_t es una martingala en diferencias, porque $E_{t-1}(U_t)=0$, donde E_{t-1} denota la esperanza condicionada al pasado de la serie hasta el instante anterior. Por tanto, resulta que $E(U_t)=0$. Además, si la distribución de ε_t es simétrica, se demuestra que $E(y_{t_1}^{n_1} \ldots y_{t_m}^{n_m})=0$ con $t_i \neq t_j$, si al menos uno de los exponentes n_i es impar (ver Milhoj, 1985). Por tanto, si existe el momento de orden cuatro de y_t, se verifica que:

$$\text{Cov}(U_t,U_{t+r})=E(U_tU_{t+r})=\begin{cases}E(y_t y_{t-k} y_{t+r} y_{t+r-k}) = 0 & \text{si } r \neq 0\\ E(U_t^2) = E(y_t^2 y_{t-k}^2) = [\gamma_Y(0)]^2 + \gamma_{Y^2}(k) & \text{si } r = 0\end{cases} \qquad [2.8]$$

donde $\gamma_Y(0)$ es la varianza de y_t, $\gamma_Y(0)=\text{Var}(y_t)=E(y_t^2)$, y $\gamma_{Y^2}(k)$ denota la función de autocovarianza de orden k de la serie de los cuadrados, $\gamma_{Y^2}(k)=\text{Cov}(y_t^2, y_{t-k}^2)$. En consecuencia, U_t es un proceso estacionario incorrelacionado con media igual a cero. Además, en virtud del corolario 6.1.1.2 de Fuller (1996), el proceso U_t es ergódico.

Con estas condiciones sobre U_t, se puede aplicar el Teorema Central del Límite para Martingalas de Billingsley (1961), en virtud del cual resulta que:

$$\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} U_t \xrightarrow{d} N(0, E(U_t^2))$$

Sustituyendo en esta expresión $E(U_t^2)$ por su valor en [2.8] se obtiene la siguiente distribución asintótica:

$$\sqrt{T}c_Y^*(k) = \frac{1}{\sqrt{T}}\sum_{t=k+1}^{T} U_t \xrightarrow{d} N(0, [\gamma_Y(0)]^2 + \gamma_{Y^2}(k)) \qquad [2.9]$$

A continuación probaremos que en el modelo ARSV(1), la varianza muestral de y_t con media conocida e igual a cero, $c_Y^*(0)$, es un estimador consistente de la varianza poblacional, $\gamma_Y(0)$, o lo que es lo mismo, que:

$$c_Y^*(0) = \frac{1}{T}\sum_{t=1}^{T} y_t^2 \xrightarrow{p} \gamma_Y(0)=E(y_t^2) \qquad [2.10]$$

Para ello, basta probar que $c_Y^*(0)$ converge en media cuadrática hacia $\gamma_Y(0)$, ya que la convergencia en media cuadrática implica la convergencia en probabilidad. Como

además $c_Y^*(0)$ es un estimador insesgado de $\gamma_Y(0)$, basta probar que:

$$\lim_{T\to\infty} \mathrm{Var}(c_Y^*(0)) = \lim_{T\to\infty} \mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} y_t^2\right) = 0 \qquad [2.11]$$

Este resultado es inmediato a partir del corolario 6.1.1.1 de Fuller (1996), que establece la convergencia hacia cero de la varianza de la media muestral para series estacionarias cuya función de autocovarianza converja hacia cero. En nuestro caso, la serie y_t^2 es estacionaria, según vimos en el capítulo 1, y su función de autocovarianza, definida en [1.18], converge hacia cero siempre que converja hacia cero $\gamma_h(k)$, donde $\gamma_h(k)$ es la función de autocovarianza del proceso subyacente h_t. Como h_t es un proceso AR(1) estacionario, esta condición se verifica y, por tanto, en virtud del corolario 6.1.1.1 de Fuller (1996), se verifica la condición [2.11], y queda probada la convergencia en probabilidad en [2.10].

La aplicación conjunta de [2.9] y [2.10] permite demostrar que:

$$\sqrt{T} r_Y^*(k) = \frac{\sqrt{T}\, c_Y^*(k)}{c_Y^*(0)} \xrightarrow{d} N(0, 1 + \frac{\gamma_{Y^2}(k)}{[\gamma_Y(0)]^2})$$

y por tanto, podemos concluir que $r_Y^*(k)$ es asintóticamente normal de media cero y varianza:

$$\mathrm{Var}(\, r_Y^*(k)\,) \sim \frac{1}{T}\left\{1 + \frac{\gamma_{Y^2}(k)}{[\gamma_Y(0)]^2}\right\} \qquad [2.12]$$

Esta expresión de la varianza asintótica de $r_Y^*(k)$ es la misma que obtienen Milhoj (1985) y Kramer y Runde (1994) para modelos ARCH.

El resultado que acabamos de probar, se seguiría verificando si suponemos que la media de y_t no es conocida y utilizamos la correlación muestral con la media

estimada, $r_Y(k)=c_Y(k)/c_Y(0)$, definida en [2.4]. La razón de que esto sea así es que, según el teorema 1 de Hannan y Heyde (1972), si la media muestral de y_t converge casi seguro hacia $E(y_t)=0$, el comportamiento asintótico de $c_Y(k)$ y $c_Y^*(k)$ es el mismo y, en consecuencia, lo es el de $r_Y(k)$ y $r_Y^*(k)$. En el modelo ARSV(1), la serie y_t es una martingala en diferencias y, por tanto, es una serie estacionaria e incorrelacionada con media cero. En virtud del mismo teorema 1 de Hannan y Heyde (1972), resulta que la media muestral de y_t, $\overline{Y}$, converge casi seguro hacia $E(y_t)=0$ y, por tanto, se prueba que $r_Y(k)$ y $r_Y^*(k)$ tienen la misma distribución asintótica.

En la expresión [2.12] observamos que el valor de la varianza asintótica de $r_Y(k)$ en modelos SV difiere de su valor habitual 1/T en procesos IID. Observamos además que esa diferencia depende únicamente de la varianza de la serie y_t, $\gamma_Y(0)$, y de la función de autocovarianza de la serie de sus cuadrados, $\gamma_{Y^2}(k)$. Las expresiones de $\gamma_Y(0)$ y $\gamma_{Y^2}(k)$ para modelos SV estacionarios aparecen en las ecuaciones [1.14] y [1.18], respectivamente. Utilizando los valores que proporcionan dichas ecuaciones, se prueba que el factor de corrección por volatilidad estocástica para $Var(r_Y(k))$ en modelos SV estacionarios admite la siguiente expresión:

$$1+\frac{\gamma_{Y^2}(k)}{[\gamma_Y(0)]^2}=1+\frac{\sigma_*^4\exp(2\mu_h+\sigma_h^2)[\exp\{\gamma_h(k)\}-1]}{\sigma_*^4\exp(2\mu_h+\sigma_h^2)}=\exp\{\gamma_h(k)\} \qquad [2.13]$$

donde μ_h, σ_h^2 y $\gamma_h(k)$ denotan la media, la varianza y la autocovarianza de orden k del proceso h_t, respectivamente. Sustituyendo [2.13] en [2.12] se obtiene la expresión de la varianza asintótica de $r_Y(k)$ en [2.7] y la demostración concluye. ■

Obsérvese que en la demostración precedente, las únicas condiciones del modelo ARSV(1) que hemos utilizado han sido la estacionariedad de y_t, la simetría de ε_t y la existencia del momento de cuarto orden de ε_t, y la condición $\gamma_h(k)\to 0$ cuando

$k\rightarrow\infty$. El modelo LMSV estacionario ($0<d<1/2$) verifica igualmente estas condiciones y, por tanto, de la proposición 2.1 se deduce el siguiente corolario.

Corolario 2.1: Sea el modelo de volatilidad estocástica LMSV estacionario definido por las siguientes ecuaciones:

$$y_t = \sigma_* \exp(\tfrac{1}{2} h_t)\varepsilon_t \qquad [2.14a]$$

$$(1-L)^d h_t = \eta_t \qquad [2.14b]$$

donde $0<d<1/2$, y σ_*, ε_t y η_t son como en la proposición 2.1. Entonces, la correlación muestral de orden k de la serie y_t, se distribuye asintóticamente como una variable normal con media cero y varianza [2.7].

Una consecuencia inmediata de los resultados anteriores es que el intervalo de confianza al 95% para contrastar incorrelación de orden k en modelos ARSV(1) y LMSV estacionarios, ya no es el habitual $\pm 1.96/\sqrt{T}$, sino este otro:

$$\pm 1.96[C_1(T,k)]^{1/2} = \pm\frac{1.96}{\sqrt{T}}\exp\{0.5\gamma_h(k)\} \qquad [2.15]$$

que corresponde a corregir $r_Y(k)$ por su varianza asintótica en [2.7].

De la expresión [2.15] se deduce, en primer lugar, que las bandas de confianza del contraste de incorrelación en presencia de volatilidad estocástica son diferentes para cada valor de k, y, en segundo lugar, que si $\gamma_h(k)$ es estrictamente mayor que cero, estas bandas son más anchas que las habituales bandas de Barlett, $\pm 1.96/\sqrt{T}$. Por tanto, éstas pueden llevarnos, en ciertas ocasiones, a tomar la decisión de rechazar erróneamente la hipótesis de incorrelación en modelos SV. Este efecto es especialmente importante en los primeros retardos, en los que la anchura de

las bandas [2.15] será mayor, pero se desvanece a medida que aumenta el orden de la correlación que se estima. De hecho, cuando $k\to\infty$, resulta que:

$$\lim_{k\to\infty} C_1(T,k) = \lim_{k\to\infty} \frac{1}{T}\exp\{\gamma_h(k)\} = \frac{1}{T} \qquad [2.16]$$

porque $\gamma_h(k)\to 0$ cuando $k\to\infty$ y, por tanto, para retardos grandes, las bandas de confianza $\pm 1.96/\sqrt{T}$ estarán bastante próximas a las reales, $\pm 1.96[C_1(T,k)]^{1/2}$, especialmente si la convergencia de $\gamma_h(k)$ hacia cero es rápida.

Por ejemplo, en el modelo ARSV(1) estacionario definido por las ecuaciones [2.6], el proceso h_t es un AR(1) cuya función de autocovarianza es $\gamma_h(k) = \frac{\sigma_\eta^2}{1-\phi^2}\phi^k$. Para valores de $\phi>0$, resulta que $\gamma_h(k)>0$ y en consecuencia $\exp\{0.5\gamma_h(k)\}>1$ y el intervalo de confianza $\pm 1.96/\sqrt{T}$ resulta demasiado "pequeño" comparado con el verdadero intervalo en [2.15], especialmente si ϕ toma valores próximos a uno.

En el modelo LMSV estacionario definido en [2.14], el proceso h_t es un ARFIMA(0,d,0) estacionario ($0<d<1/2$) cuya función de autocovarianza es $\gamma_h(k)= \sigma_\eta^2 \frac{\Gamma(1-2d)\Gamma(k+d)}{\Gamma(d)\Gamma(1-d)\Gamma(k+1-d)} \sim \sigma_\eta^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} k^{2d-1}$. Por tanto, $\gamma_h(k)$ es siempre positivo y las consecuencias de utilizar el intervalo $\pm 1.96/\sqrt{T}$ son las mismas que en el caso anterior. Sin embargo, la memoria larga asociada al modelo LMSV hace que la convergencia en [2.16] sea más lenta que en el modelo ARSV(1), debido a que $\gamma_h(k)$ converge hacia cero a un ritmo hiperbólico muy lento.

Estas características que acabamos de describir, se aprecian igualmente en los gráficos 2.1-2.4. El gráfico 2.1 y el gráfico 2.2 representan las bandas de confianza para contrastar incorrelación en modelos ARSV(1) con $\phi=0.95$ y $\phi=0.98$, respectivamente, para tres valores de la varianza, $\sigma_\eta^2=\{0.1, 0.05, 0.01\}$, y T=512. Los gráficos 2.3 y 2.4 representan las mismas bandas para modelos LMSV con d=0.45 y d=0.49,

respectivamente. En cada gráfico se representan las bandas habituales de Barlett, $\pm 1.96/\sqrt{T}$, junto con las bandas corregidas por volatilidad definidas en [2.15].

En estos gráficos observamos que la discrepancia entre las bandas corregidas y sin corregir es diferente en cada retardo y que las mayores divergencias se dan en los primeros retardos y en los modelos con mayor varianza. Sin embargo, estas diferencias disminuyen a medida que aumenta el orden de la correlación cuya significación se contrasta.

En el modelo ARSV(1) más persistente y con mayor varianza (ϕ=0.98, σ_η^2=0.1), las bandas verdaderas distan mucho de las bandas de Barlett, incluso en los retardos más alejados, y, por tanto, en este caso, el contraste habitual $\pm 1.96/\sqrt{T}$ detectaría en muchas ocasiones correlaciones espurias en y_t. En la práctica, este problema no es tan grave porque los valores habituales de σ_η^2 en las aplicaciones con datos reales son más pequeños, y el error que se comete con las bandas $\pm 1.96/\sqrt{T}$ no es muy grande.

En los modelos LMSV, las discrepancias entre las bandas corregidas y sin corregir no son tan grandes como en el modelo ARSV(1). De hecho, en el modelo con d=0.45 (gráfico 2.3) el error que se comete con las bandas de Barlett es despreciable. Sin embargo, a diferencia de lo que ocurre en el modelo autorregresivo, las diferencias entre las bandas corregidas y sin corregir apenas disminuyen al aumentar el retardo. Esto se debe precisamente a la propiedad de memoria larga, caracterizada porque la convergencia de $\gamma_h(k)$ hacia cero es muy lenta y en consecuencia también es muy lenta la convergencia en [2.16].

Gráfico 2.1. Bandas de confianza de Barlett (línea continua) y bandas corregidas por volatilidad para un modelo ARSV(1) con ϕ=0.95 y σ_{η}^2 =0.01 (····), σ_{η}^2 =0.05 (- - - -), σ_{η}^2 =0.1 (– – –).

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Gráfico 2.2. Bandas de confianza de Barlett (línea continua) y bandas corregidas por volatilidad para un modelo ARSV(1) con ϕ=0.98 y σ_{η}^2 =0.01 (····), σ_{η}^2 =0.05 (- - - -), σ_{η}^2 =0.1 (– – –).

Gráfico 2.3. Bandas de confianza de Barlett (línea continua) y bandas corregidas por volatilidad para un modelo LMSV con d=0.45 y σ_η^2 =0.01 (····), σ_η^2 =0.05 (- - - -), σ_η^2 =0.1 (– – –).

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Gráfico 2.4. Bandas de confianza de Barlett (línea continua) y bandas corregidas por volatilidad para un modelo LMSV con d=0.49 y σ_η^2 =0.01 (····), σ_η^2 =0.05 (- - - -), σ_η^2 =0.1 (– – –).

Una vez especificado el modelo SV que parametriza la heteroscedasticidad de la serie y_t, la probabilidad de error tipo I del contraste de incorrelación de las bandas de Barlett para modelos SV, puede obtenerse analíticamente. Utilizando la expresión de la varianza asintótica de $r_Y(k)$ en [2.7], la probabilidad de que $r_Y(k)$ exceda las bandas de Barlett puede calcularse, de forma aproximada, de la siguiente forma:

$$p(\text{ERROR I})= p_{H_0}\left(|r_Y(k)|\geq 1.96/\sqrt{T}\right) = p_{H_0}\left(\frac{|r_Y(k)|}{\sqrt{C_1(T,k)}} \geq \frac{1.96}{\sqrt{T\,C_1(T,k)}}\right)=$$

$$= p_{H_0}\left(\frac{|r_Y(k)|}{\sqrt{C_1(T,k)}} \geq 1.96\exp\{-0.5\gamma_h(k)\}\right) \cong 2\,[1-\Phi(1.96\exp\{-0.5\gamma_h(k)\})] \qquad [2.17]$$

donde Φ denota la función de distribución de una variable normal estándar. En el modelo ARSV(1) con $\phi>0$ y en el modelo LMSV con $d>0$, se tiene que $\gamma_h(k)>0$ y por tanto, $\exp\{-0.5\gamma_h(k)\}<1$ y $\Phi(1.96\exp\{-0.5\gamma_h(k)\})<\Phi(1.96)=0.975$. En consecuencia, la probabilidad de error de tipo I en estos modelos SV será siempre mayor que 0.05.

La tabla 2.1 recoge los valores teóricos de la probabilidad de error tipo I calculados con la fórmula [2.17], para un modelo ARSV(1) con $\sigma_\eta^2=\{0.1,0.05,0.01\}$ y distintos valores de ϕ. La tabla 2.2 recoge las mismas probabilidades para un modelo LMSV con $\sigma_\eta^2=\{0.1, 0.05, 0.01\}$ y distintos valores de d. Estas tablas confirman los resultados anteriores: para modelos ARSV(1) con parámetro autorregresivo pequeño, el error que se comete utilizando las bandas sin corregir es despreciable, pero para los modelos más próximos a la raíz unitaria ($\phi=0.95$, $\phi=0.98$) y con mayor varianza ($\sigma_\eta^2=0.1$), las diferencias entre las bandas corregidas y sin corregir se agravan y el error puede ser importante, incluso en los retardos alejados. En estos casos, la corrección por volatilidad estocástica resulta imprescindible para evitar detectar correlaciones espurias en la serie y_t. Resultados similares se obtienen en el modelo LMSV, donde los mayores errores corresponden a los modelos con d próximo a la frontera de la no estacionariedad (d=1/2) y con mayor varianza.

Tabla 2.1. Probabilidad de error tipo I del contraste de Barlett sin corregir por ARSV(1)

σ_η^2 =0.1	ϕ=0.5	ϕ=0.9	ϕ=0.95	ϕ=0.98
K=1	0.058	0.122	0.229	0.570
K=5	0.051	0.093	0.187	0.531
K=10	0.050	0.074	0.149	0.485
K=20	0.050	0.058	0.103	0.399
K=50	0.050	0.050	0.060	0.216
σ_η^2 =0.05	ϕ=0.5	ϕ=0.9	ϕ=0.95	ϕ=0.98
K=1	0.054	0.082	0.125	0.291
K=5	0.050	0.070	0.108	0.268
K=10	0.050	0.061	0.093	0.242
K=20	0.050	0.054	0.074	0.199
K=50	0.050	0.050	0.055	0.119
σ_η^2 =0.01	ϕ=0.5	ϕ=0.9	ϕ=0.95	ϕ=0.98
K=1	0.051	0.056	0.062	0.083
K=5	0.050	0.054	0.060	0.080
K=10	0.050	0.052	0.057	0.077
K=20	0.050	0.051	0.054	0.072
K=50	0.050	0.050	0.051	0.061

Tabla 2.2. Probabilidad de error tipo I del contraste de Barlett sin corregir por LMSV

σ_η^2 =0.1	d=0.2	d=0.4	d=0.45	d=0.49
K=1	0.053	0.067	0.091	0.372
K=5	0.051	0.062	0.084	0.360
K=10	0.051	0.061	0.082	0.355
K=20	0.051	0.059	0.079	0.355
K=50	0.050	0.058	0.077	0.344
σ_η^2 =0.05	d=0.2	d=0.4	d=0.45	d=0.49
K=1	0.052	0.058	0.069	0.186
K=5	0.051	0.056	0.066	0.180
K=10	0.050	0.055	0.065	0.178
K=20	0.050	0.054	0.064	0.176
K=50	0.050	0.054	0.062	0.173
σ_η^2 =0.01	d=0.2	d=0.4	d=0.45	d=0.49
K=1	0.050	0.052	0.054	0.070
K=5	0.050	0.051	0.053	0.069
K=10	0.050	0.051	0.053	0.069
K=20	0.050	0.051	0.053	0.069
K=50	0.050	0.051	0.052	0.068

En la práctica, los verdaderos valores de los parámetros del modelo se desconocen y la varianza asintótica de las correlaciones muestrales debe estimarse a partir de los datos. Como h_t no es observable, la varianza de $r_Y(k)$ se estimará a partir de las observaciones $\{y_1,\ldots,y_T\}$. En este caso, en lugar de considerar la expresión de la varianza asintótica de $r_Y(k)$ en [2.7], que depende de $\gamma_h(k)$, resulta conveniente utilizar la fórmula [2.12], que depende únicamente de $\gamma_Y(0)$ y $\gamma_{Y^2}(k)$. Como sugiere Diebold (1988), las funciones $\gamma_Y(0)$ y $\gamma_{Y^2}(k)$ se pueden estimar de forma consistente mediante sus correspondientes valores muestrales:

$$\hat{\gamma}_Y(0) = c_Y^*(0) = \frac{1}{T}\sum_{t=1}^{T} y_t^2$$

$$\hat{\gamma}_{Y^2}(k) = \frac{1}{T}\sum_{t=1}^{T-k}(y_t^2 - \overline{Y^2})(y_{t+k}^2 - \overline{Y^2}), \text{ con } \overline{Y^2} = \frac{1}{T}\sum_{t=1}^{T} y_t^2$$

y se obtiene así el siguiente estimador consistente de la varianza asintótica de $r_Y(k)$:

$$\hat{C}_1(T,k) = \frac{1}{T}\left\{1 + \frac{\hat{\gamma}_{Y^2}(k)}{[\hat{\gamma}_Y(0)]^2}\right\} \qquad [2.18]$$

A partir de este valor, se obtienen las correspondientes bandas de confianza muestrales corregidas por volatilidad para contrastar incorrelación en y_t:

$$\pm 1.96[\hat{C}_1(T,k)]^{1/2} \qquad [2.19]$$

Estas bandas son las mismas que propone Diebold (1988) para corregir por efectos ARCH, y son las que se utilizarán en las aplicaciones con datos reales.

2.2.2. Propiedades en muestras finitas del contraste de incorrelación de las bandas de Barlett en modelos SV

En esta sección se presentan los resultados de un estudio de Monte Carlo para ilustrar la validez en muestras finitas de las propiedades del contraste de incorrelación serial para modelos SV derivado en la sección anterior. En primer lugar comentaremos los resultados obtenidos para modelos ARSV(1) y después presentaremos los correspondientes a los modelos LMSV.

Todas las simulaciones se han realizado utilizando GAUSS versión 3.2 en un Pentium 166 MHz. Las realizaciones del ruido blanco gaussiano ε_t, t=1,...,T, se generan con el comando RNDNS de GAUSS. Para el modelo LMSV, las realizaciones del proceso h_t, que es un ARFIMA(0,d,0), se generan con un programa propio que utiliza el algoritmo de Davies y Harte (1987).[6] Una vez obtenidas las realizaciones de los procesos ε_t y h_t, los valores de y_t se obtienen como $y_t=\sigma_*\varepsilon_t\exp(\frac{1}{2}h_t)$, con $\sigma_*=1$.

Para el modelo ARSV(1) hemos elegido como valores de los parámetros $\phi=\{0.5, 0.9, 0.95, 0.98\}$ y $\sigma_\eta^2=0.05$. Estos valores nos permiten analizar un amplio abanico de posibilidades que incluye modelos próximos a la no estacionariedad ($\phi\approx1$) y modelos intermedios ($\phi=0.5$).[7] Además, estos valores son los mismos que hemos utilizado para calcular analíticamente las probabilidades de error tipo I que presentamos en la tabla 2.1 y, por tanto, nos permitirán comparar los valores de esa tabla con los que obtengamos empíricamente en las simulaciones.[8] Para cada modelo se analizan dos tamaños muestrales, T=512 y T=2048, y para algunos modelos

[6] Véase el apéndice A para una justificación teórica de este método de simulación.

[7] Se han analizado otros valores intermedios de ϕ sin que hubiera grandes diferencias en los resultados, por lo que sólo incluimos aquí los resultados correspondientes a $\phi=0.5$.

[8] La tabla 2.1 analizaba tres valores de σ_η^2, $\sigma_\eta^2=\{0.1, 0.05, 0.01\}$. En las simulaciones sólo incluimos uno de ellos, $\sigma_\eta^2=0.05$, porque los experimentos realizados con los otros valores proporcionan resultados muy similares y no aportan nuevas conclusiones.

concretos se prueba también un tamaño muestral más grande, T=8192.[9] En todos los casos, el nivel de significación nominal de los contrastes es el 5%.

Para cada especificación de los parámetros y cada tamaño muestral, se generan 5000 series de longitud T, y para cada una de ellas se calculan las 50 primeras autocorrelaciones muestrales de y_t, de modo que para cada retardo k, con k=1,2,…,50, se tienen 5000 valores de la correlación muestral correspondiente.

La tabla 2.3 recoge la media y la desviación estándar teórica de la distribución asintótica de $r_Y(k)$ según la expresión [2.7], con k=1,5,10,20,50, para los dos modelos más persistentes (ϕ=0.95, ϕ=0.98) y para dos tamaños muestrales (T=512, T=2048). Para el modelo más próximo a la raíz unitaria (ϕ=0.98) se presentan además los resultados con T=8192. Junto a los valores asintóticos se presentan los correspondientes valores muestrales para las 5000 réplicas. Estos últimos se obtienen de la siguiente forma:

$$\bar{r}_Y(k) = \frac{1}{5000}\sum_{i=1}^{5000} r_Y^{(i)}(k), \text{ s.d.} = \sqrt{\frac{1}{5000}\sum_{i=1}^{5000}[r_Y^{(i)}(k) - \bar{r}_Y(k)]^2}$$

donde $r_Y^{(i)}(k)$ denota la correlación muestral de orden k de y_t en la i-ésima réplica.

A juzgar por los valores que se presentan en la tabla 2.3, parece que la distribución asintótica de $r_Y(k)$ en [2.7] es una buena aproximación a la distribución empírica en muestras finitas, aunque en el modelo más próximo a la raíz unitaria (ϕ=0.98) la convergencia es más lenta y se requieren tamaños de muestras más grandes. Obsérvese que en este caso, la desviación estándar asintótica sobreestima la desviación estándar de la distribución muestral.

[9] La razón de elegir estos tamaños muestrales es que para simular el proceso ARFIMA(0,d,0) que subyace al modelo LMSV estacionario, hemos utilizado el método propuesto por Davies y Harte (1987), que se basa en transformadas de Fourier de las autocovarianzas. Trabajar con potencias de dos (T=512, 2048, 8192), nos permite utilizar la transformada rápida de Fourier y ahorrar un tiempo considerable en los cálculos.

Tabla 2.3. Media y desviación típica (entre paréntesis), asintótica y muestral, de las autocorrelaciones muestrales de la serie y_t en modelos ARSV(1) con ϕ={0.95, 0.98} y σ_η^2 =0.05.

T.muestral	T=512				T=2048				T=8192	
Parámetro	ϕ=0.95		ϕ=0.98		ϕ=0.95		ϕ=0.98		ϕ=0.98	
Correlación	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
$r_Y(1)$	0.000 (0.056)	-0.001 (0.054)	0.000 (0.082)	-0.002 (0.064)	0.000 (0.028)	0.000 (0.028)	0.000 (0.041)	0.000 (0.037)	0.000 (0.020)	0.000 (0.020)
$r_Y(5)$	0.000 (0.054)	-0.002 (0.052)	0.000 (0.078)	-0.002 (0.062)	0.000 (0.027)	0.000 (0.027)	0.000 (0.039)	0.000 (0.036)	0.000 (0.019)	0.000 (0.019)
$r_Y(10)$	0.000 (0.052)	-0.002 (0.049)	0.000 (0.074)	-0.002 (0.058)	0.000 (0.026)	-0.001 (0.025)	0.000 (0.037)	0.000 (0.034)	0.000 (0.018)	0.000 (0.018)
$r_Y(20)$	0.000 (0.049)	-0.002 (0.046)	0.000 (0.067)	-0.002 (0.051)	0.000 (0.024)	0.000 (0.024)	0.000 (0.034)	0.000 (0.031)	0.000 (0.017)	0.000 (0.016)
$r_Y(50)$	0.000 (0.045)	-0.002 (0.042)	0.000 (0.056)	-0.001 (0.044)	0.000 (0.023)	-0.001 (0.022)	0.000 (0.028)	-0.001 (0.025)	0.000 (0.014)	0.000 (0.014)

Por lo que respecta a los resultados en muestras finitas de los contrastes de incorrelación serial, la notación que utilizaremos es la siguiente. Denotaremos por P la proporción de rechazos con el contraste de incorrelación de las bandas de Barlett, es decir, la proporción de veces, en las 5000 réplicas, que $|r_Y^{(i)}(k)|>1.96/\sqrt{T}$, con i=1,…,5000. De forma análoga, P_c denotará la proporción de rechazos en las 5000 réplicas utilizando las bandas corregidas por volatilidad y, por tanto, contará la proporción de veces que $|r_Y^{(i)}(k)|>1.96[C_1(T,k)]^{1/2}$, i=1,2,…,5000, siendo $C_1(T,k)$ la varianza asintótica de $r_Y(k)$ corregida por volatilidad en [2.7]. Finalmente, $\hat{P}_c$ denotará el tamaño empírico del contraste de incorrelación que utiliza las bandas muestrales corregidas por volatilidad y, por tanto, contará la proporción de veces que $|r_Y^{(i)}(k)|>1.96[\hat{C}_1(T,k)]^{1/2}$, i=1,2,…,5000, donde $\hat{C}_1(T,k)$ es el estimador de la varianza asintótica de $r_Y(k)$ definido en [2.18]. La tabla 2.4 recoge los valores de P, P_c y $\hat{P}_c$, para los distintos modelos ARSV(1) y para los retardos k=1, 5, 10, 20 y 50.

Tabla 2.4. Tamaño empírico del contraste de incorrelación serial de las bandas de Barlett, corregidas y sin corregir por volatilidad, en un modelo ARSV(1) con σ_{η}^2 =0.05 y diferentes valores de ϕ.[(a)]

T. muestral		**T=512**				**T=2048**				**T=8192**	
Parámetros		**ϕ=0.5**	**ϕ=0.9**	**ϕ=0.95**	**ϕ=0.98**	**ϕ=0.5**	**ϕ=0.9**	**ϕ=0.95**	**ϕ=0.98**	**ϕ=0.9**	**ϕ=0.98**
K=1	P	0.048	0.071	0.102	0.167	0.050	0.080	0.118	0.239	0.081	0.267
	P_c	0.044	0.043	0.039	0.016	0.046	0.047	0.046	0.029	0.049	0.044
	$\hat{P}_c$	0.046	0.045	0.046	0.048	0.046	0.047	0.047	0.050	0.049	0.050
K=5	P	0.050	0.063	0.093	0.156	0.049	0.066	0.100	0.212	0.075	0.243
	P_c	0.050	0.044	0.038	0.017	0.049	0.048	0.049	0.036	0.057	0.048
	$\hat{P}_c$	0.051	0.047	0.045	0.048	0.048	0.047	0.047	0.048	0.056	0.053
K=10	P	0.043	0.054	0.079	0.137	0.048	0.061	0.088	0.188	0.062	0.221
	P_c	0.043	0.042	0.038	0.015	0.048	0.049	0.050	0.031	0.050	0.042
	$\hat{P}_c$	0.044	0.045	0.048	0.047	0.048	0.050	0.050	0.048	0.054	0.052
K=20	P	0.045	0.049	0.062	0.102	0.052	0.056	0.069	0.150	0.057	0.176
	P_c	0.045	0.046	0.041	0.015	0.052	0.052	0.047	0.034	0.054	0.043
	$\hat{P}_c$	0.047	0.046	0.045	0.049	0.052	0.052	0.051	0.052	0.054	0.046
K=50	P	0.041	0.038	0.038	0.053	0.045	0.047	0.053	0.094	0.047	0.108
	P_c	0.041	0.038	0.036	0.016	0.045	0.047	0.049	0.035	0.047	0.044
	$\hat{P}_c$	0.040	0.038	0.039	0.037	0.043	0.046	0.048	0.046	0.047	0.046

[(a)] Los valores P, P_c y $\hat{P}_c$ denotan, para cada k, los tamaños empíricos del contraste de incorrelación de orden k con las bandas de Barlett, con las bandas corregidas por el efecto SV teórico, definidas en [2.15], y con las bandas corregidas por el efecto SV estimado, definidas en [2.19], respectivamente. Todos los cálculos se basan en 5000 réplicas. El tamaño nominal del contraste es el 5%.

En esta tabla observamos que, cuando ϕ=0.5, el tamaño empírico que proporcionan las bandas de Barlett y las bandas corregidas por volatilidad es muy similar en todos los retardos. De hecho, el intervalo $\pm 1.96/\sqrt{T}$ mantiene el tamaño nominal prácticamente en todos los casos. Sin embargo, al aumentar el valor de ϕ, el tamaño empírico del contraste de Barlett también aumenta y se aleja del nominal 0.05. Por ejemplo, cuando ϕ=0.98, el tamaño empírico para contrastar incorrelación de orden k=10 en una muestra de tamaño T=512, casi triplica el tamaño nominal, y en una muestra de tamaño T=8192, el tamaño empírico de este contraste es casi cinco veces el nominal.

Por el contrario, las bandas corregidas por el efecto estimado de la volatilidad mantienen siempre el tamaño nominal, incluso en modelos muy persistentes y con tamaños muestrales pequeños (obsérvense los valores de $\hat{P}_c$ en el modelo con ϕ=0.98 y T=512). Nótese también que en el modelo con ϕ=0.98, cercano a la raíz unitaria y, por tanto, cercano a la no estacionariedad, el tamaño empírico del contraste de incorrelación que utiliza las bandas teóricas corregidas por volatilidad estocástica, P_c, es siempre inferior al nominal. En este caso, la convergencia de $r_Y(k)$ hacia la distribución asintótica es más lenta, pero cuando T=8192, la aproximación asintótica ya proporciona resultados aceptables; veáse la última columna de la tabla 2.4.

Por otro lado, en la tabla 2.4 observamos que las diferencias entre las bandas corregidas y sin corregir, se reducen paulatinamente a medida que aumenta el orden de la correlación cuya significación se contrasta. Por ejemplo, para k=50 y T=2048, el tamaño empírico del contraste de Barlett está muy próximo al nominal en todos los modelos, excepto en el modelo más persistente (ϕ=0.98), donde las bandas $\pm 1.96/\sqrt{T}$ son realmente inapropiadas. Estos resultados confirman los resultados teóricos de la sección anterior, en el sentido de que la utilización de las bandas de Barlett en modelos ARSV(1) muy persistentes puede conllevar la detección de correlaciones espurias.

Otro aspecto destacable de los resultados de las simulaciones es que los valores de P que se presentan en la tabla 2.4 son muy similares a los correspondientes valores teóricos de las probabilidades de error tipo I en la tabla 2.1, confirmándose una vez más la validez de los resultados asintóticos derivados en la sección anterior.

Muchas de las propiedades que acabamos de describir se aprecian claramente en los gráficos 2.5-2.8. Estos gráficos representan las distribuciones de las correlaciones muestrales, sin corregir y corregidas por volatilidad, para un modelo ARSV(1) con ϕ=0.9 y para un modelo muy próximo a la raíz unitaria (ϕ=0.98). En concreto, el gráfico 2.5 representa las distribuciones empíricas de $\sqrt{T}\, r_Y(k)$, $r_Y(k)/[C_1(T,k)]^{1/2}$ y $r_Y(k)/[\hat{C}_1(T,k)]^{1/2}$, con k=1,5,10,20,50, basadas en 5000 réplicas del modelo

ARSV(1) con parámetros $\{\phi=0.9, \sigma_\eta^2=0.05\}$ y tamaño muestral T=512. El gráfico 2.6 representa las mismas distribuciones para T=8192, y los gráficos 2.7-2.8 para un modelo con parámetros $\{\phi=0.98, \sigma_\eta^2=0.05\}$ y con T=512 y T=8192, respectivamente. Sobre las distribuciones empíricas se representa siempre la distribución N(0,1).

Las conclusiones que se derivan de estos gráficos son inmediatas. Cuando $\phi=0.9$, la distribución empírica de las correlaciones corregidas por volatilidad estocástica se ajusta perfectamente a la distribución asintótica N(0,1). Además, esta distribución apenas difiere de la distribución de $\sqrt{T}\, r_Y(k)$, excepto en los primeros retardos, donde ésta proporcionaría siempre un tamaño empírico del contraste ligeramente mayor que el nominal, en consonancia con los resultados de la tabla 2.4.

En el modelo con $\phi=0.98$, la distribución de las correlaciones muestrales corregidas muestralmente por volatilidad estocástica (tercera columna) se ajusta perfectamente a la distribución N(0,1), a diferencia de lo que ocurre con la distribución de $\sqrt{T}\, r_Y(k)$ (primera columna), que proporcionaría siempre un tamaño empírico del contraste mayor que el nominal, especialmente en los primeros retardos, que son los que habitualmente se contrastan. Sin embargo, en los retardos más alejados, las diferencias entre ambas distribuciones disminuyen considerablemente.

Por lo que respecta a la distribución de las correlaciones con la corrección teórica por volatilidad, en el gráfico 2.7 (segunda columna) se aprecia claramente que con $\phi=0.98$ y con muestras pequeñas (T=512), ésta presenta menos varianza de la esperada en la distribución asintótica, y de ahí que proporcione siempre un tamaño empírico del contraste de incorrelación menor que el nominal. En este caso, la distribución asintótica no es una buena aproximación, porque el modelo simulado es prácticamente un modelo SV integrado ($\phi=1$), en cuyo caso se violaría la hipótesis de estacionariedad y, por tanto, la proposición 2.1 dejaría de ser cierta. Sin embargo, estas diferencias entre las distribuciones empírica y asintótica se reducen al aumentar el tamaño muestral, como puede verse en el gráfico 2.8, donde las distribuciones corregidas, teórica y muestralmente, por volatilidad son casi indistinguibles.

Gráfico 2.5. Distribución empírica de las correlaciones muestrales de y_t sin corregir y corregidas, teórica y muestralmente, por volatilidad, para k=1,5,10,20,50, junto con la distribución N(0,1). El modelo simulado es un ARSV(1) con ϕ=0.9, σ_η^2=0.05, T=512.

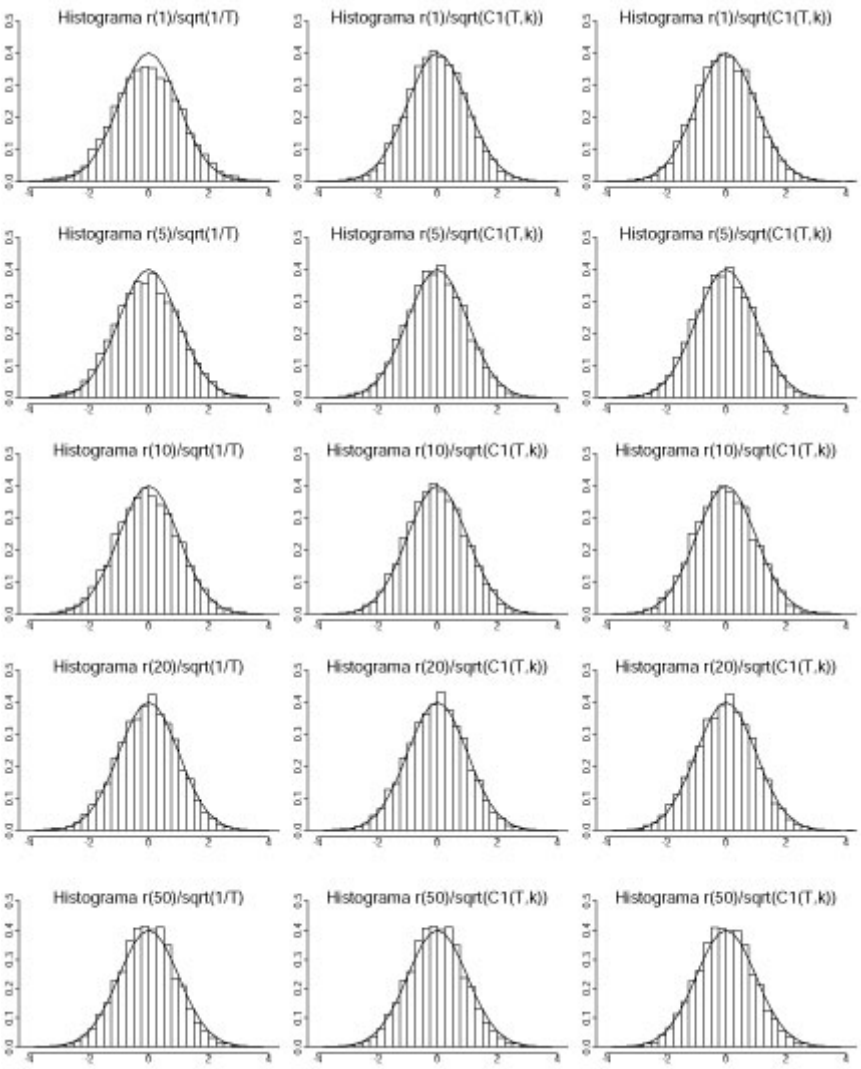


Gráfico 2.6. Distribución empírica de las correlaciones muestrales de y_t sin corregir y corregidas, teórica y muestralmente, por volatilidad, para k=1,5,10,20,50, junto con la distribución N(0,1). El modelo simulado es un ARSV(1) con ϕ=0.9, σ_η^2=0.05, T=8192.

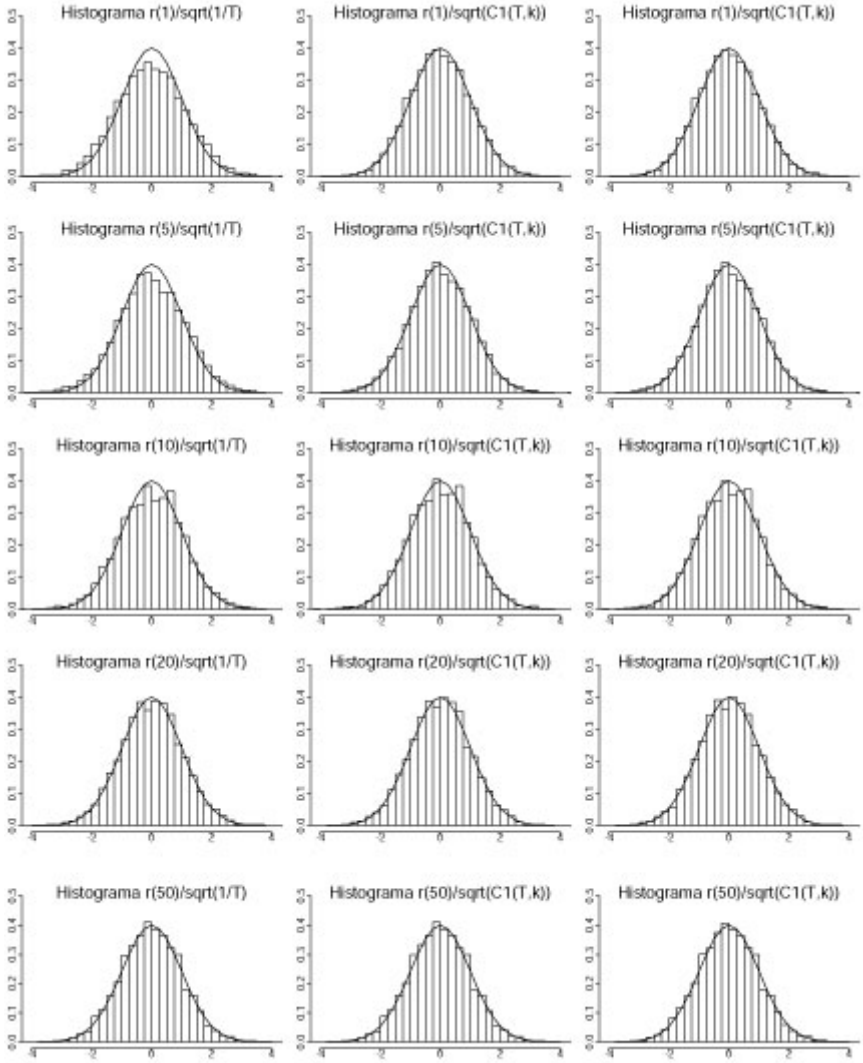

Histograma r(1)/sqrt(1/T)
Histograma r(1)/sqrt(C1(T,k))
Histograma r(1)/sqrt(C1(T,k))
Histograma r(5)/sqrt(1/T)
Histograma r(5)/sqrt(C1(T,k))
Histograma r(5)/sqrt(C1(T,k))
Histograma r(10)/sqrt(1/T)
Histograma r(10)/sqrt(C1(T,k))
Histograma r(10)/sqrt(C1(T,k))
Histograma r(20)/sqrt(1/T)
Histograma r(20)/sqrt(C1(T,k))
Histograma r(20)/sqrt(C1(T,k))
Histograma r(50)/sqrt(1/T)
Histograma r(50)/sqrt(C1(T,k))
Histograma r(50)/sqrt(C1(T,k))

Gráfico 2.7. Distribución empírica de las correlaciones muestrales de y_t sin corregir y corregidas, teórica y muestralmente, por volatilidad, para k=1,5,10,20,50, junto con la distribución N(0,1). El modelo simulado es un ARSV(1) con ϕ=0.98, σ_η^2=0.05, T=512.

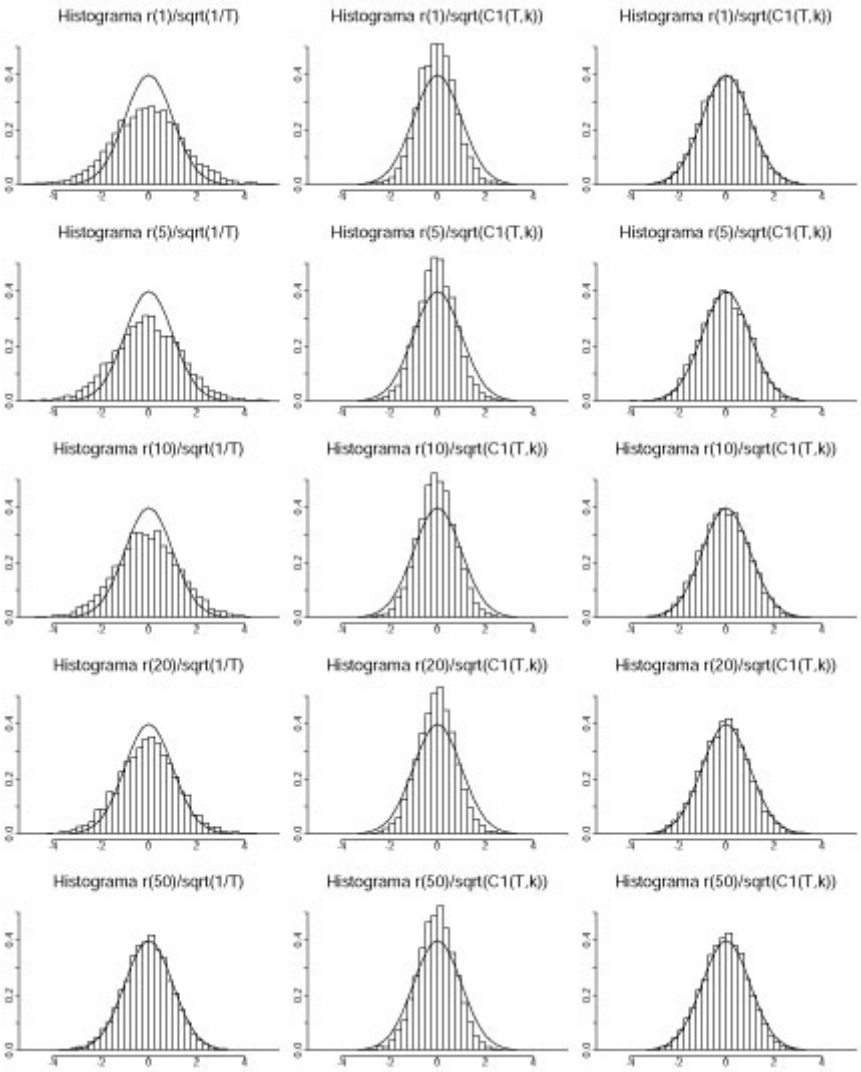
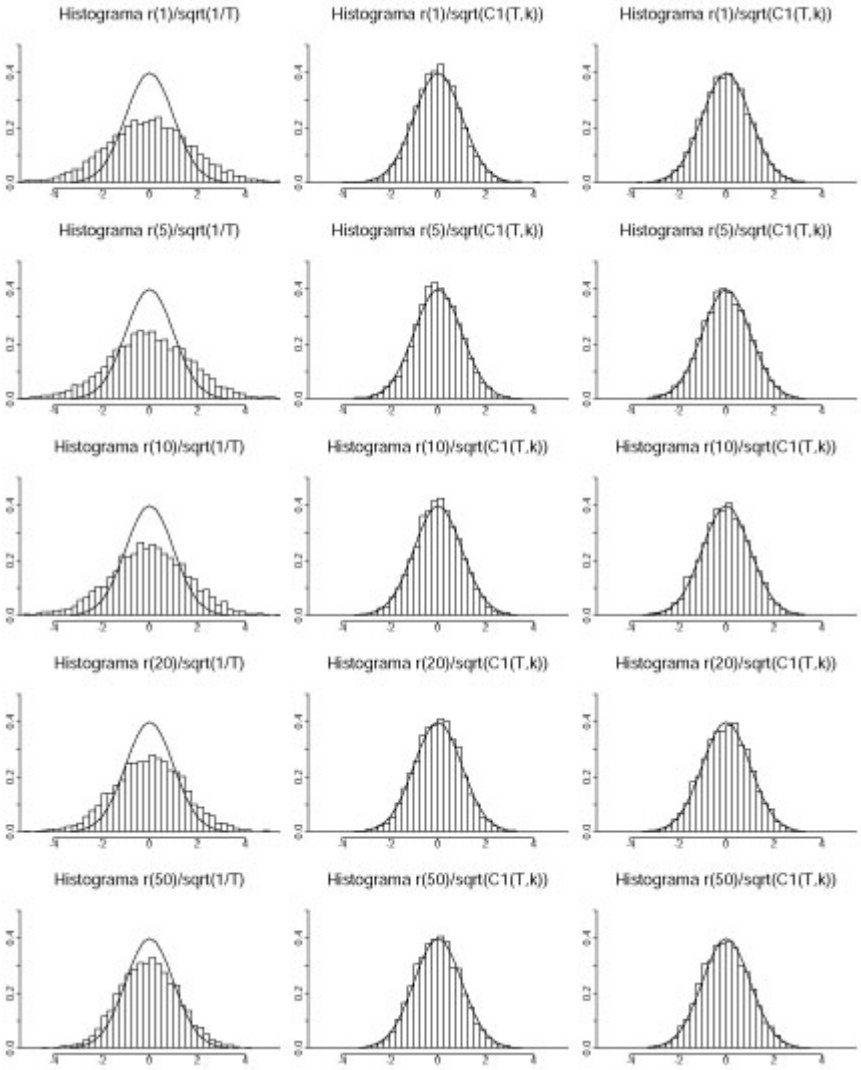


Gráfico 2.8. Distribución empírica de las correlaciones muestrales de y_t sin corregir y corregidas, teórica y muestralmente, por volatilidad, para k=1,5,10,20,50, junto con la distribución N(0,1). El modelo simulado es un ARSV(1) con ϕ=0.98, σ_η^2=0.05, T=8192.



En cualquier caso, la conclusión más importante que se deriva de los experimentos de Monte Carlo es que en los modelos ARSV(1) más persistentes, las bandas habituales, $\pm 1.96/\sqrt{T}$, son inapropiadas, mientras que las bandas corregidas muestralmente por volatilidad, $\pm 1.96[\hat{C}_1(T,k)]^{1/2}$, que son las que se utilizarán en la práctica, funcionan muy bien y mantienen siempre el tamaño nominal del contraste.

A continuación comentaremos los resultados de las simulaciones para los modelos LMSV. En este caso los parámetros analizados son d={0.2, 0.4, 0.45, 0.49} y σ_η^2=0.1. Al igual que en el modelo ARSV(1), los tamaños muestrales utilizados son T=512 y T=2048, y el número de réplicas es 5000. Para algún modelo concreto presentamos también resultados para muestras más grandes de tamaño T=8192.

La tabla 2.5 recoge la media y la desviación estándar (entre paréntesis) asintótica de $r_Y(k)$ para dos modelos concretos (d=0.4, d=0.49) y dos tamaños muestrales (T=512, T=2048), junto con los correspondientes valores muestrales para las 5000 réplicas. Para el modelo más próximo a la no estacionariedad (d=0.49), se presentan además los resultados obtenidos con T=8192.

Tabla 2.5. Media y desviación típica (entre paréntesis), asintótica y muestral, de las autocorrelaciones de la serie y_t en modelos LMSV con d={0.4, 0.49} y σ_η^2=0.1.

T.muestral	T=512				T=2048				T=8192	
Parámetro	d=0.4		d=0.49		d=0.4		d=0.49		d=0.49	
Correlación	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
$r_Y(1)$	0.000 (0.047)	-0.001 (0.045)	0.000 (0.097)	-0.001 (0.047)	0.000 (0.024)	0.000 (0.023)	0.000 (0.048)	-0.001 (0.024)	0.000 (0.024)	0.000 (0.012)
$r_Y(5)$	0.000 (0.046)	-0.002 (0.045)	0.000 (0.095)	-0.002 (0.047)	0.000 (0.023)	0.000 (0.022)	0.000 (0.047)	0.000 (0.023)	0.000 (0.024)	0.000 (0.012)
$r_Y(10)$	0.000 (0.046)	-0.002 (0.044)	0.000 (0.094)	-0.002 (0.046)	0.000 (0.023)	0.000 (0.022)	0.000 (0.047)	0.000 (0.023)	0.000 (0.024)	0.000 (0.012)
$r_Y(20)$	0.000 (0.046)	-0.003 (0.044)	0.000 (0.093)	-0.003 (0.044)	0.000 (0.023)	-0.001 (0.022)	0.000 (0.046)	-0.001 (0.023)	0.000 (0.023)	0.000 (0.012)
$r_Y(50)$	0.000 (0.046)	-0.002 (0.043)	0.000 (0.091)	-0.002 (0.043)	0.000 (0.023)	-0.001 (0.022)	0.000 (0.046)	-0.001 (0.023)	0.000 (0.023)	0.000 (0.012)

En esta tabla observamos que para d=0.4, la distribución asintótica parece una buena aproximación a la distribución en muestras finitas, incluso cuando T=512. Sin embargo, cuando el modelo es muy persistente (d=0.49), las diferencias entre los valores de la varianza asintótica y la varianza muestral son muy importantes: en todos los retardos, la desviación típica muestral de $r_Y(k)$ es aproximadamente la mitad de la desviación típica asintótica y, además, esta diferencia no disminuye al aumentar T. Por tanto, en este modelo tan próximo a la frontera de la no estacionariedad (d=1/2), la distribución asintótica es totalmente inapropiada, incluso en muestras grandes.

Lo que realmente ocurre en este caso es que al estar en la frontera de la no estacionariedad, la varianza de y_t no es finita y, por tanto, tampoco lo es la varianza de las correlaciones muestrales. Por eso la varianza asintótica es grande y la aproximación asintótica no funciona bien (el corolario 2.1 deja de ser cierto). Como veremos más adelante, esto va a provocar que el tamaño empírico del contraste que utiliza las correlaciones corregidas por el efecto teórico de la volatilidad, sea siempre menor que el nominal 0.05.

La tabla 2.6 recoge los valores de los tamaños empíricos del contraste de incorrelación de Barlett y de los dos contrastes de incorrelación definidos en [2.15] y [2.19], correspondientes a corregir por volatilidad con el factor de corrección teórico y con el factor de corrección muestral, respectivamente. Al igual que en la tabla 2.4, los tamaños empíricos de estos tres contrastes se denotan por P, P_c y $\hat{P}_c$, respectivamente, y se obtienen de la misma forma que en los modelos ARSV(1). En todos los casos el tamaño nominal es el 5%.

Tabla 2.6. Tamaño empírico del contraste de incorrelación serial de las bandas de Barlett, corregidas y sin corregir por volatilidad, en un modelo LMSV con σ_η^2 =0.1 y distintos valores de d.[(b)]

T. muestral		**T=512**				**T=2048**				**T=8192**	
Parámetros		**d=0.2**	**d=0.4**	**d=0.45**	**d=0.49**	**d=0.2**	**d=0.4**	**d=0.45**	**d=0.49**	**d=0.4**	**d=0.49**
K=1	P	0.048	0.056	0.058	0.065	0.047	0.058	0.064	0.070	0.061	0.077
	P_c	0.045	0.039	0.029	0.000	0.044	0.043	0.031	0.000	0.043	0.000
	$\hat{P}_c$	0.047	0.048	0.049	0.048	0.045	0.047	0.047	0.047	0.047	0.049
K=5	P	0.049	0.054	0.054	0.055	0.044	0.051	0.055	0.063	0.064	0.078
	P_c	0.049	0.041	0.028	0.000	0.044	0.040	0.031	0.000	0.053	0.000
	$\hat{P}_c$	0.049	0.049	0.048	0.048	0.045	0.045	0.045	0.049	0.057	0.057
K=10	P	0.043	0.046	0.048	0.050	0.052	0.055	0.058	0.063	0.058	0.069
	P_c	0.041	0.036	0.025	0.000	0.051	0.046	0.036	0.000	0.048	0.000
	$\hat{P}_c$	0.044	0.043	0.043	0.042	0.050	0.051	0.050	0.049	0.051	0.050
K=20	P	0.047	0.048	0.049	0.049	0.050	0.054	0.055	0.058	0.051	0.062
	P_c	0.046	0.038	0.026	0.000	0.049	0.043	0.032	0.000	0.045	0.000
	$\hat{P}_c$	0.047	0.047	0.047	0.047	0.049	0.048	0.048	0.049	0.048	0.048
K=50	P	0.044	0.042	0.044	0.045	0.050	0.050	0.051	0.053	0.058	0.065
	P_c	0.043	0.037	0.028	0.000	0.050	0.044	0.033	0.000	0.051	0.000
	$\hat{P}_c$	0.043	0.042	0.042	0.042	0.048	0.049	0.049	0.047	0.053	0.056

[(b)] La notación es la misma que en la tabla 2.4.

En esta tabla observamos que, al igual que ocurría en los modelos SV autorregresivos, las mayores discrepancias entre el tamaño empírico del contraste de Barlett, P, y el tamaño nominal 0.05, corresponden a los modelos más persistentes (d=0.49) y a los primeros retardos. Por otra parte, observamos que las diferencias entre el tamaño del contraste de Barlett, P, y el contraste que utiliza las bandas corregidas muestralmente, $\hat{P}_c$, no son muy grandes, y se reducen cuando aumenta el orden de la correlación cuya significación se contrasta. En estos últimos retardos, los errores que se cometerían utilizando las bandas incorrectas $\pm 1.96/\sqrt{T}$ serían muy pequeños. En la tabla 2.6, observamos también que los resultados de los contrastes apenas cambian al aumentar el tamaño muestral, debido precisamente a la propiedad de memoria larga del modelo LMSV.

En cualquier caso, los resultados más llamativos de la tabla 2.6 son los del contraste que utiliza la corrección teórica por volatilidad. En los modelos con d=0.2 y d=0.4, parece que este contraste funciona bastante bien. Sin embargo, en los modelos más persistentes, el tamaño empírico de este contraste, P_c, difiere bastante del tamaño nominal, y además esta diferencia no disminuye al aumentar T (para d=0.49, el tamaño empírico del contraste es siempre cero). Como ya hemos comentado antes, esto se debe a que el modelo con d=0.49 es prácticamente un modelo no estacionario y, por tanto, violaría una de las hipótesis del corolario 2.1, y la distribución asintótica de $r_Y(k)$ deja de ser una buena aproximación. Por el contrario, el contraste que utiliza la corrección muestral por volatilidad, $\hat{P}_c$, mantiene en todos los casos el tamaño nominal, incluso en los modelos más persistentes (d=0.49) y en muestras no muy grandes (T=512).

Una vez más, parece que las bandas que utilizan la corrección muestral por volatilidad, $\pm 1.96[\hat{C}_1(T,k)]^{1/2}$, funcionan mucho mejor que las corregidas por el efecto teórico de la volatilidad, $\pm 1.96[C_1(T,k)]^{1/2}$. Pero no sólo eso, en este caso, parece incluso que las bandas habituales, $\pm 1.96/\sqrt{T}$, son también mejores, y son casi tan buenas como las bandas corregidas muestralmente por volatilidad. Este resultado, que pudiera parecer en un principio contradictorio, realmente no lo es por lo siguiente.

Escribamos la correlación muestral corregida por la varianza asintótica 1/T y por la estimación muestral de la varianza asintótica en [2.18], de la siguiente forma:

$$\sqrt{T} r_Y(k) = \sqrt{T}\frac{c_Y(k)}{c_Y(0)} = \frac{\sum_{t=1}^{T-k}(y_t - \overline{Y})(y_{t+k} - \overline{Y})}{\frac{1}{\sqrt{T}}\sqrt{\left(\frac{1}{T}\sum_{t=1}^{T-k}(y_t - \overline{Y})^2\right)^2}}$$

$$\frac{r_Y(k)}{\sqrt{\hat{C}_1(T,k)}} = \frac{r_Y(k)}{\sqrt{\frac{1}{T}\frac{[\hat{\gamma}_Y(0)]^2 + \hat{\gamma}_{Y^2}(k)}{[\hat{\gamma}_Y(0)]^2}}} = \frac{\sum_{t=1}^{T-k}(y_t - \overline{Y})(y_{t+k} - \overline{Y})}{\frac{1}{\sqrt{T}}\sqrt{\left(\frac{1}{T}\sum_{t=1}^{T-k}(y_t - \overline{Y})^2\right)^2 + \hat{\gamma}_{Y^2}(k)}}$$

Estas dos expresiones difieren únicamente en el término de la covarianza muestral de la serie de los cuadrados, $\hat{\gamma}_{Y^2}(k)$, que aparece en el denominador. En modelos LMSV con valores pequeños de σ_η^2, la función de autocovarianza de la serie de los cuadrados, $\gamma_{Y^2}(k)$, toma valores muy pequeños, incluso en modelos muy persistentes (d=0.45, d=0.49), y además, la autocovarianza muestral de dicha serie, $\hat{\gamma}_{Y^2}(k)$, tiene siempre un sesgo negativo importante, como veremos en la sección siguiente. Por tanto, en el modelo que nos ocupa, con σ_η^2=0.1, el término $\hat{\gamma}_{Y^2}(k)$ tomará valores prácticamente iguales a cero (no hay corrección por volatilidad), y por eso las distribuciones de $\sqrt{T}\, r_Y(k)$ y $r_Y(k)/[\hat{C}_1(T,k)]^{1/2}$, son prácticamente iguales. De hecho, en los modelos LMSV con σ_η^2=0.1, la varianza condicional de la serie y_t es casi constante (no hay heteroscedasticidad), y por eso la corrección $1/\sqrt{T}$ es casi tan buena como la corrección muestral por volatilidad.

Todas las propiedades que acabamos de comentar se aprecian claramente en los gráficos 2.9-2.13. El gráfico 2.9 representa las distribuciones empíricas de $\sqrt{T}\, r_Y(k)$, $r_Y(k)/[C_1(T,k)]^{1/2}$ y $r_Y(k)/[\hat{C}_1(T,k)]^{1/2}$, para k=1,5,10,20,50, basadas en las 5000 réplicas del modelo LMSV con parámetros $\{d=0.4, \sigma_\eta^2=0.1\}$ y T=512. El gráfico 2.10 representa las mismas distribuciones para T=8192. El gráfico 2.11 representa las distribuciones de las correlaciones muestrales para un modelo LMSV con el mismo orden de integración (d=0.4), pero con mayor varianza (σ_η^2=0.5), y tamaño muestral T=8192. Finalmente, los gráficos 2.12-2.13 representan sendas distribuciones para el modelo más persistente, $\{d=0.49, \sigma_\eta^2=0.1\}$, con T=512 y T=8192, respectivamente. Sobre las distribuciones empíricas se representa la distribución N(0,1).

Gráfico 2.9. Distribución empírica de las correlaciones muestrales de y_t sin corregir y corregidas, teórica y muestralmente, por volatilidad, para k=1,5,10,20,50, junto con la distribución N(0,1). El modelo simulado es un LMSV con d=0.4, σ_η^2=0.1, T=512.

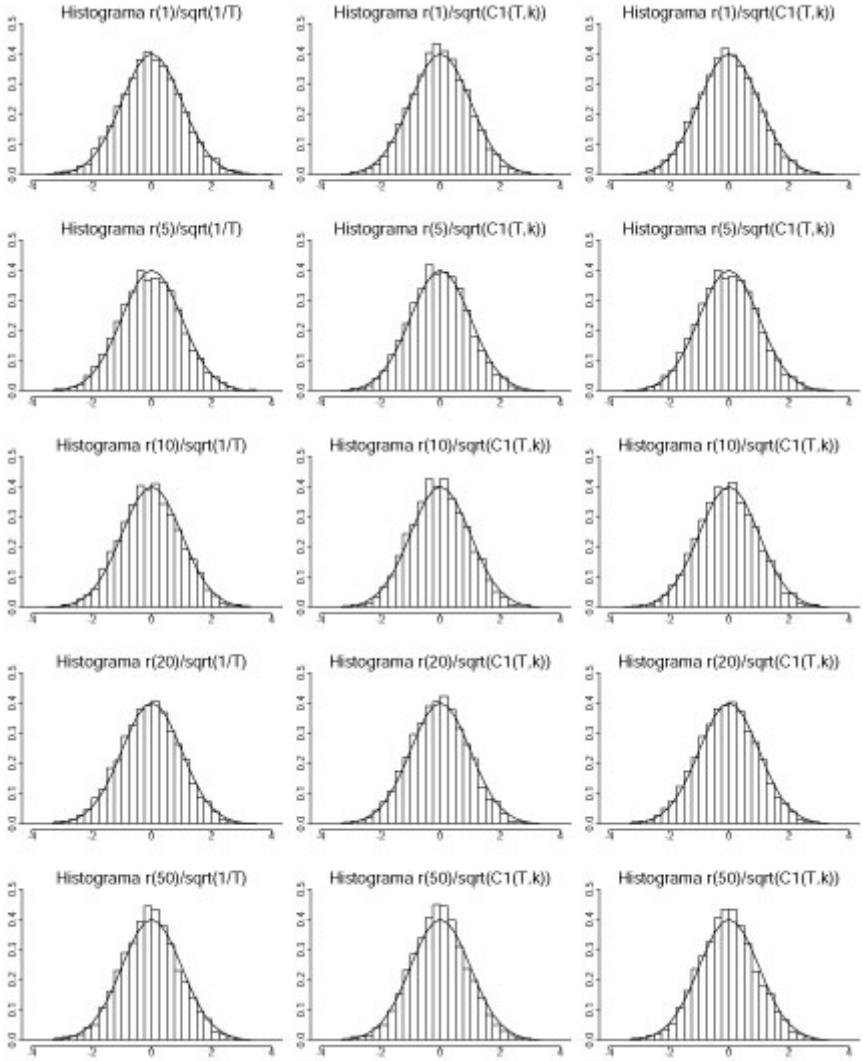


Gráfico 2.10. Distribución empírica de las correlaciones muestrales de y_t sin corregir y corregidas, teórica y muestralmente, por volatilidad, para k=1,5,10,20,50, junto con la distribución N(0,1). El modelo simulado es un LMSV con d=0.4, σ_η^2=0.1, T=8192.

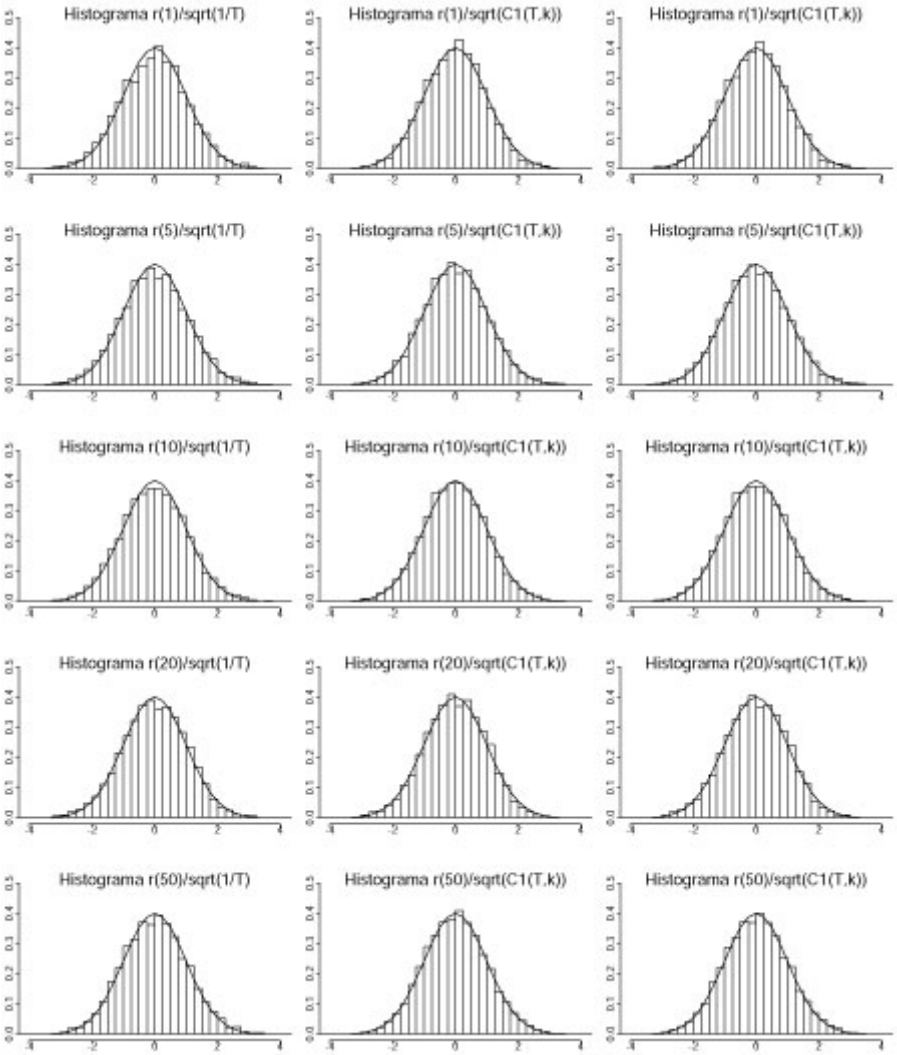


Gráfico 2.11. Distribución empírica de las correlaciones muestrales de y_t sin corregir y corregidas, teórica y muestralmente, por volatilidad, para k=1,5,10,20,50, junto con la distribución N(0,1). El modelo simulado es un LMSV con d=0.4, σ_η^2=0.5, T=8192.

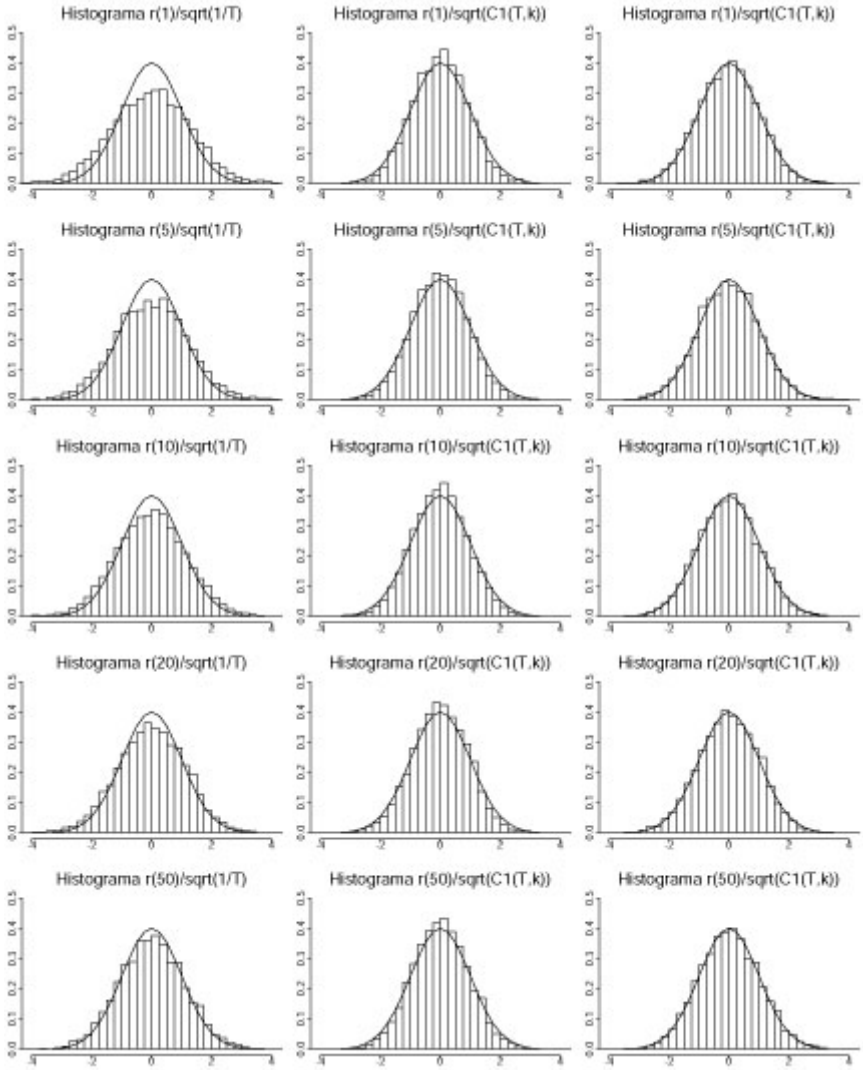


Gráfico 2.12. Distribución empírica de las correlaciones muestrales de y_t sin corregir y corregidas, teórica y muestralmente, por volatilidad, para k=1,5,10,20,50, junto con la distribución N(0,1). El modelo simulado es un LMSV con d=0.49, σ_η^2=0.1, T=512.

