

Series temporales. Estacionariedad, raíces unitarias

Alfonso Novales
Departamento de Economía Cuantitativa
Universidad Complutense

Septiembre 2017
Preliminary version
No citar sin permiso del autor
©Copyright A.Novales 2014

Contents

1	Procesos estocásticos y series temporales de datos	2
1.1	Some simple stochastic processes. Properties	3
1.2	Modelos ARMA	13
1.2.1	<i>Wald decomposition</i>	13
1.3	Stationarity, mean reversion, impulse responses	14
1.3.1	Características de una variable estacionaria: Resumen . .	17
1.3.2	Numerical exercise: Simulating simple stochastic processes	18
1.4	Ejercicios con datos financieros	21
1.5	Predicción con modelos ARMA	23
1.5.1	Forecasting with some simple processes	23
1.5.2	Expresiones analíticas para la predicción	23
1.5.3	Varianza del error de predicción	24
1.5.4	Recuperación de predicciones	24
1.5.5	Evaluación de previsiones	24
1.6	Identificación de procesos AR y MA	25
1.7	Contrastes de ausencia de autocorrelación (ruido blanco)	25
1.7.1	Coeficientes de autocorrelación	25
1.7.2	Contrastes Portmanteau	26
1.7.3	Ratios de varianza	27
1.8	Contrastes de camino aleatorio	28
1.8.1	Diferencias de varianzas	29
1.9	Variables no estacionarias. Modelos ARIMA	32
1.9.1	Tendencias deterministas y tendencias estocásticas	32
1.10	Regresión espúrea	34

1.10.1	Regresión espúrea bajo tendencias deterministas	35
1.10.2	Regresión espúrea bajo tendencias estocásticas	36
1.11	Tratamiento de tendencias deterministas	38
1.12	Tendencias estocásticas	40
1.13	Variables integradas. Cointegración	42
1.13.1	Contrastes de raíz unitaria	44
1.13.2	Raíces unitarias estacionales	44
1.14	Modelos ARIMA estacionales	45
1.15	Estimación de modelos AR y MA	45
1.16	Contrastes de especificación de un modelo ARIMA	45
1.17	Cointegración	45
1.17.1	¿Cómo tratar las relaciones entre variables en presencia de tendencias estocásticas?	45
2	Appendix 1: Valoración por simulación	47
3	Sobre simulación de procesos brownianos	49
3.1	Distribución de los cambios en precio	49
3.2	Distribución del logaritmo del precio	50
3.3	Distribución de la rentabilidad	51
4	* IGNORAR Appendix 2: Sesgo asintótico en el modelo AR(1) con autocorrelación	51

1 Procesos estocásticos y series temporales de datos

En general, contamos con observaciones históricas acerca de una o varias variables (precios, rentabilidades, etc.) y queremos calcular medidas de posición central, de dispersión y de correlación con el objeto de resumir las propiedades básicas de dichos datos.

El conjunto de datos observados define un histograma de frecuencias, o distribución muestral de frecuencias, que contiene *toda* la información disponible acerca de la variable considerada. Un histograma de frecuencias es similar a una distribución de frecuencias, pero es diferente de ella. Para entender la diferencia entre ambos, hemos de comprender el concepto de proceso estocástico, y el modo de utilizarlo en el análisis de datos de series temporales.

Un *proceso estocástico* $X_t, t = 1, 2, 3, \dots$ es una sucesión de variables aleatorias, indexadas por la variable tiempo. Las variables aleatorias pueden ser independientes entre sí o no, y pueden tener la misma distribución de probabilidad, o una distribución de probabilidad diferente.

Cada dato de una serie temporal debe interpretarse como una muestra de tamaño 1 de la distribución de probabilidad correspondiente a la variable aleatoria de ese instante. Por ej., el dato de cierre del IBEX35 (suponiendo que disponemos de datos de cierre diarios) de hoy es una *realización*, es decir, una

muestra de tamaño 1 de la variable aleatoria "precio de la cesta IBEX35" (como índice) el día de hoy. La distribución de probabilidad de esta variable hoy puede ser diferente de la variable aleatoria IBEX35 hace un año por tener, por ejemplo, una esperanza matemática menor, una volatilidad mayor, o no ser Normal, mientras que hace un año sí lo era.

Vamos a suponer generalmente que las variables X_t tienen todas la misma distribución de probabilidad, y son independientes entre sí. Este es el caso más sencillo, y constituye un proceso de *ruido blanco*. Sólo en este caso está totalmente justificado la utilización de momentos muestrales como características de "la variable X ". Esta observación debe servir como llamada de atención al lector, dada la excesiva frecuencia con que se calculan estadísticos muestrales, calculados con datos históricos, para representar características de una variable; por ej., la desviación típica de la rentabilidad bursátil de un determinado mercado.

En algunos casos, podremos suponer que X_t representa la rentabilidad de un activo que tiene las características que hemos expuesto de independencia temporal o, al menos, ausencia de autocorrelación. En otros casos, cuando dicha rentabilidad tenga autocorrelación, tendremos que especificar un modelo de evolución temporal para ella, y trabajar con la innovación de dicho proceso, que debe tener estructura de ruido blanco. La autocorrelación hace referencia a la persistencia en rentabilidades, o correlación entre el valor presente y los valores pasados de una variable.

En ocasiones, quizá sin autocorrelación en rentabilidades, se produce dependencia temporal en segundos momentos (varianzas), como consecuencia de que este momento de segundo orden es cambiante en el tiempo de modo gradual. De ese modo, la varianza en un determinado período no es independiente de la varianza en períodos previos. En tales casos, habremos de modelizar la evolución temporal de la varianza mediante algún proceso determinado, y examinaremos las rentabilidades estandarizadas, es decir, divididas por la desviación típica estimada para cada observación. Si no había autocorrelación en rentabilidades, y el modelo de varianza está bien especificado, la rentabilidades estandarizadas, así como sus cuadrados, deben estar libres de autocorrelación.

Estas cuestiones son objeto de estudio en este tema. La modelización de series temporales tiene asimismo un interés evidente para la predicción. En términos de modelización de variables financieras, el análisis de series temporales nos permite analizar, por ejemplo, si una determinada rentabilidad presenta reversión a la media. Asimismo, podemos contrastar si una determinada especificación analítica propuesta en la literatura es consistente con las características estadísticas de la serie temporal de datos de una determinada rentabilidad.

1.1 Some simple stochastic processes. Properties

A stochastic process is a sequence of random variables indexed by time. Each of the random variables in a stochastic process, corresponding to a given time index t , has its own probability distribution. These distributions can be different, and any two of the random variables in a stochastic process may either exhibit

dependence of some type or be independent from each other.

A *white noise* process is,

$$y_t = \varepsilon_t, \quad t = 1, 2, 3, \dots$$

where $\varepsilon_t, t = 1, 2, \dots$ is a sequence of independent, identically distributed zero-mean random variables, known as the *innovation* to the process. A white noise is sometimes defined by adding the assumption that ε_t has a Normal distribution. The mathematical expectation of a white noise is zero, and its variance is constant: $Var(y_t) = \sigma_\varepsilon^2$. More generally, we could consider a *white noise with drift*, by incorporating a constant term in the process,

$$y_t = a + \varepsilon_t, \quad t = 1, 2, 3, \dots$$

with mathematical expectation $E(y_t) = a$, and variance: $Var(y_t) = \sigma_\varepsilon^2$.

The future value of a white noise with drift obeys,

$$y_{t+s} = a + \varepsilon_{t+s},$$

so that, if we try to forecast any future value of a white noise on the basis of the information available¹ at time t , we would have:

$$E_t y_{t+s} = a + E_t \varepsilon_{t+s} = a,$$

because of the properties of the ε_t -process. That is, the prediction of a future value of a white noise is given by the mean of the process. In that sense, a white noise process is *unpredictable*. The prediction of such process is given by the mean of the process, with no effect from previously observed values. Because of that, the history of a white noise process is irrelevant to forecast its future values. No matter how many data points we have, we will not use them to forecast a white noise.

A *random walk with drift* is a process,

$$y_t = a + y_{t-1} + \varepsilon_t, \quad t = 1, 2, 3, \dots \quad (1)$$

so that its first differences are white noise. If $y_t = \ln(P_t)$ is the log of some market price, then its return $r_t = \ln(P_t) - \ln(P_{t-1})$, will be a white noise, as we already mentioned: $r_t = a + \varepsilon_t, \quad t = 1, 2, 3, \dots$. A random walk does not have a well defined mean or variance.

In the case of a *random walk without drift*, we have,

$$y_{t+s} = y_{t+s-1} + \varepsilon_{t+s}, \quad s \geq 1$$

so that we have the sequence of forecasts:

¹That amounts to constructing the forecast by application of the conditional expectation operator to the analytical representation of the future value being predicted, where the conditional expectation is formed with respect to the sigma algebra of events known at time t .

$$\begin{aligned} E_t y_{t+1} &= E_t y_t + E_t \varepsilon_{t+1} = y_t, \\ E_t y_{t+2} &= E_t y_{t+1} + E_t \varepsilon_{t+2} = E_t y_{t+1} = y_t \end{aligned}$$

and the same for all future variables. In this case, the history of a random walk process is relevant to forecast its future values, but only through the last observation. All data points other than the last one are ignored when forecasting a random walk process.

First order autoregressive processes

A first order autoregressive process with zero mean has the representation,

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad |\phi| < 1,$$

and can also be represented as,

$$y_t = \sum_{s=0}^{\infty} \phi^s \varepsilon_{t-s}$$

the right hand side having a finite variance under the assumption that $Var(\varepsilon_t) = \sigma_\varepsilon^2$ only if $|\phi| < 1$. If that is the case, we would have:²

$$\begin{aligned} E(y_t) &= E\left(\sum_{s=0}^{\infty} \phi^s \varepsilon_{t-s}\right) = \sum_{s=0}^{\infty} E(\phi^s \varepsilon_{t-s}) = \sum_{s=0}^{\infty} \phi^s \cdot E\varepsilon_{t-s} = 0; \\ Var(y_t) &= Var\left(\sum_{s=0}^{\infty} \phi^s \varepsilon_{t-s}\right) = \sum_{s=0}^{\infty} Var(\phi^s \varepsilon_{t-s}) = \sum_{s=0}^{\infty} \phi^{2s} Var(\varepsilon_{t-s}) = \sum_{s=0}^{\infty} \phi^{2s} \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \phi^2} \end{aligned}$$

where we have used : $Cov(\varepsilon_t, \varepsilon_s) = 0, \forall t \neq s$.

A first order autoregressive process *with constant* has the representation,

$$y_t = a + \phi y_{t-1} + \varepsilon_t, \quad |\phi| < 1,$$

Let us assume by now that the mathematical expectation exists and is finite. Under that assumption, $Ey_t = Ey_{t-1}$, and we have:

$$Ey_t = a + E(\phi y_{t-1}) + E\varepsilon_t = a + \phi Ey_t$$

so that: $Ey_t = \frac{a}{1-\phi}$. To find out the variance of the process, we can iterate on its representation:

$$\begin{aligned} y_t &= a + \phi y_{t-1} + \varepsilon_t = a + \phi(a + \phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \\ &= a(1 + \phi + \phi^2 + \dots + \phi^{s-1}) + \phi^s y_{t-s} + (\phi^{s-1} \varepsilon_{t-s+1} + \dots + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t) \end{aligned}$$

²Recordemos que la suma de una progresión geométrica: $a, ra, r^2a, r^3a, \dots, r^na$ con un número finito de términos existe siempre, con independencia del valor de la razón r . dicha suma es: $S = a \frac{1-r^{n+1}}{1-r}$. La suma de una progresión geométrica con infinitos términos existe sólo si la razón r es menor que 1 en valor absoluto, siendo entonces su suma: $S = \frac{a}{1-r}$

and if we proceed indefinitely, we get

$$y_t = a(1 + \phi + \phi^2 + \dots) + (\dots + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t)$$

since $\lim_{s \rightarrow \infty} \phi^s y_{t-s} = 0$.³ Then, taking the variance of this expression:

$$\text{Var}(y_t) = \text{Var}(\dots + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t) = \sum_{s=0}^{\infty} \phi^{2s} \sigma_{\varepsilon}^2 = \frac{\sigma_{\varepsilon}^2}{1 - \phi^2}$$

so that the variance of the y_t -process increases with the variance of the innovation, σ_{ε}^2 , but it is also higher the closer is ϕ to 1.

Predictions from a first order autoregression can be obtained by,

$$\begin{aligned} E_t y_{t+1} &= \phi E_t y_t + E_t \varepsilon_{t+1} = \phi y_t, \\ E_t y_{t+2} &= E_t (\phi y_{t+1}) + E_t \varepsilon_{t+2} = \phi^2 E_t y_{t+1} = \phi^2 y_t \end{aligned}$$

and, in general,

$$E_t y_{t+s} = \phi^s y_t, \quad s \geq 1$$

which is the reason to impose the constraint $|\phi| < 1$.

Si el proceso siguiese una estructura dependiente de su pasado a partir de un instante inicial $t = 1$ tendríamos:

$$y_t = a + \phi y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, \quad -1 < \phi < 1$$

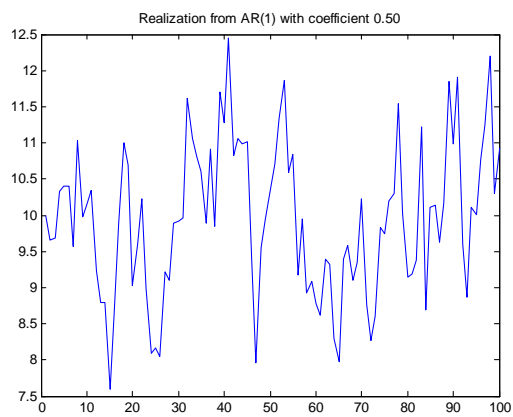
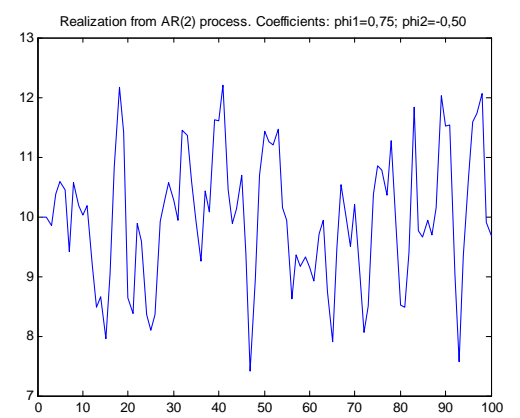
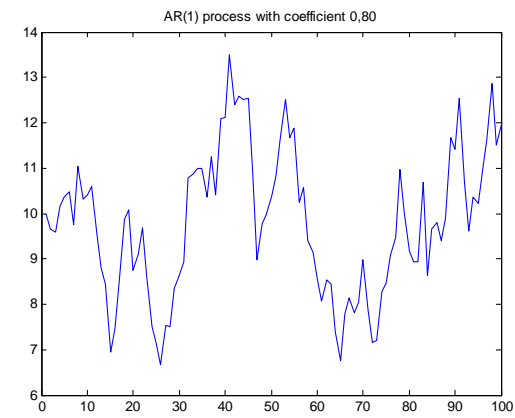
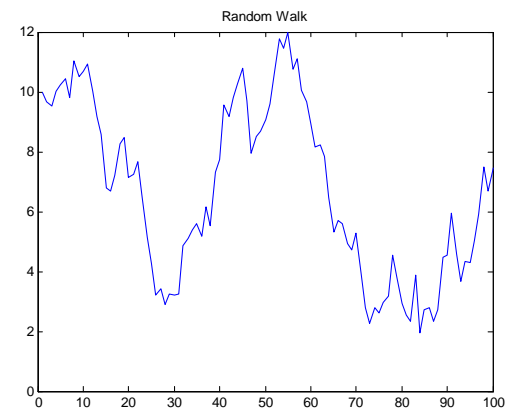
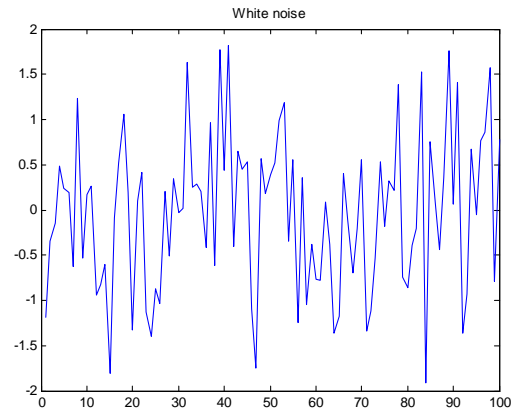
y sus propiedades serían algo distintas, con:

$$y_t = a \frac{1 - \phi^t}{1 - \phi} + \phi^t y_0 + \sum_{s=1}^t \phi^{s-1} \varepsilon_s$$

Como vemos, es un proceso en el que el valor inicial tienen una influencia que va haciéndose menor con el paso del tiempo. Lógicamente, cuando ha pasado suficiente tiempo, este proceso sería indistinguible del que antes vimos, que no tenía punto inicial.

The following figure shows a single realization of a white noise process, with an innovation with unit variance, and a realization from a random walk. The innovation in the random walk is the white noise that appears in the figure to the left. In the lower row, we have single realizations from stationary AR(1) and AR(2) processes. In all cases, the same realization of the white noise that appears in the upper row has been used as innovation.

³This is the limit of a random variable, and an appropriate limit concept must be used. It suffices to say that the power of ρ going to zero justifies the zero limit for the product random variable.



En este proceso AR(1), si ε_t sigue una distribución Normal $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, con $\text{Corr}(\varepsilon_t, \varepsilon_{t-k}) = 0 \ \forall k$, entonces y_t sigue una distribución

$$y_t \sim N\left(\frac{a}{1-\rho}, \frac{\sigma_\varepsilon^2}{1-\rho^2}\right)$$

Esta es la *distribución marginal o incondicional*, de y_t .