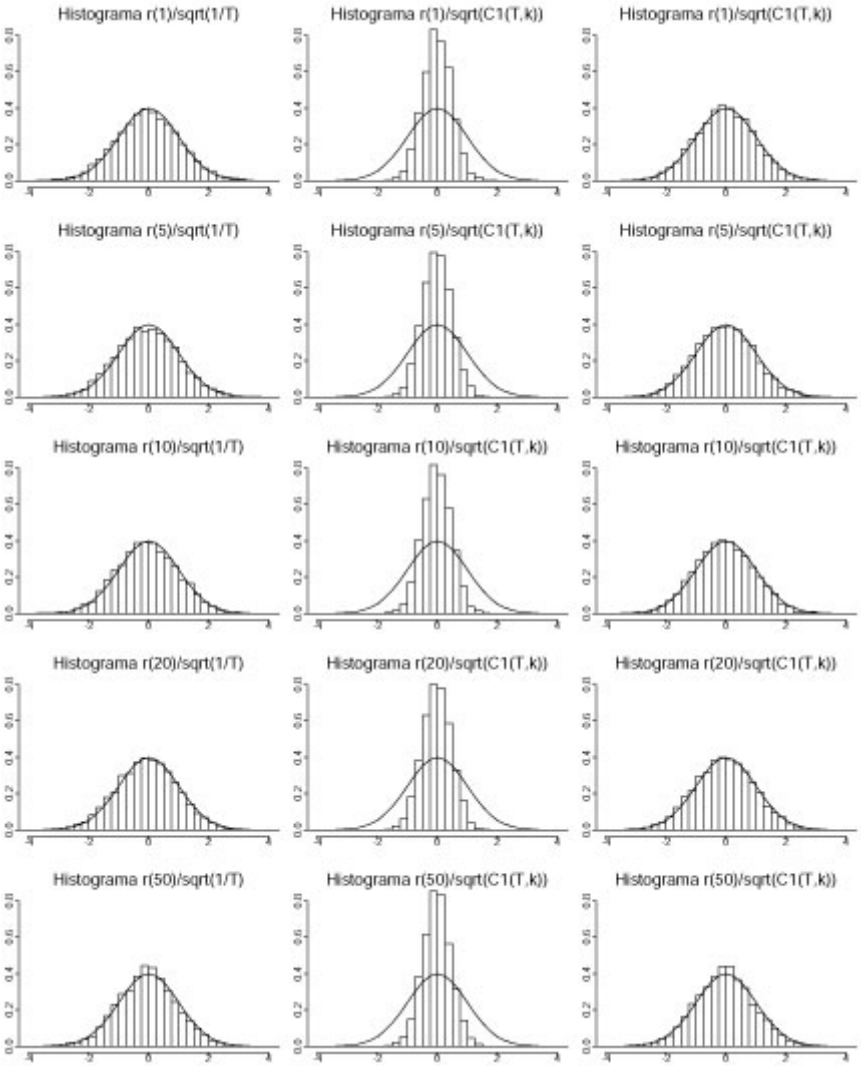
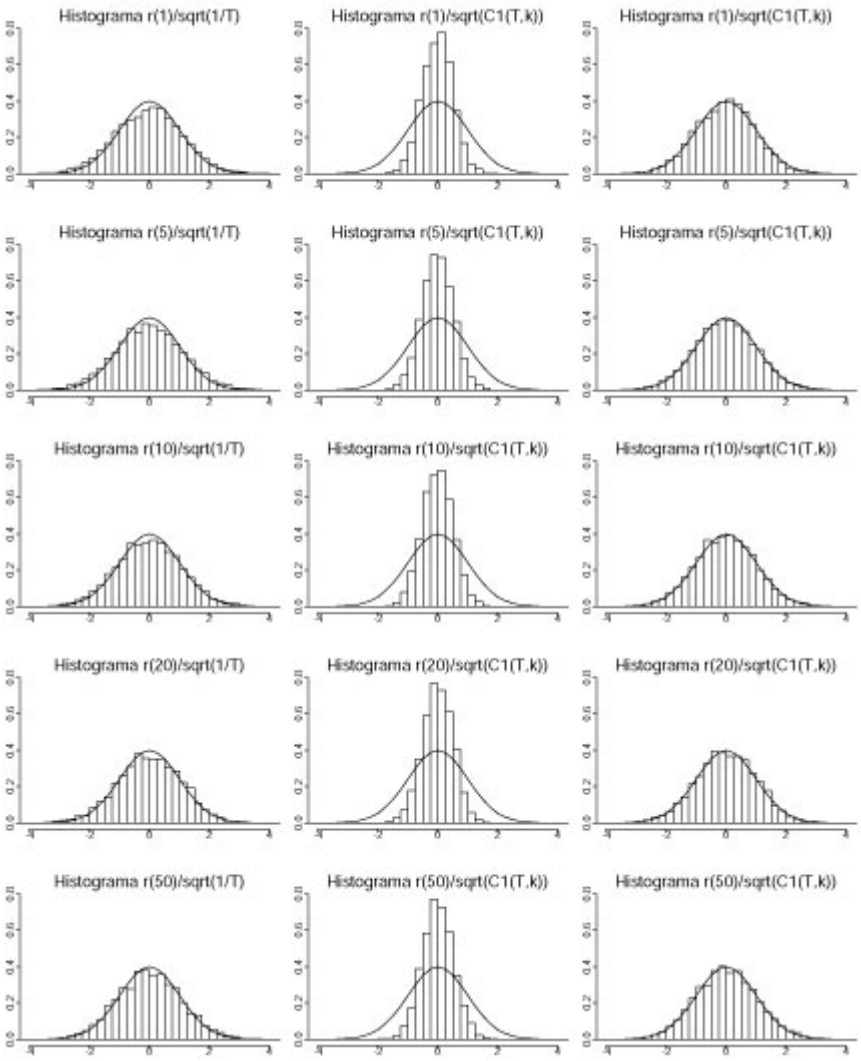


Gráfico 2.13. Distribución empírica de las correlaciones muestrales de y_t sin corregir y corregidas, teórica y muestralmente, por volatilidad, para k=1,5,10,20,50, junto con la distribución N(0,1). El modelo simulado es un LMSV con d=0.49, σ_η^2=0.1, T=8192.



Los gráficos 2.9-2.13 confirman todos los resultados anteriores para modelos LMSV. En los modelos con d=0.4, la distribución asintótica se ajusta perfectamente a la distribución empírica corregida por volatilidad, teórica y muestralmente. Además, en el modelo con σ_η^2=0.1, estas distribuciones son prácticamente iguales a la distribución de $\sqrt{T}\, r_Y(k)$, debido a que en este caso la varianza asintótica corregida por volatilidad es prácticamente igual a la varianza asintótica habitual, 1/T.

Sin embargo, en el modelo con mayor varianza (σ_η^2=0.5), representado en el gráfico 2.11, se aprecia claramente que la distribución asintótica habitual (primera columna) es totalmente inapropiada, especialmente en los primeros retardos, donde proporcionaría siempre un tamaño empírico mayor que el nominal. Sin embargo, las distribuciones corregidas por volatilidad siguen mostrando un ajuste casi perfecto a la distribución asintótica.

Por lo que respecta al modelo con d=0.49, está claro que la distribución asintótica no funciona, ni siquiera en muestras grandes (gráfico 2.13). Esto es debido a que el modelo está rayando la frontera de la no estacionariedad, lo cual invalida el corolario 2.1. Sin embargo, la distribución corregida empíricamente por volatilidad (tercera columna) se ajusta perfectamente a la distribución asintótica, para todos los retardos analizados.

En resumen, podemos concluir que la corrección muestral por volatilidad de la varianza asintótica de $r_Y(k)$ en [2.18], que es la que se utilizará en las aplicaciones con series reales, proporciona buenos resultados en modelos SV, tanto en los modelos autorregresivos como en los modelos con memoria larga, incluso en aquellos modelos muy persistentes y con tamaños muestrales no muy grandes.

2.2.3. Efectos SV sobre el estadístico de Box-Pierce

Los resultados de la sección 2.2.1 sobre la distribución asintótica de las correlaciones muestrales de y_t en modelos SV estacionarios, afectan directamente a las propiedades del estadístico de Box-Pierce, Q(K), definido en [2.5]. Recordemos que para procesos IID, la distribución asintótica de la correlación muestral de orden k de y_t es N(0,1/T), y las autocorrelaciones muestrales de distinto orden son asintóticamente independientes y, por tanto, la distribución asintótica del estadístico Q(K) en [2.5] es una χ^2 con K grados de libertad. Para modelos SV, acabamos de probar que la varianza asintótica de $r_Y(k)$ es mayor que 1/T, y en consecuencia, la distribución asintótica de Q(K) en modelos SV ya no será una variable χ^2 con K grados de libertad. Sin embargo, la corrección que debe hacerse para conseguir esa distribución es inmediata.

Según la proposición 2.1, en los modelos SV estacionarios con momento de orden cuarto finito y distribución de ε_t simétrica, se tiene que:

$$\left[C_1(T,j)\right]^{-1/2} r_Y(j) \xrightarrow{d} N(0,1)$$

donde $C_1(T,j)$ es la varianza asintótica de $r_Y(j)$ en [2.7]. A partir de este resultado, se define la siguiente versión del estadístico de Box-Pierce corregida por volatilidad estocástica, que denotaremos Q_c:

$$Q_c(K) = \sum_{j=1}^{K} \frac{[r_Y(j)]^2}{C_1(T,j)} = T\sum_{j=1}^{K} [r_Y(j)]^2 \exp\{-\gamma_h(j)\} \qquad [2.20]$$

cuya distribución asintótica es una χ^2 con K grados de libertad.

Para modelos ARSV(1) con $\phi>0$ ó modelos LMSV con $d>0$, $\gamma_h(j)$ es siempre estrictamente mayor que cero y, en consecuencia, $\exp\{-\gamma_h(j)\}<1$. Por tanto, cada sumando del estadístico $Q_c(K)$ en [2.20] es menor que el correspondiente sumando del estadístico Q(K) de Box-Pierce definido en [2.5]. Esto significa que la utilización de dicho estadístico en modelos SV conllevaría un tamaño del contraste mayor que el nominal y en algunas ocasiones se rechazaría erróneamente la hipótesis de incorrelación conjunta.

En la práctica, el proceso h_t no es observable y la varianza asintótica de $r_Y(k)$ debe estimarse a partir de las observaciones $\{y_1,\ldots,y_T\}$, según se explicó en la sección 2.2.1. Esto significa que en el estadístico $Q_c(K)$ definido en [2.20], deberá sustituirse $C_1(T,j)$ por su estimador $\hat{C}_1(T,j)$ en [2.18] y, por tanto, el estadístico de contraste será de la forma:

$$\hat{Q}_c(K) = \sum_{j=1}^{K} \frac{[r_Y(j)]^2}{\hat{C}_1(T,j)} = T\sum_{j=1}^{K} \left\{ \frac{[\hat{\gamma}_Y(0)]^2}{[\hat{\gamma}_Y(0)]^2 + \hat{\gamma}_{Y^2}(j)} \right\} [r_Y(j)]^2 \qquad [2.21]$$

donde $\hat{\gamma}_Y(0)$ y $\hat{\gamma}_{Y^2}(k)$ son la varianza muestral de y_t y la covarianza muestral de y_t^2, respectivamente.

En las simulaciones de la sección siguiente se analizan las propiedades de los estadísticos $Q_c(K)$ y $\hat{Q}_c(K)$ y se comparan con las del estadístico de Box-Pierce Q(K).

2.2.4. Propiedades en muestras finitas del contraste de incorrelación de Box-Pierce en modelos SV

En esta sección se recogen los resultados de diversos experimentos de Monte Carlo para analizar las propiedades en muestras finitas de los contrastes de incorrelación de Box-Pierce presentados en la sección anterior. En concreto

analizamos las propiedades de los estadísticos Q(K), Q_c(K) y $\hat{Q}_c(K)$, definidos en [2.5], [2.20] y [2.21], respectivamente, para K=5, 10, 20 y 50.

El diseño de los experimentos de Monte Carlo es el mismo que en la sección 2.2.2. En primer lugar se analizan modelos ARSV(1) con parámetros ϕ={0.5, 0.9, 0.95, 0.98} y σ_η^2=0.05. En cada caso se ensayan dos tamaños muestrales, T=512 y T=2048, y para algunos modelos se repite el experimento con T=8192. Para cada modelo y cada tamaño muestral se generan 5000 series de tamaño T y para cada una de ellas se calculan las 50 primeras autocorrelaciones. A partir de éstas se obtienen, en cada réplica, los valores de los correspondientes estadísticos de Box-Pierce para K=5,10,20,50, y se estima el tamaño empírico de cada contraste. El mismo experimento se repite para modelos LMSV con parámetros d={0.2, 0.4, 0.45, 0.49} y σ_η^2=0.1. En todos los casos, el nivel de significación de los contrastes será el 5%.

La tabla 2.7 presenta, para cada uno de los modelos ARSV(1) analizados y cada uno de los retardos K=5,10,20,50, los tamaños empíricos de los contrastes de incorrelación que se obtienen utilizando el estadístico de Box-Pierce y utilizando los estadísticos corregidos por volatilidad definidos en [2.20] y [2.21]. Estos valores se denotan por Q, Q_c y $\hat{Q}_c$, respectivamente, y se obtienen de la forma siguiente. Si denotamos por $Q^{(i)}$(K) el valor del estadístico de Box-Pierce para contrastar la significación conjunta de las K primeras autocorrelaciones en la i-ésima réplica, con i=1,2,…,5000, el tamaño empírico de este contraste, que denotamos por Q, es igual a la proporción de estos 5000 valores $Q^{(i)}$(K) que exceden el percentil 95% de la distribución χ_K^2; los valores de Q_c y $\hat{Q}_c$ se obtienen de forma análoga. La tabla 2.8 recoge los mismos resultados para los modelos LMSV analizados

Tabla 2.7. Tamaño empírico del contraste de incorrelación de Box-Pierce, corregido y sin corregir por volatilidad, en un modelo ARSV(1) con σ_η^2 =0.05 y diferentes valores de ϕ.[(c)]

T. muestral		T=512				T=2048				T=8192	
Parámetros		**ϕ=0.5**	**ϕ=0.9**	**ϕ=0.95**	**ϕ=0.98**	**ϕ=0.5**	**ϕ=0.9**	**ϕ=0.95**	**ϕ=0.98**	**ϕ=0.9**	**ϕ=0.98**
K=5	Q	0.054	0.098	0.171	0.338	0.051	0.101	0.203	0.486	0.099	0.562
	Q_c	0.050	0.048	0.038	0.012	0.049	0.049	0.048	0.035	0.047	0.053
	$\hat{Q}_c$	0.050	0.049	0.048	0.045	0.048	0.051	0.049	0.051	0.047	0.051
K=10	Q	0.047	0.103	0.202	0.437	0.050	0.103	0.240	0.629	0.108	0.740
	Q_c	0.045	0.049	0.043	0.014	0.048	0.050	0.051	0.046	0.048	0.055
	$\hat{Q}_c$	0.048	0.048	0.051	0.052	0.048	0.051	0.053	0.055	0.050	0.050
K=20	Q	0.052	0.094	0.211	0.503	0.048	0.094	0.269	0.748	0.105	0.854
	Q_c	0.049	0.048	0.040	0.015	0.046	0.044	0.052	0.052	0.053	0.067
	$\hat{Q}_c$	0.050	0.050	0.050	0.054	0.046	0.046	0.050	0.052	0.053	0.052
K=50	Q	0.041	0.063	0.152	0.452	0.048	0.078	0.230	0.790	0.079	0.922
	Q_c	0.041	0.040	0.043	0.012	0.047	0.048	0.050	0.057	0.047	0.085
	$\hat{Q}_c$	0.044	0.041	0.050	0.058	0.045	0.049	0.050	0.054	0.047	0.055

[(c)] Los valores Q, Q_c y $\hat{Q}_c$ denotan los tamaños empíricos del contraste de incorrelación conjunta utilizando el estadístico de Box-Pierce en [2.5], el estadístico de Box-Pierce corregido por el efecto SV teórico, definido en [2.20], y el estadístico corregido por el efecto SV estimado en [2.21], respectivamente. Los cálculos se basan en 5000 réplicas. El tamaño nominal es el 5%.

Tabla 2.8 Tamaño empírico del contraste de incorrelación de Box-Pierce, corregido y sin corregir por volatilidad, en un modelo LMSV con σ_η^2 =0.1 y distintos valores de d.[(d)]

T. muestral		T=512				T=2048				T=8192	
Parámetros		**d=0.2**	**d=0.4**	**d=0.45**	**d=0.49**	**d=0.2**	**d=0.4**	**d=0.45**	**d=0.49**	**d=0.4**	**d=0.49**
K=5	Q	0.050	0.058	0.063	0.069	0.050	0.063	0.072	0.085	0.072	0.100
	Q_c	0.046	0.035	0.020	0.000	0.047	0.039	0.024	0.000	0.045	0.000
	$\hat{Q}_c$	0.048	0.045	0.045	0.046	0.047	0.049	0.050	0.049	0.052	0.054
K=10	Q	0.050	0.063	0.068	0.075	0.051	0.068	0.084	0.101	0.070	0.118
	Q_c	0.048	0.035	0.015	0.000	0.049	0.037	0.017	0.000	0.039	0.000
	$\hat{Q}_c$	0.046	0.048	0.048	0.048	0.049	0.046	0.046	0.047	0.047	0.049
K=20	Q	0.050	0.062	0.069	0.079	0.047	0.064	0.080	0.106	0.075	0.133
	Q_c	0.048	0.032	0.010	0.000	0.044	0.034	0.012	0.000	0.037	0.000
	$\hat{Q}_c$	0.050	0.050	0.049	0.049	0.044	0.048	0.049	0.050	0.047	0.048
K=50	Q	0.042	0.049	0.056	0.065	0.055	0.078	0.098	0.121	0.076	0.163
	Q_c	0.039	0.024	0.004	0.000	0.052	0.032	0.007	0.000	0.036	0.000
	$\hat{Q}_c$	0.043	0.040	0.042	0.042	0.052	0.052	0.055	0.056	0.050	0.050

[(d)] La notación es la misma que en la tabla 2.7.

El análisis de estas tablas confirma los resultados de las secciones anteriores. En los modelos ARSV(1) con valores pequeños de ϕ, las diferencias entre el tamaño nominal y el tamaño empírico que proporcionan el estadístico de Box-Pierce sin corregir y los estadísticos corregidos por volatilidad son muy pequeñas. Sin embargo, al aumentar el valor de ϕ las diferencias aumentan considerablemente. Por ejemplo, para ϕ=0.95 y T=2048, el tamaño empírico que proporciona el estadístico de Box-Pierce es más de cuatro veces el tamaño nominal, y además esto ocurre con independencia del número de retardos que se contrasten. Por el contrario, el estadístico corregido muestralmente por volatilidad estocástica, $\hat{Q}_c(K)$, se comporta muy bien y mantiene el nivel nominal para todo K y todo el espacio paramétrico.

En los modelos más persistentes (ϕ=0.98), el estadístico con la corrección teórica, $Q_c(K)$, proporciona siempre peores resultados que $\hat{Q}_c(K)$, especialmente en muestras no muy grandes. Como ya hemos comentado en secciones anteriores, esto se debe a que estos modelos están muy próximos a la no estacionariedad (ϕ=1) y la distribución asintótica no es una buena aproximación a la distribución empírica.

Los resultados del modelo LMSV, recogidos en la tabla 2.8, muestran características similares a las del modelo ARSV(1), aunque ahora las diferencias entre el estadístico corregido $\hat{Q}_c(K)$ y el estadístico de Box-Pierce son mucho más pequeñas. Una vez más observamos que el error que se comete utilizando el estadístico de Box-Pierce es mayor cuanto más persistente es el modelo. En el caso más extremo (d=0.49) el tamaño empírico de este contraste en muestras de tamaño T=2048, dobla prácticamente el tamaño nominal para todos los retardos, mientras que el estadístico con la corrección muestral por volatilidad mantiene siempre el tamaño nominal. Por el contrario, el tamaño empírico del contraste que utiliza la corrección teórica, Q_c, difiere mucho del tamaño nominal, especialmente en los retardos alejados y en los modelos más persistentes (obsérvense los valores de Q_c para d= 0.45 y d=0.49). Además, este problema no se corrige al aumentar el tamaño muestral, debido precisamente a la propiedad de memoria larga del modelo LMSV.

Todas estas características se aprecian igualmente en los gráficos 2.14 y 2.15. El gráfico 2.14 representa la distribución empírica del estadístico de Box-Pierce, Q(K), y de los estadísticos $Q_c(K)$ y $\hat{Q}_c(K)$, correspondientes a las 5000 réplicas del modelo ARSV(1) con $\{\phi=0.9, \sigma_\eta^2=0.05\}$ y T=8192, junto con la densidad de la distribución χ_K^2, para K=5,10,20,50. El gráfico 2.15 representa las mismas distribuciones para un modelo con $\{d=0.98, \sigma_\eta^2=0.05\}$ y T=8192.

En estos gráficos se ve claramente que en los modelos ARSV(1) más próximos a la raíz unitaria (ϕ=0.98), la distribución χ_K^2 es totalmente inapropiada, ya que la distribución empírica del estadístico de Box-Pierce está sistemáticamente sesgada hacia la derecha. Esto hace que el tamaño empírico de este contraste sea siempre mayor que el nominal y, en consecuencia, su utilización puede conllevar la detección de correlaciones espurias. Esto ocurre también en el modelo con ϕ=0.9, aunque en este caso, el efecto es menos pronunciado. Por el contrario, la distribución χ_K^2 se ajusta perfectamente a la distribución empírica del estadístico corregido muestralmente, $\hat{Q}_c(K)$, para los dos modelos analizados y para todos los retardos. Sin embargo, el ajuste a la distribución empírica con la corrección teórica por volatilidad no es nada bueno en el modelo más próximo a la raíz unitaria (ϕ=0.98).

Resultados similares se observan en los gráficos 2.16-2.17, correspondientes a modelos LMSV con $\{d=0.4, \sigma_\eta^2=0.1\}$ y $\{d=0.49, \sigma_\eta^2=0.1\}$, respectivamente, y con T=8192. En este caso, las diferencias entre la distribución empírica del estadístico Q(K) y del estadístico corregido por volatilidad, $\hat{Q}_c(K)$, son mucho más pequeñas que en los modelos autorregresivos, en consonancia con los resultados presentados en la tabla 2.8. Sin embargo, en el modelo más persistente (d=0.49), las diferencias entre ambos estadísticos se agravan y comprobamos, además, que la distribución χ_K^2 no es buena aproximación a la distribución empírica de $Q_c(K)$. Por otra parte, la distribución χ_K^2 se ajusta perfectamente a la distribución empírica del estadístico $\hat{Q}_c(K)$, que es, en definitiva, el que utilizaremos en aplicaciones reales.

Gráfico 2.14. Distribución empírica de los estadísticos Q(K), $Q_c(K)$ y $\hat{Q}_c(K)$ junto con la correspondiente distribución χ^2_K , con K=5,10,20,50. El modelo simulado es un ARSV(1) con ϕ=0.9, σ^2_η=0.05,T=8192.

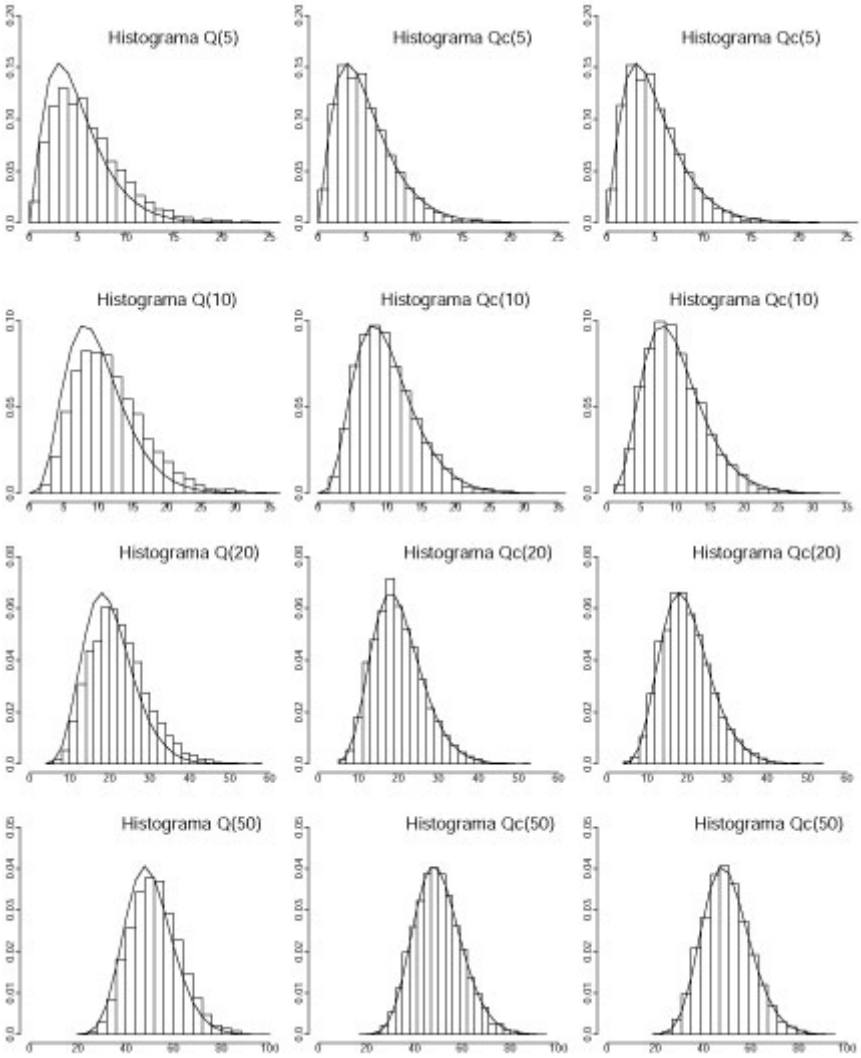


Gráfico 2.15. Distribución empírica de los estadísticos Q(K), $Q_c(K)$ y $\hat{Q}_c(K)$ junto con la correspondiente distribución χ^2_K , con K=5,10,20,50. El modelo simulado es un ARSV(1) con ϕ=0.98, σ^2_η=0.05,T=8192.

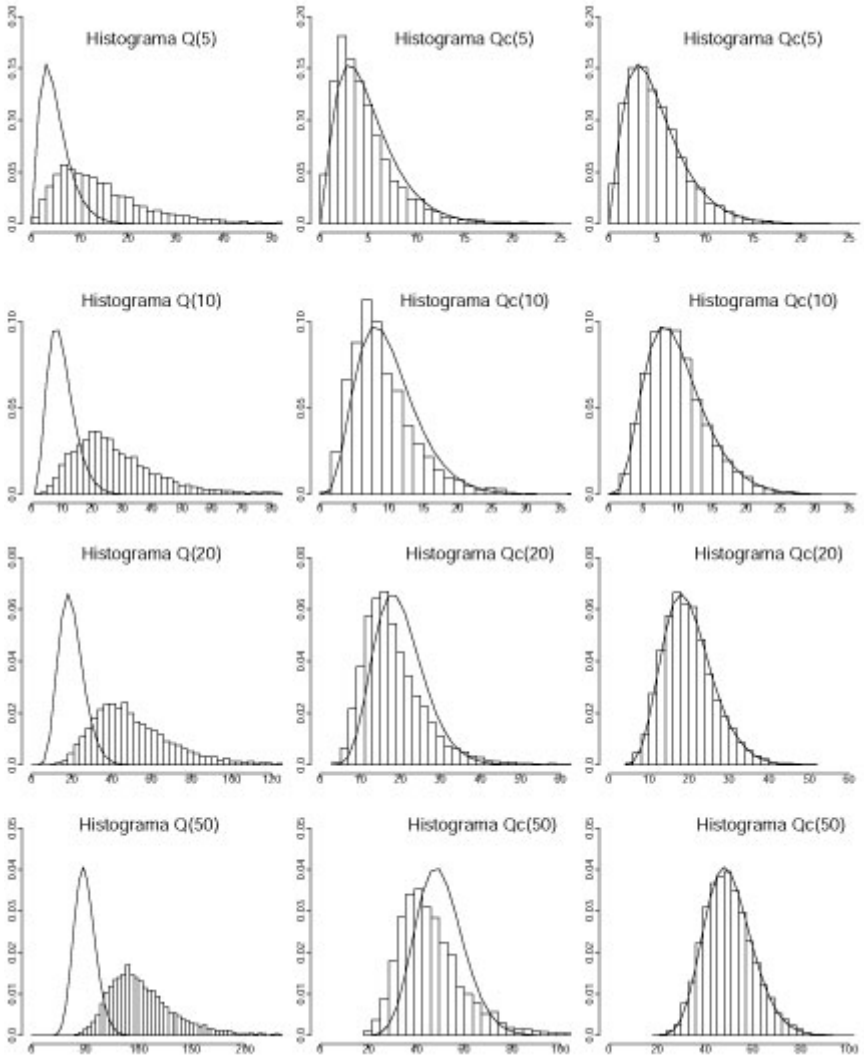


Gráfico 2.16. Distribución empírica de los estadísticos Q(K), $Q_c(K)$ y $\hat{Q}_c(K)$ junto con la correspondiente distribución χ^2_K , con K=5,10,20,50. El modelo simulado es un LMSV con d=0.4, σ^2_η=0.1, T=8192.

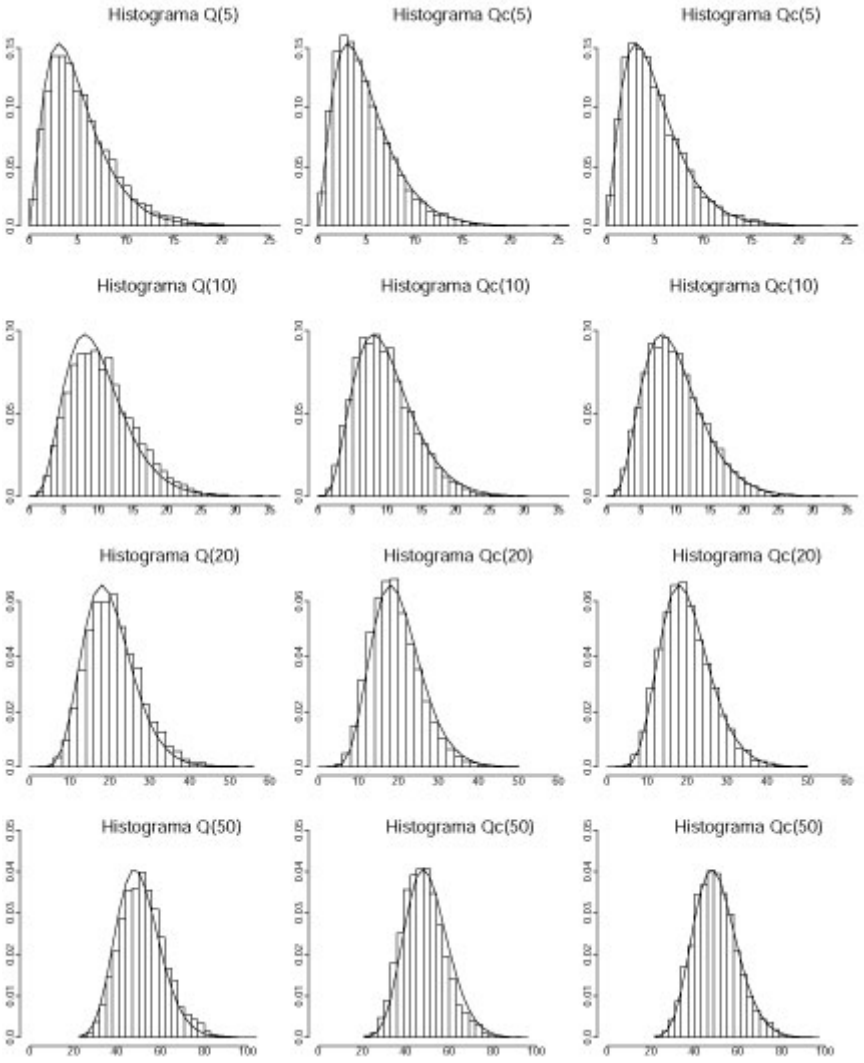
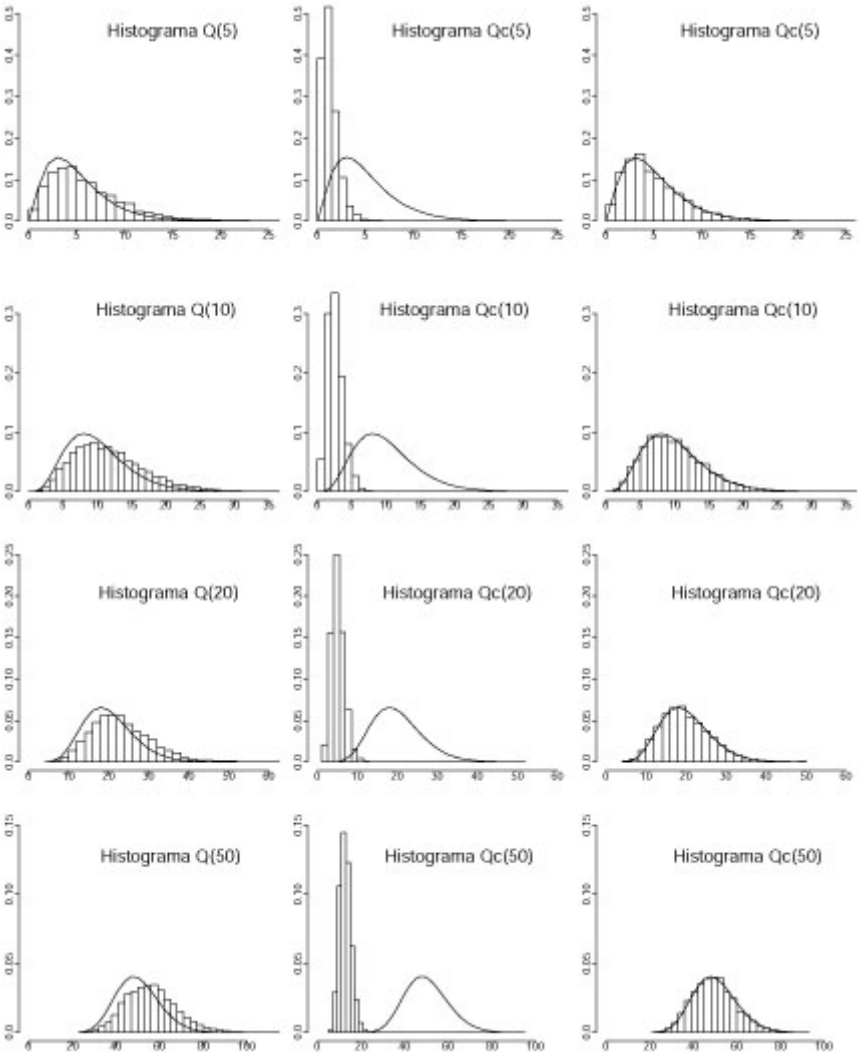


Gráfico 2.17. Distribución empírica de los estadísticos Q(K), $Q_c(K)$ y $\hat{Q}_c(K)$ junto con la correspondiente distribución χ^2_K, con K=5,10,20,50. El modelo simulado es un LMSV con d=0.49, σ^2_η=0.1, T=8192.



2.3. Autocovarianzas muestrales de la serie de logaritmos de los cuadrados

En esta sección se derivan expresiones del sesgo y de la varianza asintótica de las covarianzas muestrales para la serie $x_t=\ln(y_t^2)$ en algunos modelos SV. A pesar de que las covarianzas muestrales por sí mismas puedan tener un interés limitado, debido a que son realmente las correlaciones las que se utilizan en la detección y la identificación de modelos SV, las expresiones que se presentan en esta sección son imprescindibles para poder derivar las propiedades de las correlaciones muestrales que se presentan en la sección siguiente.

El cálculo del sesgo y de la varianza de las covarianzas muestrales de la serie $x_t=\ln(y_t^2)$ en modelos SV puede realizarse de varias formas. Una posibilidad es tratar de aplicar directamente las aproximaciones de Fuller (1996, p.316), verificando que este tipo de modelos cumple las hipótesis que allí se establecen. Dado que estas aproximaciones sólo son válidas para procesos lineales, esto supone, en cada caso, establecer la forma reducida del modelo SV y comprobar las restricciones sobre los momentos de la perturbación del proceso ARMA correspondiente. Otra forma de proceder es aplicar las fórmulas exactas que deriva Anderson (1971, pp. 448-450) para procesos débilmente estacionarios, que sólo exigen condiciones sobre la función de autocovarianza y los cumulantes de cuarto orden, sin restringirse a la clase de modelos lineales. En este último caso, los cálculos son más tediosos, pero los resultados que se obtienen son los mismos.

En el modelo ARSV(1), cualquiera de estas dos soluciones es viable. En concreto, en el apéndice B demostramos que la forma reducida del modelo ARSV(1) es un ARMA(1,1) y, por tanto, las fórmulas de Fuller (1996) pueden aplicarse directamente a la serie x_t. Sin embargo, en el modelo SV con memoria larga, la forma reducida no es inmediata y esto complica la estimación de las covarianzas. Los

resultados de Anderson (1971) tampoco se pueden aplicar, en general, porque el modelo LMSV no verifica ciertas condiciones que en ellos se establecen, debido fundamentalmente a que las autocovarianzas poblacionales de este modelo no son absolutamente sumables. En este caso, recurriremos a la literatura específica sobre modelos con memoria larga. En concreto, utilizaremos los resultados de Hosking (1996) sobre las covarianzas muestrales de series con memoria larga estacionarias, y trataremos de comprobar si sus resultados pueden extenderse al modelo LMSV. Presentaremos además los resultados de Wright (1999) sobre la distribución asintótica de las covarianzas muestrales en modelos LMSV con $0<d<1/4$.

2.3.1. Sesgo y varianza de las autocovarianzas muestrales de la serie de logaritmos de los cuadrados en el modelo ARSV(1)

Como vimos en el capítulo 1, el modelo ARSV(1) estacionario admite una representación lineal para la serie $x_t=\ln(y_t^2)$, de la forma:

$$x_t = \ln(y_t^2) = \mu + h_t + \xi_t \qquad [2.22a]$$

$$h_t = \phi\, h_{t-1} + \eta_t \qquad [2.22b]$$

donde μ es la media de x_t, $\xi_t= \ln(\varepsilon_t^2)-E[\ln(\varepsilon_t^2)]$ es un ruido blanco no gaussiano de media 0 y varianza σ_ξ^2, η_t es $NID(0,\sigma_\eta^2)$ e independiente de ξ_t, y $|\phi|<1$. En esta sección se presentan expresiones del sesgo y de la varianza asintótica de la covarianza muestral de la serie x_t definida en [2.22a].

Dada una serie estacionaria x_t con función de autocovarianza $\gamma(k)$,[10]

[10] De aquí en adelante $\gamma(k)$ denotará la función de autocovarianza de x_t, $\gamma_x(k)=Cov(x_t,x_{t+k})$, omitiéndose el subíndice x siempre que no dé lugar a confusión. Igualmente $\gamma(0)$ denotará la varianza de x_t, $\gamma(0)=Var(x_t)$.

Anderson (1971, p.448) demuestra que el sesgo de la covarianza muestral de orden k definida en [2.3], $\tilde{c}(k)$, admite una expresión de la forma:

$$E(\tilde{c}(k)) - \gamma(k) = Var(\overline{X}) - \frac{1}{T(T-k)} \sum_{s=1}^{T} \sum_{t=1}^{T-k} [\gamma(t-s) + \gamma(t-s+k)] \qquad [2.23]$$

donde $Var(\overline{X})$ puede escribirse como:

$$Var(\overline{X}) = \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} \gamma(t-s) = \frac{\gamma(0)}{T} + \frac{2}{T} \sum_{r=1}^{T-1} (1 - \frac{r}{T}) \gamma(r) \qquad [2.24]$$

siendo $\gamma(0)$ la varianza de x_t. Utilizando este resultado, vamos a demostrar la siguiente proposición:

Proposición 2.2. En el modelo ARSV(1) estacionario definido en las ecuaciones [2.22], el sesgo de la covarianza muestral de orden k para la serie $x_t=\ln(y_t^2)$, admite una expresión de la forma:

$$E(c(k)) - \gamma(k) = -\frac{k}{T} \frac{\sigma_\eta^2}{1-\phi^2} \phi^k - \frac{T-k}{T^2} \left[\sigma_\xi^2 + \frac{1+\phi}{1-\phi} \frac{\sigma_\eta^2}{1-\phi^2} \right] + O(T^{-2}) \qquad [2.25]$$

donde $c(k)$ es la covarianza muestral de x_t corregida por la media muestral, definida en [2.2], y σ_η^2 y σ_ξ^2 son las varianzas de las perturbaciones η_t y ξ_t, respectivamente.

Demostración:

En el modelo ARSV(1) definido en [2.22], con $|\phi|<1$ (condición de estacionariedad), la serie $x_t=\ln(y_t^2)$ es estacionaria y, por tanto, el sesgo de $\tilde{c}(k)$ para dicha serie puede calcularse aplicando las fórmulas [2.23] y [2.24]. Estas fórmulas dependen únicamente del valor de la varianza y de la función de

autocovarianza de x_t, cuyas expresiones para el modelo ARSV(1), aparecen en las ecuaciones [1.21] y [1.22], respectivamente, del capítulo 1. Utilizando dichas ecuaciones, la función de autocovarianza de x_t puede escribirse como:

$$\gamma(r) = \begin{cases} \sigma_h^2 + \sigma_\xi^2 & \text{si } r = 0 \\ \gamma_h(r) = \sigma_h^2 \phi^r & \text{si } r \geq 1 \end{cases} \qquad [2.26]$$

donde σ_h^2 es la varianza del proceso h_t,

$$\sigma_h^2 = \sigma_\eta^2 / (1 - \phi^2) \qquad [2.27]$$

y σ_ξ^2 es la varianza de la perturbación ξ_t. Sustituyendo en [2.24] $\gamma(r)$ por su valor en [2.26] y resolviendo los sumatorios correspondientes, se demuestra que:

$$\text{Var}(\overline{X}) = \frac{\gamma(0)}{T} + \frac{2\sigma_h^2}{T}\frac{\phi}{1-\phi} - \frac{2\sigma_h^2}{T^2}\frac{\phi}{1-\phi}\frac{1-\phi^T}{1-\phi} \qquad [2.28]$$

y se obtiene así el valor del primer sumando de la expresión [2.23].

Para evaluar el segundo sumando de la expresión [2.23], hay que tener en cuenta que éste depende del orden de la covarianza que se estime. Si $1 \leq k < T-k-1$, por ejemplo, se comprueba que dicho sumando puede escribirse como:

$$\frac{1}{T(T-k)} \sum_{s=1}^{T} \sum_{t=1}^{T-k} [\gamma(t-s) + \gamma(t-s+k)] =$$

$$= \frac{2}{T}\gamma(0) + \frac{2}{T}\sum_{r=1}^{k}\gamma(r) + \frac{2}{T(T-k)}\sum_{r=1}^{T-k-1}(T-k-r)[\gamma(r) + \gamma(k+r)]$$

y sustituyendo en esta expresión $\gamma(r)$ por su valor en [2.26] se demuestra que:

$$\frac{1}{T(T-k)}\sum_{s=1}^{T}\sum_{t=1}^{T-k}[\gamma(t-s)+\gamma(t-s+k)] =$$

$$= \frac{2}{T}\gamma(0) + \frac{4}{T}\frac{\phi}{1-\phi}\sigma_h^2 - \frac{2}{T(T-k)}\sigma_h^2(1+\phi^k)\frac{\phi}{1-\phi}\frac{1-\phi^{T-k}}{1-\phi} \qquad [2.29]$$

Finalmente, sustituyendo las expresiones [2.28] y [2.29] en la ecuación [2.23] resulta que el sesgo de $\tilde{c}(k)$, para $1 \le k < T-k-1$, puede escribirse como:

$$E(\tilde{c}(k))-\gamma(k) = -\frac{1}{T}\left[\gamma(0) + \frac{2\phi}{1-\phi}\sigma_h^2\right] + \frac{2\sigma_h^2}{T}\frac{\phi}{1-\phi}\left[\frac{k}{T(T-k)}\frac{1-\phi^T}{1-\phi} + \frac{1}{T-k}\frac{\phi^k-\phi^{T-k}}{1-\phi}\right] \qquad [2.30]$$

Para el resto de los valores de k, Anderson (1971, p.449) propone otras fórmulas específicas del sesgo de $\tilde{c}(k)$, que dependen de la varianza y la función de autocovarianza de la serie x_t. Aplicando dichas fórmulas a la serie $x_t=\ln(y_t^2)$, se comprueba que en el modelo ARSV(1), la expresión [2.30] puede utilizarse como fórmula general del sesgo de $\tilde{c}(k)$, para todo k.

En cualquier caso, la expresión [2.30] puede simplificarse si se tiene en cuenta que su último sumando es de orden $O(T^{-2})$. Se obtiene así la siguiente expresión del sesgo de $\tilde{c}(k)$:

$$E(\tilde{c}(k))-\gamma(k) = -\frac{1}{T}\left[\gamma(0) + \frac{2\phi}{1-\phi}\sigma_h^2\right] + O(T^{-2}), \text{ con } k \ge 1 \qquad [2.31]$$

donde $\gamma(0)$ representa la varianza de x_t, $\gamma(0)=\sigma_h^2+\sigma_\xi^2$, y σ_h^2 es la varianza de h_t en [2.27]. Nótese que a partir de la fórmula [2.31] se deduce que:

$$\lim_{T\to\infty} T[E(\tilde{c}(k)) - \gamma(k)] = -\left[\gamma(0) + \frac{2\phi}{1-\phi}\sigma_h^2\right] = -\sum_{r=-\infty}^{\infty}\gamma(r)$$

resultado que coincide con el que proporciona el teorema 8.3.2 de Anderson (1971).

Para obtener el sesgo de c(k), que es el estimador habitual de la autocovarianza poblacional de orden k definido en [2.2], basta tener en cuenta que $c(k)=\frac{T-k}{T}\tilde{c}(k)$ y, por tanto, resulta que:

$$E(c(k))\text{-}\gamma(k) = -\frac{k}{T}\gamma(k) + \frac{T-k}{T}[E(\tilde{c}(k)) - \gamma(k)]$$

Sustituyendo en esta expresión $E(\tilde{c}(k)) - \gamma(k)$ por su valor en [2.30], resulta que el sesgo de c(k) en el modelo ARSV(1) estacionario admite la siguiente expresión:

$$E(c(k))\text{-}\gamma(k) = -\frac{k}{T}\gamma(k) - \frac{T-k}{T^2}\left[\gamma(0) + \frac{2\phi}{1-\phi}\sigma_h^2\right] + O(T^{-2}) \qquad [2.32]$$

Finalmente, sustituyendo en esta última expresión $\gamma(0)$, $\gamma(k)$ y σ_h^2 por sus valores en [2.26] y [2.27], respectivamente, se obtiene la expresión [2.25] requerida. ■

A partir de la expresión [2.25], que depende del parámetro autorregresivo, ϕ, y de las varianzas de las dos perturbaciones, σ_η^2 y σ_ξ^2, se comprueba fácilmente que, en términos absolutos, el sesgo de c(k) es mayor cuanto mayores son los valores de estos tres parámetros del modelo. Por tanto, cuanto más persistente sea el modelo (más próximo esté ϕ a uno) y mayor sea la varianza de la volatilidad (mayor sea σ_η^2), más sesgadas serán las estimaciones de la covarianza, propiedad que se confirmará con los resultados de las simulaciones que se presentan en la siguiente sección.

La expresión del sesgo de c(k) en [2.25] se puede obtener igualmente aplicando a la serie x_t el teorema 6.2.2 de Fuller (1996) para procesos lineales. Para ello es necesario derivar la forma reducida del modelo ARSV(1) en [2.22] y demostrar que los coeficientes de dicha representación verifican las condiciones exigidas en dicho teorema.[11]

A continuación presentaremos los resultados sobre la varianza de las covarianzas muestrales de la serie $x_t=\ln(y_t^2)$. En este caso, obtener expresiones exactas como la que hemos obtenido en [2.30] para el sesgo, resulta más complicado. Por ello, y siguiendo la práctica habitual en la literatura sobre covarianzas muestrales, presentaremos una fórmula asintótica para calcular de forma aproximada la varianza de las covarianzas muestrales de la serie $x_t=\ln(y_t^2)$. Nuevamente, nuestros resultados se basan en resultados generales de Anderson (1971).

El cálculo de la varianza de las covarianzas muestrales de una serie estacionaria x_t requiere condiciones sobre sus momentos de cuarto orden. Anderson (1971, p. 444) exige, en primer lugar, que estos momentos existan y sean finitos, es decir que se verifique la condición:

$$E(x_t-\mu)^4<\infty \qquad [2.33]$$

y en segundo lugar, que verifiquen la siguiente condición de estacionariedad de cuarto orden:

$$E(x_t-\mu)(x_{t+s}-\mu)(x_{t+r}-\mu)(x_{t+q}-\mu) = E(x_0-\mu)(x_s-\mu)(x_r-\mu)(x_q-\mu) \qquad [2.34]$$

Bajo estos supuestos, Anderson (1971, p.468) demuestra que si la serie x_t verifica además las siguientes condiciones:

[11] Para obtener la forma reducida del modelo ARSV(1) véase el apéndice B.

- $$\sum_{r=-\infty}^{\infty} \gamma^2(r) < \infty \qquad [2.35]$$

- $$\left| \sum_{r=-\infty}^{\infty} \kappa_x(k,-r,k-r) \right| < \infty \qquad [2.36]$$

donde $\gamma(r)$ es la función de autocovarianza de la serie x_t y $\kappa_x(s,r,q)$ es el cumulante mixto de cuarto orden de x_t definido como:

$$\kappa_x(s,r,q) = \text{cum}_4(x_t, x_{t+s}, x_{t+r}, x_{t+q}) =$$
$$= E(x_t-\mu)(x_{t+s}-\mu)(x_{t+r}-\mu)(x_{t+q}-\mu) - [\gamma(s)\gamma(r-q) + \gamma(r)\gamma(s-q) + \gamma(q)\gamma(s-r)] \qquad [2.37]$$

entonces:

$$\lim_{T\to\infty} \frac{T^2}{T-k} \text{Var}(c^*(k)) = \sum_{r=-\infty}^{\infty} \gamma^2(r) + \sum_{r=-\infty}^{\infty} \gamma(r)\gamma(r+2k) + \sum_{r=-\infty}^{\infty} \kappa_x(k,-r,k-r) \qquad [2.38]$$

donde c*(k) es la covarianza muestral de orden k con media conocida definida en [2.1]. Utilizando éste y otros resultados de Anderson (1971), vamos a demostrar la siguiente proposición.

Proposición 2.3. En el modelo ARSV(1) estacionario definido en las ecuaciones [2.22], se verifica que:

$$\lim_{T\to\infty} \frac{T^2}{T-k} \text{Var}(c(k)) = \gamma^2(0) + 2\sigma_h^4 \frac{\phi^2}{1-\phi^2} + \sigma_h^2 \phi^{2k} \left[2\sigma_\xi^2 + \sigma_h^2 \left(\frac{1+\phi^2}{1-\phi^2} + 2k \right) \right] \qquad [2.39]$$

donde $c(k)$ es la autocovarianza muestral de x_t definida en [2.2], $\gamma(0)$ es la varianza de x_t, σ_h^2 es la varianza de h_t, y σ_ξ^2 es la varianza de la perturbación ξ_t.

Demostración:

En lugar de trabajar directamente con c(k), lo que vamos a hacer es utilizar el resultado de Anderson (1971) que hemos enunciado anteriormente, para obtener la varianza de c*(k), y después vamos a demostrar que ésta es equivalente, hasta términos de orden $O(T^{-2})$, a la varianza del estimador c(k).

En el apéndice C demostramos que en el modelo ARSV(1) estacionario, la serie $x_t=\ln(y_t^2)$ verifica las condiciones [2.33] y [2.34] sobre los momentos de cuarto orden. Además, demostramos que todos sus cumulantes de cuarto orden son cero, excepto $\kappa_x(0,0,0)=\pi^4$, y en consecuencia se verifica también la condición [2.36]. Con respecto a la condición [2.35], en el modelo ARSV(1) es fácil probar que:

$$\sum_{r=-\infty}^{\infty}\gamma^2(r)=\gamma^2(0)+2\sum_{r=1}^{\infty}\gamma^2(r)=\gamma^2(0)+2\sigma_h^4\frac{\phi^2}{1-\phi^2}<\infty \qquad [2.40]$$

Por tanto, la serie $x_t=\ln(y_t^2)$ del modelo ARSV(1) verifica todas las condiciones de Anderson (1971), y el límite de la varianza de c*(k) puede calcularse aplicando directamente a dicha serie la expresión [2.38].

A continuación evaluamos cada uno de los sumandos de la expresión [2.38]. El valor del primer sumando lo tenemos ya en [2.40]. Además, sabemos que si k>0, el tercer sumando es igual a cero porque todos los cumulantes de cuarto orden de x_t son cero. Lo único que nos faltaría entonces es obtener el valor del segundo sumando, $\sum_{r=-\infty}^{\infty}\gamma(r)\gamma(r+2k)$, donde $\gamma(r)$ denota la función de autocovarianza de x_t. Sustituyendo en este sumatorio $\gamma(r)$ por su expresión para el modelo ARSV(1) en la ecuación [2.26], se demuestra que, para k>0:

$$\sum_{r=-\infty}^{\infty}\gamma(r)\gamma(r+2k)=\sigma_h^2\phi^{2k}\left[2\sigma_\xi^2+\sigma_h^2\left(\frac{1+\phi^2}{1-\phi^2}+2k\right)\right] \qquad [2.41]$$

Finalmente, sustituyendo [2.40] y [2.41] en [2.38], resulta que, para k>0:

$$\lim_{T\to\infty}\frac{T^2}{T-k}\,\mathrm{Var}(c^*(k))=\gamma^2(0)+2\sigma_h^4\frac{\phi^2}{1-\phi^2}+\sigma_h^2\phi^{2k}\left[2\sigma_\xi^2+\sigma_h^2\left(\frac{1+\phi^2}{1-\phi^2}+2k\right)\right] \qquad [2.42]$$

Para terminar la demostración, basta probar que las varianzas de c*(k) y c(k) son asintóticamente equivalentes. Este resultado se prueba en el apéndice D, donde demostramos que la diferencia entre Var(c(k)) y Var(c*(k)) es de orden $O(T^{-2})$. Por tanto, resulta que:

$$\lim_{T\to\infty}\frac{T^2}{T-k}\,\mathrm{Var}(c(k))=\lim_{T\to\infty}\frac{T^2}{T-k}\,\mathrm{Var}(c^*(k))$$

y teniendo en cuenta la expresión [2.42], se obtiene el resultado [2.39] que queríamos demostrar. ■

La proposición 2.3 que acabamos de demostrar nos dice que, en muestras grandes, la varianza de c(k) puede obtenerse, de forma aproximada, como:

$$\mathrm{Var}(c(k))\sim\frac{T-k}{T^2}\left\{\gamma^2(0)+2\sigma_h^4\frac{\phi^2}{1-\phi^2}+\sigma_h^2\phi^{2k}\left[2\sigma_\xi^2+\sigma_h^2\left(\frac{1+\phi^2}{1-\phi^2}+2k\right)\right]\right\} \qquad [2.43]$$

donde el símbolo ~ denota que el cociente entre los términos a su derecha y su izquierda converge hacia uno cuando $T\to\infty$. En la siguiente sección analizamos la validez de la fórmula [2.43] en muestras finitas mediante un experimento de Monte Carlo que considera diferentes modelos ARSV(1) y diferentes tamaños muestrales.

Para terminar esta sección, vamos a obtener la distribución asintótica de c(k) en modelos ARSV(1) estacionarios, utilizando el mismo argumento que utiliza Wrigth (1999) para derivar la distribución asintótica de las covarianzas muestrales de x_t en el modelo LMSV con d<1/4. En concreto, vamos a demostrar el siguiente resultado.

Proposición 2.4. En el modelo ARSV(1) estacionario definido en las ecuaciones [2.22], se verifica que la distribución asintótica de $\sqrt{T}\,(c(k)-\gamma(k))$ es una Normal de media cero y varianza [2.39], donde c(k) y $\gamma(k)$ denotan la covarianza muestral y poblacional de la serie $x_t=\ln(y_t^2)$, respectivamente.

BIBLIOTECA VIRTUAL

Demostración:

La demostración sigue los pasos de Wright (1999) y se basa en los resultados de Hannan (1976) para procesos multivariantes. De hecho, lo que vamos a demostrar es que la distribución asintótica del vector $\mathbf{U}=\sqrt{T}\,\{c(0)-\gamma(0),\ c(k)-\gamma(k)\}'$ es una normal bidimensional.

Consideremos la variable bidimensional $W_t=(h_t,\xi_t)'$, donde h_t y ξ_t son la variable "estado" y la perturbación, respectivamente, de la representación en el espacio de los estados del modelo ARSV(1) definido en [2.22]. La variable W_t puede representarse como un proceso lineal bivariante de la forma:

$$W_t=\sum_{j=0}^{\infty} B(j)\upsilon_{t-j} \qquad [2.44]$$

donde:

$$W_t=\begin{bmatrix} h_t \\ \xi_t \end{bmatrix},\ \upsilon_t=\begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix},\ B(j)=\begin{bmatrix} \beta_{11}(j) & 0 \\ 0 & \beta_{22}(j) \end{bmatrix} \qquad [2.45]$$

Los elementos $\beta_{11}(j)$, j=0,1,2,…, son los coeficientes de la representación de Wold del proceso h_t y los coeficientes $\beta_{22}(j)$ son todos iguales a cero, excepto $\beta_{22}(0)=1$. Como h_t es un proceso AR(1) estacionario, se tiene $\beta_{11}(j)=\phi^j$, para todo j. Además, las perturbaciones η_t y ξ_t son variables incorrelacionadas e independientes entre sí y, por tanto, los momentos de la perturbación υ_t verifican las siguientes condiciones:

$$E(\upsilon_t)=\begin{bmatrix} E(\eta_t) \\ E(\xi_t) \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix},\ E(\upsilon_t\upsilon_s')=\begin{bmatrix} E(\eta_t\eta_s) & E(\eta_t\xi_s) \\ E(\xi_t\eta_s) & E(\xi_t\xi_s) \end{bmatrix}=\delta(t,s)\begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\xi^2 \end{bmatrix} \qquad [2.46]$$

donde $\delta(t,s)$ representa la función delta de Kronecker, $\delta(t,s)=1$ si t=s, $\delta(t,s)=0$ si $t\neq s$.

Utilizando la misma notación de Hannan (1976), denotaremos por $c_{11}(k)$ y $c_{22}(k)$ las autocovarianzas muestrales de h_t y ξ_t, respectivamente:

$$c_{11}(k)=\frac{1}{T}\sum_{t=1}^{T-k}(h_t-\bar{h})(h_{t+k}-\bar{h}),\ c_{22}(k)=\frac{1}{T}\sum_{t=1}^{T-k}(\xi_t-\bar{\xi})(\xi_{t+k}-\bar{\xi}) \qquad [2.47]$$

donde $\bar{h}=\frac{1}{T}\sum_{t=1}^{T}h_t$ y $\bar{\xi}=\frac{1}{T}\sum_{t=1}^{T}\xi_t$. Denotaremos por $c_{12}(k)$ y $c_{21}(k)$ las covarianzas muestrales cruzadas de h_t y ξ_t, definidas como:

$$c_{12}(k)=\frac{1}{T}\sum_{t=1}^{T-k}(h_t-\bar{h})(\xi_{t+k}-\bar{\xi}),\ c_{21}(k)=\frac{1}{T}\sum_{t=1}^{T-k}(\xi_t-\bar{\xi})(h_{t+k}-\bar{h}) \qquad [2.48]$$

Para k=0, $c_{11}(0)$ y $c_{22}(0)$ son las varianzas muestrales de h_t y ξ_t, respectivamente, y $c_{12}(0)=c_{21}(0)$ es la covarianza muestral entre h_t y ξ_t.

De forma análoga, $\gamma_{11}(k)$, $\gamma_{22}(k)$, $\gamma_{12}(k)$, $\gamma_{21}(k)$ denotarán los correspondientes valores poblacionales:

$$\gamma_{11}(k)=Cov(h_t,h_{t+k})=\gamma_h(k),\ \gamma_{22}(k)=Cov(\xi_t,\xi_{t+k})=\sigma_\xi^2 I_{\{k=0\}} \quad [2.49]$$

$$\gamma_{12}(k)=Cov(h_t,\xi_{t+k})=0,\ \gamma_{21}(k)=Cov(\xi_t,h_{t+k})=0 \quad [2.50]$$

donde $I_{\{k=0\}}$ es la función indicador, $I_{\{k=0\}}=1$ si $k=0$, $I_{\{k=0\}}=0$ si $k\neq 0$.

En el modelo ARSV(1) con $|\phi|<1$, la densidad espectral del proceso h_t es una función de cuadrado integrable y se comprueba que la serie W_t en [2.44] verifica las condiciones del teorema central del límite de Hannan (1976) para procesos lineales multivariantes. En virtud de dicho teorema, se demuestra que en el modelo ARSV(1) estacionario, el vector 8x1:

$$\mathbf{Z}=\sqrt{T}\ \{c_{11}(0)-\gamma_{11}(0),\ c_{12}(0),\ c_{21}(0),\ c_{22}(0)-\gamma_{22}(0),\ c_{11}(k)-\gamma_{11}(k),\ c_{12}(k),\ c_{21}(k),\ c_{22}(k)\}' \quad [2.51]$$

con $k>0$, converge asintóticamente hacia una variable Normal con vector de medias igual a cero y con una estructura de covarianzas dada por los siguientes elementos:

$$\lim_{T\to\infty} T\,Cov(c_{ab}(i)-\gamma_{ab}(i),c_{cd}(j)-\gamma_{cd}(j)) =$$

$$=\sum_{u=-\infty}^{\infty}\left[\gamma_{ac}(u)\gamma_{bd}(u+i-j)+\gamma_{ad}(u)\gamma_{bc}(u+i+j)\right]+$$

$$+\sum_{u=-\infty}^{\infty}\left[\sum_{p=1}^{2}\sum_{q=1}^{2}\sum_{r=1}^{2}\sum_{s=1}^{2}\kappa_{pqrs}\sum_{\ell=0}^{\infty}\beta_{ap}(\ell)\beta_{bq}(\ell+i)\beta_{cr}(\ell+u)\beta_{ds}(\ell+u+j)\right] \quad [2.52]$$

donde i=0,k, j=0,k, a=1,2; b=1,2; c=1,2; d=1,2, y κ_{pqrs} es el cumulante de cuarto orden mixto entre las componentes de υ_t:

$\kappa_{pqrs}=$

$=E(\upsilon_t^{(p)}\upsilon_t^{(q)}\upsilon_t^{(r)}\upsilon_t^{(s)})-E(\upsilon_t^{(p)}\upsilon_t^{(q)})E(\upsilon_t^{(r)}\upsilon_t^{(s)})-E(\upsilon_t^{(p)}\upsilon_t^{(r)})E(\upsilon_t^{(q)}\upsilon_t^{(s)})-$
$E(\upsilon_t^{(p)}\upsilon_t^{(s)})E(\upsilon_t^{(q)}\upsilon_t^{(r)})$

con p=1,2; q=1,2; r=1,2; s=1,2, y donde $\upsilon_t^{(1)}=\eta_t$ y $\upsilon_t^{(2)}=\xi_t$.

Para evaluar la expresión [2.52], tendremos en cuenta, en primer lugar, que debido a la independencia entre las perturbaciones η_t y ξ_t, y a la hipótesis de normalidad sobre η_t, el único cumulante distinto de cero es $\kappa_{2222} = E(\xi_t^4) - 3[E(\xi_t^2)]^2 = \pi^4$. En segundo lugar, utilizaremos que $\beta_{12}(j)=\beta_{21}(j)=0$, para todo j, $\beta_{22}(j)=0$ para $j\neq 0$, y $\beta_{22}(0)=1$. Esto significa que el último sumando de la expresión [2.52] sólo será distinto de cero cuando $i=j=\ell=u=0$ y a=b=c=d=2, en cuyo caso tomará el valor $\kappa_{2222}[\beta_{22}(0)]^4=\pi^4$.

Sustituyendo en [2.52] estos valores y los de las expresiones [2.49] y [2.50], se comprueba que los únicos elementos de la matriz de varianzas y covarianzas del vector **Z** distintos de cero son aquellos en los que {a=b=c=d=1}, {a=c=1,b=d=2}, {a=c=2,b=d=1},{a=d=1,b=c=2},{a=d=2,b=c=1} y {a=b=c=d=2}. A continuación evaluaremos cada uno de estos elementos:

- Si {a=b=c=d=1},

$$\lim_{T\to\infty} T\,\mathrm{Var}(c_{11}(k)) = \sum_{u=-\infty}^{\infty}[\gamma_h^2(u)+\gamma_h(u)\gamma_h(u+2k)],\ \text{para } k\geq 0$$

$$\lim_{T\to\infty} T\,\mathrm{Cov}(c_{11}(0),c_{11}(k)) = 2\sum_{u=-\infty}^{\infty}\gamma_h(u)\gamma_h(u+2k)$$

- Si {a=c=1,b=d=2} ó {a=c=2,b=d=1},

$$\lim_{T\to\infty} T\,\mathrm{Var}(c_{12}(k)) = \lim_{T\to\infty} T\,\mathrm{Var}(c_{21}(k)) = \gamma_{11}(0)\gamma_{22}(0) = \sigma_h^2\,\sigma_\xi^2,\ \text{para } k\geq 0$$

$$\lim_{T\to\infty} T\,\mathrm{Cov}(c_{12}(0),c_{12}(k)) = \lim_{T\to\infty} T\,\mathrm{Cov}(c_{21}(0),c_{21}(k)) = \gamma_{11}(k)\gamma_{22}(0) = \sigma_\xi^2\gamma_h(k)$$

- Si {a=d=1,b=c=2} ó {a=d=2,b=c=1},

$$\lim_{T\to\infty} T\,\mathrm{Cov}(c_{12}(0), c_{21}(0)) = \lim_{T\to\infty} T\,\mathrm{Cov}(c_{21}(0), c_{12}(0)) = \sigma_\xi^2\,\sigma_h^2$$

$$\lim_{T\to\infty} T\,\mathrm{Cov}(c_{12}(k), c_{21}(k)) = \lim_{T\to\infty} T\,\mathrm{Cov}(c_{21}(k), c_{12}(k)) = \sigma_\xi^2\gamma_h(2k), \text{ para } k>0$$

$$\lim_{T\to\infty} T\,\mathrm{Cov}(c_{12}(0), c_{21}(k)) = \lim_{T\to\infty} T\,\mathrm{Cov}(c_{21}(0), c_{12}(k)) = \sigma_\xi^2\gamma_h(k)$$

- Si {a=b=c=d=2},

$$\lim_{T\to\infty} T\,\mathrm{Var}(c_{22}(k)) = \begin{cases} 2\sigma_\xi^4 + \pi^4 & \text{si } k = 0 \\ \sigma_\xi^4 & \text{si } k \neq 0 \end{cases}$$

$$\lim_{T\to\infty} T\,\mathrm{Cov}(c_{22}(0), c_{22}(k)) = 0, \text{ para } k>0$$

donde $\gamma_h(k)$ es la función de autocovarianza de h_t, y σ_h^2 y σ_ξ^2 son las varianzas de h_t y ξ_t, respectivamente.

Por tanto, la matriz de varianzas y covarianzas asintótica 8x8 del vector **Z** definido en [2.51], puede escribirse de forma particionada como:

$$\Sigma = \left[\begin{array}{c|c} A_{00} & A_{0k} \\ \hline A_{0k} & A_{kk} \end{array}\right]$$

donde A_{00}, A_{0k} y A_{kk} son matrices 4x4 con los siguientes elementos:

$$A_{00} = \begin{bmatrix} 2\sum_{u=-\infty}^{\infty}\gamma_h^2(u) & 0 & 0 & 0 \\ 0 & \sigma_\xi^2\sigma_h^2 & \sigma_\xi^2\sigma_h^2 & 0 \\ 0 & \sigma_\xi^2\sigma_h^2 & \sigma_\xi^2\sigma_h^2 & 0 \\ 0 & 0 & 0 & 2\sigma_\xi^4 + \pi^4 \end{bmatrix}$$

$$A_{0k} = \begin{bmatrix} 2\sum_{u=-\infty}^{\infty}\gamma_h(u)\gamma_h(u+2k) & 0 & 0 & 0 \\ 0 & \sigma_\xi^2\gamma_h(k) & \sigma_\xi^2\gamma_h(k) & 0 \\ 0 & \sigma_\xi^2\gamma_h(k) & \sigma_\xi^2\gamma_h(k) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_{kk} = \begin{bmatrix} \sum_{u=-\infty}^{\infty}[\gamma_h^2(u)+\gamma_h(u)\gamma_h(u+2k)] & 0 & 0 & 0 \\ 0 & \sigma_\xi^2\sigma_h^2 & \sigma_\xi^2\gamma_h(2k) & 0 \\ 0 & \sigma_\xi^2\gamma_h(2k) & \sigma_\xi^2\sigma_h^2 & 0 \\ 0 & 0 & 0 & \sigma_\xi^4 \end{bmatrix}$$

BIBLIOTECA VIRTUAL

A partir de la distribución del vector **Z**, vamos a obtener la distribución asintótica de las covarianzas muestrales de la serie $x_t=\ln(y_t^2)$. Para ello, tendremos en cuenta que, según la ecuación [2.22a], la serie x_t puede escribirse como una suma, $x_t=\mu+h_t+\xi_t$, donde los procesos h_t y ξ_t son independientes. Esto significa que la covarianza muestral de x_t, que hemos denotado por c(k), se puede escribir en función de las autocovarianzas y las covarianzas muestrales cruzadas entre h_t y ξ_t, de la siguiente forma:

$$c(k) = c_{11}(k) + c_{12}(k) + c_{21}(k) + c_{22}(k), \text{ para } k\geq 0$$

donde $c_{11}(k)$, $c_{12}(k)$, $c_{21}(k)$, $c_{22}(k)$ son las covarianzas definidas en [2.47] y [2.48]. Además, por la expresión [2.26] sabemos que la función de autocovarianza poblacional de x_t, que hemos denotado por $\gamma(k)$, se puede escribir como:

$$\gamma(k) = \gamma_{11}(k) + \gamma_{22}(k) = \gamma_h(k) + \sigma_\xi^2 I_{\{k=0\}}$$

donde $I_{\{k=0\}}$ es la función indicador definida previamente.

A partir de estas expresiones de c(k) y $\gamma(k)$, se comprueba que el vector:

$$\mathbf{U}=\sqrt{T}\,\{c(0)-\gamma(0),c(k)-\gamma(k)\}'$$

puede escribirse como una transformación lineal del vector **Z** definido en [2.51] de la forma $\mathbf{U}=\Omega\mathbf{Z}$, donde Ω es la matriz 2x8:

$$\Omega=\left[\begin{array}{cccc|cccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]=\left[\begin{array}{c|c} \underline{1}_v' & \underline{0}_v' \\ \hline \underline{0}_v' & \underline{1}_v' \end{array}\right]$$

BIBLIOTECA VIRTUAL

donde $\underline{1}_v'=(1,1,1,1)$ y $\underline{0}_v'=(0,0,0,0)$. Según acabamos de demostrar, el vector **Z** definido en [2.51] tiene una distribución asintótica Normal multivariante con vector de medias cero y matriz de varianzas y covarianzas $\sum$. Utilizando los resultados clásicos para transformaciones lineales de variables normales multivariantes, se demuestra que el vector $\mathbf{U}=\Omega\mathbf{Z}$ se distribuye asintóticamente como una Normal bidimensional con vector de medias cero y matriz de varianzas y covarianzas:

$$\mathbf{V}=\Omega\sum\Omega'=\left[\begin{array}{c|c} \underline{1}_v' & \underline{0}_v' \\ \hline \underline{0}_v' & \underline{1}_v' \end{array}\right]\left[\begin{array}{c|c} A_{00} & A_{0k} \\ \hline A_{0k} & A_{kk} \end{array}\right]\left[\begin{array}{c|c} \underline{1}_v & \underline{0}_v \\ \hline \underline{0}_v & \underline{1}_v \end{array}\right]=\left[\begin{array}{cc} v_{00} & v_{0k} \\ v_{0k} & v_{kk} \end{array}\right] \qquad [2.53]$$

donde:

$$v_{00}=\underline{1}_v' A_{00}\underline{1}_v=2\sum_{u=-\infty}^{\infty}\gamma_h^2(u)+4\sigma_\xi^2\sigma_h^2+2\sigma_\xi^4+\pi^4 \qquad [2.54]$$

$$v_{0k}=\underline{1}_v' A_{0k}\underline{1}_v=2\sum_{u=-\infty}^{\infty}\gamma_h(u)\gamma_h(u+k)+4\sigma_\xi^2\gamma_h(k) \qquad [2.55]$$

$$v_{kk}=\underline{1}_v' A_{kk}\underline{1}_v=\sum_{u=-\infty}^{\infty}[\gamma_h^2(u)+\gamma_h(u)\gamma_h(u+2k)]+2\sigma_\xi^2[\sigma_h^2+\gamma_h(2k)]+\sigma_\xi^4 \qquad [2.56]$$

Utilizando nuevamente las propiedades de la distribución normal multivariante, se demuestra que la variable $\sqrt{T}\{c(k)-\gamma(k)\}$, que es la segunda componente del vector **U**, tiene una distribución Normal unidimensional $N(0,v_{kk})$. Para obtener una expresión de v_{kk} en función de los parámetros del modelo ARSV(1), basta sustituir en la expresión [2.56] $\gamma_h(u)$ por su valor, $\gamma_h(u)=\sigma_h^2\phi^u$, y σ_h^2 por su valor en [2.27]. Se obtiene así una expresión de la varianza asintótica de $\sqrt{T}\,c(k)$ que es precisamente la que obtuvimos en [2.39]. ■

2.3.2. Propiedades en muestras finitas de las covarianzas muestrales de la serie de logaritmos de los cuadrados en el modelo ARSV(1)

Para analizar la validez en muestras finitas de los resultados de la sección anterior, se han realizado simulaciones del modelo ARSV(1) con los siguientes valores de los parámetros: $\phi=\{0.9, 0.95, 0.98\}$ y $\sigma_\eta^2=\{0.5, 0.1, 0.01\}$[12]. En todos los casos se experimentan tres tamaños muestrales, T=512, T=1024 y T=4096, elegidos como potencias de dos por las razones expuestas en la sección 2.2. Para cada modelo y cada tamaño muestral se realizan 5000 réplicas, y en cada réplica se calculan las covarianzas muestrales, hasta de orden 50, de la serie $x_t=\ln(y_t^2)$.

Al igual que en las simulaciones de la sección 2.2, los valores de y_t se obtienen como $y_t=\sigma_*\varepsilon_t\exp(\frac{1}{2}h_t)$, con $\sigma_*=1$, donde las realizaciones del ruido blanco gaussiano ε_t, t=1,...,T, se generan con el comando RNDNS de GAUSS. Haciendo $x_t=\ln(y_t^2)$ se consigue la serie $\{x_1,\ldots,x_T\}$ a partir de la cual se obtienen las covarianzas y las correlaciones muestrales cuyas propiedades queremos analizar.

---

[12] Elegimos estos valores de ϕ y σ_η^2 porque corresponden a los que habitualmente se obtienen en aplicaciones empíricas del modelo ARSV(1). Véase por ejemplo Sandmann y Koopman (1998).

La notación que utilizaremos en esta sección es la siguiente. Si denotamos por $c^{(i)}(k)$ la covarianza muestral de orden k de la serie x_t en la i-ésima réplica, entonces el valor medio, el sesgo y la desviación típica muestral de c(k) sobre las 5000 réplicas vienen dados por las siguientes expresiones:

$$\bar{c}(k) = \frac{1}{5000}\sum_{i=1}^{5000} c^{(i)}(k)\,,\ sesgo = \bar{c}(k) - \gamma(k)\,,\ \text{s.d.} = \sqrt{\frac{1}{5000}\sum_{i=1}^{5000}[c^{(i)}(k) - \bar{c}(k)]^2} \qquad [2.57]$$

respectivamente, donde $\gamma(k)$ es la covarianza poblacional de orden k de la serie x_t. La misma notación se utilizará más adelante cuando presentemos los resultados de las covarianzas muestrales del modelo LMSV.

Las tablas 2.9-2.11 recogen los valores del sesgo y la desviación estándar de la covarianza muestral de la serie x_t en los modelos ARSV(1) con ϕ=0.9, ϕ=0.95 y ϕ=0.98, respectivamente. Para cada modelo se presentan los valores del sesgo teórico asintótico y de la desviación estándar asintótica, obtenidos a partir de las fórmulas [2.25] y [2.43], respectivamente, junto a los correspondientes valores muestrales calculados a partir de los resultados de las 5000 réplicas, según se indica en [2.57]. Se incluyen además los valores de la función de autocovarianza poblacional de la serie $x_t=\ln(y_t^2)$ para los cuatro retardos analizados, k=1,10,20,50.

Tabla 2.9. Sesgo y desviación estándar de c(k) para la serie $\ln(y_t^2)$ en el modelo ARSV(1) con ϕ=0.9 y diferentes valores de σ_η^2, junto con las autocovarianzas teóricas $\gamma(k)$ de dicha serie. [(a)]

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	$\gamma(k)$	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
σ_η^2=0.5	1	2.368	-0.112 (0.627)	-0.117 (0.605)	-0.056 (0.444)	-0.050 (0.436)	-0.014 (0.222)	-0.015 (0.221)
	10	0.918	-0.123 (0.526)	-0.131 (0.506)	-0.062 (0.374)	-0.059 (0.365)	-0.016 (0.188)	-0.014 (0.189)
	20	0.320	-0.116 (0.478)	-0.120 (0.462)	-0.059 (0.341)	-0.057 (0.333)	-0.015 (0.172)	-0.017 (0.170)
	50	0.014	-0.098 (0.453)	-0.096 (0.427)	-0.052 (0.329)	-0.056 (0.328)	-0.013 (0.168)	-0.014 (0.164)
σ_η^2=0.1	1	0.474	-0.030 (0.275)	-0.032 (0.272)	-0.015 (0.195)	-0.011 (0.195)	-0.004 (0.098)	-0.004 (0.095)
	10	0.184	-0.032 (0.255)	-0.037 (0.251)	-0.016 (0.181)	-0.016 (0.182)	-0.004 (0.091)	-0.003 (0.092)
	20	0.064	-0.030 (0.247)	-0.033 (0.246)	-0.016 (0.176)	-0.017 (0.177)	-0.004 (0.089)	-0.006 (0.089)
	50	0.003	-0.027 (0.238)	-0.029 (0.235)	-0.014 (0.173)	-0.014 (0.176)	-0.004 (0.088)	-0.004 (0.088)
σ_η^2=0.01	1	0.047	-0.012 (0.222)	-0.013 (0.223)	-0.006 (0.157)	-0.003 (0.157)	-0.002 (0.079)	-0.002 (0.077)
	10	0.031	-0.012 (0.219)	-0.016 (0.217)	-0.006 (0.155)	-0.006 (0.157)	-0.002 (0.078)	-0.001 (0.079)
	20	0.006	-0.011 (0.216)	-0.014 (0.218)	-0.006 (0.154)	-0.009 (0.156)	-0.002 (0.078)	-0.003 (0.079)
	50	0.000	-0.011 (0.210)	-0.015 (0.210)	-0.006 (0.152)	-0.006 (0.154)	-0.002 (0.078)	-0.002 (0.078)

[(a)] Para cada retardo y cada tamaño muestral se presentan el sesgo y la desviación estándar (entre paréntesis) asintóticos calculados a partir de las fórmulas [2.25] y [2.43], respectivamente, junto con los correspondientes valores simulados de Monte Carlo: el sesgo medio y su desviación estándar muestral (entre paréntesis) sobre todas las réplicas. El número de réplicas es 5000.

Tabla 2.10. Sesgo y desviación estándar de c(k) para la serie $\ln(y_t^2)$ en el modelo ARSV(1) con ϕ=0.95 y diferentes valores de σ_η^2, junto con las autocovarianzas teóricas γ(k) de dicha serie. [(b)]

Tamaño muestral			**T=512**		**T=1024**		**T=4096**	
Parámetros	k	γ(k)	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
σ_η^2=0.5	1	4.872	-0.409 (1.494)	-0.410 (1.385)	-0.205 (1.057)	-0.190 (1.005)	-0.051 (0.529)	-0.052 (0.526)
	10	3.070	-0.452 (1.369)	-0.458 (1.261)	-0.228 (0.973)	-0.218 (0.922)	-0.057 (0.488)	-0.056 (0.486)
	20	1.838	-0.456 (1.222)	-0.459 (1.105)	-0.232 (0.873)	-0.225 (0.819)	-0.059 (0.440)	-0.061 (0.431)
	50	0.395	-0.400 (1.034)	-0.394 (0.873)	-0.210 (0.751)	-0.219 (0.703)	-0.054 (0.383)	-0.057 (0.372)
σ_η^2=0.1	1	0.974	-0.089 (0.406)	-0.092 (0.387)	-0.045 (0.287)	-0.039 (0.279)	-0.011 (0.144)	-0.011 (0.141)
	10	0.614	-0.098 (0.375)	-0.103 (0.359)	-0.050 (0.267)	-0.047 (0.260)	-0.012 (0.134)	-0.012 (0.134)
	20	0.368	-0.099 (0.347)	-0.101 (0.330)	-0.050 (0.248)	-0.051 (0.240)	-0.013 (0.125)	-0.014 (0.124)
	50	0.079	-0.087 (0.314)	-0.089 (0.289)	-0.046 (0.228)	-0.047 (0.225)	-0.012 (0.116)	-0.012 (0.114)
σ_η^2=0.01	1	0.097	-0.018 (0.228)	-0.019 (0.228)	-0.009 (0.161)	-0.005 (0.161)	-0.002 (0.081)	-0.002 (0.079)
	10	0.061	-0.018 (0.223)	-0.023 (0.222)	-0.009 (0.159)	-0.009 (0.160)	-0.002 (0.080)	-0.002 (0.081)
	20	0.037	-0.018 (0.220)	-0.021 (0.221)	-0.009 (0.157)	-0.012 (0.158)	-0.002 (0.079)	-0.004 (0.080)
	50	0.008	-0.017 (0.212)	-0.021 (0.212)	-0.009 (0.154)	-0.009 (0.156)	-0.002 (0.079)	-0.002 (0.079)

[(b)] La notación es la misma que en la tabla 2.9.

Tabla 2.11. Sesgo y desviación estándar de c(k) para la serie $\ln(y_t^2)$ en el modelo ARSV(1) con ϕ=0.98 y diferentes valores de σ_η^2, junto con las autocovarianzas teóricas γ(k) de dicha serie. [c]

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	γ(k)	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
σ_η^2=0.5	1	12.374	-2.470 (5.593)	-2.277 (4.451)	-1.236 (3.957)	-1.163 (3.506)	-0.309 (1.979)	-0.314 (1.922)
	10	10.317	-2.605 (5.452)	-2.446 (4.294)	-1.324 (3.874)	-1.254 (3.412)	-0.331 (1.944)	-0.335 (1.887)
	20	8.429	-2.685 (5.209)	-2.549 (3.992)	-1.366 (3.720)	-1.317 (3.249)	-0.346 (1.874)	-0.335 (1.811)
	50	4.598	-2.661 (4.446)	-2.568 (2.984)	-1.390 (3.228)	-1.380 (2.707)	-0.359 (1.645)	-0.371 (1.574)
σ_η^2=0.1	1	2.475	-0.502 (1.172)	-0.466 (0.950)	-0.251 (0.829)	-0.232 (0.742)	-0.063 (0.415)	-0.063 (0.403)
	10	2.063	-0.529 (1.139)	-0.501 (0.919)	-0.267 (0.809)	-0.253 (0.720)	-0.067 (0.406)	-0.067 (0.395)
	20	1.686	-0.544 (1.088)	-0.519 (0.852)	-0.277 (0.777)	-0.268 (0.685)	-0.070 (0.391)	-0.073 (0.379)
	50	0.920	-0.539 (0.934)	-0.524 (0.651)	-0.282 (0.678)	-0.279 (0.581)	-0.073 (0.345)	-0.075 (0.331)
σ_η^2=0.01	1	0.248	-0.059 (0.263)	-0.057 (0.253)	-0.029 (0.186)	-0.024 (0.181)	-0.007 (0.093)	-0.007 (0.092)
	10	0.206	-0.061 (0.257)	-0.062 (0.246)	-0.031 (0.183)	-0.029 (0.180)	-0.008 (0.092)	-0.007 (0.092)
	20	0.169	-0.063 (0.251)	-0.063 (0.241)	-0.032 (0.179)	-0.033 (0.176)	-0.008 (0.090)	-0.009 (0.090)
	50	0.092	-0.062 (0.236)	-0.065 (0.224)	-0.032 (0.171)	-0.032 (0.169)	-0.008 (0.087)	-0.008 (0.086)

[c] La notación es la misma que en la tabla 2.9.

Las conclusiones que podemos sacar a la vista de los resultados que se presentan en las tablas 2.9-2.11 son las siguientes:

- El sesgo de c(k) siempre es negativo: la covarianza muestral subestima siempre la función de autocovarianza poblacional. Sin embargo, para los casos relevantes empíricamente (σ_η^2=0.1, 0.01) este sesgo es despreciable en muestras grandes.

- Como era de esperar, el tamaño del sesgo, en valor absoluto, y la desviación estándar disminuyen al aumentar T. Además se comprueba que en todos los casos, la tasa de convergencia de la desviación estándar muestral es $\sqrt{T}$.

- Fijados T y k, tanto el sesgo absoluto como la desviación estándar decrecen al disminuir el valor de σ_η^2, obteniéndose siempre los valores más pequeños cuando σ_η^2=0.01. Realmente, lo que ocurre en estos modelos es que la serie x_t es casi un ruido blanco homocedástico y las covarianzas son prácticamente cero.

- En general, el sesgo asintótico y el sesgo simulado difieren en centésimas o milésimas y, por tanto, podemos considerar que la fórmula [2.25] es una buena aproximación para los tamaños muestrales investigados en este trabajo.[13] Lo mismo puede decirse de la aproximación a la varianza asintótica en [2.43], que proporciona, en general, valores muy similares a los de Monte Carlo, excepto en alguno de los modelos más persistentes (ϕ=0.98).

- La máximas diferencias entre los resultados asintóticos y los simulados se observan siempre en el modelo más próximo a la raíz unitaria (ϕ=0.98) y con mayor volatilidad (σ_η^2=0.5), y especialmente en muestras pequeñas (T=512). Sin embargo, este modelo no es muy relevante en aplicaciones empíricas con series financieras reales, donde el número de observaciones disponibles suele ser mucho mayor y σ_η^2 suele tomar valores más pequeños.

- Comparando los resultados de las tres tablas observamos que, tanto el sesgo como la varianza de las covarianzas muestrales, aumentan al aumentar el valor de ϕ. Por ejemplo, en modelos con $\sigma_\eta^2=0.5$, cuando $\phi=0.95$ ($\phi=0.98$) la desviación estándar de c(k) es casi tres (diez) veces mayor que cuando $\phi=0.9$, y lo mismo ocurre con el sesgo. Esto significa que cuanto más persistente es el modelo (más cerca está h_t de la raíz unitaria) peor se estima la covarianza de x_t y en consecuencia, peor se estimará la correlación, como veremos en la sección 2.4.

En general, podemos concluir que la distribución asintótica de las covarianzas muestrales obtenida en la sección anterior, es una buena aproximación a la distribución muestral de las covarianzas muestrales de la serie $x_t=\ln(y_t^2)$ para el modelo ARSV(1).

2.3.3. Sesgo y varianza de las autocovarianzas muestrales de la serie de logaritmos de los cuadrados en el modelo LMSV

En esta sección se presentan expresiones del sesgo y la varianza de la covarianza muestral de la serie $x_t=\ln(y_t^2)$ en el modelo de volatilidad estocástica con memoria larga (LMSV) estacionario. Como vimos en el capítulo 1, en este modelo el proceso h_t que define la volatilidad ya no es un proceso autorregresivo, sino un proceso ARFIMA(0,d,0), donde d puede ser un número no entero. El modelo LMSV, al igual que el modelo ARSV(1), admite una representación en el espacio de los estados de la forma:

$$x_t = \ln(y_t^2) = \mu + h_t + \xi_t \qquad [2.58a]$$

[13] Experimentos realizados con otros tamaños muestrales (T=256,T=8192) y otros valores de σ_η^2 proporcionan resultados similares.

$$(1\text{-}L)^d h_t = \eta_t \qquad [2.58b]$$

donde las perturbaciones η_t y ξ_t son como en el modelo ARSV(1).

Cuando $0<d<1/2$, el proceso x_t en [2.58a] es estacionario y tiene memoria larga, en el sentido de que sus autocovarianzas decrecen hiperbólicamente hacia cero y no son absolutamente sumables. En particular, se prueba que la función de autocovarianza de la serie x_t verifica la siguiente condición (ver sección 1.3.2):

$$\gamma(k) \sim \sigma_\eta^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} k^{2d-1}, \text{ cuando } k\to\infty \qquad [2.59]$$

BIBLIOTECA VIRTUAL

donde Γ denota la función gamma y el símbolo $\sim$ significa que el cociente entre los términos a su derecha e izquierda converge hacia uno cuando $k\to\infty$. De la expresión [2.59] se deduce que cuando $0<d<1/2$, $\sum_{k=-\infty}^{\infty} |\gamma(k)| = \infty$, y se dice que la serie x_t tiene memoria larga.

Este comportamiento de la función de autocovarianza de x_t hace que algunos de los resultados que hemos aplicado al modelo ARSV(1) ya no sean válidos para el modelo LMSV. Incluso los resultados generales de Anderson (1971), que podrían aplicarse al sesgo de la covarianza muestral y a la varianza de la covarianza muestral en modelos con $d<1/4$, conllevarían unos cálculos muy engorrosos porque las series ya no son geométricas, sino hiperbólicas, y esto complica las operaciones. Por esta razón, al trabajar con el modelo LMSV vamos a recurrir directamente a la literatura específica para procesos con memoria larga, y en particular a los resultados de Hosking (1996) sobre propiedades asintóticas de la media, la covarianza y la correlación muestral, y al resultado de Wright (1999) sobre la distribución de las covarianzas muestrales en modelos LMSV con $d<1/4$.

El sesgo de la covarianza muestral de la serie x_t en el modelo LMSV se

obtiene de forma inmediata a partir del siguiente teorema de Hosking (1996).

Teorema 3(a) de Hosking (1996):

Sea x_t una serie estacionaria cuya función de autocovarianza verifica la condición:

$$\gamma(k) \sim \lambda k^{-\alpha}, \text{ con } 0<\alpha<1,\ \lambda>0, \text{ cuando } k\to\infty \qquad [2.60]$$

Entonces, el sesgo asintótico de la autocovarianza muestral de orden k de la serie x_t viene dado por la siguiente expresión:

$$E(c(k))-\gamma(k) \sim \frac{-2\lambda}{(1-\alpha)(2-\alpha)} T^{-\alpha}, \text{ cuando } T\to\infty \qquad [2.61]$$

donde c(k) *es la autocovarianza muestral de x_t definida en* [2.2].

En el modelo LMSV estacionario (0<d<1/2), la función de autocovarianza de la serie x_t verifica la condición [2.59], que es precisamente un caso particular de la condición [2.60], donde $\lambda = \sigma_\eta^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} > 0$ y α=1-2d (0<α<1). Por tanto, el teorema de Hosking (1996) que acabamos de enunciar se puede aplicar directamente al modelo LMSV estacionario, y el sesgo de c(k) se podrá calcular utilizando la fórmula [2.61]. Sustituyendo en dicha fórmula las constantes λ y α por sus respectivos valores para el modelo LMSV, se obtiene la siguiente aproximación al sesgo asintótico de c(k) para la serie $x_t=\ln(y_t^2)$:

$$E(c(k)) - \gamma(k) \sim -\frac{\sigma_\eta^2}{d(1+2d)} \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} T^{2d-1}, \text{ cuando } T\to\infty \qquad [2.62]$$

Obsérvese que esta aproximación no depende de k, lo que significa que la magnitud

del sesgo de las covarianzas muestrales de la serie x_t en el modelo LMSV es independiente del orden de la covarianza que se estime.

Por lo que respecta al cálculo de la varianza de las covarianzas muestrales, el teorema 3 de Hosking (1996) proporciona fórmulas asintóticas para series con memoria larga que verifiquen, además de la condición [2.60], las siguientes condiciones:

- $x_t = \mu + \sum_{j=0}^{\infty} \psi_j z_{t-j}$ con $z_t \sim IID(0, \sigma_z^2)$ [2.63]
- $\psi_j \sim \delta\, j^{d-1}$, cuando $j \to \infty$ [2.64]
- $E(z_t^4) = \sigma_z^4(3+\kappa_4) < \infty$ [2.65]

Como acabamos de ver, el modelo LMSV verifica la condición [2.60], y por tanto, para poder aplicar a este modelo las fórmulas de la varianza de c(k) de Hosking (1996), habría que comprobar que la serie x_t en [2.58a] verifica además las condiciones [2.63]-[2.65]. La condición [2.65] es una condición necesaria para asegurar que exista el momento de cuarto orden de x_t y sea finito. En el apéndice C se demuestra que en el modelo LMSV, $E(x_t-\mu)^4<\infty$, y, por tanto, debe ocurrir también que $E(z_t^4)<\infty$. Con respecto a las condiciones [2.63] y [2.64], sabemos que por ser x_t una serie débilmente estacionaria, podemos asegurar que admite una representación de Wold de la forma $x_t = \mu + \sum_{j=0}^{\infty} \psi_j z_{t-j}$, con $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ y z_t ruido blanco. Sin embargo, esta representación no proporciona la forma explícita de los coeficientes $\{\psi_j\}$, lo que impide verificar la condición [2.64]. Para ello habría que derivar la forma reducida del modelo estructural dado por las ecuaciones [2.58a] y [2.58b] y verificar que los coeficientes de dicha representación verifican esa condición, pero esta solución es bastante complicada.

Granger y Marmol (1998) analizan el comportamiento de las covarianzas y las

correlaciones muestrales de procesos ARFIMA "contaminados" por un ruido blanco gaussiano, y concluyen que las expresiones del sesgo y la varianza obtenidas por Hosking (1996) también son aplicables a este tipo de procesos. Para ello aproximan los coeficientes $\{\psi_j\}$ de la representación de Wold de la serie contaminada mediante los coeficientes de la representación del proceso ARFIMA. En nuestro caso, la serie x_t del modelo LMSV en [2.58] puede considerarse como un proceso ARFIMA h_t "contaminado" por un ruido no gaussiano ξ_t. Sin embargo, dado que los coeficientes $\{\psi_j\}$ de la representación lineal de h_t sólo son una aproximación a los de x_t, no podemos afirmar que los resultados de Granger y Marmol (1998) puedan aplicarse directamente a la covarianza muestral de la serie x_t en el modelo LMSV.

Alternativamente, Wright (1999) deriva la distribución asintótica de las covarianzas muestrales de la serie $x_t=\ln(y_t^2)$ para un modelo LMSV con $d<1/4$, aplicando los resultados de Hannan (1976) para series multivariantes. En particular, de los resultados de Wright se deduce que si $d<1/4$, la covarianza muestral de orden k de la serie x_t es asintóticamente normal con media $\gamma(k)$ y varianza:

$$\mathrm{Var}(c(k)) \sim \frac{1}{T}\left\{\sum_{u=-\infty}^{\infty}[\gamma_h^2(u)+\gamma_h(u)\gamma_h(u+2k)]+2\sigma_\xi^2[\sigma_h^2+\gamma_h(2k)]+\sigma_\xi^4\right\} \quad [2.66]$$

donde c(k) es la covarianza muestral de x_t definida en [2.2], σ_h^2 y σ_ξ^2 son las varianzas de h_t y ξ_t, respectivamente, y $\gamma_h(k)$ es la función de autocovarianza del proceso h_t.

El argumento que utiliza Wright (1999) para demostrar este resultado es precisamente el que nosotras hemos tomado como referencia para demostrar la normalidad asintótica de las covarianzas muestrales de la serie x_t en el modelo ARSV(1) y, por tanto, creemos innecesario volver a repetirlo aquí. Tan sólo es necesaria alguna puntualización.

La demostración de Wright (1999) se basa en aplicar el teorema central del

límite de Hannan (1976) a las covarianzas muestrales de la variable bidimensional $W_t=(h_t,\xi_t)'$, donde h_t y ξ_t son la variable "estado" y la perturbación, respectivamente, de la representación en el espacio de los estados del modelo LMSV en [2.58].

La aplicación del teorema de Hannan (1976) requiere que la densidad espectral de cada uno de los procesos que forman el vector W_t sea una función de cuadrado integrable. En el modelo LMSV, el proceso h_t es un ARFIMA(0,d,0) y sólo cumple esa condición cuando $d<1/4$. En este intervalo, la serie $W_t=(h_t,\xi_t)'$ verifica todas las condiciones del teorema central del límite de Hannan (1976) y, por tanto, siguiendo el mismo razonamiento que en el modelo ARSV(1), se demuestra que el vector $\mathbf{U}=\sqrt{T}\{c(0)-\gamma(0),c(k)-\gamma(k)\}'$ converge asintóticamente hacia una Normal bidimensional con vector de medias cero y matriz de varianzas y covarianzas [2.53].

A partir de esta distribución, se demuestra de forma inmediata que la distribución asintótica de la variable $\sqrt{T}\{c(k)-\gamma(k)\}$ es Normal, y se comprueba que la varianza asintótica de c(k) admite una expresión como la de la ecuación [2.66]. Esta misma expresión se obtendría aplicando el resultado general de Anderson (1971, p.468) que hemos utilizado para demostrar la proposición 2.3, ya que el modelo LMSV con $d<1/4$ verifica las condiciones [2.35] y [2.36] y, por tanto, la fórmula [2.38] que aplicamos en el modelo ARSV(1) también es válida en este caso.

El mayor inconveniente del resultado de Wright (1999) es que los valores $d<1/4$ apenas tienen interés en aplicaciones con series financieras reales, ya que éstas suelen mostrar mayor persistencia en la volatilidad y, por tanto, estarán asociadas a modelos con valores de d más cercanos a 1/2. Sin embargo, hasta este momento, no conocemos ningún resultado que permita derivar la varianza asintótica de c(k) en modelos LMSV con $1/4\leq d<1/2$. Sin duda alguna éste es uno de los problemas que requieren mayor investigación.

A modo de ejercicio y simplemente como ilustración, supondremos que las conclusiones de Marmol y Granger (1998) para procesos ARFIMA "contaminados" por un ruido, son válidas para el modelo LMSV. En ese caso, los resultados de

Hosking (1996) serían directamente aplicables a la serie x_t y tendríamos las siguientes expresiones para la varianza de c(k) en el intervalo $1/4 \leq d < 1/2$:

$$\text{Var}(c(k)) \sim 4\left[\sigma_\eta^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\right]^2 T^{-1} \ln(T) \text{ , si } d=1/4$$

$$\text{Var}(c(k)) \sim 2K_2\left[\sigma_\eta^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\right]^2 T^{4d-2} \text{ , si } 1/4<d<1/2 \qquad [2.67]$$

donde K_2 es una constante definida como:

$$K_2 = \int_0^1\int_0^1 \left[|x-y|^{-\alpha} - \frac{x^{1-\alpha}+(1-x)^{1-\alpha}+y^{1-\alpha}+(1-y)^{1-\alpha}}{1-\alpha} + \frac{2}{(1-\alpha)(2-\alpha)} \right]^2 dxdy =$$

$$= \frac{-1}{(1-2\alpha)(1-\alpha)} - \frac{4}{(2-\alpha)^2(3-2\alpha)} - \frac{4[\Gamma(1-\alpha)]^2}{\Gamma(4-2\alpha)} \qquad [2.68]$$

En la sección siguiente comparamos los valores de la varianza asintótica de c(k) que proporcionaría la expresión [2.67] en modelos LMSV con $1/4<d<1/2$, con la varianza muestral de c(k) obtenida en muestras finitas. Asimismo, estudiamos la validez en muestras finitas de la fórmula [2.66] de la varianza asintótica de c(k), derivada por Wright (1999) para modelos LMSV con $d<1/4$.

2.3.4. Propiedades en muestras finitas de las covarianzas muestrales de la serie de logaritmos de los cuadrados en el modelo LMSV

En esta sección se analiza la validez en muestras finitas de los resultados asintóticos obtenidos en la sección anterior. Las tablas 2.12-2.15 recogen los valores del sesgo y la desviación estándar de la covarianza muestral de la serie $x_t=\ln(y_t^2)$ en un

modelo LMSV con d=0.2, d=0.4, d=0.45 y d=0.49, respectivamente. Para cada valor de d se consideran tres valores de la varianza, $\sigma_\eta^2=\{1, 0.5, 0.1\}$ y para cada modelo se analizan tres tamaños muestrales, T=512, T=1024 y T=4096. En todos los casos se obtienen 5000 réplicas.

Los valores de d elegidos representan un modelo con d<1/4, al cual pueden aplicarse los resultados asintóticos de Wright (1999), y tres modelos próximos a la frontera de la no estacionariedad (d=1/2), que son los de mayor interés en aplicaciones empíricas; véanse, por ejemplo, los valores de d obtenidos por Ray y Tsay (2000) al aplicar modelos LMSV a diferentes series de rendimientos diarios de acciones e índices bursátiles de la Bolsa de Nueva York.

Con respecto a los valores de σ_η^2, el valor $\sigma_\eta^2=1$ no es muy realista, pero se ha elegido como referencia por ser el único considerado en el trabajo de Breidt, Crato y de Lima (1998) sobre estimación de modelos LMSV, y por homogeneizar los resultados con los de los experimentos de Monte Carlo que se presentan en el capítulo siguiente. El valor $\sigma_\eta^2=0.01$, analizado en los modelos ARSV(1), ahora no se considera porque para este valor de σ_η^2, la serie x_t del modelo LMSV se confunde con un ruido blanco y esto complica la interpretación de los resultados.[14]

Para cada modelo y cada tamaño muestral, las tablas 2.12-2.15 presentan los valores asintóticos del sesgo y de la desviación estándar de la serie $x_t=\ln(y_t^2)$, junto con el sesgo medio y la desviación estándar muestral para todas las réplicas, calculados éstos según se indicó en [2.57]. En todos los modelos analizados, el sesgo asintótico de c(k) se obtiene a partir de la fórmula [2.62]. Sin embargo, en el modelo con d=0.2, la desviación estándar asintótica de c(k) se calcula utilizando la expresión [2.66], mientras que en los modelos con 1/4<d<1/2, se utiliza, a modo de prueba, la expresión [2.67]. En todos los casos, se presenta además el valor de la función de autocovarianza poblacional de la serie $x_t=\ln(y_t^2)$ para los cuatro retardos analizados.

Esta última se obtiene utilizando la fórmula [1.58] del capítulo 1.

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

[14] Los resultados están disponibles a petición del interesado.

Tabla 2.12. Sesgo y desviación estándar de c(k) para la serie $\ln(y_t^2)$ en el modelo LMSV con d=0.2 y diferentes valores de σ_η^2, junto con las autocovarianzas teóricas $\gamma(k)$ de dicha serie. [(a)]

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	$\gamma(k)$	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
$\sigma_\eta^2=1$	1	0.275	-0.024 (0.280)	-0.040 (0.271)	-0.016 (0.198)	-0.021 (0.193)	-0.008 (0.099)	-0.010 (0.098)
	10	0.070	-0.024 (0.274)	-0.041 (0.270)	-0.016 (0.193)	-0.019 (0.194)	-0.008 (0.097)	-0.009 (0.095)
	20	0.046	-0.024 (0.273)	-0.037 (0.267)	-0.016 (0.193)	-0.021 (0.187)	-0.008 (0.096)	-0.008 (0.095)
	50	0.027	-0.024 (0.272)	-0.036 (0.251)	-0.016 (0.192)	-0.019 (0.188)	-0.008 (0.096)	-0.007 (0.094)
$\sigma_\eta^2=0.5$	1	0.137	-0.012 (0.248)	-0.026 (0.241)	-0.008 (0.175)	-0.013 (0.173)	-0.003 (0.088)	-0.006 (0.087)
	10	0.035	-0.012 (0.245)	-0.028 (0.244)	-0.008 (0.173)	-0.012 (0.175)	-0.003 (0.086)	-0.005 (0.086)
	20	0.023	-0.012 (0.244)	-0.025 (0.242)	-0.008 (0.173)	-0.013 (0.175)	-0.003 (0.086)	-0.005 (0.086)
	50	0.013	-0.012 (0.244)	-0.023 (0.227)	-0.008 (0.173)	-0.010 (0.170)	-0.003 (0.086)	-0.003 (0.085)
$\sigma_\eta^2=0.1$	1	0.027	-0.002 (0.224)	-0.015 (0.218)	-0.002 (0.158)	-0.007 (0.157)	-0.001 (0.079)	-0.004 (0.078)
	10	0.007	-0.002 (0.223)	-0.018 (0.223)	-0.002 (0.158)	-0.005 (0.160)	-0.001 (0.079)	-0.003 (0.079)
	20	0.005	-0.002 (0.223)	-0.014 (0.0221)	-0.002 (0.158)	-0.007 (0.155)	-0.001 (0.079)	-0.002 (0.079)
	50	0.003	-0.002 (0.223)	-0.014 (0.208)	-0.002 (0.158)	-0.004 (0.156)	-0.001 (0.079)	-0.001 (0.078)

[(a)] Para cada retardo y cada tamaño muestral se presentan el sesgo y la desviación estándar (entre paréntesis) asintóticos calculados a partir de las fórmulas [2.62] y [2.66], respectivamente, junto con los correspondientes valores simulados de Monte Carlo: el sesgo medio y su desviación estándar muestral (entre paréntesis) sobre todas las réplicas. El número de réplicas es 5000.

Tabla 2.13. Sesgo y desviación estándar de c(k) para la serie $\ln(y_t^2)$ en el modelo LMSV con d=0.4 y diferentes valores de σ_η^2, junto con las autocovarianzas teóricas $\gamma(k)$ de dicha serie. [(b)]

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	$\gamma(k)$	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
$\sigma_\eta^2=1$	1	1.380	-0.554 (0.219)	-0.570 (0.369)	-0.483 (0.190)	-0.488 (0.285)	-0.366 (0.144)	-0.369 (0.182)
	10	0.877	-0.554 (0.219)	-0.578 (0.352)	-0.483 (0.190)	-0.490 (0.277)	-0.366 (0.144)	-0.369 (0.176)
	20	0.763	-0.554 (0.219)	-0.578 (0.337)	-0.483 (0.190)	-0.495 (0.262)	-0.366 (0.144)	-0.370 (0.173)
	50	0.636	-0.554 (0.219)	-0.579 (0.302)	-0.483 (0.190)	-0.497 (0.249)	-0.366 (0.144)	-0.370 (0.170)
$\sigma_\eta^2=0.5$	1	0.690	-0.277 (0.109)	-0.292 (0.281)	-0.241 (0.095)	-0.247 (0.209)	-0.183 (0.072)	-0.186 (0.119)
	10	0.438	-0.277 (0.109)	-0.297 (0.275)	-0.241 (0.095)	-0.247 (0.205)	-0.183 (0.072)	-0.186 (0.116)
	20	0.382	-0.277 (0.109)	-0.295 (0.268)	-0.241 (0.095)	-0.250 (0.197)	-0.183 (0.072)	-0.186 (0.114)
	50	0.318	-0.277 (0.109)	-0.295 (0.246)	-0.241 (0.095)	-0.250 (0.192)	-0.183 (0.072)	-0.185 (0.113)
$\sigma_\eta^2=0.1$	1	0.138	-0.055 (0.022)	-0.069 (0.223)	-0.048 (0.019)	-0.054 (0.162)	-0.037 (0.014)	-0.039 (0.081)
	10	0.088	-0.055 (0.022)	-0.071 (0.227)	-0.048 (0.019)	-0.052 (0.163	-0.037 (0.014)	-0.039 (0.081)
	20	0.076	-0.055 (0.022)	-0.068 (0.225)	-0.048 (0.019)	-0.054 (0.158)	-0.037 (0.014)	-0.039 (0.081)
	50	0.064	-0.055 (0.022)	-0.068 (0.210)	-0.048 (0.019)	-0.051 (0.158)	-0.037 (0.014)	-0.037 (0.080)

[(b)] Para cada retardo y cada tamaño muestral se presentan el sesgo y la desviación estándar (entre paréntesis) asintóticos calculados a partir de las fórmulas [2.62] y [2.67], respectivamente, junto con los correspondientes valores simulados de Monte Carlo: el sesgo medio y su desviación estándar muestral (entre paréntesis) sobre todas las réplicas. El número de réplicas es 5000.

Tabla 2.14. Sesgo y desviación estándar de c(k) para la serie ln(y_t^2) en el modelo LMSV con d=0.45 y diferentes valores de σ_η^2, junto con las autocovarianzas teóricas γ(k) de dicha serie. (c)

Tamaño muestral			**T=512**		**T=1024**		**T=4096**	
Parámetros	k	γ(k)	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
σ_η^2=1	1	2.980	-1.875 (0.321)	-1.891 (0.453)	-1.749 (0.300)	-1.755 (0.377)	-1.523 (0.261)	-1.527 (0.290)
	10	2.376	-1.875 (0.321)	-1.902 (0.431)	-1.749 (0.300)	-1.760 (0.366)	-1.523 (0.261)	-1.528 (0.285)
	20	0.217	-1.875 (0.321)	-1.905 (0.408)	-1.749 (0.300)	-1.767 (0.348)	-1.523 (0.261)	-1.529 (0.280)
	50	0.023	-1.875 (0.321)	-1.910 (0.355)	-1.749 (0.300)	-1.774 (0.323)	-1.523 (0.261)	-1.531 (0.275)
σ_η^2=0.5	1	1.490	-0.937 (0.161)	-0.952 (0.315)	-0.875 (0.150)	-0.880 (0.246)	-0.761 (0.130)	-0.765 (0.166)
	10	1.188	-0.937 (0.161)	-0.959 (0.305)	-0.875 (0.150)	-0.882 (0.240)	-0.761 (0.130)	-0.765 (0.163)
	20	0.108	-0.937 (0.161)	-0.958 (0.294)	-0.875 (0.150)	-0.886 (0.230)	-0.761 (0.130)	-0.765 (0.160)
	50	1.011	-0.937 (0.161)	-0.961 (0.265)	-0.875 (0.150)	-0.888 (0.220)	-0.761 (0.130)	-0.765 (0.158)
σ_η^2=0.1	1	0.298	-0.188 (0.032)	-0.201 (0.227)	-0.175 (0.030)	-0.181 (0.166)	-0.152 (0.026)	-0.155 (0.086)
	10	0.238	-0.188 (0.032)	-0.204 (0.230)	-0.175 (0.030)	-0.179 (0.166)	-0.152 (0.026)	-0.155 (0.085)
	20	0.222	-0.188 (0.032)	-0.201 (0.227)	-0.175 (0.030)	-0.181 (0.161)	-0.152 (0.026)	-0.155 (0.085)
	50	0.202	-0.188 (0.032)	-0.201 (0.212)	-0.175 (0.030)	-0.179 (0.160)	-0.152 (0.026)	-0.153 (0.084)

(c) La notación de esta tabla es la misma que en la tabla 2.13.

Tabla 2.15. Sesgo y desviación estándar de c(k) para la serie ln(y_t^2) en el modelo LMSV con d=0.49 y diferentes valores de σ_η^2, junto con las autocovarianzas teóricas $\gamma(k)$ de dicha serie. [(d)]

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	$\gamma(k)$	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
$\sigma_\eta^2=1$	1	15.719	-14.312 (0.451)	-14.328 (0.566)	-14.114 (0.444)	-14.121 (0.506)	-13.730 (0.432)	-13.734 (0.456)
	10	15.022	-14.312 (0.451)	-14.344 (0.539)	-14.114 (0.444)	-14.129 (0.493)	-13.730 (0.432)	-13.736 (0.452)
	20	14.815	-14.312 (0.451)	-14.350 (0.507)	-14.114 (0.444)	-14.140 (0.471)	-13.730 (0.432)	-13.739 (0.446)
	50	14.546	-14.312 (0.451)	-14.361 (0.432)	-14.114 (0.444)	-14.152 (0.433)	-13.730 (0.432)	-13.743 (0.438)
$\sigma_\eta^2=0.5$	1	7.859	-7.156 (0.225)	-7.171 (0.362)	-7.057 (0.222)	-7.064 (0.302)	-6.864 (0.216)	-6.868 (0.243)
	10	7.511	-7.156 (0.225)	-7.180 (0.350)	-7.057 (0.222)	-7.066 (0.295)	-6.864 (0.216)	-6.869 (0.241)
	20	7.408	-7.156 (0.225)	-7.181 (0.333)	-7.057 (0.222)	-7.072 (0.282)	-6.864 (0.216)	-6.870 (0.237)
	50	7.273	-7.156 (0.225)	-7.186 (0.294)	-7.057 (0.222)	-7.077 (0.265)	-6.864 (0.216)	-6.871 (0.233)
$\sigma_\eta^2=0.1$	1	1.572	-1.431 (0.045)	-1.445 (0.232)	-1.411 (0.044)	-1.417 (0.171)	-1.373 (0.043)	-1.376 (0.094)
	10	1.502	-1.431 (0.045)	-1.448 (0.234)	-1.411 (0.044)	-1.416 (0.170)	-1.373 (0.043)	-1.375 (0.093)
	20	1.482	-1.431 (0.045)	-1.446 (0.230)	-1.411 (0.044)	-1.419 (0.166)	-1.373 (0.043)	-1.375 (0.093)
	50	1.455	-1.431 (0.045)	-1.446 (0.214)	-1.411 (0.044)	-1.417 (0.164)	-1.373 (0.043)	-1.374 (0.092)

[(d)] La notación de esta tabla es la misma que en la tabla 2.13.

Los aspectos más destacables de los resultados que se presentan en las tablas 2.12-2.15 son los siguientes:

- El sesgo de c(k) siempre es negativo y, por tanto, comprobamos una vez más que las covarianzas muestrales subestiman sus análogos poblacionales.
- Como era de esperar, fijados los valores de d y σ_η^2, el sesgo absoluto y la desviación estándar disminuyen al aumentar T. Sin embargo, este decrecimiento es muy lento, especialmente cuando el modelo es muy persistente (d=0.49). A diferencia de lo que ocurría en el modelo ARSV(1), en los modelos LMSV con 1/4<d<1/2, la tasa de convergencia de la desviación estándar muestral ya no es $\sqrt{T}$, aunque sí los es en el modelo con d<1/4, en consonancia con el resultado asintótico de Wright (1999).
- Para cada modelo y cada tamaño muestral, el sesgo y la desviación estándar empíricos son prácticamente constantes para todos los retardos. Esto también ocurre con los valores asintóticos, cuyas expresiones [2.62] y [2.67] no dependen de k, excepto en el modelo con d<1/4, donde la expresión [2.66] sí depende de k.
- Fijado el valor de d y el tamaño muestral, la estimación de la covarianza es mejor cuanto menor es σ_η^2, obteniéndose siempre los valores mínimos del sesgo y de la desviación estándar en los modelos con σ_η^2=0.1. Sin embargo, como ya hemos comentado anteriormente, esto se debe a que en este caso, el proceso LMSV es casi indistinguible de un ruido blanco y el sesgo es despreciable.
- Fijados T y σ_η^2, el sesgo y la varianza de c(k) aumentan al aumentar d. Esto significa, una vez más, que cuanto más persistente es el modelo, menos fiable es la covarianza muestral como estimador de la covarianza poblacional y, en consecuencia, menos fiable será el correlograma, como veremos en la sección 2.4.
- Con respecto a la validez de las aproximaciones asintóticas, podemos decir que la fórmula del sesgo de c(k) en [2.62] funciona bastante bien para todos los modelos

y todos los tamaños muestrales aquí analizados.[15] Apenas hay diferencias entre los valores del sesgo asintótico y el sesgo muestral, excepto en algunos casos con muestras no muy grandes (T=512).

- La expresión de la varianza asintótica de c(k) en [2.66] para modelos con d<1/4, parece una buena aproximación a los valores obtenidos en muestras finitas, y confirma la validez del resultado asintótico de Wright (1999).
- Sin embargo, la aproximación a la varianza asintótica de c(k) en [2.68] para modelos LMSV con 1/4<d<1/2, funciona realmente mal, especialmente cuando σ_η^2=0.1 y d está muy próximo a 1/2 (el valor simulado en algunos casos es casi diez veces el valor asintótico). Esto confirma lo que ya apuntamos en la sección anterior, en el sentido de que los resultados de Hosking (1996) no son directamente aplicables al modelo LMSV y, por tanto, para 1/4≤d<1/2, no tenemos aún ningún resultado asintótico que nos permita obtener un valor aproximado de la varianza de c(k).

2.4. Autocorrelaciones muestrales de la serie de logaritmos de los cuadrados

En esta sección se analizan las propiedades de las correlaciones muestrales de la serie $x_t=\ln(y_t^2)$ en modelos SV. Para el modelo ARSV(1), se derivan expresiones del sesgo y la varianza asintótica de las correlaciones muestrales de dicha serie y se obtiene su distribución asintótica. Para el modelo LMSV, se deriva la distribución asintótica de las correlaciones muestrales de la serie $x_t=\ln(y_t^2)$ cuando d<1/4. Para los modelos con 1/4≤d<1/2, aún no disponemos de ningún resultado asintótico y, por

[15] Se han analizado otros tamaños muestrales obteniéndose resultados similares.

tanto, sólo presentaremos resultados en muestras finitas. No obstante, los resultados de nuestras simulaciones prueban, en primer lugar, que las expresiones que deriva Hosking (1996) para procesos ARFIMA no son directamente aplicables al modelo LMSV, y en segundo lugar, cuestionan la validez de las fórmulas propuestas por Marmol y Granger (1998) para calcular el sesgo y la varianza asintóticos de r(k) en procesos ARFIMA "contaminados" por un ruido no gaussiano.

Los resultados sobre el sesgo de las correlaciones muestrales en modelos lineales que aparecen en Fuller (1996, p.317), Kendall, Stuart y Ord (1968, pp. 550-551) y Marriot y Pope (1954), entre otros, se basan todos en la idea siguiente. La función de autocorrelación poblacional, $\rho(k)=\gamma(k)/\gamma(0)$, se estima mediante la correlación muestral, $r(k)=c(k)/c(0)$, y, por tanto, la diferencia entre ellas puede escribirse como:

$$r(k)-\rho(k) = \frac{1}{c(0)} \{[c(k)-\gamma(k)]-\rho(k)[c(0)-\gamma(0)]\}$$

Esta expresión permite interpretar la diferencia $r(k)-\rho(k)$ como una función de $(c(0),c(k))$. Expandiendo esta función en serie de Taylor sobre el valor $(\gamma(0),\gamma(k))$ hasta términos de orden dos, y aplicando los resultados de Fuller (1996, sec. 5.4) sobre expansiones de Taylor para esperanzas matemáticas, se obtiene una expresión aproximada del sesgo de r(k) en función del sesgo y la varianza de las covarianzas muestrales. Sustituyendo en esa expresión la esperanza y la varianza de las covarianzas muestrales por sus expresiones derivadas previamente, se obtiene la fórmula del sesgo de r(k) en función de los parámetros del modelo.

El mismo tipo de argumento es el que utiliza Hosking (1996) para derivar el sesgo de las correlaciones muestrales en procesos ARFIMA, y ésta será también nuestra forma de proceder en el caso de los modelos de volatilidad estocástica.

2.4.1. Sesgo y varianza de las autocorrelaciones muestrales de la serie de logaritmos de los cuadrados en el modelo ARSV(1)

Fuller (1996, p.317) demuestra que si x_t es un proceso lineal definido como $x_t = \sum_{j=-\infty}^{\infty} \psi_j z_{t-j}$, con coeficientes $\{\psi_j\}$ absolutamente sumables, z_t I.I.D.$(0, \sigma_z^2)$ y $E(z_t^6)=\eta\sigma_z^6$, entonces el sesgo y la varianza de la correlación muestral de orden k de la serie x_t se pueden calcular de la siguiente forma:

$$E(r(k))-\rho(k)= = \frac{-k}{T}\rho(k) - \frac{1-\rho(k)}{\gamma(0)}\mathrm{Var}(\overline{X}) + \frac{\rho(k)\mathrm{Var}(c^*(0)) - \mathrm{Cov}(c(0),c(k))}{[\gamma(0)]^2} + O(T^{-2}) \qquad [2.69]$$

$$\mathrm{Var}(r(k))= = \frac{1}{\gamma^2(0)}[\rho^2(k)\mathrm{Var}(c^*(0)) + \frac{T^2}{(T-k)^2}\mathrm{Var}(c^*(k)) - 2\rho(k)\frac{T}{T-k}\mathrm{Cov}(c^*(0),c^*(k))] + O(T^{-2}) \qquad [2.70]$$

donde $\gamma(0)$ y $\rho(k)$ son la varianza y la autocorrelación poblacional de orden k de x_t, respectivamente; $c^*(0)$ es la varianza muestral de x_t con media conocida, según se definió en [2.1]; y $c(0)$ y $c(k)$ son la varianza y la covarianza muestral de x_t con la media estimada, según se definieron en [2.2]. A partir de este resultado, vamos a demostrar la siguiente proposición.

Proposición 2.5. En el modelo ARSV(1) estacionario definido en las ecuaciones [2.22], el sesgo y la varianza de la correlación muestral de orden k para la serie $x_t=\ln(y_t^2)$, admiten una expresión de la forma:

$$E(r(k))-\rho(k)=-\frac{1}{T}\left[1+\frac{\sigma_h^2}{\gamma(0)}\left(\frac{2\phi}{1-\phi}+(k+1)\phi^k\right)\right]+\frac{2\rho(k)}{T}\frac{\sigma_h^2}{\gamma(0)}\left(\frac{1+2\phi}{1+\phi}-k\right)+$$

$$+\frac{\rho(k)}{T}\frac{\pi^4}{\gamma^2(0)}+\frac{4\rho(k)}{T}\frac{\sigma_h^4}{\gamma^2(0)}\frac{\phi^2}{1-\phi^2}+O(T^{-2}) \qquad [2.71]$$

$$Var(r(k))=\frac{1}{T}+\frac{2}{T}\frac{[1+2\rho^2(k)]}{\gamma^2(0)}\frac{\phi^2}{1-\phi^2}\sigma_h^4+\frac{1}{T}\frac{\rho(2k)}{\gamma(0)}\left[2\sigma_\xi^2+\sigma_h^2\left(\frac{1+\phi^2}{1-\phi^2}+2k\right)\right]-$$

$$-\frac{4}{T}\frac{\rho^2(k)}{\gamma(0)}\left[2\sigma_\xi^2+\sigma_h^2\left(\frac{1+\phi^2}{1-\phi^2}+k\right)\right]+\frac{\rho^2(k)}{T}\left[2+\frac{\pi^4}{\gamma^2(0)}\right]+O(T^{-2}) \qquad [2.72]$$

donde r(k) es la autocorrelación muestral de x_t definida en [2.4]; $\gamma(0)$ y $\rho(k)$ son la varianza y la función de autocorrelación poblacional de x_t, $\gamma(0)=\sigma_h^2+\sigma_\xi^2$ y $\rho(k)=\phi^k\frac{\sigma_h^2}{\sigma_h^2+\sigma_\xi^2}$, respectivamente; σ_h^2 es la varianza de h_t, $\sigma_h^2=\frac{\sigma_\eta^2}{1-\phi^2}$; y σ_ξ^2 es la varianza de la perturbación ξ_t.

BIBLIOTECA VIRTUAL

Demostración:

En el apéndice B se demuestra que la forma reducida del modelo ARSV(1) corresponde a un proceso ARMA(1,1) de la forma:

$$(1-\phi L)(x_t-\mu)=(1+\theta L)z_t$$

donde las perturbaciones z_t son I.I.D. y los parámetros (ϕ,θ) están sujetos a las restricciones $\{0<\phi<1,-1<\theta<0\}$ ó $\{-1<\phi<0,0<\theta<1\}$. A partir de esta representación se demuestra que la serie x_t puede escribirse como un proceso MA(∞) de la forma $x_t=\mu+\sum_{j=0}^{\infty}\psi_j z_{t-j}$, con z_t I.I.D. y con coeficientes $\psi_0=1$, $\psi_1=\phi+\theta$, y $\psi_j=\phi^{j-1}\psi_1$, para $j>1$. Estos coeficientes son absolutamente sumables.

Además, en el apéndice C se comprueba que el momento de orden seis de x_t es finito y, en consecuencia, lo es también el de z_t. Por tanto, la serie x_t verifica las condiciones del resultado de Fuller (1996, p.317) que hemos enunciado previamente y, en consecuencia, el sesgo y la varianza asintóticos de r(k) para dicha serie se pueden obtener aplicando directamente las fórmulas [2.69] y [2.70].

En primer lugar, obtendremos el valor del sesgo de r(k) según la expresión [2.69]. Para ello es necesario obtener previamente los valores de Var($\overline{X}$), Var(c*(0)) y Cov(c(0),c(k)). El valor de Var($\overline{X}$) para el modelo ARSV(1) ya lo obtuvimos en [2.28] y el valor de Var(c*(0)) lo obtendremos a continuación. Por lo que respecta al cálculo de la covarianza entre las covarianzas muestrales, en el apéndice D demostramos que la diferencia entre Cov(c(0),c(k)) y Cov(c*(0),c*(k)) es de orden $O(T^{-2})$ y, por tanto, si utilizamos Cov(c*(0),c*(k)) en lugar de Cov(c(0),c(k)) en la fórmula [2.69], obtendremos una expresión asintóticamente equivalente del sesgo de r(k), hasta términos de orden $O(T^{-2})$. En consecuencia, para evaluar la expresión [2.69] sólo nos resta calcular Var(c*(0)) y Cov(c*(0),c*(k)). Empezaremos por el primero de ellos.

La varianza muestral con media conocida, c*(0), es un estimador insesgado de la varianza poblacional, $\gamma(0)$, y, por tanto, $\text{Var}(c^*(0))=E\{[c^*(0)]^2\}-[\gamma(0)]^2$. Esto significa que para obtener el valor de la varianza de c*(0), sólo es necesario calcular $E\{[c^*(0)]^2\}$, que puede escribirse como:

$$E\{[c^*(0)]^2\}=\frac{1}{T^2}\left\{\sum_{t=1}^{T}E[(x_t-\mu)^4]+2\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}E[(x_t-\mu)^2(x_{t+j}-\mu)^2]\right\} \qquad [2.73]$$

Por tanto, el cálculo de $E\{[c^*(0)]^2$ se reduce a obtener $E[(x_t-\mu)^4]$ y $E[(x_t-\mu)^2(x_{t+j}-\mu)^2]$.

Utilizando los resultados del apéndice C sobre los momentos de cuarto orden de x_t para el modelo ARSV(1), se comprueba que:

$$E[(x_t-\mu)^4]=3[\gamma(0)]^2+\pi^4$$
$$E[(x_t-\mu)^2(x_{t+j}-\mu)^2]= [\gamma(0)]^2+2[\gamma(j)]^2$$

donde $\gamma(j)$ es la función de autocovarianza de x_t en el retardo j, y $\gamma(0)$ es la varianza de x_t. Sustituyendo estos dos valores en la fórmula [2.73], se obtiene:

$$E\{[c^*(0)]^2\} = \frac{1}{T}[\pi^4 + 3\gamma^2(0)] + \frac{T-1}{T}\gamma^2(0) + \frac{4}{T^2}\sum_{t=1}^{T-1}\sum_{r=1}^{T-t}\gamma^2(j)$$

Finalmente, sustituyendo en esta expresión $\gamma(j)$ por su valor en [2.26] y resolviendo los sumatorios correspondientes, se obtiene la siguiente expresión para $E\{[c^*(0)]^2\}$:

$$E\{[c^*(0)]^2\} = \gamma^2(0) + \frac{2}{T}\gamma^2(0) + \frac{\pi^4}{T} + \frac{4}{T}\frac{\phi^2}{1-\phi^2}\sigma_h^4 - \frac{4}{T^2}\frac{\phi^2}{1-\phi^2}\sigma_h^4\frac{1-\phi^{2T}}{1-\phi^2}$$

y por tanto:

$$\text{Var}(c^*(0))=E\{[c^*(0)]^2\}-[\gamma(0)]^2=\frac{1}{T}\left[2\gamma^2(0) + \pi^4 + 4\sigma_h^4\frac{\phi^2}{1-\phi^2}\right] + O(T^{-2}) \qquad [2.74]$$

Si se quiere obtener una expresión de Var(c*(0)) en función de los tres parámetros del modelo, ϕ, σ_η^2 y σ_ξ^2, basta sustituir en [2.74] las funciones $\gamma(0)$ y σ_h^2 por sus respectivos valores en [2.26] y [2.27].

A continuación obtendremos el valor de Cov(c*(0),c*(k)) en el modelo ARSV(1). Esto requiere cálculos más tediosos. Para empezar, c*(k) es un estimador sesgado de $\gamma(k)$, ya que $E(c^*(k))=\frac{T-k}{T}\gamma(k)$ y por tanto:

$$\text{Cov}(c^*(0),c^*(k))=E(c^*(0)c^*(k)) - \frac{T-k}{T}\gamma(0)\gamma(k)$$

donde $E(c^*(0)c^*(k))$ admite una expresión de la forma:

$$E(c^*(0)c^*(k))= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T-k}E[(x_t-\mu)^2(x_s-\mu)(x_{s+k}-\mu)]$$

Teniendo en cuenta la condición [2.34] de estacionariedad en los momentos de cuarto orden, se comprueba que esta última expresión puede escribirse como:

$$\begin{aligned}E(c^*(0)c^*(k)) = \frac{1}{T^2}\Bigg\{ & 2(T-k)E[(x_t-\mu)^3(x_{t+k}-\mu)]+ \\ & +2\sum_{j=1}^{T-k-1}(T-k-j)E[(x_t-\mu)^2(x_{t+j}-\mu)(x_{t+k+j}-\mu)]+ \\ & +(T-k)\sum_{j=1}^{k-1}E[(x_t-\mu)(x_{t+k}-\mu)(x_{t+j}-\mu)^2]\Bigg\}\end{aligned} \qquad [2.75]$$

Por tanto, el cálculo de $E(c^*(0)c^*(k))$ se reduce a evaluar ciertos momentos cruzados de orden cuatro de la serie x_t.

Utilizando nuevamente los resultados del apéndice C sobre los momentos de orden cuatro de x_t en el modelo ARSV(1), se comprueba fácilmente que:

$$\begin{aligned}&E[(x_t-\mu)^3(x_{t+k}-\mu)] = 3\gamma(0)\gamma(k)\\ &E[(x_t-\mu)^2(x_{t+j}-\mu)(x_{t+k+j}-\mu)] = \gamma(0)\gamma(k) + 2\gamma(j)\gamma(k+j)\\ &E[(x_t-\mu)(x_{t+k}-\mu)(x_{t+j}-\mu)^2] = \gamma(0)\gamma(k) + 2\gamma(j)\gamma(k-j)\end{aligned}$$

y sustituyendo estos valores en la fórmula [2.75] se obtiene:

E(c*(0)c*(k))=

$$= \frac{1}{T^2}\left[(T-k)(T+4)\gamma(0)\gamma(k) + 4\sum_{j=1}^{T-k-1}(T-k-j)\gamma(j)\gamma(k+j) + 2(T-k)\sum_{j=1}^{k-1}\gamma(j)\gamma(k-j)\right]$$

Sustituyendo finalmente en esta ecuación $\gamma(j)$ por su valor en [2.26] y evaluando los sumatorios correspondientes, resulta que:

E(c*(0)c*(k))=

$$= \frac{(T+4)(T-k)}{T^2}\gamma(0)\gamma(k) + \frac{2(T-k)}{T^2}\sigma_h^4\phi^k\left(\frac{2\phi^2}{1-\phi^2} + k - 1\right) - \frac{4}{T^2}\sigma_h^4\phi^k\frac{\phi^2}{1-\phi^2}\frac{1-\phi^{2(T-k)}}{1-\phi^2}$$

BIBLIOTECA VIRTUAL

Por tanto, podemos concluir que la covarianza entre las covarianzas muestrales es:

$$\text{Cov}(c^*(0),c^*(k)) = E(c^*(0)c^*(k)) - \frac{T-k}{T}\gamma(0)\gamma(k) =$$

$$= \frac{(T-k)}{T^2}\left[4\gamma(0)\gamma(k) + 2\sigma_h^4\phi^k\left(\frac{2\phi^2}{1-\phi^2} + k - 1\right)\right] + O(T^{-2}),\ \text{si } k \geq 1 \qquad [2.76]$$

Sustituyendo en esta última expresión $\gamma(0)$, $\gamma(k)$ y σ_h^2 por sus valores en [2.26] y [2.27], respectivamente, se obtendría una expresión de Cov(c*(0),c*(k)) en términos los tres parámetros del modelo: las varianzas de las dos perturbaciones, σ_η^2 y σ_ξ^2, y del parámetro autorregresivo, ϕ.

Las expresiones de Var(c*(0)) y Cov(c*(0),c*(k)) que acabamos de presentar en [2.74] y [2.76], respectivamente, también podrían haberse obtenido aplicando directamente las correspondientes fórmulas de Anderson (1971, pp. 450-452), que dependen únicamente de la función de autocovarianza de x_t y del cumulante de cuarto orden de x_t definido en [2.37]. Aplicando estas fórmulas a la serie x_t del modelo ARSV(1), se obtienen resultados idénticos a los obtenidos anteriormente.

Una vez que hemos probado que $Cov(c(0),c(k))=Cov(c^*(0),c^*(k))+O(T^{-2})$ (véase el apéndice D) y hemos obtenido las expresiones de $Var(\overline{X})$, $Var(c^*(0))$ y $Cov(c^*(0),c^*(k))$ en [2.28], [2.74] y [2.76], respectivamente, sustituimos sus valores en la ecuación [2.69]. Después de hacer algunas operaciones, obtenemos la fórmula del sesgo de la autocorrelación muestral de orden k de la serie $x_t=\ln(y_t^2)$ en [2.71] que queríamos demostrar.

Para obtener la varianza asintótica de r(k), basta sustituir en la fórmula [2.70] los valores de $Var(c^*(0))$, $Var(c^*(k))$ y $Cov(c^*(0),c^*(k))$, por sus expresiones [2.74], [2.43] y [2.76], respectivamente. Se obtiene así una expresión de la forma:

$$\begin{aligned}Var(r(k)) = {} & \frac{1}{\gamma^2(0)}\rho^2(k)\frac{1}{T}\left[2\gamma^2(0)+\pi^4+4\frac{\phi^2}{1-\phi^2}\sigma_h^4\right]+ \\ & +\frac{1}{\gamma^2(0)}\frac{1}{T-k}\left\{\gamma^2(0)+2\frac{\phi^2}{1-\phi^2}\sigma_h^4+\sigma_h^2\phi^{2k}\left[2\sigma_\xi^2+\sigma_h^2\left(\frac{1+\phi^2}{1-\phi^2}+k\right)\right]\right\}- \\ & -\frac{1}{\gamma^2(0)}2\rho(k)\frac{1}{T}\left[4\gamma(0)\gamma(k)+2\sigma_h^4\left(\frac{2\phi^2}{1-\phi^2}+k-1\right)\right]+O(T^{-2}) \qquad [2.77]\end{aligned}$$

Considerando en esta expresión 1/(T-k)~1/T, y simplificando algunos términos, se obtiene la expresión [2.72] requerida. ■

A continuación vamos a demostrar un resultado que proporciona la distribución asintótica de r(k) en modelos ARSV(1) estacionarios.

Proposición 2.6. En el modelo ARSV(1) estacionario definido en las ecuaciones [2.22], la correlación muestral de orden k, r(k), de la serie $x_t=\ln(y_t^2)$ se distribuye asintóticamente como una variable Normal de media $\rho(k)$ y varianza [2.72], donde $\rho(k)$ denota la función de autocorrelación poblacional de x_t.

Demostración:

La distribución asintótica de r(k) se obtiene a partir de la distribución asintótica de las covarianzas muestrales {c(0),c(k)} de la siguiente forma.

Definamos los vectores $\mathbf{c}=\{c(0),c(k)\}'$ y $\boldsymbol{\gamma}=\{\gamma(0),\gamma(k)\}'$, donde c(0) y c(k) son la varianza y la covarianza muestral de orden k de la serie $x_t=\ln(y_t^2)$ y $\gamma(0)$ y $\gamma(k)$ son sus análogos poblacionales. Sea f una función real de $\Re^2$ en $\Re$ definida como $f(\mathbf{x})=f((x_1,x_2)')=x_2/x_1$. Entonces, la correlación muestral y la función de autocorrelación poblacional de orden k de la serie x_t pueden escribirse como $r(k)=f(\mathbf{c})=c(k)/c(0)$ y $\rho(k)=f(\boldsymbol{\gamma})=\gamma(k)/\gamma(0)$, respectivamente.

En la proposición 2.4 hemos demostrado que:

$$\sqrt{T}\,(\mathbf{c}-\boldsymbol{\gamma})=\sqrt{T}\,\{c(0)-\gamma(0),\, c(k)-\gamma(k)\}' \xrightarrow{d} N(0,V)$$

donde V es la matriz de varianzas y covarianzas definida en [2.53], cuyos elementos aparecen definidos en las expresiones [2.54]-[2.56]. Utilizando un resultado clásico sobre convergencia de funciones diferenciables de vectores asintóticamente normales (Serfling, 1980, p.122), se demuestra que, bajo ciertas condiciones de diferenciabilidad sobre la función f, se tiene que:

$$\sqrt{T}\,(f(\mathbf{c})-f(\boldsymbol{\gamma})) = \sqrt{T}\,\{r(k)-\rho(k)\} \xrightarrow{d} N(0,DVD') \qquad [2.78]$$

donde D es el vector de las derivadas parciales de f evaluadas en $\boldsymbol{\gamma}$:

$$D=\left[\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2}\right]_{x=\boldsymbol{\gamma}}$$

Para verificar las condiciones que permiten aplicar este resultado, es suficiente probar que cada una de las derivadas parciales de f es continua en $\boldsymbol{\gamma}$ y que

no todas se anulan en $\boldsymbol{\gamma}$. En nuestro caso, las derivadas parciales de f son de la forma $\frac{\partial f(x)}{\partial x_1} = -\frac{x_2}{(x_1)^2}, \frac{\partial f(x)}{\partial x_2} = \frac{1}{x_1}$, y se comprueba que ambas son continuas en $\boldsymbol{\gamma}$ y que ninguna de ellas se anula en $\boldsymbol{\gamma}$. De hecho, resulta que:

$$D=\left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}\right]_{x=\boldsymbol{\gamma}} = \left[-\frac{\gamma(k)}{\gamma^2(0)}, \frac{1}{\gamma(0)}\right] = \frac{1}{\gamma(0)}[-\rho(k),1] \qquad [2.79]$$

Por tanto, el resultado [2.78] se puede aplicar y se demuestra que:

$$\sqrt{T}\,(r(k)-\rho(k)) \xrightarrow{d} N(0,v)$$

donde:

$$v=DVD'=\frac{1}{\gamma^2(0)}[v_{00}\,\rho^2(k) - 2\,v_{0k}\,\rho(k) + v_{kk}\,] \qquad [2.80]$$

donde v_{00}, v_{0k} y v_{kk} aparecen definidos en las ecuaciones [2.54]-[2.56].

Para poder obtener una expresión de v en función de las correlaciones poblacionales de x_t, basta probar que las expresiones de v_{00}, v_{0k} y v_{kk} en [2.54]-[2.56] pueden escribirse en función de la varianza y la función de autocovarianza poblacional de x_t, de la siguiente forma:

$$v_{00} = \pi^4 + 2\sum_{u=-\infty}^{\infty}\gamma^2(u) \qquad [2.81]$$

$$v_{0k} = 2\sum_{u=-\infty}^{\infty}\gamma(u)\gamma(u+k)\text{, si } k>0 \qquad [2.82]$$

$$v_{kk} = \sum_{u=-\infty}^{\infty}\gamma^2(u) + \sum_{u=-\infty}^{\infty}\gamma(u)\gamma(u+2k)\text{, si } k>0 \qquad [2.83]$$

donde $\gamma(k)$ es la función de autocovarianza poblacional de x_t definida en [2.26]. Sustituyendo estas expresiones en [2.80] se obtiene la siguiente fórmula para la varianza asintótica de $\sqrt{T}\,(r(k)-\rho(k))$:

$$v = \sum_{u=-\infty}^{\infty}\left[2\rho^2(k)\rho^2(u) - 4\rho(k)\rho(u)\rho(u+k) + \rho^2(u) + \rho(u)\rho(u+2k)\right] + \pi^4\frac{\rho^2(k)}{\gamma^2(0)} \qquad [2.84]$$

donde $\rho(k)$ y $\gamma(0)$ son la función de autocorrelación poblacional y la varianza de la serie $x_t=\ln(y_t^2)$, respectivamente.

Compárese esta expresión con la fórmula habitual para procesos lineales que aparece, por ejemplo, en Anderson (1971, p.489), Brockwell y Davies(1991, p.221) o Fuller (1996, p.317). El último término de la expresión [2.84] es el término extra debido a la no normalidad del ruido ξ_t, que hace que haya un cumulante de cuarto orden de x_t distinto de cero, $\kappa_x(0,0,0)=E(\xi_t^4)-3[E(\xi_t^2)]^2=\pi^4$.

En la práctica, el valor asintótico de la varianza de r(k) se calcula a partir de la expresión [2.84] haciendo las siguientes simplificaciones en los sumatorios:

$$\text{Var}(r(k)) \sim \frac{1}{T}v =$$

$$= \frac{1}{T}\left\{1 + 2\rho^2(k) + 2\rho(2k) + \pi^4\frac{\rho^2(k)}{\gamma^2(0)} + 2[1+2\rho^2(k)]\sum_{u=1}^{\infty}\rho^2(u) - 4\rho(k)\sum_{u=0}^{k}\rho(u)\rho(u-k) - \right.$$

$$\left. - 8\rho(k)\sum_{u=1}^{\infty}\rho(u)\rho(u+k) + \sum_{u=1}^{2k-1}\rho(u)\rho(2k-u) + 2\sum_{u=1}^{\infty}\rho(u)\rho(u+2k)\right\}$$

[2.85]

Sustituyendo en esta expresión $\rho(k)$ por su valor para la serie $x_t=\ln(y_t^2)$ en la ecuación [1.23] se obtiene la expresión [2.72] requerida. ■

Obsérvese que la expresión [2.85] de la varianza de r(k) puede utilizarse para construir de forma recursiva las bandas de confianza para contrastar incorrelación en la serie $x_t=\ln(y_t^2)$. Partiendo de la hipótesis de incorrelación, $H_0:\rho(k)=0$, para todo k, se va contrastando la significación de cada una de las correlaciones, empezando por la primera (k=1) e incorporando sucesivamente aquellas que vayan resultando significativas.

Por ejemplo, bajo la hipótesis de incorrelación $H_0:\rho(k)=0$, para todo $k\geq 1$, la varianza asintótica de r(k), para todo $k\geq 1$, es 1/T, y se obtendrían las bandas de confianza habituales al 95% para las correlaciones muestrales de x_t, dadas por:

$$\pm 1.96/\sqrt{T}$$

Si la hipótesis de incorrelación se rechaza porque r(1) está fuera de estas bandas de confianza, la hipótesis nula se reformula como $H_0:\rho(k)=0$, para todo $k\geq 2$. En este caso, la varianza asintótica de r(k) se calcula con la expresión [2.85], que, bajo la hipótesis nula, toma el siguiente valor para r(1):

$$\frac{1}{T}\left\{1+2\rho^2(1)+\pi^4\frac{\rho^2(1)}{\gamma^2(0)}+2[1+2\rho^2(1)]\rho^2(1)-8\rho^2(1)+\rho^2(1)\right\}$$

A partir de este valor se obtendrían las siguientes bandas de confianza al 95%:

$$\pm\frac{1.96}{\sqrt{T}}\left\{1-3[r(1)]^2+\pi^4\frac{[r(1)]^2}{[c(0)]^2}+4[r(1)]^4\right\}$$

Si la hipótesis nula, $H_0:\rho(k)=0$ para todo $k\geq 2$, se rechaza porque r(2) se sale de estas bandas, se replantea dicha hipótesis como $H_0:\rho(k)=0$, para todo $k\geq 3$. Bajo la

nueva hipótesis nula, la varianza de r(2) viene dada nuevamente por la fórmula [2.85], a partir de la cual se obtienen las siguientes bandas de confianza al 95%:

$$\pm\frac{1.96}{\sqrt{T}}\left\{1-3[r(2)]^2+\pi^4\frac{[r(2)]^2}{[c(0)]^2}+2[r(1)]^2+4r(2)[r(1)]^2+4[r(2)]^2[r(1)]^2+4[r(2)]^4\right\}$$

Procediendo de esta forma, se podrían ir construyendo las bandas de confianza para todos los retardos y se podría contrastar así la hipótesis de incorrelación en x_t para cualquier orden. A partir de un cierto retardo, las correlaciones de la serie x_t serán prácticamente cero y estas bandas serán prácticamente constantes.

2.4.2. Propiedades en muestras finitas de las correlaciones muestrales de la serie de los cuadrados y de la serie de logaritmos de los cuadrados en el modelo ARSV(1)

En esta sección se recogen los resultados de un experimento de Monte Carlo para analizar la validez en muestras finitas de los resultados asintóticos de la sección anterior. El diseño del experimento y los modelos ARSV(1) simulados, son los mismos que en la sección 2.3.2. Para cada modelo y cada tamaño muestral se obtienen 5000 réplicas, y en cada réplica se calculan las correlaciones muestrales, hasta de orden 50, de la serie $x_t=\ln(y_t^2)$. El sesgo y la varianza muestral de las correlaciones sobre todas las réplicas se calcula utilizando las expresiones en [2.57], pero sustituyendo c(k) y $\gamma(k)$ por r(k) y $\rho(k)$, respectivamente.

Las tablas 2.16-2.18 recogen los valores del sesgo y de la desviación estándar muestral de r(k) sobre todas las réplicas, en distintos modelos ARSV(1), con ϕ=0.9, ϕ=0.95 y ϕ=0.98, respectivamente. Junto a los valores muestrales se presentan los correspondientes valores asintóticos del sesgo y la varianza de r(k), obtenidos a partir

de las fórmulas [2.71] y [2.72], respectivamente. Además, se incluyen los valores de la función de autocovarianza poblacional, $\rho(k)$, de la serie $x_t=\ln(y_t^2)$ obtenidos a partir de la expresión [1.23] del capítulo 1.

Los resultados numéricos se complementan con una serie de gráficos que facilitan su interpretación. Los gráficos 2.18-2.20 representan, para cada uno de los tres modelos ARSV(1) analizados, la función de autocorrelación teórica, $\rho(k)$, para $k=1,2,\ldots,50$, junto con los correspondientes valores del correlograma medio sobre las 5000 réplicas, $\bar{r}(1),\bar{r}(2),...,\bar{r}(50)$. Estos últimos valores se calculan como $\bar{r}(k)=\frac{1}{5000}\sum_{i=1}^{5000} r^{(i)}(k)$, donde $r^{(i)}(k)$ es la correlación muestral de orden k en la i-ésima réplica. En cada uno de estos gráficos, cada fila corresponde a un tamaño muestral, y dentro de cada fila, los tres correlogramas representados corresponden, de izquierda a derecha, a modelos con $\sigma_\eta^2=0.5$, $\sigma_\eta^2=0.1$ y $\sigma_\eta^2=0.01$, respectivamente.

Tabla 2.16. Sesgo y desviación estándar de r(k) para la serie ln(y_t^2) en el modelo ARSV(1) con ϕ=0.9 y diferentes valores de σ_η^2, junto con las autocorrelaciones teóricas ρ(k) de dicha serie. [(a)]

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	ρ(k)	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
σ_η^2=0.5	1	0.313	-0.013 (0.065)	-0.014 (0.064)	-0.006 (0.046)	-0.007 (0.045)	-0.002 (0.023)	-0.003 (0.023)
	10	0.121	-0.017 (0.065)	-0.018 (0.062)	-0.009 (0.046)	-0.009 (0.045)	-0.002 (0.023)	-0.002 (0.023)
	20	0.042	-0.017 (0.063)	-0.017 (0.060)	-0.008 (0.045)	-0.008 (0.043)	-0.002 (0.022)	-0.002 (0.022)
	50	0.002	-0.014 (0.063)	-0.013 (0.056)	-0.007 (0.045)	-0.007 (0.043)	-0.002 (0.022)	-0.002 (0.022)
σ_η^2=0.1	1	0.087	-0.005 (0.050)	-0.006 (0.049)	-0.003 (0.035)	-0.002 (0.035)	-0.001 (0.018)	-0.001 (0.017)
	10	0.034	-0.006 (0.047)	-0.007 (0.046)	-0.003 (0.033)	-0.003 (0.033)	-0.001 (0.017)	-0.001 (0.017)
	20	0.012	-0.006 (0.046)	-0.006 (0.045)	-0.003 (0.033)	-0.003 (0.032)	-0.001 (0.016)	-0.001 (0.016)
	50	0.001	-0.005 (0.046)	-0.005 (0.043)	-0.003 (0.032)	-0.003 (0.032)	-0.001 (0.016)	-0.001 (0.016)
σ_η^2=0.01	1	0.010	-0.002 (0.045)	-0.002 (0.044)	-0.001 (0.032)	-0.001 (0.031)	-0.000 (0.016)	-0.000 (0.015)
	10	0.004	-0.002 (0.044)	-0.003 (0.043)	-0.001 (0.031)	-0.001 (0.031)	-0.000 (0.016)	-0.000 (0.016)
	20	0.001	-0.002 (0.044)	-0.003 (0.043)	-0.001 (0.031)	-0.002 (0.031)	-0.000 (0.016)	-0.000 (0.016)
	50	0.000	-0.002 (0.044)	-0.003 (0.042)	-0.001 (0.031)	-0.001 (0.031)	-0.000 (0.016)	-0.000 (0.016)

[(a)] Para cada retardo y cada tamaño muestral se presentan el sesgo y la desviación estándar (entre paréntesis) asintóticos calculados a partir de las fórmulas [2.71] y [2.72], respectivamente, junto con los correspondientes valores simulados de Monte Carlo: el sesgo medio y su desviación estándar muestral (entre paréntesis) sobre todas las réplicas. El número de réplicas es 5000.