Los momentos condicionales de este proceso son,

$$E_{t-1}(y_t) = a + \rho y_{t-1}; \quad \text{Var}(y_t) = \sigma_\varepsilon^2$$

Condional en la historia pasada de y_t , sin incluir el dato de fecha t , la *distribución de probabilidad condicional* de y_t es,

$$y_t/\Omega_{t-1} \sim N(a + \rho y_{t-1}, \sigma_u^2).$$

La varianza condicional de y_t es igual a la varianza de u_t , σ_u^2 , mientras que la varianza incondicional de y_t es siempre mayor que σ_u^2 . Por tanto, la varianza condicional, que es constante, es siempre inferior a la varianza incondicional. La diferencia entre ambas es tanto mayor cuanto más se aproxima el valor numérico del coeficiente ρ a -1 ó $+1$. En ambos límites, además, la varianza del proceso autoregresivo de primer orden tiende a infinito. La varianza incondicional de y_t es tanto mayor cuanto más se acerque el parámetro ρ a 1, creciendo dicha varianza sin límite. Sin embargo, la varianza condicional es siempre σ_u^2 , con independencia del valor numérico del parámetro ρ .

La esperanza condicional $E(y_t/y_{t-1}) = a + \rho y_{t-1}$ es cambiante en el tiempo, en función de la información disponible en cada instante. La esperanza incondicional $E(y_t) = \frac{a}{1-\rho}$ es la mejor predicción que podríamos proporcionar del proceso y_t sin disponer de información muestral para el mismo, pero conociendo la estructura estocástica de dicho proceso, incluidos los valores numéricos de los parámetros. Si dispusiéramos de información muestral pero ignorásemos el proceso estocástico que sigue y_t , podríamos sustituir $\frac{a}{1-\rho}$ por la media muestral. Esta es la mejor predicción en cuanto a minimizar el error cuadrático medio de la predicción, pero no es la predicción óptima, que es $E(y_t/y_{t-1}) = a + \rho y_{t-1}$ y requiere estimaciones de los parámetros. Esta predicción minimiza el error cuadrático medio condicional.

Autocorrelación, persistencia. Funciones de autocorrelación simple y parcial

La autocorrelación de un proceso estocástico se mide a través de las funciones de autocorrelación simple y parcial. The parameter ρ is sometimes known as the *persistence* of the process. As the previous expression shows, an increase or decrease in y_t will show up in any future y_{t+s} , although the influence of that y_t -value will gradually disappear over time, according to the value of ρ . A value of ρ close to 1 will introduce high persistence in the process, the opposite being true for ρ close to zero.

The *autocovariance function* is defined as the set of covariances between the values a given process at two different points in time. At this point, we make

the assumption that the covariance between the values of y_t at two points in time depends just on the distance between them, i.e.:

$$Cov(y_t, y_{t+s}) = Cov(y_p, y_{p+s}), \forall p, t$$

In the case of the first order autoregressive process AR(1) at two points in time, the covariance function is:

$$Cov(y_t, y_{t+s}) = \rho^s Var(y_t), \quad s \geq 0,$$

so that the linear correlation is:

$$Corr(y_t, y_{t+s}) = \frac{Cov(y_t, y_{t+s})}{Var(y_t)} = \rho^s$$

which dies away at a rate of ρ . The set of correlations obtained for $s = 0, 1, 2, \dots$ is the *simple autocorrelation function*. In an autoregressive process with a value of ρ close to 1, the correlation of y_t with past values will be sizeable for a number of periods, although it decreases to zero at the rate ρ . If $\rho > 0$, all values of the autocorrelation function will be positive. If $\rho < 0$, the autocorrelation function will go to zero alternating in sign.

The *partial autocorrelation function* of a process is defined by the coefficient in the last lag in a sequence of regressions of y_t on a set of lags $(y_{t-1}, y_{t-2}, \dots, y_{t-p})$. To estimate the partial autocorrelation function, we run regressions:

$$\begin{aligned} y_t &= \beta_{01} + \beta_{11}y_{t-1} + u_{1t} \\ y_t &= \beta_{02} + \beta_{12}y_{t-1} + \beta_{22}y_{t-2} + u_{2t} \\ y_t &= \beta_{03} + \beta_{13}y_{t-1} + \beta_{23}y_{t-2} + \beta_{33}y_{t-3} + u_{3t} \end{aligned}$$

and the sequence $\{\beta_{11}, \beta_{22}, \beta_{33}, \dots\}$ is the partial autocorrelation function of y_t . In a first order autoregressive process, the partial autocorrelation function is $\{\beta, 0, 0, \dots\}$. It has only a non-zero value, the first one, being equal to the slope coefficient in the autoregressive process.

Las funciones de autocorrelación simple y parcial no varían por la introducción de una constante en el modelo autoregresivo.

Cyclical fluctuations

Another important characteristic of economic time series is the possibility that they exhibit cyclical fluctuations. In fact, first order autoregressive processes may display a shape similar to that of many economic time series, although to produce regular cycles we need a second order autoregressive processes,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t,$$

with ε_t being an innovation, a sequence of independent and identically distributed over time. Using the lag operator: $B^s y_t = y_{t-s}$ in the representation of the process:

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = (1 - \phi_1 B - \phi_2 B^2) y_t = \varepsilon_t,$$

The dynamics of this process is characterized by the roots of its characteristic equation,

$$1 - \phi_1 B - \phi_2 B^2 = (1 - \lambda_+ B)(1 - \lambda_- B) = 0$$

which are given by:

$$\lambda_+, \lambda_- = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}$$

Stationary second order autoregressions have the two roots of the characteristic equation smaller than 1. A root greater than one in absolute size will produce an explosive behavior. As we will see below, a root equal to 1 signals nonstationarity, although the sample realization will not be explosive. It will display extremely persistent fluctuations, very rarely crossing its mean, as it was the case with a random walk. This is very clear in the similar representation of a random walk: $(1 - B) y_t = \varepsilon_t$.

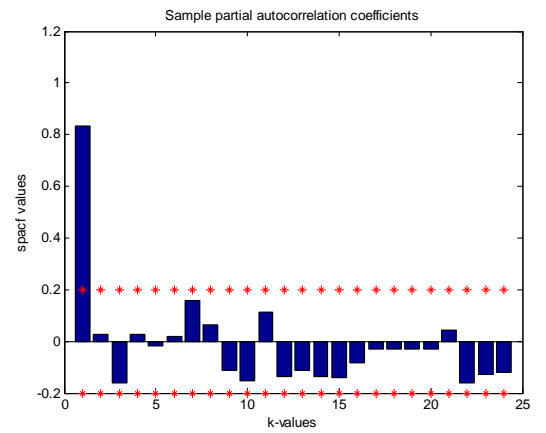
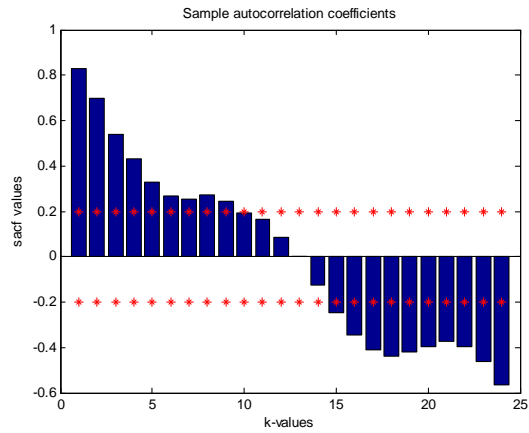
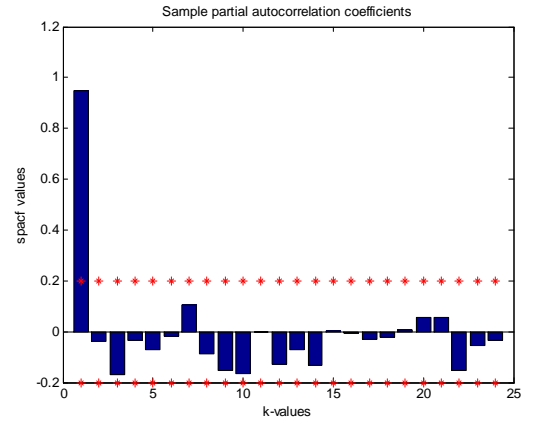
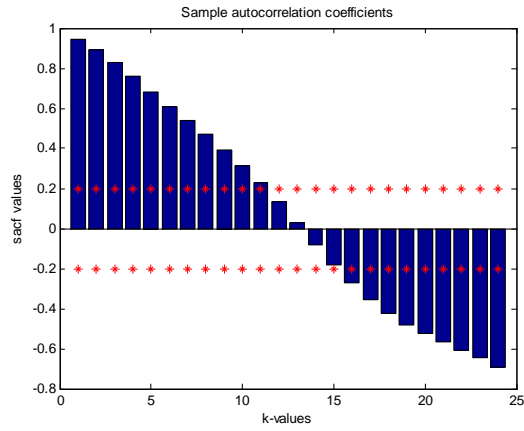
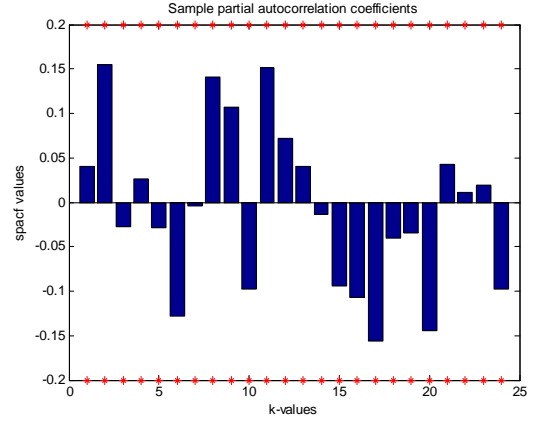
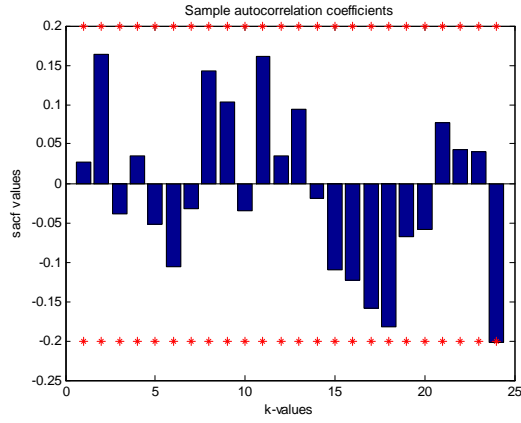
Since the characteristic equation is now of second degree, it might have as roots two conjugate complex numbers. When that is the case, the autoregressive process displays cyclical fluctuations. The response of y_t to an innovation ε_t will also display cyclical fluctuations, as we will see in dynamic macroeconomic models below.

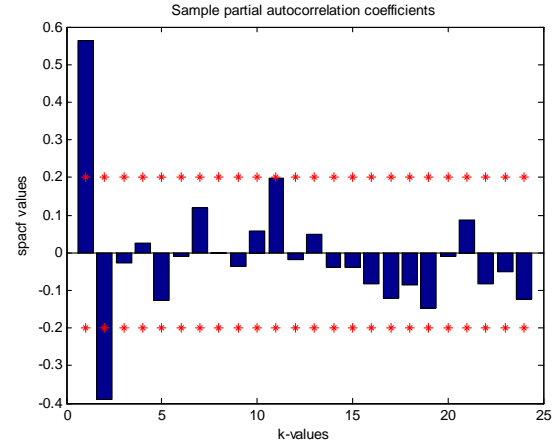
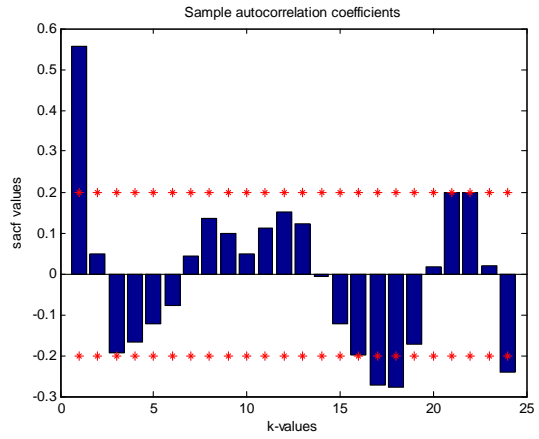
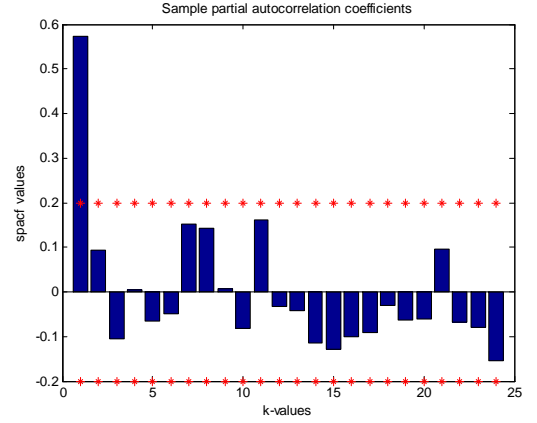
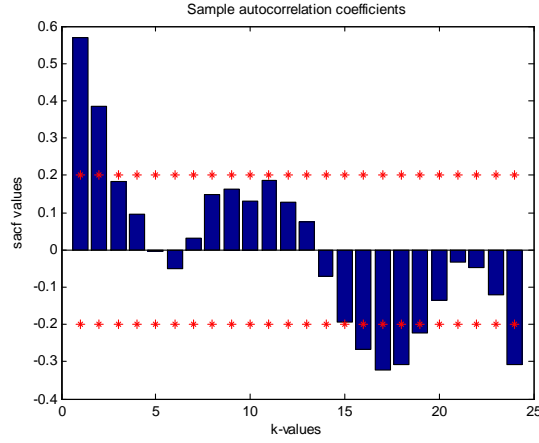
A condition for cycles to arise is $\phi_2 < 0$. When cycles exist, they will have a period T given by:

$$\cos \frac{2\pi}{T} = \frac{\phi_1}{2\sqrt{-\phi_2}}$$

with decay factor: $d = \sqrt{-\phi_2}$.

The following figure shows simple and partial autocorrelation functions for the time series above: a white noise, a random walk, an AR(1) with coefficient 0.80, an AR(1) with coefficient 0.50, and an AR(2) process with coefficients 0.75 and -0.50.





Si la innovación del proceso sigue una distribución Normal

El modelo MA(1)

El modelo MA(1) tiene como estructura estocástica:

$$y_t = a + \varepsilon_t - \theta\varepsilon_{t-1}$$

con propiedades:

$$E(y_t) = a; \text{Var}(y_t) = (1 + \theta^2)\sigma_\varepsilon^2$$

función de autocovarianza:

$$\text{Cov}(y_t, y_{t-1}) = -\theta\sigma_\varepsilon^2; \text{Cov}(y_t, y_{t-s}) = 0, \forall s > 1$$

y función de autocorrelación simple:

$$\text{Corr}(y_t, y_{t-1}) = -\frac{\theta}{1 + \theta^2}; \text{Corr}(y_t, y_{t-s}) = 0, \forall s > 1$$

Estos momentos poblacionales pueden utilizarse para estimar el modelo aplicando el Método Generalizado de Momentos, que consistiría en resolver el sistema que resulta al igualar los valores numéricos de tres momentos muestrales a las expresiones analíticas de dichos momentos en la población:

$$\begin{aligned}\frac{1}{T} \sum_{t=1}^T y_t &= a; \\ \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2 &= (1 + \theta^2) \sigma_\varepsilon^2 \\ \frac{1}{T-1} \sum_{t=2}^T (y_t - \bar{y})(y_{t-1} - \bar{y}) &= -\theta \sigma_\varepsilon^2\end{aligned}$$

para encontrar valores numéricos de $a, \theta, \sigma_\varepsilon^2$. Dividiendo las dos últimas ecuaciones tenemos:

$$-\frac{\theta}{1 + \theta^2} = \frac{\frac{1}{T-1} \sum_{t=2}^T (y_t - \bar{y})(y_{t-1} - \bar{y})}{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2} = \rho_1$$

siendo ρ_1 el primer valor de la función de autocorrelación. Por tanto, tenemos la ecuación cuadrática:

$$\rho_1 \theta^2 + \theta + \rho_1 = 0$$

cuya solución nos da la estimación de θ , obteniendo la estimación de σ_ε^2 a partir de cualquiera de las dos ecuaciones que hemos utilizado. Nótese que la ecuación cuadrática tendrá dos soluciones, pero sólo una será válida, estando en el intervalo $(-1, 1)$.

1.2 Modelos ARMA

Sección 13.5 en Econometría – Novales

1.2.1 *Wald decomposition*

Any covariance-stationary stochastic process y_t admits the presentation:

$$y_t = f(t) + \sum_{s=0}^{\infty} a_s \varepsilon_{t-s}$$

with ε_t white noise ($E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0$, if $t \neq s$), and $a_r = 0$ for $s > r$, for some r , possibly infinite, with $\sum_{s=0}^{\infty} a_s^2 < \infty$. In this representation $f(t)$ is a purely deterministic function which it is therefore perfectly predictable from its own past. Other than for this component, the stochastic process y_t admits a possibly infinite MA representation.

But we also know that such a MA representation can be inverted, provided the roots of the lag polynomial satisfy the appropriate requirements, to obtain a possibly infinite AR representation. A stochastic process is said to have a unit root if the characteristic equation of its AR representation has such a root. A stochastic process may have more than one unit root. The first difference of a process having a unit root is stationary. A stochastic process is said to be integrated of order d , if its d -th order difference is stationary.

1.3 Stationarity, mean reversion, impulse responses

A stochastic process is stationary when the distribution of k -tuples $(y_{t_1}, y_{t_2}, \dots, y_{t_k})$ is the same with independence of the value of k and of the time periods t_1, t_2, \dots, t_k considered. It is a property of any *stationary stochastic process* that the forecast of a future value converges to its mean as the forecast horizon goes to infinity. This is obviously fulfilled in the case of a white noise process. Another characteristic is that any time realization crosses the sample mean often, while a nonstationary process would spend arbitrarily large periods of time at either side of its sample mean. As we have seen above for the first order autoregression, the simple autocorrelation function of a stationary process, made up by the sequence of correlations between any two values of the process, will go to zero relatively quickly, while for processes close to nonstationarity, the autocorrelation function will dye away very slowly.

When they are not subject to an stochastic innovation,⁴ stationary autoregressive processes converge smoothly and relatively quickly to their mathematical expectation. The y_t -process will converge to $\frac{a}{1-\rho}$ either from above or from below, depending on whether the initial value, y_0 , is above or below $\frac{a}{1-\rho}$. The speed of convergence is given by the autoregressive coefficient. When the process is subject to a nontrivial innovation, the convergence in the mean of the process will not be easily observed. This is the case because the process experiences a shock through the innovation process every period, which would start a new convergence that would overlap with the previous one, and so on. Under normal circumstances we will just see a time realization exhibiting fluctuations around the mathematical expectation of the process, unless the process experiences a huge innovation, or the starting condition y_0 is far enough from $\frac{a}{1-\rho}$, in units of its standard deviation, $\sqrt{\frac{\sigma_\varepsilon^2}{1-\rho^2}}$.