Tabla 2.17. Sesgo y desviación estándar de r(k) para la serie $\ln(y_t^2)$ en el modelo ARSV(1) con ϕ=0.95 y diferentes valores de σ_η^2, junto con las autocorrelaciones teóricas ρ(k) de dicha serie. [(b)]

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	ρ(k)	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
σ_η^2=0.5	1	0.484	-0.031 (0.083)	-0.032 (0.082)	-0.016 (0.059)	-0.016 (0.057)	-0.004 (0.029)	-0.005 (0.029)
	10	0.305	-0.045 (0.098)	-0.046 (0.093)	-0.023 (0.069)	-0.022 (0.066)	-0.006 (0.034)	-0.006 (0.034)
	20	0.183	-0.050 (0.103)	-0.049 (0.094)	-0.025 (0.073)	-0.024 (0.068)	-0.006 (0.037)	-0.007 (0.036)
	50	0.039	-0.047 (0.106)	-0.041 (0.085)	-0.023 (0.075)	-0.023 (0.068)	-0.006 (0.038)	-0.006 (0.036)
σ_η^2=0.1	1	0.164	-0.014 (0.062)	-0.015 (0.060)	-0.007 (0.044)	-0.007 (0.043)	-0.002 (0.022)	-0.002 (0.022)
	10	0.103	-0.017 (0.060)	-0.018 (0.058)	-0.008 (0.042)	-0.008 (0.041)	-0.002 (0.021)	-0.002 (0.021)
	20	0.062	-0.018 (0.058)	-0.018 (0.054)	-0.009 (0.041)	-0.009 (0.039)	-0.002 (0.020)	-0.003 (0.020)
	50	0.013	-0.016 (0.055)	-0.015 (0.048)	-0.008 (0.039)	-0.008 (0.037)	-0.002 (0.019)	-0.002 (0.019)
σ_η^2=0.01	1	0.019	-0.003 (0.045)	-0.004 (0.045)	-0.002 (0.032)	-0.001 (0.032)	-0.000 (0.016)	-0.000 (0.016)
	10	0.012	-0.004 (0.045)	-0.004 (0.044)	-0.002 (0.032)	-0.002 (0.032)	-0.001 (0.016)	-0.000 (0.016)
	20	0.007	-0.004 (0.045)	-0.004 (0.044)	-0.002 (0.032)	-0.002 (0.031)	-0.001 (0.016)	-0.001 (0.016)
	50	0.002	-0.004 (0.044)	-0.004 (0.042)	-0.002 (0.031)	-0.002 (0.031)	-0.001 (0.016)	-0.000 (0.016)

[(b)] La notación de esta tabla es la misma que en la tabla 2.16.

Tabla 2.18. Sesgo y desviación estándar de r(k) para la serie $\ln(y_t^2)$ en el modelo ARSV(1) con ϕ=0.98 y diferentes valores de σ_η^2, junto con las autocorrelaciones teóricas ρ(k) de dicha serie. (c)

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	ρ(k)	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
σ_η^2=0.5	1	0.705	-0.072 (0.097)	-0.071 (0.100)	-0.036 (0.069)	-0.035 (0.068)	-0.009 (0.034)	-0.010 (0.034)
	10	0.588	-0.108 (0.130)	-0.107 (0.127)	-0.054 (0.092)	-0.053 (0.090)	-0.014 (0.046)	-0.014 (0.046)
	20	0.480	-0.137 (0.156)	-0.132 (0.142)	-0.068 (0.110)	-0.067 (0.104)	-0.017 (0.055)	-0.018 (0.055)
	50	0.262	-0.176 (0.199)	-0.155 (0.145)	-0.088 (0.141)	-0.084 (0.123)	-0.022 (0.070)	-0.022 (0.068)
σ_η^2=0.1	1	0.332	-0.060 (0.110)	-0.055 (0.095)	-0.030 (0.078)	-0.028 (0.071)	-0.007 (0.039)	-0.008 (0.038)
	10	0.277	-0.069 (0.115)	-0.064 (0.097)	-0.034 (0.081)	-0.033 (0.073)	-0.009 (0.041)	-0.009 (0.040)
	20	0.226	-0.076 (0.117)	-0.070 (0.094)	-0.038 (0.083)	-0.036 (0.073)	-0.009 (0.041)	-0.010 (0.040)
	50	0.123	-0.084 (0.117)	-0.073 (0.080)	-0.042 (0.083)	-0.040 (0.069)	-0.011 (0.041)	-0.011 (0.039)
σ_η^2=0.01	1	0.048	-0.011 (0.050)	-0.011 (0.049)	-0.006 (0.036)	-0.005 (0.035)	-0.001 (0.018)	-0.001 (0.017)
	10	0.040	-0.012 (0.050)	-0.012 (0.047)	-0.006 (0.035)	-0.006 (0.034)	-0.001 (0.018)	-0.001 (0.017)
	20	0.033	-0.013 (0.049)	-0.012 (0.046)	-0.006 (0.035)	-0.006 (0.034)	-0.002 (0.017)	-0.002 (0.017)
	50	0.018	-0.013 (0.048)	-0.013 (0.043)	-0.007 (0.034)	-0.006 (0.033)	-0.002 (0.017)	-0.002 (0.017)

(c) La notación de esta tabla es la misma que en la tabla 2.16.

Gráfico 2.18. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie $\ln(y_t^2)$ en un modelo ARSV(1) con ϕ=0.9 y distintos valores de σ_η^2.

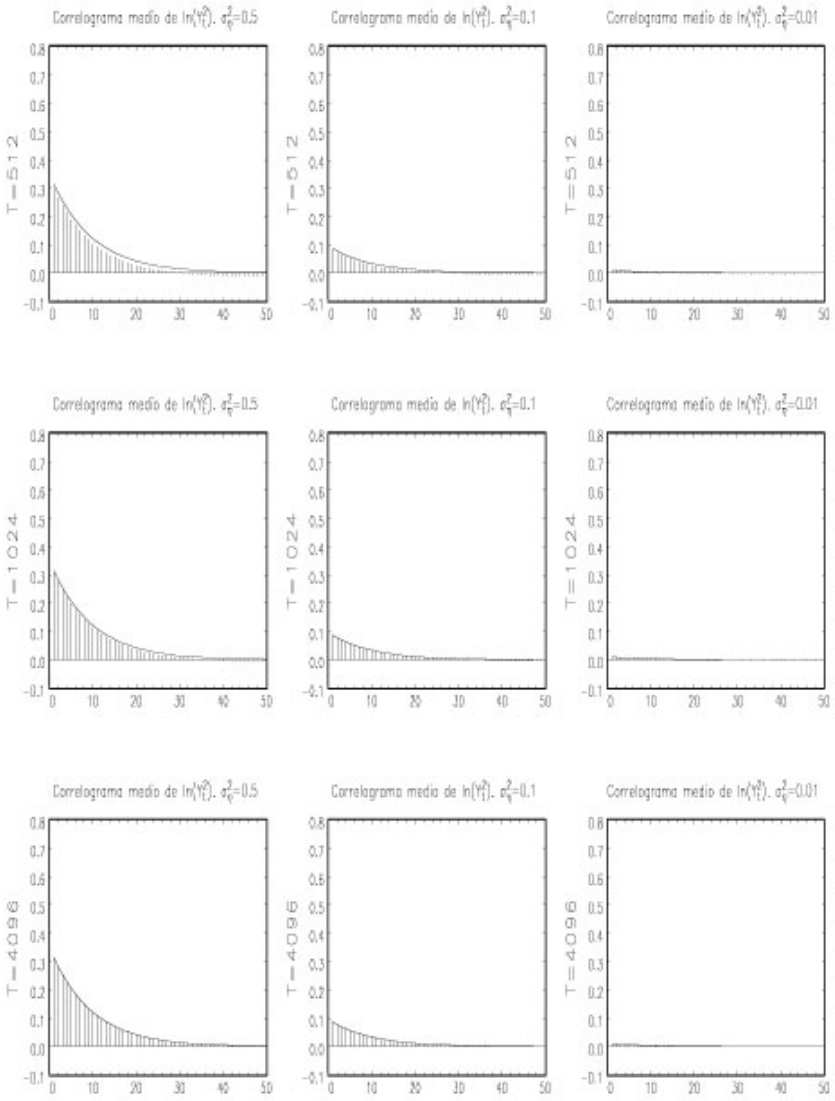


Gráfico 2.19. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie $\ln(y_t^2)$ en un modelo ARSV(1) con ϕ=0.95 y distintos valores de σ_η^2.

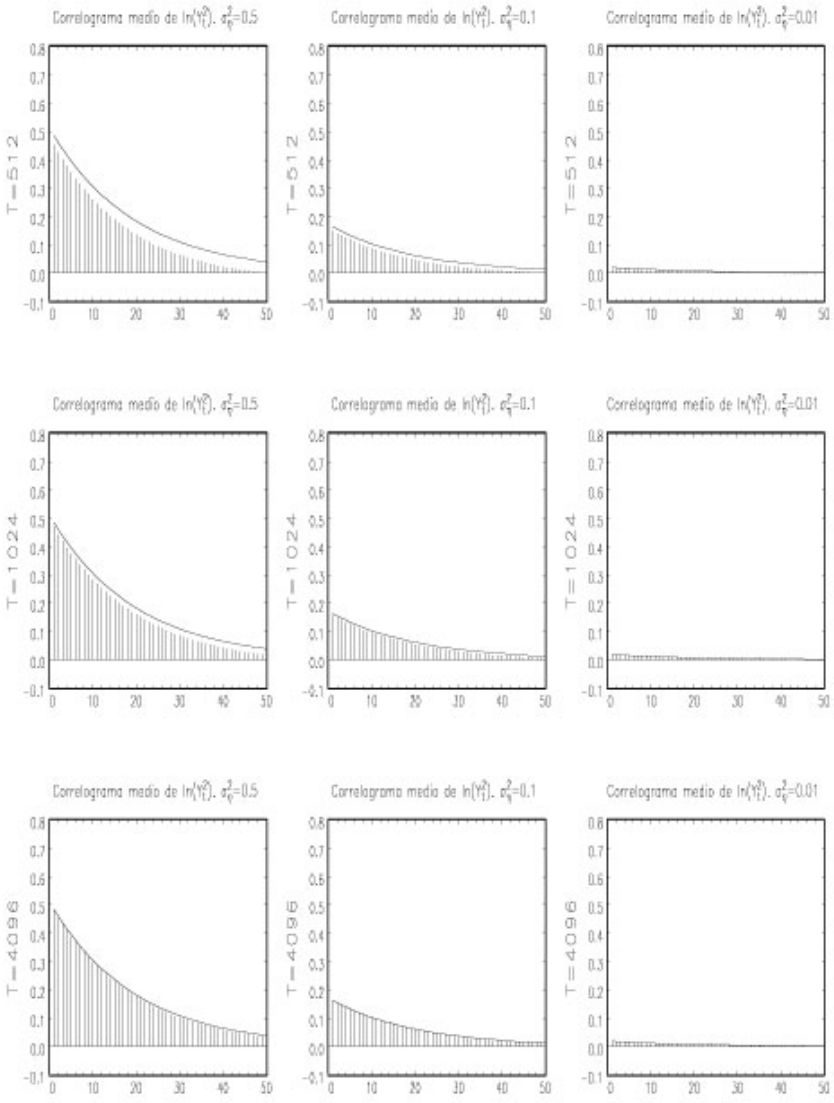
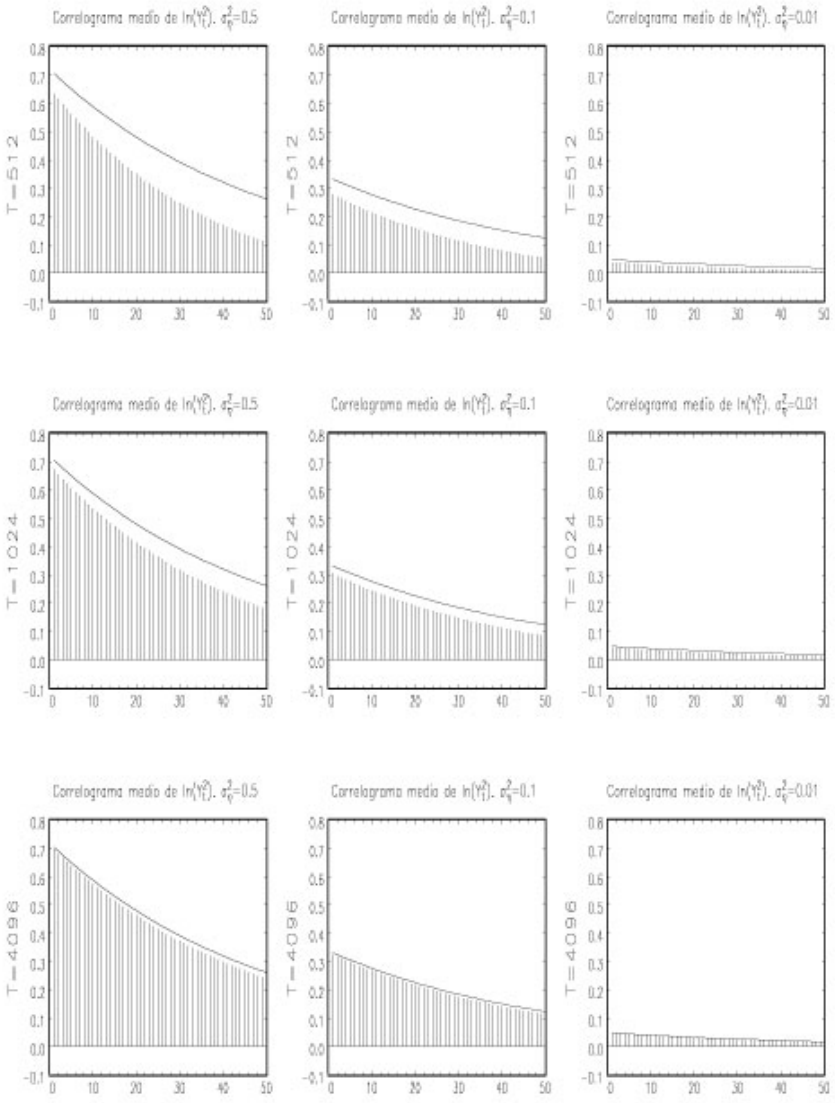


Gráfico 2.20. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie $\ln(y_t^2)$ en un modelo ARSV(1) con ϕ=0.98 y distintos valores de σ_η^2.



Las conclusiones que se derivan del análisis de las correlaciones muestrales de la serie $x_t=\ln(y_t^2)$ son similares a las ya comentadas para las covarianzas muestrales. Nuevamente, lo más destacable es que el sesgo es siempre negativo (el correlograma siempre subestima la función de autocorrelación poblacional) y que la magnitud del sesgo y la desviación estándar decrecen al disminuir el valor de σ_η^2 y al aumentar T. De hecho, se comprueba que la tasa de convergencia de la desviación estándar muestral de las correlaciones muestrales es $\sqrt{T}$.

Estas características se aprecian claramente en el gráfico 2.20, correspondiente al modelo más persistente (ϕ=0.98), donde observamos que para T=4096 y/o σ_η^2=0.01, la función de autocorrelación teórica se ajusta perfectamente al correlograma medio de x_t, a diferencia de lo que ocurre cuando T=512, donde el sesgo de r(k) es considerable para los valores mayores de σ_η^2.

Además de estas características, cabría destacar como aspectos específicos del comportamiento de las correlaciones muestrales en modelos ARSV(1) los siguientes:

- En general, el sesgo en las autocorrelaciones muestrales de la serie $x_t=\ln(y_t^2)$ es despreciable, máxime cuando σ_η^2 es muy pequeño. De hecho, cuando σ_η^2=0.01, la propia serie x_t es indistinguible de un ruido blanco homocedástico y la desviación estándar de r(k) en estos casos coincide prácticamente con el valor habitual $1/\sqrt{T}$. En los gráficos 2.18-2.20 se ve claramente que la función de autocorrelación de x_t en los modelos con σ_η^2=0.01 (última columna) es prácticamente cero en todos los retardos. De hecho, se comprueba que sus valores están siempre dentro de las bandas de confianza habituales, $\pm 1.96/\sqrt{T}$, por lo que en ninguno de estos casos rechazaríamos la hipótesis de ruido blanco.
- Sin embargo, cuando ϕ=0.98 (gráfico 2.20), en el modelo con mayor volatilidad (σ_η^2=0.5) y en las muestras de tamaño T=512, aparecen sesgos importantes en las

correlaciones de $x_t=\ln(y_t^2)$. Pero como ya comentamos anteriormente, valores tan grandes de σ_η^2 no son habituales en aplicaciones con series financieras reales.

- En general, los valores del sesgo asintótico y del sesgo empírico no difieren mucho. Incluso cuando T=512, las diferencias son muy pequeñas, y podemos concluir que la aproximación asintótica del sesgo en [2.71] funciona bastante bien.
- Igualmente, las diferencias entre los valores de las desviaciones estándar asintótica y muestral son, en general, despreciables, excepto en alguno de los modelos más persistentes (ϕ=0.98) y en muestras no muy grandes (T=512). Lo que ocurre en estos casos es que los modelos están rozando la no estacionariedad y, por tanto, violan prácticamente una de las hipótesis en las que se sustenta la proposición 2.5, que es precisamente la estacionariedad de la serie x_t.
- En todos los casos resulta evidente que el valor habitual $1/\sqrt{T}$ subestima la desviación típica de las correlaciones muestrales de x_t, excepto en aquellos modelos en los que la varianza de la volatilidad es tan pequeña (σ_η^2=0.01) que son prácticamente ruido blanco. Esto significa que las bandas de Barlett son muy estrechas y podrían detectar correlaciones espurias en la serie x_t. En estos casos sería más apropiado construir las bandas de confianza corregidas, según se indicó en la sección anterior.

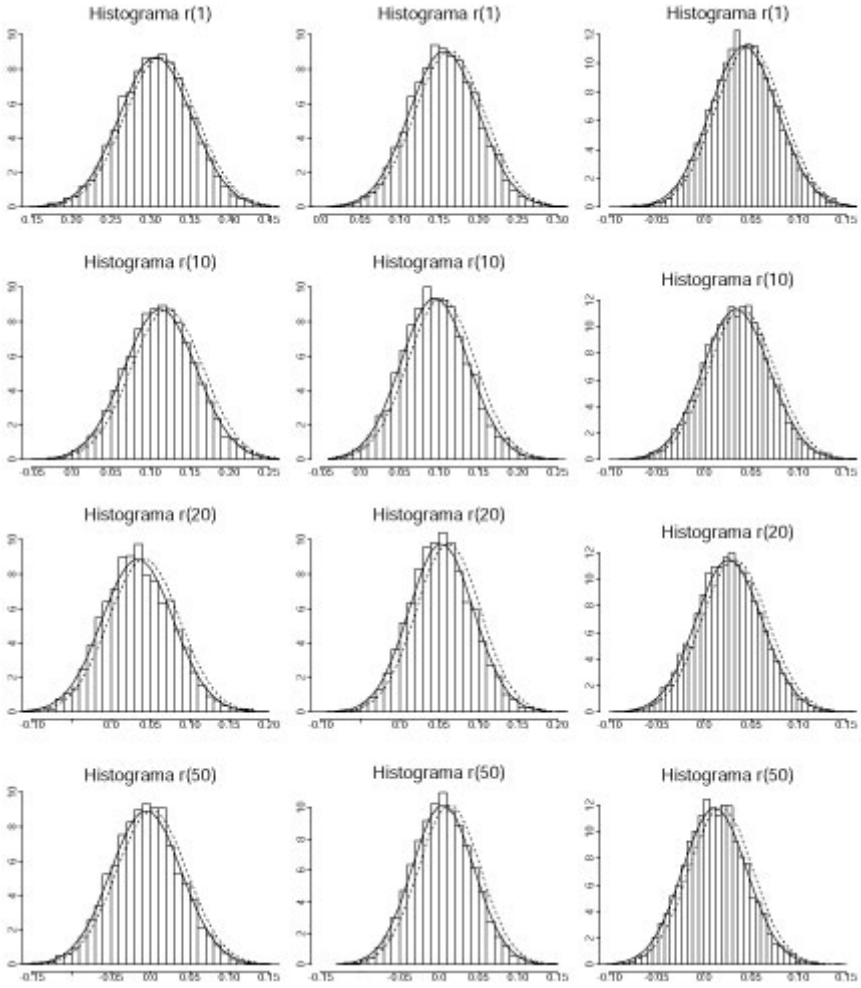
Si comparamos los resultados de las tres tablas, observamos que el sesgo y la varianza de r(k) aumentan al aumentar el valor de ϕ. Por ejemplo, en el modelo con σ_η^2=0.5 y ϕ=0.98, la desviación estándar de r(k) es prácticamente el doble que en el mismo modelo con ϕ=0.9, y lo mismo ocurre con el sesgo, que es siete veces mayor. Esto significa que cuanto más persistente es el modelo (más cerca está de la raíz unitaria) peor se estima la correlación y menos "fiable" es el correlograma, especialmente en muestras pequeñas. Este resultado se aprecia claramente comparando los correlogramas del gráfico 2.18 (ϕ=0.9) con los del gráfico 2.20 (ϕ=0.98). Mención aparte merecen los modelos con σ_η^2=0.01, en los que apenas hay diferencia

entre los valores del sesgo y la desviación típica de r(k) para los distintos valores de ϕ. Pero como ya hemos comentado anteriormente, esto se debe a que en estos modelos la propia serie x_t es casi ruido blanco, en cuyo caso la desviación típica de r(k) es $1/\sqrt{T}$ y el sesgo es $E(r(k)) = -\frac{T-k}{T(T-1)} + O(T^{-2})$; ver Fuller (1996, p. 318).

Para completar el análisis de las correlaciones muestrales de la serie $x_t=\ln(y_t^2)$, el gráfico 2.21 muestra la distribución empírica de r(k) sobre las 5000 réplicas, con k=1, 10, 20, 50, para tres modelos ARSV(1) concretos con parámetros $\{\phi=0.9, \sigma_\eta^2=0.5\}$, $\{\phi=0.95, \sigma_\eta^2=0.1\}$ y $\{\phi=0.98, \sigma_\eta^2=0.01\}$, y para muestras de tamaño T=1024. Para cada modelo se representa la distribución empírica de r(k) sobre las 5000 réplicas. Sobre esta distribución se representa, con línea discontinua, la densidad de una variable $N(\rho(k),Var(r(k)))$, donde $\rho(k)$ es la correlación teórica de orden k de la serie x_t en [1.23] y Var(r(k)) es la varianza asintótica de r(k) en [2.72]. Además, con línea continua se representa la densidad de una variable normal $N(\rho(k)+sesgo(r(k)),Var(r(k)))$, donde el sesgo de r(k) se obtiene utilizando la expresión [2.71]. Este gráfico sugiere que la distribución asintótica corregida por el sesgo es una buena aproximación a la distribución empírica de las correlaciones muestrales de la serie $\ln(y_t^2)$ en modelos ARSV(1).

Gráfico 2.21. Distribución empírica de las correlaciones muestrales de la serie $x_t=\ln(y_t^2)$, para k=1,10,20,50, junto con la distribución asintótica (línea punteada) y la distribución asintótica corregida por sesgo (línea continua), para tres modelos ARSV(1) con T=1024 y 5000 réplicas.

(a) $\{\phi=0.9, \sigma_\eta^2=0.5\}$ **(b)** $\{\phi=0.95, \sigma_\eta^2=0.1\}$ **(c)** $\{\phi=0.98, \sigma_\eta^2=0.01\}$



Para terminar esta sección, comentaremos los resultados de las correlaciones de la serie y_t^2. Para esta serie aún no tenemos resultados asintóticos y, por tanto, sólo presentamos los resultados de los experimentos de Monte Carlo. La tabla 2.19 recoge el sesgo y la desviación típica muestral de las correlaciones muestrales de y_t^2 sobre las 5000 réplicas, para los distintos modelos ARSV(1) analizados. Los gráficos 2.22-2.24 representan, para cada uno de estos modelos, la función de autocorrelación teórica, $\rho(k)$, para k=1,2,…,50, junto con los correspondientes valores del correlograma medio sobre las 5000 réplicas, $\bar{r}(1),\bar{r}(2),...,\bar{r}(50)$. Cada fila de estos gráficos representa un tamaño muestral, y dentro de cada fila, los correlogramas corresponden a modelos con σ_η^2=0.5, σ_η^2=0.1 y σ_η^2=0.01, respectivamente. La función de autocorrelación de y_t^2 se calcula con la fórmula [1.19] del capítulo 1.

Los resultados obtenidos para la serie y_t^2 difieren en algunos aspectos de los obtenidos para la serie $\ln(y_t^2)$, especialmente en los modelos próximos a la raíz unitaria. En la tabla 2.19 observamos, en primer lugar, que el comportamiento del sesgo depende del valor de σ_η^2. En general, el sesgo es positivo cuando σ_η^2=0.5 y pasa a ser negativo para valores más pequeños de σ_η^2. Una vez más, el sesgo y la varianza aumentan cuanto mayor es ϕ y mayor es σ_η^2. En el modelo más persistente ($\phi\approx1$) y con mayor volatilidad (σ_η^2=0.5), el sesgo no es nada despreciable y, lo que es peor, parece que no disminuye al aumentar el tamaño muestral, como tampoco parece que disminuya la varianza. Este comportamiento se aprecia claramente en los gráficos 2.22-2.24, donde puede verse que la función de autocorrelación teórica en los modelos con σ_η^2=0.5 (columna izquierda), subestima siempre el correlograma medio. Además, cuando ϕ=0.98 y σ_η^2=0.5, el correlograma parece sugerir que hay memoria larga en y_t^2 cuando realmente no la hay. En todos los casos analizados, la varianza de las correlaciones muestrales es mayor que 1/T y, por tanto, las bandas de Barlett serían demasiados estrechas y podrían detectar correlaciones espurias en la serie y_t^2.

Tabla 2.19. Sesgo y desviación estándar empíricos de r(k) para la serie y_t^2 en diversos modelos ARSV(1).

Tamaño muestral		**T=512**			**T=1024**			**T=4096**		
Parámetros	k	ϕ=0.90	ϕ=0.95	ϕ=0.98	ϕ=0.90	ϕ=0.95	ϕ=0.98	ϕ=0.90	ϕ=0.95	ϕ=0.98
σ_η^2=0.5	1	0.000 (0.126)	0.015 (0.142)	0.032 (0.153)	0.006 (0.112)	0.022 (0.132)	0.039 (0.146)	0.009 (0.091)	0.026 (0.120)	0.044 (0.138)
	10	0.005 (0.060)	0.031 (0.075)	0.062 (0.088)	0.010 (0.051)	0.039 (0.072)	0.071 (0.085)	0.010 (0.035)	0.038 (0.058)	0.073 (0.079)
	20	-0.003 (0.044)	0.012 (0.053)	0.033 (0.061)	-0.001 (0.035)	0.018 (0.046)	0.041 (0.058)	0.003 (0.022)	0.018 (0.034)	0.044 (0.052)
	50	-0.007 (0.037)	-0.008 (0.035)	-0.002 (0.035)	-0.004 (0.028)	-0.003 (0.029)	0.006 (0.035)	-0.000 (0.016)	0.002 (0.017)	0.010 (0.027)
σ_η^2=0.1	1	-0.014 (0.079)	-0.032 (0.097)	-0.053 (0.112)	-0.008 (0.061)	-0.020 (0.080)	-0.035 (0.099)	-0.003 (0.036)	-0.007 (0.056)	-0.013 (0.082)
	10	-0.011 (0.054)	-0.027 (0.070)	-0.045 (0.087)	-0.006 (0.042)	-0.014 (0.060)	-0.025 (0.079)	-0.002 (0.023)	-0.004 (0.039)	-0.004 (0.063)
	20	-0.008 (0.046)	-0.022 (0.057)	-0.037 (0.070)	-0.004 (0.035)	-0.011 (0.047)	-0.017 (0.064)	-0.001 (0.019)	-0.003 (0.030)	0.000 (0.052)
	50	-0.006 (0.042)	-0.015 (0.047)	-0.029 (0.049)	-0.004 (0.031)	-0.009 (0.035)	-0.014 (0.047)	-0.000 (0.017)	-0.002 (0.021)	0.002 (0.036)
σ_η^2=0.01	1	-0.003 (0.048)	-0.006 (0.053)	-0.024 (0.064)	-0.002 (0.035)	-0.003 (0.039)	-0.014 (0.050)	0.001 (0.020)	-0.000 (0.019)	-0.004 (0.028)
	10	-0.004 (0.044)	-0.007 (0.048)	-0.025 (0.065)	-0.002 (0.032)	-0.004 (0.036)	-0.014 (0.047)	-0.001 (0.018)	-0.002 (0.018)	-0.004 (0.027)
	20	-0.003 (0.044)	-0.006 (0.046)	-0.025 (0.055)	-0.001 (0.031)	-0.003 (0.033)	-0.014 (0.043)	-0.001 (0.017)	-0.000 (0.016)	-0.004 (0.025)
	50	-0.003 (0.042)	-0.006 (0.042)	-0.023 (0.046)	-0.001 (0.030)	-0.003 (0.031)	-0.014 (0.036)	0.000 (0.016)	-0.000 (0.016)	-0.004 (0.021)

Gráfico 2.22. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie y_t^2 en un modelo ARSV(1) con ϕ=0.9 y distintos valores de σ_η^2.

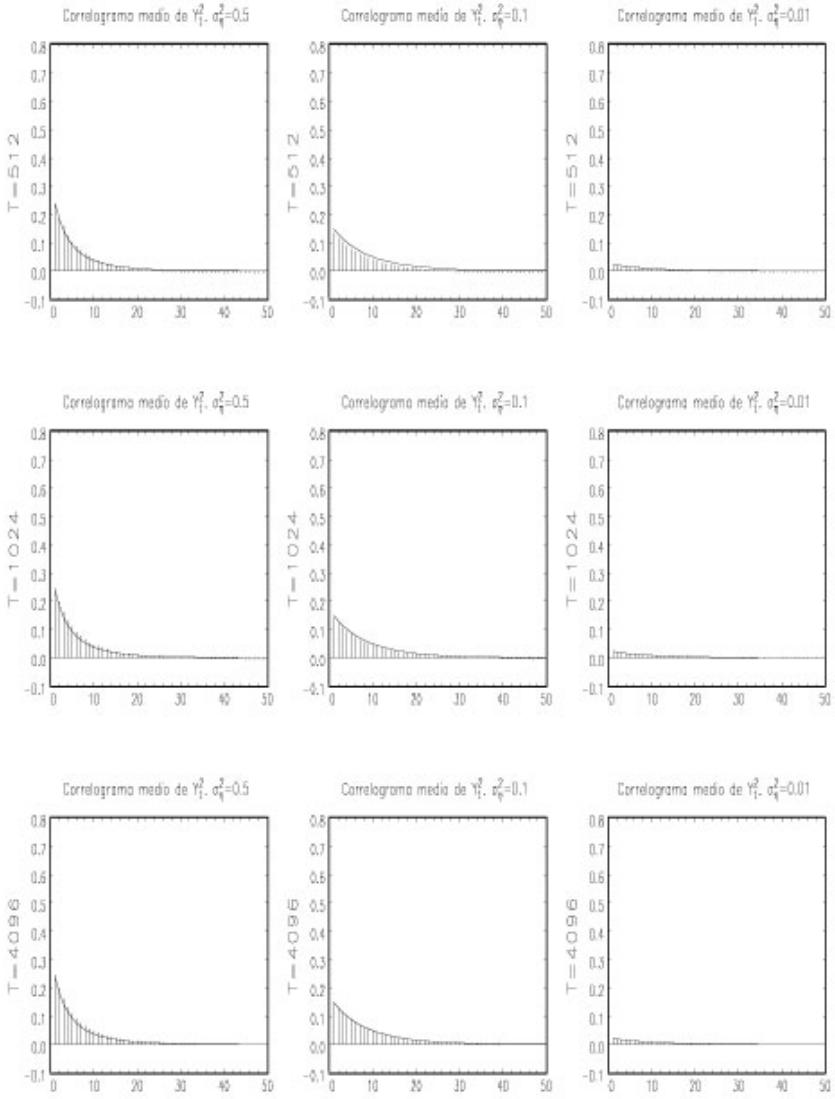


Gráfico 2.23. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie y_t^2 en un modelo ARSV(1) con ϕ=0.95 y distintos valores de σ_η^2.

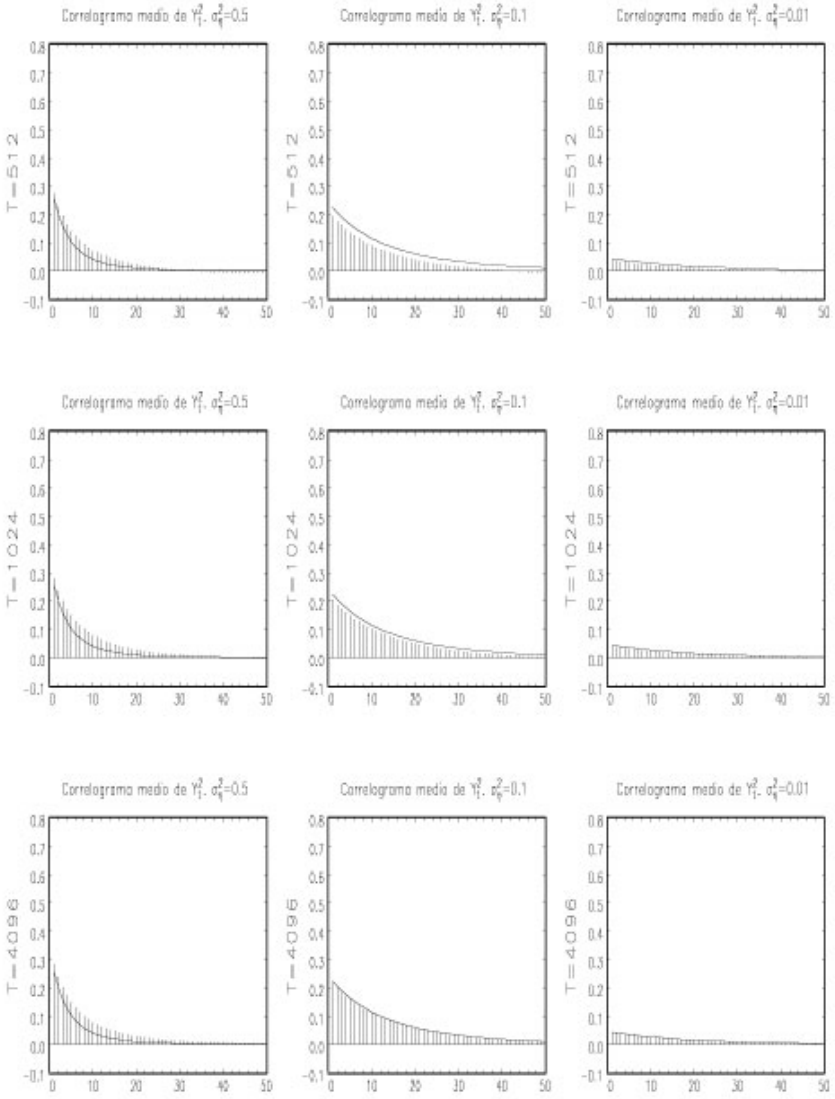
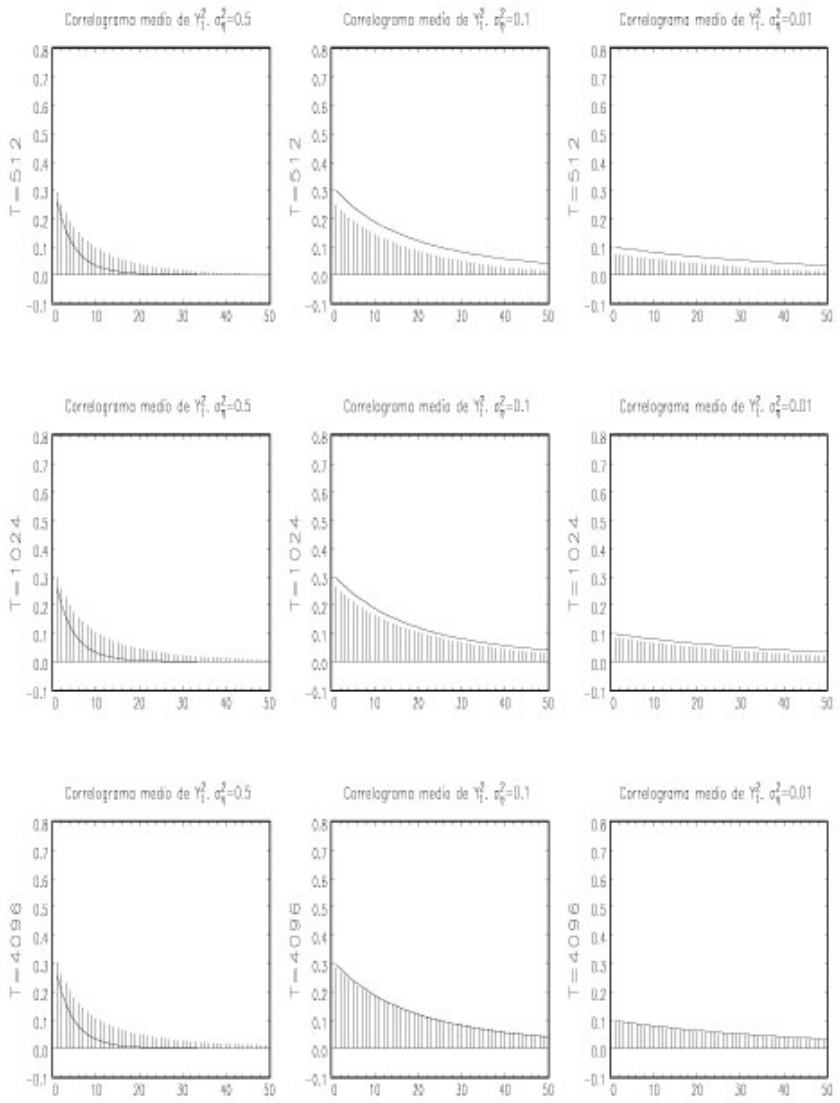


Gráfico 2.24. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie y_t^2 en un modelo ARSV(1) con ϕ=0.98 y distintos valores de σ_η^2.



2.4.3.

Sesgo y varianza de las autocorrelaciones muestrales de la serie de logaritmos de los cuadrados en el modelo LMSV

Los únicos resultados teóricos que nosotras conocemos sobre propiedades asintóticas de las correlaciones muestrales en procesos con memoria larga son los que se demuestran en Hosking (1996). Sin embargo, la extensión de estos resultados al modelo LMSV no parece sencilla, ya que en el modelo LMSV, la serie $x_t=\ln(y_t^2)$ es un proceso ARFIMA "contaminado" por un ruido no gaussiano, y no se verifican algunos de los supuestos en los que se sustentan dichos resultados.

Alternativamente, los resultados de Wright (1999) sobre las covarianzas muestrales en modelos LMSV, pueden utilizarse para derivar la distribución asintótica de las correlaciones muestrales de la serie x_t en modelos LMSV con d<1/4. Este argumento es el que utilizamos en la siguiente proposición.

Proposición 2.7. En el modelo LMSV estacionario con d<1/4, la correlación muestral r(k) de la serie $x_t=\ln(y_t^2)$ se distribuye asintóticamente como una variable Normal de media $\rho(k)$ y varianza:

$$\frac{1}{T}\left\{\sum_{u=-\infty}^{\infty}\left[2\rho^2(k)\rho^2(u)-4\rho(k)\rho(u)\rho(u+k)+\rho^2(u)+\rho(u)\rho(u+2k)\right]+\pi^4\frac{\rho^2(k)}{\gamma^2(0)}\right\} \quad [2.86]$$

donde $\gamma(0)$ y $\rho(k)$ son la varianza y la función de autocorrelación poblacional de x_t, $\gamma(0)=\sigma_h^2+\sigma_\xi^2$ y $\rho(k)=\frac{\Gamma(1-d)\Gamma(k+d)}{\Gamma(d)\Gamma(k+1-d)}\frac{\sigma_h^2}{\sigma_h^2+\sigma_\xi^2}$, respectivamente; σ_ξ^2 es la varianza de la perturbación ξ_t; y σ_h^2 es la varianza de h_t, $\sigma_h^2=\frac{\Gamma(1-2d)}{[\Gamma(1-d)]^2}\sigma_\eta^2$.

Demostración:

La demostración de esta proposición utiliza los mismos argumentos que

hemos utilizado para demostrar la proposición 2.6 de la sección anterior. Por esta razón, creemos que no es necesario repetir aquí todos los pasos de la demostración y sólo haremos algunas matizaciones.

Recordemos que la distribución asintótica de r(k) se puede obtener a partir de la distribución asintótica de las covarianzas muestrales utilizando el teorema sobre convergencia de funciones diferenciables de vectores asintóticamente normales de Serfling (1980, p.122).

Como vimos en la sección 2.4.1, este resultado se puede aplicar a las correlaciones muestrales y poblacionales porque r(k) y $\rho(k)$ son funciones de la forma $r(k)=f(\{c(0),c(k)\}')=c(k)/c(0)$ y $\rho(k)=f(\{\gamma(0),\gamma(k)\}')=\gamma(k)/\gamma(0)$, donde la función f verifica ciertas condiciones de diferenciabilidad.

La aplicación del resultado de Serfling (1980, p.122) exige además que las covarianzas muestrales tengan una distribución asintótica normal multivariante. Este requisito también se cumple en el modelo LMSV con d<1/4, porque, según vimos en la sección 2.3.3, la distribución asintótica del vector $\sqrt{T}\{c(0)-\gamma(0),c(k)-\gamma(k)\}'$ en el modelo con d<1/4, es normal bidimensional N(0,V), donde V es una matriz 2x2 cuyos elementos aparecen definidos en [2.54]-[2.56].

Aplicando entonces el resultado de Serfling (1980) a las covarianzas muestrales de la serie x_t en el modelo LMSV con d<1/4, resulta que:

$$\sqrt{T}\,(r(k)-\rho(k)) \xrightarrow{d} N(0,DVD')$$

donde D es el vector de las derivadas parciales de la función f definidas en [2.79]. Resolviendo el producto de matrices DVD', se obtiene la expresión de la varianza asintótica de $\sqrt{T}\,(r(k)-\rho(k))$ en [2.84], que depende de la función de autocorrelación poblacional de x_t y del cumulante de cuarto orden de x_t, $\kappa_x(0,0,0)=\pi^4$. Diviendo esta expresión entre T, se obtiene de forma inmediata la expresión [2.86] que permite obtener de forma aproximada la varianza de r(k) en modelos LMSV con d<1/4. ■

A diferencia de lo que ocurría en el modelo ARSV(1), en el modelo LMSV no es fácil obtener el valor exacto de la suma de las series que aparecen en la expresión [2.86] porque, en este caso, las series ya no son geométricas y esto complica las operaciones. Por ello, seguiremos la práctica habitual de truncar los sumatorios en un valor suficientemente grande a partir del cual los demás sumandos sean despreciables. Los resultados que se presentan en la sección siguiente corresponden a truncar los sumatorios en M=90000.

Al igual que en el modelo ARSV(1) estacionario, la expresión [2.86] puede utilizarse para construir de forma recursiva las bandas de confianza para contrastar incorrelación en la serie $x_t=\ln(y_t^2)$ en modelos LMSV con d<1/4. La forma de construir estas bandas se explicó en la sección 2.4.1.

Por lo que respecta al cálculo de la varianza asintótica de r(k) en modelos LMSV con 1/4≤d<1/2 y al cálculo del sesgo de r(k) para cualquier modelo LMSV con 0<d<1/2, aún no disponemos de ningún resultado que nos proporcione sus expresiones asintóticas.

Los únicos resultados que conocemos a este respecto y que, en principio, podrían tratar de aplicarse al modelo LMSV son los de Hosking (1996). Este autor deriva expresiones asintóticas del sesgo y la varianza de r(k) para series con memoria larga estacionarias que verifiquen la condición [2.60] y las condiciones [2.63]-[2.65]. Bajo estos supuestos, los resultados de Hosking (1996) aplicados al modelo LMSV proporcionarían los siguientes valores del sesgo y la varianza de r(k):

$$E(r(k))-\rho(k) \sim -\frac{[1-\rho(k)]}{d(1+2d)}\frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\frac{\sigma_\eta^2}{\gamma(0)}T^{2d-1}, \text{ si } 0<d<1/2 \qquad [2.87]$$

$$Var(r(k)) \sim 2K_2\left[\sigma_\eta^2\frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\right]^2\frac{[1-\rho(k)]^2}{\gamma^2(0)}T^{4d-2}, \text{ si } 1/4<d<1/2 \qquad [2.88a]$$

$$\mathrm{Var}(r(k)) \sim 4\left[\sigma_\eta^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}\right]^2 \frac{[1-\rho(k)]^2}{\gamma^2(0)} \frac{\ln(T)}{T}, \text{ si } d=1/4, \qquad [2.88b]$$

$$\mathrm{Var}(r(k)) \sim \frac{1}{T}\sum_{u=-\infty}^{\infty}[\rho^2(u)+\rho(u)\rho(u+2k)+2\rho^2(u)\rho^2(k)-4\rho(k)\rho(u)\rho(u+k)],$$

$$\text{si } 0<d<1/4, \qquad [2.88c]$$

donde $\gamma(0)$ y $\rho(k)$ son la varianza y la función de autocorrelación de x_t, y σ_h^2 es la varianza de h_t.

La validez de estas fórmulas pasa por comprobar que la serie $x_t=\ln(y_t^2)$ en el modelo LMSV cumple las condiciones [2.60], [2.63]-[2.65], pero según vimos en la sección 2.3.3, el modelo LMSV verifica sólo algunas de ellas. En concreto, parece difícil comprobar la condición [2.64], que requiere derivar la forma reducida del modelo estructural dado por las ecuaciones [2.58a] y [2.58b].

Por otra parte, en la sección 2.3.3 comprobamos que las fórmulas de Hosking (1996) para calcular la varianza de las covarianzas muestrales no eran directamente aplicables al modelo LMSV, precisamente porque éste no verificaba algunas de las condiciones requeridas. Por esta misma razón, tampoco podrían utilizarse directamente las fórmulas [2.87] y [2.88] para calcular el sesgo y la varianza de las correlaciones muestrales en el modelo LMSV, ya que éstas se derivan en base a las mismas condiciones. De hecho, si comparamos la expresión [2.88c] de la varianza de r(k) para modelos con $d<1/4$ con la que nosotras hemos obtenido en [2.86], observamos que difieren en el último término que recoge, precisamente, la existencia de un cumulante de cuarto orden de x_t distinto de cero, $\kappa_x(0,0,0)=\pi^4$. Por tanto, queda claro que la fórmula de Hosking (1996) no es directamente aplicable a nuestro modelo, porque no incluye ese término adicional.

A pesar de la dudosa validez de los resultados de Hosking (1996) para la serie x_t del modelo LMSV, vamos a evaluar, a modo de ejercicio, las expresiones del sesgo y la varianza de r(k) en [2.87] y [2.88], respectivamente, y en la sección siguiente vamos a comparar estos valores con los que se obtienen en muestras finitas. Como veremos más adelante, los resultados de las simulaciones confirman que estas fórmulas no *funcionan* para el modelo LMSV: las diferencias entre los valores teóricos y los valores simulados son considerables, incluso para muestras de tamaño T=4096.

Esto significa que el problema de obtener una expresión para el sesgo de las autocorrelaciones muestrales de la serie x_t en el modelo LMSV y para la varianza en modelos con $d \geq 1/4$, debería abordarse desde otra perspectiva, y no tratando de comprobar si este modelo verifica las condiciones que permitan aplicar los resultados de Hosking (1996). Éste es sin duda alguna uno de los temas en los que seguiremos trabajando en un futuro próximo.

2.4.4. Propiedades en muestras finitas de las correlaciones muestrales de la serie de los cuadrados y de la serie de logaritmos de los cuadrados en el modelo LMSV

En esta sección se presentan los resultados de las simulaciones para las correlaciones muestrales de la serie $x_t=\ln(y_t^2)$ y de la serie y_t^2 en distintos modelos LMSV. El diseño del experimento y los modelos analizados son los mismos que en la sección 2.3.4. En primer lugar, comentaremos los resultados correspondientes a la serie $x_t=\ln(y_t^2)$ y finalmente resumiremos los resultados para y_t^2.

Las tablas 2.20-2.23 presentan los valores del sesgo y la desviación estándar de la correlación muestral, r(k), de la serie $x_t=\ln(y_t^2)$, para $d=\{0.2, 0.4, 0.45, 0.49\}$, respectivamente, con k=1,10,20,50. Junto a estos valores se presentan los correspondientes valores del sesgo asintótico y la desviación estándar asintótica. En

todos los casos, el sesgo se calcula a partir de la fórmula [2.87]. Para el modelo LMSV con d=0.2, la desviación típica asintótica se obtiene como la raíz cuadrada de la expresión [2.86], y para el resto de los modelos se utiliza la fórmula [2.88a], aún sabiendo que teóricamente no es aplicable al modelo LMSV.

Los resultados numéricos se complementan con los siguientes gráficos. Los gráficos 2.25-2.27 representan, para d={0.4, 0.45, 0.49}, respectivamente, los correlogramas medios de la serie $\ln(y_t^2)$ sobre las 5000 réplicas junto con la función de autocorrelación teórica de dicha serie, $\rho(k)$, con k=1,2,…,50. En cada gráfico, cada fila corresponde a un tamaño muestral, y dentro de cada fila, los tres correlogramas representados corresponden, de izquierda a derecha, a $\sigma_\eta^2=0.5$, $\sigma_\eta^2=0.1$ y $\sigma_\eta^2=0.01$, respectivamente. La función de autocorrelación teórica de $\ln(y_t^2)$ se obtiene utilizando la fórmula [1.60] del capítulo 1.

Los gráficos correspondientes al modelo con d=0.2 no se incluyen porque para este modelo las correlaciones son prácticamente cero en todos los retardos, incluso en el modelo con $\sigma_\eta^2=1$, y no se aprecia ninguna diferencia.

Tabla 2.20. Sesgo y desviación estándar de r(k) para la serie ln(y_t^2) en el modelo LMSV con d=0.2 y diferentes valores de σ_η^2, junto con las autocovarianzas teóricas $\rho(k)$ de dicha serie. [(a)]

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	$\rho(k)$	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
σ_η^2=1	1	0.046	-0.004 (0.048)	-0.006 (0.045)	-0.002 (0.034)	-0.003 (0.032)	-0.001 (0.017)	-0.002 (0.016)
	10	0.012	-0.004 (0.047)	-0.007 (0.045)	-0.003 (0.033)	-0.003 (0.032)	-0.001 (0.017)	-0.002 (0.016)
	20	0.008	-0.004 (0.047)	-0.006 (0.044)	-0.003 (0.033)	-0.004 (0.031)	-0.001 (0.017)	-0.001 (0.016)
	50	0.004	-0.004 (0.047)	-0.006 (0.041)	-0.003 (0.033)	-0.003 (0.031)	-0.001 (0.017)	-0.001 (0.016)
σ_η^2=0.5	1	0.025	-0.002 (0.046)	-0.005 (0.044)	-0.001 (0.033)	-0.002 (0.031)	-0.001 (0.016)	-0.001 (0.016)
	10	0.006	-0.002 (0.046)	-0.005 (0.044)	-0.001 (0.032)	-0.002 (0.032)	-0.001 (0.016)	-0.001 (0.016)
	20	0.004	-0.002 (0.046)	-0.005 (0.044)	-0.001 (0.032)	-0.003 (0.031)	-0.001 (0.016)	-0.001 (0.016)
	50	0.002	-0.002 (0.046)	-0.004 (0.041)	-0.001 (0.032)	-0.002 (0.031)	-0.001 (0.016)	-0.001 (0.015)
σ_η^2=0.1	1	0.005	-0.001 (0.045)	-0.003 (0.043)	0.000 (0.032)	-0.001 (0.031)	0.000 (0.016)	-0.001 (0.015)
	10	0.001	-0.001 (0.045)	-0.003 (0.044)	0.000 (0.031)	-0.001 (0.032)	0.000 (0.016)	-0.001 (0.016)
	20	0.001	-0.001 (0.044)	-0.003 (0.044)	0.000 (0.031)	-0.001 (0.031)	0.000 (0.016)	-0.000 (0.016)
	50	0.001	-0.001 (0.044)	-0.003 (0.041)	0.000 (0.031)	-0.001 (0.031)	0.000 (0.016)	-0.000 (0.015)

[(a)] Para cada retardo y cada tamaño muestral se presentan el sesgo y la desviación estándar (entre paréntesis) asintóticos calculados a partir de las fórmulas [2.87] y [2.86], respectivamente, junto con los correspondientes valores simulados de Monte Carlo: el sesgo medio y su desviación estándar muestral (entre paréntesis) sobre todas las réplicas. El número de réplicas es 5000.

Tabla 2.21. Sesgo y desviación estándar de r(k) para la serie $\ln(y_t^2)$ en el modelo LMSV con d=0.4 y diferentes valores de σ_η^2, junto con las autocorrelaciones teóricas $\rho(k)$ de dicha serie. [(b)]

Tamaño muestral			**T=512**		**T=1024**		**T=4096**	
Parámetros	k	$\rho(k)$	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
$\sigma_\eta^2=1$	1	0.197	-0.064 (0.025)	-0.072 (0.053)	-0.055 (0.022)	-0.061 (0.040)	-0.042 (0.017)	-0.046 (0.024)
	10	0.125	-0.069 (0.027)	-0.080 (0.053)	-0.060 (0.024)	-0.067 (0.041)	-0.046 (0.018)	-0.049 (0.025)
	20	0.109	-0.071 (0.028)	-0.081 (0.051)	-0.061 (0.024)	-0.069 (0.039)	-0.047 (0.018)	-0.050 (0.025)
	50	0.091	-0.072 (0.028)	-0.082 (0.046)	-0.063 (0.025)	-0.070 (0.037)	-0.048 (0.019)	-0.051 (0.024)
$\sigma_\eta^2=0.5$	1	0.116	-0.041 (0.016)	-0.046 (0.048)	-0.036 (0.014)	-0.039 (0.036)	-0.027 (0.011)	-0.029 (0.020)
	10	0.074	-0.043 (0.017)	-0.049 (0.048)	-0.038 (0.015)	-0.040 (0.035)	-0.028 (0.011)	-0.030 (0.019)
	20	0.064	-0.044 (0.017)	-0.049 (0.047)	-0.038 (0.015)	-0.041 (0.034)	-0.029 (0.011)	-0.030 (0.019)
	50	0.054	-0.044 (0.017)	-0.049 (0.043)	-0.038 (0.015)	-0.042 (0.033)	-0.029 (0.011)	-0.030 (0.019)
$\sigma_\eta^2=0.1$	1	0.027	-0.011 (0.004)	-0.013 (0.044)	-0.009 (0.004)	-0.010 (0.032)	-0.007 (0.003)	-0.008 (0.016)
	10	0.017	-0.011 (0.004)	-0.014 (0.044)	-0.009 (0.004)	-0.010 (0.032)	-0.007 (0.003)	-0.018 (0.016)
	20	0.015	-0.011 (0.004)	-0.013 (0.044)	-0.009 (0.004)	-0.011 (0.031)	-0.007 (0.003)	-0.008 (0.016)
	50	0.012	-0.011 (0.004)	-0.013 (0.041)	-0.009 (0.004)	-0.010 (0.031)	-0.007 (0.003)	-0.007 (0.016)

[(b)] Para cada retardo y cada tamaño muestral se presentan el sesgo y la desviación estándar (entre paréntesis) asintóticos calculados a partir de las fórmulas [2.87] y [2.88a], respectivamente, junto con los correspondientes valores simulados de Monte Carlo: el sesgo medio y su desviación estándar muestral (entre paréntesis) sobre todas las réplicas. El número de réplicas es 5000.

Tabla 2.22. Sesgo y desviación estándar de r(k) para la serie $\ln(y_t^2)$ en el modelo LMSV con d=0.45 y diferentes valores de σ_η^2, junto con las autocorrelaciones teóricas $\rho(k)$ de dicha serie. (c)

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	$\rho(k)$	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
$\sigma_\eta^2=1$	1	0.348	-0.143 (0.024)	-0.187 (0.060)	-0.133 (0.026)	-0.170 (0.047)	-0.116 (0.020)	-0.143 (0.033)
	10	0.277	-0.158 (0.027)	-0.208 (0.060)	-0.148 (0.025)	-0.188 (0.049)	-0.128 (0.022)	-0.158 (0.035)
	20	0.259	-0.162 (0.028)	-0.213 (0.058)	-0.151 (0.026)	-0.194 (0.047)	-0.132 (0.022)	-0.162 (0.035)
	50	0.236	-0.167 (0.029)	-0.220 (0.051)	-0.156 (0.027)	-0.201 (0.045)	-0.136 (0.023)	-0.167 (0.035)
$\sigma_\eta^2=0.5$	1	0.221	-0.108 (0.019)	-0.129 (0.052)	-0.101 (0.017)	-0.118 (0.040)	-0.088 (0.015)	-0.100 (0.025)
	10	0.176	-0.114 (0.020)	-0.137 (0.051)	-0.107 (0.018)	-0.124 (0.039)	-0.093 (0.016)	-0.106 (0.025)
	20	0.164	-0.116 (0.020)	-0.139 (0.050)	-0.108 (0.019)	-0.127 (0.038)	-0.094 (0.016)	-0.107 (0.025)
	50	0.150	-0.118 (0.020)	-0.141 (0.045)	-0.110 (0.019)	-0.129 (0.037)	-0.096 (0.016)	-0.109 (0.025)
$\sigma_\eta^2=0.1$	1	0.056	-0.033 (0.006)	-0.037 (0.044)	-0.031 (0.005)	-0.033 (0.032)	-0.027 (0.005)	-0.029 (0.016)
	10	0.045	-0.034 (0.006)	-0.038 (0.044)	-0.032 (0.005)	-0.034 (0.032)	-0.028 (0.005)	-0.029 (0.016)
	20	0.042	-0.034 (0.006)	-0.038 (0.044)	-0.032 (0.005)	-0.034 (0.031)	-0.028 (0.005)	-0.029 (0.016)
	50	0.038	-0.034 (0.006)	-0.038 (0.041)	-0.032 (0.005)	-0.034 (0.031)	-0.028 (0.005)	-0.029 (0.016)

(c) La notación de esta tabla es la misma que en la tabla 2.21.

Tabla 2.23. Sesgo y desviación estándar de r(k) para la serie ln(y_t^2) en el modelo LMSV con d=0.49 y diferentes valores de σ_η^2, junto con las autocorrelaciones teóricas $\rho(k)$ de dicha serie. (d)

Tamaño muestral			T=512		T=1024		T=4096	
Parámetros	k	$\rho(k)$	Asintótico	MonteCarlo	Asintótico	MonteCarlo	Asintótico	MonteCarlo
σ_η^2=1	1	0.738	-0.176 (0.006)	-0.542 (0.067)	-0.174 (0.005)	-0.519 (0.055)	-0.169 (0.005)	-0.479 (0.043)
	10	0.706	-0.198 (0.006)	-0.612 (0.068)	-0.195 (0.006)	-0.584 (0.059)	-0.190 (0.006)	-0.541 (0.047)
	20	0.696	-0.204 (0.006)	-0.632 (0.066)	-0.202 (0.006)	-0.605 (0.057)	-0.196 (0.006)	-0.556 (0.048)
	50	0.683	-0.213 (0.007)	-0.658 (0.058)	-0.210 (0.007)	-0.631 (0.055)	-0.204 (0.006)	-0.580 (0.048)
σ_η^2=0.5	1	0.600	-0.219 (0.007)	-0.485 (0.056)	-0.216 (0.007)	-0.469 (0.045)	-0.210 (0.007)	-0.442 (0.032)
	10	0.573	-0.233 (0.007)	-0.518 (0.056)	-0.230 (0.007)	-0.501 (0.045)	-0.224 (0.007)	-0.471 (0.034)
	20	0.565	-0.237 (0.008)	-0.528 (0.054)	-0.234 (0.007)	-0.511 (0.044)	-0.228 (0.007)	-0.480 (0.034)
	50	0.555	-0.243 (0.008)	-0.541 (0.048)	-0.240 (0.008)	-0.523 (0.042)	-0.233 (0.007)	-0.492 (0.034)
σ_η^2=0.1	1	0.239	-0.166 (0.005)	-0.214 (0.045)	-0.164 (0.005)	-0.209 (0.033)	-0.159 (0.005)	-0.202 (0.018)
	10	0.229	-0.168 (0.005)	-0.218 (0.045)	-0.166 (0.005)	-0.212 (0.033)	-0.161 (0.005)	-0.204 (0.018)
	20	0.226	-0.169 (0.005)	-0.219 (0.045)	-0.167 (0.005)	-0.213 (0.032)	-0.162 (0.005)	-0.205 (0.018)
	50	0.222	-0.170 (0.005)	-0.220 (0.041)	-0.167 (0.005)	-0.214 (0.032)	-0.163 (0.005)	-0.206 (0.017)

(d) La notación de esta tabla es la misma que en la tabla 2.21.

Gráfico 2.25. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie $\ln(y_t^2)$ en un modelo LMSV con d=0.4 y distintos valores de σ_η^2.

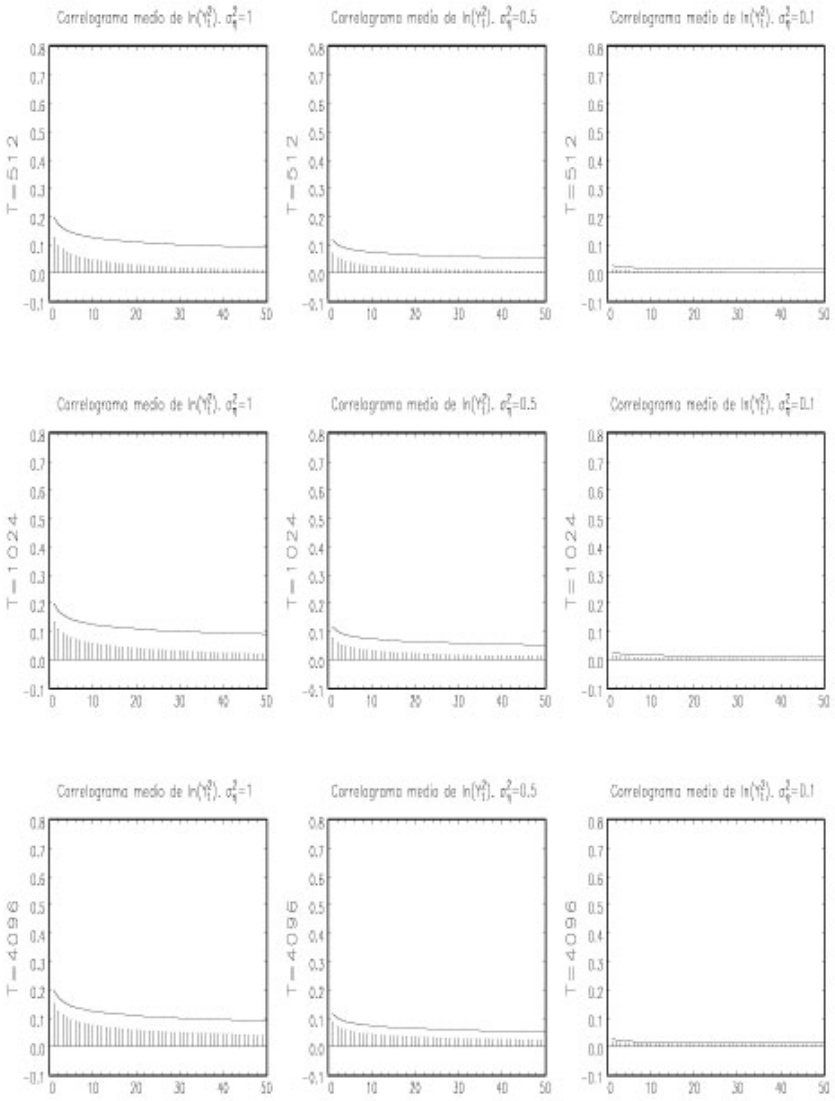


Gráfico 2.26. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie $\ln(y_t^2)$ en un modelo LMSV con d=0.45 y distintos valores de σ_η^2.

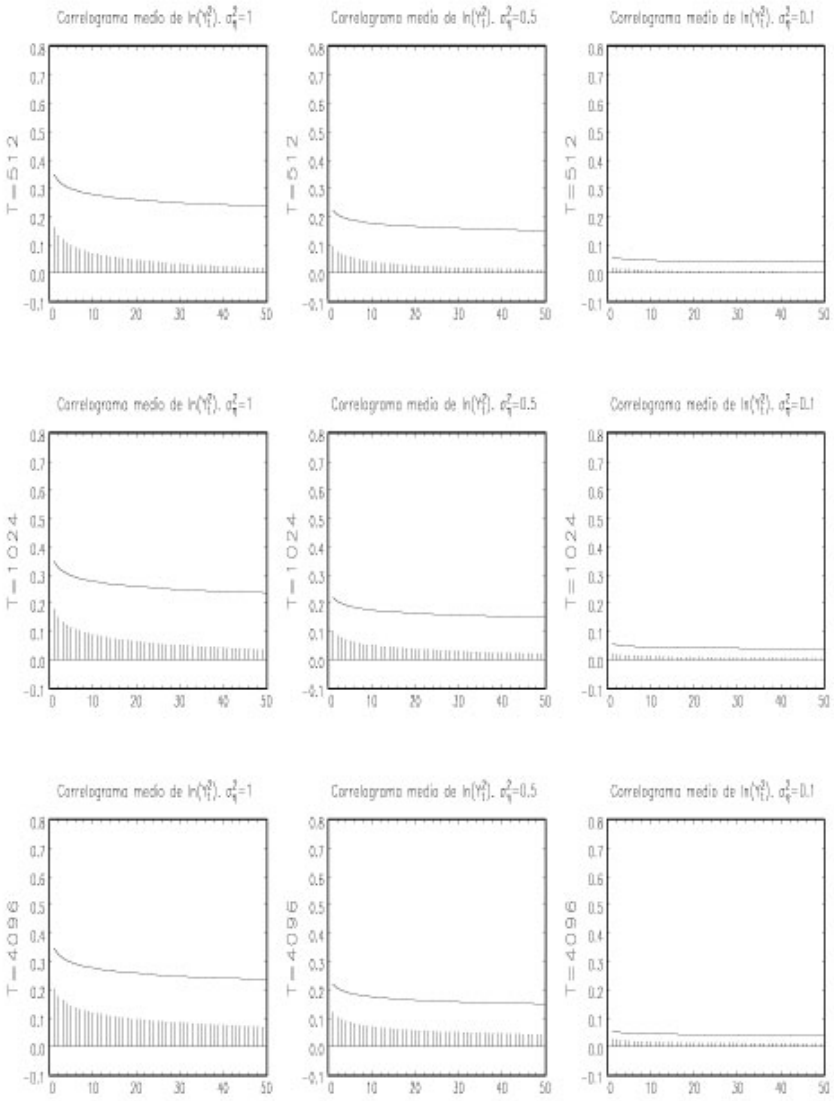
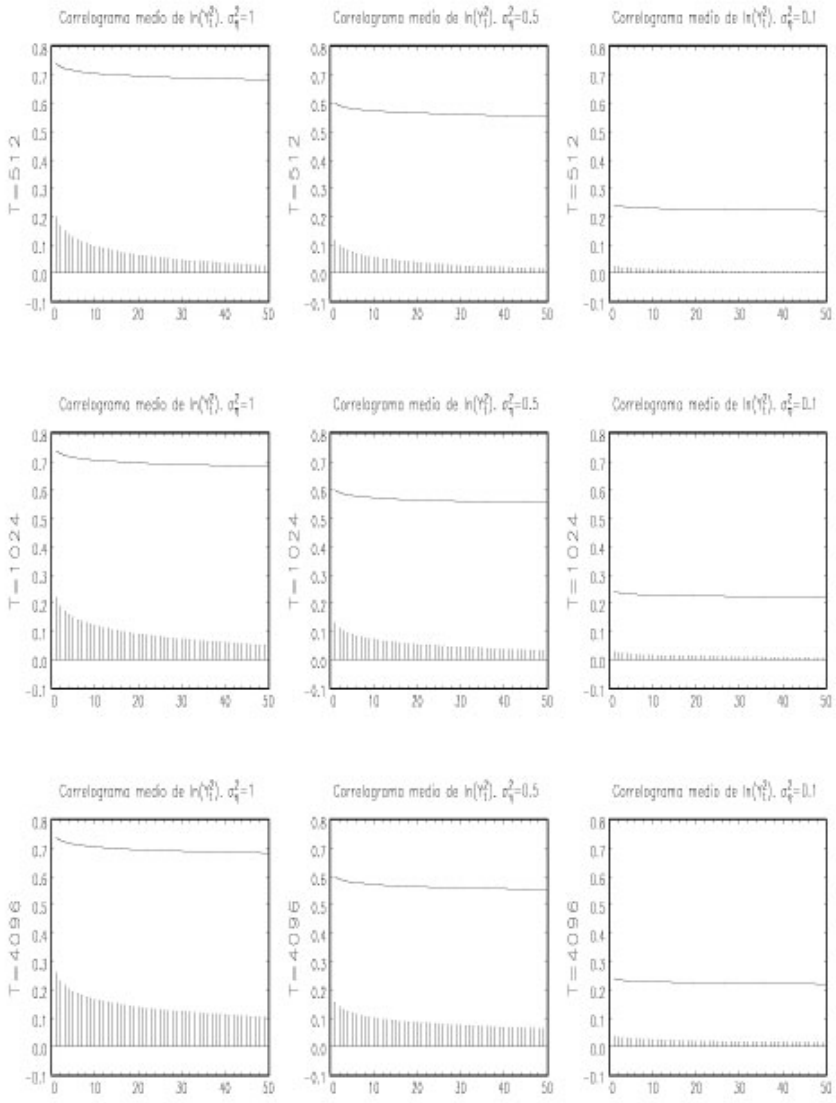


Gráfico 2.27. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie $\ln(y_t^2)$ en un modelo LMSV con d=0.49 y distintos valores de σ_η^2.



Los resultados de las correlaciones muestrales de la serie $x_t=\ln(y_t^2)$ reflejan muchas de las características observadas en las covarianzas muestrales. Una vez más, destaca el hecho de que el sesgo sea siempre negativo y mayor cuanto más persistente es el modelo. Conviene señalar que en el caso de las correlaciones, éste es un problema realmente serio, porque el correlograma es la herramienta más utilizada en la identificación y validación de modelos de series temporales. Concretamente, en los modelos con d=0.45 y d=0.49, la magnitud del sesgo del correlograma es tal que cuestiona su utilización para detectar o validar modelos de volatilidad estocástica con memoria larga, incluso en muestras de tamaño grande; véanse los gráficos 2.26-2.27.

Junto a este resultado, que es sin duda el más importante de cara a las aplicaciones empíricas, existen algunos otros que también merecen ser comentados:

- Al igual que ocurría en el modelo ARSV(1), el sesgo y la desviación estándar simulados de r(k) aumentan con k, y lo mismo ocurre con los valores asintóticos: las correlaciones más alejadas se estiman peor que las primeras.
- Cuando d=0.2, e incluso cuando d=0.4 y σ_η^2 es pequeño, el sesgo asintótico y el sesgo simulado son ambos despreciables y difieren muy poco uno del otro. En este caso, parece que la fórmula [2.87] del sesgo funciona bastante bien. Sin embargo, lo que ocurre en estos casos es que la propia serie x_t es casi ruido blanco y el sesgo es prácticamente cero. Obsérvese que en el modelo con d=0.2, los valores de la desviación típica de las correlaciones muestrales coinciden prácticamente con el valor habitual $1/\sqrt{T}$ para series incorrelacionadas y homocedásticas.
- A medida que el modelo se aproxima a la frontera de la no estacionariedad (d=1/2), el sesgo empírico aumenta drásticamente y la fórmula asintótica deja de funcionar. En estos casos, las diferencias entre el sesgo asintótico y el sesgo empírico son considerables, incluso en muestras grandes.

- El sesgo empírico disminuye con σ_η^2 (comparar por columnas los correlogramas de los gráficos 2.25-2.27), sin embargo, esto no es cierto, en general, para el sesgo asintótico. De hecho, derivando en la expresión [2.87] con respecto a σ_η^2, se comprueba que el tamaño del sesgo, $|E(r(k))-\rho(k)|$, decrece con σ_η^2 si $\rho(k)\leq 0.5$, pero cuando $\rho(k)>0.5$, el sesgo aumenta con σ_η^2 hasta alcanzar un máximo en $\sigma_\eta^2 = \frac{\sigma_\xi^2}{2\rho(k)-1}\frac{[\Gamma(1-d)]^2}{\Gamma(1-2d)}$, y luego decrece. Por ejemplo, con d=0.49, el sesgo absoluto máximo se obtendría en $\sigma_\eta^2=0.38$, lo que justifica que el sesgo asintótico obtenido con $\sigma_\eta^2=0.5$ sea mayor que con $\sigma_\eta^2=0.1$. Esto confirma, una vez más, que la fórmula del sesgo en [2.87] no es válida para el modelo LMSV.
- En el modelo con d=0.2, la fórmula [2.86] de la varianza asintótica funciona bien.
- Sin embargo, en los modelos con d>1/4, las diferencias entre la desviación estándar simulada y asintótica son enormes para todos los retardos y todos los tamaños muestrales. Esto confirma que las fórmulas de Hosking (1996) no son directamente aplicables al modelo LMSV y acentúa la necesidad de encontrar fórmulas alternativas para el sesgo y la varianza de r(k) en modelos con d>1/4.

Finalmente, presentamos los resultados de las simulaciones para la serie y_t^2. La tabla 2.24 recoge los valores del sesgo y la desviación típica muestral de las correlaciones muestrales de la serie y_t^2, sobre las 5000 réplicas, para los modelos LMSV con d={0.4,0.45,0.49}. Los resultados numéricos se completan con unos gráficos (gráficos 2.28-2.30) que representan, para cada modelo y cada tamaño muestral analizado, el correlograma medio sobre las 5000 réplicas junto con la función de autocorrelación poblacional de y_t^2. Los valores de esta función se obtienen a partir de la expresión [1.56] del capítulo 1. Los resultados correspondientes al modelo con d=0.2 no se recogen porque en este caso la función de autocorrelación de y_t^2 es prácticamente cero en todos los retardos, y esto dificulta la interpretación.

Tabla 2.24. Sesgo y desviación estándar empíricos de r(k) para la serie y_t^2 en diversos modelos LMSV.

Tamaño muestral		**T=512**			**T=1024**			**T=4096**		
Parámetros	k	d=0.4	d=0.45	d=0.49	d=0.4	d=0.45	d=0.49	d=0.4	d=0.45	d=0.49
σ_η^2=1	1	-0.025	-0.038	-0.031	-0.019	-0.030	-0.022	-0.011	-0.019	-0.008
		(0.082)	(0.091)	(0.099)	(0.069)	(0.079)	(0.088)	(0.050)	(0.062)	(0.073)
	10	-0.032	-0.047	-0.040	-0.025	-0.038	-0.030	-0.016	-0.025	-0.015
		(0.054)	(0.059)	(0.063)	(0.045)	(0.051)	(0.055)	(0.030)	(0.038)	(0.047)
	20	-0.035	-0.051	-0.045	-0.027	-0.041	-0.033	-0.017	-0.027	-0.018
		(0.047)	(0.050)	(0.069)	(0.039)	(0.043)	(0.046)	(0.028)	(0.033)	(0.043)
	50	-0.036	-0.053	-0.048	-0.028	-0.043	-0.036	-0.019	-0.030	-0.021
		(0.040)	(0.041)	(0.063)	(0.034)	(0.036)	(0.038)	(0.023)	(0.027)	(0.038)
σ_η^2=0.5	1	-0.045	-0.086	-0.113	-0.036	-0.075	-0.100	-0.025	-0.058	-0.079
		(0.071)	(0.078)	(0.084)	(0.055)	(0.063)	(0.071)	(0.036)	(0.044)	(0.052)
	10	-0.046	-0.091	-0.121	-0.038	-0.080	-0.108	-0.026	-0.063	-0.086
		(0.052)	(0.057)	(0.061)	(0.042)	(0.047)	(0.052)	(0.026)	(0.032)	(0.039)
	20	-0.047	-0.093	-0.125	-0.039	-0.082	-0.110	-0.027	-0.064	-0.088
		(0.049)	(0.051)	(0.054)	(0.037)	(0.042)	(0.045)	(0.024)	(0.030)	(0.035)
	50	-0.047	-0.094	-0.126	-0.039	-0.082	-0.112	-0.028	-0.066	-0.091
		(0.042)	(0.043)	(0.044)	(0.034)	(0.037)	(0.039)	(0.022)	(0.026)	(0.031)
σ_η^2=0.1	1	-0.025	-0.064	-0.214	-0.019	-0.057	-0.205	-0.014	-0.048	-0.191
		(0.051)	(0.054)	(0.057)	(0.037)	(0.040)	(0.043)	(0.020)	(0.023)	(0.026)
	10	-0.024	-0.065	-0.220	-0.020	-0.059	-0.211	-0.014	-0.049	-0.197
		(0.046)	(0.048)	(0.050)	(0.034)	(0.036)	(0.039)	(0.018)	(0.021)	(0.024)
	20	-0.024	-0.065	-0.222	-0.020	-0.059	-0.213	-0.014	-0.049	-0.199
		(0.045)	(0.046)	(0.048)	(0.032)	(0.034)	(0.036)	(0.017)	(0.020)	(0.023)
	50	-0.024	-0.066	-0.224	-0.019	-0.059	-0.215	-0.014	-0.050	-0.201
		(0.041)	(0.042)	(0.042)	(0.032)	(0.033)	(0.035)	(0.017)	(0.019)	(0.022)

Gráfico 2.28. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie y_t^2 en un modelo LMSV con d=0.4 y distintos valores de σ_η^2.

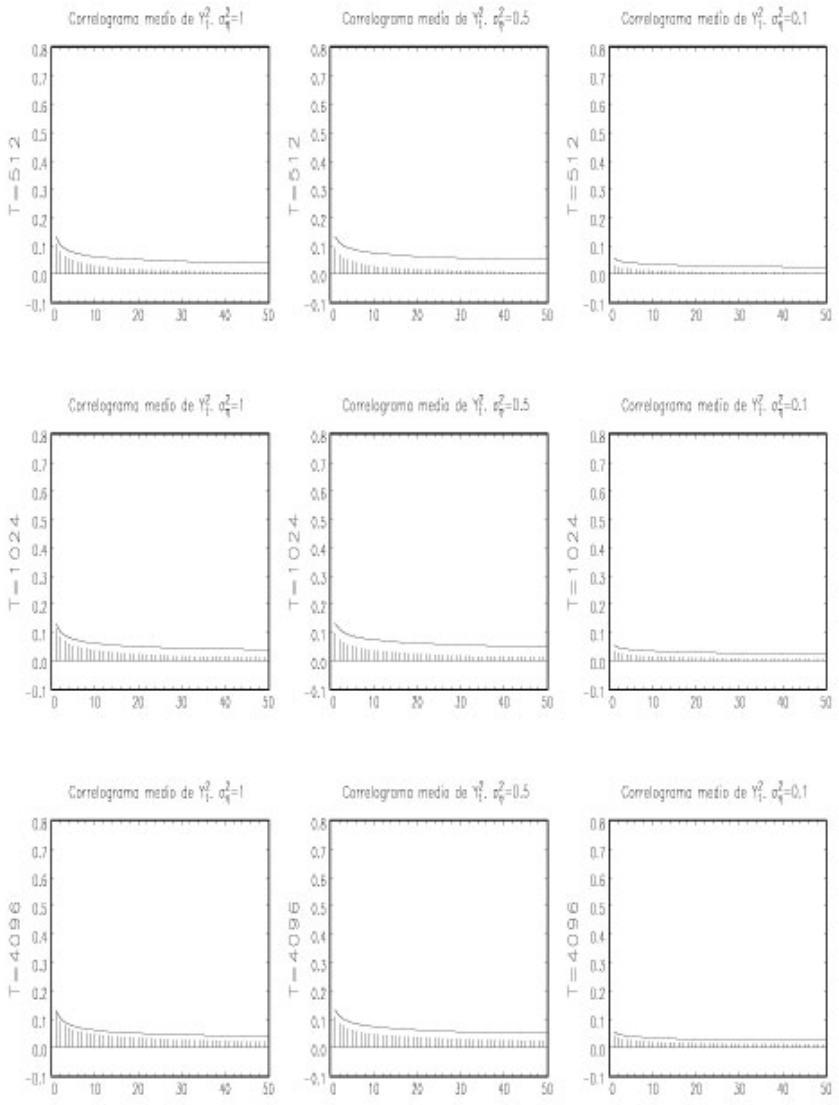


Gráfico 2.29. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie y_t^2 en un modelo LMSV con d=0.45 y distintos valores de σ_η^2.

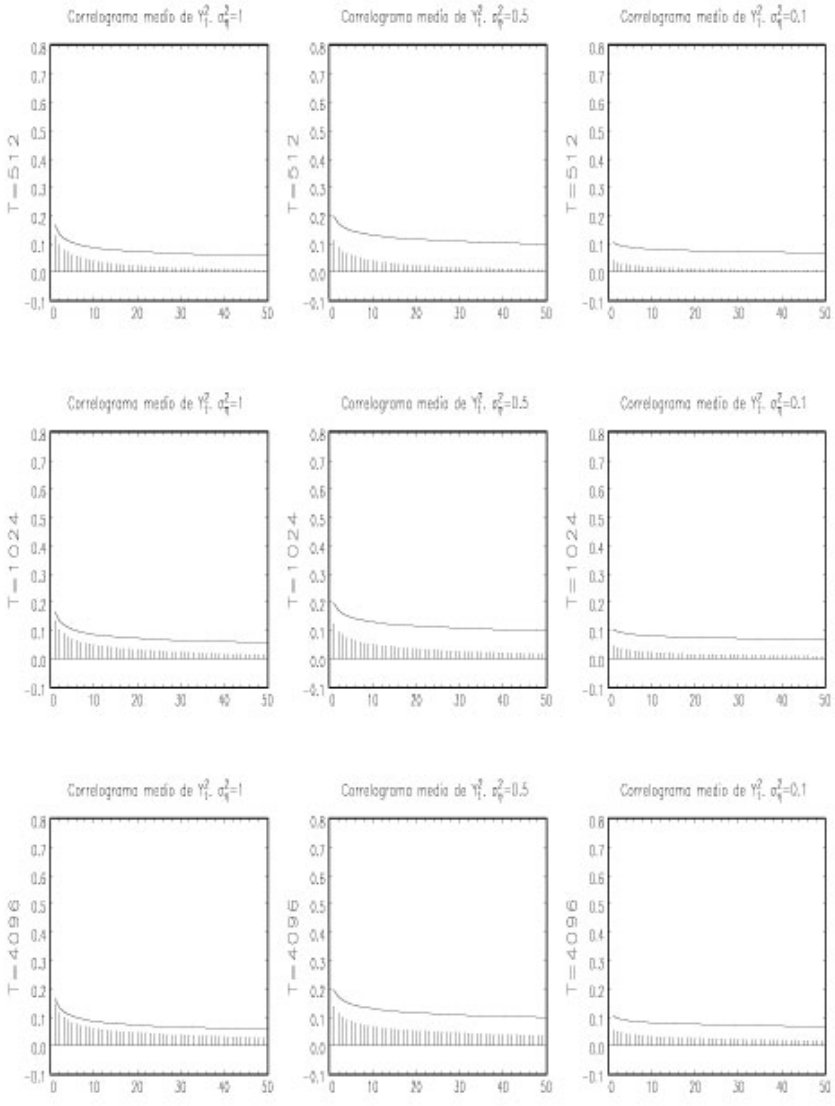
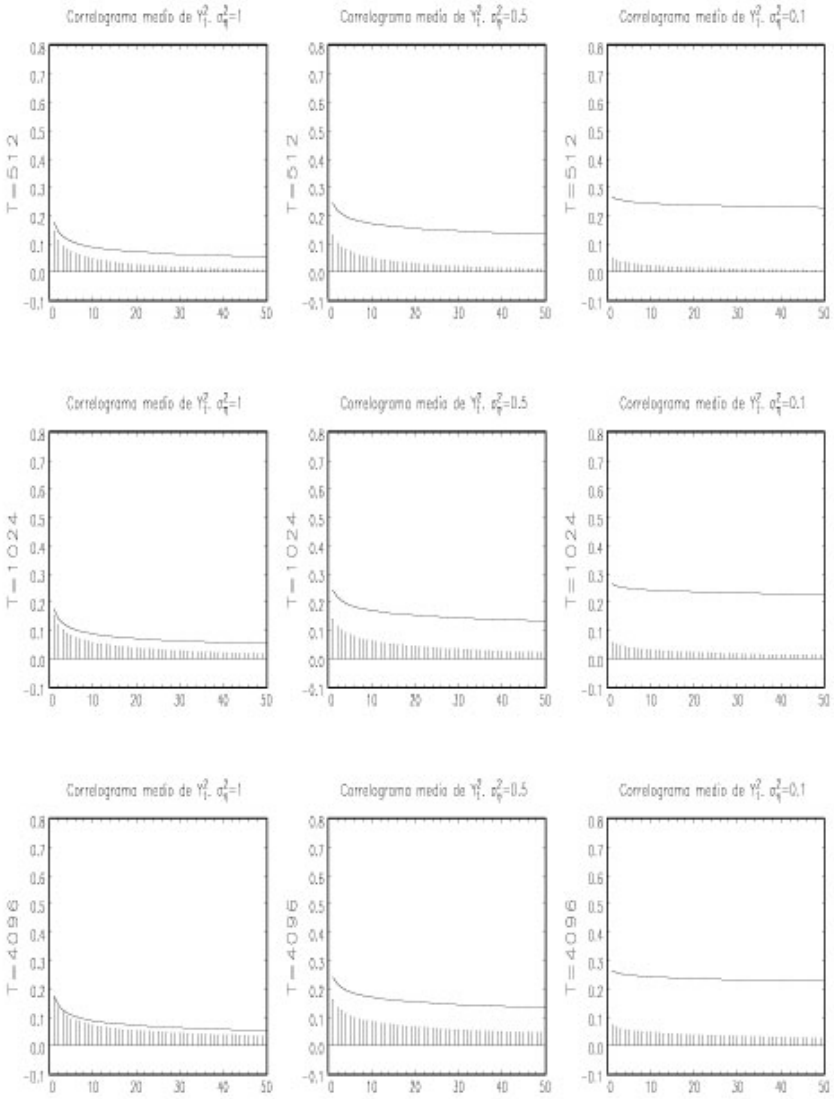


Gráfico 2.30. Correlograma medio y función de autocorrelación poblacional (línea continua) de la serie y_t^2 en un modelo LMSV con d=0.49 y distintos valores de σ_η^2.



Las características más notables que se derivan del análisis de la tabla 2.24 y de los gráficos 2.28-2.30, son las siguientes:

- Una vez más, observamos que el correlograma subestima siempre la función de autocorrelación poblacional.
- El comportamiento del sesgo es diferente dependiendo del valor de d. Por ejemplo, cuando d=0.4 y d=0.45, los mayores valores del sesgo de r(k) se obtienen en los modelos con σ_η^2=0.5, pero cuando d=0.49, el sesgo aumenta continuamente a medida que disminuye el valor de σ_η^2, obteniéndose los valores mayores del sesgo cuando σ_η^2=0.1.
- Como era de esperar, tanto el sesgo como la desviación típica de r(k) disminuyen al aumentar el tamaño muestral. Esto se aprecia claramente comparando por filas los correlogramas de los gráficos 2.28-2.30. Sin embargo, el ritmo de convergencia de la desviación típica es mucho más lento que $\sqrt{T}$.
- Nuevamente, el sesgo y la varianza de r(k) aumentan con d, confirmando lo dicho anteriormente: cuanto más persistente es el modelo, menos fiable es el correlograma como estimador de la función de autocorrelación poblacional.

Capítulo 3.

ESTIMACIÓN DEL MODELO LMSV

3.1. Introducción

El modelo de volatilidad estocástica con memoria larga (LMSV), definido en el capítulo 1, es una generalización del modelo de volatilidad estocástica autorregresivo en el que el proceso que define la ecuación de la volatilidad puede ser un proceso ARMA fraccionalmente integrado (ARFIMA). Podría pensarse entonces, como primera aproximación al problema de estimación del modelo LMSV, en tratar de aplicar alguno de los métodos de estimación propuestos para el modelo autorregresivo. Sin embargo, la extensión no es inmediata. Pensemos, por ejemplo, en el método de estimación pseudo-máximo verosímil propuesto por Harvey, Ruiz y Shephard (1994). Este método utiliza la representación en el espacio de los estados del modelo SV autorregresivo transformado, dado por las ecuaciones:

$$x_t = \ln(y_t^2) = \mu + h_t + \xi_t \qquad [3.1a]$$

$$h_t = \phi\, h_{t-1} + \eta_t \qquad [3.1b]$$

donde [3.1a] es la ecuación de medida y [3.1b] es la ecuación de transición. Con esta estructura, el modelo puede estimarse por pseudo-máxima verosimilitud, tratando la

perturbación del modelo, ξ_t, como si fuera normal, y utilizando el filtro de Kalman para obtener la descomposición de la verosimilitud en los errores de predicción. Una vez obtenidas las estimaciones de los parámetros, el filtro de Kalman y el correspondiente algoritmo de predicción, permiten además obtener una estimación suavizada de la propia serie de la volatilidad.

En su formulación más sencilla, el modelo LMSV transformado admite la siguiente representación lineal:

$$x_t = \ln(y_t^2) = \mu + h_t + \xi_t \qquad [3.2a]$$

$$(1-L)^d h_t = \eta_t \qquad [3.2b]$$

BIBLIOTECA VIRTUAL

donde ξ_t es un ruido blanco no gaussiano, d es el parámetro de integración fraccional y η_t es un ruido blanco NID$(0, \sigma_\eta^2)$ independiente de ξ_t. Esta representación también puede interpretarse como un modelo estructural en el espacio de los estados, donde [3.2a] es la ecuación de medida y [3.2b] es la ecuación de transición. La ecuación de medida nos dice que la serie x_t es la suma de un proceso h_t más una perturbación ξ_t no gaussiana, y la ecuación de transición nos dice que ese proceso h_t es un ARFIMA(0,d,0). En principio, el filtro de Kalman sólo podría aplicarse en la estimación de este modelo si la correspondiente representación AR(∞) (o la correspondiente representación MA(∞)) del proceso h_t se trunca en un retardo suficientemente grande. Este procedimiento multiplicaría el número de parámetros a estimar y llevaría implícito un cierto grado de imprecisión, especialmente en los modelos más persistentes (d$\approx$1/2), ya que en estos casos ningún proceso autorregresivo o de medias móviles, por muy alto que sea su orden, puede reproducir las propiedades dinámicas de un proceso ARFIMA. Por estos motivos, el atractivo que tenía la estimación pseudo-máximo verosímil a través del filtro de Kalman para el modelo de volatilidad estocástica autorregresivo, se pierde en el modelo SV con memoria larga.

Sin embargo, el trabajo reciente de Chan y Palma (1998) sobre estimación de procesos ARFIMA abre nuevas expectativas sobre la posibilidad de aplicar el filtro de Kalman a la estimación del modelo LMSV. Chan y Palma (1998) demuestran que la función de verosimilitud exacta de un modelo ARFIMA gaussiano puede obtenerse a través del filtro de Kalman en un número finito de pasos, a pesar de que la representación en el espacio de los estados de dicho modelo es necesariamente infinita. Además, estos autores proponen un método de estimación máximo verosímil aproximado, que consiste en truncar la representación MA(∞) del modelo ARFIMA en primeras diferencias y maximizar la función de verosimilitud del modelo truncado, utilizando una representación en el espacio de los estados de dicho proceso. Los estimadores que se obtienen son consistentes y asintóticamente normales y eficientes. La extensión de los resultados de Chan y Palma (1998) al modelo LMSV constituye un tema de investigación en el que aún estamos trabajando.

Si no se utiliza la representación en el espacio de los estados, la estimación pseudo-máximo verosímil del modelo LMSV en el dominio temporal requiere cálculos tediosos. Cada vez que se evalúe la función de verosimilitud de $\{x_1,\ldots,x_T\}$ habrá que invertir la matriz de varianzas-covarianzas, de orden TxT, lo cual puede llegar a ser impracticable en muestras grandes. Véase, por ejemplo, Sowell (1992a) para un desarrollo riguroso de este método en procesos ARFIMA para la media, y Cheung y Diebold (1994) y Smith, Sowell y Zin (1997) para una discusión de las ventajas e inconvenientes del mismo.

Harvey (1998) y Breidt, Crato y de Lima (1998) proponen una forma alternativa de estimar el modelo LMSV, relativamente fácil de aplicar y razonablemente eficiente, que consiste en maximizar la aproximación de Whittle a la función de verosimilitud gaussiana de x_t en el dominio de las frecuencias. Como vimos en la sección 1.2, este método ha sido estudiado con rigor en el contexto de los modelos ARFIMA y existen ya muchos resultados teóricos sobre sus propiedades. Fox y Taqqu (1986), Dahlhaus (1989) y Giriatis y Surgalis (1990) demuestran, bajo diferentes supuestos, que dichos estimadores son consistentes y asintóticamente

normales. Además, Cheung y Diebold (1994) comprueban, mediante simulaciones de Monte Carlo, que estos estimadores se comportan bien, para muestras de tamaño medio/grande, en procesos donde la media sea desconocida, que es lo habitual en aplicaciones empíricas. Para modelos LMSV, Breidt, Crato y de Lima (1998) demuestran que los estimadores pseudo-máximo verosímiles en el dominio de las frecuencias son consistentes. Además, estos autores comprueban, para ciertos valores de los parámetros, que dichos estimadores tienen buenas propiedades en muestras finitas. Sin embargo, este último resultado debe interpretarse con cautela porque los valores que utilizan Breidt, Crato y de Lima (1998) en sus simulaciones no son muy realistas y, como veremos más adelante, el comportamiento de estos estimadores se deteriora considerablemente cuando los parámetros toman valores similares a los obtenidos en la estimación de series reales. Harvey (1998) propone además un algoritmo para obtener una estimación suavizada de la volatilidad en modelos LMSV. Este algoritmo requiere la inversión de una matriz TxT, lo cual puede resultar inviable en muestras grandes. En estos casos, el mismo autor sugiere que, debido a que los pesos dados a las observaciones distantes son muy pequeños, el suavizado puede hacerse con los pesos obtenidos para una muestra más pequeña sin que se pierda mucha precisión.

La estimación máximo verosímil del modelo LMSV también puede realizarse utilizando un enfoque bayesiano. Hsu y Breidt (1997), por ejemplo, proponen un método de estimación mediante muestreo de Gibbs que permite obtener la distribución a posteriori de los parámetros del modelo LMSV y la estimación suavizada de la volatilidad. So (1999) estudia el problema de estimación de modelos ARFIMA(0,d,0) "contaminados" por un ruido gaussiano, y propone técnicas bayesianas de estimación MCMC (Markov Chain Monte Carlo) que podrían aplicarse al modelo LMSV. Aunque las propiedades de estos métodos sean óptimas, estas técnicas son computacionalmente muy costosas y no son fáciles de extender a modelos más complicados. Como señalan Hsu y Breidt (1997), existen serios problemas de aplicación cuando el tamaño muestral aumenta o cuando el proceso ARFIMA contiene alguna componente autorregresiva o de medias móviles. Recientemente, Chan y Petris (2000) han propuesto un método de

estimación para el modelo LMSV que recoge la idea de Chan y Palma (1998) de utilizar la representación en el espacio de los estados del modelo ARFIMA truncado en primeras diferencias. Sin embargo, no se conoce ningún resultado sobre las propiedades de este estimador.

Wright (1999) ha propuesto un estimador del modelo LMSV estacionario basado en el método de los momentos generalizados (GMM) y ha demostrado que dicho estimador es $\sqrt{T}$ -consistente y asintóticamente normal en el rango $-1/2<d<1/4$. Según el autor, uno de los atractivos de este estimador es, precisamente, que su distribución asintótica es conocida. Sin embargo, la utilidad práctica de este resultado puede ser limitada porque las estimaciones del parámetro d en series reales suelen estar cerca de la frontera de la no estacionariedad (d=1/2), donde la distribución asintótica del estimador GMM es desconocida. Además, si se analizan los resultados del estudio de Monte Carlo en Wright (1999) se observa, en primer lugar, que la distribución asintótica no es una buena aproximación a la distribución en muestras finitas, ni siquiera para tamaños muestrales grandes (T=4000), y en segundo lugar, que el estimador GMM se compara desfavorablemente con el estimador pseudo-máximo verosímil en el dominio de las frecuencias, ya que ambos tienen varianzas similares pero este último tiene un sesgo mucho menor.

Como alternativa a los métodos de estimación paramétricos que acabamos de describir, Deo y Hurvich (1998) han propuesto un estimador semiparamétrico para el parámetro de memoria, d, basado en la idea original de Geweke y Porter-Hudak (1983) para modelos ARFIMA. Este estimador se obtiene a partir de un modelo de regresión para el logaritmo del periodograma de la serie transformada, $x_t=\ln(y_t^2)$, utilizando sólo las m primeras frecuencias en torno al origen. Bajo ciertas condiciones sobre m, Deo y Hurvich (1998) derivan las expresiones del sesgo y de la varianza asintótica del estimador, y concluyen que el sesgo es siempre negativo y que éste depende de m y del cociente entre las varianzas de h_t y ξ_t, de forma que cuanto menor es ese cociente y/o mayor es m, mayor es el sesgo. Además, estos autores demuestran que, cuando m verifica ciertas condiciones, que dependen de d, el estimador $\hat{d}_{GPH}$ es asintóticamente

normal, pero con una tasa de convergencia $\sqrt{m}$ inferior a la $\sqrt{T}$ típica de los estimadores paramétricos gaussianos. Aunque estos autores presentan resultados importantes para la estimación del parámetro d, no presentan ningún resultado sobre la estimación de los otros parámetros del modelo ni sobre la estimación de la volatilidad.

En este capítulo nos centraremos en el estudio del método de estimación pseudo-máximo verosímil (PMV) en el dominio de las frecuencias propuesto por Harvey (1998) y Breidt, Crato y de Lima (1998). De forma adicional, recogemos también los resultados de nuestras primeras investigaciones sobre la posible aplicación de los resultados de Chan y Palma (1998) a la estimación máximo verosímil del modelo LMSV mediante el filtro de Kalman.

Salvo que se indique lo contrario, trabajaremos fundamentalmente con el modelo LMSV básico, en el que el proceso subyacente h_t es un proceso ARFIMA(0,d,0). Trabajar con este modelo, además del interés que pueda tener por sí mismo, nos permite simplificar la notación y estudiar en profundidad el problema de estimación de los dos parámetros del modelo, d y σ_η^2. Posteriormente, analizaremos los problemas de identificación y estimación en el modelo SV autorregresivo con memoria larga (ARLMSV), en el que el proceso subyacente h_t es un ARFIMA(1,d,0) con una componente autorregresiva. En cualquier caso, los resultados que se presentan pueden extenderse de forma inmediata a un modelo más general en el que h_t sea un proceso ARFIMA(p,d,q). De hecho, Breidt, Crato y de Lima (1998) demuestran la consistencia de los estimadores PMV para este modelo más general.

El capítulo se organiza como sigue. En la sección 3.2 se describe el método de estimación PMV en el dominio de las frecuencias y se presentan las propiedades asintóticas derivadas por Breidt, Crato y de Lima (1998). Se discuten además las posibles ventajas de este método frente a la estimación máximo verosímil en el dominio temporal. La sección 3.3 recoge los resultados de un extenso estudio de Monte Carlo que analiza las propiedades de los estimadores PMV en muestras finitas. Nuestro estudio complementa el realizado por Breidt, Crato y de Lima (1998) en varios aspectos. En primer lugar, incorporamos resultados para el estimador de la

varianza de la perturbación del proceso h_t, que es precisamente el que peores propiedades tiene en muestras finitas. En segundo lugar, estudiamos el comportamiento de los estimadores para un rango más amplio de valores de los parámetros, que incluye valores más realistas que los que se presentan en Breidt, Crato y de Lima (1998) y modelos no estacionarios. Nuestro estudio nos permite detectar los problemas que existen cuando se estiman modelos próximos a las fronteras de la no estacionariedad ($d\approx 1/2$) y/o de la homocedasticidad ($\sigma_\eta^2 \approx 0$), en cuyo caso será necesario disponer de tamaños muestrales extremadamente grandes para que las inferencias sean fiables. En la sección 3.3 también discutimos el problema de identificación que existe en el modelo SV autorregresivo y con memoria larga cuando el proceso que define la volatilidad tiene casi una raíz unitaria. A través de simulaciones descubrimos que en estos casos el estimador PMV no va a ser capaz de distinguir la procedencia de la no estacionariedad. En la sección 3.4 se describe el método de estimación PMV en el dominio temporal mediante el filtro de Kalman y se presenta un pequeño estudio de Monte Carlo para comparar sus propiedades en muestras finitas con las del estimador PMV en el dominio de las frecuencias. Finalmente, en la sección 3.5 se describe el método propuesto por Harvey (1998) para obtener una estimación suavizada de la volatilidad en modelos LMSV y se presentan algunos resultados con series simuladas. Se resuelve además el problema de adaptación de este método a series muy largas, donde la propuesta original es inviable.

3.2. Estimación pseudo-máximo verosímil en el dominio de las frecuencias

El método de estimación pseudo-máximo verosímil en el dominio de las frecuencias para los modelos LMSV, propuesto simultáneamente por Harvey (1998) y Breidt, Crato y de Lima (1998), aprovecha el hecho de que los modelos LMSV admitan una representación en el espacio de los estados para la serie $x_t=\ln(y_t^2)$. En el

modelo LMSV básico, esa representación viene dada por las ecuaciones [3.2a]-[3.2b], donde μ es la media de la serie x_t, $\xi_t = \ln(\varepsilon_t^2) - E[\ln(\varepsilon_t^2)]$ es un ruido blanco no gaussiano de media 0 y varianza σ_ξ^2, cuyas propiedades dependen de las de ε_t[1], η_t es $NID(0, \sigma_\eta^2)$ e independiente de la perturbación ε_t, y d es el orden de integración fraccional, $0 \leq d \leq 1$.[2] Cuando d=0, el proceso h_t se reduce a un ruido blanco η_t y cuando d=1 se obtiene el modelo SV con paseo aleatorio en [1.25], analizado en Harvey, Ruiz y Shephard (1994). Por tanto, los valores de interés en nuestro caso son $0<d<1$.

La serie x_t en [3.2a] puede interpretarse como la suma de un proceso ARFIMA(0,d,0) gaussiano más una perturbación, ξ_t, que es un ruido blanco de media 0 y varianza σ_ξ^2, cuya distribución no es normal. El método de estimación propuesto por Harvey (1998) y Breidt, Crato y de Lima (1998) consiste en maximizar la aproximación de Whittle a la función de verosimilitud de $\{x_1,\ldots,x_T\}$ en el dominio de las frecuencias, calculada como si ξ_t tuviese una distribución normal $N(0, \sigma_\xi^2)$. Es decir, la estimación del modelo [3.2] se realizará tratando dicho modelo "como si fuera gaussiano", y por eso los estimadores que se obtienen se denominan *pseudo*-máximo verosímiles (PMV).

Según vimos en el capítulo 1, la aproximación discreta de Whittle a la función de verosimilitud en el dominio de las frecuencias para un proceso x_t gaussiano y estacionario es, salvo constantes,

$$\ln L(\psi;\mathbf{x}) = -\frac{1}{2T}\sum_{j=1}^{T-1}\left[\ln f(\lambda_j;\psi) + \frac{I(\lambda_j)}{f(\lambda_j;\psi)}\right] \qquad [3.3]$$

donde ψ es el vector de parámetros a estimar, T es el tamaño muestral, $\mathbf{x}=(x_1,\ldots,x_T)'$ es

[1] Véase la sección 1.1.2.1 para una descripción de las propiedades de ξ_t.

[2] Conservamos aquí la notación de Harvey (1998) que considera el modelo LMSV definido en $0 \leq d \leq 1$, a diferencia de Breidt, Crato y de Lima (1998) que consideran el rango $-1/2<d<1/2$, donde el modelo es estacionario e invertible.

el vector de observaciones, $\lambda_j=\frac{2\pi j}{T}$, $f(\lambda_j;\psi)$ es la densidad espectral de x_t evaluada en λ_j e $I(\lambda_j)$ es la j-ésima ordenada del periodograma de x_t,

$$I(\lambda_j)=\frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{-i\lambda_j t}\right|^2 , j=1,2,\ldots,T-1^3$$

Además, como las funciones $I(\lambda_j)$ y $f(\lambda_j;\psi)$ son simétricas con respecto a π, no es necesario evaluarlas en todas las frecuencias, lo cual simplifica considerablemente el cálculo de la función de verosimilitud. Así, los estimadores PMV en el dominio de las frecuencias se obtendrán maximizando con respecto a ψ la función $\ln L(\psi;\mathbf{x})$ en [3.3] o, equivalentemente, minimizando la función:

$$\tilde{L}_W(\psi;\mathbf{x})=\frac{1}{T}\sum_{j=1}^{[T/2]}\delta_j\left[\ln f(\lambda_j;\psi)+\frac{I(\lambda_j)}{f(\lambda_j;\psi)}\right] \qquad [3.4]$$

donde el subíndice W hace referencia a Whittle, [T/2] denota la parte entera de T/2, y δ_j es la función:

$$\delta_j=\begin{cases}1/2 & \text{si } j=[T/2] \text{ y T es par}\\ 1 & \text{en el resto}\end{cases}$$

Las condiciones de primer orden para minimizar la función $\tilde{L}_W(\psi;\mathbf{x})$ en [3.4] se obtendrían resolviendo la ecuación $\frac{\partial \tilde{L}_W(\psi;\mathbf{x})}{\partial\psi}=0$. Habitualmente, esta ecuación no es lineal en ψ y el problema de minimizar $\tilde{L}_W(\psi;\mathbf{x})$ se resuelve con métodos numéricos.

[3] Obsérvese que en esta definición del periodograma la serie x_t no aparece corregida por su media muestral. La razón es que esa corrección es inocua para las frecuencias j=1,2,…,T-1, que son las únicas que aparecen en la verosimilitud.

Entre la gran variedad de algoritmos que existen para minimizar numéricamente una función, se ha elegido el método de Newton-Raphson. Éste es un método iterativo en el cual las soluciones se van actualizando según la siguiente expresión:

$$\tilde{\psi} = \hat{\psi} - \left[\frac{\partial^2 \tilde{L}_W(\psi;\mathbf{x})}{\partial\psi\,\partial\psi'}\right]^{-1}_{|\psi=\hat{\psi}} \left[\frac{\partial \tilde{L}_W(\psi;\mathbf{x})}{\partial\psi}\right]_{|\psi=\hat{\psi}}$$

donde $\hat{\psi}$ es el estimador actual, $\tilde{\psi}$ es el estimador actualizado, y $\frac{\partial \tilde{L}_W(\hat{\psi};\mathbf{x})}{\partial\psi}$ y $\frac{\partial^2 \tilde{L}_W(\hat{\psi};\mathbf{x})}{\partial\psi\,\partial\psi'}$ son el gradiente y el hessiano, respectivamente, ambos evaluados en $\hat{\psi}$.

Como la función $\tilde{L}_W(\psi;\mathbf{x})$ en [3.4] sólo depende de los parámetros a través de $f(\lambda_j;\psi)$, el gradiente y el hessiano dependerán únicamente de las derivadas parciales de $f(\lambda_j;\psi)$. En concreto, diferenciando $\tilde{L}_W(\psi;\mathbf{x})$ con respecto a ψ se obtiene la siguiente expresión del gradiente:

$$\frac{\partial \tilde{L}_W(\psi;\mathbf{x})}{\partial\psi} = \sum_{j=1}^{[T/2]} \delta_j \frac{[f_j - I(\lambda_j)]}{f_j^2} \frac{\partial f_j}{\partial\psi} \qquad [3.5]$$

donde $f_j=f(\lambda_j;\psi)$, y diferenciando una segunda vez se obtiene la expresión del hessiano:

$$\frac{\partial^2 \tilde{L}_W(\psi;\mathbf{x})}{\partial\psi\,\partial\psi'} = \sum_{j=1}^{[T/2]} \delta_j \left\{\frac{[f_j - I(\lambda_j)]}{f_j^2} \frac{\partial^2 f_j}{\partial\psi\,\partial\psi'} - \frac{[f_j - 2I(\lambda_j)]}{f_j^3} \frac{\partial f_j}{\partial\psi} \frac{\partial f_j}{\partial\psi'}\right\} \qquad [3.6]$$

Los valores para inicializar el método de Newton Raphson, se pueden obtener mediante un procedimiento de "búsqueda" que localice el mínimo sobre la superficie de la función $\tilde{L}_W(\psi;\mathbf{x})$ evaluada en una malla de valores del espacio paramétrico. Si

fuera necesario pueden utilizarse transformaciones para restringir el campo de variación a un intervalo finito sobre el que hacer la búsqueda.[4]

Para obtener los estimadores pseudo-máximo verosímiles del modelo LMSV supondremos adicionalmente que ε_t es N(0,1). En este caso, la varianza de la perturbación $\xi_t=\ln(\varepsilon_t^2)-E[\ln(\varepsilon_t^2)]$ tiene un valor conocido, $\sigma_\xi^2=\pi^2/2\approx 4.93$, y, por tanto, el vector de parámetros a estimar para el modelo LMSV básico, por ejemplo, es $\psi=(d,\sigma_\eta^2)$. Si no suponemos para ε_t una distribución conocida, podemos considerar σ_ξ^2 como un parámetro más del modelo y estimarlo conjuntamente con d y σ_η^2, lo cual podría aumentar la eficiencia. Ruiz (1994b) presenta una discusión de este problema en el modelo ARSV(1) y en el modelo SV integrado.

Una de las propiedades más atractivas del método de estimación PMV en el dominio de las frecuencias es que es invariante a la verdadera, pero desconocida, media del proceso, μ, debido a que la ordenada del periodograma de x_t en la frecuencia cero no se utiliza en la estimación.[5] Esto significa que el método de estimación PMV en el dominio de las frecuencias evita estimar la media de x_t, lo cual puede suponer una gran ventaja cuando se compara este método con su análogo en el dominio temporal.

La estimación en el dominio temporal requiere estimar μ en primer lugar, utilizando, por ejemplo, la media muestral, $\hat{\mu}=\overline{X}$, y después maximizar la función de verosimilitud para la serie centrada, $(x_t-\overline{X})$. La memoria larga asociada al modelo LMSV con d>0 puede hacer que la media muestral converja hacia μ muy despacio y, en consecuencia, los estimadores PMV en el dominio temporal pueden no tener buenas propiedades en muestras finitas. Cheung y Diebold (1994) presentan una discusión interesante sobre este problema en el ámbito de los procesos ARFIMA. Nosotras presentamos en la sección 3.4 algunos resultados para el modelo LMSV básico.

---

[4] Véase Harvey (1990, p. 130).

[5] Es bien conocido que la ordenada del periodograma en la frecuencia cero es la única que depende de la media. Véase por ejemplo Priestley (1981, p. 417).

El método de estimación PMV en el dominio de las frecuencias que acabamos de describir, requiere en su planteamiento que la serie x_t sea estacionaria para que su densidad espectral esté definida. Como vimos en el capítulo 1, la condición para que un modelo de volatilidad estocástica sea estacionario es que lo sea el proceso h_t. En el modelo LMSV básico, el proceso h_t es un ARFIMA(0,d,0) y, por tanto, la condición de estacionariedad es $d<1/2$. Si $d\geq 1/2$, h_t ya no es estacionario y en consecuencia tampoco lo es x_t, aunque sí lo son sus primeras diferencias. Es necesario pues tratar de forma diferente la estimación de modelos estacionarios ($d<1/2$) y no estacionarios ($d\geq 1/2$).

3.2.1. Estimación del modelo LMSV estacionario

En el intervalo $0<d<1/2$, el proceso h_t que define la volatilidad del modelo LMSV es estacionario y, en consecuencia, también lo es la serie x_t. Por tanto, el método de estimación PMV en el dominio de las frecuencias puede aplicarse directamente sobre la serie $\{x_1,\ldots,x_T\}$. Los estimadores PMV de ψ se obtendrán minimizando con respecto a ψ la función $\tilde{L}_W(\psi;\mathbf{x})$ en [3.4], donde $f(\lambda_j;\psi)$ es la función de densidad espectral de x_t. Si el modelo que se considera es el modelo LMSV básico en el que h_t es un proceso ARFIMA(0,d,0), la densidad espectral de x_t es de la forma:

$$f(\lambda_j;\psi)=\frac{\sigma_\eta^2}{2\pi}[2(1-\cos\lambda_j)]^{-d}+\frac{\sigma_\xi^2}{2\pi} \qquad [3.7]$$

con $\psi=(d,\sigma_\eta^2)$ y $\sigma_\xi^2=\pi^2/2$. Si el modelo que se estima es, por ejemplo, un modelo ARLMSV, en el que h_t es un ARFIMA(1,d,0) de la forma $(1-\phi L)(1-L)^d h_t=\eta_t$, entonces la densidad espectral de x_t es:

$$f(\lambda_j;\psi)=\frac{\sigma_\eta^2}{2\pi}\frac{[2(1-\cos\lambda_j)]^{-d}}{(1-2\phi\cos\lambda_j+\phi^2)}+\frac{\sigma_\xi^2}{2\pi}$$

donde $\psi=(\phi,d,\sigma_\eta^2)$ y $\sigma_\xi^2=\pi^2/2$. Derivando la función $f(\lambda_j;\psi)$ con respecto a ψ y sustituyendo en [3.5] y [3.6] se obtienen las expresiones del gradiente y el hessiano necesarias para implementar el algoritmo de Newton-Raphson y obtener el mínimo de la función $\tilde{L}_W(\psi;\mathbf{x})$.

Uno de los problemas de la estimación de modelos LMSV es que los parámetros a estimar están sujetos a ciertas restricciones ($|\phi|<1$, $0<d<1/2$, $\sigma_\eta^2>0$). Como posible solución a este problema algunos autores sugieren poner "barreras" sobre el espacio paramétrico de forma que los parámetros se mantengan suficientemente alejados de los valores de la frontera. Hauser (1999) adopta esta solución para la estimación de un modelo ARFIMA(1,d,0) imponiendo $d\leq 0.4965$ y $|\phi|<0.9965$. Jacquier, Polson y Rossi (1994) utilizan este mismo criterio en la estimación del modelo ARSV(1) imponiendo $\sigma_\eta>0.001$ y $|\phi|<0.995$. Con estas barreras se asegura que tanto el modelo que se estima como los estimadores que se obtienen estén bien definidos, cosa que no ocurriría en los valores de la frontera ($\phi=1$ ó $d=1/2$) donde el modelo pasaría a ser no estacionario y tendría propiedades muy diferentes a las del modelo estacionario que se quiere estimar. Con esta solución, los problemas de estimación que pudieran surgir en la frontera del espacio paramétrico, principalmente en la frontera de la no estacionariedad ($d=1/2$, $\phi=1$), se detectarían en las "barreras".

Otros autores aplican la estimación máximo verosímil en el dominio de las frecuencias sin imponer ninguna restricción sobre el parámetro d; véase por ejemplo Smith, Sowell y Zin (1997) para modelos ARFIMA y Breidt, Crato y de Lima (1998) para modelos LMSV. Estos autores argumentan que la función de densidad espectral del proceso ARFIMA(0,d,0) definida en [1.31], puede evaluarse incluso para $d\geq 1/2$, donde el proceso no es estacionario. En consecuencia, también puede evaluarse en $d\geq 1/2$ la función de densidad espectral de x_t en [3.7] y, por tanto, sobre ese intervalo se puede obtener un valor de la función $\tilde{L}_W(\psi;\mathbf{x})$ definida en [3.4]. Desde un punto de vista teórico, este argumento puede no ser muy riguroso porque, si bien es cierto que la función $f(\lambda_j;\psi)$ puede evaluarse en $d\geq 1/2$, en este intervalo el proceso h_t que define la

volatilidad ya no es estacionario y, por tanto, $f(\lambda_j;\psi)$ ya no sería la densidad espectral de x_t y la función que maximizamos dejaría de ser la función de verosimilitud de x_t.

Nosotras hemos decidido resolver el problema de las restricciones utilizando ciertas transformaciones sobre los parámetros, con el fin de conseguir un vector paramétrico no restringido y poder así aplicar correctamente el algoritmo de Newton-Raphson sobre los parámetros transformados.[6]

Breidt, Crato y de Lima (1998) demuestran el siguiente teorema que asegura la consistencia de los estimadores PMV en el dominio de las frecuencias para el modelo LMSV estacionario general, en el que el proceso h_t es un ARFIMA(p,d,q).

Teorema 1 de Breidt, Crato y de Lima (1998):

Sea el modelo LMSV estacionario transformado, dado por la ecuación:

$$x_t = \ln(y_t^2) = \mu + h_t + \xi_t$$

donde h_t es un proceso ARFIMA(p,d,q) de la forma,

$$(1-L)^d \phi(L) h_t = \theta(L)\eta_t$$

donde η_t *es* $NID(0,\sigma_\eta^2)$, $\phi(L)=1-\phi_1 L-\ldots-\phi_p L^p$ *y* $\theta(L)=1+\theta_1 L+\ldots+\theta_q L^q$ *son dos polinomios de retardos de orden p y q, respectivamente, cuyas raíces están fuera del círculo unidad, y el parámetro d está en el intervalo –1/2<d<1/2. Supongamos que el vector paramétrico* $\psi=(d,\sigma_\eta^2,\sigma_\xi^2,\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q)$ *es un elemento del espacio paramétrico compacto* Θ*, y supongamos que se verifica la siguiente condición de identificabilidad: si* $f(\lambda;\psi_1)=f(\lambda;\psi_2)$ *para todo* $\lambda \in [0,\pi]$, *entonces* $\psi_1=\psi_2$, *donde* $f(\lambda;\psi)$ *es la densidad espectral de* x_t.

Sea $\hat{\psi}_T$ *el valor que minimiza sobre* Θ *la aproximación a la función de verosimilitud gaussiana en el dominio de las frecuencias,* $\tilde{L}_W(\psi;\mathbf{x})$, *y sea* ψ_0 *el verdadero valor del parámetro. Entonces* $\hat{\psi}_T$ *converge casi seguro hacia* ψ_0

[6] Véase Box, Davies y Swann (1969, p. 42), Murray (1979, p. 42) y Harvey (1990, pp. 204-205), para una justificación de este criterio.

Obsérvese que este resultado es válido para el modelo LMSV general en el que h_t es un ARFIMA(p,d,q), y no sólo para el modelo LMSV básico en [3.2], que es un caso particular del anterior. Obsérvese también que Breidt, Crato y de Lima (1998) consideran el intervalo $-1/2<d<1/2$, donde el proceso h_t es estacionario ($d<1/2$) e invertible ($d>-1/2$), y en consecuencia también lo es x_t. Harvey (1998) define el modelo LMSV básico sobre el intervalo $0\leq d\leq 1$ y, por tanto, en este caso, el rango que define la estacionariedad es $0\leq d<1/2$. Este rango incluye el valor $d=0$ (ruido blanco) y los valores $0<d<1/2$, que son los que realmente producen memoria larga. En cualquier caso, el método de estimación PMV que acabamos de describir es igualmente válido tanto si se considera el intervalo $-1/2<d<1/2$ como $0\leq d<1/2$. La única diferencia de tipo práctico será el tipo de transformaciones utilizadas para recoger las distintas restricciones sobre el parámetro d.

3.2.2. Estimación del modelo LMSV no estacionario

El método de estimación *pseudo*-máximo verosímil que hemos descrito en la sección anterior sólo se puede aplicar, desde un punto de vista teórico, a procesos estacionarios, ya que requiere que la función de densidad espectral de x_t esté definida. Cuando d toma valores mayores o iguales que 1/2, el proceso h_t que define la ecuación de la volatilidad del modelo LMSV ya no es estacionario, y tampoco lo es la serie x_t.

Si consideramos el modelo LMSV básico tal como lo define Harvey (1998), el parámetro d varía en $0\leq d\leq 1$ y, por tanto, en el intervalo $1/2\leq d\leq 1$ el proceso h_t no es estacionario, y tampoco lo es x_t. Si $d=1$, el modelo LMSV se reduce al modelo SV con paseo aleatorio y puede estimarse por pseudo-máxima verosimilitud a través del filtro de Kalman, según se explica en Harvey, Ruiz y Shephard (1994) y Ruiz (1994b). Si $1/2\leq d<1$, h_t es un proceso fraccionalmente integrado no estacionario. Sin embargo, sus primeras diferencias constituyen un proceso estacionario de la forma, $(1-L)^{d^*}\Delta h_t = \eta_t$, con $d^*=d-1$ y $-1/2\leq d^*<0$. Por tanto, tomando primeras diferencias en el modelo

LMSV básico definido en [3.2] se consigue el siguiente modelo SV estacionario:

$$\Delta x_t = \Delta h_t + \Delta \xi_t \qquad [3.8a]$$

$$(1-L)^{d^*}\Delta h_t = \eta_t, \qquad \text{con } -1/2 \le d^* < 0 \qquad [3.8b]$$

donde Δ es el operador de primeras diferencias y d*=d-1, con $-1/2 \le d^* < 0$.

Trabajando con las observaciones del modelo en primeras diferencias, que ya es estacionario, podemos obtener los estimadores *pseudo*-máximo verosímiles de los parámetros (d^*, σ_η^2) según se explicó en la sección anterior, y simplemente haciendo $\hat{d} = \hat{d}^* + 1$, obtendremos el estimador del parámetro d. Precisamente esta forma de estimar d es la que proponen Smith, Sowell y Zin (1997) para modelos ARFIMA, pero utilizando el estimador máximo verosímil de d* en el dominio temporal.

La función que hay que minimizar para estimar el modelo [3.8] en primeras diferencias sigue siendo la función $\tilde{L}_W(\psi;\mathbf{x})$ definida en [3.4], pero ahora el vector de parámetros es $\psi=(d^*, \sigma_\eta^2)$, $I(\lambda_j)$ es el periodograma de la serie Δx_t y $f(\lambda_j;\psi)$ es la función de densidad espectral de dicha serie, que viene dada por la siguiente expresión:

$$f(\lambda_j;\psi) = \frac{\sigma_\eta^2}{2\pi}[2(1-\cos\lambda_j)]^{-d^*} + \frac{\sigma_\xi^2}{2\pi}[2(1-\cos\lambda_j)] \qquad [3.9]$$

con $\sigma_\xi^2 = \pi^2/2$. El algoritmo de Newton-Raphson se implementa como en el caso estacionario, con valores iniciales obtenidos de una búsqueda en dos dimensiones sobre el espacio paramétrico, y utilizando transformaciones para tratar las restricciones sobre los valores de los parámetros. El gradiente y el hessiano se calculan aplicando las fórmulas [3.5] y [3.6] a la serie en primeras diferencias, Δx_t, y utilizando la expresión de su densidad espectral en [3.9].

Aunque no existe aún ningún resultado teórico sobre cuáles son las propiedades de los estimadores obtenidos a partir del modelo en primeras diferencias, los resultados de las simulaciones que se presentan en la sección 3.3 son bastante

halagüeños. De forma heurística podría argumentarse que si el parámetro d* está en el intervalo $-1/2<d^*<1/2$, el estimador PMV de d* en el dominio de las frecuencias, $\hat{d}^*$, sería consistente en virtud del teorema 1 de Breidt, Crato y de Lima (1998) y, por tanto, también sería consistente el estimador de d obtenido como $\hat{d}=\hat{d}^*+1$, con $1/2<d<3/2$. Sin embargo, no está claro aún si el teorema de Breidt, Crato y de Lima (1998) puede aplicarse al modelo en primeras diferencias, ya que al tomar primeras diferencias hemos introducido cierta correlación en la perturbación del nuevo modelo. De hecho la covarianza de orden uno de la perturbación $\Delta\xi_t$ en [3.8a] es $Cov(\Delta\xi_t,\Delta\xi_{t-1}) = -\sigma_\xi^2 \neq 0$, y en consecuencia $\Delta\xi_t$ ya no es ruido blanco y la densidad espectral de Δx_t ya no es de la misma forma que la de x_t.

El problema de la estimación de modelos LMSV no estacionarios es sin duda uno de los temas que necesita más investigación. Pensemos que ni tan siquiera en los modelos ARFIMA para la media este problema está resuelto, aunque son varios los trabajos recientes que están aportando soluciones a este respecto; véase, por ejemplo, Velasco (1999a, 1999b) y Velasco y Robinson (2000). La mayor dificultad en la estimación de modelos con memoria larga no estacionarios estriba, probablemente, en encontrar un método de estimación válido para todo $0<d<1$, debido a la ruptura de las propiedades del proceso ARFIMA en d=1/2.

Una posible forma de unificar la estimación para todo el rango $0<d<1$ sería trabajar siempre con el modelo en primeras diferencias, independientemente de que el modelo original sea o no estacionario. Smith, Sowell y Zin (1997) han propuesto esta estrategia de diferenciar la serie, incluso innecesariamente, como forma de corregir el sesgo que tiene el estimador máximo verosímil en el dominio temporal para modelos ARFIMA. Sin embargo, no parece que ésta vaya a ser una estrategia muy adecuada para el estimador en el dominio de las frecuencias del modelo LMSV, a juzgar por los resultados de nuestras simulaciones que se recogen en la sección 3.3.2.

3.3. Propiedades de los estimadores PMV en muestras finitas

En esta sección se presentan los resultados de un extenso estudio de Monte Carlo para analizar las propiedades en muestras finitas de los estimadores PMV del modelo LMSV estacionario y del modelo no estacionario. En primer lugar, nuestro trabajo se centrará en el modelo LMSV básico en el que h_t es un ARFIMA(0,d,0), sin componentes autorregresivos y de medias móviles. La elección de este modelo, además del interés que pueda tener en sí mismo, nos permite estudiar en profundidad las propiedades de los estimadores de d y σ_η^2 para valores próximos a las fronteras de la no estacionariedad (d=1/2) y de la homocedasticidad (σ_η^2=0), que son los de mayor relevancia empírica.

Breidt, Crato y de Lima (1998) presentan resultados para modelos LMSV más generales donde h_t es un ARFIMA(p,d,q), con p=1, q=1 y también con d<0. Nosotras hemos realizado simulaciones de este tipo de modelos únicamente para el caso p=1, q=0 y d>0, que corresponde al modelo que hemos denominado ARLMSV, en el que h_t es un proceso ARFIMA(1,d,0). Este modelo es precisamente el que utilizan Henry y Payne (1997), Breidt, Crato y de Lima (1998) y Chan y Petris (2000) en las aplicaciones empíricas con datos reales. Al estudiar este modelo, lo que más nos interesa es analizar las propiedades de los estimadores PMV cuando los parámetros del modelo están próximos a la frontera de la no estacionariedad (d≈1/2 ó ϕ≈1).

3.3.1. Propiedades de los estimadores en el modelo LMSV estacionario

Para analizar las propiedades en muestras finitas de los estimadores PMV en el modelo LMSV básico estacionario, donde h_t es un ARFIMA(0,d,0) con 0<d<1/2, los valores del parámetro d que hemos analizado son d={0.2, 0.4, 0.45, 0.49}. La elección

de estos valores no es arbitraria: d=0.2 y d=0.4 son los valores analizados por Breidt, Crato y de Lima (1998), lo que nos permitirá comparar nuestros resultados con los suyos. Además, hemos incorporado valores de d próximos a la no estacionariedad, que representan modelos más persistentes y son los habituales en aplicaciones empíricas. Para cada valor del parámetro d se analizan tres valores de σ_η^2, $\sigma_\eta^2=\{1, 0.5, 0.1\}$. En la mayoría de los casos el valor $\sigma_\eta^2=1$ es excesivo (genera un modelo con excesiva volatilidad), pero hemos preferido incluirlo como referencia porque éste es el único valor de σ_η^2 que analizan Breidt, Crato y de Lima (1998). Consideramos siempre dos tamaños muestrales, T=1024 y T=4096, y para algunos modelos concretos se considera además un tamaño muestral mayor, T=8192, que nos permita analizar el comportamiento de los estimadores PMV en muestras muy grandes. Todos los modelos se simulan suponiendo normalidad para η_t y ε_t. Además, en la estimación se supone que el parámetro σ_ξ^2 es conocido, $\sigma_\xi^2=\pi^2/2$.

Para cada especificación $\{d, \sigma_\eta^2\}$ se generan 3000 muestras de tamaño T, y para cada una de ellas se obtienen las estimaciones PMV de los parámetros d y σ_η^2 según se explicó en la sección anterior. A partir de las estimaciones obtenidas en las 3000 réplicas, el sesgo y la desviación típica muestral de las estimaciones se calculan de la siguiente forma. Si denotamos por $\hat{d}_i$ la estimación de d para la i-ésima réplica, con i=1,2,…,3000, el valor medio estimado, el sesgo y la desviación típica se estimarán de la siguiente forma:

$$\bar{d} = \frac{1}{3000}\sum_{i=1}^{3000}\hat{d}_i \text{ , Sesgo} = \bar{d} - d_0, \text{ Desviación típica} = \sqrt{\frac{1}{3000}\sum_{i=1}^{3000}(\hat{d}_i - \bar{d})^2} \qquad [3.10]$$

respectivamente, donde d_0 denota el verdadero valor del parámetro d. La misma formulación se utilizará para σ_η^2.

En el modelo LMSV estacionario ($0<d<1/2$), el proceso h_t es un ARFIMA estacionario y se puede generar utilizando el método propuesto por Davies y Harte (1987). Al implementar este método, que se basa en transformadas de Fourier de las autocovarianzas, puede ahorrarse un tiempo considerable si se utiliza en los cálculos la transformada rápida de Fourier. Ésta es precisamente la razón por la que los tamaños muestrales analizados en este trabajo son potencias de dos.

Todas las simulaciones se han realizado utilizando GAUSS versión 3.2 sobre un Pentium 166 MHz. Las realizaciones del ruido blanco gaussiano ε_t, $t=1,...,T$, se generan con el comando RNDNS de GAUSS. Una vez obtenidas las realizaciones de los procesos ε_t y h_t, los valores de $\{y_1,...,y_T\}$ se calculan según la ecuación $y_t=\sigma_*\varepsilon_t\exp(\frac{1}{2}h_t)$, con $\sigma_*=1$. Haciendo $x_t=\ln(y_t^2)$ se consigue la serie $\{x_1,\ldots,x_T\}$, a partir de la cual se realiza la estimación según se ha explicado en la sección 3.2.

Para minimizar la función $\tilde{L}_W(\psi;\mathbf{x})$ en [3.4] utilizamos el procedimiento OPTMUM de GAUSS con el algoritmo NEWTON. Hemos probado también otros algoritmos como BFGS, por ejemplo, y no hemos encontrado diferencias importantes en los resultados, aunque sí en la velocidad de convergencia: NEWTON es mucho más rápido que otros. Para que la convergencia sea más rápida, se proporciona al programa las subrutinas para calcular analíticamente el gradiente y el hessiano, según las expresiones [3.5] y [3.6], respectivamente. En algunos casos hemos utilizado también derivadas numéricas, sin que hubiera ninguna diferencia en los resultados. Se ha probado también a utilizar en las iteraciones de Newton-Raphson otra matriz distinta del hessiano, como el producto exterior del gradiente que se utiliza en el algoritmo de Berndt, Hall, Hall y Hausman, pero tampoco se obtuvieron diferencias en los resultados.

La tabla 3.1 muestra, para los tres tamaños muestrales analizados, el sesgo y la desviación típica muestral (entre paréntesis) de las estimaciones de d y σ_η^2, calculados con las expresiones [3.10]. Los verdaderos valores de los parámetros aparecen en la casilla izquierda de la fila correspondiente.

Tabla 3.1. Sesgo y desviación típica (entre paréntesis) muestral de los estimadores PMV de d y σ_η^2 en el dominio de las frecuencias para el modelo LMSV estacionario y no estacionario. 3000 réplicas.

Parámetros	**d**			$\boldsymbol{\sigma_\eta^2}$		
Modelo LMSV estacionario	**T=1024**	**T=4096**	**T=8192**	**T=1024**	**T=4096**	**T=8192**
$d=0.2, \sigma_\eta^2=1$	-0.006 (0.096)	-0.005 (0.043)		0.020 (0.413)	0.030 (0.209)	
$d=0.2, \sigma_\eta^2=0.5$	-0.002 (0.146)	-0.009 (0.073)		0.028 (0.361)	0.027 (0.191)	
$d=0.2, \sigma_\eta^2=0.1$	-0.026 (0.171)	-0.027 (0.157)	0.000 (0.148)	0.121 (0.265)	0.056 (0.149)	0.036 (0.115)
$d=0.4, \sigma_\eta^2=1$	-0.003 (0.070)	-0.001 (0.037)		0.048 (0.391)	0.028 (0.201)	
$d=0.4, \sigma_\eta^2=0.5$	-0.009 (0.094)	-0.002 (0.052)	-0.001 (0.037)	0.069 (0.330)	0.030 (0.171)	0.021 (0.124)
$d=0.4, \sigma_\eta^2=0.1$	-0.100 (0.178)	-0.026 (0.111)	-0.011 (0.081)	0.137 (0.265)	0.061 (0.137)	0.035 (0.095)
$d=0.45, \sigma_\eta^2=1$	-0.010 (0.058)	-0.002 (0.034)		0.068 (0.360)	0.029 (0.191)	
$d=0.45, \sigma_\eta^2=0.5$	-0.022 (0.076)	-0.005 (0.045)	-0.002 (0.033)	0.098 (0.303)	0.038 (0.160)	0.022 (0.115)
$d=0.45, \sigma_\eta^2=0.1$	-0.101 (0.161)	-0.032 (0.086)	-0.017 (0.063)	0.141 (0.258)	0.060 (0.124)	0.034 (0.083)
$d=0.49, \sigma_\eta^2=1$	-0.024 (0.045)	-0.010 (0.024)		0.112 (0.327)	0.057 (0.165)	
$d=0.49, \sigma_\eta^2=0.5$	-0.036 (0.060)	-0.016 (0.033)		0.131 (0.278)	0.062 (0.136)	
$d=0.49, \sigma_\eta^2=0.1$	-0.105 (0.141)	-0.043 (0.068)	-0.027 (0.048)	0.149 (0.250)	0.064 (0.111)	0.039 (0.071)
Modelo LMSV no estacionario	**T=1024**	**T=4096**	**T=8192**	**T=1024**	**T=4096**	**T=8192**
$d=0.75, \sigma_\eta^2=1$	0.009 (0.072)	0.001 (0.035)		0.028 (0.346)	0.016 (0.172)	
$d=0.75, \sigma_\eta^2=0.5$	0.026 (0.096)	0.004 (0.045)		0.040 (0.273)	0.015 (0.128)	
$d=0.75, \sigma_\eta^2=0.1$	0.044 (0.152)	0.020 (0.081)	0.015 (0.056)	0.053 (0.156)	0.012 (0.064)	0.004 (0.040)

Comentaremos en primer lugar la **estimación del parámetro d** en modelos estacionarios (0<d<1/2). Los aspectos más relevantes de la estimación de este parámetro son los siguientes:

- El sesgo de d siempre es negativo: el método PMV en el dominio de las frecuencias siempre subestima el parámetro d. Cheung y Diebold (1994) obtienen este mismo resultado en la estimación de modelos ARFIMA para la media y Breidt, Crato y de Lima (1998) y Wright (1999) en el propio modelo LMSV.
- Al aumentar T se reducen notablemente el sesgo y la varianza (consistencia). Por ejemplo, si T=8192 el sesgo es diez veces menor que con T=1024, y la varianza se reduce más de la mitad. De hecho se comprueba que la tasa de convergencia de la desviación estándar es $\sqrt{T}$.
- Fijados T y σ_η^2, el sesgo parece que aumenta ligeramente con el valor de d, pero la varianza disminuye claramente al aumentar d. Estos resultados concuerdan nuevamente con los de Cheung y Diebold (1994) para modelos ARFIMA.
- Para un valor fijo de d, tanto el sesgo como la varianza del estimador PMV de d aumentan al disminuir el valor de σ_η^2 (menos volatilidad). Los peores resultados corresponden siempre a σ_η^2=0.1.

Este último resultado es similar a los resultados que obtienen otros autores para el modelo ARSV(1). Por ejemplo, Ruiz (1994b) y Sandman y Koopman (1998) comprueban que el estimador PMV de los parámetros del modelo ARSV(1) se deteriora (aumenta su sesgo y su varianza) a medida que disminuye la volatilidad, pero además observan que esto ocurre no sólo con el estimador PMV, sino con otros estimadores propuestos para el modelo ARSV(1). Esto puede hacernos pensar que más que un problema del método de estimación, se trata de un problema del modelo. En este sentido, no debemos olvidar que en cualquiera de los modelos SV, la serie $x_t=\ln(y_t^2)$ es igual a un proceso lineal h_t con perturbación η_t, "contaminado" por un ruido blanco ξ_t que suponemos tiene varianza conocida $\sigma_\xi^2 = \pi^2/2 \approx 4.93$. Cuando la varianza de η_t es muy pequeña, la varianza del proceso h_t es también pequeña y mucho

menor que la varianza de ξ_t y, por tanto, el "ruido" ξ_t contamina el modelo de tal forma que no nos permite determinar con precisión la estructura dinámica del proceso h_t (la "señal"). De hecho, si observamos la expresión de la función de autocorrelación de x_t en [1.23] comprobamos que cuando σ_ξ^2 es mucho mayor que σ_h^2, el cociente de las varianzas, σ_ξ^2/σ_h^2, puede ser de tal magnitud que la función de autocorrelación de x_t sea prácticamente igual a cero, independientemente de cuál sea el valor del otro parámetro del modelo (ϕ en el modelo ARSV(1) ó d en el modelo LMSV), y esto puede acarrear un problema de identificación.

Recordemos que el teorema de Breidt, Crato y de Lima (1998) demuestra la consistencia de los estimadores PMV del modelo LMSV bajo una condición de identificabilidad que garantice que distintos valores de los parámetros proporcionan distintos valores de la densidad espectral (o equivalentemente de la función de autocorrelación). Si en el modelo LMSV se tiene $\sigma_\eta^2 \approx 0$, la función de densidad espectral del proceso h_t es prácticamente igual a cero para distintos valores de d y puede ocurrir que para algún $\psi_1 \neq \psi_2$ resulte que $f(\lambda;\psi_1) \cong f(\lambda;\psi_2)$, siendo $\psi=(d,\sigma_\eta^2)$. Este problema de identificación puede ser importante en aplicaciones empíricas donde lo habitual es obtener este tipo de valores $\sigma_\eta^2 \approx 0$.

Por lo que respecta a la **estimación de σ_η^2** en modelos LMSV estacionarios (d<1/2), las conclusiones que podemos deducir de los resultados de la tabla 3.1 son, en ciertos aspectos, similares a las que hemos comentado para la estimación de d. A continuación se resumen las más importantes:

- El sesgo y la varianza de $\hat{\sigma}_\eta^2$ se reducen considerablemente al aumentar T, y se comprueba que la tasa de convergencia de la desviación estándar es $\sqrt{T}$.
- Con T=8192, los estimadores PMV son bastante adecuados, con poco sesgo y poca dispersión.
- Para σ_η^2 y T fijos, el sesgo aumenta ligeramente con d pero su varianza disminuye.

- Para d fijo, el sesgo de $\hat{\sigma}_\eta^2$ siempre aumenta al disminuir σ_η^2, pero su varianza disminuye.

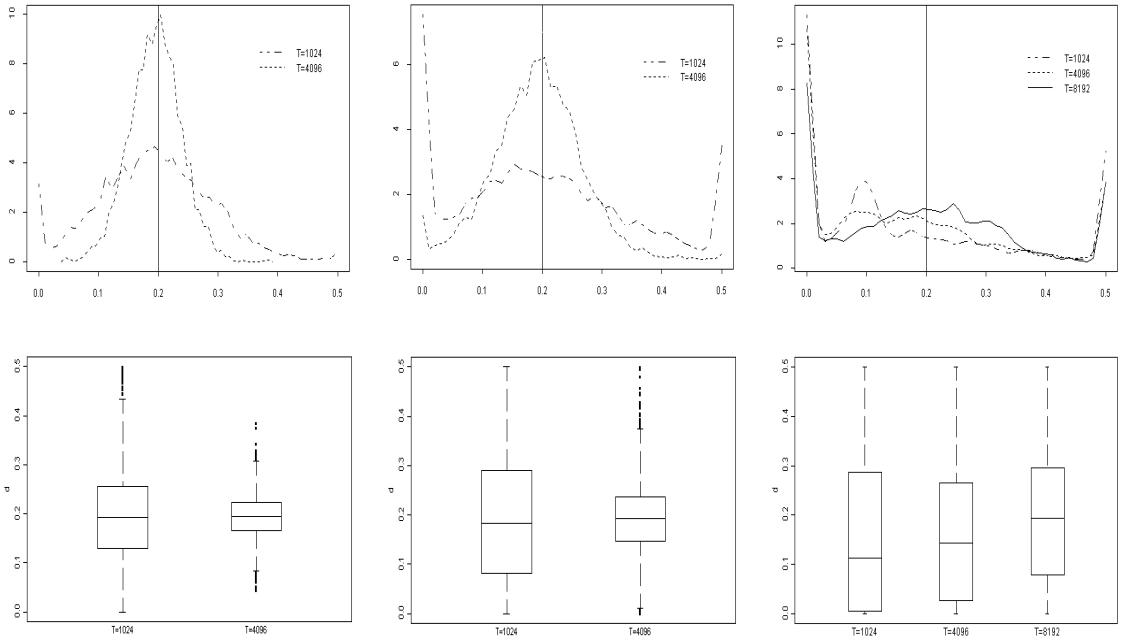
Este último resultado debe interpretarse con cautela, ya que, como señalan Jacquier, Polson y Rossi (1994), la reducción en la dispersión de $\hat{\sigma}_\eta^2$ puede producirse, y de hecho así es, porque muchas muestras proporcionan el mínimo en la frontera inferior del espacio paramétrico, $\sigma_\eta^2=0$, y esto hace que el estimador parezca favorable en términos de varianza. Breidt y Carriquiry (1996) observan también que los métodos de estimación pseudo-máximo verosímil proporcionan estimaciones de σ_η^2 muy sesgadas siempre que el verdadero valor de este parámetro sea muy pequeño y la serie x_t sea prácticamente indistinguible de un ruido blanco, lo cual corrobora nuestros resultados.

Para completar los resultados del experimento de Monte Carlo en el modelo LMSV estacionario, el gráfico 3.1 representa la distribución muestral de las estimaciones de d (paneles superiores) y σ_η^2 (paneles inferiores) obtenidas en las 3000 réplicas, para d=0.2 y para cada uno de los tres valores de σ_η^2 analizados, $\sigma_\eta^2=1$ (izquierda), $\sigma_\eta^2=0.5$ (centro) y $\sigma_\eta^2=0.1$ (derecha). El gráfico 3.2 y el gráfico 3.3 representan las mismas distribuciones para d=0.4 y d=0.45, respectivamente. Los resultados para d=0.49 no se presentan porque son muy similares a los que se obtienen con d=0.45 y no aportan nada nuevo. En cada caso se representa la distribución para los dos tamaños muestrales analizados, T=1024 y T=4096, y en aquellos modelos para los que hemos repetido las simulaciones con T=8192, se representan las distribuciones de los estimadores para los tres tamaños muestrales. En la primera fila de cada panel se representa la estimación semiparamétrica de la densidad[7] del estimador correspondiente, con una línea vertical sobre el verdadero valor del parámetro y en la segunda fila se representa el diagrama de caja para cada tamaño muestral.

[7] La estimación semiparamétrica se realiza con el comando `density` de S-PLUS utilizando la anchura de banda que sugiere Silverman (1990, sección 3.4.2). Se han probado diferentes Kernel sin que hubiera diferencias sustanciales en los resultados, por lo que finalmente hemos utilizado la opción por defecto de S-PLUS.

Gráfico 3.1. Densidad estimada y diagrama de caja de los estimadores PMV de d y σ_η^2 en el dominio de las frecuencias para el modelo LMSV con d=0.2 y σ_η^2 ={1 (izquierda), 0.5 (centro), 0.1 (derecha)}. 3000 réplicas.

3.1(a) Estimación de d



3.1(b) Estimación de σ_η^2

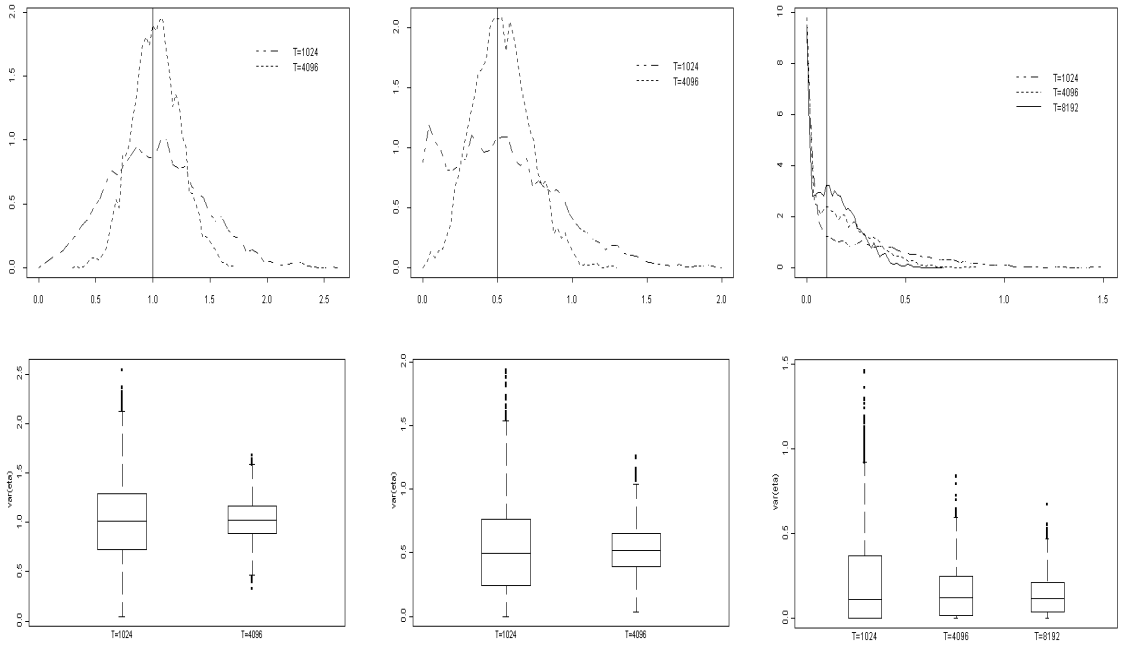
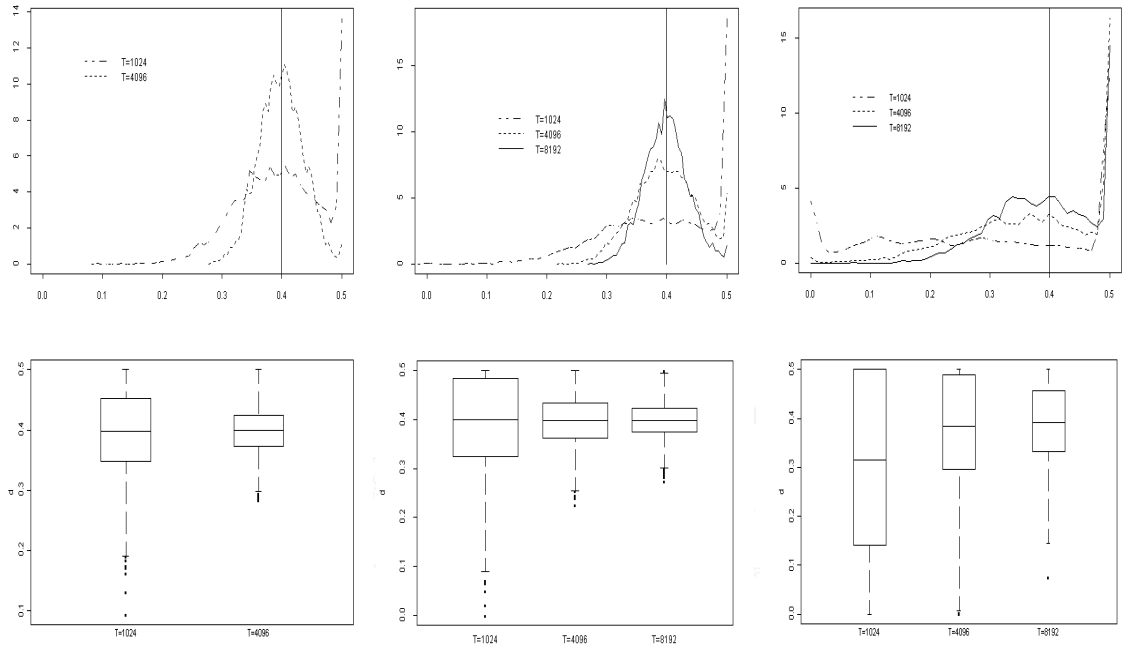


Gráfico 3.2. Densidad estimada y diagrama de caja de los estimadores PMV de d y σ_η^2 en el dominio de las frecuencias para el modelo LMSV con d=0.4 y σ_η^2 ={1 (izquierda), 0.5 (centro), 0.1 (derecha)}. 3000 réplicas.