Mean reversion [Simple simulation.xls]

The property of converging to the mean after any stochastic shock is called *mean reversion*, and is characteristic of stationary processes. In stationary processes, any shock tends to be corrected over time. This cannot be appreciated because shocks to y_t are just the values of the innovation process, which take place every period. So, the process of mean reversion following a shock gets disturbed by the next shock, and so on. But the stationary process will always react to shocks as trying to return to its mean. Alternatively, a non stationary process will tend to depart from its mean following any shock. As a

⁴That is, if the innovation ε_t has zero variance.

consequence, the successive values of the innovation process ε_t will take y_t every time farther away from its mean.

An alternative way of expressing this property is through the effects of purely transitory shocks or innovations. A stationary process has transitory responses to purely transitory innovations. On the contrary, a nonstationary process may have permanent responses to purely transitory shocks. So, if a stationary variable experiences a one-period shock, its effects may be felt longer than that, but will disappear after a few periods. The effects of such a one-period shock on a nonstationary process will be permanent. A white noise is just an innovation process. The value taken by the white noise process is the same as that taken by its innovation. Hence, the effects of any innovation last as long as the innovation itself, reflecting the stationary of this process.

The situation with a random walk is quite different. A random walk takes a value equal to the one taken the previous period, plus the innovation. Hence, any value of the innovation process gets accumulated in successive values of the random walk. The effects of any shock last forever, reflecting the nonstationary nature of this process.

In a stationary first order autoregression, any value of the innovation ε_t gets incorporated into y_t that same period. It will also have an effect of size $\rho\varepsilon_t$ on y_{t+1} . This is because $y_{t+1} = \rho y_t + \varepsilon_{t+1}$ so, even if $\varepsilon_{t+1} = 0$, the effect of ε_t would still be felt on y_{t+1} through the effect it previously had on y_t .

Impulse response function

This argument suggests how to construct what we know as an *impulse response function*. In the case of a single variable, as it is the case we consider in this section, the impulse response is obtained by setting the innovation to zero every period except by one, in which the impulse is produced. At that time, the innovation takes a unit value.⁵ The impulse response function is the difference between the values taken by the process after the impulse in its innovation, and the values that would have prevailed without the impulse. The response of a white noise to an impulse in its own innovation is a single unit peak at the time of the impulse, since the white noise is every period equal to its innovation, which is zero except at that time period.

Consider now a stationary first order autoregression. A unit innovation at time t^* will have a unit response at that time period, and a response of size ρ^s each period $t + s$, gradually decreasing to zero. If the impulse has size Δ , the response will be $\rho^s\Delta$.

En efecto, supongamos que hasta el instante t^* , el proceso se encuentra en su nivel de equilibrio, que es su esperanza matemática: $y_t = \frac{\alpha}{1-\rho}$. En $t = t^*$, se produce un impulso de tamaño Δ . Esto significa que la innovación del proceso ha sido 0 para $t < t^*$, y tenemos $\varepsilon_{t^*} = \Delta$ and $\varepsilon_t = 0, \forall t > t^*$. El efecto es:

⁵When working with several variables, responses can be obtained for impulses in more than one variable. To make the size of the responses comparable, each innovation is supposed to take a value equal to its standard deviation, which may be quite different for innovations in different variables.

$$y_{t^*} = \alpha + \rho y_{t^*-1} + \Delta = \frac{\alpha}{1-\rho} + \Delta$$

Como consecuencia, tendremos:

$$\begin{aligned} y_{t^*+1} &= \alpha + \rho y_{t^*} = \alpha + \rho(y_{t^*-1} + \Delta) = \alpha + \rho y_{t^*-1} + \rho\Delta = \frac{\alpha}{1-\rho} + \rho\Delta \\ y_{t^*+2} &= \alpha + \rho y_{t^*+1} = \alpha + \rho(\alpha + \rho y_{t^*-1} + \rho\Delta) = \frac{\alpha}{1-\rho} + \rho^2\Delta \\ y_{t^*+3} &= \alpha + \rho y_{t^*+2} = \frac{\alpha}{1-\rho} + \rho^3\Delta \end{aligned}$$

En ausencia del impulso, habríamos tenido: $y_t = \frac{\alpha}{1-\rho}$ en todos los periodos, anteriores y posteriores a t^* . Por tanto, la función de respuesta al impulso es en este caso: $FRI = (\Delta, \rho\Delta, \rho^2\Delta, \rho^3\Delta, \dots)$, decreciendo gradualmente hacia cero. El retardo mediano es el número de periodos que deben transcurrir para que se deje sentir sobre y_t el 50% del efecto de largo plazo, que es la suma de todos los efectos, $(1-\rho)^{-1}\Delta$. El retardo medio viene definido por: $\sum_i i f_i / \sum_i f_i$. En el proceso AR(1) el retardo medio es $1/(1-\rho)$; por ejemplo, si $\rho = 0,90$, entonces el período medio es igual a 10. El retardo mediano de un proceso AR(1) es $s = .50/\ln \rho$, donde s incluye el periodo en que se produce el impulso; es decir, una vez calculado s , el retardo mediano se interpretaría como $s - 1$ periodos después de aquél en que se produce la perturbación.

La función de respuesta de un proceso MA(1): $y_t = a + \varepsilon_t - \theta\varepsilon_{t-1}$ a un impulso de tamaño Δ es: $(\Delta, -\theta\Delta, 0, 0, \dots)$. Si consideramos un proceso MA de orden 2: $y_t = a + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$, su función de respuesta a un impulso de tamaño Δ es: $(\Delta, -\theta_1\Delta, -\theta_2\Delta, 0, 0, \dots)$.

In the case of a general random walk, a zero innovation would lead to a random walk growing constantly at a rate defined by the drift a from a given initial condition y_0 . If at time t^* the innovation takes a unit value, the random walk will increase by that amount at time t^* , but also at any future time. So the impulse response is in this case a *step function*, that takes the value 1 at t^* and at any time after that. As we can see, even though the impulse is transitory, lasting for just one period, its effect is permanent.

En ocasiones tiene interés calcular la función de respuesta a un escalón, siendo éste una perturbación constante a partir de un determinado instante t^* . Es sencillo probar que la función de respuesta de un proceso AR(1) a una perturbación de tipo escalón no es sino la acumulacion de la funcion de respuesta al impulso, por lo que en el caso de un proceso AR(1), dicha respuesta resulta: $(\Delta, (1+\rho)\Delta, (1+\rho+\rho^2)\Delta, \dots)$, que es una función creciente (si ρ es positivo) que converge a $(1-\rho)^{-1}\Delta$, que es mayor en valor absoluto que la perturbación inicial, Δ . Como vemos, el efecto de largo plazo de una perturbación de tipo escalón coincide con el efecto total de una perturbación de tipo impulso.

1.3.1 Características de una variable estacionaria: Resumen

Una variable estacionaria tiene generalmente varianza finita (salvo que obedezca a una distribución que como la Cauchy, o la t-Student con 2 grados de libertad, carece de este momento); más precisamente, su varianza no cambia con el paso del tiempo y, desde luego, no tiende a infinito. Una perturbación transitoria sobre una variable estacionaria tiene efectos puramente transitorios; pueden durar varios períodos, pero sus efectos terminan desapareciendo. Los valores sucesivos de su función de autocorrelación convergen rápidamente hacia cero, excepto quizá en los retardos de carácter estacional. la serie temporal correspondiente a una variable estacionaria no deambula durante períodos largos de tiempo a un mismo lado de su media muestral, sino que cruza frecuentemente dicho nivel medio. El número medio de períodos que transcurre entre dos cruces consecutivos del nivel medio muestral es pequeño.

Por el contrario, una perturbación de carácter transitorio sobre una variable no estacionaria tiene efectos permanentes. La función de autocorrelación de una variable no estacionaria converge a cero muy lentamente, y su serie temporal muestra claramente largos períodos de tiempo en que deambula sin cruzar su nivel medio.

Characteristics of a stationary process:

- It has finite variance
- Its simple and partial autocorrelation functions converge to zero quickly
- The time series realization crosses its sample mean level often
- A one-period shock has purely transitory effects
- Its forecast converges to its mathematical expectations as the forecast horizon goes to infinity

Characteristics of a non-stationary stochastic process

- Its variance increases with the sample size
- Its autocorrelation functions do not go to zero quickly
- The number of periods between successive crosses with its sample mean is infinity
- A one-period shock has permanent effects
- Its forecast does not converge to its mathematical expectation as the forecast horizon goes to infinity

1.3.2 Numerical exercise: Simulating simple stochastic processes

The *Simple_simulation.xls* EXCEL book presents simulations of some of these simple stochastic processes. Column A in the *Simulations* spreadsheet contains a time index. Column B contains a sample realization of random numbers extracted from a $N(0,1)$ distribution. This has been obtained from EXCEL using the sequence of keys: *Tools/Data Analysis/Random Number Generator* and selecting as options in the menu *number of variables =1, observations = 200*, a *Normal* distribution with expectation 0 and variance 1, and selecting the appropriate output range in the spreadsheet.

A well constructed random number generator produces independent realizations of the chosen distribution. We should therefore have in column B 200 independent data points from a $N(0,1)$, which can either be interpreted as a sample of size 200 from a $N(0,1)$ population, or as a single time series realization from a white noise where the innovation follows a $N(0,1)$ probability distribution. The latter is the interpretation we will follow. At the end of the column, we compute the sample mean and standard deviation, with values of 0.07 and 1.04, respectively, si bien estos valores numéricos cambiarán con cada realización de la serie temporal. These are estimates of the 0 mathematical expectation and unit standard deviation with this sample. Below that, we present the standard deviation of the first and the last 100 observations, of 1.09 and .98. Estimates of the variance obtained with the full sample or with the two sub-samples seem reasonable. A different sample would lead to different numerical estimates.

Since we compare in this spreadsheet the characteristic of processes that differ in the numerical values of their parameters, we use in simulating a time series realization from each process the same time series for the innovation, the one we have in column B.

Panel 2 contains sample realizations from three different random walks without drift. The only parameter in such processes is the variance of the innovation, which takes values 1, 25 and 100, respectively. At a difference of a white noise, an initial condition is needed to generate a time series for a random walk, because of the time dependence between successive observations, as can be seen in (1). We take $y_1 = 100$ as initial condition in all cases. The three sample realizations are graphed in the *RandomWalks* spreadsheet. All exhibit extreme persistence, crossing the sample mean just once in 200 observations. We know by construction that these three processes lack a well defined mean and have a time increasing variance. We can always compute sample averages and standard deviations, as shown in the spreadsheet at the end of the series, but it is not advisable to try to interpret such statistics. In particular, in this case, by drawing different realization for the white noise in column B, the reader can easily check how sample mean and standard deviations may drastically change. In fact, standard deviations are calculated in the spreadsheet for the first and last 100 sample observations, and they can turn out to be very different, and different from the theoretical result: $Var(y_t) = t\sigma_\varepsilon^2$. The point is that we cannot estimate that time-varying moment with much precision.

Panel 3 compares a random walk to three first-order autoregressive processes, with autoregressive coefficients of 0.99, 0.95 and 0.30. As mentioned above, a random walk can be seen as the limit of a first order autoregression, as the autoregressive coefficient converges to 1, although the limit presents some discontinuity since, theoretically, autoregressive processes are stationary so long as the autoregressive coefficient is below 1 in absolute value, while the random walk is nonstationary. The autoregressive processes will all have a well-defined mean and variance, and we can see how the variance increases with the value of the autoregressive coefficient, as expected. The sample time series realizations for the four processes are displayed in the *AR-processes* spreadsheet, where it can be seen that sample differences between the autoregressive process with the 0.99 coefficient and the random walk are minor, in spite of the theoretical differences between the two processes. In particular, the autoregressive process crosses its sample mean in very few occasions. That is also the case for the 0.95-autoregressive process, although its mean reverting behavior is very clear at the end of the sample. On the other hand, the time series realization from the 0.30-autoregressive process exhibits the typical behavior in a clearly stationary process, crossing its sample mean repeatedly.

Panel 4 presents sample realizations from two white noise processes with drift and $N(0,1)$ innovations. As shown in the enclosed graph, both fluctuate around their mathematical expectation, which is the value of the constant defining the drift, crossing their sample means very often. Panel 5 contains time series realizations for two random walk processes with drift. These show in the graph in the form of what could look as deterministic trends. This is because the value of the drifts, of 1.0 and 3.0, respectively, is large, relative to the innovation variance which is of 25 in both cases. If the value of the drift is reduced, or the variance of the innovation increased, the shape of the time series would be different, since the fluctuations would then dominate over the accumulated effect of the drift, as the reader can check by reducing the numerical values of the drift parameters used in the computation of these two columns. A similar result arises by significantly increasing the innovation variance. What are the differences between both cases in terms of the values taken by the process?

Panel 6 presents realizations of a stationary first order autoregression with coefficient of .90. In the second case we have not included an innovation process, so that it can be considered as a deterministic autoregression. It is interesting to see in the enclosed graph the behavior of a stationary process: starting from an initial condition, in the absence of an innovation, the process will always converge smoothly to its mathematical expectation. That is not so obvious in the stochastic autoregression, just because the innovation variance, of 25, is large relative to the distance between the initial condition, 150, and the mathematical expectation, 100. The reader can check how reducing the standard deviation used in column S from 5 to 0.5, the pattern of the time series changes drastically, and the convergence process then becomes evident. The mathematical expectation of these two autoregressive processes is $E(y_t) = 100$. The reader can change the initial condition to a value below $E(y_t)$, like $y_1 = 80$, and see how the convergence to the mean is obtained from below.

Panel 7 contains realizations for second order autoregressions. The first two columns present sample realizations from stationary autoregressions,

$$\text{Model 1: } y_t = 10 + .6y_{t-1} + .3y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, 1) \quad (2)$$

$$\text{Model 2: } y_t = 30 + 1.2y_{t-1} - .5y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, 1) \quad (3)$$

and are represented in an enclosed graph. The two time series display fluctuations around their sample mean of 100, which they cross a number of times. The second time series, represented in red in the graph can be seen to exhibit a more evident stationary behavior, with more frequent crosses with the mean. The next three columns present realizations for nonstationary second order autoregressions. There is an important difference between them: the first two correspond to processes:

$$\text{Model 3 : } y_t = .7y_{t-1} + .3y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, 1) \quad (4)$$

$$\text{Model 4 : } y_t = 1.5y_{t-1} - .5y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, 1) \quad (5)$$

that contain exactly a unit root, the second one being stable.⁶ The roots of the characteristic equation for Model 3 are 1 and -0.3, while those for Model 2 are 1 and 0.5. The last autoregression

$$\text{Model 5 : } y_t = .3y_{t-1} + 1.2y_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, 1) \quad (6)$$

has a root greater than one, which produces an explosive behavior. The two roots are -0.95 and 1.25.

The *Impulse responses* spreadsheet contains the responses to a unit shock for the stochastic processes considered above: a random walk, three first-order autoregressions, two stationary second-order autoregressions, and three nonstationary second-order autoregressions. To compute responses to an impulse innovation at $t = 11$, the innovation in each process is supposed to take a zero value for ten periods, to be equal to 1, the standard deviation assumed for the innovation in all cases at $t^* = 11$, and be again equal to zero afterwards. We compare that to the case when the innovation is zero at all time periods. Impulse responses are computed as the difference between the time paths followed by each process under the scenario with a shock at $t^* = 11$, and in the absence of that shock. The first-order autoregressions are supposed to start from an initial condition $y_1 = 100$, when their mathematical expectations is zero, so in the absence of any shock, they follow a smooth trajectory gradually converging to zero at a speed determined by its autoregressive coefficient. The second order autoregressions are assumed to start from $y_1 = y_2 = 100$, which is also

⁶The two polynomials can be written as $1 - a_1B - a_2B^2 = (1 - B)(1 - \lambda B)$, the second root being $1/\lambda$. The reader just need to find the value of λ in each case.

their mathematical expectations. So, in the absence of any shock, the processes would stay at that value forever.⁷

The first graph to the right displays impulse responses for a random walk as well as for the three first order autoregressions considered above, with coefficients 0.99, 0.95 and 0.30. A random walk has the constant, permanent impulse response that we mentioned above when describing this process. The responses of the first order autoregressions can be seen to gradually decrease to zero from the initial unit value. The response is shorter the lower it is the autoregressive coefficient. For high autoregressive coefficients, the process shows strong persistence, which makes the effects of the shock to last longer.

The second graph shows the impulse responses of the two stationary second-order autoregressions. As the reader can easily check, the characteristic equation for Model 1 has roots -0.32 and 0.92, so it is relatively close to nonstationarity. The characteristic equation for Model 2 has roots $0.6 \pm 0.37417i$, with modulus 0.5. This difference shows up in a much more persistent response of Model 1. The complex roots of Model 2 explain the oscillatory behavior of the impulse response of this model.

The third graph displays impulse responses for the three nonstationary second order autoregressions. In the two cases when there is a unit root (Models 3 and 4), the graph shows a permanent response to the purely transitory, one-period shock. The response of Model 5 is explosive because of having one root above 1, and its values are shown on the right Y-axis.