3.2(a) Estimación de d



3.2(b) Estimación de σ_η^2

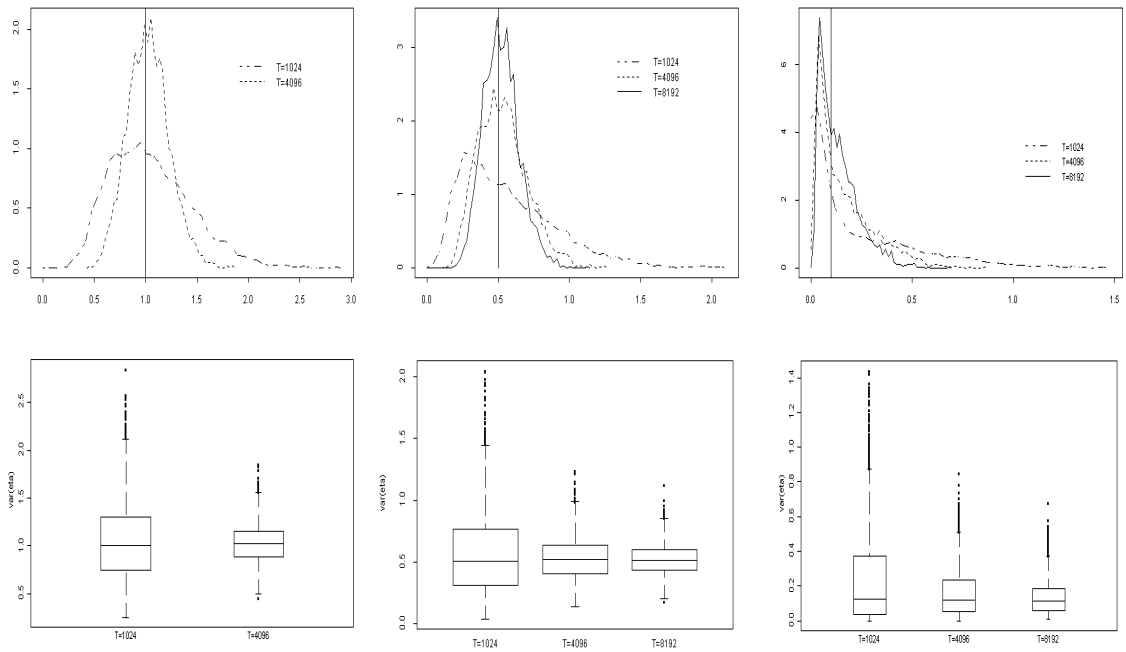
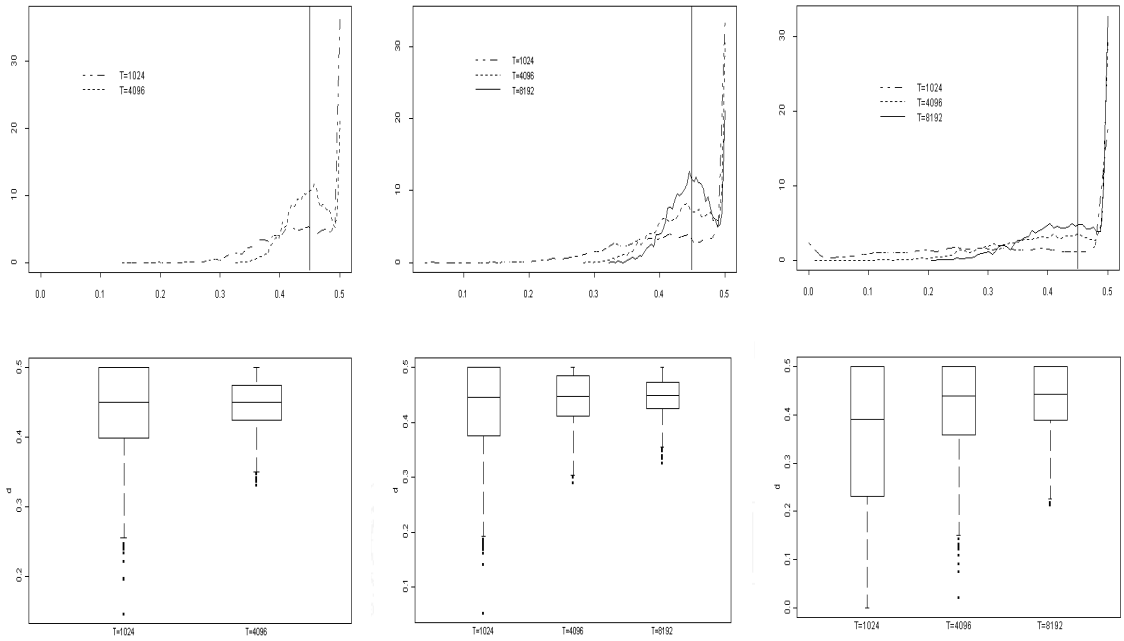
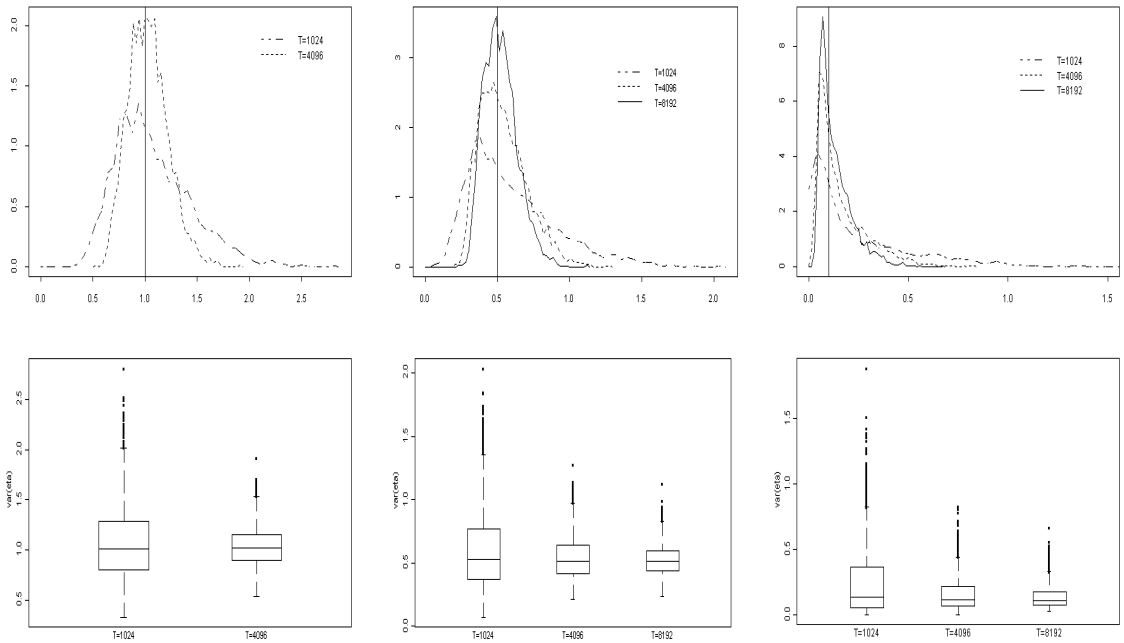


Gráfico 3.3. Densidad estimada y diagrama de caja de los estimadores PMV de d y σ_η^2 en el dominio de las frecuencias para el modelo LMSV con d=0.45 y σ_η^2 ={1 (izquierda), 0.5 (centro), 0.1 (derecha)}. 3000 réplicas.

3.3(a) Estimación de d



3.3(b) Estimación de σ_η^2



Los gráficos 3.1 - 3.3 confirman los resultados anteriores y además nos permiten descubrir algunas características que no se percibían en los resultados de la tabla 3.1 y que pueden tener gran interés en aplicaciones empíricas.

Por ejemplo, descubrimos que en los modelos más persistentes (d≈1/2) muchas estimaciones del parámetro d se acumulan en la frontera de la no estacionariedad, d=1/2, algo que no se había observado en el trabajo de Breidt, Crato y de Lima (1998) debido a que estos autores estiman el modelo sin restricciones. Esta concentración en torno al valor 1/2 significa que existe una gran probabilidad de estimar $\hat{d}=1/2$ (modelo no estacionario) cuando el proceso generador de los datos es realmente un proceso estacionario. En los modelos menos persistentes (d=0.2, d=0.4) y con mayor volatilidad ($\sigma_\eta^2=1, \sigma_\eta^2=0.5$) este efecto se amortigua al aumentar el tamaño muestral. Por ejemplo, en el gráfico 3.2(a) vemos que para el modelo $\{d=0.4, \sigma_\eta^2=0.5\}$ la estimación mejora considerablemente con T=8192. Sin embargo, en los modelos más persistentes (d=0.45) y/o con menor volatilidad ($\sigma_\eta^2=0.1$) sigue habiendo una concentración de masa en la frontera de la no estacionariedad, incluso con T=8192, como puede verse en el gráfico 3.3(a). En estos casos sería necesario un tamaño muestral enorme para poder estimar con fiabilidad el parámetro d.

Con respecto a la estimación de σ_η^2, en los paneles inferiores de los gráficos 3.1-3.3 observamos, en primer lugar, que el tamaño muestral corrige la asimetría de la distribución de $\hat{\sigma}_\eta^2$, y que para los valores mayores de σ_η^2, el estimador PMV tiene propiedades razonables cuando T=4096. Sin embargo, para los modelos con poca volatilidad ($\sigma_\eta^2=0.1$), la distribución muestral de $\hat{\sigma}_\eta^2$ está sesgada negativamente hacia el cero de tal forma que en estos casos estimaremos $\hat{\sigma}_\eta^2 \approx 0$ con mucha probabilidad.

Aunque no hemos sido capaces de obtener analíticamente la distribución asintótica de los estimadores PMV, los gráficos parecen indicar una distribución normal. Además, los resultados de la tabla 3.1 sugieren que la convergencia puede ser $\sqrt{T}$.

Para tener una visión más de la precisión de los estimadores PMV en los modelos analizados anteriormente, los gráficos 3.4 - 3.6 representan las estimaciones de (d, σ_η^2) obtenidas en las 3000 réplicas de los experimentos de Monte Carlo, para d=0.2, d=0.4 y d=0.45, respectivamente. En cada gráfico se presentan los resultados de los dos tamaños muestrales, T=1024 (panel superior) y T=4096 (panel inferior). Las líneas que cruzan representan los verdaderos valores de los parámetros.

En estos gráficos observamos un fuerte grado de correlación negativa entre los estimadores de d y σ_η^2, similar al que observa Shephard (1996) entre los estimadores de ϕ y σ_η^2 en el modelo ARSV(1). Además en los gráficos correspondientes a σ_η^2=0.1 (derecha), se aprecia claramente el fenómeno al que aludíamos anteriormente: muchas muestras estiman $\hat{\sigma}_\eta^2$=0 (obsérvese cuántos puntos están sobre el eje de abscisas) y esto puede producir una reducción engañosa en la dispersión del estimador $\hat{\sigma}_\eta^2$, aunque este efecto se reduce notablemente al aumentar T.

Comparando unos gráficos con otros se observa que el tamaño muestral siempre aumenta la precisión de los estimadores PMV y que éstos son más eficientes cuanto mayor es d, aunque su comportamiento empeora a medida que disminuye σ_η^2, lo cual confirma todos los resultados anteriores.

Nuestras conclusiones sobre las propiedades de los estimadores PMV en el dominio de las frecuencias parecen contradecir los resultados de Breidt, Crato y de Lima (1998), pero realmente no es así. Lo que ocurre es que estos autores analizan las propiedades de $\hat{d}$ en el modelo LMSV básico con d={0.2, 0.4} y σ_η^2=1, y en estos casos, el estimador PMV de d y σ_η^2 tiene buenas propiedades, como puede verse en la columna izquierda de los gráficos 3.1(a) y 3.2(a).

Gráfico 3.4. Estimaciones PMV de (d, σ_η^2) en el modelo LMSV con d=0.2. En el eje de abscisas se representa d y en el de ordenadas σ_η^2. Las líneas que cruzan indican los verdaderos valores de los parámetros. Los gráficos del panel superior corresponden a T=1024 y los del panel inferior a T=4096.

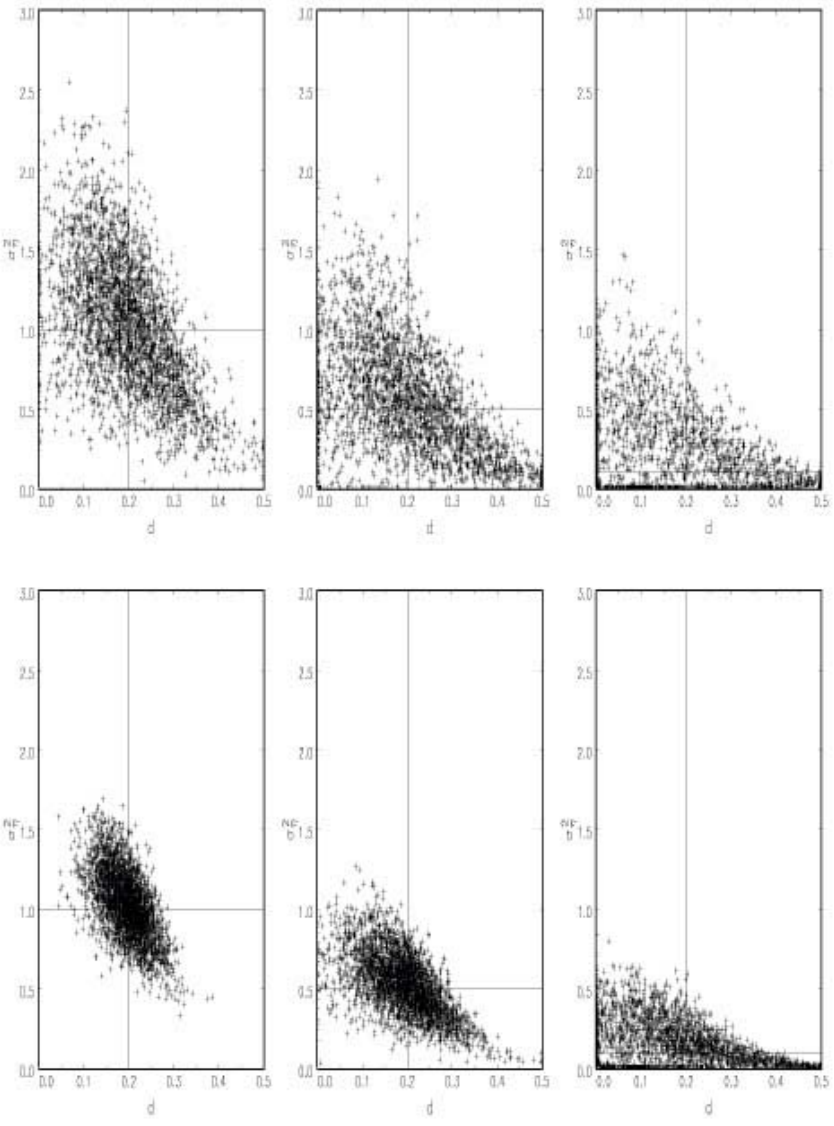


Gráfico 3.5. Estimaciones PMV de (d, σ_η^2) en el modelo LMSV con d=0.4. En el eje de abscisas se representa d y en el de ordenadas σ_η^2. Las líneas que cruzan indican los verdaderos valores de los parámetros. Los gráficos del panel superior corresponden a T=1024 y los del panel inferior a T=4096.

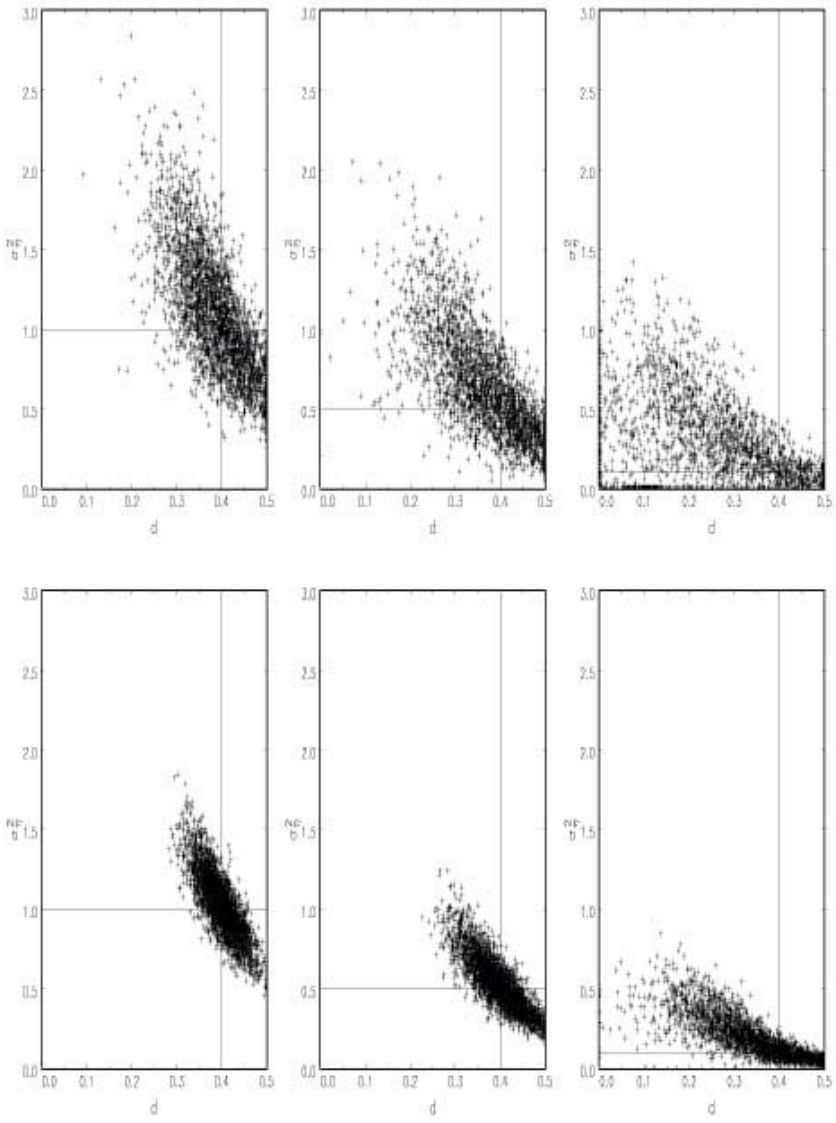
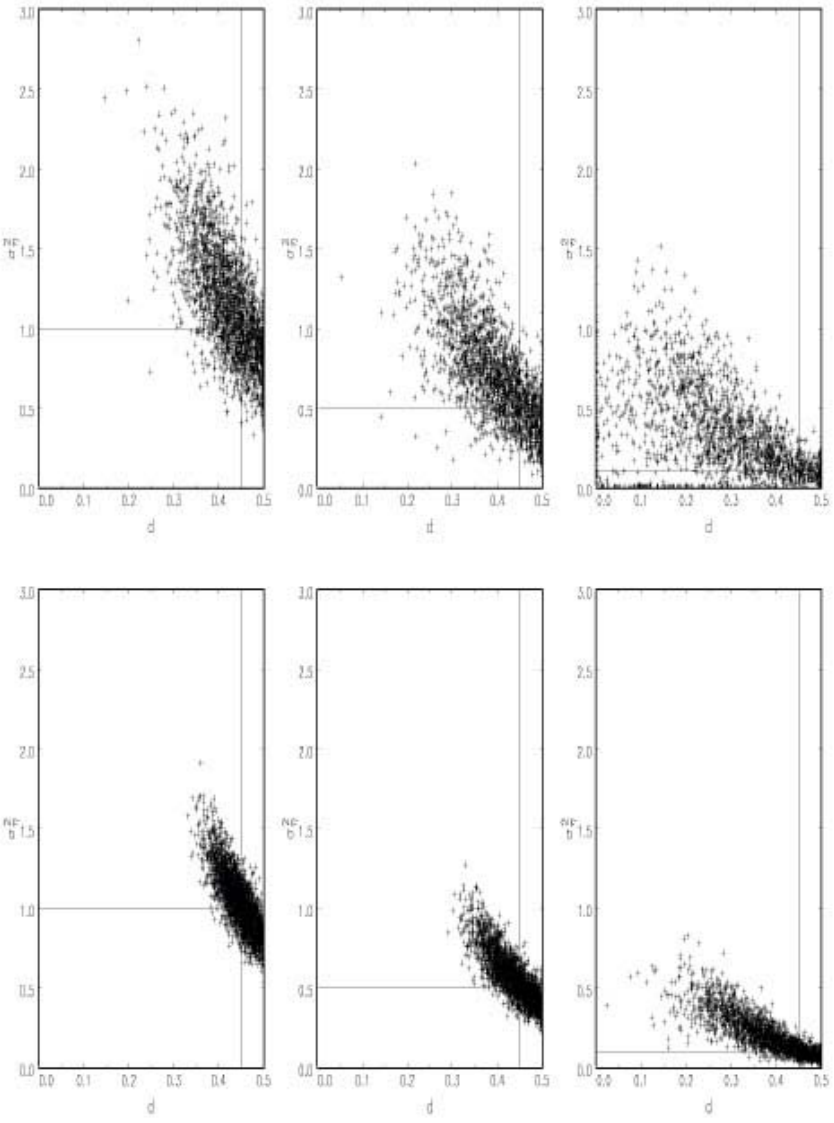


Gráfico 3.6. Estimaciones PMV de (d, σ_η^2) en el modelo LMSV con d=0.45. En el eje de abscisas se representa d y en el de ordenadas σ_η^2. Las líneas que cruzan indican los verdaderos valores de los parámetros. Los gráficos del panel superior corresponden a T=1024 y los del panel inferior a T=4096.



3.3.2. Propiedades de los estimadores en el modelo LMSV no estacionario

BIBLIOTECA VIRTUAL

Analizaremos a continuación las propiedades de los estimadores PMV en el dominio de las frecuencias para el modelo LMSV no estacionario ($1/2 \leq d < 1$), en el que la estimación se realiza sobre el modelo en primeras diferencias, según se explicó en la sección 3.2. Para ello hemos generado series con $d=0.75$ y $\sigma_\eta^2 = \{1, 0.5, 0.1\}$ y tamaños muestrales $T=1024$ y $T=4096$. Para el modelo con $\sigma_\eta^2 = 0.1$ se considera además $T=8192$. Al igual que en el caso estacionario, para cada especificación $\{d, \sigma_\eta^2\}$ se generan 3000 muestras de tamaño T, y se calcula el sesgo y la desviación típica muestral de las estimaciones sobre las 3000 réplicas utilizando las fórmulas [3.10].

En este caso, el algoritmo de Davies y Harte (1987) no puede aplicarse directamente para simular el proceso h_t, porque h_t ahora no es estacionario y el algoritmo requiere que el proceso que se simula lo sea. Este problema se puede solucionar de varias formas. Una posibilidad es truncar el polinomio de la representación AR(∞) del proceso h_t en un retardo P suficientemente grande y simular el proceso AR(P) resultante. Éste es el procedimiento que utilizan, por ejemplo, Baillie, Bollerslev y Mikkelsen (1996) truncando en P=1000. Otra posibilidad es simular con el algoritmo de Davies y Harte (1987) el proceso en primeras diferencias, $\Delta h_t = (1-L)^{-d^*}\eta_t$, con $d^* = d-1$ y $-1/2 \leq d^* < 0$, que sí es estacionario, y reconstruir después el

proceso h_t haciendo sumas de forma recursiva. Ambas soluciones se han ensayado sin que hayamos encontrado grandes diferencias en los resultados.

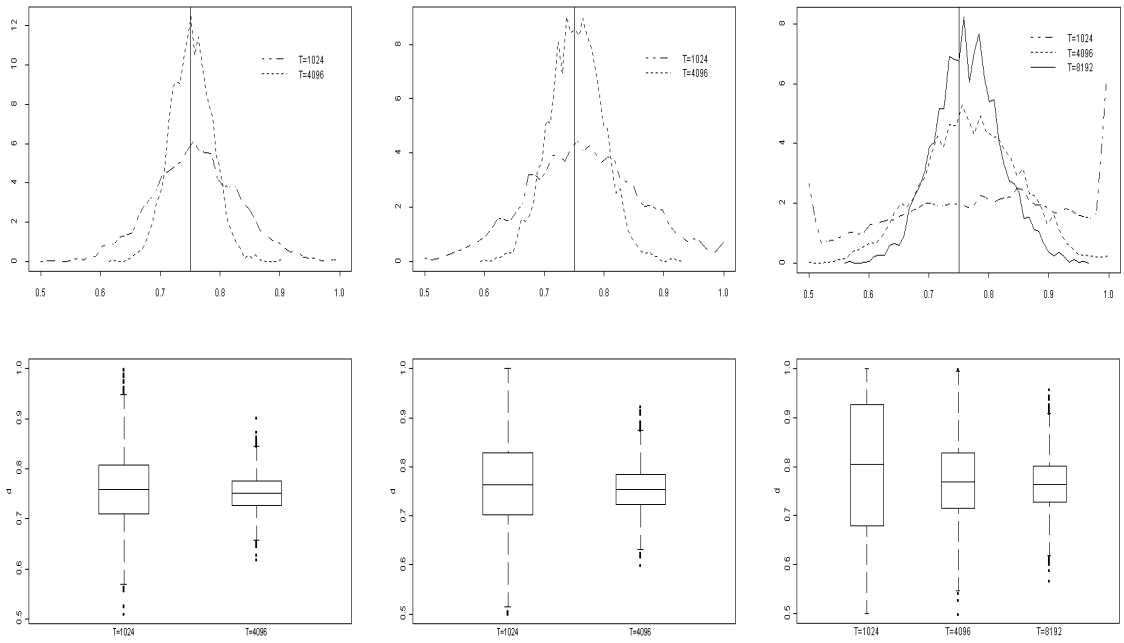
Los resultados de las simulaciones del modelo no estacionario, se recogen en las últimas filas de la tabla 3.1 presentada anteriormente, y en los gráficos 3.7 y 3.8, que se presentan a continuación. El gráfico 3.7 representa la distribución muestral de las estimaciones de d y σ_η^2 para d=0.75 y para $\sigma_\eta^2=\{1, 0.5, 0.1\}$. El gráfico 3.8 representa las estimaciones de (d, σ_η^2) obtenidas en las 3000 réplicas para T=1024 y T=4096.

A primera vista, el rasgo más notable es que la estimación PMV del modelo no estacionario parece tener mejores propiedades que en el caso estacionario. En concreto, de los resultados de la tabla 3.1 deducimos que:

- El estimador de d tiene siempre un sesgo pequeño pero positivo, a diferencia de lo que ocurría en el caso estacionario, donde el método PMV siempre subestima d.
- Para un valor σ_η^2 fijo, el sesgo y la desviación típica de $\hat{\sigma}_\eta^2$ son, en general, menores que en cualquiera de los modelos estacionarios con el mismo valor de σ_η^2.
- Nuevamente la estimación mejora siempre con el tamaño muestral, y la reducción de la desviación típica se produce a un ritmo $\sqrt{T}$.
- Al igual que en los modelos estacionarios, cuanto menor es σ_η^2 peor se estima d, pero a cambio se reduce la varianza de $\hat{\sigma}_\eta^2$. Como comentamos anteriormente este efecto puede ser debido a una acumulación de estimaciones en $\sigma_\eta^2=0$.

Gráfico 3.7. Densidad estimada y diagrama de caja de los estimadores PMV de d y σ_η^2 en el dominio de las frecuencias para el modelo LMSV con d=0.75 y σ_η^2 ={1 (izquierda), 0.5 (centro), 0.1 (derecha)}. 3000 réplicas.

3.7(a) Estimación de d



3.7(b) Estimación de σ_η^2

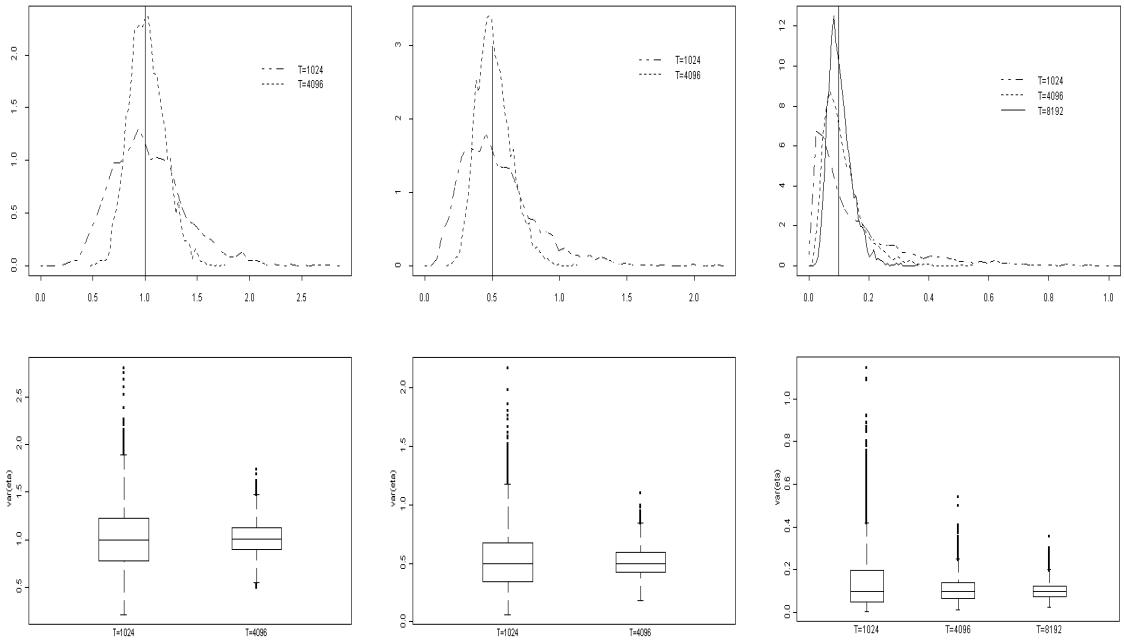
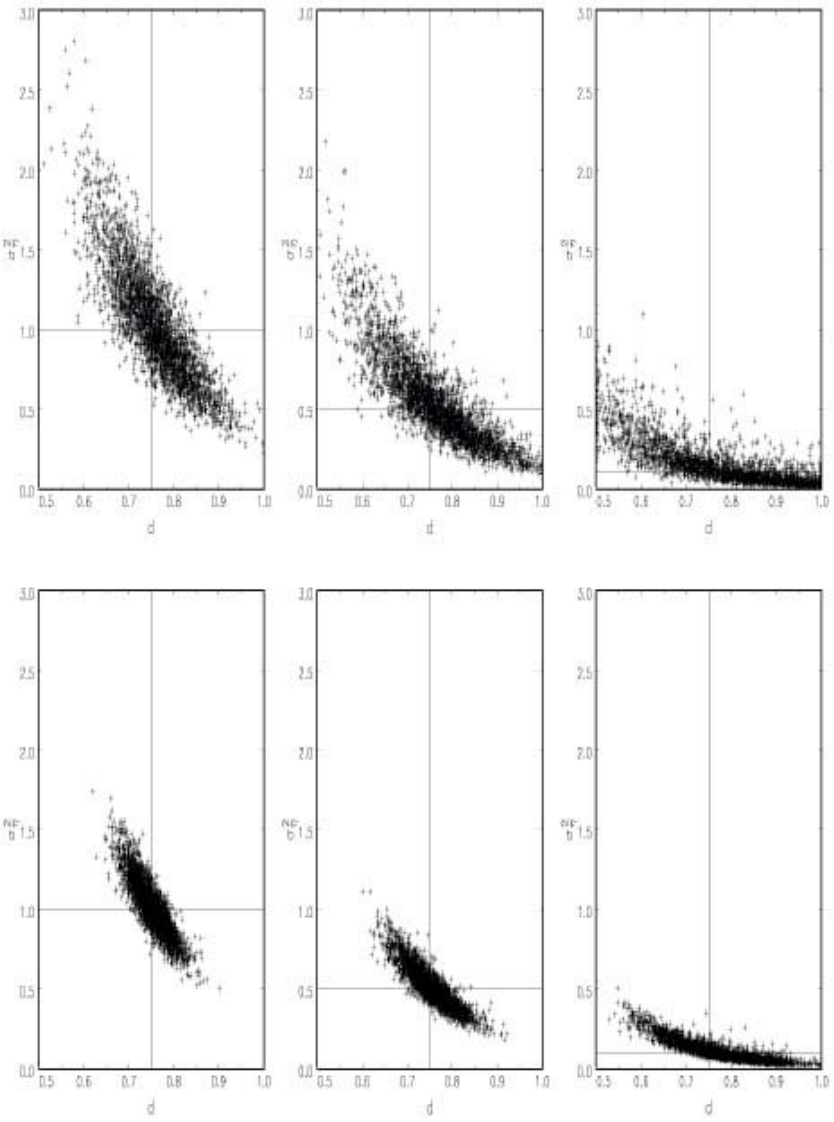


Gráfico 3.8. Estimaciones PMV de (d, σ_η^2) en el modelo LMSV con d=0.75. En el eje de abscisas se representa d y en el de ordenadas σ_η^2. Las líneas que cruzan indican los verdaderos valores de los parámetros. Los gráficos del panel superior corresponden a T=1024 y los del panel inferior a T=4096.



BIBLIOTECA VIRTUAL

Lo que más llama la atención en el gráfico 3.7 es que el parámetro d se estima mucho mejor ahora que en el caso estacionario: la densidad del estimador PMV de d y σ_η^2 en modelos con $\sigma_\eta^2=1$ ó $\sigma_\eta^2=0.5$ parece aproximarse bien a la densidad "gaussiana", incluso cuando T=1024, y los problemas de la frontera que aparecen en el modelo con $\sigma_\eta^2=0.1$ se corrigen notablemente aumentando el tamaño muestral a T=4096. Estas características también pueden apreciarse en el gráfico 3.8.

Los buenos resultados de la estimación del modelo no estacionario podrían hacernos pensar en estimar siempre el modelo LMSV en primeras diferencias como estrategia para mejorar los resultados obtenidos en el modelo estacionario. Sin embargo, los resultados de nuestras simulaciones apuntan en sentido contrario. Concretamente, hemos simulado 3000 series de un modelo LMSV básico estacionario con $\{d=0.4, \sigma_\eta^2=0.5\}$ y dos tamaños muestrales, T=1024 y T=4096, y hemos estimado después el modelo en primeras diferencias, obteniendo los siguientes resultados. El sesgo medio y la desviación típica muestral (entre paréntesis) para $\hat{d}$ son 0.170 (0.188) y 0.098 (0.110) cuando T=1024 y T=4096, respectivamente, y el sesgo medio y la desviación típica muestral para $\hat{\sigma}_\eta^2$ son –0.070 (0.327) y –0.083 (0.182) cuando

T=1024 y T=4096, respectivamente. Comparando estos resultados con los que obtuvimos estimando el mismo modelo sin diferenciar (tabla 3.1), se observa que la estimación es ahora mucho peor: el sesgo de $\hat{d}$ ahora es claramente mayor y positivo, y también se ha incrementado su varianza.

Obsérvese que al tomar primeras diferencias en el modelo LMSV estacionario, el modelo se está sobrediferenciando innecesariamente sobre el rango $0<d<1/2$, donde la estacionariedad ya está asegurada. Esto puede resultar arriesgado porque el modelo en primeras diferencias que se obtiene en este caso ya no es invertible. El parámetro $d^*=d-1$ toma valores en el intervalo $-1<d^*<-1/2$ y, por tanto, sobrepasa la frontera de la invertibilidad, donde parece que el estimador PMV en el dominio de las frecuencias tiene muy malas propiedades.

En el ámbito de los procesos ARFIMA, este comportamiento de los estimadores PMV ya era conocido. Smith, Sowell y Zin (1997) comprueban que el estimador PMV de d en el dominio de las frecuencias para modelos ARFIMA no invertibles presenta sistemáticamente un sesgo muy grande y positivo, y la estimación resulta poco fiable. Estas características coinciden precisamente con las que nosotras hemos observado en la estimación de d^* en el modelo LMSV, y de ahí que la estimación del parámetro d, con $0<d<1/2$, obtenida a partir de d^*, tampoco sea muy fiable. Esto nos hace pensar que estimar siempre el modelo LMSV en primeras diferencias no va a solucionar los problemas del estimador PMV en el dominio de las frecuencias, aunque sin duda este tema requiere mayor investigación.

BIBLIOTECA VIRTUAL
MIGUEL DE CERVANTES

3.3.3. Propiedades de los estimadores en el modelo ARLMSV

Se presentan a continuación algunos resultados sobre la estimación del modelo SV autorregresivo y con memoria larga (ARLMSV). Recordemos que en este modelo el proceso h_t que define la volatilidad es un ARFIMA(1,d,0) y, por tanto, el modelo transformado admite una representación lineal para la serie $\ln(y_t^2)$ de la forma:

$$x_t = \ln(y_t^2) = \mu + h_t + \xi_t \qquad [3.11a]$$

$$(1-\phi L)(1-L)^d h_t = \eta_t \qquad [3.11b]$$

Este modelo, que ha sido ajustado a series financieras reales por Henry y Payne (1997), Breidt, Crato y de Lima (1998) y Chan y Petris (2000), presenta un problema de identificación cuando la ecuación que define la volatilidad tiene una raíz unitaria. Así, el modelo SV integrado, en el que h_t es un paseo aleatorio, puede obtenerse a partir del modelo [3.11] tomando $\{\phi=1,d=0\}$ ó $\{\phi=0,d=1\}$.

Nuestro principal objetivo al analizar el modelo ARLMSV es precisamente descubrir qué problemas de estimación pueden surgir debido a esta falta de identificación de los parámetros del modelo, cuando el proceso que define la volatilidad está cerca de la frontera de la no estacionariedad.

Para ilustrar el efecto que este problema pueda tener sobre las propiedades en muestras finitas de los estimadores PMV en el dominio de las frecuencias, hemos simulado modelos ARLMSV próximos a la no estacionariedad en los dos sentidos: bien tomando ϕ muy próximo a 1, o tomando d casi igual a 1/2. En concreto, hemos simulado series con $\phi=\{0.8, 0.9, 0.99\}$, d=0.2 y $\sigma_\eta^2=0.1$, y series con $\phi=0.2$, d=$\{0.4, 0.45, 0.49\}$ y $\sigma_\eta^2=0.1$. Con este diseño es posible comprobar si el estimador PMV puede discernir cuál es la causa de la no estacionariedad en un modelo ARLMSV.

Como en casos anteriores, se consideran tres tamaños muestrales, T=1024, T=4096 y T=8192, y para cada especificación $\{\phi,d,\sigma_\eta^2,T\}$ se realizan 3000 réplicas.

Los resultados del experimento de Monte Carlo se resumen en la tabla 3.2 y en el gráfico 3.9. La tabla 3.2 recoge el sesgo y la desviación típica muestral de las estimaciones para cada uno de los modelos analizados. El gráfico 3.9 representa la estimación semiparamétrica de la densidad de las estimaciones de ϕ (panel izquierdo), d (panel central) y σ_η^2 (panel derecho) obtenidas en las 3000 réplicas, para cuatro modelos seleccionados y para los tres tamaños muestrales analizados. La línea vertical representa el verdadero valor del parámetro. Los casos seleccionados son representativos de todos los demás modelos.

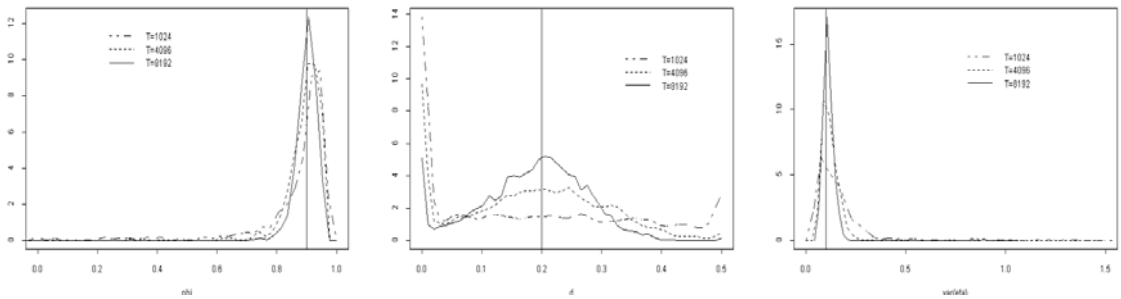
BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Tabla 3.2. Sesgo y desviación típica (entre paréntesis) muestral de los estimadores PMV de ϕ, d y σ_η^2 en el dominio de las frecuencias para el modelo ARLMSV. 3000 réplicas.

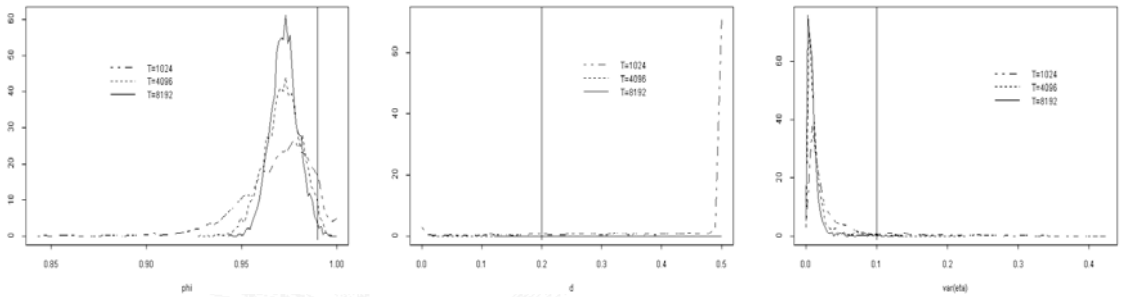
Parámetros	**ϕ**			**d**			**σ_η^2**		
Modelo ARLMSV	**T=1024**	**T=4096**	**T=8192**	**T=1024**	**T=4096**	**T=8192**	**T=1024**	**T=4096**	**T=8192**
ϕ=0.8,d=0.2, σ_η^2=0.1	-0.117 (0.326)	-0.058 (0.200)	-0.026 (0.121)	-0.035 (0.160)	-0.016 (0.111)	-0.010 (0.078)	0.131 (0.270)	0.055 (0.137)	0.026 (0.079)
ϕ=0.9,d=0.2, σ_η^2=0.1	-0.052 (0.190)	-0.013 (0.089)	-0.003 (0.050)	-0.024 (0.171)	-0.020 (0.122)	-0.016 (0.088)	0.071 (0.177)	0.024 (0.071)	0.013 (0.041)
ϕ=0.99,d=0.2, σ_η^2=0.1	-0.022 (0.019)	-0.018 (0.010)	-0.018 (0.007)	0.224 (0.139)	0.275 (0.070)	0.286 (0.049)	-0.065 (0.052)	-0.087 (0.018)	-0.090 (0.013)
ϕ=0.2,d=0.4, σ_η^2=0.1	0.167 (0.520)	0.073 (0.478)	0.027 (0.425)	-0.136 (0.189)	-0.070 (0.131)	-0.034 (0.092)	0.116 (0.271)	0.057 (0.156)	0.041 (0.121)
ϕ=0.2,d=0.45, σ_η^2=0.1	0.174 (0.530)	0.072 (0.471)	0.026 (0.424)	-0.161 (0.181)	-0.065 (0.117)	-0.032 (0.081)	0.120 (0.271)	0.059 (0.155)	0.044 (0.123)
ϕ=0.2,d=0.49, σ_η^2=0.1	0.169 (0.533)	0.083 (0.466)	0.047 (0.421)	-0.171 (0.174)	-0.067 (0.105)	-0.036 (0.066)	0.127 (0.275)	0.061 (0.157)	0.045 (0.125)

Gráfico 3.9. Densidad estimada de los estimadores PMV en el dominio de las frecuencias de ϕ (panel izquierdo), d (panel central) y σ_{η}^2 (panel derecho) para varios modelos ARLMSV. 3000 réplicas.

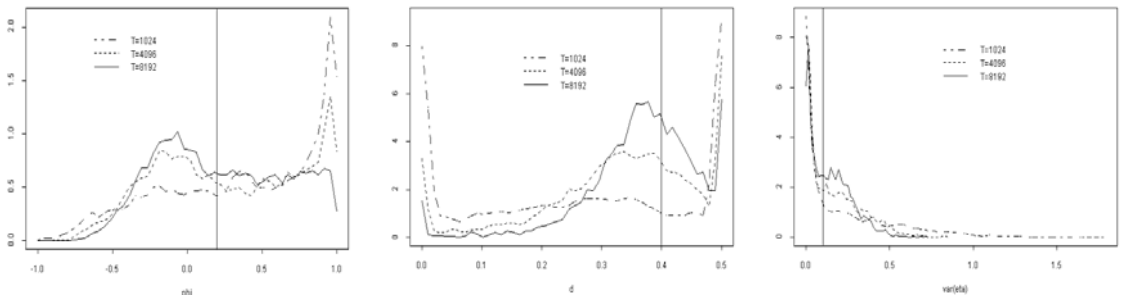
3.9(a) Modelo ARLMSV con ϕ=0.9, d=0.2, σ_{η}^2 =0.1



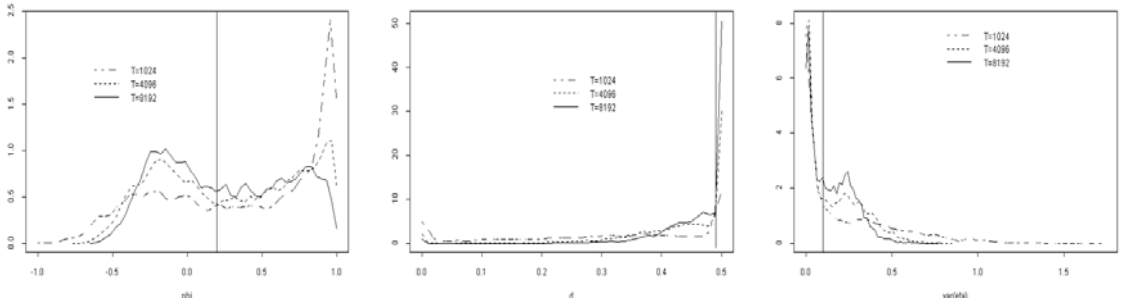
3.9(b) Modelo ARLMSV con ϕ=0.99, d=0.2, σ_{η}^2 =0.1



3.9(c) Modelo ARLMSV con ϕ=0.2, d=0.4, σ_{η}^2 =0.1



3.9(d) Modelo ARLMSV con ϕ=0.2, d=0.49, σ_{η}^2 =0.1



Los resultados de este experimento son sorprendentes. Comencemos por analizar los tres modelos en los que ϕ está muy próximo a 1. En los modelos con ϕ=0.8 ó ϕ=0.9 y $\{d=0.2, \sigma_\eta^2=0.1\}$, la estimación de los parámetros d y σ_η^2 mejora con respecto a la estimación que se obtenía en el correspondiente modelo LMSV básico con los mismos valores de d y σ_η^2. Además parece que la estimación es mejor cuanto mayor es ϕ. La varianza de $\hat{d}$ disminuye, especialmente en muestras grandes, y el sesgo, que sigue siendo negativo, también disminuye. Se reduce también la dispersión de $\hat{\sigma}_\eta^2$, especialmente en el modelo con ϕ=0.9, donde el cambio es realmente importante con respecto al modelo LMSV básico, y disminuye también considerablemente el sesgo. A estas mismas conclusiones se puede llegar comparando las densidades de $\hat{d}$ y $\hat{\sigma}_\eta^2$ en el gráfico 3.1 (paneles de la derecha) con las densidades correspondientes en el gráfico 3.9(a), en el que se aprecia claramente que la inclusión del parámetro ϕ=0.9 ha mejorado la precisión de la estimación de d y σ_η^2.

Sin embargo, cuando se fuerza un valor de ϕ muy próximo a la raíz unitaria, la situación cambia completamente. Cuando ϕ=0.99, el parámetro d no puede estimarse con ninguna precisión. El sesgo de $\hat{d}$ cambia repentinamente de signo y pasa a ser siempre positivo y muy grande, y además no disminuye con T. Por ejemplo, con T=8192, se estima $\hat{d}$=0.486, cuando el verdadero valor del parámetro es d=0.2. Este efecto se aprecia claramente en la densidad de $\hat{d}$, que se representa en el panel central del gráfico 3.9(b): la distribución está sesgada hacia el valor d=1/2, con una gran concentración de masa en torno a ese valor. De hecho, en las simulaciones con T=8192, de las 3000 réplicas realizadas, 2585 (más del 85%) estimaron $\hat{d}$=1/2. Es decir, el método PMV sobrestima d tratando de capturar la no estacionariedad debida a la presencia de una raíz casi unitaria en el polinomio autorregresivo. Por lo que respecta a la estimación del parámetro σ_η^2, podemos decir que cuando ϕ=0.99, el sesgo de $\hat{\sigma}_\eta^2$ pasa a ser negativo y tampoco decrece con T. Igualmente, el parámetro ϕ siempre se subestima y la magnitud de su sesgo tampoco disminuye al aumentar T,

como puede verse en el panel izquierdo del gráfico 3.9(b). Recordemos una vez más que los estimadores PMV en el dominio de las frecuencias son consistentes, según el teorema de Breidt, Crato y de Lima (1998), bajo una condición que asegure la identificación de los parámetros del modelo, algo que no se cumple en este caso.

Analicemos ahora qué ocurre en los modelos ARLMSV cuando la causa de la no estacionariedad es el parámetro d, que toma valores muy próximos a 1/2. Los resultados de este experimento se recogen en las últimas filas de la tabla 3.2 y en los gráficos 3.9(c) – 3.9(d). En este caso parece que los problemas de estimación surgen con el parámetro autorregresivo. A diferencia de lo que ocurría antes, el parámetro ϕ presenta ahora un sesgo positivo y de mayor magnitud, que parece aumentar ligeramente cuanto más próximo está d al valor de la frontera de la no estacionariedad. Esto significa que el método PMV siempre sobrestima ϕ en su intento de capturar una raíz unitaria que pudiera estar causando la no estacionariedad del modelo. De hecho, en los gráficos 3.9(c) y 3.9(d) se observa que la densidad estimada de $\hat{\phi}$ presenta una concentración de masa en torno al uno mayor de la esperada. Por lo que respecta a la estimación de los parámetros d y σ_η^2, y comparando los resultados con los que se obtenían para los correspondientes modelos LMSV sin parámetro autorregresivo, podemos concluir que el método PMV sigue subestimando el parámetro d, aunque ahora el sesgo es mucho mayor y también ha aumentado su dispersión. Sin embargo, la estimación de σ_η^2 prácticamente no ha cambiado. Curiosamente, los resultados que se obtienen en la estimación de los tres parámetros del modelo ARLMSV con $\{\phi=0.2, \sigma_\eta^2=0.1\}$ son prácticamente iguales para los tres valores de d analizados, d={0.4, 0.45, 0.49}, sin importar la mayor o menor proximidad a la no estacionariedad.

En los dos experimentos realizados hemos comprobado que el método de estimación PMV en el dominio de las frecuencias es incapaz de detectar la causa de la no estacionariedad en un modelo ARLMSV, tanto si ésta proviene de una raíz unitaria en el polinomio autorregresivo, como si se trata de un orden de integración fraccionario próximo a 1/2. Estos resultados pueden tener implicaciones importantes

en aplicaciones empíricas con datos reales, donde parece que lo habitual es estimar precisamente modelos próximos a la no estacionariedad; véase por ejemplo Breidt, Crato y de Lima (1998), Ray y Tsay (2000) y los resultados del capítulo 4.

3.4. Resultados preliminares de la estimación PMV en el dominio temporal con el Filtro de Kalman

Chan y Palma (1998) proponen una representación canónica infinita en el espacio de los estados para procesos ARFIMA gaussianos,[8] y demuestran que a pesar de esta representación infinita, la función de verosimilitud de estos procesos puede evaluarse en un numero finito de pasos utilizando el filtro de Kalman, lo cual facilita, en principio, la obtención de los estimadores máximo verosímiles en el dominio temporal.

En el modelo LMSV básico definido en [3.2], el proceso h_t que define la volatilidad es un ARFIMA(0,d,0) y, por tanto, h_t admite una representación en el espacio de los estados como la propuesta por Chan y Palma (1998). Utilizando este resultado, se demuestra fácilmente que el modelo LMSV en [3.2] admite una representación infinita en el espacio de los estados de la forma:

$$x_t = \mu + \mathbf{Z}\,\alpha_t + \xi_t \qquad [3.12a]$$

$$\alpha_t = \mathbf{T}\,\alpha_{t-1} + \mathbf{R}\,\eta_t \qquad [3.12b]$$

donde μ es la media de la serie x_t, ξ_t es un ruido blanco no gaussiano de media 0 y

[8] Véase Hannan y Deistler (1988, p. 8) para una descripción de la representación canónica en el espacio de los estados de procesos lineales.

varianza σ_ξ^2, η_t es NID$(0, \sigma_\eta^2)$ e independiente de ξ_t, α_t es el vector de estado:

$$\alpha_t = \begin{bmatrix} h_t \\ E_t(h_{t+1}) \\ E_t(h_{t+2}) \\ \vdots \end{bmatrix}$$

donde $E_t(h_s)$ representa la esperanza de h_s condicionada por $\{h_t, h_{t-1}, h_{t-2}, \ldots\}$, y **Z**, **T** y **R** son las matrices del sistema:

$$\mathbf{Z} = [1, 0, 0, \ldots], \mathbf{T} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \mathbf{R} = \begin{bmatrix} 1 \\ \varphi_1 \\ \varphi_2 \\ \vdots \end{bmatrix}$$

donde $\{1, \varphi_1, \varphi_2, \ldots\}$ son los coeficientes de la representación MA(∞) del proceso h_t:

$$\varphi_0 = 1, \ \varphi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)}, \ k = 1, 2, \ldots$$

definidos en la ecuación [1.30] del capítulo 1.

La representación en el espacio de los estados del modelo LMSV en [3.12] difiere en varios aspectos de la representación canónica que utilizan Chan y Palma (1998) para los procesos ARFIMA gaussianos. En primer lugar, la perturbación de la ecuación de medida, ξ_t, y la perturbación de la ecuación de transición, η_t, son diferentes y están incorrelacionadas. En segundo lugar, la ecuación de medida [3.12a] contiene una constante distinta de cero, μ, que representa la media de la serie x_t, y que no aparece en el modelo de Chan y Palma (1998). Finalmente, la perturbación ξ_t no tiene una distribución normal y, por tanto, el modelo no es condicionalmente gaussiano. En este sentido, los estimadores de los parámetros del modelo [3.12] que se obtendrían maximizando la función de verosimilitud gaussiana de $\{x_1, \ldots, x_T\}$ serían

estimadores *pseudo*-máximo verosímiles, y por tanto, las propiedades asintóticas que se derivan en Chan y Palma (1998) no serían directamente aplicables a la estimación del modelo LMSV. Estudiar las propiedades de estos estimadores PMV en el dominio temporal para el modelo LMSV, es un tema pendiente para investigaciones futuras.

A pesar de las diferencias que existen entre la representación en el espacio de los estados del proceso ARFIMA gaussiano, propuesta por Chan y Palma (1998), y la representación en el espacio de los estados del modelo LMSV en [3.12], vamos a demostrar que, para una muestra de tamaño T, la función de verosimilitud de este último modelo, calculada como si ξ_t fuera normal, puede obtenerse utilizando sólamente las T primeras componentes del vector α_t. Por tanto, los estimadores *pseudo*-máximo verosímiles de los parámetros del modelo LMSV en el dominio temporal, pueden obtenerse en un número finito de pasos. Para ello, describiremos previamente la forma de las ecuaciones del filtro de Kalman en el modelo LMSV.

La función de verosimilitud gaussiana de la serie $\{x_1,\ldots,x_T\}$ para el modelo LMSV se puede calcular utilizando la descomposición en los errores de predicción, de la siguiente forma:[9]

$$\mathbb{L}_T(\psi,\mu) = \ln L\,(\psi,\mu;\mathbf{x}) = -\frac{1}{2}\left[T\ln(2\pi) + \sum_{t=1}^{T}\ln f_t + \sum_{t=1}^{T}\frac{\nu_t^2}{f_t^2}\right] \qquad [3.13]$$

donde T es el tamaño muestral, μ es la media de x_t, ψ es el vector de parámetros, $\psi=(d,\sigma_\eta^2,\sigma_\xi^2)$, $\mathbf{x}=(x_1,\ldots,x_T)'$ es el vector de las observaciones, ν_t es el error de predicción:

$$\nu_t = x_t - \hat{x}_{t/t-1} \qquad [3.14]$$

donde $\hat{x}_{t/t-1}$ la predicción de x_t un paso hacia adelante, y f_t es el error cuadrático

[9] Véase, por ejemplo, Harvey (1989, p.126).

medio del error de predicción:

$$f_t=E(x_t-\hat{x}_{t/t-1})^2$$

El filtro de Kalman proporciona de forma recursiva los valores del error de predicción, ν_t, y del error cuadrático medio, f_t, para t=1,2,…,T, necesarios para calcular $L_T(\psi,\mu)$. Los estimadores pseudo-máximo verosímiles de los parámetros del modelo se obtendrían después maximizando $L_T(\psi,\mu)$ con respecto a (ψ,μ) mediante algún algoritmo de optimización numérica.

Denotemos por $\hat{\alpha}_{t/t-1}$ la predicción del vector de estado un paso hacia adelante dada la información hasta el instante t-1, y sea $P_{t/t-1}$ la correspondiente matriz de varianzas y covarianzas del error, $P_{t/t-1}=E_{t-1}(\alpha_t-\hat{\alpha}_{t/t-1})(\alpha_t-\hat{\alpha}_{t/t-1})'=\sigma_\eta^2\Omega_t$. Adaptando al modelo LMSV las ecuaciones de predicción y adaptación del filtro de Kalman según se establecen en Harvey (1989, p.106), se obtienen las siguientes recursividades para la predicción del estado y la varianza del error:

$$\hat{\alpha}_{t+1/t}[i]=\hat{\alpha}_{t/t-1}[i+1]+\frac{\sigma_\eta^2}{f_t}\omega_{i+1,1}^{(t)}\nu_t\,,\ i=1,2,\ldots. \qquad [3.15a]$$

$$\omega_{i,j}^{(t+1)}=\omega_{i+1,j+1}^{(t)}-\frac{\sigma_\eta^2}{f_t}\omega_{i+1,1}^{(t)}\omega_{j+1,1}^{(t)}+\varphi_{i-1}\varphi_{j-1}\,,\ i,j=1,2,\ldots,T \qquad [3.15b]$$

donde $\hat{\alpha}_{t/t-1}[i]$ denota el i-ésimo elemento de $\hat{\alpha}_{t/t-1}$, $\omega_{i,j}^{(t)}$ denota el elemento (i,j) de la matriz Ω_t y f_t denota el error cuadrático medio del error de predicción. Teniendo en cuenta que la predicción de x_t un paso hacia adelante admite la siguiente expresión:

$$\hat{x}_{t/t-1}=\mu+\mathbf{Z}\,\hat{\alpha}_{t/t-1}=\mu+\hat{\alpha}_{t/t-1}[1] \qquad [3.16]$$

se demuestra de forma inmediata que f_t puede escribirse como:

$$f_t = \sigma_\eta^2 \omega_{1,1}^{(t)} + \sigma_\xi^2 \qquad [3.17]$$

Para inicializar el filtro de Kalman es necesario establecer condiciones iniciales. En el modelo LMSV con 0<d<1/2, esas condiciones son:

$$\hat{\alpha}_{1/0} = E(\alpha_t) = (0,0,\ldots)'$$

$$P_{1/0} = E(\alpha_t \alpha_t') = \sigma_\eta^2 \Omega_1$$

donde $\Omega_1 = (\omega_{i,j}^{(1)})_{i,j}$ con $\omega_{i,j}^{(1)} = \sum_{k=0}^{\infty} \varphi_{i-1+k} \varphi_{j-1+k}$, siendo φ_j los coeficientes del vector **R**. Partiendo de estas condiciones iniciales, los valores de f_t y ν_t, con t=1,2,..,T, se obtienen sustituyendo sucesivamente en las ecuaciones [3.17], [3.16] y [3.14] $\hat{\alpha}_{t/t-1}[1]$ y $\omega_{1,1}^{(t)}$ por sus valores en [3.15a] y [3.15b], respectivamente. Finalmente, sustituyendo ν_t y f_t en [3.13] se obtiene la función de verosimilitud $L_T(\psi,\mu)$.

El problema del procedimiento que acabamos de describir es que el vector de estado, α_t, es infinito y, en consecuencia, las matrices del filtro de Kalman también serían de dimensión infinita. Sin embargo, vamos a demostrar, siguiendo los pasos de la demostración de Chan y Palma (1998) para procesos ARFIMA, que la función de verosimilitud $L_T(\psi,\mu)$ en [3.13], puede calcularse utilizando solamente las T primeras componentes del vector estado en la representación [3.12]. Para ello, basta probar que los valores de ν_t y f_t, necesarios para calcular la función $L_T(\psi,\mu)$, coinciden con los valores que se obtienen utilizando solamente las primeras T componentes del vector α_t.

Sea $\tilde{\alpha}_{1/0}$ el vector formado por las T primeras componentes de $\hat{\alpha}_{1/0}$, es decir:

$$\tilde{\alpha}_{1/0} = (\hat{\alpha}_{1/0}[1], \hat{\alpha}_{1/0}[2], \ldots, \hat{\alpha}_{1/0}[T])'$$

y sea $\tilde{P}_{1/0}$ la matriz TxT definida como:

$$\widetilde{P}_{1/0}=\{E(\alpha_1[i]-\widetilde{\alpha}_{1/0}[i])(\alpha_1[j]-\widetilde{\alpha}_{1/0}[j])\}_{i,j=1,2,\ldots,T}=\sigma_\eta^2\,\widetilde{\Omega}_1$$

donde $\widetilde{\Omega}_1=(\widetilde{\omega}_{i,j}^{(1)})_{i,j=1,2,..,T}$. Entonces resulta que $\widetilde{\alpha}_{1/0}[i]=\hat{\alpha}_{1/0}[i]$, para i=1,2,…,T, y $\widetilde{\omega}_{i,j}^{(1)}=\omega_{i,j}^{(1)}$, para i,j=1,2,..,T. Partiendo de estas condiciones iniciales, las ecuaciones de predicción y adaptación del filtro de Kalman para el modelo LMSV considerando sólo las T primeras componentes de α_t son:

$$\widetilde{\alpha}_{t+1/t}[i]=\begin{cases}\widetilde{\alpha}_{t/t-1}[i]+\dfrac{\sigma_\eta^2}{\widetilde{f}_t}\,\widetilde{\omega}_{i+1,1}^{(t)}\,\widetilde{\nu}_t, & \text{si } 1\le i\le T-1\\ 0 & \text{si } i=T\end{cases} \qquad [3.18a]$$

$$\widetilde{\omega}_{i,j}^{(t)}=\begin{cases}\widetilde{\omega}_{i+1,j+1}^{(t)}+\varphi_{i-1}\varphi_{j-1}-\dfrac{\sigma_\eta^2}{\widetilde{f}_t}\,\widetilde{\omega}_{i+1,1}^{(t)}\,\widetilde{\omega}_{j+1,1}^{(t)} & \text{si } 1\le i,j\le T-1\\ \varphi_{i-1}\varphi_{j-1} & \text{si } i=T \text{ ó } j=T\end{cases} \qquad [3.18b]$$

donde:

$$\widetilde{\nu}_t=x_t-\widetilde{x}_{t/t-1} \qquad [3.19a]$$

$$\widetilde{x}_{t/t-1}=\mu+\mathbf{Z}\widetilde{\alpha}_{t/t-1}=\mu+\widetilde{\alpha}_{t/t-1}[1] \qquad [3.19b]$$

$$\widetilde{f}_t=E(x_t-\widetilde{x}_{t/t-1})^2=\sigma_\eta^2\,\widetilde{\omega}_{1,1}^{(t)}+\sigma_\xi^2 \qquad [3.19c]$$

Comparando las ecuaciones [3.15] y [3.18] se comprueba que, fijado un valor de t, con t=1,2,..,T,

$$\widetilde{\alpha}_{t/t-1}[i]=\hat{\alpha}_{t/t-1}[i], \text{ cuando } i=1,2,..,T-t+1,$$
$$\widetilde{\omega}_{i,j}^{(t)}=\omega_{i,j}^{(t)}, \text{ cuando } i,j=1,2,\ldots,T-t+1,$$

y comparando a continuación las ecuaciones [3.14], [3.16] y [3.17], con las ecuaciones

[3.19], resulta que $\tilde{v}_t = v_t$ y $\tilde{f}_t = f_t$, para todo t=1,2,…,T. Esto significa que la función de verosimilitud que se obtendría utilizando solamente las T primeras componentes del vector estado,

$$\tilde{L}_T(\psi,\mu) = -\frac{1}{2}\left[T\ln(2\pi) + \sum_{t=1}^{T} \ln \tilde{f}_t + \sum_{t=1}^{T} \frac{\tilde{v}_t^2}{\tilde{f}_t^2} \right] \qquad [3.20]$$

donde $\tilde{v}_t$ y $\tilde{f}_t$ son los valores en [3.19], coincide con la función $L_T(\psi,\mu)$ en [3.13] y, por tanto, ésta puede calcularse utilizando únicamente las primeras T componentes del vector α_t.

Obsérvese que la media del proceso, μ, aparece como un parámetro más en la función de verosimilitud a través del error de predicción. Una solución habitual en estos casos es sustituir μ por un estimador suyo, por ejemplo la media muestral, $\hat{\mu} = \overline{X}$, y aplicar el filtro de Kalman a la serie centrada $(x_t - \overline{X})$, y después maximizar con respecto a ψ la función de verosimilitud correspondiente. Sin embargo, la memoria larga asociada al modelo LMSV con d>0 puede hacer que la media muestral converja muy despacio hacia μ y, en consecuencia, los estimadores PMV en el dominio temporal calculados de esta forma pueden no tener buenas propiedades en muestras finitas. Otra posibilidad es estimar μ junto a los otros parámetros del modelo, maximizando la función de verosimilitud en [3.20] con respecto a (ψ,μ).

Con el fin de analizar el efecto que pueden tener estas dos soluciones sobre los estimadores PMV de d y σ_η^2 en el dominio temporal, y además poder comparar las propiedades de estos estimadores con sus análogos en el dominio de las frecuencias, hemos realizado un pequeño experimento de Monte Carlo que describimos a continuación.

Se han simulado series del modelo LMSV básico estacionario con $0<d<1/2$, suponiendo normalidad para η_t y ε_t. Los valores de los parámetros elegidos son

$\{d=0.4, \sigma_\eta^2=1\}$, $\{d=0.4, \sigma_\eta^2=0.1\}$ y $\{d=0.2, \sigma_\eta^2=0.1\}$, y los tamaños muestrales son T=64 y T=256. El modelo $\{d=0.4, \sigma_\eta^2=1\}$ se ha elegido como caso representativo en el que la estimación PMV en el dominio de las frecuencias funciona bien, según vimos en las simulaciones de la sección anterior. Por el contrario, $\{d=0.4, \sigma_\eta^2=0.1\}$ y $\{d=0.2, \sigma_\eta^2=0.1\}$ representan dos modelos donde la estimación en el dominio de las frecuencias tiene malas propiedades. De esta forma podremos ver si la estimación en el dominio temporal puede competir con la estimación en el dominio de las frecuencias en aquellos modelos donde ésta funciona bien y comprobaremos si la estimación en el dominio temporal mejora sustancialmente los resultados para aquellos modelos donde el estimador en el dominio de las frecuencias no es muy fiable.

Para cada especificación $\{d, \sigma_\eta^2\}$ se generan 100 muestras de tamaño T, y para cada una de ellas se obtienen las estimaciones PMV de los parámetros d y σ_η^2 utilizando del filtro de Kalman con las dos opciones mencionadas anteriormente (con media conocida y estimando la media), y las estimaciones PMV en el dominio de las frecuencias. Supondremos siempre que el parámetro σ_ξ^2 es conocido, $\sigma_\xi^2=\pi^2/2$. A partir de las estimaciones obtenidas en las 100 réplicas, el sesgo y la desviación típica muestral se calculan según se explicó en la sección 3.3. Obsérvese que tanto los tamaños muestrales como el número de réplicas en este experimento se han reducido notablemente con respecto a las simulaciones de las secciones anteriores. Esta reducción se debe a que la estimación con el filtro de Kalman requiere calcular y almacenar una matriz TxT en cada una de las T iteraciones, y esto ralentiza enormemente los cálculos computacionales.

Denotaremos por PMV1a el estimador máximo verosímil en el dominio temporal mediante el filtro de Kalman con la media μ conocida, $\mu=E(\ln(\varepsilon_t^2))\cong -1.27$. PMV1b denotará el estimador en el dominio temporal que se obtiene sustituyendo μ por la media muestral y maximizando la función de verosimilitud para la serie $(x_t-\overline{X})$. Se ha realizado el mismo experimento considerando μ como un parámetro más del

modelo y maximizando conjuntamente la función de verosimilitud de x_t con respecto a $\psi=(\mu,d,\sigma_\eta^2)$ mediante el filtro de Kalman, obteniéndose los mismos resultados que en el caso anterior. Finalmente, PMV2 denotará el estimador PMV en el dominio de las frecuencias cuyas propiedades se analizaron en las dos secciones anteriores.

La tabla 3.3 recoge los resultados de las simulaciones para los tres estimadores, junto con el tiempo de computación requerido en cada caso.

Tabla 3.3. Sesgo y desviación típica (entre paréntesis) muestral de los estimadores PMV en el dominio temporal, con media conocida (PMV1a) y con media estimada (PMV1b), y del estimador PMV en el dominio de las frecuencias (PMV2), para el modelo LMSV. 100 réplicas.

Parámetros	**PMV1a**		**PMV1b**		**PMV2**	
$\{d=0.4,\ \sigma_\eta^2=1\}$	**T=64**	**T=256**	**T=64**	**T=256**	**T=64**	**T=256**
	1h 23’	39h 25’	55’	45h 40’	6’’	11’’
Estimación de d	-0.073	-0.045	-0.236	-0.131	-0.063	-0.032
	(0.168)	(0.097)	(0.112)	(0.113)	(0.168)	(0.134)
Estimación de σ_η^2	0.244	0.176	0.248	0.288	0.249	0.179
	(1.158)	(0.680)	(1.157)	(0.683)	(1.123)	(0.712)
$\{d=0.4,\ \sigma_\eta^2=0.1\}$	**T=64**	**T=256**	**T=64**	**T=256**	**T=64**	**T=256**
	1h 10’	78h 46’	52’	44h 16’	5’’	11’’
Estimación de d	-0.194	-0.106	-0.243	-0.232	-0.095	-0.110
	(0.162)	(0.168)	(0.128)	(0.133)	(0.180)	(0.185)
Estimación de σ_η^2	0.582	0.244	0.559	0.269	0.457	0.262
	(0.947)	(0.489)	(0.922)	(0.493)	(0.827)	(0.488)
$\{d=0.2,\ \sigma_\eta^2=1\}$	**T=64**	**T=256**	**T=64**	**T=256**	**T=64**	**T=256**
	1h 6’	64h	50’	30h 33’	5’’	12’’
Estimación de d	-0.008	-0.016	-0.042	-0.054	0.093	0.026
	(0.133)	(0.148)	(0.117)	(0.123)	(0.187)	(0.186)
Estimación de σ_η^2	0.421	0.250	0.450	0.244	0.450	0.244
	(0.837)	(0.483)	(0.812)	(0.479)	(0.824)	(0.480)

Lo que más llama la atención en esta tabla es, en primer lugar, la enorme diferencia entre los resultados del estimador PMV1b, que requiere la estimación de la media, y los resultados de los otros dos estimadores, que no requieren tal estimación.

En segundo lugar, destaca la diferencia entre los tiempos de computación de los estimadores en el dominio temporal (PMV1a, PMV1b) y del estimador en el dominio de las frecuencias (PMV2), que es muchísimo más rápido.

Analizando los resultados de la tabla 3.3, queda claro que al estimar la media y trabajar con la serie centrada, $(x_t - \overline{X})$, se introduce un sesgo negativo enorme en el estimador PMV1b del parámetro d, que en ningún caso compensa la disminución en la varianza. Si bien es cierto que este sesgo disminuye al aumentar T, el tiempo que requeriría el cálculo de este estimador en muestras grandes lo hace impracticable. Observamos también que el estimador PMV1a es, en general, ligeramente superior al estimador en el dominio de las frecuencias, aunque las diferencias entre ambos no son muy importantes. Sin embargo, no debemos olvidar que PMV1a no es un estimador factible, porque supone que la media del proceso x_t es conocida, algo que no ocurre en la práctica. Incluso aunque la media fuera conocida, la diferencia entre los tiempos que requiere el cálculo de cada uno de ellos es tal que favorecería al estimador PMV2. Estos resultados son similares a los que obtienen Cheung y Diebold (1994) y Smith, Sowell y Zin (1997) para procesos ARFIMA.

A la vista de estos resultados podemos concluir que, a pesar de que la función de verosimilitud en el dominio temporal para el modelo LMSV pueda calcularse en un número finito de pasos, la estimación del parámetro d es tan sesgada y el tiempo que requiere la computación de los estimadores es tan grande, que este método no parece muy recomendable en aplicaciones con datos reales.

En el ámbito de los procesos ARFIMA, Chan y Palma (1998) proponen una alternativa a este método que consiste en truncar la representación MA(∞) del proceso en primeras diferencias y maximizar la función de verosimilitud de este modelo truncado. Si el proceso se trunca después de incluir m componentes, las matrices que aparecen ahora en las ecuaciones del filtro de Kalman serán de orden mxm, con m<T, y el algoritmo para calcular la función de verosimilitud será de orden Txmxm. Chan y Palma (1998) prueban que los estimadores máximo verosímiles que se obtienen a partir de este modelo truncado, son consistentes, asintóticamente normales y

asintóticamente eficientes, bajo ciertas condiciones sobre el valor de m. Por ejemplo, para una muestra de tamaño T=10000, estos autores prueban que truncando en un valor tan pequeño como m=14, el estimador máximo verosímil aproximado funciona realmente bien.

La aplicación de estos resultados al modelo LMSV no es tan inmediata como pudiera parecer, ya que en el modelo LMSV en primeras diferencias, definido en [3.8], la perturbación $\Delta\xi_t$ que aparece en la ecuación de medida ya no es ruido blanco y esto contradice uno de los supuestos básicos para la aplicación del filtro de Kalman. Sin embargo, no es descartable que pueda encontrarse una representación en el espacio de los estados del modelo en primeras diferencias que satisfaga las condiciones necesarias para poder aplicar los resultados de Chan y Palma (1998) y permita utilizar el filtro de Kalman para estimar por máxima verosimilitud el modelo truncado. Éste es precisamente uno de los temas en los que pensamos seguir trabajando.

3.5. Estimación suavizada de la volatilidad

Cuando se ajusta un modelo para representar la evolución dinámica de la volatilidad de una serie, existe un interés en obtener estimaciones de la propia volatilidad. Harvey (1998) propone un método para obtener una estimación suavizada de la volatilidad en el modelo LMSV. Este método consiste en lo siguiente. Consideremos la ecuación [3.2a] escrita en notación matricial de la siguiente forma:

$$\mathbf{x} = \mu\,\mathbf{i} + \mathbf{h} + \boldsymbol{\xi}$$

donde $\mathbf{x}$ es un vector Tx1 que contiene las observaciones $x_t=\ln(y_t^2)$, t=1,2,…,T, μ es la media del proceso x_t, $\mathbf{i}$ es un vector de unos Tx1, y $\mathbf{h}$ y $\boldsymbol{\xi}$ son vectores Tx1 con los valores de h_t y ξ_t, t=1,2,…,T, respectivamente.

Si el proceso h_t es estacionario (d<1/2), el estimador lineal mínimo cuadrático de h_t viene dado por la siguiente expresión:

$$\tilde{\mathbf{h}} = V_h V^{-1}(\mathbf{x} - \mu\,\mathbf{i}) + \mu\,\mathbf{i} = V_h V^{-1}\mathbf{x} + (I - V_h V^{-1})\,\mu\,\mathbf{i} \qquad [3.21]$$

donde V_h y V son las matrices de varianzas y covarianzas de h_t y x_t, respectivamente. Teniendo en cuenta que la perturbación ξ_t está incorrelacionada con h_t, resulta que $V=V_h+V_\xi$, donde V_ξ es la matriz de varianzas y covarianzas de ξ_t. Además, ξ_t está serialmente incorrelacionada (es ruido blanco) y, por tanto, $V_\xi=\sigma_\xi^2 I$, donde I es la matriz identidad TxT y σ_ξ^2 es la varianza de ξ_t. Sustituyendo estas expresiones en [3.21] se obtiene la siguiente expresión para la estimación suavizada de h_t:

$$\tilde{\mathbf{h}} = \mathbf{x} - \sigma_\xi^2 V^{-1}(\mathbf{x} - \mu\,\mathbf{i}) \qquad [3.22]$$

En la práctica la media μ es desconocida y se estima utilizando la media muestral de las observaciones $\{x_1,\ldots,x_T\}$.

Cuando el proceso h_t no es estacionario (d≥1/2), la fórmula de suavizado [3.21] se aplica al modelo en primeras diferencias definido en [3.8]. De esta forma se obtiene una estimación suavizada de la serie Δh_t. Obsérvese que en este caso V_h representaría la matriz de varianzas y covarianzas del proceso $\Delta h_t \sim$ ARFIMA(0,d*,0) con d*=d-1, y V_ξ sería la matriz de varianzas y covarianzas de la perturbación $\Delta\xi_t$, cuya varianza es $\{2\sigma_\xi^2\}$. Esta perturbación ya no es ruido blanco porque su autocovarianza de orden uno es $\mathrm{Cov}(\Delta\xi_t,\Delta\xi_{t-1}) = -\sigma_\xi^2 \neq 0$ y, por tanto, en el modelo LMSV no estacionario, V_ξ sería una matriz de la forma:

$$V_\xi = \sigma_\xi^2 \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix} \qquad [3.23]$$

Obsérvese también que en el modelo en primeras diferencias, el término correspondiente a la media, μ, desaparece y la ecuación de suavizado [3.22] se convierte en:

$$\tilde{\mathbf{h}}^* = \mathbf{x}^* - V_\xi V^{-1} \mathbf{x}^* \qquad [3.24]$$

donde $\tilde{\mathbf{h}}^*$ es un vector que contiene las estimaciones suavizadas de Δh_t, t=2,…,T; $\mathbf{x}^*$ es un vector con las observaciones de Δx_t, t=2,…,T; V_ξ es la matriz de varianzas y covarianzas de $\Delta\xi_t$ en [3.23]; y V es la matriz de varianzas y covarianzas de Δx_t. Las estimaciones suavizadas de h_t se obtienen de forma recursiva a partir de las componentes del vector $\tilde{\mathbf{h}}^*$ según las fórmulas:

$$\tilde{h}_{t/T} = \tilde{h}_{t-1/T} + \tilde{h}^*_{t/T},\ t=2,\ldots,T$$
$$\tilde{h}_{1/T} = 0$$

Una vez obtenido el vector $\tilde{\mathbf{h}}$ con las estimaciones suavizadas del proceso h_t para toda la muestra, la estimación suavizada de la volatilidad subyacente, que denotaremos por $\hat{\sigma}_{t/T}$, se obtiene como el producto:

$$\hat{\sigma}_{t/T} = \hat{\sigma}_* \tilde{\sigma}_t \qquad [3.25]$$

con t=1,2,…,T, donde $\hat{\sigma}_*$ es un estimador del factor de escala y $\tilde{\sigma}_t$ se obtiene como:

$$\tilde{\sigma}_t = \exp(\tfrac{1}{2}\,\tilde{h}_{t/T}),\ \text{para } t=1,2,\ldots,T \qquad [3.26]$$

Harvey y Shephard (1993) sugieren que el estimador $\hat{\sigma}_*$ sea la desviación típica muestral de las observaciones y_t corregidas por heteroscedasticidad. Es decir:

$$\hat{\sigma}_* = \sqrt{\frac{1}{T}\sum_{t=1}^{T} \tilde{y}_t^2}$$

donde:

$$\tilde{y}_t = y_t / \tilde{\sigma}_t = y_t \exp(-\tfrac{1}{2}\,\tilde{h}_{t/T})$$

BIBLIOTECA VIRTUAL

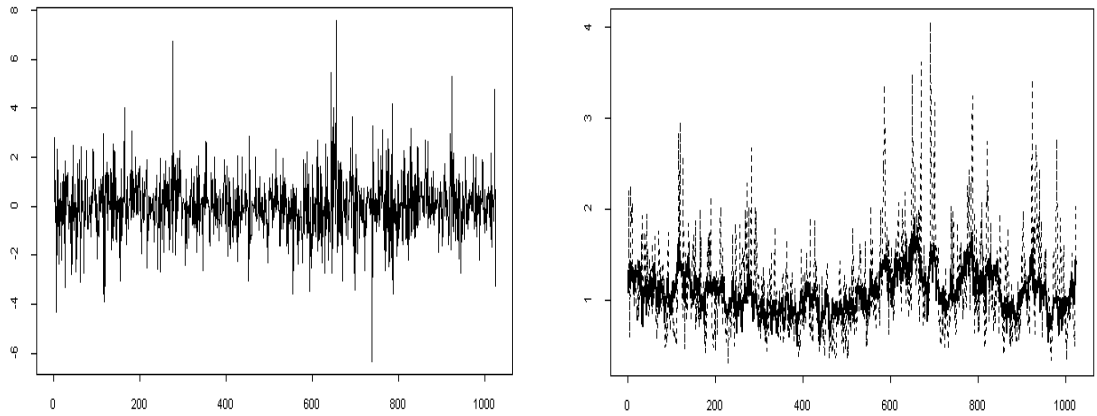
siendo $\tilde{\sigma}_t$ el estimador en [3.26].

Para comprobar el comportamiento del método de suavizado que acabamos de describir, se han generado series $\{y_t\}_{t=1,\ldots,T}$ a partir de tres modelos LMSV con parámetros $\{d=0.4, \sigma_\eta^2=0.5\}$, $\{d=0.49, \sigma_\eta^2=0.1\}$ y $\{d=0.75, \sigma_\eta^2=0.1\}$, respectivamente, y $\sigma_*=1$. De esta forma, las series generadas representan un modelo estacionario ($d<1/2$), un modelo en la frontera de la no estacionariedad ($d\approx 1/2$) y un modelo claramente no estacionario ($d\geq 1/2$). En los tres casos se utiliza T=1024.

Las tres series simuladas y las correspondientes estimaciones de su volatilidad se representan en el gráfico 3.10. Para cada especificación de (d, σ_η^2), el panel izquierdo del gráfico representa la serie simulada, $\{y_t\}_{t=1,\ldots,T}$, y el panel derecho representa la volatilidad de dicha serie, σ_t, junto con la estimación suavizada de la volatilidad, $\hat{\sigma}_{t/T}$, para t=1,2,…,T, obtenida con el algoritmo de suavizado descrito previamente. El algoritmo de suavizado se ha aplicado a los modelos estimados, utilizando las estimaciones PMV en el dominio de las frecuencias de los parámetros del modelo, que han sido $\{\hat{d}=0.348, \hat{\sigma}_\eta^2=0.504\}$, $\{\hat{d}=0.494, \hat{\sigma}_\eta^2=0.16\}$ y $\{\hat{d}=0.731, \hat{\sigma}_\eta^2=0.129\}$, respectivamente.

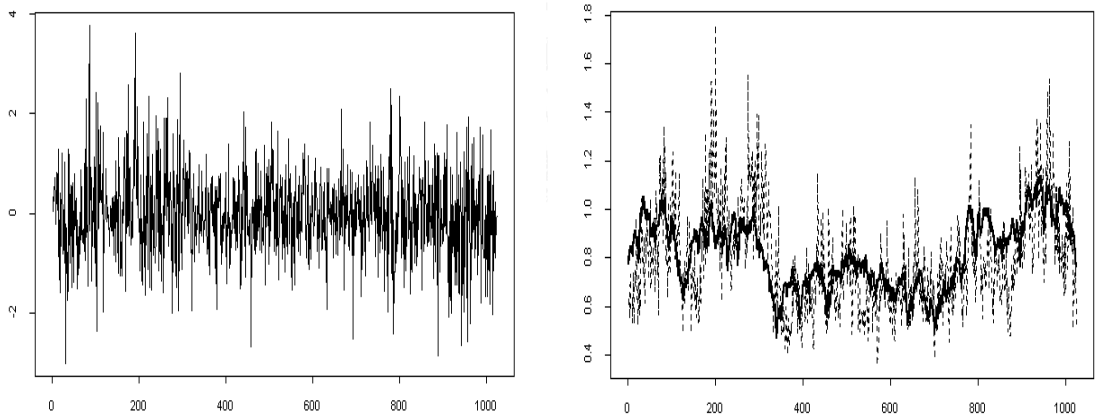
Gráfico 3.10. Series y_t generadas por diferentes modelos LMSV (panel izquierdo) junto con su volatilidad σ_t (panel derecho línea punteada) y la estimación suavizada de la volatilidad $\hat{\sigma}_{t/T}$ (panel derecho negrita)

3.10(a) Modelo LMSV con {d=0.4, σ_η^2=0.5}

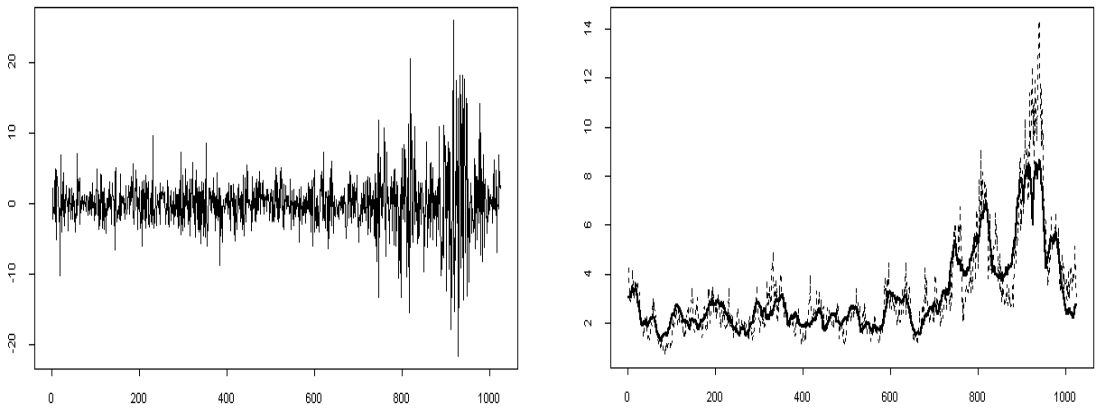


BIBLIOTECA VIRTUAL

3.10(b) Modelo LMSV con {d=0.49, σ_η^2=0.1}



3.10(c) Modelo LMSV con {d=0.75, σ_η^2=0.1}

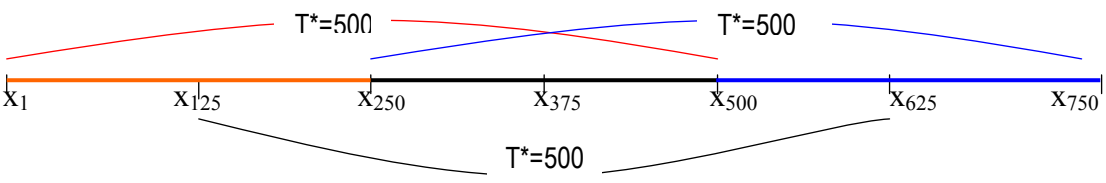


A la vista de estos gráficos podemos concluir que el método de suavizado propuesto por Harvey (1998) reproduce con bastante exactitud la evolución dinámica de la volatilidad, tanto en modelos estacionarios como en modelos no estacionarios, al menos para las series aquí analizadas. Sin embargo, este procedimiento presenta un problema de tipo práctico, ya que al programar las ecuaciones del suavizado es necesario guardar e invertir matrices TxT, y esto puede ser impracticable en muestras grandes. Para estos casos, Harvey (1998) propone una solución que sólo requiere invertir una matriz de orden menor que T, aunque finalmente no la implementa.

La propuesta de Harvey se basa en observar lo siguiente. Cuando el tamaño muestral T es muy grande, se comprueba que la matriz del algoritmo de suavizado en [3.21], $V_h V^{-1}$, asigna unos pesos muy pequeños a las observaciones muy distantes, tan pequeños que si el suavizado se realizara utilizando unos pesos calculados para una muestra más pequeña, digamos de tamaño T*, con T*<T, y se completaran los pesos restantes de forma adecuada, apenas se perdería precisión en la estimación.

Para completar esa matriz de pesos más pequeña, de orden T*xT*, y conseguir una matriz de pesos para toda la muestra, de orden TxT, se propone desplazar por bloques la matriz T*xT* sobre la diagonal principal de una matriz TxT, y completar con ceros los pesos que faltan en los extremos. Con esta estrategia, sólo habrá sido necesario invertir una matriz T*xT*, con T*<T.

Por ejemplo, si T=750 y los pesos de la matriz de suavizado han sido calculados para una muestra de tamaño T*=500, la matriz de pesos del suavizado para toda la muestra, $\{x_1,\ldots,x_{750}\}$, que es de orden TxT, se puede construir desplazando la matriz T*xT* en tres bloques sobre la diagonal y completando con ceros los pesos de las observaciones más distantes. En las zonas donde haya solapamientos, se eligen aquellos pesos más próximos a las observaciones centrales según el siguiente esquema:



Así, para obtener el suavizado de h_t, para t=1,2,…,250, los pesos asignados a las observaciones $\{x_1,\ldots,x_{500}\}$ serán los de la fila t correspondiente de la matriz T*xT*, y los pesos asignados a las restantes observaciones, $\{x_{501},\ldots,x_{750}\}$ serán cero. Para obtener la estimación suavizada de h_t, con t=251,…,500, asignaremos pesos cero a las observaciones extremas, $\{x_1,\ldots,x_{125}\}$ y $\{x_{626},\ldots,x_{750}\}$, y a las observaciones centrales, $\{x_{126},\ldots,x_{625}\}$, les asignaremos los pesos de la fila correspondiente de la matriz T*xT*. Finalmente, la estimación suavizada de h_t, para t=501,…,750, se obtiene asignando pesos cero a las primeras observaciones, $\{x_1,\ldots,x_{250}\}$, y asignando a las observaciones restantes, $\{x_{256},\ldots,x_{750}\}$ los pesos de la fila correspondiente de la matriz T*xT*.

En resumen, si denotamos por $(w_{ij})_{i,j=1,\ldots,T^*}$ los pesos calculados con la muestra de tamaño T*=500, la matriz de pesos para toda la muestra (T=750), sería de la forma:

$$\begin{bmatrix}
w_1 \cdots w_{125} & w_{126} \cdots w_{250} & w_{251} \cdots w_{375} & w_{376} \cdots w_{500} & 0 \cdots 0 & 0 \cdots 0 \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
w_{125} \cdots w_1 & w_2 \cdots w_{126} & w_{127} \cdots w_{251} & w_{252} \cdots w_{376} & 0 \cdots 0 & 0 \cdots 0 \\
w_{126} \cdots w_2 & w_1 \cdots w_{125} & w_{126} \cdots w_{250} & w_{251} \cdots w_{375} & 0 \cdots 0 & 0 \cdots 0 \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
w_{250} \cdots w_{126} & w_{125} \cdots w_1 & w_2 \cdots w_{126} & w_{127} \cdots w_{251} & 0 \cdots 0 & 0 \cdots 0 \\
0 \cdots 0 & w_{126} \cdots w_2 & w_1 \cdots w_{125} & w_{126} \cdots w_{250} & w_{251} \cdots w_{375} & 0 \cdots 0 \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
0 \cdots 0 & w_{250} \cdots w_{126} & w_{125} \cdots w_1 & w_2 \cdots w_{126} & w_{127} \cdots w_{251} & 0 \cdots 0 \\
0 \cdots 0 & w_{251} \cdots w_{127} & w_{126} \cdots w_2 & w_1 \cdots w_{125} & w_{126} \cdots w_{250} & 0 \cdots 0 \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
0 \cdots 0 & w_{375} \cdots w_{251} & w_{250} \cdots w_{126} & w_{125} \cdots w_1 & w_2 \cdots w_{126} & 0 \cdots 0 \\
0 \cdots 0 & 0 \cdots 0 & w_{251} \cdots w_{127} & w_{126} \cdots w_2 & w_1 \cdots w_{125} & w_{126} \cdots w_{250} \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
0 \cdots 0 & 0 \cdots 0 & w_{375} \cdots w_{251} & w_{250} \cdots w_{126} & w_{125} \cdots w_1 & w_2 \cdots w_{126} \\
0 \cdots 0 & 0 \cdots 0 & w_{376} \cdots w_{250} & w_{251} \cdots w_{127} & w_{126} \cdots w_2 & w_1 \cdots w_{125} \\
\vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots & \vdots \quad \vdots \\
0 \cdots 0 & 0 \cdots 0 & w_{500} \cdots w_{376} & w_{375} \cdots w_{251} & w_{250} \cdots w_{126} & w_{125} \cdots w_1
\end{bmatrix}$$

[3.27]

Para tener una idea de la precisión que se consigue con esta solución, hemos calculado los pesos verdaderos que corresponderían a la muestra de tamaño T=750, y los pesos que se obtendrían considerando sólo una muestra de tamaño T*=500 en un modelo muy persistente, con $\{d=0.45, \sigma_\eta^2=0.1\}$. El gráfico 3.11(a) representa los verdaderos pesos calculados con toda la muestra, que se asignarían a las observaciones $\{x_1,\ldots,x_T\}$ para obtener las estimaciones de $\tilde{h}_t$, para t=75, 225, 375, 600. Es decir, este gráfico representa los valores de las filas 75, 225, 375, 600 de la matriz $V_h V^{-1}$ en [3.21]. El gráfico 3.11(b) representa los pesos que recibirían las observaciones $\{x_1,\ldots,x_T\}$ si obtenemos los pesos del suavizado únicamente para una muestra de tamaño T*=500, y completamos los pesos restantes de la matriz TxT según se indica en [3.27]. Es decir, los valores que se representan en el gráfico 3.11(b) son los que aparecerían en las filas 75, 225, 375, 600, respectivamente, de la matriz de pesos TxT en [3.27].

A la vista de estos gráficos queda claro que la diferencia entre los pesos verdaderos y los pesos que se obtienen al adoptar la solución que propone Harvey (1998), es despreciable. Piénsese, además, que en nuestro ejemplo, los tamaños muestrales elegidos no son muy grandes. Cuanto mayor sea T*, menores serán los pesos asignados a las observaciones extremas y mejor será la aproximación. Por ejemplo, con T*=1200, los pesos asignados a los extremos en la matriz T*xT* ya son prácticamente cero y la pérdida de precisión es mínima, pues para cualquier tamaño muestral T mayor que T*, los verdaderos pesos que asignaría el algoritmo a las observaciones a partir de T*, también son casi idénticos a cero.

Gráfico 3.11. Pesos del suavizado para $\widetilde{h}_{75}$, $\widetilde{h}_{225}$, $\widetilde{h}_{375}$, $\widetilde{h}_{600}$ en un modelo LMSV con d=0.45, σ_{η}^{2} =0.1.

3.11(a) Pesos exactos para una muestra de tamaño T=750

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

3.11(b) Pesos aproximados para una muestra de tamaño T=750 calculados con T*=500

Capítulo 4.

APLICACIÓN EMPÍRICA: MODELIZACIÓN DE LA VOLATILIDAD DEL INDICE IBEX-35

BIBLIOTECA VIRTUAL
MIGUEL DE CERVANTES

4.1. Análisis preliminar de los rendimientos del IBEX-35

Los datos que vamos a analizar corresponden a la serie de rendimientos diarios del índice IBEX-35 de la Bolsa de Madrid, en el periodo comprendido entre el 5/1/1987 y el 30/12/1998, con un total de 2991 observaciones. El rendimiento diario del índice se define como la variación porcentual en el logaritmo del precio de cierre del índice entre dos días consecutivos de mercado. Por tanto, el rendimiento diario para el día t, que denotamos por R_t, se calcula como:

$$R_t=100x(\ln P_t-\ln P_{t-1})$$

donde P_t denota el valor del índice del día t. Los gráficos 4.1 y 4.2 representan la evolución de las series P_t y R_t, respectivamente. La tabla 4.1 resume las principales características descriptivas de la serie R_t y la tabla 4.2 presenta las autocorrelaciones muestrales de dicha serie junto a las correlaciones muestrales de la serie de sus cuadrados, de los valores absolutos y del logaritmo de los cuadrados. Estas tres últimas series se han elegido porque son las que suelen considerarse como representativas de la volatilidad de R_t; véase Crato y Ray (2000).

Gráfico 4.1. Evolución diaria del IBEX-35 en el periodo 5/1/87-30/12/98

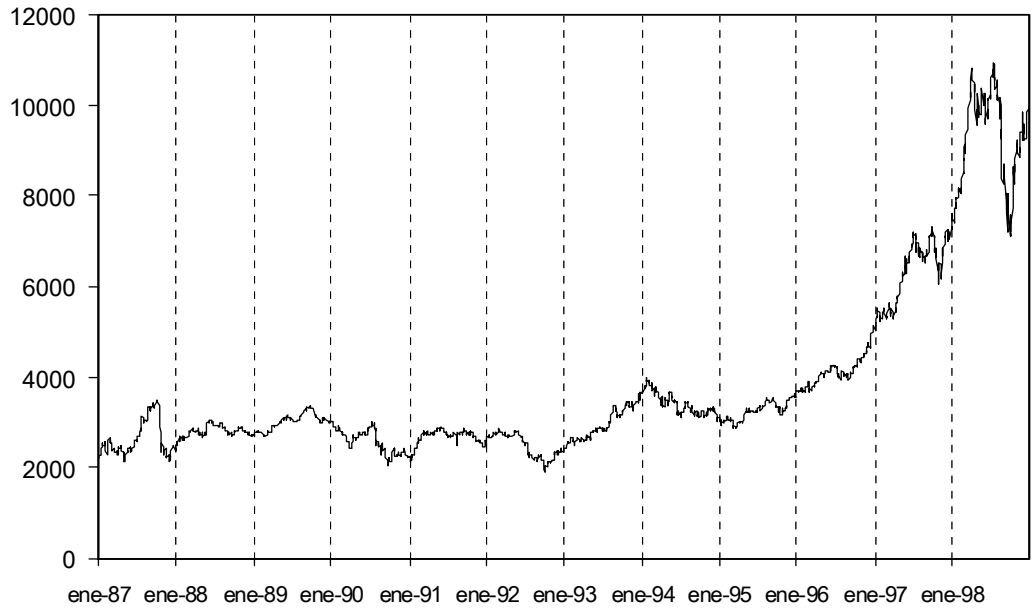


Gráfico 4.2. Rendimientos diarios del IBEX-35 en el periodo 7/1/87-30/12/98

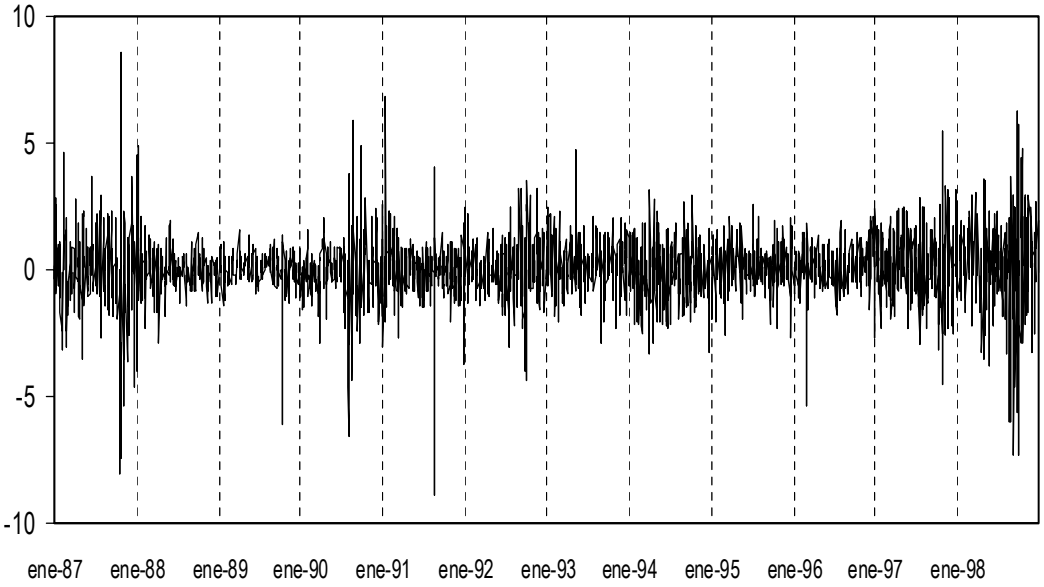


Tabla 4.1. Momentos muestrales de los rendimientos diarios del IBEX-35 en el periodo 7/1/87-30/12/98

Tamaño muestral	Media	Desviación típica	Asimetría	Curtosis	Mínimo	Máximo	Contraste de Normalidad
2991	0.0501	1.2421	-0.4870*	9.4118*	-8.8758	8.5918	5241.751*

* Valores significativos al 5%.

Tabla 4.2. Autocorrelaciones muestrales de los rendimientos del IBEX-35 en el periodo 7/1/87-30/12/98

Retardo Serie	k=1	k=2	k=3	k=4	k=5	k=10	k=20	k=50	k=100
R_t	0.158*	0.002	-0.005	0.026	-0.003	0.057*	-0.038*	-0.016	-0.007
[s.e.$_c$]	[0.029]	[0.033]	[0.027]	[0.032]	[0.029]	[0.028]	[0.024]	[0.022]	[0.020]
R_t^2	0.184*	0.275*	0.141*	0.249*	0.181*	0.166*	0.084*	0.047*	0.029
$\|R_t\|$	0.241*	0.271*	0.222*	0.213*	0.229*	0.227*	0.167*	0.113*	0.087*
$\ln(R_t^2)$	0.101*	0.139*	0.142*	0.092*	0.100*	0.112*	0.089*	0.076*	0.055*

* Valores significativos al 5% respecto a la distribución N(0,1/T).

Nota: Los valores s.e.$_c$ de esta tabla denotan el error estándar de las correlaciones muestrales corregido por volatilidad según la expresión [4.1].

Las características más destacables de la serie de rendimientos del IBEX-35 que se desprenden del análisis del gráfico 4.2 y de las tablas 4.1 y 4.2, son las siguientes:

1) En el gráfico 4.2 observamos que la serie de rendimientos se mueve alrededor de un nivel medio cero, que es bastante estable en toda la muestra. Por el contrario, la varianza de la serie presenta cambios a lo largo del tiempo, alternándose los periodos de poca volatilidad, caracterizados por cambios pequeños en el valor de los rendimientos, con periodos de mucha volatilidad, en los que el margen de variación de los rendimientos en torno a la media es muy amplio. Por ejemplo, a finales de 1987 hay mucha volatilidad debido a la fuerte caída de la Bolsa de Nueva York en octubre de ese año; en agosto de 1990 hay otro periodo de mucha

incertidumbre a causa de la crisis del Golfo; y en el verano de 1998 la bolsa presenta mucha volatilidad debido a las crisis financieras asiática y rusa.

2) En el gráfico 4.2 se observa además la presencia de algunas observaciones individuales posiblemente atípicas, con valores extremadamente grandes, positivos o negativos, dentro de periodos de relativa calma. Por ejemplo, el día 19/8/1991 se observa un valor atípico especialmente negativo, que corresponde a la fuerte caída que sufrieron todas las bolsas internacionales por el golpe militar que derribó a Gorbachov. El acontecimiento afectó también al valor del índice en los días sucesivos que tardó el mercado en recuperarse. Hay además otros tres valores atípicos que destacan porque se producen dentro de periodos de poca volatilidad, y que corresponden al lunes 16/10/1989 (impacto del mini-*crash* del viernes 13 en Wall Street), al día 13/5/1993 (devaluación de la peseta en el seno del Sistema Monetario Europeo) y al día 4/3/1996 (caída de la bolsa tras conocerse la "escasa" victoria electoral del PP).

3) En la tabla 4.1 destaca el valor del coeficiente de curtosis, 9.4118, que es significativamente mayor que 3, lo que indica que la distribución de los rendimientos tiene colas más anchas que la distribución normal, como es habitual en series financieras. Además, el valor del estadístico de Jarque-Bera para contrastar normalidad es 5241.751, con un p-valor asociado de 0.000 que rechaza claramente la hipótesis de normalidad.

4) Con respecto a las correlaciones muestrales de la tabla 4.2, observamos, en primer lugar, que existe mucha más correlación en la serie de los cuadrados y en los rendimientos absolutos que en la propia serie R_t, en consonancia con los resultados de Taylor (1986) y Ding, Granger y Engle (1993). Esta correlación en los cuadrados evidencia el agrupamiento de los periodos de volatilidad antes mencionado, y sugiere la presencia de heteroscedasticidad condicionada en la serie R_t (Bollerslev, 1986).

5) Como vimos en la sección 2.2, la presencia de heteroscedasticidad condicionada invalida los contrastes de incorrelación habituales debido a que la varianza

asintótica de las autocorrelaciones muestrales es mayor que 1/T. Por esta razón, en la tabla 4.2 se incluye la desviación estándar corregida por heteroscedasticidad (s.e.$_c$) calculada, según se explicó en la sección 2.2, utilizando la siguiente expresión:

$$\text{s.e.}_c = \frac{1}{\sqrt{T}} \left\{ 1 + \frac{\hat{\gamma}_2(k)}{[\hat{\gamma}(0)]^2} \right\}^{1/2} \qquad [4.1]$$

donde $\hat{\gamma}(0)$ es la varianza muestral de la propia serie R_t y $\hat{\gamma}_2(k)$ es la covarianza muestral de orden k de la serie de sus cuadrados. Obsérvese en la tabla 4.2 que la correlación de orden uno de la serie de los rendimientos es positiva y pequeña (r(1)=0.158), pero significativamente distinta de cero al 95%, cuando se compara con las bandas corregidas por heteroscedasticidad, $\{\pm 1.96\ \text{s.e.}_c\}$. Esto indica que existe cierta estructura dinámica en la media de R_t. La presencia de una primera correlación positiva y significativa en los rendimientos, es una característica común a muchas series de activos financieros; véase por ejemplo Taylor (1986), Poon y Taylor (1992), Ding, Granger y Engle (1993) y Ruiz (1994a).

6) Finalmente, en la tabla 4.2 observamos que las autocorrelaciones de las tres series "representativas" de la volatilidad (la serie de los cuadrados, la serie de los rendimientos absolutos y la serie del logaritmo de los cuadrados) toman siempre valores positivos y no muy grandes, pero significativamente distintos de cero y, además, decrecen lentamente. Para el logaritmo de los rendimientos al cuadrado, por ejemplo, la autocorrelación de orden k=10 es de magnitud similar a la correlación de orden uno, y la autocorrelación de orden k=100 es aún significativa. Estas características sugieren la existencia de memoria larga en la volatilidad del IBEX-35; véase Ding, Granger y Engle (1993).

Resumiendo, podemos decir que la serie de los rendimientos diarios del IBEX-35 en el periodo analizado presenta muchas de las características que algunos

autores consideran ya como "típicas" de las series financieras,[1] a saber: ausencia o escasa correlación en la propia serie; exceso de curtosis; varianza no constante a lo largo del tiempo, caracterizada por la alternancia de periodos de calma con periodos de mucha volatilidad; mayor correlación en los cuadrados de los rendimientos que en los rendimientos mismos; y persistencia en las volatilidades, con correlaciones que decrecen muy lentamente hacia cero en las series de los cuadrados y de los rendimientos absolutos.

Como vimos en los capítulos anteriores, existen diversos modelos SV capaces de modelizar estas características que definen la dinámica de la volatilidad de la serie de rendimientos del IBEX-35. El objetivo de este capítulo es analizar empíricamente dicha serie, viendo qué modelo se ajusta mejor a su comportamiento dinámico.

Siguiendo la práctica habitual en la literatura sobre modelos SV, vamos a "limpiar" previamente la serie de rendimientos de cualquier estructura dinámica en la media, y después vamos a modelizar la estructura dinámica de los momentos de segundo orden de la serie "filtrada".

Para ello, se ha estimado, en primer lugar, un modelo MA(1) para la serie R_t que recoja esa primera autocorrelación positiva observada en los niveles. A los residuos de este modelo se les ha ajustado un modelo ARSV(1) que recoja la dependencia observada empíricamente en la volatilidad. Si el modelo estimado fuera adecuado, los residuos estandarizados con la volatilidad estimada deberían ser un ruido blanco gaussiano de media cero y varianza uno. Sin embargo, al analizar esos residuos se ha detectado la presencia de ciertos valores que exceden 4.5 veces su desviación típica. Esos valores corresponden precisamente a los días 16/10/89 (t=688), 19-21/8/1991 (t=1141,1142,1143), 13/5/1993 (t=1576) y 4/3/1996 (t=2283), en los que, como ya hemos comentado en el punto 2, la serie R_t toma un valor excesivamente grande o excesivamente pequeño, dentro de periodos de poca volatilidad. Dichos valores son considerados como observaciones atípicas.

[1] Véase por ejemplo Bollerslev, Engle y Nelson (1994) y Ghysels, Harvey y Renault (1996).

Una vez que se han detectado las observaciones atípicas, éstas se igualan al valor medio de toda la serie sin dichas observaciones, y se trabaja ya con la serie de los rendimientos corregida por atípicos, que seguiremos denotando por R_t.

Para modelizar la estructura dinámica en la media de esta serie R_t "limpia" de atípicos, hemos ajustado un modelo MA(1) que recoja esa autocorrelación significativa de orden uno antes mencionada, obteniéndose el siguiente resultado:

$$\hat{R}_t = \underset{(0.026)}{0.052} + a_t + \underset{(0.029)}{0.168}\, a_{t-1} \qquad [4.2]$$

Los valores que aparecen entre paréntesis son las desviaciones estándar calculadas utilizando la corrección por heteroscedasticidad de White (1980). Los dos parámetros del modelo son significativos al 5%.

A partir de este momento, los residuos del modelo [4.2] serán la serie objeto de nuestro análisis, que denotaremos por y_t.[2] Por tanto, de aquí en adelante, cuando mencionemos la serie de rendimientos del índice nos estaremos refiriendo ya a esta serie y_t. Antes de entrar propiamente en la modelización de la volatilidad de esta serie, describimos a continuación algunas de sus características más importantes.

La tabla 4.3 recoge los principales momentos muestrales de la serie y_t junto con el valor del estadístico de Jarque-Bera para contrastar normalidad en dicha serie. Obsérvese cómo han disminuido los valores de los coeficientes de curstosis y de asimetría de y_t comparados con los de la serie original R_t, aunque éstos siguen siendo significativos. En la tabla 4.4 se presentan, para ciertos retardos, las autocorrelaciones muestrales de la serie y_t y de las series y_t^2, $|y_t|$, $\ln(y_t^2)$. La tabla 4.5 presenta, en la primera fila, los valores del estadístico Q(K) de Box-Ljung para contrastar la significación conjunta de las K primeras autocorrelaciones de la serie y_t, y en la

[2] La modelización de los atípicos y de la estructura dinámica en la media de los rendimientos, se podría realizar estimando simultáneamente un modelo MA(1) junto con seis variables ficticias, una para cada atípico, y trabajando después con los residuos resultantes de este ajuste. Los resultados obtenidos de esta forma son muy similares a los que se presentan aquí.

segunda fila, los valores del estadístico de Box-Pierce corregido por heterocedasticidad, $Q_c(K)$. Los valores de este estadístico se obtienen utilizando la formulación que se presentó en la sección 2.2 del capítulo 2. Todos los valores se acompañan del p-valor correspondiente a una distribución χ^2_{K-1}. Las tres últimas filas de la tabla 4.5 recogen los valores del estadístico de Box-Ljung para contrastar incorrelación en la serie de los cuadrados, en la serie de los valores absolutos, y en la serie del logaritmo de los cuadrados. Estos tres estadísticos se denotan por $Q_2(K)$, $Q_A(K)$ y $Q_{ln}(K)$, respectivamente. Junto a ellos aparecen los p-valores de una distribución χ^2_K; véase McLeod y Li (1983) para una justificación de esta distribución.

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Tabla 4.3. Momentos muestrales de la serie y_t

Media	Desviación típica	Asimetría	Curtosis	Mínimo	Máximo	Contraste de Normalidad
0.0000	1.1985	-0.2740*	8.3205*	-8.0925	8.2082	3565.28*

* Valores significativos al 5%.

Tabla 4.4. Autocorrelaciones muestrales de las series y_t, y_t^2, $|y_t|$ y $\ln(y_t^2)$.

Retardo Serie	k=1	k=2	k=3	k=4	k=5	k=10	k=20	k=50	k=100
y_t	0.002	0.011	-0.010	0.033	-0.007	0.060*	-0.049*	-0.018	-0.011
s.e.$_c$	[0.030]	[0.034]	[0.028]	[0.033]	[0.030]	[0.029]	[0.025]	[0.022]	[0.021]
y_t^2	0.218*	0.328*	0.188*	0.307*	0.234*	0.210*	0.117*	0.053*	0.041*
$\|y_t\|$	0.262*	0.302*	0.266*	0.256*	0.279*	0.253*	0.195*	0.114*	0.093*
$\ln(y_t^2)$	0.131*	0.183*	0.172*	0.107*	0.142*	0.129*	0.118*	0.091*	0.077*

* Valores significativos al 5% respecto a la distribución N(0,1/T).

Nota: Los valores s.e.$_c$ de esta tabla denotan el error estándar de las correlaciones muestrales corregido por volatilidad según la expresión [4.1].

Tabla 4.5. Estadístico de Box-Ljung para la serie y_t y las series y_t^2, $|y_t|$ y $\ln(y_t^2)$.

Retardo / Estadístico	K=5	K=10	K=20	K=50	K=100
Q(K)	4.032	20.267*	45.782*	82.375*	135.65*
	(0.402)	(0.016)	(0.001)	(0.002)	(0.008)
$Q_c(K)$	1.276	7.691	20.023	42.765	86.052
	(0.865)	(0.565)	(0.393)	(0.723)	(0.820)
$Q_2(K)$	1018.1*	1729.1*	2287.4*	2853.8*	3015.7*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$Q_A(K)$	1122.3*	1997.2*	3120.6*	4776.7*	5813.6*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$Q_{ln}(K)$	335.68*	609.28*	1043.2*	1936.3*	2830.9*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

* Valores significativos al 5% respecto a la correspondiente distribución χ^2.

Nota: Q(K) denota el estadístico de Box-Ljung para contrastar incorrelación hasta de orden K en y_t. $Q_c(K)$ denota el estadístico de Box-Pierce con la corrección por volatilidad, según se definió en la sección 2.2. $Q_2(K)$, $Q_A(K)$ y $Q_{ln}(K)$ denotan el estadístico de Box-Ljung para contrastar incorrelación en las series de los cuadrados, de los valores absolutos y del logaritmo de los cuadrados, respectivamente. Entre paréntesis se representa el p-valor de la distribución χ^2 correspondiente.

Como era de esperar, la serie y_t apenas presenta ya correlación en los niveles, pero sigue teniendo muchas de las características que habíamos observado en la serie original de los rendimientos del índice: mayor correlación en los cuadrados que en el nivel de la serie, y persistencia en la volatilidad, que se manifiesta con la presencia de correlaciones que decrecen lentamente hacia cero en la serie de los cuadrados, en la serie de los valores absolutos y en la serie del logaritmo de los cuadrados.

A pesar de la escasa correlación que existe en y_t, algunos resultados de las tablas 4.4 y 4.5 parecen indicar que el modelo MA(1) que hemos ajustado no ha eliminado completamente la correlación en los niveles. Sin embargo, conviene recordar que, en presencia de heteroscedasticidad condicionada, el nivel de significación nominal de los contrastes de incorrelación habituales es muy engañoso, según vimos en la sección 2.2, y, por tanto, los valores de los estadísticos de dichos contrastes deben interpretarse con cautela. En efecto, si se utiliza el error estándar de

las autocorrelaciones muestrales corregido por heteroscedasticidad, s.e.$_c$, y se aplica el estadístico de Box-Pierce corregido, $Q_c(K)$, se comprueba que las correlaciones de la serie y_t en la tabla 4.4 no son significativas (ninguna excede, en valor absoluto, 1.96 veces la desviación típica corregida s.e.$_c$) y los valores del estadístico $Q_c(K)$ en la tabla 4.5 en ningún caso rechazan la hipótesis de incorrelación en los niveles. Por tanto, podemos concluir que la serie y_t está incorrelacionada.

En la tabla 4.4 también podemos observar que al eliminar las observaciones atípicas y modelizar la estructura dinámica en la media, los valores de las correlaciones de las series y_t^2, $|y_t|$ y $\ln(y_t^2)$, se han incrementado notablemente con respecto a los valores correspondientes que aparecen en la tabla 4.2.

Muchas de las características que acabamos de describir se aprecian igualmente en el gráfico 4.3. Este gráfico representa, de izquierda a derecha y de arriba hacia abajo, los correlogramas de las cuatro series analizadas: y_t, y_t^2, $|y_t|$ y $\ln(y_t^2)$. El correlograma de la serie y_t se representa con las bandas de confianza al 95% corregidas por heteroscedastidad, $\{\pm 1.96\,\text{s.e.}_c\}$.

En el gráfico 4.3 se comprueba, efectivamente, que utilizando las bandas corregidas por volatilidad ya no se rechaza la hipótesis de incorrelación en la serie y_t. Además, este gráfico muestra claramente la existencia de correlación en la serie de los cuadrados, ya que las correlaciones de esta serie son significativamente positivas, incluso hasta de orden k=50. Esta persistencia en las correlaciones se agudiza en las series $|y_t|$ y $\ln(y_t^2)$, que presentan valores positivos y muy significativos, incluso para los retardos más alejados, y muestran un decrecimiento lento hacia cero, especialmente notable en la serie del logaritmo de los cuadrados. Obsérvese que las correlaciones muestrales de esta última serie son pequeñas en magnitud pero apenas decrecen con k (la correlación de orden k=100 es aún 0.077).

Gráfico 4.3. Correlogramas para diversas transformaciones de la serie de rendimientos del IBEX35 durante el periodo 7/1/87-30/12/98: (a) rendimientos; (b) rendimientos cuadrados; (c) rendimientos absolutos; (d) logaritmo de los rendimientos cuadrados.

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Este comportamiento de las correlaciones de las series y_t^2, $|y_t|$ y $\ln(y_t^2)$ sugiere que la volatilidad del IBEX-35 podría tener la propiedad de la memoria larga o podría ser no estacionaria.

Para contrastar la hipótesis de memoria larga en la volatilidad del IBEX-35, hemos aplicado a las series y_t^2 y $\ln(y_t^2)$ algunos de los contrastes de memoria larga descritos en la sección 1.2. En concreto, hemos utilizado el contraste de Geweke y Porter-Hudak (1983), que denotamos GPH, el contraste R/S de Lo (1991), y el contraste de los multiplicadores de Lagrange (LM) de Lobato y Robinson (1998). En nuestro caso, estos contrastes deben interpretarse como contrastes de memoria corta (H_0:d=0) frente a alternativas muy generales que no especifican ninguna forma concreta de memoria larga. En Breidt, Crato y de Lima (1998), Lobato y Savin (1998) y Olmeda (1998) pueden verse aplicaciones similares de estos contrastes para detectar la existencia de memoria larga en las volatilidades de los rendimientos diarios de distintos activos financieros.

Como vimos en el capítulo 1, el contraste GPH es un contraste de significación sobre el parámetro d, con una distribución t-Student, que se deriva de la estimación de un modelo de regresión lineal para el logaritmo del periodograma realizado sólo sobre las primeras m ordenadas. Como no existe una regla clara para elegir m, utilizaremos los valores habituales, $m=[T^{\alpha}]$, con α=0.45, 0.5, 0.6, donde T denota el tamaño muestral. Para $m=[T^{0.5}]$, que es el truncado óptimo propuesto por GPH, repetiremos la estimación eliminando además las frecuencias más bajas, hasta $l=[T^{0.1}]$, como en Breidt, Crato y de Lima (1998). No obstante, Deo y Hurvich (1998) prueban que en modelos LMSV, truncar inferiormente empeora las propiedades del estimador GPH.

El contraste R/S de Lo (1991) es un contraste de memoria corta basado en un estadístico de la forma $V_T(q)=Q_T(q)/\sqrt{T}$, donde q es un parámetro de truncamiento y $Q_T(q)$ es el estadístico de rango reescalado definido en [1.46]. Ante la falta de un criterio claro para elegir el valor de q, hemos seleccionado q=25, 50 y 100, y q=q*,

siendo $q^*=\left(\frac{3T}{2}\right)^{1/3}\left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{2/3}$ el retardo de truncado óptimo de Andrews (1991) para modelos AR(1) con parámetro autorregresivo ρ. En este último caso, las ponderaciones que utilizamos para construir el estadístico $Q_T(q)$ no son las habituales de Newey-West, sino las del propio Andrews, según se indica en Lo (1991, p.1302).

Por último, el contraste semiparamétrico de Lobato y Robinson (1998) utiliza un estadístico de los multiplicadores de Lagrange, de la forma $LM=t^2$, donde el estadístico t, definido en la sección 1.2.3, depende de las primeras m ordenadas del periodograma, siendo $0<m<T/2$. Siguiendo las pautas de Lobato y Savin (1998), los valores de m que hemos elegido son m=30, 50, 70, 100.

La tabla 4.6 recoge los resultados de estos contrastes para las series y_t^2 y $\ln(y_t^2)$.

Tabla 4.6. Contrastes unilaterales de memoria larga (H_0:d=0 *vs* H_1:d>0) para la volatilidad del IBEX-35

Contraste GPH (1983)								
Serie	y_t^2				$\ln(y_t^2)$			
m= l =	$T^{0.45}$	$T^{0.5}$	$T^{0.6}$	$T^{0.5}$ $T^{0.1}$	$T^{0.45}$	$T^{0.5}$	$T^{0.6}$	$T^{0.5}$ $T^{0.1}$
$\hat{d}_{GPH}$	0.2818	0.3802	0.3992	0.3725	0.6241	0.5466	0.4574	0.5546
t	2.8615	4.5522	7.7639	3.5086	6.1070	6.0224	6.5213	4.8630
p-valor	0.0021	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000
Contraste R/S de Lo (1991)								
Serie	y_t^2				$\ln(y_t^2)$			
q=	q=25	q=50	q=100	q=q*=9	q=25	q=50	q=100	q=q*=9
$V_T(q)$	2.2533	1.8780	1.6247	3.0800	2.5711	2.0263	1.5795	3.9117
p-valor	0.0015	0.0227	0.0974	0.0000	0.0001	0.0084	0.1222	0.0000
Contraste LM de Robinson y Lobato (1998)								
Serie	y_t^2				$\ln(y_t^2)$			
m=	m=30	m=50	m=70	m=100	m=30	m=50	m=70	m=100
t	2.6764	4.8610	6.2038	8.6390	5.6971	9.0527	10.747	14.035
p-valor	0.0037	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

A juzgar por los resultados de esta tabla, la evidencia en contra de la memoria corta en ambas series parece incuestionable, si bien es algo más clara en la serie $\ln(y_t^2)$ que en la serie y_t^2. Esto confirma las características de los correlogramas en el gráfico 4.3. Para la serie $\ln(y_t^2)$, el procedimiento GPH estima siempre $\hat{d}>1/2$ y, por tanto, podría estarse detectando no estacionariedad en lugar de memoria larga. Por lo que respecta al contraste R/S, el estadístico $V_T(q^*)$ computado con los pesos de Andrews (1991), es significativo para ambas series. Sin embargo, la significación de $V_T(q)$ disminuye a medida que aumenta el valor de q, en consonancia con los resultados de Lo (1991). Obsérvense por ejemplo los p-valores correspondientes a q=100. El contraste LM unilateral de Lobato y Robinson (1998) corrobora los resultados anteriores, rechazando siempre la hipótesis nula H_0:d=0 en favor de la alternativa de memoria larga H_1:d>0. Al igual que en los casos anteriores, la evidencia de memoria larga es mucho más clara en la serie $\ln(y_t^2)$ que en la serie y_t^2.

A pesar de que parece evidente que existe memoria larga en la volatilidad del IBEX-35, los resultados de los contrastes GPH, R/S y LM que acabamos de presentar deben interpretarse con cierta cautela, porque sus propiedades en modelos con memoria larga para la volatilidad apenas se conocen, y pudieran estar detectando raíces unitarias. Deo y Hurvich (1998) estudian las propiedades asintóticas y en muestras finitas del estimador GPH de d en modelos LMSV, y demuestran que tiene siempre un sesgo negativo, que se incrementa cuando el cociente de las varianzas señal-ruido, σ_h^2/σ_ξ^2, disminuye y/o cuando aumenta el número de ordenadas m que se utilizan en la regresión. Breidt, Crato y de Lima (1998) estudian las propiedades en muestras finitas de los contrastes GPH y R/S en modelos LMSV y concluyen que estos contrastes detectan bien las alternativas de memoria larga muy persistentes, en las que el parámetro d está muy próximo a 1/2 y la varianza σ_η^2 no es extremadamente pequeña. Sin embargo, estos autores comprueban que el contraste GPH con valores grandes de m puede proporcionar evidencia espuria de memoria larga en modelos SV, cuando el logaritmo de la volatilidad es un proceso próximo a la raíz unitaria.

Aún siendo conscientes de estas limitaciones, los resultados de los contrastes que aparecen en la tabla 4.6 confirman las características muestrales observadas en los correlogramas de las series y_t^2 y $\ln(y_t^2)$, y proporcionan razones suficientes para pensar que los modelos SV con raíces unitarias o los modelos SV con memoria larga, descritos en las secciones 1.1 y 1.3, respectivamente, pueden ser apropiados para modelizar la volatilidad de los rendimientos diarios del índice IBEX-35.

4.2. Estimación de modelos SV para la volatilidad del IBEX-35

BIBLIOTECA VIRTUAL

En esta sección se presentan los resultados de las estimaciones de diversos modelos SV para la serie de los rendimientos diarios del IBEX-35 en el periodo comprendido entre el 7/1/1987 y el 30/12/1998. Como ya comentamos en la sección anterior, la serie que vamos a modelizar no es la serie original de los rendimientos del índice, sino la serie de los residuos, y_t, que se obtiene al ajustar a los rendimientos del índice corregidos por atípicos el modelo MA(1) en [4.2]. Según hemos visto en la sección anterior, la serie y_t se caracteriza por la ausencia de correlación en los niveles, el exceso de curtosis y la persistencia en las correlaciones de la serie de los cuadrados y de la serie del logaritmo de los cuadrados.

Entre los modelos SV capaces de reproducir este comportamiento hemos estimado, en primer lugar, un modelo ARSV(1) de la forma:

$$y_t = \sigma_* \exp(\tfrac{1}{2} h_t)\varepsilon_t \qquad [4.3a]$$

$$h_t = \phi\, h_{t-1} + \eta_t \qquad [4.3b]$$

donde σ_* es un factor de escala, $|\phi|<1$, ε_t es NID(0,1) y η_t es NID$(0,\sigma_\eta^2)$ e independiente de ε_t. Los resultados de la estimación de este modelo (con la desviación típica entre paréntesis) han sido los siguientes:

$$\hat{\phi} = \underset{(0.0057)}{0.9898}\,,\ \hat{\sigma}_{\eta}^{2} = \underset{(0.0042)}{0.0168}\,,\ \hat{\sigma}_{*} = \underset{(0.0374)}{0.9297} \qquad [4.4]$$

Los parámetros $(\phi, \sigma_{\eta}^{2})$ se han estimado utilizando el método de *pseudo*-máxima verosimilitud (PMV) propuesto por Harvey, Ruiz y Shephard (1994), aplicado a la serie y_t estandarizada con su media y su desviación típica muestral. La desviación estándar de los estimadores en [4.4] se ha obtenido aplicando los resultados de Ruiz (1994b). El parámetro de escala, σ_*, se ha estimado con la desviación típica muestral de las observaciones corregidas por heteroscedasticidad, $\tilde{y}_t = y_t \exp(-\frac{1}{2}\tilde{h}_{t/T})$, donde $\tilde{h}_{t/T}$ es la estimación suavizada del proceso h_t. La desviación estándar del estimador $\hat{\sigma}_*$ se ha calculado utilizando los resultados de Harvey y Shephard (1993).

En la primera columna de la tabla 4.7 aparecen los principales momentos muestrales de las observaciones estandarizadas para el modelo ARSV(1), obtenidas éstas como $\hat{\varepsilon}_t = y_t / \hat{\sigma}_{t/T}$, para t=1,2,…,T, donde $\hat{\sigma}_{t/T}$ es la estimación suavizada de la volatilidad de y_t, calculada como:

$$\hat{\sigma}_{t/T} = \hat{\sigma}_{*} \exp(\tfrac{1}{2}\tilde{h}_{t/T}) \qquad [4.5]$$

según se explicó en el capítulo 3. La tabla 4.7 recoge, además, los valores de otros estadísticos importantes, como son las correlaciones muestrales de las observaciones estandarizadas y el estadístico de Box-Ljung para contrastar incorrelación hasta de orden K en dicha serie. Finalmente, la tabla 4.7 presenta los valores de las correlaciones muestrales de la serie de los cuadrados de las observaciones estandarizadas, y los valores del estadístico de Box-Ljung para contrastar incorrelación hasta de orden K en dicha serie. Estos últimos se denotan por $Q_2(K)$.

Tabla 4.7. Momentos muestrales de los rendimientos del índice IBEX-35 estandarizados con las estimaciones suavizadas de la volatilidad para los diferentes modelos SV estimados.

Modelo	**ARSV(1)**	**RWSV**	**LMSV**	**ARLMSV**
Parámetros estimados	$\hat{\sigma}_{\eta}^2$ =0.0168 $\hat{\phi}$ =0.9898 $\hat{\sigma}_*$ =0.9297	$\hat{\sigma}_{\eta}^2$ =0.0099 $\hat{\sigma}_*$ =0.9484	$\hat{\sigma}_{\eta}^2$ =0.0906 $\hat{d}$ =0.7538 $\hat{\sigma}_*$ =1.5112	$\hat{\sigma}_{\eta}^2$ =0.0155 $\hat{\phi}$ =0.6632 $\hat{d}$ =0.7035 $\hat{\sigma}_*$ =1.5485
Media	0.0101	0.0074	0.0113	0.0122
Varianza	1.0000	1.0000	0.9999	0.9999
Asimetría	-0.0284	-0.0441	-0.0183	-0.0127
Curtosis	3.5620**	3.7070**	3.3941**	3.4380**
Correlaciones				
1	0.0097	0.0097	0.0105	0.0104
2	0.0320	0.3280	0.0305	0.0307
3	-0.0071	-0.0055	-0.0084	-0.0088
4	0.0270	0.2530	0.0275	0.0277
5	-0.0061	-0.0067	-0.0036	-0.0032
10	0.0376	0.0388	0.0363	0.0362
50	0.0002	-0.0017	0.0027	0.0033
100	0.0060	0.0061	0.0076	0.0075
Q(10)	13.00	13.64	12.60	12.56
Q(50)	41.75	41.69	42.39	42.57
Q(100)	89.50	88.88	91.17	91.53
Correlaciones de los cuadrados				
1	0.0709**	0.0894**	0.0419*	0.0455*
2	0.0676**	0.0923**	0.0378*	0.0320
3	0.0306	0.0392*	0.0217	0.0173
4	0.0342	0.0542**	0.0129	0.0053
5	0.0335	0.0464*	0.0290	0.0231
10	0.0206	0.0258	0.0225	0.0210
50	0.0011	0.0011	-0.0011	-0.0017
100	0.0300	0.0320	0.0256	0.0249
Q_2(10)	43.15**	80.06**	17.89	15.50
Q_2(50)	101.70**	144.61**	71.37*	68.43*
Q_2(100)	150.41**	198.87**	117.67	113.39

** Valores significativos al 1%. * Valores significativos al 5%

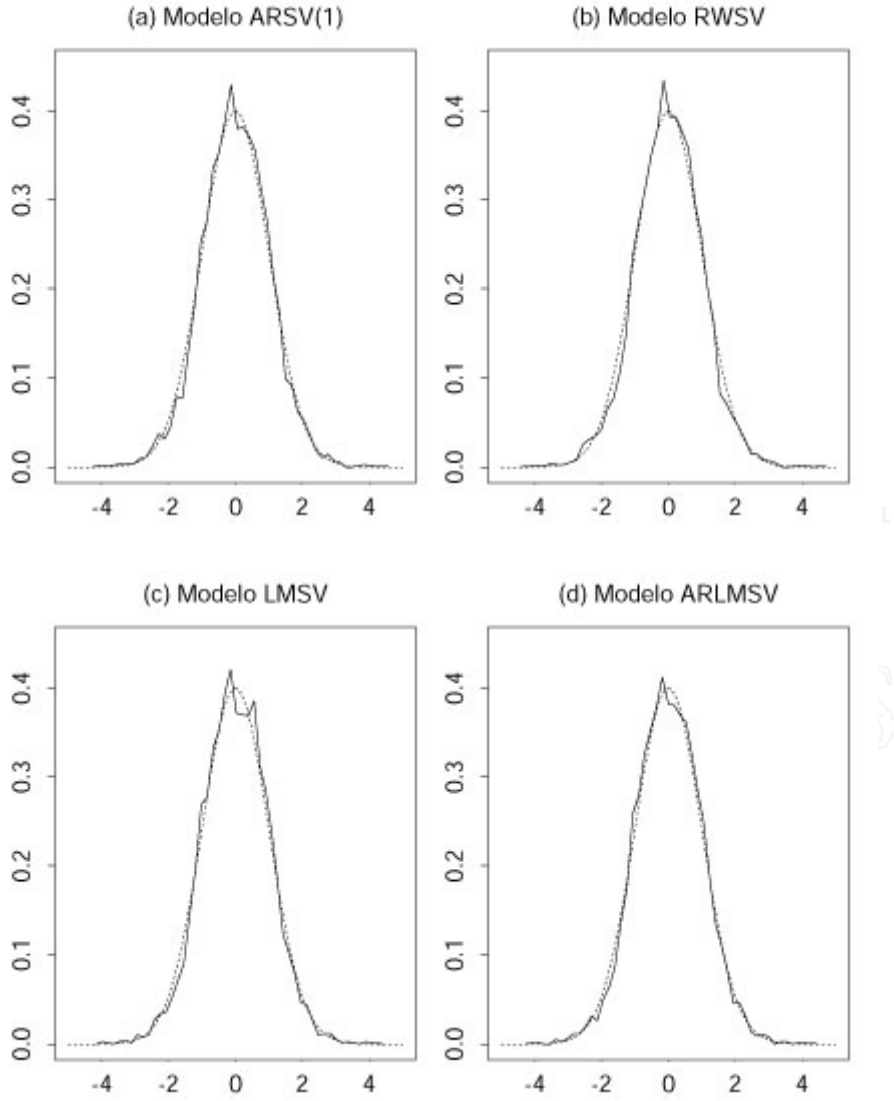
Nota: Q(K) denota el estadístico de Box-Ljung para contrastar incorrelación hasta de orden K en las observaciones estandarizadas; Q_2(K) denota el estadístico de Box-Ljung para contrastar incorrelación hasta de orden K en los cuadrados de las observaciones estandarizadas.

Si nos fijamos en los resultados de la estimación del modelo ARSV(1) de la tabla 4.7, observamos, en primer lugar, que los valores del estadístico Q(K) para contrastar la existencia de autocorrelación en la media condicionada ya no son significativos comparados con la distribución χ^2_K, y tampoco lo es, a nivel individual, ninguna de las correlaciones de dicha serie. Sin embargo, las dos primeras autocorrelaciones de la serie de los cuadrados de las observaciones estandarizadas para el modelo ARSV(1), aún son significativas, aunque eso sí, son menores que en la serie original (compárense estos valores con las correlaciones de y_t^2 en la tabla 4.4). Lo mismo puede decirse de los valores del estadístico de Box-Ljung para la serie de los cuadrados, $Q_2(K)$, que han disminuido considerablemente con respecto a los valores de la serie original en la tabla 4.5. Sin embargo, los valores de $Q_2(K)$ siguen siendo significativos y, por tanto, nos están indicando que aún existe cierto grado de correlación en los momentos de segundo orden que el modelo ARSV(1) no ha sido capaz de modelizar.

Otro aspecto destacable de los resultados que aparecen en la tabla 4.7 es que los valores de los coeficientes de asimetría y curtosis de las observaciones estandarizadas se han reducido notablemente, en comparación con los valores correspondientes de la tabla 4.3, aunque el coeficiente de curtosis en el modelo ARSV(1) sigue siendo significativamente mayor que 3. Esto significa que los residuos del ajuste aún tienen exceso de curtosis.

El gráfico 4.4(a) representa la densidad estimada de las observaciones estandarizadas para el modelo ARSV(1) junto con la densidad normal estándar. En general, parece que la distribución normal se ajusta bastante bien a la distribución empírica, excepto en la parte central, donde existe una concentración de valores alrededor del cero superior a la que se esperaría si $\hat{\varepsilon}_t$ siguiera una distribución normal. También se observa que las colas de la distribución empírica son ligeramente más anchas que las de la distribución normal, confirmándose así el exceso de curtosis en las observaciones estandarizadas del modelo ARSV(1) detectado previamente.

Gráfico 4.4. Densidad estimada de las observaciones estandarizadas para los diferentes modelos SV estimados junto con la densidad normal estándar.



Como era de esperar, el parámetro autorregresivo ϕ del modelo ARSV(1) se ha estimado muy próximo a uno, indicando una fuerte persistencia de los *shocks* en la volatilidad del IBEX-35. En consecuencia, hemos ajustado a la serie diaria de los rendimientos del índice un modelo RWSV, que incluye de forma específica una raíz unitaria (ϕ=1) en la ecuación de la volatilidad [4.3b], según vimos en la sección 1.1.2 del capítulo 1. Los resultados de la estimación PMV de este modelo han sido:

$$\hat{\sigma}_{\eta}^{2} = \underset{(0.0024)}{0.0099}, \hat{\sigma}_{*} = \underset{(0.0436)}{0.9484} \qquad [4.6]$$

Los valores de la desviación estándar de los estimadores se han obtenido nuevamente utilizando los resultados de Harvey y Shephard (1993) y Ruiz (1994b).

Los resultados de la estimación de este modelo, junto con los principales momentos muestrales de las observaciones estandarizadas correspondientes, aparecen en la segunda columna de la tabla 4.7. Los resultados de esta tabla muestran, en primer lugar, que los ajustes de los modelos ARSV(1) y RWSV son muy similares. Este resultado es lógico debido a que el valor estimado de ϕ en el modelo ARSV(1) está muy próximo a la unidad. Sin embargo, al restringir ϕ=1 en el modelo RWSV, se pierde algo de ajuste (como era de esperar), y aparecen autocorrelaciones significativas en los cuadrados de las observaciones estandarizadas, lo que podría estar indicando que un modelo estacionario es más adecuado.

En la tabla 4.7 observamos también que, al igual que ocurría con el modelo autorregresivo, el coeficiente de curtosis de las observaciones estandarizadas se ha reducido considerablemente con respecto al que se tenía en la serie original, aunque aún sigue siendo significativo, lo que indica que los residuos del ajuste siguen teniendo exceso de curtosis. Obsérvese la densidad estimada de las observaciones estandarizadas para el modelo RWSV en el gráfico 4.4(b).

Por lo que respecta a los modelos con memoria larga, en primer lugar hemos estimado un modelo LMSV básico estacionario de la forma:

$$y_t = \sigma_* \exp(\tfrac{1}{2} h_t)\, \varepsilon_t \qquad [4.7a]$$
$$(1-L)^d h_t = \eta_t \qquad [4.7b]$$

donde $0<d<1/2$ y las perturbaciones ε_t y η_t son como en los casos anteriores. Los parámetros de este modelo, (d, σ_η^2), se estiman por *pseudo*-máxima verosimilitud en el dominio de las frecuencias, según se explicó en el capítulo 3.

Al ajustar el modelo [4.7] a los rendimientos diarios del IBEX-35, resulta que la estimación del parámetro d se obtiene en la frontera de la no estacionariedad ($\hat{d}=1/2$). Según Harvey (1998), este resultado puede indicar que sería más apropiado utilizar un modelo no estacionario ($d>1/2$), aunque también cabría pensar que el hecho de haber estimado d en la frontera de la no estacionariedad es sencillamente una consecuencia de las propiedades de los estimadores PMV en modelos con memoria larga. Baste recordar aquí que en modelos LMSV estacionarios y con poca varianza ($\sigma_\eta^2 \approx 0$), la distribución del estimador PMV de d tiene una concentración de masa en torno a 1/2 que hace muy probable estimar erróneamente $\hat{d}=1/2$; véase la sección 3.3 del capítulo 3.

En cualquier caso, hemos ajustado a la serie de rendimientos del IBEX-35 un modelo LMSV no estacionario de la forma [4.7], pero con $d>1/2$, y hemos obtenido las siguientes estimaciones:

$$\hat{d}=0.7538,\ \hat{\sigma}_\eta^2=0.0906,\ \hat{\sigma}_*=1.5112 \qquad [4.8]$$

Los parámetros (d, σ_η^2) se han estimado por *pseudo*-máxima verosimilitud en el dominio de las frecuencias a partir del modelo lineal transformado en primeras diferencias, según se explicó en la sección 3.2.2. El parámetro σ_* se ha estimado con la desviación típica muestral de las observaciones corregidas por heteroscedasticidad, $\tilde{y}_t = y_t \exp(-\tfrac{1}{2}\tilde{h}_{t/T})$, donde $\tilde{h}_{t/T}$ es la estimación suavizada del proceso h_t, obtenida con el algoritmo de suavizado propuesto por Harvey (1998), que se describió en la

sección 3.5. Como vimos en dicha sección, este algoritmo presenta un problema de tipo práctico en muestras grandes, ya que al programar las ecuaciones del suavizado para h_t es necesario guardar e invertir matrices de orden TxT. En nuestro caso, el tamaño muestral T=2991 es lo suficientemente grande como para que el algoritmo sea impracticable. Para solventar este problema, hemos realizado el suavizado utilizando unos pesos calculados con una muestra más pequeña, digamos de tamaño T*, con T*<T, y asignando pesos iguales a cero a las observaciones más distantes, desde T* hasta T. Según vimos en la sección 3.5, con este procedimiento apenas se pierde precisión en la estimación y los resultados son muy fiables. En nuestro caso, los pesos se calcularon sobre una muestra de tamaño T*=1201.

La tercera columna de la tabla 4.7 recoge los principales momentos muestrales de la serie de las observaciones estandarizadas, $\hat{\varepsilon}_t = y_t / \hat{\sigma}_{t/T}$, para el modelo LMSV, junto con las correlaciones muestrales de la serie de sus cuadrados.

Los aspectos más destacables de los resultados de la estimación del modelo LMSV son los siguientes. En primer lugar, observamos que los momentos muestrales de los residuos estandarizados son muy similares a los obtenidos con los modelos ARSV(1) y RWSV, aunque el modelo LMSV ha conseguido reducir aún más el exceso de curtosis. Pero lo más importante, sin duda alguna, es la reducción que se ha producido en los valores de las correlaciones muestrales de la serie de los cuadrados (compárense con las correlaciones originales de y_t^2 en la tabla 4.4). Ninguna de estas autocorrelaciones es significativa al 1%, comparada con las bandas de Barlett correspondientes, y únicamente las dos primeras autocorrelaciones son significativamente positivas, si el contraste se realiza con un nivel de significación del 5%. Este comportamiento de las correlaciones se manifiesta igualmente en los valores del estadístico de Box-Ljung para la serie de los cuadrados: los valores de $Q_2(K)$ son muchísimo menores que los que obtuvimos con los modelos ARSV(1) y RWSV, y ya no son significativos al 1%. El único valor significativo al 5% es $Q_2(20)$, que rechazaría la hipótesis de incorrelación conjunta en la serie de los cuadrados para los 20 primeros retardos. Sin embargo, este valor dejaría de ser

significativo si el contraste se realizara con un nivel de significación del 1%. Queda claro entonces que el modelo LMSV modeliza la estructura dinámica de la volatilidad mejor que cualquiera de los dos modelos anteriores, especialmente en los retardos más alejados. Esto es razonable porque precisamente la propiedad de la memoria larga se manifiesta en estos retardos.

Finalmente, hemos ajustado a los rendimientos diarios del IBEX-35 un modelo SV autorregresivo con memoria larga (ARLMSV) de la forma:

$$y_t = \sigma_* \exp(\tfrac{1}{2} h_t)\, \varepsilon_t \qquad [4.9a]$$

$$(1-\phi L)(1-L)^d h_t = \eta_t \qquad [4.9b]$$

con $|\phi|<1$ y $0<d<1/2$. Este modelo engloba como casos particulares los tres modelos anteriores. El ajuste del modelo [4.9] produce nuevamente una estimación del parámetro d en la frontera de la no estacionariedad. En concreto se obtiene $\hat{\phi}=0.8722$, $\hat{d}=0.5$, $\hat{\sigma}_\eta^2=0.0081$. Por esta razón, hemos repetido la estimación considerando un modelo ARLMSV no estacionario con $d>1/2$. Los resultados de la estimación PMV de este modelo han sido los siguientes:

$$\hat{\phi}=0.6632,\ \hat{d}=0.7035,\ \hat{\sigma}_\eta^2=0.0155,\ \hat{\sigma}_*=1.5485 \qquad [4.10]$$

Al igual que en el caso anterior, el parámetro σ_* se ha estimado con la desviación típica muestral de las observaciones corregidas por heteroscedasticidad, utilizando para ello el algoritmo de suavizado propuesto por Harvey (1998), con pesos calculados sobre una muestra de tamaño $T^*=1201$.

Obsérvese que al presentar los resultados de la estimación del modelo LMSV y el modelo ARLMSV, en ninguno de ellos se proporciona el valor de la desviación estándar de los estimadores. Esto se debe a que aún no se conoce la distribución asintótica de los estimadores PMV en el dominio de las frecuencias para los modelos SV con memoria larga.

Si comparamos los resultados obtenidos en los dos modelos con memoria larga, comprobamos que ambos proporcionan estimaciones de d y σ_* muy similares, aunque la introducción del parámetro ϕ en el modelo ARLMSV ha reducido considerablemente el valor estimado de σ_η^2.

La cuarta columna de la tabla 4.7 recoge los principales momentos muestrales de las observaciones estandarizadas correspondientes al modelo ARLMSV. Éstas presentan características similares a las del modelo LMSV, aunque ahora parece que se ha corregido aún más la correlación en los cuadrados. Además, en el gráfico 4.4 parece que el ajuste de la distribución de las observaciones estandarizadas del modelo ARLMSV a la distribución normal es ligeramente mejor que el del modelo LMSV, aunque las diferencias son muy leves.

En general, los resultados presentados hasta este momento parecen indicar que los modelos SV con memoria larga proporcionan mejores ajustes que los modelos ARSV(1) y RWSV. Sin embargo, a partir de los resultados obtenidos resulta difícil distinguir entre los dos modelos con memoria larga ajustados.

Una posible forma de elegir entre ambos modelos sería realizar un contraste de hipótesis que nos permita evaluar la significación del componente autorregresivo en el modelo ARLMSV. Recuérdese que el modelo LMSV es un caso particular del modelo ARLMSV, con ϕ=0 fijo, y por tanto, el contraste de significación sobre ϕ podría plantearse como un contraste de razón de verosimilitud de la hipótesis H_0:LMSV(ϕ=0) frente a la alternativa H_1:ARLMSV($\phi\neq0$). En nuestro caso, el valor del estadístico de este contraste es LR=2[ln($\hat{\psi}$)-ln($\hat{\psi}_0$)]=1.0885, que no es significativo con respecto a la distribución χ_1^2. En consecuencia, la hipótesis nula no se rechaza y nos decantaríamos por el modelo LMSV. Una vez más, conviene señalar que este resultado debe interpretarse con cautela, porque la distribución del estadístico LR no es conocida en modelos de volatilidad estocástica y aquí estamos asumiendo la teoría asintótica estándar.

Finalmente, el gráfico 4.5 representa la serie de los rendimientos absolutos del índice IBEX-35 junto con la estimación suavizada de la volatilidad (en negrita) de los rendimientos del índice, $\hat{\sigma}_{t/T}$, obtenida con cada uno de los cuatro modelos estimados previamente. En este gráfico observamos que el comportamiento general de la volatilidad estimada es muy similar para los cuatro modelos, y reproduce adecuadamente los cambios que experimentan los rendimientos del IBEX-35 en el periodo analizado. La volatilidad comienza con niveles bastante altos, que corresponden al periodo de incertidumbre asociado a la crisis bursátil de 1987; a este periodo le sigue un periodo de calma, hasta el comienzo de la crisis del Golfo en agosto de 1990 (aproximadamente en torno a t=890), donde la volatilidad vuelve a aumentar considerablemente; después sigue un periodo sin grandes cambios, en el que la serie presenta sucesivas subidas y bajadas, no muy bruscas; al final de la muestra, la volatilidad se incrementa y experimenta un aumento espectacular de nivel que corresponde al momento de mayor volatilidad del índice, debido a las crisis financieras asiática y rusa durante los meses del verano de 1998. Estos movimientos en la volatilidad del IBEX-35 son precisamente los movimientos que ya habíamos observado en la propia serie de rendimientos representada en el gráfico 4.2.

Si comparamos en el gráfico 4.5 unos modelos con otros, observamos, en primer lugar, que apenas existe diferencia entre la volatilidad estimada por los modelos ARSV(1) y RWSV, siendo esta última algo más suave. Este resultado era de esperar dada la similitud en las estimaciones de los parámetros de ambos modelos. Por lo que respecta a las volatilidades estimadas por los modelos con memoria larga, representados en los gráficos 4.5(c) y 4.5(d), podemos decir que éstas reproducen fielmente la dinámica de los rendimientos del IBEX-35, especialmente en los momentos de mayor volatilidad, donde el ajuste que proporcionan estos modelos es claramente superior al obtenido con los modelos ARSV(1) y RWSV. Aunque resulta difícil de apreciar en este gráfico, la volatilidad estimada por el modelo ARLMSV es algo más suave que la del modelo LMSV, debido a que el valor estimado de σ_η^2 en el primero ($\hat{\sigma}_\eta^2$=0.0155) es más pequeño que en el modelo LMSV ($\hat{\sigma}_\eta^2$=0.0906).

Gráfico 4.5. Estimaciones suavizadas de la volatilidad del IBEX-35 para diferentes modelos SV junto con los rendimientos absolutos diarios del índice.

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Otra forma posible de diagnosticar la capacidad de los modelos estimados para reproducir las propiedades dinámicas de los rendimientos del índice, sería comparar los momentos muestrales de la propia serie de rendimientos con los momentos poblacionales implícitos de y_t para los modelos estimados. El inconveniente de este diagnóstico es que los momentos implícitos sólo están definidos para aquellos modelos que son estacionarios, y en nuestro caso, sólo uno de los modelos estimados lo es: el modelo ARSV(1). Esto significa que la comparación de los momentos muestrales y poblacionales de y_t sólo es posible para este modelo.

La tabla 4.8 compara los valores de la varianza muestral y el coeficiente de curtosis de la serie y_t y los valores de las correlaciones muestrales de la serie de sus cuadrados, con los valores teóricos correspondientes que se obtendrían con el modelo ARSV(1) de parámetros [4.4]. Las fórmulas para calcular estos momentos teóricos se dedujeron en la sección 1.1.2 del capítulo 1. Concretamente, las expresiones de la varianza y del coeficiente de curtosis de y_t para el modelo ARSV(1) aparecen en las ecuaciones [1.14] y [1.16], respectivamente. La fórmula de la función de autocorrelación de y_t^2 aparece en la ecuación [1.19]. Sustituyendo en estas fórmulas los valores de los parámetros estimados en [4.4] se obtienen los valores que aparecen en la última fila de la tabla 4.8.