1.4 Ejercicios con datos financieros

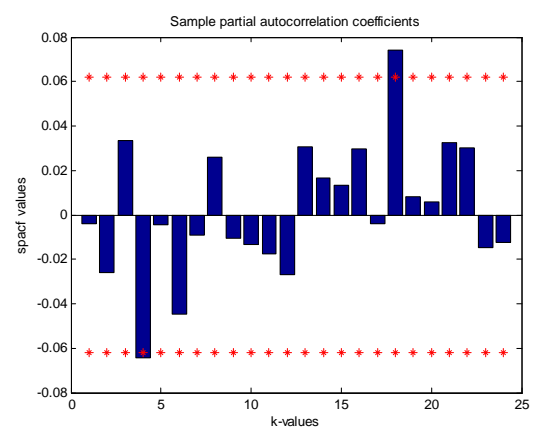
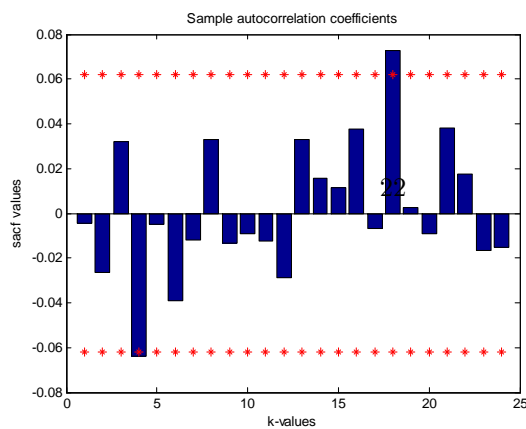
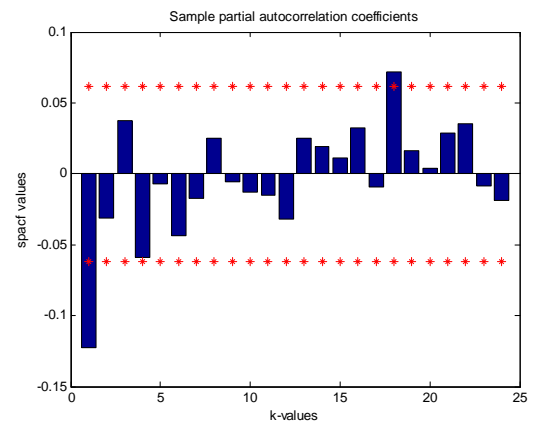
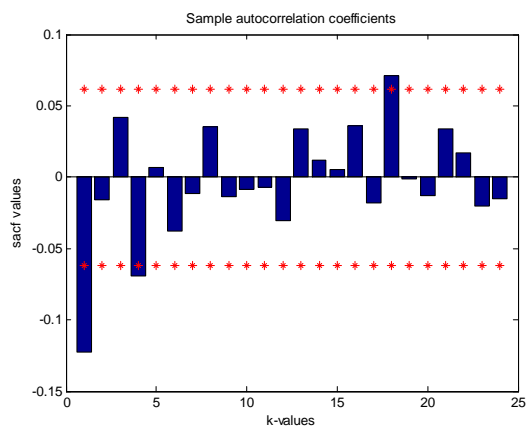
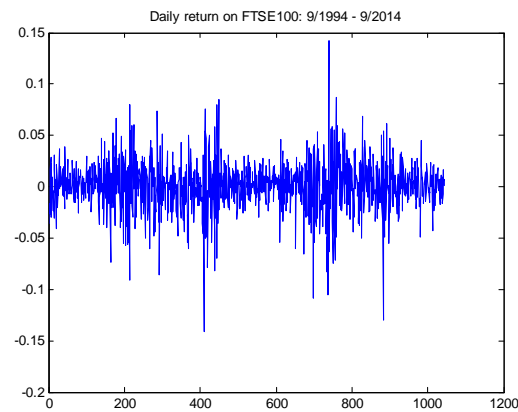
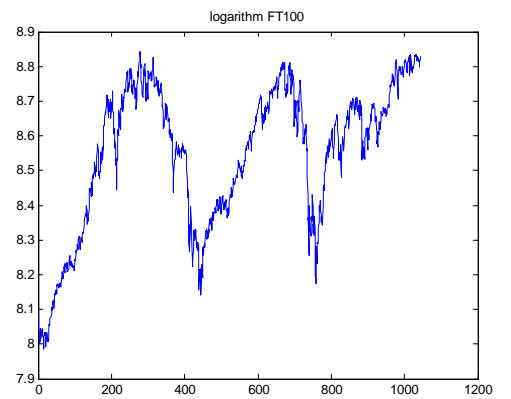
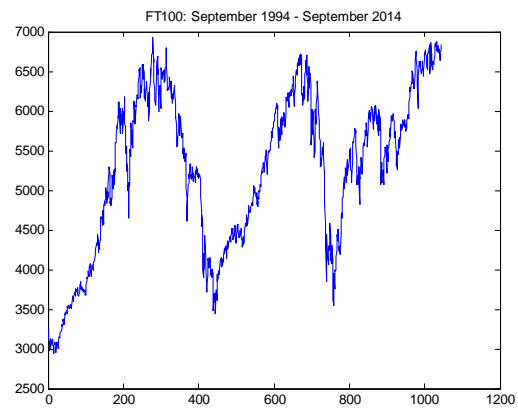
The first row of the figure displays closing weekly data on the FTSE100 index from 5/9/1994 to 8/9/2014. The two market crisis of 2001 (crisis of dot-com firms) and september 2008 (fall of Lehman brothers) can be clearly seen in the graph. The series is clearly nonstationary, as it is it also the case with its logarithmic transformation. Daily returns are stationary in the mean, and their simple and partial autocorrelation functions suggest initially an AR(1) structure. We estimate it:

$$r_t = \underset{(1.05)}{0.0806} \cdot 10^{-2} - \underset{(4.00)}{0.1232} r_{t-1} + u_t$$

with $\sigma_u^2 = 6.20 \cdot 10^{-4}$, $\sigma_u = 0.0249$, and t-Student statistics in parenthesis.

The simple and partial autocorrelation functions for the residuals do not show indication of any further structure.

⁷We could have done otherwise, like starting the first-order autoregressions at their mathematical expectation, and the second-order autoregressions outside their expected values. The reader can experiment with these changes.



1.5 Predicción con modelos ARMA

1.5.1 Forecasting with some simple processes

Forecasting an AR(1) process: A future value of the first order autoregression can be represented:

$$y_{t+s} = a + \rho y_{t+s-1} + \varepsilon_{t+s}, \quad |\rho| < 1, \quad s \geq 1,$$

which can be iterated to,

$$y_{t+s} = a(1 + \rho + \rho^2 + \dots + \rho^{s-1}) + \rho^s y_t + (\rho^{s-1} \varepsilon_{t+1} + \rho^{s-2} \varepsilon_{t+2} + \dots + \varepsilon_{t+s})$$

so that, taking expectations based on information up to time t , its forecast is given by,

$$E_t y_{t+s} = a \frac{1 - \rho^s}{1 - \rho} + \rho^s y_t = \rho^s y_t + (1 - \rho^s) E_t y$$

Hence, as the forecast horizon goes to infinity, the forecast converges to,

$$\lim E_t y_{t+s} = \frac{a}{1 - \rho}$$

the mean of the process.

Forecasting an MA(1) process:

Future values of an MA(1) process respond to:

$$\begin{aligned} y_{t+1} &= a + \varepsilon_{t+1} - \theta \varepsilon_t \Rightarrow E_t y_{t+1} = a - \theta \varepsilon_t \\ y_{t+2} &= a + \varepsilon_{t+2} - \theta \varepsilon_{t+1} \Rightarrow E_t y_{t+2} = a \Rightarrow E_t y_{t+s} = a, \forall s > 1 \end{aligned}$$

1.5.2 Expresiones analíticas para la predicción

En un proceso $AR(2)$:

$$\begin{aligned} E_T y_{T+1} &= E_T (\delta + \phi_1 y_T + \phi_2 y_{T-1} + \varepsilon_{T+1}) = \delta + \phi_1 y_T + \phi_2 y_{T-1} \\ E_T y_{T+2} &= E_T (\delta + \phi_1 y_{T+1} + \phi_2 y_T + \varepsilon_{T+2}) = \delta + \phi_1 (\delta + \phi_1 y_T + \phi_2 y_{T-1}) + \phi_2 y_T = \\ &= \delta (1 + \phi_1) + (\phi_1^2 + \phi_2) y_T + \phi_1 \phi_2 y_{T-1} \\ E_T y_{T+3} &= \delta (1 + \phi_1 + \phi_1^2) + \phi_1 (\phi_1^2 + 2\phi_2) y_T + (\phi_1^2 \phi_2 + \phi_2^2) y_{T-1} \end{aligned}$$

Para procesos autoregresivos de orden superior a 1 no existe una forma analítica sencilla para representar las predicciones a distintos horizontes. Como vemos, en un proceso autoregresivo de orden p , la información muestral que es relevante para el cálculo de predicciones se resume en las p últimas observaciones de la serie temporal.

Un proceso MA(1) sin constante, invertible, puede escribirse:

$$y_{T+1} = -\theta y_T - \theta^2 y_{T-1} - \theta^3 y_{T-2} - \dots + \varepsilon_{T+1}$$

por lo que:

$$\begin{aligned} E_T y_{T+1} &= -\theta y_T - \theta^2 y_{T-1} - \theta^3 y_{T-2} - \dots \\ E_T y_{T+2} &= -\theta E_T y_{T+1} - \theta^2 y_T - \theta^3 y_{T-1} - \theta^4 y_{T-2} - \dots = \\ &= -\theta (-\theta y_T - \theta^2 y_{T-1} - \theta^3 y_{T-2} - \dots) - \theta^2 y_T - \theta^3 y_{T-1} - \theta^4 y_{T-2} - \dots = 0 \\ E_T y_{T+k} &= 0, \forall k \geq 2 \end{aligned}$$

lo cual es evidente, recordando que la estructura del proceso MA(1) es:

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

por lo que:

$$\begin{aligned} E_T y_{T+1} &= E_T \varepsilon_{T+1} - \theta E_T \varepsilon_T \\ E_T y_{T+2} &= E_T \varepsilon_{T+2} - \theta E_T \varepsilon_{T+1} = 0 \\ E_T y_{T+k} &= 0, \forall k \geq 2 \end{aligned}$$

1.5.3 Varianza del error de predicción

Sección 13.9 en Econometría - Novales

1.5.4 Recuperación de predicciones

1.5.5 Evaluación de previsiones

Para evaluar la calidad de un conjunto de s previsiones obtenidas a partir de un modelo, es habitual utilizar:

- Raíz del Error Cuadrático Medio:

$$RMSE = \sqrt{\frac{1}{s} \sum_{i=1}^s (\hat{y}_{t+i} - y_{t+i})^2}$$

- Error Absoluto Medio:

$$MAE = \sum_{i=1}^s |\hat{y}_{t+i} - y_{t+i}|$$

Cuando se evalúa la calidad de un conjunto de predicciones acerca de variables que oscilan en torno a una media no nula, o variables no estacionarias, suelen utilizarse sus versiones porcentuales:

- Raíz del Error Cuadrático Medio Porcentual:

$$RMSE = \sqrt{\frac{1}{s} \sum_{i=1}^s \left(\frac{\hat{y}_{t+i} - y_{t+i}}{y_{t+i}} \right)^2}$$

- Error Porcentual Absoluto Medio:

$$MAE = \sum_{i=1}^s \left| \frac{\hat{y}_{t+i} - y_{t+i}}{y_{t+i}} \right|$$

que no conviene utilizar cuando se trabaja con rentabilidades, por ejemplo.

1.6 Identificación de procesos AR y MA

Secciones 13.3 y 13.4 en Econometría – Novales

En un modelo con m parámetros que han sido estimados con n datos, el criterio de información de Akaike es:

$$AIC = -\frac{2 \ln L}{n} + \frac{2m}{n}$$

siendo L^* el valor maximizado de la función de verosimilitud, mientras que el criterio de información de Schwarz, de bases bayesianas, es:

$$BIC = -\frac{2 \ln L}{n} + \frac{m \ln(n)}{n}$$

1.7 Contrastes de ausencia de autocorrelación (ruido blanco)

1.7.1 Coeficientes de autocorrelación

La manera más directa de contrastar si un determinado proceso estocástico tiene naturaleza de camino aleatorio, o lo que es lo mismo, satisface la hipótesis de martingala, es analizando los coeficientes de correlación entre dos variables constituyentes de dicho proceso estocástico en dos instantes distintos de tiempo. Bajo la hipótesis nula, todos los coeficientes de autocorrelación deberían ser nulos, lo que complica en cierta medida el diseño del contraste, pues el número de hipótesis a contrastar es potencialmente ilimitado. Pero, una vez más, para llevar a cabo el contraste nos habremos de servir de sus análogos muestrales, en cuyo cálculo perdemos observaciones muestrales, por el hecho de tener que retardar la serie temporal de datos.

Fuller (1976) caracteriza la distribución asintótica del vector de los m primeros coeficientes de autocorrelación, siendo ésta Normal multivariante:

$$\sqrt{T} \hat{\rho}_k \rightarrow N(0, 1) \quad (7)$$

En muestras finitas, si el proceso estocástico tiene estructura de camino aleatorio (RW1-*Taylor antiguo*), con varianza finita σ^2 y momento de orden seis proporcional a σ^6 , se tiene:

$$\begin{aligned} E(\hat{\rho}_k) &= -\frac{T-k}{T(T-1)} + O(T^{-2}) \\ \text{Var}(\hat{\rho}_k) &= \frac{T-k}{T^2} + O(T^{-2}) \\ \text{Cov}(\hat{\rho}_k, \hat{\rho}_l) &= O(T^{-2}) \end{aligned}$$

Por tanto, los coeficientes de autocorrelación muestrales de un camino aleatorio están sesgados a la baja.⁸

Para evitar el sesgo en el estimador de los coeficientes de autocorrelación, Fuller (1976) propone la corrección:

$$\tilde{\rho}_k = \hat{\rho}_k + \frac{T-k}{(T-1)^2} (1 - \hat{\rho}_k^2),$$

con:

$$\begin{aligned} \frac{T}{\sqrt{T-k}} \tilde{\rho}_k &\rightarrow N(0, 1) \\ E(\tilde{\rho}_k) &= O(T^{-2}) \end{aligned}$$

A menudo, nos interesa pronunciarnos sobre la hipótesis nula: $H_0 : \rho_k = 0$ para $k = q+1, q+2, \dots$. Si los datos han sido generados por un proceso MA(q), entonces la varianza de $\hat{\rho}_k$ puede aproximarse por:

$$\text{Var}(\hat{\rho}_k) = \frac{1}{T} \left(1 + 2 \sum_{i=1}^{k-1} \rho_i^2 \right), \quad k > 1$$

por lo que, bajo H_0 , tenemos: $\text{Var}(\hat{\rho}_k) = 1/T$, el mismo resultado que surge de la distribución asintótica que antes vimos.

1.7.2 Contrastes Portmanteau

Para recoger adecuadamente un concepto de camino aleatorio que implica que todos los coeficientes de autocorrelación son cero, Box y Pierce (1970) propusieron un contraste conjunto basado en el estadístico:

⁸El sesgo negativo en la estimación $\hat{\rho}_k$ proviene de que su cálculo se basa en los productos de desviaciones respecto de la media muestral en distintos períodos de tiempo, t y $t+k$. Pero la suma de tales desviaciones es cero, luego desviaciones positivas tenderán a venir seguidas en el futuro por desviaciones negativas, y viceversa, por lo que su producto tenderá a ser negativo. En muestras pequeñas, esto es importante: con $T = 10$, $E(\hat{\rho}_k) = -\frac{T-k}{T(T-1)} = -\frac{10-k}{90} \approx -0,10$.

$$Q_m \equiv T \sum_{k=1}^m \rho_k^2$$

Bajo la hipótesis nula $H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0$, (RW1), (7) implica que: $\hat{Q}_m \equiv T \sum_{k=1}^m \hat{\rho}_k^2$ se distribuye como una χ_m^2 .

Ljung y Box (1978) propusieron una corrección para muestras finitas:

$$Q'_m \equiv T(T+2) \sum_{k=1}^m \frac{\rho_k^2}{T-k}$$

Al agregar los coeficientes de autocorrelación al cuadrado, el contraste tiene potencia frente a diversas alternativas. Sin embargo, la elección del orden m es ambigua, y puede condicionar los resultados del contraste. Si se utiliza un m pequeño, puede no detectarse la autocorrelación de orden superior. Si se utiliza m grande, el contraste pierde potencia debido a la acumulación de autocorrelaciones no significativas. Si se dispone de una alternativa concreta, puede diseñarse un contraste con mejores propiedades estadísticas.

1.7.3 Ratios de varianza

Supongamos que la serie temporal que consideramos es la rentabilidad de un determinado activo financiero. Recordando que las rentabilidades continuas son aditivas, es decir, que la rentabilidad sobre 2 períodos es la suma de las rentabilidades sobre cada uno de ellos: $r_t^2 = r_t^1 + r_{t-1}^1$, tenemos la razón de varianzas a 2 períodos,

$$RV(2) = \frac{Var(r_t^2)}{2Var(r_t^1)} = \frac{Var(r_t^1 + r_{t-1}^1)}{2Var(r_t^1)} = \frac{2Var(r_t^1) + 2Cov(r_t^1, r_{t-1}^1)}{2Var(r_t^1)} = 1 + \rho_1$$

que está determinada por el primer valor de la función de autocorrelación simple de r_t .

Si las rentabilidades son ruido blanco (*white noise*), el coeficiente de autocorrelación de orden 1 es igual a cero, y la razón de varianzas es igual a 1. Con autocorrelación positiva, la razón de varianzas será mayor que uno, siendo inferior a la unidad si las rentabilidades están negativamente autocorrelacionadas, lo que no es muy frecuente en datos financieros.

Para contrastar la significatividad de este estadístico, puede utilizarse la distribución asintótica:

$$\sqrt{2T}(RV(2) - 1) \sim N(0, 2)$$

como más abajo veremos, lo que sugiere que,

$$RV(2) \sim N\left(1, \frac{1}{T}\right)$$

por lo que, manteniendo un 95% de confianza, la *razón de varianzas de orden 2* no debería separarse de 1,0 en más del doble de $\sqrt{1/T}$.

Existe un contraste más amplio, que incorpora los coeficientes de autocorrelación hasta orden q . La razón de varianzas es entonces:

$$RV(q) = \frac{Var(r_t^q)}{qVar(r_t^1)} = 1 + 2 \sum_{i=1}^{q-1} \left(1 - \frac{i}{q}\right) \rho_i$$

que muestra que el ratio de varianzas $RV(q)$ es una combinación lineal de los $q - 1$ primeros coeficientes de autocorrelación, tomados con ponderaciones decrecientes. En el caso $q = 2$ tenemos la expresión que antes vimos para $RV(2)$. Nuevamente, si el proceso es ruido blanco, el ratio de varianzas $RV(q)$ es igual a 1 para todo q .

Si, por ejemplo, se trata de un proceso autoregresivo de primer orden,

$$r_t = \phi r_{t-1} + \varepsilon_t$$

se tiene:

$$RV(q) = 1 + 2 \sum_{i=1}^{q-1} \left(1 - \frac{i}{q}\right) \rho_i = 1 + \frac{2}{1 - \phi} \left[\phi - \frac{\phi^q}{q} - \frac{\phi - \phi^q}{q(1 - \phi)} \right]$$

una expresión que puede utilizarse para diseñar un contraste de camino aleatorio teniendo una estructura AR(1) como hipótesis alternativa.

1.8 Contrastes de camino aleatorio

Las estructuras de camino aleatorio han sido objeto de análisis en Finanzas. Si el precio de un activo sigue una estructura de camino aleatorio, tendremos:

$$P_t = P_{t-1} + \varepsilon_t$$

con ε_t ruido blanco, es decir: $E_t \varepsilon_{t+1} = 0$, y $E_t \varepsilon_{t+s} = 0, \forall s \geq 1$, por lo que:

$$E_t P_{t+s} = P_t$$

afirmandose en tal caso que el precio actual del activo resume toda la información existente que pueda anticipar el precio futuro. En efecto, como muestra la expresión anterior, si P_t es una camino aleatorio, entonces su predicción a cualquier horizonte futuro es igual al precio actual, el último precio observado. Nótese que no importa cuanta información pasada tengamos, no juega ningún papel en la formación de la predicción del precio futuro. Dicho de otro modo, el precio actual resume entonces toda la información disponible (es un estadístico suficiente, en la nomenclatura estadística, respecto de la predicción del precio futuro). Este comportamiento caracteriza asimismo el concepto de un *mercado eficiente*.

Existen en la literatura distintas definiciones de camino aleatorio, no todas equivalentes entre sí. Una definición requiere que las rentabilidades sean independientes e idénticamente distribuidas. Esta definición, que puede resultar interesante en determinados contextos, no lo es tanto cuando tratamos la posible predictibilidad de la serie de rentabilidades. La razón es que la mayoría de los contrastes de la misma se basarán en la hipótesis de Normalidad y examinarán la ausencia de autocorrelación en la rentabilidad. Pero una estructura de heterocedasticidad condicional (la varianza depende en cada período de las realizaciones recientes de la serie de rentabilidades) introduce dependencia temporal. Si contrastamos únicamente la ausencia de autocorrelación en una rentabilidad, podemos no rechazar la hipótesis nula, porque las rentabilidades carezcan de autocorrelación, a pesar de que hay dependencia temporal entre ellas.

Una definición más general [Granger y Mortensen (1970)] se basa en las condiciones: a) esperanza matemática constante y b) ausencia de correlación serial. En este caso, la predicción lineal óptima de una rentabilidad futura es su esperanza incondicional, que estamos suponiendo constante.

Si pretendemos contrastar la hipótesis de que la serie de rentabilidades sigue una estructura de camino aleatorio, tenemos que introducir condiciones adicionales [ver Lo y MacKinlay (1988)].

Una tercera definición [Samuelson (1965)] es: $E[r_{t+1}/I_t] = \mu$ para cierta constante μ y para todo instante t y toda historia pasada: $I_t = \{r_{t-i}, i \geq 0\}$.

La tercera definición implica la segunda, siempre que las rentabilidades tengan varianza finita. La diferencia entre ambas es menor. Los contrastes de camino aleatorio que utilizan funciones de autocorrelación se basan en la segunda definición. Suponiendo varianza finita, si un test de este tipo rechaza la segunda definición, rechaza también la tercera definición.

El modelo: $P_t = P_{t-1} + \varepsilon_t$ incorpora dos hipótesis: i) $H0 : \beta = 1$ frente a $H1 : \beta \neq 1$ en el modelo: $P_t = \beta P_{t-1} + \varepsilon_t$, y si $H0$ no se rechaza: ii) la diferencia $P_t - P_{t-1}$ carece de autocorrelación. Por tanto, un enfoque para este contraste se basa en aplicar un test de raíz unitaria, que describiremos más adelante, a la representación autoregresiva de P_t , que puede ser de orden 1 o de orden superior. Si no se rechaza la hipótesis de existencia de una raíz unitaria en dicha representación autoregresiva, aplicaríamos un contraste de ausencia de autocorrelación, como los vistos en esta sección, a la diferencia $P_t - P_{t-1}$.

1.8.1 Diferencias de varianzas

A partir de una serie de precios $P_t, t = 0, 1, \dots, 2T$, de longitud $2T+1$, denotemos por p_t a la serie de logaritmos, $p_t = \ln(P_t), t = 1, 2, \dots, 2T$. Si suponemos que los datos han sido generados por el modelo:

$$p_t - p_{t-1} = \mu + \varepsilon_t, \varepsilon_t \sim i.i.d., N(0, \sigma^2),$$

consideremos los estimadores habituales de μ y σ^2 :

$$\begin{aligned}\hat{\mu} &= \bar{r} = \frac{1}{2T} \sum_{i=1}^{2T} (p_i - p_{i-1}) = \frac{1}{2T} (p_{2T} - p_0) \\ \hat{\sigma}_a^2 &= \frac{1}{2T} \sum_{i=1}^{2T} (p_i - p_{i-1} - \hat{\mu})^2\end{aligned}$$

junto con un segundo estimador de la varianza:

$$\hat{\sigma}_b^2 = \frac{1}{2T} \sum_{i=1}^T (p_{2i} - p_{2i-2} - 2\hat{\mu})^2$$

que se basa en que bajo la hipótesis nula, p_t sigue un camino aleatorio. Nótese que el segundo estimador de la varianza utiliza unicamente los incrementos correspondientes a las observaciones de orden par. Por tanto, la suma que en él aparece tiene la mitad de términos que en el primer estimador. El estimador toma la mitad de la varianza de los incrementos mencionados. Esto se debe a que bajo la hipótesis nula de camino aleatorio, la esperanza matemática y la varianza crecen linealmente con la amplitud del intervalo temporal para el que se calculan. Ambos estimadores son consistentes. Los estimadores de la varianza tienen distribución asintótica:

$$\begin{aligned}\sqrt{2T} (\hat{\sigma}_a - \sigma^2) &\sim N(0, 2\sigma^4) \\ \sqrt{2T} (\hat{\sigma}_b - \sigma^2) &\sim N(0, 4\sigma^4)\end{aligned}$$

Como $\hat{\sigma}_a^2$ es un estimador asintóticamente eficiente bajo la hipótesis nula (camino aleatorio, *RW1*), podemos utilizar el clásico argumento de Hausman, para concluir que la varianza asintótica de la diferencia de un estimador consistente y un estimador asintóticamente eficiente es igual a la diferencia de las varianzas asintóticas de ambos estimadores.

Por tanto, si denotamos $D\hat{V}(2) = \hat{\sigma}_b^2 - \hat{\sigma}_a^2$, tenemos el estadístico de *Diferencia de varianzas* de orden 2:

$$\sqrt{2T} D\hat{V}(2) \sim N(0, 2\sigma^4) \Rightarrow \sqrt{T} D\hat{V}(2) \sim N(0, \sigma^4)$$

por lo que la hipótesis nula de camino aleatorio puede contrastarse utilizando cualquier estimador consistente de σ^4 como, por ejemplo: $\hat{\sigma}^4 = (\hat{\sigma}_a^2)^2$. Entonces, el estadístico estandarizado, $\sqrt{T} D\hat{V}(2) / \sqrt{\hat{\sigma}^4} \equiv \sqrt{T} D\hat{V}(2) / \hat{\sigma}_a^2$ sigue una distribución $N(0,1)$ bajo la hipótesis nula. Lo y MacKinley (1988) sugieren utilizar un contraste de camino aleatorio basado en esta distribución. La hipótesis nula del contraste es que la primera diferencia del logaritmo carece de autocorrelación. Por tanto, se aplicará generalmente a los precios de activos financieros, siendo un contraste de ausencia de correlación en rentabilidades.

Por la aditividad de las rentabilidades logarítmicas, el estadístico de razón de varianzas de orden 2 puede obtenerse mediante $R\hat{V}(2) = \hat{\sigma}_b^2 / \hat{\sigma}_a^2$, por lo que:

$$R\hat{V}(2) - 1 = \frac{\hat{\sigma}_b^2 - \hat{\sigma}_a^2}{\hat{\sigma}_a^2} = \frac{D\hat{V}(2)}{\sqrt{(\hat{\sigma}_a^2)^2}}$$

y, utilizando una aproximación de Taylor de primer orden o mediante el llamado *método delta*, podemos deducir para el Ratio de Varianza la misma distribución asintótica que antes adelantamos:⁹

$$\sqrt{2T}(R\hat{V}(2) - 1) \sim N(0, 2) \Rightarrow \sqrt{T}(R\hat{V}(2) - 1) \sim N(0, 1)$$

Aunque suele preferirse el estadístico *Ratio de varianzas* al de *Diferencia de varianzas*, por estar el primero libre de escala, ambos conducen a las mismas conclusiones.

La potencia de este tipo de contrastes aumenta si se reduce la posible presencia de heterocedasticidad en los datos.

Las definiciones y estadísticos pueden extenderse a intervalos de más de 2 períodos, con muestras de tamaño qT :

$$\begin{aligned}\hat{\mu} &= \bar{r} = \frac{1}{qT} \sum_{t=1}^{qT} (p_t - p_{t-1}) = \frac{1}{qT} (p_{qT} - p_0) \\ \hat{\sigma}_a^2 &= \frac{1}{qT} \sum_{t=1}^{qT} (p_t - p_{t-1} - \hat{\mu})^2 \\ \hat{\sigma}_b^2(q) &= \frac{1}{qT} \sum_{t=1}^T (p_{qt} - p_{qt-q} - q\hat{\mu})^2\end{aligned}$$

con las distribuciones:

$$\sqrt{qT}D\hat{V}(q) \sim N(0, 2(q-1)\sigma^4)$$

$$\sqrt{qT}(R\hat{V}(q) - 1) \sim N(0, 2(q-1))$$

siendo q el número de períodos. Dos refinamientos mejoran las propiedades de muestras finitas de estos contrastes. Uno consiste en utilizar periodos solapados de amplitud q :

$$\hat{\sigma}_c^2(q) = \frac{1}{q^2T} \sum_{t=q}^{Tq} (p_t - p_{t-q} - q\hat{\mu})^2$$

⁹Método delta: Si $\sqrt{T}(\hat{\theta} - \theta_0) \sim N(0, \Sigma)$, entonces, si $f(\theta)$ es una función no lineal de θ , tenemos: $\sqrt{T}(f(\hat{\theta}) - f(\theta_0)) \sim N\left(0, \frac{\partial f}{\partial \theta} \Sigma \frac{\partial f}{\partial \theta}'\right)$. En este caso, $\theta = \hat{\sigma}_b^2 - \hat{\sigma}_a^2$, y $f(\theta) = \frac{\theta}{\hat{\sigma}_a^2}$. Por tanto, $\frac{\partial f}{\partial \theta} = 1/\hat{\sigma}_a^2$, y $\frac{\partial f}{\partial \theta} \Sigma \frac{\partial f}{\partial \theta}' = \frac{1}{\hat{\sigma}_a^2} 2\sigma^4 \frac{1}{\hat{\sigma}_a^2}$.

y el segundo en corregir un sesgo en los estimadores $\hat{\sigma}_a^2$ y $\hat{\sigma}_c^2$ antes de dividir uno por otro. Los estimadores insesgados son:

$$\begin{aligned}\bar{\sigma}_a^2 &= \frac{1}{qT-1} \sum_{t=1}^{qT} (p_t - p_{t-1} - \hat{\mu})^2 \\ \bar{\sigma}_c^2(q) &= \frac{1}{q^2T} \sum_{t=q}^{Tq} (p_t - p_{t-q} - q\hat{\mu})^2\end{aligned}$$

con $m = q(nq - q + 1) \left(1 - \frac{q}{nq}\right)$ para luego calcular con ellos los estadísticos:

$$\begin{aligned}DV(q) &= \bar{\sigma}_c^2(q) - \bar{\sigma}_a^2 \\ VR(q) &= \frac{\bar{\sigma}_c^2(q)}{\bar{\sigma}_a^2}\end{aligned}$$

1.9 Variables no estacionarias. Modelos ARIMA

Secciones 13.6 y 13.7 en Econometría – Novales

La no estacionariedad de las variables involucradas en una regresión es uno de las situaciones que requiere una consideración más cuidadosa en el análisis de regresión. La ausencia de estacionariedad se produce con mucha frecuencia en variables económicas; además, como vamos a ver, sus implicaciones en la estimación de modelos de regresión pueden ser bastante negativas e importantes. Por último, su detección y tratamiento no son siempre evidentes.

1.9.1 Tendencias deterministas y tendencias estocásticas

La ausencia de estacionariedad en variables económicas puede reflejarse mediante la presencia de tendencias estocásticas o de *tendencias deterministas* en los precios de mercado, a través de volatilidad cambiante en el tiempo, etc.. Una *tendencia estocástica* es un componente estocástico cuya varianza tiende a infinito con el paso del tiempo. Una *tendencia determinista* es una función exacta del tiempo, generalmente lineal o cuadrática, lo que hace que el valor de la variable crezca o disminuya constantemente; si la tendencia es lineal, la variable tenderá a más o menos infinito; si la tendencia es cuadrática o de orden superior, la variable puede estar acotada.

Si una variable presenta una *tendencia determinista lineal*, su valor esperado tenderá a aumentar o disminuir continuamente, con lo que será imposible mantener el supuesto de que la esperanza matemática de la sucesión de variables aleatorias que configura el proceso estocástico correspondiente a dicha variable, es constante. En consecuencia, tampoco podrá mantenerse que la distribución de probabilidad de dichas variables es la misma a través del tiempo. Sin embargo, si efectuamos una correcta especificación de la estructura de dicha

tendencia, podrá estimarse y extraerse del precio, para obtener una variable estacionaria, que no presentaría las dificultades antes mencionadas.

Mayor dificultad entraña el caso en que una variable precio incluye una tendencia estocástica pues, en tal caso, su esperanza y varianza no están definidas. La presencia de una tendencia estocástica requiere transformar la variable, generalmente en primeras diferencias temporales, o tomando las diferencias entre las observaciones correspondientes a una misma estación cronológica, en el caso de una variable estacional. La transformación mediante diferencias resulta bastante natural en el análisis de datos financieros, por cuanto que la primera diferencia del logaritmo de un precio, en logaritmos, es la *rentabilidad* del activo, lo que hace que la transformación logarítmica sea utilizada muy frecuentemente.

Como prácticamente ningún precio o índice financiero es estacionario, el uso indiscriminado de un estadístico como la varianza o la desviación típica como indicador de riesgo conduce a medidas de volatilidad sesgadas al alza. Consideremos un modelo muy popular en el análisis de mercados financieros, el camino aleatorio:

$$y_t = \mu + y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots$$

que evoluciona a partir de un valor inicial y_0 dado, donde ε_t es un *ruido blanco*: sucesión de variables aleatorias, independientes, con media constante (que suponemos cero), y varianza asimismo constante σ_ε^2 . Mediante sucesivas sustituciones, este proceso puede escribirse, de modo equivalente:

$$y_t = y_0 + t\mu + \sum_{s=1}^t \varepsilon_s$$

En consecuencia, un camino aleatorio y_t tiene varianza creciente en el tiempo:

$$Var(y_t) = t\sigma_\varepsilon^2$$

Ello se debe a que el último sumando en la representación anterior es un ejemplo de *tendencia estocástica*. Cuanto mayor sea el número de observaciones consideradas, mayor será la varianza muestral del camino aleatorio: un camino aleatorio tiene menor varianza a lo largo de una hora que a lo largo de un día, a lo largo de un día que a lo largo de una semana, etc..

Esto es lo que ocurrirá con la inmensa mayoría de los precios cotizados en los mercados financieros. Aunque la presencia de tendencias estocásticas se produce generalmente junto con estructuras más complejas que la de un camino aleatorio, la implicación acerca de una varianza creciente con el tiempo se mantiene cuando se añaden a ésta componentes autoregresivos o de medias móviles para y_t . Para evitarlo, caracterizamos la volatilidad de un mercado o de un activo analizando el comportamiento de la rentabilidad que ofrece a lo largo del tiempo, no de su precio o cotización.

1.10 Regresión espúrea

El problema de la regresión espúrea fue analizado por Granger y Newbold (1974), quienes mostraron la posibilidad de que, en determinadas situaciones, estimaciones mínimocuadráticas de un modelo de regresión lineal que sugieren una estrecha relación entre variable dependiente y variables independientes, están reflejando, en realidad, una relación espúrea o ficticia, que en realidad no existe. Es evidente que tal posibilidad sería extremadamente peligrosa, tanto en la estimación de coeficientes de impacto o elasticidades, como en la contrastación de hipótesis teóricas. Lo que suele ignorarse con demasiada frecuencia es que las condiciones para que una regresión sea espúrea se dan con mucha frecuencia en la investigación aplicada en Economía, en general, y en Finanzas, en particular.

Comenzamos describiendo el tipo de dificultades a que puede dar lugar la ausencia de estacionariedad de las variables dependiente e independiente en un modelo de regresión lineal. Para ello, pensemos en el siguiente ejercicio: comenzamos simulando dos ruidos blancos independientes, $\varepsilon_{x_t}, \varepsilon_{y_t}, t = 1, 2, \dots, T$, a partir de distribuciones de probabilidad Normal, con esperanza matemática $\mu_{\varepsilon_x}, \mu_{\varepsilon_y}$ (por ejemplo, iguales a cero) y varianzas $\sigma_{\varepsilon_x}^2, \sigma_{\varepsilon_y}^2$; el coeficiente de correlación muestral entre las series temporales resultantes será, por construcción, muy reducido, si bien no exactamente igual a cero.

Nota: Cuanto mayor sea el tamaño muestral, más probable es que dicha correlación sea igual a cero, debido a que la correlación muestral, es decir, la correlación entre las dos series temporales simuladas es, por la ley de los grandes números, un estimador consistente de su análogo poblacional, que es el coeficiente de correlación teórico entre los dos procesos $\varepsilon_{x_t}, \varepsilon_{y_t}$, que es igual a cero. Por tanto, al aumentar T , la distribución de probabilidad del coeficiente de correlación muestral se concentra alrededor de cero.

El gráfico de ambas variables presentará una pauta oscilando alrededor de su media muestral que, por la misma razón apuntada para el coeficiente de correlación, serán próximas, si bien no iguales, a $\mu_{\varepsilon_x}, \mu_{\varepsilon_y}$. Observaremos que cada serie temporal cruza repetidamente su nivel medio. Si estimamos una regresión del tipo:

$$\varepsilon_{y_t} = \beta_0 + \beta_1 \varepsilon_{x_t} + u_t, \quad t = 1, 2, \dots, T$$

deberíamos obtener una estimación de β_1 no significativamente diferente de cero, y un R^2 prácticamente nulo. En efecto, salvo por el error estadístico, así ocurre cuando llevamos a cabo un ejercicio de simulación de Monte Carlo: al 95% de confianza, el habitual contraste tipo t rechazará la hipótesis nula de ausencia de capacidad explicativa de ε_{x_t} $H_0 : \beta_1 = 0$ aproximadamente en un 5% de los casos, y el valor mediano del coeficiente de determinación R^2 para todas las simulaciones es muy reducido. El término constante sólo resultaría significativo si en la generación de las series temporales, hemos utilizado valores diferentes de $\mu_{\varepsilon_x}, \mu_{\varepsilon_y}$.

Este resultado no se ve afectado significativamente en ningún otro sentido por la presencia de tales términos constantes, ni tampoco por cambios en el valor de las respectivas varianzas. Al variar el valor relativo de $\sigma_{\varepsilon_y}^2 / \sigma_{\varepsilon_x}^2$ tan

sólo se observa un comportamiento algo errático del tamaño del contraste de significación del parámetro β_0 . En definitiva, en esta primera parte del ejercicio tendremos el resultado que esperaríamos: una regresión no significativa, excepto en lo relativo al nivel escogido para el contraste.

1.10.1 Regresión espúrea bajo tendencias deterministas

A continuación, añadimos una tendencia lineal determinista a cada una de ellos,

$$\begin{aligned}y_t^* &= at + \varepsilon_{y_t} \\x_t^* &= bt + \varepsilon_{x_t}\end{aligned}$$

donde a y b son constantes arbitrarias y t es una tendencia determinista, es decir, una variable que aumenta cada período en una cantidad constante, Δ .

Si calculamos el coeficiente de correlación muestral entre x_t e y_t , apreciaremos que es elevado. Esto es sorprendente porque, como muestran las expresiones anteriores, cada variable es la suma de un componente de naturaleza determinista, que no experimenta ninguna fluctuación aleatoria, y un segundo componente de naturaleza estocástica. El coeficiente de correlación debería indicar la asociación estadística entre ambas variables, que es lo mismo que la asociación entre sus componentes estocásticos, es decir, entre sus innovaciones. Pero dicha correlación debería ser, por construcción, prácticamente igual a cero, en contra del resultado que se obtiene cuando se lleva a cabo este ejercicio de simulación. En todo caso, tal elevada correlación no refleja ninguna relación real entre las variables, por lo que se denomina *correlación espúrea*.

Como consecuencia de la misma, si se estima una regresión lineal, tomando cualquiera de estas variables como variable dependiente y la otra como independiente,

$$y_t^* = \beta_0 + \beta_1 x_t^* + v_t, \quad t = 1, 2, \dots, T$$

los resultados cambian sustancialmente: se obtiene un R -cuadrado elevado pues, como ya sabemos, es igual al cuadrado del coeficiente de correlación entre ambas variables, a la vez que una pendiente β_1 aparentemente significativa, de acuerdo con el criterio habitual de utilizar su estadístico tipo *t-Student*. Ambas cosas ocurrirán en un elevado porcentaje de las simulaciones que realicemos, para distintas series temporales de $\varepsilon_{x_t}, \varepsilon_{y_t}, t = 1, 2, \dots, T$. Por consiguiente, creeríamos que la capacidad explicativa de la variable x_t^* sobre y_t^* es muy importante. Este resultado es sorprendente, por cuanto que las variables y_t^*, x_t^* tienen la misma estructura estocástica que $\varepsilon_{x_t}, \varepsilon_{y_t}$, por lo que ambas relaciones deberían proporcionar resultados análogos. Esta apariencia ficticia de capacidad explicativa es lo que se conoce como *regresión espúrea*.

El grado de correlación observado entre y_t^*, x_t^* depende de dos factores: la similitud entre las constantes a y b , y la relación entre ambas y las desviaciones típicas de los ruidos blancos originales, $\varepsilon_{x_t}, \varepsilon_{y_t}$. Si, por ejemplo, fijamos el valor numérico de b en $b = 1$, y vamos tomando en cada ejercicio de simulación valores:

$a = [0, 1; 0, 5; 0, 9; 1; 3; 10; 100]$ el coeficiente de determinación resultante, R^2 aumenta monótonicamente con el valor de a . Es decir, la mayor correlación no se obtiene cuando hay a ambos lados de la igualdad la misma pendiente, lo que equivaldría a utilizar $a = 1$, sino que dicha correlación aumenta con a . Esto se debe a que, según aumenta a , cada vez hay más tendencia determinista en y_t^* , en el sentido de que ésta predomina sobre el componente estocástico ε_{y_t} , y dicha tendencia determinista puede explicarse muy adecuadamente mediante el componente análogo de x_t^* .

1.10.2 Regresión espúrea bajo tendencias estocásticas

En su trabajo pionero, Granger y Newbold (1974) trataron el problema de no estacionariedad producido por la presencia de tendencias estocásticas o raíces unitarias. Para ello, realizaron el siguiente ejercicio: a partir de la simulación de dos ruidos blancos independientes que tendrán, por construcción, como antes, un coeficiente de correlación muestral muy reducido, añadieron una *raíz unitaria* o *tendencia estocástica* a cada uno de ellos,

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_{y_t} \\ x_t &= x_{t-1} + \varepsilon_{x_t} \end{aligned}$$

obteniendo que el coeficiente de correlación entre x_t e y_t era muy próximo a la unidad. Esto es sorprendente, por cuanto que, a partir de condiciones iniciales conocidas, los valores de ambas variables en cada instante de tiempo pueden escribirse como,

$$y_t = y_0 + \sum_{s=1}^t \varepsilon_{y_s}$$

que indican que la evolución temporal de cada una de las variables se debe a la acumulación temporal de sus innovaciones. Por tanto, la naturaleza estocástica de cada variable está totalmente determinada por la naturaleza de sus innovaciones. Si ε_{x_t} y ε_{y_t} son independientes, entonces también deberían serlo x_t e y_t , en contra de los valores obtenidos para sus coeficientes de correlación muestrales en repetidas simulaciones. En todo caso, nuevamente, tal elevada correlación no refleja ninguna relación real entre las variables, por lo que se denomina *correlación espúrea*.

Si estimamos una regresión lineal entre estas variables, en cualquier orden, tendremos de nuevo un R -cuadrado elevado y una pendiente significativa, de acuerdo con el criterio habitual de utilizar su estadístico tipo *t-Student*, pero la evidencia de capacidad explicativa proporcionada por esta regresión sería *espúrea*.

Si las series temporales obtenidas mediante simulación para las innovaciones o ruidos blancos ε_{x_t} y ε_{y_t} tuviesen correlación distinta de cero, las variables x_t e y_t de los ejemplos anteriores mostrarían correlaciones muestrales similares

a las que se encuentran en los ejercicios de simulación descritos. En ese caso, los elevados coeficientes de correlación no serían tan engañosos, si bien serían numéricamente más altos de lo que la correlación entre x_t e y_t haría esperar.

En un ejercicio de simulación como el descrito, Granger y Newbold encontraron una frecuencia aproximada de rechazos de la hipótesis nula $H_0 : \beta_1 = 0$ del 76%. La frecuencia de rechazos de la capacidad explicativa global de la regresión se eleva muy significativamente al aumentar el número de variables explicativas independientes con estructura de ruido blanco. Nuevamente los coeficientes de determinación son muy elevados, lo que sorprende, pues realmente, x_t no explica apenas a y_t . El estadístico de Durbin-Watson habitualmente utilizado para contrastar ausencia de autocorrelación se reduce hacia cero, por lo que la combinación de este hecho con un elevado R^2 suele utilizarse como indicio de una regresión espúrea.

Ejercicio de simulación

En la pestaña *Datos* del archivo *Espureo.xls*, se han generado dos series temporales correspondientes a una población Normal $N(0,1)$. El generador de números aleatorios de Excel produce observaciones independientes entre sí, por lo que ambas series temporales se obtienen asimismo de manera independiente. La correlación poblacional entre ellas es cero, si bien la correlación muestral, al final de ambas variables, es de 0,021. Tampoco la media y desviación típica muestrales de cada variable son exactamente 0 y 1, como sus valores teóricos, si bien no difieren mucho de ellos. El coeficiente de asimetría teórico, así como el exceso de curtosis (exceso respecto de 3.0, que es la curtosis de toda población Normal), deberían ser ambos igual a cero lo que, nuevamente, ocurre sólo con carácter aproximado.

Hemso estimado una regresión para explicar la primera de estas variables por la segunda y una constante. El coeficiente de determinación es el cuadrado de su coeficiente de correlación y, por tanto, muy reducido. La estimación del coeficiente asociado a la variable explicativa aparece como no significativamente diferente de cero, de acuerdo con el estadístico t habitual. El gráfico que presenta el ajuste de la recta a la nube de puntos muestra una línea sin apenas pendiente, y una nube de puntos bastante circular. Ambos hechos reflejan una escasa correlación: una pendiente no significativa sugiere que la variable explicativa puede cambiar de valor sin que la variable dependiente cambie; una nube de puntos circular muestra que el rango de valores de cada una de las dos variables asociado a un valor determinado de la otra es muy amplio. En consecuencia, un valor numérico de cualquiera de ellas apenas nos informa acerca del valor de la otra variable. Esta es la manifestación de la ausencia de correlación entre ambas.

Lo contrario ocurre al estimar una regresión lineal entre las variables, una vez que se ha añadido una tendencia determinista a cada una de ellas. Para ello, en la pestaña *Determinista* se han generado dos nuevas variables, sumando una tendencia lineal a la variable Y incrementando en 0,05 unidades su valor cada periodo, y aumentando 0,10 unidades a la variable X en cada periodo. Aunque el componente estocástico en ambas variables es el mismo de antes, la nube de puntos entre ambas tiene un perfil totalmente distinto, siendo prácti-

camente una linea recta [ver junto a la regresión estimada, al final de ambas series temporales]. La correlación entre ambas variables se ha incrementado hasta 0,790. Esto se debe a que el componente tendencial predomina sobre el estocástico; como consecuencia, la regresión estimada entre ambas variables muestra un coeficiente de determinación muy próximo a la unidad. Lo que es quizá más preocupante, es que la pendiente estimada en dicha regresión, que es sustancialmente más elevada que la estimada con las variables originales, aparece como claramente significativa, sugiriendo una importante capacidad explicativa a la variable independiente, contrariamente a lo que detectamos en la primera regresión. Aquí cabe discutir si este es un resultado razonable: podría argumentarse que ambas variables tienen un componente tendencial importante y que, en ese sentido, no es sorprendente que el coeficiente de determinación entre ambas sea elevado. Es cierto, pero sólo refleja la relación entre los componentes deterministas, que no son los que deben concentrar la atención del analista: si hay componentes deterministas en las variables dependiente y explicativas, el analista debería indagar las razones que explican la presencia simultánea de tales elementos en las variables. Esta sería uno de los elementos del análisis; el segundo consistiría en evaluar la relación entre los componentes estocásticos en ambas variables; éste elemento es importante, pues nos proporciona información acerca del impacto que sobre la variable dependiente puede tener una intervención sobre alguna de las variables explicativas, o una alteración exógena en su valor numérico. La dificultad con la regresión anterior estriba en que si no se lleva a cabo este análisis por componentes, el examen mecánico de los resultados de la regresión sugeriría que la variable dependiente reacciona a fluctuaciones inesperadas en la variable explicativa, cuando no es así; tal conclusión sería un error.

Posteriormente, hemos mantenido los mismos elementos tendenciales de ambas variables, pero hemos incrementado de manera apreciable el componente aleatorio en ellas. Siendo tales componentes variables aleatorias de esperanza matemática igual a cero, su tamaño queda representado por su desviación típica, que era unitaria para ambas en las regresiones anteriores. En la pestaña *Determinista*(2) hemos generado otras dos variables con desviaciones típicas 3 y 5; la correlación entre ellas desciende, si bien no de manera dramática, situándose en 0,237.

1.11 Tratamiento de tendencias deterministas

De las dos situaciones descritas en el apartado anterior, es algo más sencilla de tratar la presencia de tendencias deterministas, cuando se anticipa correctamente que la presencia de las mismas es la única causa de no estacionariedad de las variables que se pretende relacionar, es decir, cuando las variables tienen estructura,

$$y_t = \alpha_0 + \alpha_1 t + \varepsilon_{y_t} \quad (8)$$

$$x_t = \delta_0 + \delta_1 t + \varepsilon_{x_t} \quad (9)$$

Para ello, hay dos posibilidades: la primera consiste en incorporar en el modelo de regresión una tendencia determinista lineal como variable explicativa,

$$y_t = \alpha + \gamma t + \beta x_t + u_t \quad (10)$$

en la que el coeficiente estimado $\hat{\beta}$ y su desviación típica serán, aproximadamente, los mismos que habríamos estimado en la regresión,

$$\varepsilon_{y_t} = \eta_0 + \eta_1 \varepsilon_{x_t} \quad (11)$$

En la regresión (10) que hemos sugerido todavía tendrá un R -cuadrado muy elevado, debido a la capacidad explicativa que el término γt tiene sobre y_t , debido a la presencia de una tendencia determinista en esta última variable. Este término aparecerá como claramente significativo, con un estadístico t muy elevado.

Esto significa que si ambas innovaciones son independientes, en la regresión (10) se tendrá un coeficiente reducido en magnitud, y estadísticamente no significativo, en términos de su estadístico t de Student.

Esto es distinto del resultado que se obtiene en la estimación de la regresión habitual,

$$y_t = \alpha + \beta x_t + u_t \quad (12)$$

cuando las variables tienen estructura (9), (8). En dicha regresión se tendría un R -cuadrado elevado, una estimación numérica de β relativamente elevada, y un estadístico t para dicho coeficiente, claramente por encima de 2,0 en valor absoluto, sugiriendo que la capacidad de x_t para explicar y_t es significativa, contrariamente a lo que, en realidad, ocurre.

La diferenciación elimina las tendencias deterministas. Si consideramos una tendencia determinista:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

la primera diferencia del proceso es:

$$\Delta y_t = y_t - y_{t-1} = (\beta_1 - \beta_2) + (\varepsilon_t - \varepsilon_{t-1})$$

sin tendencia, aunque hemos introducido una estructura de media móvil, MA(1), en el término de error de la regresión resultante.

De este modo, si el precio de un determinado activo tiene una tendencia temporal determinista lineal, su primera diferencia estará libre de dicha tendencia. Un proceso con una tendencia determinista cuadrática sigue trayectorias con formas parabólicas, cóncavas o convexas. Su primera diferencia presentará una tendencia lineal, mientras que su segunda diferencia estará libre de tendencia.¹⁰

¹⁰Si consideramos una tendencia determinista de segundo grado:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$$

cuya primera diferencia es:

Un proceso con una tendencia determinista cuadrática sigue trayectorias con formas parabólicas, cóncavas o convexas, dependiendo del signo del coeficiente del término de segundo grado. Su primera diferencia presentará una tendencia lineal, mientras que su segunda diferencia estará libre de tendencia. Un proceso con una tendencia determinista representada por un polinomio de grado tres puede tener ciclos. La primera diferencia de este proceso tendrá una tendencia cuadrática.

Por tanto, aparentemente, una solución en el caso en que sospechamos que puede haber tendencias deterministas en las variables que pretendemos relacionar, consistiría en estimar la posible relación entre ellas después de haber tomado diferencias temporales. Sin embargo, con dicha transformación perdemos bastante información acerca de las fluctuaciones de corto plazo en las variables, por lo que los procedimientos anteriormente descritos son más recomendables. Sin embargo, aunque en datos financieros pueda apreciarse la presencia de alguna tendencia determinista de grado 1 o 2, esa no es la causa principal de no estacionariedad, sino la presencia de raíces unitarias.

Ejercicio de simulación

Simule 300 observaciones de dos ruidos blancos independientes, con distribuciones $N(\mu_x, \sigma_x^2), N(\mu_y, \sigma_y^2)$, como observaciones muestrales para las innovaciones $\varepsilon_{x_t}, \varepsilon_{y_t}$. A continuación, genere observaciones para una tendencia determinista t . Los valores numéricos para las variables x e y se obtienen añadiendo la tendencia t , multiplicada por sendas constantes a, b , a las respectivas innovaciones, para reproducir las estructuras (9), (8).

El ejercicio consiste en comparar el coeficiente de correlación que se obtiene para $\varepsilon_{x_t}, \varepsilon_{y_t}$, que será muy reducido, con el que se obtiene entre x_t e y_t , que debería ser similar, pero será, sin embargo, muy elevado. En segundo lugar, deben estimarse regresiones análogas a (11), (12), (10) y compare los resultados que obtiene en cada una de ellas. ¿A qué cree que se debe las diferencias entre dichos resultados? Estime ahora una regresión con las primeras diferencias de las variables y compare los resultados con los anteriormente obtenidos.

1.12 Tendencias estocásticas

A random walk as the limit of an AR(1) process

As the coefficient ρ in an AR(1) process approaches 1, the first order autoregression becomes a random walk, that has an infinite variance. This is because if we repeat for the random walk the same iterated substitutions argument we made for an AR(1), we get,

$$\Delta y_t = y_t - y_{t-1} = (\beta_1 - \beta_2) + 2\beta_2 t + (\varepsilon_t - \varepsilon_{t-1})$$

siendo su segunda diferencia:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - 2y_{t-1} + y_{t-2} = 2\beta_2 + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2})$$

$$\begin{aligned}
y_t &= a + y_{t-1} + \varepsilon_t = a + (a + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \\
&= as + y_{t-s} + (\varepsilon_{t-s+1} + \dots + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t)
\end{aligned}$$

At a difference from a stationary AR(1) process, the past term y_{t-s} does not die away no matter how far we move back into the past, and the variance of the sum in brackets increases without bound as we move backwards in time. The random walk process has an infinite variance. Sometimes, it can be assumed that there is a known initial condition y_0 . The random walk process can then be represented:

$$\begin{aligned}
y_t &= a + y_{t-1} + \varepsilon_t = a + (a + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \\
&= \dots = at + y_0 + (\varepsilon_1 + \dots + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t)
\end{aligned} \tag{13}$$

with $E(y_t) = y_0 + ta$ and $Var(y_t) = t\sigma_\varepsilon^2$. Hence, both moments change over time, the variance increasing without any bound.

Ello se debe a que el último sumando en la representación anterior, $\sum_{s=1}^t \varepsilon_s$, es un ejemplo de *tendencia estocástica*. *Una tendencia estocástica es un componente de una variable, cuya varianza tiende a infinito con el tamaño de la muestra*. Esto es lo que sucede con el camino aleatorio: cuanto mayor sea el número de observaciones consideradas, mayor será la varianza muestral del camino aleatorio: un camino aleatorio tiene menor varianza a lo largo de una hora que a lo largo de un día, a lo largo de un día que a lo largo de una semana, etc.. El aumento de la varianza a lo largo del tiempo no tiene nada que ver con el término at que aparece en (13) que, siendo determinista, tiene varianza cero.

La tendencia estocástica aparece debido al coeficiente unitario del retardo de y_t en el modelo AR(1) que explica el comportamiento de esta variable. En el lenguaje estadístico, se dice que el proceso y_t tiene una *raíz unitaria*.

Esto es lo que ocurrirá con la inmensa mayoría de los precios cotizados en los mercados financieros. La propiedad de que la varianza de un camino aleatorio crece sin límite con el tiempo se mantiene cuando se añaden a la estructura de camino aleatorio componentes autoregresivos o de medias móviles para y_t . En general, de hecho, muchos procesos relativamente complejos incorporan en su estructura un camino aleatorio. Para evitar los problemas estadísticos que plantea tratar con procesos que incorporan tendencias estocásticas, caracterizamos la volatilidad de un mercado o de un activo analizando el comportamiento de la rentabilidad que ofrece a lo largo del tiempo, no analizando su precio.

If we compare in the same graph time series realizations of a random walk together with some stationary autoregressive processes, it will be hard to tell which is the process with an infinite variance. That means that if we test the null hypothesis that a time series comes from a random walk process against the alternative hypothesis that it comes from an AR(1) process, the tests might have low power, being difficult to reject the random walk hypothesis when the true process is an AR(1) with a high ρ coefficient.

En general, decimos que un proceso estocástico tiene una raíz unitaria cuando la ecuación característica del polinomio de retardos que representa su estructura autoregresiva, tiene una raíz unitaria. Por ejemplo, teniendo en cuenta el operador de retardos: $Ly_t = y_{t-1}$, $L^s y_t = y_{t-s}$, el proceso AR(1) puede escribirse: $(1 - \rho L)y_t = \alpha + \varepsilon_t$. La posible presencia de una constante no añade nada a esta argumentación. La ecuación característica de su polinomio de retardos es $1 - \rho L = 0$, que tiene por solución: $L = 1/\rho$. Si $\rho = 1$, entonces el proceso AR(1) tiene una raíz unitaria. Si $\rho < 1$, la ecuación característica tiene una raíz mayor que 1 y el proceso es estacionario.¹¹ Lo contrario sucede si $\rho > 1$.

Un proceso AR(2) puede representarse:

$$(1 - \rho_1 L - \rho_2 L^2)y_t = \alpha + \varepsilon_t$$

y el proceso será estacionario si la ecuación característica: $1 - \rho_1 L - \rho_2 L^2 = 0$ tiene dos raíces reales mayores que 1. Si tiene dos raíces complejas conjugadas, su modulo deberá ser mayor que 1 para que el proceso sea estacionario. De modo similar puede analizarse la estacionariedad en modelos más complejos. En todo caso, esta depende exclusivamente de las propiedades de la estructura autoregresiva de proceso, pues el componente de media móvil no afecta a la estacionariedad del mismo. Es decir, el proceso: $y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t - \theta \varepsilon_{t-1}$ será estacionario si y sólo si el proceso $y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t$ lo es.

En un proceso AR(1), la tendencia estocástica aparece debido al coeficiente unitario del retardo de y_t en la ecuación que explica el comportamiento de esta variable, por lo que una tendencia estocástica se conoce asimismo como una *raíz unitaria*. Con más generalidad, recordemos que por la descomposición de Wald, todo proceso estacionario acepta una representación autoregresiva, quizá de orden infinito,

$$y_t = \alpha_0 + \sum_{s=1}^{\infty} \alpha_s y_{t-s} = \alpha(L) y_t$$

donde L denota el operador de retardos, definido como $L^j y_t = y_{t-j}$. Si obtenemos las raíces de dicho polinomio de retardos, podremos escribir, $\alpha(L) = \prod_{i=1}^p (1 - a_i L) \prod_{j=1}^q (1 - b_j L - c_j L^2)$, donde los últimos factores tienen como raíces dos números complejos conjugados. una raíz unitaria es un factor del primer tipo, con $a_i = 1$. En el lenguaje estadístico, se dice que el proceso y_t tiene una raíz unitaria.

1.13 Variables integradas. Cointegración

Una variable y_t se dice que es integrada de orden d , lo cual representamos como $y_t \sim I(d)$ si es preciso diferenciar serie temporal d veces para conseguir una transformación que sea estacionaria. Una variable estacionaria es integrada de orden 0. Dicho de otro modo una variable es integrada de orden d si el polinomio

¹¹Nótese que la estacionariedad requiere que el coeficiente dle proceso AR(1) sea *menor* que 1, lo que equivale a que el polinomio de retardos tenga una raíz *mayor* que 1.

característico asociado a su representación autoregresiva tiene exactamente d raíces unitarias. Si una serie temporal es $I(d)$, entonces puede escribirse:

$$y_t = (1 - L)^d z_t$$

con $z_t \sim I(0)$, estacionario.

Consideremos dos variables integradas: $x_t \sim I(d)$, $z_t \sim I(e)$ con $e > d$. Toda variable que sea combinación lineal de ambas variables es, en general, $I(e)$, puesto que habría que diferenciar dicha combinación e veces para conseguir una variable estacionaria. Lo sorprendente es que si las dos variables fuesen integradas del mismo orden, d , podría haber situaciones en que una combinación lineal fuese integrada de orden inferior a d . Se dice entonces que ambas variables están cointegradas de orden $d - b$.

El caso quizá mas interesante es: $x_t \sim I(1)$, $z_t \sim I(1)$. en general, si $y_t = ax_t + bz_t$, entonces $y_t \sim I(1)$. Pero es posible en algunos casos que $y_t \sim I(0)$ sea estacionaria. En tal caso, se dice que x_t y z_t están cointegradas. Por ejemplo, supongamos que el IBEX35 y el Dow Jones son variables $I(1)$ ¿Existirá alguna combinación lineal de ambos índices que sea estacionario? Si esto sucediera, interpretariamos que la raíz unitaria que contienen ambas variables es la misma y, por tanto, una cartera que tomase posiciones en ambos índices de acuerdo con los coeficientes de dicha combinación lineal tendría un precio de mercado estacionario, mostrando reversion a la media (o a su tendencia, si tiene comportamiento alcista). Se diría entonces que ambos mercados mantienen una relación de equilibrio a largo plazo. Esta denominación pretende recoger el hecho de que cualquier distanciamiento *permanente* respecto de su nivel de equilibrio de uno de dichos activos x_t , debe ir acompañado de otro distanciamiento, asimismo *permanente*, en el otro activo z_t , de tal manera que su combinación lineal $y_t = ax_t + bz_t$ permanezca próxima a su nivel de equilibrio. Pero no impide en absoluto que se produzcan desviaciones *transitorias* en ambos activos que separen a y_t *transitoriamente* de su nivel de equilibrio. Entre otras cosas, esto sugeriría que podríamos tomar posiciones en ambos mercados (muy posiblemente de signo opuesto) para aprovechar desviaciones *transitorias* en ambos mercados respecto de la relación de equilibrio que mantendrían a largo plazo. Este es el fundamento estadístico de la negociación por pares, que ha suscitado mucho interés en los últimos años.

Se puede tener cointegración entre varias variables, incluso con diferente orden de integración. Entonces, están cointegradas si existe una combinación lineal de las variables de mayor orden de integración que están cointegradas con las variables con menor orden de integración del grupo de variables. Por ejemplo, si $x_t \sim I(2)$, $z_t \sim I(2)$, $w_t \sim I(1)$, las tres variables están cointegradas si existen coeficientes a, b tales que $v_t = ax_t + bz_t \sim I(1)$ y $cv_t + w_t \sim I(0)$.

Si x_t y z_t están cointegradas, entonces x_t y z_{t-i} también están cointegradas.

Si $x_t \sim I(1)$ y $z_t \sim I(1)$ están cointegradas, entonces también están cointegradas con cualquier variable estacionaria.

Si x_t y z_t están cointegradas, entonces o bien x_t causa en el sentido de Granger a z_t , o z_t causa en el sentido de Granger a x_t , o suceden ambas cosas

a la vez. Esto se debe a la existencia del *modelo de corrección del error*.

El estimador MCO de una relación de cointegración es superconsistente, es decir, converge al verdadero valor de la parámetro al aumentar el tamaño de la muestra mas rapidamente que lo hace el estimador MCO de un modelo de regresión entre variables estacionarias. Además, su consistencia no requiere la ausencia de correlación entre variables explicativas de la regresión de cointegración y el término de error. De hecho, en la regresión de cointegracion puede tomarse como variable dependiente cualquiera de las variables del conjunto cuya cointegración se esta contrastando. Sin embargo, dicho estimador es sesgado en muestras finitas, y su distribución limite no es estándar. El tamaño del sesgo está relacionado con $1-R^2$ de la regresión de cointegración, de modo que si el R^2 es alto, dicho sesgo será pequeño.

Engle y Granger (1987) demostraron que el estimador en dos etapas del modelo de corrección del error es asimismo consistente.

1.13.1 Contrastes de raíz unitaria

Secciones 14.2 a 14.5 en Econometría - Novales

1.13.2 Raíces unitarias estacionales

Sección 14.6 en Econometría - Novales

- [ver referencia recomendada sobre raíces unitarias y contrastes de raíz unitaria]

1.14 Modelos ARIMA estacionales

Sección 13.8 en Econometría - Novales

1.15 Estimación de modelos AR y MA

Sección 13.10 en Econometría - Novales

1.16 Contrastes de especificación de un modelo ARIMA

Sección 13.11 en Econometría - Novales

1.17 Cointegración

Se dice que dos variables x_t, y_t están cointegradas si teniendo ambas una raíz unitaria, existe una constante β tal que $x_t - \beta y_t$ es estacionaria. Evidentemente, existe otra constante $\gamma = 1/\beta$ tal que $\alpha x_t - y_t$ es una variable estacionaria.

Si ambas variables están cointegradas, existe una única relación de cointegración. En efecto, supongamos que además de ser estacionaria $x_t - \beta y_t$, también lo fuese $x_t - \delta y_t$ para un cierto δ . Restando ambas variables tendríamos que $(\delta - \beta)y_t$ debería ser estacionaria, ya que la diferencia de dos variables estacionarias lo es. Pero si y_t es no estacionaria, su producto por cualquier constante tampoco será estacionario, alcanzándose así una contradicción.

La interpretación de la cointegración es que la tendencia estocástica contenida en ambas variables es la misma, de modo que puede encontrarse una combinación lineal (no necesariamente con coeficiente 1), que carece de dicha tendencia estocástica. Puesto que la tendencia estocástica refleja un comportamiento de largo plazo, decimos que ambas variables mantienen una relación de equilibrio a largo plazo, de la que experimentan únicamente desviaciones de carácter transitorio.

1.17.1 ¿Cómo tratar las relaciones entre variables en presencia de tendencias estocásticas?

La diferenciación, que vimos que elimina las tendencias deterministas lineales, elimina también las tendencias estocásticas, pues tendríamos,

$$y_t - y_{t-1} = \Delta y_t = a + \varepsilon_t, \quad t = 1, 2, \dots$$

con $E(y_t) = a$, $Var(y_t) = \sigma_\varepsilon^2$, para todo t .

Un proceso estocástico puede tener asimismo varias raíces unitarias. Los tipos de interés ya son rentabilidades, por lo que tienen, generalmente, un orden de no estacionariedad (es decir, un número de tendencias) menos que las

series de índices bursátiles o de precios de derivados, por ejemplo. En ocasiones, sin embargo, algunas series de precios son no estacionarias de orden 2 (tienen 2 raíces unitarias), por lo que incluso las rentabilidades pueden ser no estacionarias, presentando una raíz unitaria.

De acuerdo con la discusión que hemos llevado a cabo en las secciones anteriores, el procedimiento a seguir en caso de presencia de raíces unitarias en las variables de una regresión lineal simple es claro. Si tanto x_t como y_t son variables $I(1)$, es decir, tienen una raíz unitaria, entonces el tratamiento que hemos de aplicar depende de si están o no cointegradas. Si lo están, hemos de estimar un *modelo de corrección de error* (ver capítulo sobre VAR, Vectores Autoregresivos). Si las dos variables no están cointegradas, hemos de estimar un modelo en diferencias.

En el ejercicio de simulación descrito más arriba, la estimación de la relación en primeras diferencias,

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + v_t, \quad t = 1, 2, \dots, T$$

arrojará un coeficiente β_1 no significativo y un coeficiente de determinación muy reducido en la mayoría de las simulaciones.

Esto significa, entre otras cosas, que la recomendación de tratar la no estacionariedad diferenciando las variables, no es correcta. Tal sugerencia es válida cuando, existiendo raíces unitarias en ambas variables, no están cointegradas. Cuando están cointegradas, el modelo que se estime debe especificar la relación de corto plazo entre ambas variables, pero también la relación de largo plazo que pueda existir entre ellas. Este es el *modelo de corrección del error*, que incorpora un *término de corrección de error*:

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \gamma(y_{t-1} - x_{t-1}) + v_t, \quad t = 1, 2, \dots, T$$

Aún así, subsisten algunos matices:

- **Modelo uniecuacional:** como hemos comentado anteriormente, la cointegración entre variables no dice nada acerca de la posible relación de causalidad entre ambas variables. De hecho, de acuerdo con el teorema de representación de Engle y Granger, el modelo de relación entre ambas variables es un modelo de corrección de error, que es un modelo de dos ecuaciones, una para y_t en diferencias, y otra para x_t en diferencias. En ambas aparece el *término de corrección de error* retardado como una de las variables explicativas, debiendo esperar que tome signo opuesto en cada una de las dos ecuaciones, según como se haya definido dicho término, por las razones antes expuestas. Además de dicho término, aparecerán posiblemente algunos retardos de las diferencias, tanto de x_t como de y_t . Sin embargo, es práctica habitual utilizar tal representación para especificar un modelo de regresión con una única ecuación, como proponen Engle y Granger (19xx). Al actuar así, hemos de interpretar que estamos estimando por separado tan sólo una de las ecuaciones del modelo de corrección de error, lo cual puede hacernos perder eficiencia en la estimación,

salvo si: a) las innovaciones en las dos ecuaciones están incorrelacionadas, o b) las dos ecuaciones tuvieran exactamente las mismas variables explicativas.

- ¿Qué diferencias? Ya sabemos que, en caso de cointegración, el modelo a estimar es una relación entre las variables x_t e y_t en diferencias. En muchos casos, el investigador dispone de observaciones mensuales o trimestrales de variables como el consumo agregado, el PIB de un país, la inversión, un agregado monetario, etc. Estas variables tienen, generalmente, una raíz unitaria, por lo que, en caso de querer relacionar dos de ellas, y en presencia de cointegración, deberíamos estimar un modelo de corrección de error. Sin embargo, no sólo la primera diferencia, es decir, la variación entre meses o trimestres sucesivos, $y_t - y_{t-1}$, sino la diferencia anual, $y_t - y_{t-4}$ en el caso de datos trimestrales, o $y_t - y_{t-12}$ en el caso de datos anuales, también son variables $I(0)$, es decir, estacionarias. Por tanto, el modelo de corrección de error puede especificarse en unas u otras diferencias, siempre que seamos consistentes en tratar tanto x_t como y_t de igual manera. Y, sin embargo, las propiedades estadísticas de unas u otras diferencias son bien diferentes; por ejemplo, su volatilidad es muy distinta. Además, es perfectamente concebible que la variación anual (es decir, la tasa interanual) de inflación esté correlacionada con la tasa interanual de crecimiento monetario, a la vez que las tasas de variación intermensuales (es decir, mes a mes) de ambas variables, no muestren una relación significativa. Por consiguiente, no sería lo mismo estimar un modelo de relación,

$$\Delta_{12}y_t \equiv y_t - y_{t-12} = \beta_0 + \beta_1 (x_t - x_{t-12}) + \gamma (y_{t-1} - \delta_0 - \delta_1 x_{t-1}) + u_t$$

que un modelo,

$$\Delta y_t \equiv y_t - y_{t-1} = \beta_0 + \beta_1 (x_t - x_{t-1}) + \gamma (y_{t-1} - \delta_0 - \delta_1 x_{t-1}) + u_t$$

de los que no cabe esperar resultados comparables. Tampoco debe pensarse que es ésta una cuestión estadística. Por el contrario, es el propio investigador quien debe decidir si piensa que la relación entre las variables se debe a las fluctuaciones que experimentan en períodos breves de tiempo, como un mes, o en períodos más amplios, como un año.

2 Appendix 1: Valoración por simulación

La valoración de una opción mediante simulación se ajusta a la idea general de simular el precio del subyacente desde el instante en que se valora la opción hasta el vencimiento de la misma. La práctica de valoración de derivados por simulación del precio del subyacente ha alcanzado un alto grado de sofisticación

computacional. Sin embargo, la calidad de la valoración resultante depende tanto de la complejidad y realismo sofisticación del modelo como de la calidad del modelo estadístico utilizado en la simulación del precio del subyacente. Hay dos componentes a tener en cuenta,

1. un buen modelo de evolución temporal del precio del subyacente
2. unas hipótesis adecuadas acerca de las características estocásticas de la innovación de dicho proceso.

Así, por ejemplo, suponer Normalidad de la innovación, al generar sendas futuras para el precio del subyacente, cuando existe evidencia de asimetría y curtosis puede generar errores de aproximación muy importantes al llevar a cabo la valoración. Es importante disponer de un modelo dinámico del precio del subyacente suficientemente bueno y utilizar una distribución de probabilidad adecuada al generar realizaciones simuladas para la innovación del proceso. Estas serán dos de las cuestiones a las que prestaremos atención en este curso.

En el caso de una opción Europea, basta con considerar en cada simulación el precio resultante al final de la serie temporal simulada, que coincide con el instante de vencimiento de la opción. Mediante un elevado número de realizaciones simuladas, podemos aproximar la distribución de probabilidad del precio del activo subyacente al vencimiento de la opción. De este modo, obtenemos el valor intrínseco de la opción a vencimiento para cada realización y, por tanto, una aproximación a la distribución de probabilidad de dicho valor intrínseco. De dicha distribución de probabilidad inferimos un precio actual para la opción a partir de un determinado mecanismo de valoración: una posibilidad es calcular la esperanza matemática de la distribución de probabilidad del valor intrínseco a vencimiento, y descontarlo al instante en que se efectúa la valoración.

En el caso de otros tipos de opciones, puede utilizarse, en general, un procedimiento análogo, si bien teniendo en cuenta a) todos los posibles instantes de ejercicio, b) el valor intrínseco en cada uno de ellos, c) el descuento apropiado a utilizar.

Sin embargo, hay otras posibilidades: una, interesante, consistiría en considerar los tipos de interés como estocásticos, y simular simultáneamente los tipos de interés y el precio del subyacente, una vez que hubiéramos recogido en el modelo la dependencia entre ambos. Por ejemplo,

$$\begin{aligned} y_t &= \beta_0 + \beta_1 r_t + \varepsilon_{yt} \\ r_t &= \alpha_0 + \alpha_1 r_{t-1} + \varepsilon_{2t} \end{aligned}$$

con $(\varepsilon_{1t}, \varepsilon_{2t}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$. Téngase en cuenta que, en un modelo de estas características, la relación entre los tipos de interés r_t y la rentabilidad del activo subyacente y_t se produce por dos vías: una, explícita, por la presencia de los tipos en la ecuación de la rentabilidad; otra, implícita, por la correlación entre las innovaciones de ambas ecuaciones.

En el caso en que la rentabilidad y los tipos tengan estructuras de volatilidad condicional no trivial, entonces podríamos establecer un modelo *ARCH* bivariante, en el que se pueden recoger las dependencias tanto entre rentabilidades como entre volatilidades.

3 Sobre simulación de procesos brownianos

De acuerdo con un proceso generalizado de Wiener, el precio de un activo evolucionaría de acuerdo con,

$$dS_t = \mu dt + \sigma dz$$

donde dz denota un proceso de Wiener básico, es decir, un proceso estocástico de Markov, con una variación promedio igual a cero, y una *tasa de varianza* anual igual a 1.0. Se conoce también como movimiento Browniano: 1) su variación durante un intervalo pequeño de tiempo, de amplitud Δt es $\Delta z = \varepsilon \Delta t$, siendo ε independiente en el tiempo, Normal(0,1), 2) los valores de Δz son independientes en el tiempo.

Sin embargo, el proceso anterior no recoge algunas de las características importantes de los precios de los activos financieros. Más interesante resulta el supuesto $\Delta S = \mu S \Delta t$ que, en el límite se convierte en, $dS = \mu S dt$, es decir, $dS/S = \mu dt$, de modo que,

$$S_T = S_0 e^{\mu T}$$

Para incorporar volatilidad, suponemos que la variabilidad en la rentabilidad porcentual del activo es independiente del precio del mismo, es decir, que la volatilidad del precio es proporcional a su nivel,

$$dS_t = \mu S_t dt + \sigma S_t dz$$

3.1 Distribución de los cambios en precio

Si el precio de un activo que no paga dividendos tiene una volatilidad anual del 30% y una rentabilidad esperada del 15% anual, compuesta de modo continuo, el proceso que sigue es,

$$\frac{dS_t}{S_t} = \mu dt + \sigma dz = 0.15dt + 0.30dz$$

y sobre intervalos finitos, pero cortos de tiempo,

$$\frac{\Delta S_t}{S_t} = 0.15\Delta t + 0.30\varepsilon_t \sqrt{\Delta t}$$

Por ejemplo, para un intervalo de una semana (una fracción 0,0192 de un año), si el precio inicial es 100, la variación semanal sería,

$$\Delta S_t = 100 \left(0.15 (0.0192) + 0.30 \varepsilon_t \sqrt{0.0192} \right) = 0.288 + 4.16 \varepsilon_t$$

siguiendo una distribución $Normal(0.288, 4.16)$. Al cabo de un año,

$$\Delta S_t = S_0 (0.15 + 0.30 \varepsilon_t) = 15 + 30 \varepsilon_t$$

con una distribución $Normal(15, 30)$, si bien esta extrapolación temporal es bastante más cuestionable.

Deberíamos obtener una muestra de la innovación ε_t , y sustituir repetidamente en la ecuación anterior. Así podríamos generar una distribución de frecuencias (probabilidad) del precio al final del intervalo de tiempo deseado.

Generalmente, el valor de un derivado sobre una acción es independiente del valor de μ . Por el contrario el valor de σ es clave, y normalmente oscila entre 20% y 40%.

Puede probarse que la volatilidad anual del precio de una acción es igual a la desviación típica de la rentabilidad ofrecida por dicho acción, compuesta de modo continuo, durante un año.

3.2 Distribución del logaritmo del precio

Asimismo, el lema de Ito implica que,

$$d \ln S_t = \left(\mu - \frac{\sigma^2}{2} \right) dt + \sigma dz$$

siguiendo por tanto un proceso de Wiener generalizado. Por tanto, la variación en $\ln S$ se distribuye $N\left(\left(\mu - \frac{\sigma^2}{2}\right)T, \sigma\sqrt{T}\right)$, y tenemos,

$$\ln S_T \sim Normal\left(\ln S_0 + \left(\mu - \frac{\sigma^2}{2}\right)T, \sigma\sqrt{T}\right)$$

En consecuencia,

$$E(S_T) = S_0 e^{\mu T}; \quad Var(S_T) = S_0^2 e^{2\mu T} (e^{\sigma^2 T} - 1)$$

que muestra que μ puede interpretarse como una tasa de rentabilidad esperada.

Al cabo de un año, tendríamos,

$$E(S_T) = 100e^{.30 \times 1} = 134.99; \quad Var(S_T) = 100^2 e^{2 \times .30 \times 1} (e^{.3^2} - 1) = 1715.97; \quad DT(S_T) = 41.42$$

3.3 Distribución de la rentabilidad

La propia rentabilidad cumple, por definición,

$$E(S_T) = S_0 e^{\eta T}$$

por lo que,

$$\eta = \frac{1}{T} \ln \frac{S_T}{S_0}$$

y utilizando la distribución de $\ln S_T$, tenemos que,

$$\eta \sim \left(N \left(\mu - \frac{\sigma^2}{2}, \frac{\sigma}{\sqrt{T}} \right) \right)$$

4 * IGNORAR Appendix 2: Sesgo asintótico en el modelo AR(1) con autocorrelación

Consideramos otra vez la estimación del modelo AR(1) con residuos autocorrelacionados, asimismo mediante una estructura AR(1):

$$\begin{aligned} y_t &= \alpha + \beta y_{t-1} + u_t, \quad t = 1, 2, \dots, |\beta| < 1 \\ u_t &= \rho u_{t-1} + \varepsilon_t = 0 \quad |\rho| < 1, \quad E\varepsilon_t = 0, \quad E(\varepsilon_t \varepsilon_{t-s}) = 0 \quad \forall s \neq 0, \quad E(\varepsilon_t^2) = \sigma_\varepsilon^2 \end{aligned}$$

Comencemos notando que, dados los límites que hemos supuesto para los valores numéricos de los parámetros β y ρ , y_t y u_t admiten las representaciones:

$$y_t = \sum_{s=0}^{\infty} \beta^s u_{t-s}; \quad u_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i};$$

Ahora bien, también tenemos una representación de y_t en función de la innovación del modelo:

$$y_t = (\beta + \rho) y_{t-1} - \beta \rho y_{t-2} + \varepsilon_t \quad (14)$$

Si multiplicamos esta expresión por y_{t-1} , sumamos de 3 a T y dividimos por $\sum_2^T y_{t-1}^2$, tenemos,

$$\hat{\beta}_{MCO} = \frac{\frac{1}{T} \sum_3^T y_t y_{t-1}}{\frac{1}{T} \sum_3^T y_{t-1}^2} = (\beta + \rho) - \beta \rho \frac{\frac{1}{T} \sum_3^T y_{t-1} y_{t-2}}{\frac{1}{T} \sum_3^T y_{t-1}^2} + \frac{\frac{1}{T} \sum_3^T \varepsilon_t y_{t-1}}{\frac{1}{T} \sum_3^T y_{t-1}^2}$$

Ahora, tomando límites en probabilidad y notando que:

$$p \lim \frac{1}{T} \sum_3^T y_t y_{t-1} = E(y_t y_{t-1}) = p \lim \frac{1}{T} \sum_3^T y_{t-1} y_{t-2}$$

tenemos:

$$p \lim \hat{\beta}_{MCO} = (\beta + \rho) - (\beta \rho) p \lim \hat{\beta}_{MCO} + \frac{p \lim \frac{1}{T} \sum_3^T \varepsilon_t y_{t-1}}{p \lim \frac{1}{T} \sum_3^T y_{t-1}^2}$$

La última fracción es igual a cero, ya que el numerador converge a $E(\varepsilon_t y_{t-1})$ y ya hemos visto que y_{t-1} es combinación lineal de $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$, pero no de ε_t .

Así, tenemos,

$$p \lim \hat{\beta}_{MCO} = \frac{\beta + \rho}{1 + \beta \rho} = \beta + \frac{\rho(1 - \beta^2)}{1 + \beta \rho}$$

que, como puede comprobarse, está entre -1 y 1 .

Por tanto,

$$\text{Sesgo a sin tótico}(\hat{\beta}_{MCO}) = p \lim \hat{\beta}_{MCO} - \beta = \frac{\rho(1 - \beta^2)}{1 + \beta \rho}$$

de modo que si $\rho > 0$, el estimador $\hat{\beta}_{MCO}$ sobreestima a β , subestimando el verdadero valor del parámetro cuando $\rho < 0$.

Consideremos ahora el estimador de ρ :

$$\hat{\rho}_{MCO} = \frac{\frac{1}{T} \sum_3^T \hat{u}_t \hat{u}_{t-1}}{\frac{1}{T} \sum_3^T \hat{u}_{t-1}^2}$$

siendo $\hat{u}_t = y_t - \hat{\beta}_{MCO} \cdot y_{t-1}$.

Para obtener el límite en probabilidad del estimador $\hat{\rho}_{MCO}$ notemos, en primer lugar, que:

$$\begin{aligned} \hat{u}_{t-1}^2 &= (y_{t-1} - \hat{\beta} y_{t-2})^2 = y_{t-1}^2 - 2\hat{\beta} y_{t-1} y_{t-2} + \hat{\beta}^2 y_{t-2}^2 = \\ &= y_{t-1}^2 - 2\hat{\beta} (\hat{\beta} y_{t-2} + \hat{u}_{t-1}) y_{t-2} + \hat{\beta}^2 y_{t-2}^2 \end{aligned}$$

Dividimos por T , sumamos y tomamos límites en probabilidad, obteniendo:

$$\begin{aligned} p \lim \frac{1}{T} \sum_3^T \hat{u}_{t-1}^2 &= p \lim \frac{1}{T} \sum_3^T y_{t-1}^2 - 2p \lim (\hat{\beta}^2) p \lim \left(\frac{1}{T} \sum_3^T y_{t-2}^2 \right) - \\ &\quad - 2p \lim (\hat{\beta}) p \lim \left(\frac{1}{T} \sum_3^T \hat{u}_{t-1} y_{t-2} \right) + p \lim (\hat{\beta}^2) p \lim \left(\frac{1}{T} \sum_3^T y_{t-2}^2 \right) \end{aligned}$$

es decir,

$$p \lim \frac{1}{T} \sum_3^T \hat{u}_{t-1}^2 = \left[1 - (p \lim \hat{\beta})^2 \right] p \lim \frac{1}{T} \sum_3^T y_{t-1}^2 - 2p \lim (\hat{\beta}) p \lim \left(\frac{1}{T} \sum_3^T \hat{u}_{t-1} y_{t-2} \right)$$

pero el último término es igual a $E(\hat{u}_{t-1}y_{t-2})$, que es igual a cero por ser y_{t-2} una variable explicativa en la regresión en la que se generó el residuo de mínimos cuadrados \hat{u}_{t-1} .

En segundo lugar:

$$\begin{aligned}\hat{u}_t\hat{u}_{t-1} &= \left(y_t - \hat{\beta}_{MCO} \cdot y_{t-1}\right) \left(y_{t-1} - \hat{\beta}_{MCO} \cdot y_{t-2}\right) = \\ &= y_t y_{t-1} - \hat{\beta}_{t-1}^2 y_{t-1}^2 - \hat{\beta}_{t-1} y_t y_{t-2} + \hat{\beta}_{t-1}^2 y_{t-1} y_{t-2}\end{aligned}$$

de modo que:

$$\begin{aligned}p \lim \frac{1}{T} \sum_3^T \hat{u}_t \hat{u}_{t-1} &= \left(1 + p \lim \hat{\beta}^2\right) p \lim \frac{1}{T} \sum_3^T y_{t-1} y_{t-2} - \left(p \lim \hat{\beta}\right) p \lim \frac{1}{T} \sum_3^T y_{t-1}^2 \\ &\quad - \left(p \lim \hat{\beta}\right) p \lim \frac{1}{T} \sum_3^T y_t y_{t-2}\end{aligned}\tag{15}$$

Pero, utilizando la representación (14), tenemos:

$$p \lim \frac{1}{T} \sum_3^T y_t y_{t-2} = p \lim \frac{1}{T} \sum_3^T (\beta + \rho) y_{t-1} y_{t-2} - \beta \rho p \lim \frac{1}{T} \sum_3^T y_{t-2}^2 + p \lim \frac{1}{T} \sum_3^T \varepsilon_t y_{t-2}\tag{16}$$

donde el último término converge a $E(\varepsilon_t y_{t-2})$, que es igual a cero porque y_{t-2} depende de ε_{t-2} y anteriores, pero no de ε_t .

Sustituyendo (16) en (15) tenemos:

$$\begin{aligned}p \lim \frac{1}{T} \sum_3^T \hat{u}_t \hat{u}_{t-1} &= \left[1 + \left(p \lim \hat{\beta}\right)^2 - \left(p \lim \hat{\beta}\right) (\beta + \rho)\right] p \lim \frac{1}{T} \sum_3^T y_t y_{t-1} - p \lim \hat{\beta} \left(p \lim \frac{1}{T} \sum_3^T y_{t-1}^2\right) + \\ &= \left[1 + \left(p \lim \hat{\beta}\right)^2 - \left(p \lim \hat{\beta}\right) (\beta + \rho)\right] \left(p \lim \hat{\beta}\right) p \lim \frac{1}{T} \sum_3^T y_{t-1}^2 - p \lim \hat{\beta} \left(p \lim \frac{1}{T} \sum_3^T y_{t-1}^2\right) + \\ &= \left[\left(p \lim \hat{\beta}\right)^3 - \left(p \lim \hat{\beta}\right)^2 (\beta + \rho) + \beta \rho \left(p \lim \hat{\beta}\right)\right] p \lim \frac{1}{T} \sum_3^T y^2\end{aligned}$$

de modo que:

$$p \lim (\hat{\rho}_{MCO}) = \frac{\left(p \lim \hat{\beta}\right)^3 - \left(p \lim \hat{\beta}\right)^2 (\beta + \rho) + \beta \rho \left(p \lim \hat{\beta}\right)}{1 - \left(p \lim \hat{\beta}\right)^2}$$

Proposition 1 $p \lim (\hat{\rho}_{MCO}) = \beta \rho \frac{\beta + \rho}{1 + \beta \rho}$

Demostración:

$$\begin{aligned}
p \lim (\hat{\rho}_{MCO}) &= \frac{1}{1 - \left(\frac{\beta + \rho}{1 + \beta\rho}\right)^2} \frac{\beta + \rho}{1 + \beta\rho} \left[\left(\frac{\beta + \rho}{1 + \beta\rho}\right)^2 - \frac{(\beta + \rho)^2}{1 + \beta\rho} + \beta\rho \right] = \\
&= \frac{1}{1 - \left(\frac{\beta + \rho}{1 + \beta\rho}\right)^2} \frac{\beta + \rho}{1 + \beta\rho} \left[\left(\frac{\beta + \rho}{1 + \beta\rho}\right)^2 - 1 \right] (-\beta\rho) = \beta\rho \frac{\beta + \rho}{1 + \beta\rho}
\end{aligned}$$

Corollary 2 *En consecuencia, tenemos: Sesgo a sin tótico ($\hat{\rho}_{MCO}$) = $p \lim \hat{\rho}_{MCO} - \rho = -\rho \frac{1 - \beta^2}{1 + \beta\rho}$*

Corollary 3 *$p \lim (\hat{\beta}_{MCO} + \hat{\rho}_{MCO}) = \beta + \rho$*

Como muestra el último corolario, la suma de ambos parámetros se estima consistentemente, a pesar del sesgo asintótico en que se incurra al estimar por MCO cada uno de ellos. Una vez más, hay que observar que, aunque el problema de correlación no nula entre variables explicativas y término de error se produce tan sólo en la primera ecuación del modelo, que tiene a y_t por variable dependiente, ambos parámetros se estiman inconsistentemente.

Hay que notar también que la estimación $\hat{\rho}_{MCO}$ puede subestimar de modo apreciable la autocorrelación existente en u_t . Además, no es posible evaluar dicha autocorrelación únicamente a partir de $\hat{\rho}_{MCO}$ pues dicha estimación numérica puede estar muy sesgada hacia cero. Como consecuencia, el estadístico Durbin-Watson está sesgado asintóticamente a la baja cuando $\rho > 0$.

Proposition 4 *Sesgo a sin tótico ($\hat{\alpha}_{MCO}$) = $p \lim \hat{\alpha}_{MCO} - \alpha = -\alpha\rho \frac{1 + \beta}{1 + \beta\rho}$*

Demostración:

$$\begin{aligned}
\text{Sesgo a sin tótico } (\hat{\alpha}_{MCO}) &= p \lim \hat{\alpha}_{MCO} - \alpha = p \lim \left[\left(1 - \hat{\beta}_{MCO}\right) \bar{y} \right] - \alpha = \left(1 - p \lim \hat{\beta}_{MCO}\right) E y - \alpha = \\
&= \left(1 - \beta - \rho \frac{1 - \beta^2}{1 + \beta\rho}\right) \frac{\alpha}{1 - \beta} - \alpha = \left(1 - \rho \frac{1 + \beta}{1 + \beta\rho} - 1\right) \alpha = -\rho\alpha \frac{1 + \beta}{1 + \beta\rho}
\end{aligned}$$