Tabla 4.8. Momentos muestrales de la serie y_t junto con los momentos teóricos implícitos de y_t para el modelo ARSV(1) estimado con parámetros $\hat{\phi}$ =0.9898, $\hat{\sigma}_\eta^2$ =0.0168, $\hat{\sigma}_*$ =0.9297.

Momentos	Varianza	Curtosis	Correlaciones de y_t^2						
			K=1	K=2	K=3	K=4	K=5	K=10	K=100
Muestrales	1.4364	8.3205	0.2184	0.3284	0.1885	0.3069	0.2336	0.2098	0.0413
ARSV(1)	1.5127	6.8645	0.2164	0.2132	0.2100	0.2069	0.2038	0.1894	0.0590

A la vista de estos resultados, parece que el modelo ARSV(1) reproduce con bastante fidelidad los momentos muestrales de los rendimientos del IBEX-35, excepto la curtosis, cuyo valor muestral es mucho mayor que el que proporciona el modelo estimado. Igualmente, las correlaciones muestrales de y_t^2 son siempre mayores que las correlaciones implícitas del modelo ARSV(1) estimado. Esto último podría estar reflejando el problema del sesgo que tienen las correlaciones muestrales de y_t^2 en modelos SV próximos a la raíz unitaria, según vimos en el capítulo 2. De hecho, si corregimos las correlaciones muestrales por su sesgo correspondiente, el panorama cambia, como veremos a continuación.

El gráfico 4.6 representa las autocorrelaciones muestrales de las series y_t^2 y $\ln(y_t^2)$ junto con la función de autocorrelación teórica de cada una de estas series para el modelo ARSV(1) estimado. Los valores teóricos de estas funciones se obtienen utilizando las fórmulas [1.19] y [1.23], respectivamente. Además, en este gráfico se representan también las funciones de autocorrelación teóricas corregidas por su sesgo. En otras palabras, este gráfico representa, para cada una de las series analizadas, los valores de sus correlaciones muestrales, r(k), junto con los valores $\rho(k)$ y $\rho(k)+\text{sesgo}(r(k))$, donde $\rho(k)$ es la correlación teórica de orden k y sesgo(r(k)) es el sesgo teórico de la correlación muestral de orden k de la serie correspondiente.

Para la serie $\ln(y_t^2)$, el sesgo de r(k) se obtiene analíticamente a partir de la expresión [2.71] utilizando los valores estimados de los parámetros. Sin embargo, para la serie y_t^2 no tenemos ninguna expresión analítica que nos permita calcular el sesgo de sus correlaciones muestrales. En este caso, el sesgo se ha estimado mediante simulaciones de la forma siguiente. Utilizando las estimaciones de los parámetros del modelo ARSV(1) en [4.4], simulamos 3000 series de longitud T=2991, y para cada una de ellas calculamos la función de autocorrelación muestral de la serie y_t^2. Para cada retardo k, el sesgo medio de r(k) se obtiene como $[\frac{1}{3000}\sum_{i=1}^{3000} r^{(i)}(k)]-\rho(k)$, donde $r^{(i)}(k)$ es la correlación muestral de orden k de y_t^2 en la i-ésima réplica.

Gráfico 4.6. Autocorrelaciones muestrales para la serie de los cuadrados de los rendimientos y del logaritmo de los cuadrados de los rendimientos del IBEX-35 junto con las funciones de autocorrelación teóricas correspondientes al modelo ARSV(1) estimado, corregidas y sin corregir por su sesgo.

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Las conclusiones que podemos sacar del gráfico 4.6 son las siguientes. En general, la función de autocorrelación que proporciona el modelo ARSV(1) estimado refleja bastante bien el comportamiento de las correlaciones muestrales de las dos series analizadas. Además, al corregir las correlaciones muestrales por su sesgo, el ajuste mejora considerablemente, especialmente en la serie y_t^2. En la serie $\ln(y_t^2)$ el sesgo del correlograma no es tan importante. En este último caso, parece que las correlaciones muestrales están, en general, por encima de las correlaciones implícitas por el modelo ARSV(1) ajustado, lo que podría estar indicando que el ajuste con un modelo de memoria larga sería mejor.

A la vista de los resultados obtenidos, podríamos decir que el modelo ARSV(1) estimado modeliza bastante bien el comportamiento de los rendimientos diarios del IBEX-35. Sin embargo, cuando analizamos el comportamiento de las observaciones estandarizadas para los cuatro modelos estimados, hemos visto que la inclusión de integración fraccional en la parte autorregresiva del modelo mejoraba claramente los resultados. Sin embargo, no podemos establecer ningún juicio comparativo en la estimación de las correlaciones, porque el hecho de que ninguno de los modelos fraccionarios estimados sea estacionario nos impide comparar las correlaciones implícitas de estos modelos con las del modelo ARSV(1).

En cualquier caso, nuestras conclusiones no pueden tomarse como definitivas, porque muchas de las herramientas estadísticas que hemos utilizado para analizar el comportamiento de los residuos estandarizados, tienen propiedades aún desconocidas en el ámbito de los modelos SV con memoria larga y, por tanto, los valores de estos estadísticos y las consecuencias que de ellos se derivan, deben interpretarse aún con mucha cautela. Sólo cuando conozcamos las propiedades teóricas de estos estadísticos para los modelos LMSV y sepamos cómo corregir correctamente sus distribuciones asintóticas, estaremos en condiciones de comparar equitativamente, con las mismas herramientas, los diferentes modelos ajustados. Sin lugar a dudas este es uno de los campos en los que seguiremos investigando.

CONCLUSIONES

Los objetivos de este trabajo han sido, en primer lugar, analizar algunos de los problemas de estimación que presenta el modelo de Volatilidad Estocástica con Memoria Larga y, en segundo lugar, discutir otros problemas más generales, que no se habían planteado hasta este momento en el ámbito de los modelos de Volatilidad Estocástica, como son la identificación y validación de este tipo de modelos.

El trabajo ha comenzado, necesariamente, con una revisión de los principales modelos de volatilidad propuestos en la literatura econométrica, con especial atención a los modelos de Volatilidad Estocástica, o modelos SV, y en particular al modelo de Volatilidad Estocástica con Memoria Larga, o modelo LMSV. Esto nos ha llevado a una revisión previa de los modelos con memoria larga para la media, especialmente de los modelos fraccionalmente integrados, o modelos ARFIMA, que resultan imprescindibles para comprender la dinámica de los modelos LMSV.

Por lo que respecta al problema de identificación y validación de modelos SV, nuestra investigación se ha centrado en el estudio de las propiedades estadísticas del correlograma, por ser ésta una de las herramientas más utilizadas en la modelización de series temporales. Concretamente, en la modelización de la volatilidad, el comportamiento del correlograma de la serie de los cuadrados suele utilizarse como un primer indicio de la dependencia temporal en los momentos de segundo orden y de la necesidad de modelos no lineales. Además, el correlograma de los cuadrados de los residuos suele utilizarse para validar el modelo correspondiente.

En los modelos de Volatilidad Estocástica, o modelos SV, el problema de la identificación viene determinado, en gran parte, por el hecho de que la función de autocorrelación de las series y_t^2 y $\ln(y_t^2)$ puede tomar valores difícilmente distinguibles de los de un ruido blanco, si la varianza del logaritmo de la volatilidad es muy pequeña, algo que ocurre con bastante frecuencia en las aplicaciones empíricas. Este problema se agrava aún más cuando se consideran las autocorrelaciones muestrales, que presentan sistemáticamente un sesgo negativo, especialmente importante en los modelos más persistentes.

En este trabajo se analizan, en primer lugar, las propiedades estadísticas del correlograma de la serie y_t en presencia de volatilidad estocástica y el efecto que estas propiedades tienen sobre los contrastes habituales de incorrelación serial que de ellas se derivan. En concreto, se prueba que, en los modelos SV estacionarios, la varianza asintótica de las correlaciones muestrales es mayor que la habitual 1/T y el estadístico de Box-Pierce tampoco tiene la distribución habitual. En consecuencia, se propone una corrección por volatilidad de ambos estadísticos.

En segundo lugar, se analiza el comportamiento de las covarianzas y las correlaciones muestrales de la serie de los cuadrados y de la serie del logaritmo de los cuadrados, tanto en el modelo SV autorregresivo como en el modelo SV con memoria larga. En concreto, se obtienen las expresiones del sesgo y de la varianza asintóticos de las covarianzas muestrales de la serie $\ln(y_t^2)$ en ambos modelos. Además se obtiene la distribución asintótica de las covarianzas muestrales para el modelo ARSV(1) y para el modelo LMSV con $d<1/4$. Se derivan también las fórmulas del sesgo y de la varianza asintóticos de las correlaciones muestrales de la serie $\ln(y_t^2)$ en el modelo ARSV(1), y se obtiene su distribución asintótica. Con respecto al modelo LMSV, se comprueba que los resultados de Hosking (1996) sobre las correlaciones muestrales en modelos ARFIMA no son directamente aplicables al modelo LMSV. Sin embargo, utilizando los resultados de Wright (1999), se obtiene

la varianza asintótica de las correlaciones muestrales de la serie $\ln(y_t^2)$ y se deriva su distribución asintótica para modelos con d<1/4.

Como conclusiones más importantes de nuestra investigación con respecto al problema de las correlaciones en modelos SV, destacaríamos las siguientes:

- Las bandas de confianza para contrastar incorrelación en modelos de volatilidad estocástica son diferentes para cada retardo y son siempre mayores que las habituales bandas de Barlett $\pm 1.96/\sqrt{T}$. Por tanto, la utilización de éstas puede conllevar, en ciertas ocasiones, a la decisión de rechazar erróneamente la hipótesis de incorrelación. Este efecto es especialmente importante en los primeros retardos y en los modelos más persistentes y con mayor varianza. Conclusiones similares se obtienen para el estadístico de Box-Pierce.
- Los contrastes de incorrelación corregidos muestralmente por la presencia de volatilidad estocástica, propuestos en este trabajo, proporcionan buenos resultados en muestras finitas. En todos los experimentos de Monte Carlo realizados, los contrastes propuestos mantienen siempre el tamaño nominal, incluso en los modelos más persistentes y con tamaños muestrales moderados.
- En los modelos de volatilidad estocástica ARSV(1) y LMSV, la covarianza muestral de la serie $\ln(y_t^2)$ siempre subestima su análogo poblacional. Su sesgo y su varianza son mayores cuanto más persistente es el modelo y mayor es la volatilidad. Las mismas conclusiones se obtienen para las correlaciones.
- En el modelo SV con memoria larga, el sesgo de las correlaciones muestrales de las series y_t^2 y $\ln(y_t^2)$ es de tal magnitud, que cuestiona la validez del correlograma de dichas series para la identificación y la validación del modelo. De hecho, su utilización puede conllevar la selección de un modelo erróneo, ya que, con los tamaños de muestra habituales, detectar la presencia de memoria larga a través del correlograma de dichas series es extremadamente difícil, incluso cuando el modelo es muy persistente.

Por lo que respecta a la estimación del modelo de volatilidad estocástica con memoria larga (LMSV), nuestro trabajo se ha centrado principalmente en las propiedades del método de estimación pseudo-máximo verosímil (PMV) en el dominio de las frecuencias, propuesto simultáneamente por Breidt, Crato y de Lima (1998) y Harvey (1998).

Tomando como referencia el trabajo de Breidt, Crato y de Lima (1998) sobre propiedades en muestras finitas del estimador PMV, sus resultados se extienden en varios sentidos. En primer lugar, se analiza un rango más amplio de valores de los parámetros que incluye modelos más realistas y modelos no estacionarios. En segundo lugar, se presentan las propiedades de los estimadores para *todos* los parámetros del modelo, incluido el estimador de la varianza de la perturbación del proceso que modeliza la volatilidad, que es precisamente el que peores propiedades tiene en muestras finitas.

Nuestro estudio ha permitido descubrir que existen serios problemas en la estimación PMV de modelos de volatilidad estocástica con memoria larga cuando los parámetros toman valores próximos a las fronteras de la no estacionariedad y/o de la homocedasticidad. En estos casos, que corresponden a los modelos más realistas, será necesario disponer de muestras extremadamente grandes para poder hacer inferencias fiables.

Nuestro trabajo supone también una llamada de atención sobre el problema de identificación que existe en el modelo SV autorregresivo y con memoria larga (ARLMSV) cuando el proceso que define la volatilidad tiene casi una raíz unitaria. En este caso, el estimador PMV es incapaz de distinguir la procedencia de la no estacionariedad, lo que puede conllevar la estimación de un modelo erróneo. Este resultado puede tener implicaciones importantes en aplicaciones empíricas con datos reales, donde parece que lo habitual es estimar precisamente modelos próximos a la no estacionariedad.

Los resultados más novedosos de nuestro trabajo sobre la estimación PMV en el dominio de las frecuencias para el modelo LMSV, se pueden resumir en los siguientes puntos:

- En los modelos LMSV más persistentes y/o con menor volatilidad, es muy probable estimar $\hat{d}=1/2$ (proceso no estacionario), aunque el proceso generador de los datos sea realmente estacionario, debido a que la distribución del estimador PMV del parámetro d tiene una gran concentración de masa en la frontera de la no estacionariedad, incluso en muestras muy grandes.
- En general, la distribución del estimador PMV de la varianza de la perturbación del logaritmo de la volatilidad, $\hat{\sigma}_\eta^2$, es asimétrica y sesgada negativamente hacia cero, pero mejora considerablemente al aumentar el tamaño muestral y el valor de σ_η^2. Sin embargo, en los modelos con poca volatilidad, la distribución muestral de $\hat{\sigma}_\eta^2$ es tal, que se estimará $\hat{\sigma}_\eta^2 \approx 0$ con mucha probabilidad.
- La estimación PMV de modelos LMSV no estacionarios, que se realiza sobre el modelo en primeras diferencias, parece tener mejores propiedades que en el caso estacionario.
- Cuando en el modelo SV autorregresivo y con memoria larga, el valor de ϕ está muy próximo a la raíz unitaria, el método PMV sobrestima d tratando de capturar la no estacionariedad debida a la presencia de la raíz casi unitaria. La distribución de $\hat{d}$ aparece sesgada hacia el valor de la frontera, d=1/2, y presenta una gran concentración de masa en torno a ese valor, que provoca un sesgo positivo muy grande en las estimaciones de d, que además no disminuye con T. Por otra parte, el parámetro ϕ se subestima y su sesgo negativo tampoco disminuye al aumentar el tamaño muestral.
- Cuando la no estacionariedad proviene del parámetro de integración fraccional d, los problemas de estimación del modelo ARLMSV surgen en el parámetro autorregresivo ϕ. A diferencia de lo que ocurría en el caso anterior, $\hat{\phi}$ presenta

ahora un sesgo positivo y mayor cuanto más próximo está d a la frontera de la no estacionariedad (d=1/2). En este caso, el método PMV siempre sobrestima ϕ en su intento de capturar una posible raíz unitaria que pudiera estar causando la no estacionariedad del modelo. En consecuencia, la distribución de $\hat{\phi}$ presenta una concentración de masa en torno a uno mayor de la esperada.

Estas conclusiones que acabamos de presentar parecen contradecir los resultados de trabajos anteriores, como el de Breidt, Crato y de Lima (1998), en el que se afirma que los estimadores PMV del modelo LMSV tienen buenas propiedades en muestras finitas. Sin embargo, esta aparente contradicción no es tal. Lo único que ocurre es que para los valores de los parámetros analizados en esos trabajos, el estimador PMV en el dominio de las frecuencias tiene, efectivamente, buenas propiedades, como puede comprobarse en las simulaciones recogidas en el capítulo 3.

Como alternativa al método de estimación PMV en el dominio de las frecuencias, se propone un método de estimación PMV en el dominio temporal mediante el filtro de Kalman, basado en la propuesta de Chan y Palma (1998) para procesos ARFIMA. Los resultados preliminares del estudio de Monte Carlo presentado en este trabajo, parecen indicar que esta alternativa es inviable para los tamaños muestrales habituales en aplicaciones empíricas. Además, cuando la media del proceso es desconocida, el estimador compara desfavorablemente con respecto a su análogo en el dominio de las frecuencias, no sólo en tiempo, sino en términos de sesgo y varianza.

Otra aportación de esta tesis es haber resuelto el problema de cómo adaptar el método propuesto por Harvey (1998) para obtener la estimación suavizada de la volatilidad en modelos LMSV, cuando el tamaño de la muestra es lo suficientemente grande como para hacer inviable la propuesta original, que requiere la inversión de matrices TxT. El método se ha implementado satisfactoriamente tanto en series simuladas como en la serie de los rendimientos diarios del índice IBEX-35.

Por lo que respecta a la modelización de la volatilidad del índice IBEX-35, los resultados presentados en este trabajo parecen indicar, en general, que los modelos SV con memoria larga estimados proporcionan mejores ajustes que los modelos ARSV(1) y RWSV. Además, las volatilidades estimadas por los modelos con memoria larga reproducen mejor la dinámica de los rendimientos del IBEX-35, especialmente en los periodos de mayor volatilidad, donde el ajuste que proporcionan estos modelos es claramente superior. Sin embargo, no disponemos aún de un criterio claro que permita elegir entre el modelo LMSV y el modelo ARLMSV ajustados.

En cualquier caso, estas conclusiones no pueden tomarse como definitivas, porque las propiedades de las correlaciones de los residuos estandarizados, utilizados para comparar los distintos modelos ajustados, aún son desconocidas en el ámbito de los modelos SV, y por tanto, los valores de estos estadísticos y las consecuencias que de ellos se derivan, deben interpretarse con cautela.

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

APÉNDICES

A. Simulación del proceso ARFIMA(0,d,0)

En este apéndice se describe el método propuesto por Davies y Harte (1987) para generar una serie de observaciones de un proceso ARFIMA(0,d,0) gaussiano.

Sean $\gamma(0)$, $\gamma(1)$, ... los valores de la función de autocovarianza teórica del proceso ARFIMA(0,d,0). Sea f_k la transformada finita de Fourier de orden k de la secuencia

$$\{\gamma(0), \gamma(1),..., \gamma(N-1), \gamma(N), \gamma(N-1),..., \gamma(1)\}$$

definida como:

$$f_k = \sum_{j=0}^{N-1} \gamma(j) \exp\{i \lambda_k j\} + \sum_{j=N}^{2N-1} \gamma(2N-j) \exp\{i \lambda_k j\} \qquad [A.1]$$

donde $\lambda_k = \frac{2\pi k}{2N}$ es la k-ésima frecuencia de Fourier, con k=0,1,…,2N-1. Lo que vamos a ver es que a partir de estos valores f_k se puede construir una sucesión de variables aleatorias $\{X_0, X_1, \ldots, X_N\}$ cuya función de autocovarianza es precisamente $\gamma(k)$, con k=0,1,…,N.

En primer lugar demostraremos que los valores de f_k, con k=0,1,…,2N-1, son números reales y que $f_{2N-k}=f_k$, para k=1,...,N-1. Para probar la primera propiedad reescribimos f_k como:

$$f_k = \sum_{j=0}^{N-1} \gamma(j) \exp\{i\lambda_k j\} + \sum_{j=1}^{N} \gamma(j) \exp\{i\lambda_k (2N - j)\} \qquad [A.2]$$

Teniendo en cuenta que:

$$\exp\{i\lambda_k N\} = \exp\{i\pi k\} = (-1)^k$$
$$\exp\{i\lambda_k (2N-j)\} = \exp\{i2\pi k\} \exp\{-i\lambda_k j\} = \exp\{-i\lambda_k j\} \qquad [A.3]$$

y sustituyendo estas expresiones en [A.2], f_k se puede escribir como:

$$f_k = \gamma(0) + \sum_{j=1}^{N-1} \gamma(j) \exp\{i\lambda_k j\} + \sum_{j=1}^{N-1} \gamma(j) \exp\{-i\lambda_k j\} + (-1)^k \gamma(N)$$

que nos lleva a la siguiente expresión de f_k:

$$f_k = \gamma(0) + (-1)^k \gamma(N) + 2\sum_{j=1}^{N-1} \gamma(j) \cos(\lambda_k j) \qquad [A.4]$$

para k=0,1,...,2N-1, y con esto queda demostrado que f_k es una función real.

Para probar que $f_{2N-k}=f_k$, con k=1,...,N-1, escribimos en primer lugar la expresión de f_{2N-k} que, según la ecuación [A.4], es:

$$f_{2N-k} = \gamma(0) + (-1)^{2N-k} \gamma(N) + 2\sum_{j=1}^{N-1} \gamma(j) \cos(\lambda_{2N-k} j)$$

pero como $\cos(\lambda_{2N-k}\, j) = \cos(\frac{2\pi(2N-k)}{2N} j) = \cos(2\pi - \lambda_k\, j) = \cos(\lambda_k\, j)$, resulta que:

$$f_{2N-k} = \gamma(0) + (-1)^k \gamma(N) + 2\sum_{j=1}^{N-1} \gamma(j)\cos(\lambda_k\, j) \qquad [A.5]$$

para k=1,2,...,N-1. Comparando las expresiones [A.5] y [A.4] se comprueba $f_{2N-k}=f_k$. Esta propiedad tiene consecuencias importantes en la práctica, ya que al hacer las simulaciones sólo necesitaremos calcular las primeras N+1 transformadas de Fourier, $f_0,f_1,...,f_N$, que se obtienen aplicando la fórmula [A.4].

Para que el método de simulación propuesto por Davies y Harte (1987) funcione, los valores $\{f_k\}$ tienen que ser además no negativos. Si algún f_k es negativo, el método falla. En nuestros ejemplos todos los valores fueron positivos.

Consideremos ahora una sucesión de variables aleatorias normales complejas[1] e independientes, $\{Z_k\}$, k=0,1,...,2N-1, con partes real e imaginaria independientes y satisfaciendo las siguientes condiciones:

- Z_0 ,Z_N son variables con distribución N(0,2),
- $Z_k=U_k+iV_k$, con $U_k \sim N(0,1)$, $V_k \sim N(0,1)$, para k=1,...,N-1,
- $Z_k=Z_{2N-k}$, para k=N+1,...,2N-1.

Bajo estos supuestos, demostraremos que la serie $\{X_0,X_1,\ldots,X_N\}$, definida como:

$$X_j = \frac{1}{2\sqrt{N}} \sum_{k=0}^{2N-1} Z_k \sqrt{f_k}\, \exp\{i\lambda_j k\}, \text{ con } j=0,1,\ldots,N \qquad [A.6]$$

es real y tiene la función de autocovarianza requerida, $\gamma(k)$.

[1] Ver Brillinger (1981, pp. 89, 507) para la definición de una distribución Normal compleja.

Veamos en primer lugar que las variables X_j, j=0,1,..,N son reales. El argumento para demostrar esta propiedad es igual al anterior. Se escribe X_j como:

$$X_j =$$

$$= \frac{1}{2\sqrt{N}}\left[Z_0 \sqrt{f_0} + (-1)^j Z_N \sqrt{f_N} + \sum_{k=1}^{N-1} Z_k \sqrt{f_k}\, \exp\{i\lambda_j k\} + \sum_{k=1}^{N-1} Z_{2N-k} \sqrt{f_{2N-k}}\, \exp\{i\lambda_j (2N-k)\} \right]$$

Por los resultados anteriores sabemos que $f_{2N-k}=f_k$, y que $Z_{2N-k}=Z_k$, por definición, para k=1,...,N-1. Utilizando estas propiedades y la expresión [A.3], se comprueba que:

$$X_j = \frac{1}{2\sqrt{N}}\left[Z_0 \sqrt{f_0} + (-1)^j Z_N \sqrt{f_N} + \sum_{k=1}^{N-1} \sqrt{f_k} \left(Z_k \exp\{i\lambda_j k\} + \overline{Z}_k \exp\{-i\lambda_j k\} \right) \right] =$$

$$= \frac{1}{2\sqrt{N}}\left[Z_0 \sqrt{f_0} + (-1)^j Z_N \sqrt{f_N} + 2\sum_{k=1}^{N-1} \sqrt{f_k} \left(U_k \cos\{\lambda_j k\} - V_k \sin\{\lambda_j k\} \right) \right]$$

Puesto que las variables Z_0, Z_N, y U_k, V_k, con k=1,…,N-1, son todas reales, las variables X_j, con j=0,1,…,N, también lo serán.

Veamos ahora que las autocovarianzas de la serie $\{X_0,X_1,\ldots,X_N\}$ coinciden con las dadas, $\gamma(0)$, $\gamma(1)$,..., $\gamma(N)$. La prueba de esta propiedad resulta más sencilla si se trabaja con la notación compleja de X_j definida en [A.6]. Puesto que $E(Z_k)=0$ para todo k, entonces $E(X_j)=0$, para todo j, y la función de autocovarianza de esta serie puede escribirse como:

$$\mathrm{Cov}(X_j, X_{j+\tau}) = \mathrm{E}(X_j \overline{X}_{j+\tau}) =$$

$$= \mathrm{E}\left(\frac{1}{2\sqrt{N}} \sum_{k=0}^{2N-1} Z_k \sqrt{f_k}\, \exp\{i\lambda_j k\} \right) \left(\frac{1}{2\sqrt{N}} \sum_{r=0}^{2N-1} \overline{Z}_r \sqrt{f_r}\, \exp\{i\lambda_{j+\tau} r\} \right) =$$

$$= \frac{1}{4N} \sum_{k=0}^{2N-1} \sum_{r=0}^{2N-1} \sqrt{f_k}\, \sqrt{f_r}\, \exp\{i\lambda_j (k-r)\} \exp\{-i\lambda_\tau r\} \mathrm{E}(Z_k \overline{Z}_r) \qquad [A.7]$$

Además, como las variables Z_k son variables independientes con media cero y varianza igual a 2, se tiene que:

$$\mathrm{E}(Z_j \overline{Z}_r) = \begin{cases} 0 & \text{si } r \neq k \\ 2 & \text{si } r = k \end{cases}$$

BIBLIOTECA VIRTUAL

Sustituyendo estos valores en [A.7] resulta que los únicos sumandos distintos de cero son aquellos en los que r=k y, por tanto, puede escribirse:

$$\mathrm{Cov}(X_j, X_{j+\tau}) = \frac{1}{2N} \sum_{k=0}^{2N} f_k \exp\{-i\lambda_\tau k\}, \text{ para } \tau=0,1,\ldots,N$$

Esta última expresión significa que $\mathrm{Cov}(X_j, X_{j+\tau})$ es la transformada finita de Fourier de los valores $\{f_0,...,f_N\}$. Pero precisamente estos valores se definieron en [A.1] como la transformada inversa de Fourier de $\{\gamma(0),\gamma(1),..., \gamma(N)\}$. Dada la relación que existe entre la transformada finita de Fourier y su inversa[2], se concluye que:

$$\mathrm{Cov}(X_j, X_{j+\tau}) = \gamma(\tau), \text{ para } \tau=0,1,...,N,$$

y por tanto la serie $\{X_0, X_1, \ldots, X_N\}$ definida en [A.6] tiene la distribución requerida.

[2] Ver Bloomfield (1976, cap.3) para más detalles sobre la transformada de Fourier y su inversa.

B. Forma reducida del modelo ARSV(1)

Como vimos en la sección 1.1.2.1, el modelo SV autorregresivo de orden uno, denominado ARSV(1), admite una representación en el espacio de los estados para la serie del logaritmo de los cuadrados, de la forma:

$$x_t = \ln(y_t^2) = \mu + h_t + \xi_t \qquad [B.1a]$$

$$h_t = \phi h_{t-1} + \eta_t \qquad [B.1b]$$

donde η_t es NID$(0, \sigma_\eta^2)$, μ es la media de x_t y ξ_t= ln(ε_t^2)-E[ln(ε_t^2)] es un ruido blanco no gaussiano de media 0 y varianza σ_ξ^2, e independiente de η_t. Esta formulación permite interpretar [B.1a] y [B.1b] como las ecuaciones de medida y de transición, respectivamente, de un modelo estructural para la serie x_t=ln(y_t^2). Para obtener la forma reducida de este modelo, basta con resolver dichas ecuaciones.

En primer lugar, se despeja h_t en la ecuación [B.1b] y se sustituye su valor en [B.1a], obteniéndose la siguiente expresión de x_t:

$$(x_t - \mu) = \frac{\eta_t}{(1 - \phi L)} + \xi_t$$

o equivalentemente

$$(1-\phi L)(x_t-\mu) = \eta_t + \xi_t - \phi\xi_{t-1} \qquad [B.2]$$

Para determinar la estructura dinámica de este nuevo modelo, analizamos la función de autocorrelación de la serie $(1-\phi L)(x_t-\mu)$ a partir de su expresión en [B.2]. Para ello obtenemos la varianza y la función de autocovarianza de dicha serie:

$$\mathrm{Var}[(1-\phi L)(x_t-\mu)] = \mathrm{Var}(\eta_t + \xi_t - \phi\xi_{t-1}) = \sigma_\eta^2 + (1+\phi^2)\sigma_\xi^2 \qquad [B.3]$$

$$\mathrm{Cov}([(1-\phi L)(x_t-\mu)],[(1-\phi L)(x_{t-k}-\mu)]) =$$

$$= \mathrm{Cov}(\eta_t + \xi_t - \phi\xi_{t-1}, \eta_{t-k} + \xi_{t-k} - \phi\xi_{t-k}) = \begin{cases} -\phi\sigma_\xi^2 & \text{si} \quad k=1 \\ 0 & \text{si} \quad k \geq 2 \end{cases} \qquad [B.4]$$

Dividiendo [B.4] entre [B.3] se obtiene la función de autocorrelación de $(1-\phi L)(x_t-\mu)$:

$$\rho(k;(1-\phi L)(x_t-\mu)) = \begin{cases} -\dfrac{\phi}{q+1+\phi^2} & \text{si } k=1 \\ 0 & \text{si } k \geq 2 \end{cases} \qquad [B.5]$$

donde q denota el cociente de las varianzas de las perturbaciones, $q=\sigma_\eta^2/\sigma_\xi^2$.

La función de autocorrelación en [B.5] tiene la misma estructura que la de un proceso MA(1): todas las autocorrelaciones son iguales a cero, excepto la primera, que es la única distinta de cero. Por tanto, la serie $(1-\phi L)(x_t-\mu)$ es un proceso MA(1), y, en consecuencia, la serie $(x_t-\mu)$ será un proceso ARMA(1,1). Este proceso admite una representación de la forma:

$$(1-\phi L)(x_t-\mu) = z_t + \theta\, z_{t-1} \qquad [B.6]$$

donde la perturbación z_t es I.I.D.$(0,\sigma_z^2)$, porque las perturbaciones η_t y ξ_t son I.I.D. y además son independientes entre sí, y donde los parámetros $\{\phi,\theta\}$ verifican ciertas restricciones. Estas restricciones sobre los parámetros se obtienen igualando la función de autocorrelación en [B.5] a la función de autocorrelación de un proceso MA(1) con parámetro θ, a saber:

$$\rho(k;MA(1)) = \begin{cases} \dfrac{\theta}{1+\theta^2} & \text{si } k = 1 \\ 0 & \text{si } k \geq 2 \end{cases} \qquad [B.7]$$

Resolviendo en θ la siguiente ecuación, resultante de igualar [B.5] y [B.7]:

$$\phi\theta^2 + (q+1+\phi^2)\theta + \phi = 0 \qquad [B.8]$$

con las condiciones $|\phi|<1$ y $|\theta|<1$, para asegurar estacionariedad e invertibilidad, resulta que la única solución posible en θ es:

$$\theta = \frac{-(q+1+\phi^2) + \sqrt{(q+1+\phi^2)^2 - 4\phi^2}}{2\phi}$$

con las siguientes restricciones sobre los parámetros del modelo: $\{0<\phi<1,-1<\theta<0\}$ ó $\{-1<\phi<0,0<\theta<1\}$.

Con estas restricciones, el modelo ARMA(1,1) definido en [B.6] es estacionario e invertible, y admite una representación lineal MA(∞) de la forma:

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j z_{t-j}$$

donde z_t es I.I.D. y los coeficientes $\{\psi_j\}$ son $\psi_0=1$, $\psi_1=\phi+\theta$, y $\psi_j=\phi^{j-1}\psi_1$ para $j>1$. Estos coeficientes son absolutamente sumables, porque $|\phi|<1$.

C. Cumulantes de cuarto orden y momentos de orden superior de la serie $\ln(y_t^2)$ en modelos SV

Sea x_t un proceso estacionario con media μ, tal que los momentos de orden cuatro existen y verifican la condición de estacionariedad de cuarto orden establecida en [2.34]. Se define el cumulante mixto de cuarto orden de x_t como:

$$\kappa_x(s,r,q) = E(x_t-\mu)(x_{t+s}-\mu)(x_{t+r}-\mu)(x_{t+q}-\mu) - [\,\gamma(s)\gamma(r-q)+\gamma(r)\gamma(s-q)+\gamma(q)\gamma(s-r)\,] \quad [C.1]$$

donde $\gamma(r)$ denota la autocovarianza de orden r del proceso x_t.

Veamos a continuación cómo son los cumulantes de cuarto orden de la serie $x_t=\ln(y_t^2)$ en el modelo ARSV(1) estacionario. Según las ecuaciones [B.1], en este modelo, la serie $(x_t-\mu)$ es la suma de un proceso AR(1) estacionario gaussiano de media cero, h_t, y un ruido blanco no gaussiano, ξ_t, con media cero y varianza σ_ξ^2, e independiente de h_t. A partir de la expresión [B.1a] se deducen de forma sencilla los momentos de cuarto orden de x_t, como veremos a continuación.

Para que el momento de orden cuatro de x_t exista y sea finito, debe verificarse la condición $E(x_t-\mu)^4<\infty$. En el modelo ARSV(1) esta condición está asegurada porque $E(x_t-\mu)^4$ puede escribirse como:

$$E(x_t-\mu)^4 = E(h_t+\xi_t)^4 = E(h_t^4)+6E(h_t^2)E(\xi_t^2)+E(\xi_t^4) \quad [C.2]$$

y los momentos de segundo y cuarto orden de h_t y ξ_t son finitos. De hecho, utilizando los resultados de Abramowitz y Stegun (1970, p.943), se comprueba que $E(\xi_t^2)=\pi^2/2$ y $E(\xi_t^4)=3\sigma_\xi^4+\pi^4$, porque ξ_t es el logaritmo de una variable χ_1^2. Además, sabemos que $E(h_t^2)=\sigma_h^2$ y $E(h_t^4)=3\sigma_h^4$, porque la distribución de h_t es normal $N(0,\sigma_h^2)$. Sustituyendo estos valores en [C.2], el momento de orden cuatro centrado de x_t puede escribirse como:

$$E(x_t-\mu)^4 = 3\sigma_h^4+6\sigma_h^2\sigma_\xi^2+3\sigma_\xi^4+\pi^4 = 3[\gamma(0)]^2+\pi^4 \quad [C.3]$$

donde $\gamma(0)=Var(x_t)=\sigma_h^2+\sigma_\xi^2$. Por tanto, se verifica que $E(x_t-\mu)^4<\infty$. De forma similar se demuestra que $x_t=\ln(y_t^2)$ verifica la condición de estacionariedad en los momentos de cuarto orden establecida en [2.34], por verificarla las series h_t y ξ_t.

Para calcular los cumulantes de cuarto orden de la serie x_t en el modelo ARSV(1), sustituimos en la fórmula [C.1] la serie $(x_t-\mu)$ por su expresión para dicho modelo, $x_t-\mu=h_t+\xi_t$. Teniendo en cuenta que h_t y ξ_t son variables independientes, y que además las variables ξ_t están incorrelacionadas, se demuestra que el cumulante mixto de cuarto orden de x_t puede escribirse como:

$$\kappa_x(s,r,q) =$$
$$=E(h_t h_{t+s} h_{t+r} h_{t+q}) + E(h_t h_{t+s})E(\xi_{t+r} \xi_{t+q}) + E(h_t h_{t+q})E(\xi_{t+s} \xi_{t+r}) + E(h_t h_{t+r})E(\xi_{t+s} \xi_{t+q})+$$
$$+E(h_{t+s} h_{t+q})E(\xi_t \xi_{t+r}) + E(h_{t+s} h_{t+r})E(\xi_t \xi_{t+q}) + E(h_{t+q} h_{t+r})E(\xi_t \xi_{t+s}) +E(\xi_t \xi_{t+s} \xi_{t+r} \xi_{t+q}).$$
$$- [\gamma(s)\gamma(r-q) + \gamma(r)\gamma(s-q) + \gamma(q)\gamma(s-r)] \qquad [C.4]$$

donde $\gamma(r)$ denota la función de autocovarianza de la serie $x_t=\ln(y_t^2)$. En las ecuaciones [1.21] y [1.22], demostramos que esta función puede escribirse como:

$$\gamma(0) = Var(x_t)=\sigma_h^2+\sigma_\xi^2$$
$$\gamma(r) = \gamma_h(r), \text{ para } r>0$$

donde $\gamma_h(r)$ denota la función de autocovarianza del proceso h_t. Además, por ser h_t un proceso estacionario gaussiano de media cero, todos sus cumulantes de cuarto orden son cero (ver Anderson, 1971, p.444), y resulta que:

$$E(h_t h_{t+s} h_{t+r} h_{t+q}) = \gamma_h(s)\gamma_h(r-q) + \gamma_h(r)\gamma_h(s-q) + \gamma_h(q)\gamma_h(s-r) \qquad [C.5]$$

Sustituyendo estas expresiones en [C.4] se obtienen los siguientes valores de κ_x:

- Si r=q≠s, s≠0, q≠0,

  $\kappa_x(s,q,q) = E(h_t h_{t+s} h_{t+q}^2) + \sigma_\xi^2 E(h_t h_{t+s}) - [\gamma(s)\gamma(0) + 2\gamma(q)\gamma(s-q)] =$

  $= \sigma_h^2 \gamma_h(s) + 2\gamma_h(q)\gamma_h(s-q) + \sigma_\xi^2 \gamma_h(s) - \sigma_h^2 \gamma_h(s) - \sigma_\xi^2 \gamma_h(s) - 2\gamma_h(k)\gamma_h(r-k) = 0$

  De forma similar se prueba que $\kappa_x(q,s,q) = 0$ y $\kappa_x(q,q,s) = 0$, si s≠0, q≠0, y q≠s.

- Si r=q≠s, s=0

  $\kappa_x(0,q,q) = E(h_t^2 h_{t+q}^2) + 2\sigma_h^2 \sigma_\xi^2 + \sigma_\xi^4 - \{[\gamma(0)]^2 + 2[\gamma(q)]^2\} =$

  $= \sigma_h^4 + 2[\gamma_h(q)]^2 + 2\sigma_h^2 \sigma_\xi^2 + \sigma_\xi^4 - (\sigma_h^2 + \sigma_\xi^2)^2 - 2[\gamma_h(q)]^2 = 0$

  De forma similar se prueba que $\kappa_x(q,0,q)=0$ y $\kappa_x(q,q,0)=0$, si q≠0.

- Si s=q=r≠0,

  $\kappa_x(s,s,s) = E(h_t h_{t+s}^3) + 3\sigma_\xi^2 E(h_t h_{t+s}) - 3\gamma(0)\gamma(s) = 3(\sigma_h^2 + \sigma_\xi^2)\gamma_h(s) - 3\gamma(0)\gamma_h(s) = 0$

- Si s=q=r=0,

  $\kappa_x(0,0,0) = E(x_t-\mu)^4 - 3[\gamma(0)]^2 = \pi^4$, según la expresión [C.3]

- Si s=0, r≠q, r≠0, q≠0,

  $\kappa_x(0,r,q) = E(h_t^2 h_{t+r} h_{t+q}) + \sigma_\xi^2 E(h_{t+r} h_{t+q}) - [\gamma(0)\gamma(r-q) + 2\gamma(r)\gamma(q)] =$

  $= \sigma_h^2 \gamma_h(r-q) + 2\gamma_h(r)\gamma_h(q) + \sigma_\xi^2 \gamma_h(r-q) - \gamma(0)\gamma_h(r-q) - 2\gamma_h(r)\gamma_h(q) = 0$

  Igualmente, $\kappa_x(s,0,q)=0$ si s≠q, s≠0, q≠0, y $\kappa_x(s,r,0)=0$ si s≠r, s≠0, r≠0.

- Si r=q=0, s≠0,

  $\kappa_x(s,0,0) = E(h_t^3 h_{t+s}) + 3\sigma_\xi^2 E(h_t h_{t+s}) - 3\gamma(s)\gamma(0) =$

  $= 3\sigma_h^2 \gamma_h(s) + 3\sigma_\xi^2 \gamma_h(s) - 3\gamma_h(s)\gamma(0) = 3(\sigma_h^2 + \sigma_\xi^2)\gamma_h(s) - 3\gamma_h(s)\gamma(0) = 0$

  De forma similar se prueba que $\kappa_x(0,s,0)=0$ y $\kappa_x(0,0,s)=0$, si s≠0.

- Por último, si s≠r≠q≠0,

  $\kappa_x(s,r,q) = E(h_t h_{t+s} h_{t+r} h_{t+q}) - [\gamma(s)\gamma(r-q) + \gamma(r)\gamma(s-q) + \gamma(q)\gamma(s-r)] =$

$= E(h_t h_{t+s} h_{t+r} h_{t+q}) - [\gamma_h(s) \gamma_h(r-q) + \gamma_h(r) \gamma_h(s-q) + \gamma_h(q) \gamma_h(s-r)] = 0$, según la ecuación [C.5].

De todo anterior se deduce que en el modelo ARSV(1) todos los cumulantes mixtos de cuarto orden de x_t son cero, excepto $\kappa_x(0,0,0) = \pi^4$.

En el modelo LMSV se obtienen los mismos resultados por lo siguiente. La igualdad [C.5] se verifica para cualquier proceso estacionario gaussiano con momento de cuarto orden finito y, por tanto, se verifica también para el proceso h_t del modelo LMSV, que es un ARFIMA(0,d,0) gaussiano. En consecuencia, si el momento de cuarto orden de h_t existe y es finito, se verifican todos los resultados anteriores. El momento de orden cuarto de h_t es finito en virtud del resultado de Anderson (1971, p.466), porque los coeficientes de la representación MA(∞) del proceso ARFIMA son de cuadrado sumable y el momento de cuarto orden de η_t es finito. En consecuencia, todos los resultados de este apéndice son válidos para el modelo LMSV.

A continuación veremos que el momento de orden seis de x_t es finito. Para ello, escribimos el momento de orden seis de x_t en función de los momentos de h_t y ξ_t de la siguiente forma:

$$E(x_t-\mu)^6 = E(h_t + \xi_t)^6 =$$
$$=E(h_t^6)+6E(h_t^5 \xi_t)+15E(h_t^4 \xi_t^2)+30E(h_t^3 \xi_t^3)+15E(h_t^2 \xi_t^4)+6E(h_t \xi_t^5)+E(\xi_t^6) \qquad [C.6]$$

Teniendo en cuenta que los procesos h_t y ξ_t son independientes, y que todos los momentos de orden impar de h_t son cero, porque h_t es gaussiano, la expresión [C.6] se reduce a calcular:

$$E(x_t-\mu)^6 =E(h_t^6)+15E(h_t^4)E(\xi_t^2)+15E(h_t^2)E(\xi_t^4)+E(\xi_t^6) \qquad [C.7]$$

Teniendo en cuenta que h_t tiene una distribución $N(0,\sigma_h^2)$, resulta que $E(h_t^2)=\sigma_h^2$, $E(h_t^4)=3\sigma_h^4$ y $E(h_t^6)=15\sigma_h^6$. Por otra parte, como $\xi_t=\ln(\varepsilon_t^2)-E(\ln(\varepsilon_t^2))$, donde ε_t es N.I.D.(0,1), en virtud de los resultados de Abramowitz y Stegun (1970, pp. 258, 807-810, 943) se comprueba que:

$$E(\xi_t^6)=10\mu\,E(\xi_t^4)+15\sigma_\xi^2\pi^4+10[E(\xi_t^3)]^2+8\pi^6$$

donde $\mu=E(\ln(\varepsilon_t^2))=\psi(1/2)-\ln(1/2)$, $E(\xi_t^2)=\sigma_\xi^2=\pi^2/2$, $E(\xi_t^3)=-14\zeta(3)$ y $E(\xi_t^4)=3\sigma_\xi^4+\pi^4$, donde ψ es la función psi de Euler y ζ es la función zeta de Rienman. Por tanto, todos los momentos de orden dos, cuatro y seis que aparecen en [C.7] son finitos y, en consecuencia, lo es también el momento de orden seis de x_t.

D. Condiciones de Anderson (1971, p. 471)

En este apéndice vamos a demostrar que la diferencia entre la varianza de c*(k) y la varianza de c(k) en el modelo ARSV(1) es de orden $O(T^{-2})$, donde c*(k) y c(k) son las covarianzas muestrales de x_t definidas en [2.1] y [2.2], respectivamente. Para ello vamos a utilizar el siguiente resultado de Anderson (1971, pp.470-471).

Sea x_t un proceso estacionario con momento de cuarto orden finito verificando la condición de estacionariedad de cuarto orden establecida en [2.34]. Sea $\gamma(k)$ su función de autocovarianza y $\kappa_x(s,r,q)$ el cumulante mixto de cuarto orden definido en [C.1]. Si $\sum_{-\infty}^{\infty}\gamma(k)<\infty$ y $\left|\sum_{s,r,q=-\infty}^{\infty}\kappa_x(s,r,q)\right|<\infty$, se demuestra que:

$$(T-k)[\mathrm{Cov}(C(g),C(k))-\mathrm{Cov}(\tilde{C}(g),\tilde{C}(k))]=O(1/(T-g)) \qquad [D.1]$$

donde C(k) y $\tilde{C}(k)$ son dos estimadores de $\gamma(k)$ definidos como:

$$C(k) = \frac{1}{T-k}\sum_{t=1}^{T-k}(x_t - \mu)(x_{t+k} - \mu)$$

$$\tilde{C}(k) = \frac{1}{T-k}\sum_{t=1}^{T-k}(x_t - \overline{X}_k)(x_{t+k} - \overline{X}_{k+})$$

respectivamente, siendo $\overline{X}_k$ la media de las T-k primeras observaciones de x_t, y $\overline{X}_{k+}$ la media de las últimas T-k observaciones de x_t. Por tanto, según la ecuación [D.1], en muestras grandes puede utilizarse Cov(C(g),C(k)) como una aproximación de $\mathrm{Cov}(\tilde{C}(g),\tilde{C}(k))$. Además, se demuestra que $\mathrm{Cov}(\tilde{c}(g),\tilde{c}(k))$ puede tratarse de forma similar, siendo $\tilde{c}(k)$ la covarianza muestral de x_t definida en [2.3].

De este resultado de Anderson (1971) se deducen las siguientes conclusiones:

1. Si $\mathrm{Cov}(\tilde{c}(g),\tilde{c}(k))$ pueda tratarse de forma similar a $\mathrm{Cov}(\tilde{C}(g),\tilde{C}(k))$, esto significa que la diferencia (T-g)[Cov(C(g),C(k))-$\mathrm{Cov}(\tilde{c}(g),\tilde{c}(k))$] también es de orden O(1/(T-g)). Por tanto, cuando g=k, resulta que:

   $$(T-k)[\mathrm{Var}(C(k))-\mathrm{Var}(\tilde{c}(k))]=O(1/(T-k))$$

   y, por tanto, existe una constante M>0 tal que:

   $$(T-k)^2|\mathrm{Var}(C(k))-\mathrm{Var}(\tilde{c}(k))|\leq M, \text{ para todo } k. \qquad [D.2]$$

   Cuando g=0, resulta que:

   $$(T-k)[\mathrm{Cov}(C(0),C(k))-\mathrm{Cov}(\tilde{c}(0),\tilde{c}(k))]=O(1/(T-k))$$

   y, por tanto, existe un M>0 tal que:

$$(T-k)^2|Cov(C(0),C(k))-Cov(\tilde{c}(0),\tilde{c}(k))|\leq M, \text{ para todo k.} \qquad [D.3]$$

2. Los estimadores de la autocovarianza que nos interesan no son C(k) y $\tilde{c}(k)$, sino los estimadores c*(k) y c(k), definidos en [2.1] y [2.2], respectivamente. Sin embargo, sabemos que entre ellos existe la siguiente relación:

$$\tilde{c}(k)=\frac{T}{T-k}c(k),\ C(k)=\frac{T}{T-k}c^*(k) \qquad [D.4]$$

Sustituyendo estos valores en [D.2], resulta que $T^2|Var(c^*(k))-Var(c(k))|\leq M$, para todo k, y por tanto,

$$Var(c^*(k))-Var(c(k))=O(T^{-2})$$

Esto significa que, en muestras grandes, Var(c*(k)) puede utilizarse como una aproximación de Var(c(k)).

3. Por lo que respecta a la covarianza entre las covarianzas muestrales, sustituyendo las expresiones [D.4] en [D.3] se tiene que:

$$T(T-k)|Cov(c(0),c(k))-Cov(c^*(0),c^*(k))|\leq M, \text{ para todo k}$$

Además, la sucesión {T/(T-k)} es una sucesión de números positivos que converge hacia uno cuando $T\to\infty$, y por tanto está acotada. De ello se deduce que:

$$T^2|Cov(c(0),c(k))-Cov(c^*(0),c^*(k))|=$$
$$=[T/(T-k)]\{T(T-k)|Cov(c(0),c(k))-Cov(c^*(0),c^*(k))|\}\leq M'$$

para una cierta constante M', y por tanto resulta que:

$$Cov(c^*(0),c^*(k))=Cov(c(0),c(k))+O(T^{-2}).$$

Esto significa que, en muestras grandes, Cov(c*(0),c*(k)) puede utilizarse como una aproximación a Cov(c(0),c(k)).

De los puntos anteriores se deduce que si una serie estacionaria x_t con momentos de cuarto orden finito y verificando la condición de estacionariedad de cuarto orden [2.34], verifica que $\sum_{-\infty}^{\infty}\gamma(k)<\infty$ y $\left|\sum_{s,r,q=-\infty}^{\infty}\kappa_x(s,r,q)\right|<\infty$, donde $\gamma(k)$ es la función de autocovarianza y $\kappa_x(s,r,q)$ es el cumulante de cuarto orden, resulta que:

- $Var(c(k))=Var(c^*(k))+O(T^{-2})$ [D.5]
- $Cov(c(0),c(k))=Cov(c^*(0),c^*(k))+O(T^{-2})$ [D.6]

En el modelo ARSV(1), la serie $x_t=\ln(y_t^2)$ verifica las tres condiciones anteriores, ya que:

(a) $\sum_{-\infty}^{\infty}\gamma(k)=\gamma(0)+2\sum_{r=1}^{\infty}\sigma_h^2\phi^k<\infty$.

(b) Los cumulantes de cuarto orden de x_t son todos cero, excepto $\kappa_x(0,0,0)=\pi^4$, según hemos visto en el apéndice C, y por tanto la suma de los cumulantes está acotada.

(c) Los momentos de cuarto orden de x_t son finitos y verifican la condición de estacionariedad [2.34] porque la verifican los de los procesos h_t y ξ_t; véase apéndice C.

En consecuencia, para el modelo ARSV(1), se cumplen las igualdades [D.5] y [D.6] y, por tanto, en muestras grandes, Var(c*(k)) y Cov(c*(0),c*(k)) pueden utilizarse como aproximaciones a Var(c(k)) y Cov(c(0),c(k)), respectivamente.

BIBLIOGRAFÍA

Abramowitz, M. y N. C. Stegun (1970) *Handbook of Mathematical Functions*. Dover Publications Inc., New York.

Agiakloglou, C. y P. Newbold (1994) "Lagrange multiplier tests for fractional difference". *Journal of Time Series Analysis*, 15, 253-262.

Agiakloglou, C., P. Newbold y M. Wohar (1993) "Bias in an estimator of the fractional difference parameter". *Journal of Time Series Analysis*, 14, 235-246.

Andersen, T.G. y T. Bollerslev (1997) "Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns". *Journal of Finance*, LII, 975-1005.

Andersen, T.G. y B. Sorensen (1996) "GMM estimation of a stochastic volatility model: a Monte Carlo study". *Journal of Business and Economic Statistics,* 14, 328-352.

Andersen, T.G., T. Bollerslev, F.X. Diebold y P. Labys (1999) "The distribution of exchange rate volatility". *Wharton Financial Institutions Center Working Paper* 99-08 y *NBER Working Paper* 6961.

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*. Wiley, New York.

Anderson, T.W. (1992) "The asymptotic distribution of autocorrelation coefficients". En K.V. Madia (ed.), *The Art of Statistical Science: A tribute to G.S. Watson*, 9-25. Wiley, New York.

Anderson, T.W. y A.M. Walker (1964) "On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process". *Annals of Mathematical Statistics*, 35, 1296-1303.

Andrews, D. (1991) "Heteroskedasticity and autocorrelation consistent covariance matrix estimation". *Econometrica*, 59, 817-858.

Baillie, R.T. (1996) "Long memory processes and fractional integration in econometrics". *Journal of Econometrics*, 73, 5-59.

Baillie, R.T. y T. Bollerslev (1989) "The message in daily exchange rates: a conditional-variance tale". *Journal of Business and Economic Statistics*, 7, 297-305.

Baillie, R.T. y T. Bollerslev (1994) "The long memory property of forward premium". *Journal of International Money and Finance*, 13, 565-571.

Baillie, R.T, y R.P. DeGennaro (1990) "Stock returns and volatility". *Journal of Financial and Quantitative Analysis*, 25, 203-215.

Baillie, R.T., T. Bollerslev y H.O. Mikkelsen (1996) "Fractionally integrated autoregressive conditional heteroskedaticity". *Journal of Econometrics*, 74, 3-30.

Bera, A.K. y M.L. Higgins (1993) "A survey of ARCH models: properties, estimation and testing". *Journal of Economic Surveys*, 7, 305-366.

Beran, J. (1994) *Statistics for Long-memory Processes*. Chapman and Hall, New York.

Billingsley, P. (1961) "The Lindeberg-Lévy theorem for martingales". *Proceedings of the American Statistical Society*, 12, 788-792.

Blasco, N. y R. Santamaría (1994a) "Memoria de largo plazo en el mercado de valores español: una aproximación mediante el análisis R/S". *Investigaciones Económicas*, XVIII, 571-583.

Blasco, N. y R. Santamaría (1994b) "¿Son las rentabilidades de las acciones series fraccionalmente integradas?. Resultados en el mercado de valores español". *Revista de Economía Aplicada*, 6, 5-28.

Bloomfield, P. (1973) "An exponential model for the spectrum of a scalar time series". *Biometrika*, 60, 217-226.

Bloomfield, P. (1976) *Fourier Analysis of Time Series: An Introduction*. Wiley, New York.

Bollerslev, T. (1986) "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. (1987) "A conditionally heteroscedastic time series model for speculative prices and rates of return". *Review of Economics and Statistics*, 69, 542-547.

Bollerslev, T. (1988) "On the correlation structure for the generalized autoregressive conditional heteroskedastic process". *Journal of Time Series Analysis*, 9, 121-131.

Bollerslev, T. y H.O. Mikkelsen (1996) "Modeling and pricing long memory in stock market volatility". *Journal of Econometrics*, 73, 151-184.

Bollerslev, T. y J.M. Wooldridge (1992) "Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances". *Econometric Reviews*, 11, 143-172.

Bollerslev, T., R.Y. Chou y K.F. Kroner (1992) "ARCH modeling in finance: a review of the theory and empirical evidence". *Journal of Econometrics*, 52, 5-59.

Bollerslev, T., R.F. Engle y D.B. Nelson (1994) "ARCH models". En R.F. Engle y D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, 2959-3038. Amsterdam, North Holland.

Box, M.J., D. Davies y W.H. Swann (1969) *Non-linear Optimization Techniques*. ICI Monograph, No 5. Oliver and Boyd, Edinburg.

Breidt, F.J. y A.L. Carriquiry (1996) "Improved quasi-maximum likelihood estimation for stochastic volatility models". En J.L. Lee, W.O. Johnson y A. Zellner (eds.), *Modelling and Prediction: Honouring Seymour Geisser*, 228-247. Springer Verlag, New York.

Breidt, F.J., N. Crato y P.J.F. de Lima (1998) "The detection and estimation of long-memory in stochastic volatility". *Journal of Econometrics*, 83, 325-348.

Brillinger, D.R. (1981) *Time Series: Data Analysis and Theory*. Holt, Rinehart y Winston, New York.

Brockwell, P.J. y R.A. Davis (1991) *Time Series: Theory and Methods*. 2ª ed. Springer Verlag, Berlin.

Chan, N.H. y W. Palma (1998) "State space modeling of long memory processes". *Annals of Statistics*, 26, 719-740.

Chan, N.H. y G. Petris (2000) "Long memory stochastic volatility: a bayesian approach". *Communications in Statistics. Theory and Methods*, 29, 1367-1378.

Cheung, Y.W. (1993a) "Long memory in foreign-exchange rates". *Journal of Business and Economic Statistics*, 11, 93-101.

Cheung, Y.W. (1993b) "Tests for fractional integration: a Monte Carlo investigation". *Journal of Time Series*, 14, 331-345.

Cheung, Y.W. y F.X. Diebold (1994) "On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean". *Journal of Econometrics*, 62, 301-316.

Chou, R.Y. (1988) "Volatility persistence and stocks valuations: some empirical evidence using GARCH". *Journal of Applied Econometrics*, 3, 279-294.

Chung, C.F. (1994) "A note on calculating the autocovariances of the fractionally integrated ARMA models". *Economics Letters*, 45, 293-297.

Chung, C.F. y R.T. Baillie (1993) "Small sample bias in conditional sum of squares estimators of fractionally integrated ARMA models". *Empirical Economics*, 18, 791-806.

Crato, N y P.J.F. de Lima (1994) "Long-range dependence in the conditional variance of stock returns". *Economics Letters*, 45, 281-285.

Crato, N. y B. Ray (2000) "Memory in returns and volatilities of commodity futures contracts". *Journal of Futures Markets*, 20, 525-543.

Dacorogna, M.M., U.A. Müller, R.J. Nagler, R.B. Olsen y O.V. Pictet (1993) "A geographical model for the daily and weekly seasonal volatility in the foreign exchange market". *Journal of International Money and Finance*, 12, 413-438.

Dahlhaus, R. (1989) "Efficient parameter estimation for self-similar processes". *The Annals of Statistics*, 17, 1749-1766.

Danielsson, J. (1994) "Stochastic volatility in asset prices: estimation with simulated maximum likelihood". *Journal of Econometrics*, 61, 375-400.

Davies, R.B. y D.S. Harte (1987) "Tests for Hurst effect". *Biometrika*, 74, 95-101.

De Gooijer, J.G. (1980) "Exact moments of the sample autocorrelations from series generated by general ARIMA processes of order (p,d,q), d=0 or 1". *Journal of Econometrics*, 14, 365-379.

Delgado, M.A. y P.M. Robinson (1993) "New methods for the analysis of long-memory time series: application to Spanish inflation". *Journal of Empirical Finance*, 1, 83-106.

Deo, R. y C. Hurvich (1998) "On the log-periodogram estimator of the memory parameter in long memory stochastic volatility". *Statistics and Operational Research Working Paper* SOR-98-4, New York University.

Diebold, F.X. (1988) *Empirical Modelling of Exchange Rate Dynamics*. Springer Verlag, New York.

Diebold, F.X y G.D. Rudebusch (1989) "Long-memory and persistence in aggregate output." *Journal of Monetary Economics*, 24, 189-209.

Diebold, F.X. y G.D. Rudebusch (1991) "On the power of Dickey-Fuller tests against fractional alternatives". *Economics Letters*, 35, 155-160.

Ding, Z. y C.W.J. Granger (1996) "Modeling volatility persistence of speculative returns: a new approach". *Journal of Econometrics*, 73, 185-215.

Ding, Z., C.W.J. Granger y R.F. Engle (1993) "A long memory property of stock market returns and a new model". *Journal of Empirical Finance*, 1, 83-106.

Dolado, J.J. y F. Marmol (1997) "On the properties of the Dickey-Pantula test against fractional alternatives". *Economics Letters*, 57, 11-16.

Engle, R.F. (1982) "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation". *Econometrica*, 50, 987-1007.

Engle, R.F. (1983) "Estimates of the variance of U.S. inflation based upon the ARCH model". *Journal of Money, Credit and Banking*, 15, 286-301.

Engle, R.F. (1995) *ARCH: Selecting Readings*. Oxford University Press, Oxford.

Engle, R.F. y T. Bollerslev (1986) "Modelling the persistence of conditional variance". *Econometric Reviews*, 5, 1-50, 81-87.

Fox, R. y M.S. Taqqu (1986) "Large sample properties of parameter estimates for strongly dependent stationary gaussian time series". *Annals of Statistics*, 14, 221-238.

Fuller, W.A. (1996) *Introduction to Statistical Time Series*. 2ª ed. Wiley, New York.

Geweke, J. y S. Porter-Hudak (1983) "Estimation and application of long memory time series models". *Journal of Time Series Analysis*, 4, 221-238.

Ghysels, E., A.C. Harvey y E. Renault (1996) "Stochastic Volatility". En G.S. Maddala y C.R. Rao (eds.), *Handbook of Statistics*, vol. 14, 119-191. North Holland, Amsterdam.

Giriatis, L. y D. Surgalis (1990) "A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle´s estimate". *Probability Theory and Related Fields*, 86, 87-104.

Gradshteyn, I.S. e I.M. Ryzhik (1980) *Table of Integrals, Series and Products*. Academic Press, San Diego.

Granger, C.W.J. (1966) "The typical spectral shape of an economic variable". *Econometrica*, 34, 150-161.

Granger, C.W.J. (1980) "Long memory relationships and the aggregation of dynamic models". *Journal of Econometrics*, 14, 227-238.

Granger, C.W.J. (1990) "Aggregation of time series variables: a survey". En T. Baker y M.H. Pesaran (eds.), *Disaggregation in Econometric Modelling*, 17-34. Routledge, London y New York.

Granger, C.W.J. y R. Joyeux (1980) "An introduction to long-memory time series models and fractional differencing". *Journal of Time Series Analysis*, 1, 15-29.

Granger, C.W.J. y F. Marmol (1998) "The correlogram of a long memory process plus a simple noise". *Working Paper* 98-14, Universidad Carlos III de Madrid.

Hannan, E.J. (1976) "The asymptotic distribution of serial covariances". *Annals of Statistics*, 4, 396-399.

Hannan, E.J. y M. Deistler (1988) *Statistical Theory of Linear Systems*. Wiley, New York.

Hannan, E.J. y C.C. Heyde (1972) "On limit theorems for quadratic functions of discrete time series". *Annals of Mathematical Statistics*, 43, 2058-2066.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.

Harvey, A.C. (1990) *The Econometric Analysis of Time Series*. 2ª ed. Phillip Allan, New York.

Harvey, A.C. (1993) *Time Series Models*. 2ª ed. Harvester Wheatsheaf, New York.

Harvey, A.C. (1998) "Long memory in stochastic volatility". En J. Knight y S. Satchell (eds.), *Forecasting Volatility in Financial Markets*, 307-320. Butterworth-Haineman, Oxford.

Harvey, A.C. y N.G. Shephard (1993) "Estimation and testing of stochastic volatility models". *STICERD Econometrics Discussion Paper* EM/93/268, London School of Economics.

Harvey, A.C. y N.G. Shephard (1996) "Estimation of an asymmetric stochastic volatility model for asset returns". *Journal of Business and Economic Statistics*, 14, 429-434.

Harvey, A.C., E. Ruiz y N.G. Shephard (1994) "Multivariate stochastic variance models". *Review of Economic Studies*, 61, 247-264.

Hassler, U. y J. Wolters (1994) "On the power of unit root tests against fractional alternatives". *Economics Letters*, 45, 1-5.

Haubrich, J.G. y A.W. Lo (1989) "The sources and nature of long-term memory in the business cycle". *NBER Working Paper* 2951.

Hauser, M.A. (1999) "Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo study". *Journal of Statistical Planning and Inference*, 80, 229-255.

He, C. y T. Terasvirta (1999a) "Properties of moments of a family of GARCH processes". *Journal of Econometrics*, 92, 173-192.

He, C. y T. Terasvirta (1999b) "Fourth moment structure of the GARCH(p,q) process". *Econometric Theory*, 15, 824-846.

He, S. (1996) "A note on the asymptotic normality of sample autocorrelations for a linear stationary sequence". *Journal of Multivariate Analysis*, 2, 182-188.

Henry, M. y R. Payne (1997) "An investigation of long-range dependence in intra-day foreign exchange rate volatility". *Discussion Paper* 264, Financial Markets Group, London School of Economics.

Hosking, J.R.M. (1981) "Fractional differencing". *Biometrika*, 68, 165-176.

Hosking, J.R.M. (1996) "Asymptotic distributions of the sample mean, autocovariances and autocorrelations of long-memory time series". *Journal of Econometrics*, 73, 261-284.

Hosoya, Y. (1996) "The quasi-likelihood approach to statistical inference on multiple time-series with long-range dependence". *Journal of Econometrics*, 73, 217-236.

Hosoya, Y. (1997) "A limit theorem for long-range dependence and statistical inference on related models". *The Annals of Statistics*, 25, 105-137.

Hsieh, D.A. (1989) "Modeling heteroscedasticity in daily foreign-exchange rates". *Journal of Business and Economic Statistics*, 7, 307-317.

Hsu, N-J. y F.J. Breidt (1997) "Bayesian analysis of long memory stochastic volatility models". *Technical report* 97-4, Department of Statistics, Iowa State University.

Hurst, H.E. (1951) "Methods for storage capacity of reservoirs". *Trans. Am. Soc. Civil Engineers*, 116, 770-799.

Hurvich, C.M., R. Deo y J. Brodsky (1998) "The mean square error of Geweke and Porter-Hudak estimator of the memory parameter in a long-memory time series". *Journal of Time Series Analysis*, 19, 19-46.

Jacquier, E., N.G. Polson y P.E. Rossi (1994) "Bayesian analysis of stochastic volatility models". *Journal of Business and Economic Statistics*, 12, 371-417.

Johnson, N.L. y S. Kotz (1970) *Continuous Uivariate Distributions, 2*. Wiley, New York.

Karanasos, M (1999) "The second moment and the autocovariance function of the squared errors of the GARCH model". *Journal of Econometrics*, 90, 63-76.

Kendall, M.G. (1954) "Note on bias in the estimation of autocorrelation". *Biometrika*, 41, 403-404.

Kendall, S.M., A. Stuart y J.K. Ord (1968) *The Advanced Theory of Statistics*, vol. III. Griffin, London.

Kim, S., N. Shephard y S. Chib (1998) "Stochastic volatility: likelihood inference and comparison with ARCH models". *Review of Economic Studies*, 65, 361-393.

Kramer, W. y R. Runde (1994) "Some pitfalls in using empirical autocorrelations to test for zero correlation among common stock returns". En J. Kaehler y P. Kugler (eds.), *Econometric Analysis of financial Markets*, 1-10. Springer-Verlag, New York.

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

Lee, D. y P. Schmidt (1996) "On the power of the KPSS test of stationarity against fractionally integrated alternatives". *Journal of Econometrics*, 73, 285-302.

Lee, H.S. y C. Amsler (1997) "Consistency of the KPSS unit root test against fractionally integrated alternative". *Economics Letters*, 55, 151-160.

Lee, S.W. y B.E. Hansen (1994) "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator". *Econometric Theory*, 10, 539-569.

Li, W.K. y T.K. Mak (1994) "On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity". *Journal of Time Series Analysis*, 15, 627-636.

Li, W.K. y A.I. McLeod (1986) "Fractional time series modelling". *Biometrika*, 73, 217-221

Lo, A.W. (1991). "Long-term memory in stock market prices". *Econometrica*, 59, 1279-1313.

Lobato, I. y P.M. Robinson (1998) "A nonparametric test for I(0)". *Review of Economic Studies*, 65, 475-495.

Lobato, I. y N.E. Savin (1998) "Real and spurious long-memory properties of stock-market data". *Journal of Business and Economic Statistics*, 16, 261-283.

Lumsdaine, R.L. (1996) "Asymptotic properties of the quasi-maximum likelihood estimator in the GARCH(1,1) and IGARCH(1,1) models". *Econometrica*, 64, 575-596.

Mandelbrot, B.B. (1963) "The variation of certain speculative prices". *Journal of Business*, 36, 394-419.

Mandelbrot, B.B. y J.W. Van Ness (1968) "Fractional Brownian motions, fractional noises and applications". *SIAM Review*, 10, 4, 422-437.

Marmol, F. (1998) "Searching for fractional evidence using combined unit root tests". *Working Paper* 98-39, Universidad Carlos III de Madrid.

Marriot, F.H. y J.A. Pope (1954) "Bias in the estimation of autocorrelations". *Biometrika*, 41, 390-402.

Melino, A. y M. Turnbull (1990) "Pricing foreign currency options with stochastic volatility". *Journal of Econometrics*, 45, 239-265.

Milhoj, A. (1985) "The moment structure of ARCH processes". *Scandinavian Journal of Statistics*, 12, 281-292.

McLeod, A.I. y W.K. Li (1983) "Diagnostic cheking ARMA time series models using squared residual autocorrelation". *Journal of Time Series Analysis*, 4, 269-273.

Murray, W. (1979) *Numerical Methods for Unconstrained Optimization*. Academic Press, London.

Nelson, D.B. (1988) "The time series behaviour of stock market volatility and returns". Tesis doctoral. Massachusetts Institute of Technology, Cambridge, MA.

Nelson, D.B. (1990) "Stationarity and persistence in the GARCH(1,1) model". *Econometric Theory*, 6, 318-334.

Nelson, D.B. (1991) "Conditional heteroscedasticity in asset returns: a new approach". *Econometrica*, 59, 347-370.

Nelson, D.B. y C.Q. Cao (1992) "Inequality constraints in the univariate GARCH model". *Journal of Business and Economic Statistics*, 10, 229-235.

Newbold, P. y C. Agiakloglou (1993) "Bias in the sample autocorrelations of fractional noise". *Biometrika*, 80, 698-702.

Novales, A. y M. Gracia-Diez (1993) "Guía para la estimación de modelos ARCH". *Estadística Española*, vol. 35, 132, 5-38.

Olmeda, I. (1998) "Algunos resultados sobre memoria de largo plazo en series bursátiles". *Moneda y Crédito*, 207, 145-203.

Palm, F.C. (1996) "GARCH models for volatility". En G.S. Maddala y C.R. Rao (eds.), *Handbook of Statistics,* vol. 14, 209-240. Elsevier, Amsterdam.

Poon, S.H. y S.J. Taylor (1992) "Stock returns and volatility: an empirical study of the UK stock market". *Journal of Banking and Finance*, 16, 37-59.

Priestley, M.B. (1981) *Spectral Analysis and Time Series*. Academic Press, London.

Ray, B. y R. Tsay (2000) "Long-range dependence in daily stock volatilities". *Journal of Business and Economic Statistics*, 18, 254-262.

Robinson, P.M. (1978) "Statistical inference for random coefficient autoregressive model". *Scandinavian Journal of Statistics*, 5, 163-168.

Robinson, P.M. (1991) "Testing for strong serial correlation and dynamic conditional heteroskedaticity in multiple regression". *Journal of Econometrics*, 47, 67-84.

Robinson, P.M. (1994a) "Efficient tests of nonstationary hypotheses". *Journal of the American Statistical Association*, 89, 1420-1437.

Robinson, P.M. (1994b) "Time series with strong dependence". En C.A. Sims (ed.), *Advances in Econometrics: Sixth World Congress,* vol. 1, 47-96. Cambridge University Press, Cambridge.

Robinson, P.M. (1995a) "Log-periodogram regression of time series with long-run dependence". *Annals of Statistics*, 23, 1048-1072.

Robinson, P.M. (1995b) "Gaussian semiparametric estimation of long-range dependence". *Annals of Statistics*, 23, 1630-1661.

Robinson P.M. y P. Zaffaroni (1997) "Modelling nonlinearity and long memory in time series". En C.D. Cutler y D.T. Kaplan (eds.), *Nonlinear Dynamics and Time Series.* Fields Institute Communications, 11, 161-170.

Robinson P.M. y P. Zaffaroni (1998) "Nonlinear time series with long memory: a model for stochastic volatility". *Journal of Statistical Planning and Inference*, 68, 359-371.

Roy, R. (1989) "Asymptotic covariance structure of serial correlations in multivariate time series". *Biometrika*, 76, 824-827.

Ruiz, E. (1994a) "Modelos para series temporales heterocedásticas". *Cuadernos Económicos de I.C.E.*, 56, 73-108.

Ruiz, E. (1994b) "Quasi-maximum likelihood estimation of stochastic volatility models". *Journal of Econometrics*, 63, 289-306.

Sandmann, G. y S.J. Koopman (1998) "Estimation of stochastic volatility models via Monte Carlo maximum likelihood". *Journal of Econometrics*, 87, 271-301.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Shephard, N.G. (1996) "Statistical aspects of ARCH and stochastic volatility". En D.R. Cox, D.V. Hinkley y O.E. Barndorff-Nielsen (eds.), *Time Series Models in Econometrics, Finance and other fields*, 1-67. Chapman and Hall, London.

Shephard, N y M.K. Pitt (1997) "Likelihood analysis of non-gaussian parameter-driven models". *Biometrika*, 84, 653-667.

Silverman, B.W. (1990) *Density Estimation for Statistics and Data Analysis*. 2ª ed. Chapman and Hall, London.

Smith, A.A., F. Sowell y S.E. Zin (1997) "Fractional integration with drift: estimation in small samples". *Empirical Economics*, 22, 103-116.

So, M.K.P. (1999) "Time series with additive noise". *Biometrika*, 86, 474-482.

Sowell, F. (1990) "The fractional unit root distribution". *Econometrica*, 58, 495-505.

Sowell, F. (1992a) "Maximum likelihood estimation of stationary univariate fractionally integrated time series models". *Journal of Econometrics*, 53, 165-188.

Sowell, F. (1992b) "Modelling long-run behaviour with the fractional ARIMA model". *Journal of Monetary Economics*, 29, 277-302.

Taylor, S. (1986) *Modelling Financial Time Series.* Wiley, New York.

Taylor, S. (1994) "Modeling stochastic volatility: a review and comparative study". *Mathematical Finance*, 4, 183-204.

Velasco, C. (1999a) "Gaussian semiparametric estimation of nonstationary time series". *Journal of Time Series Analysis*, 20, 87-127.

Velasco, C. (1999b) "Non-stationary log-periodogram regression". *Journal of Econometrics*, 91, 325-371.

Velasco, C. (2000) "Non-gaussian log-periodogram regression". *Econometric Theory*, 16, 44-79.

Velasco, C. y P.M. Robinson (2000) "Whittle pseudo-maximum likelihood estimates of nonstationary time series". *Journal of the American Statistical Society*, próximamente.

Weiss, A.A. (1986) "Asymptotic theory for ARCH models: estimation and testing". *Econometric Theory*, 2, 107-131.

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

Whittle, P. (1962) "Gaussian estimation in stationary time series". *Bull. Internat. Statistical Institute*, 39, 105-129.

Wright, J.H. (1999) "A new estimator of the fractionally integrated stochastic volatility model". *Economics Letters*, 63, 295-303.

Yajima, Y. (1985) "On estimation of long-memory time series models". *Australian Journal of Statistics*, 27, 303-320.

Zaffaroni, P. (1997) "Gaussian estimation of long-range dependent volatility in asset prices". *STICERD Econometrics Discussion Paper* EM/97/329. London School of Economics.

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES