Regresión con series temporales

Miguel Jerez

Universidad Complutense de Madrid

Septiembre 2014

Índice

- Introducción
- Correlación espuria
- Transformaciones de datos
- Cointegración
- Modelos deterministas
- Autocorrelación
- Ideas principales

Introducción (I): Series temporales y datos de sección cruzada

La característica distintiva de las series temporales, en comparación con datos de corte transversal, es que la muestra tiene un orden natural: del pasado al futuro. Esto determina qué relaciones son posibles y cuáles no:

Relación entre edad y estatura de un conjunto de niños:

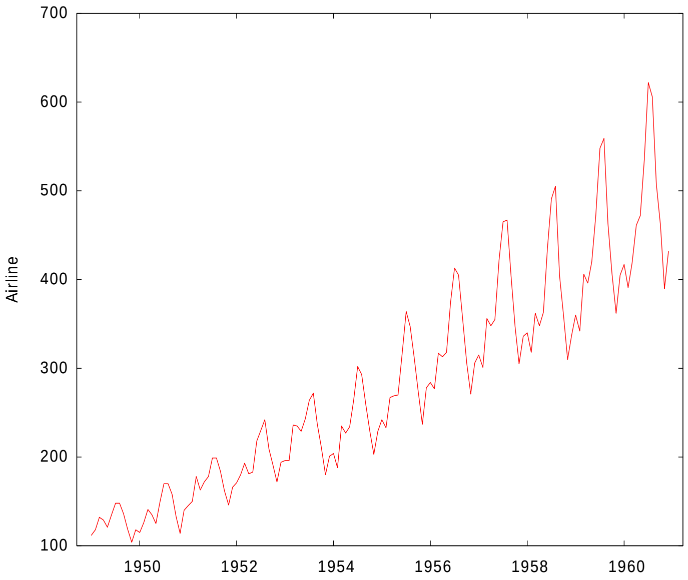
- El momento en que se miden las observaciones es irrelevante
- La muestra puede ordenarse de distintas formas
- La edad o la estatura de un individuo no dependen de los valores correspondientes de los individuos adyacentes
- La dirección de causalidad se conoce en este caso (la edad causa la estatura) pero no hay un procedimiento empírico para determinarla

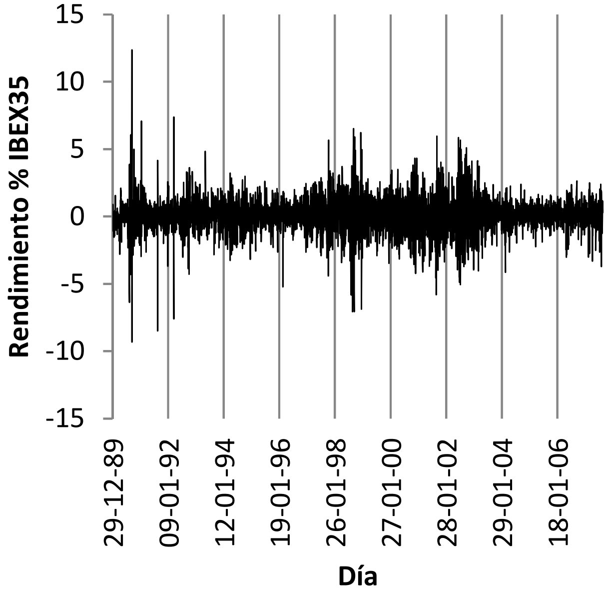
Relación entre ventas y publicidad de una empresa:

- Los datos están "fechados"
- El orden natural de los datos es el cronológico
- Es razonable que las ventas de hoy dependan de las ventas de ayer
- Es razonable que las ventas de hoy dependan de la publicidad de hoy y de la de ayer
- Existe un fundamento objetivo para la causalidad: X causa a Y si el pasado de X ayuda a predecir (o está correlado con) el presente de Y

Introducción (II): Características de las series económicas

Asimismo, el orden de la muestra da lugar a características específicas de las series temporales como la tendencia, la estacionalidad o el “volatility clustering”:





Las series económicas de baja frecuencia (mensuales, trimestrales…) muestran: (a) tendencia, (b) estacionalidad y (c) una variabilidad que crece con el nivel de la serie.

Ejemplo: Pasajeros de líneas aéreas internacionales; dataset: Airline

Los rendimientos financieros observados con alta frecuencia (por horas, días…) no son estacionales y suelen tener: (a) una media estable y (b) períodos de alta y baja volatilidad

Ejemplo: Rendimientos diarios del índice IBEX35; dataset: CAPM

Introducción (III): Agenda

Nos centraremos en las series del primer tipo

Teniendo en cuenta sus características habituales (tendencia, estacionalidad ...) los puntos principales de nuestra agenda serán:

1. Comprender el riesgo de crear una correlación espuria, es decir, una aparente relación entre dos variables que no tienen conexión lógica

2. Aprender a evitar este riesgo mediante un tratamiento adecuado de las variables del modelo. Hay tres enfoques principales:

   - La transformación de los datos, para estabilizar su media y su varianza
   - Añadir variables explicativas cointegradas con la variable endógena
   - Añadir variables explicativas deterministas para capturar la tendencia y estacionalidad

3. Aprender a detectar y modelizar posibles autocorrelaciones residuales

Índice

- Introducción
- Correlación espuria
- Transformaciones de datos
- Cointegración
- Modelos deterministas
- Autocorrelación
- Ideas principales

Correlación espuria (I): Concepto

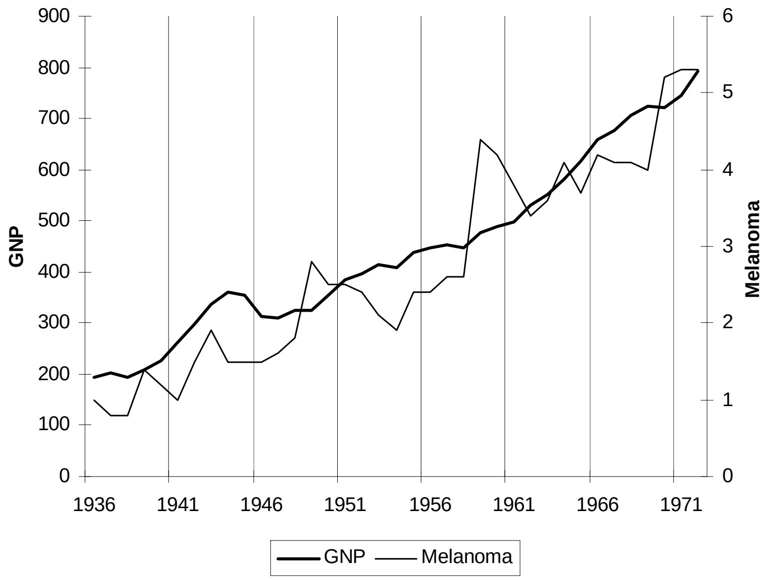
- Una correlación espuria es una relación empírica entre dos acontecimientos sin conexión lógica
- Las correlaciones espurias pueden producirse con datos de corte transversal o series temporales
  - **Ejemplo**: En 1952 J. Neyman analizó la relación entre la tasa de nacimientos y la población de cigüeñas en varias regiones, encontrando una elevada correlación entre ambas variables
  - **Ejemplo**: Utilizando datos anuales para el período 1866-1911, G. Udny Yule encontró que el coeficiente de correlación entre la tasa de mortalidad en Inglaterra-Gales y el porcentaje de matrimonios en la iglesia de Inglaterra era de 0.95
- En series temporales las correlaciones espurias son frecuentes, simplemente porque muchas series tienen tendencia
- La pregunta es: ¿cómo podemos distinguir las relaciones reales de las espurias?
- La respuesta será: una relación espuria es la que desaparece al darle a la tendencia un tratamiento adecuado

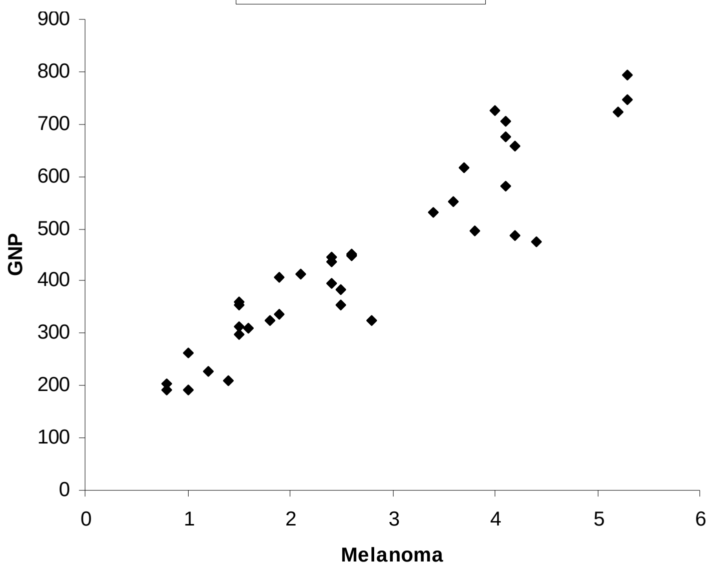
Correlación espuria (II.a): Ejemplo

Las cifras muestran la serie anual (1936-1972) de:

- EE.UU. PNB en miles de millones de US$ corrientes, y
- La incidencia de melanoma en la población masculina de Connecticut (datos ajustados para la edad)

Ambas series tienen aparentemente una fuerte relación lineal, pero relacionarlas es obviamente absurdo





Correlación espuria (II.b): Ejemplo

El primer cuadro muestra los resultados de dos regresiones de la incidencia del melanoma sobre GNP. Como puede verse:

- El R^2 es muy elevado
- Todos los parámetros estimados son significativos, y
- Implican que un caso adicional de melanoma aumenta el PNB en 119 billones of US$

Si estimamos una regresión en primeras diferencias

$$\nabla GNP_t = GNP_t - GNP_{t-1}$$

$$\nabla Melanoma_t = Melanoma_t - Melanoma_{t-1}$$

…la relación espuria desaparece, ya que estaba causada por la tendencia de las series

```
Model 1: OLS, using observations 1936-1972 (T = 37)
Dependent variable: GDP

              coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const         118.566       23.7290      4.997    1.62e-05  ***
  Melanoma      118.981        7.81415    15.23     5.22e-017 ***

Mean dependent var   443.6730   S.D. dependent var   171.4417
Sum squared resid    138787.6   S.E. of regression   62.97110
R-squared            0.868836   Adjusted R-squared   0.865088
F(1, 35)             231.8413   P-value(F)           5.22e-17
Log-likelihood      -204.7517   Akaike criterion     413.5034
Schwarz criterion    416.7252   Hannan-Quinn         414.6392
rho                  0.554021   Durbin-Watson        0.879122

Model 2: OLS, using observations 1937-1972 (T = 36)
Dependent variable: d_GDP

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const         16.5684       3.17933     5.211     9.14e-06 ***
  d_Melanoma     0.706295     6.58576     0.1072    0.9152

Mean dependent var   16.65278   S.D. dependent var   18.22001
Sum squared resid    11614.98   S.E. of regression   18.48289
R-squared            0.000338   Adjusted R-squared  -0.029064
F(1, 34)             0.011502   P-value(F)           0.915224
Log-likelihood      -155.0594   Akaike criterion     314.1187
Schwarz criterion    317.2858   Hannan-Quinn         315.2241
rho                  0.356257   Durbin-Watson        1.262415
```

Índice

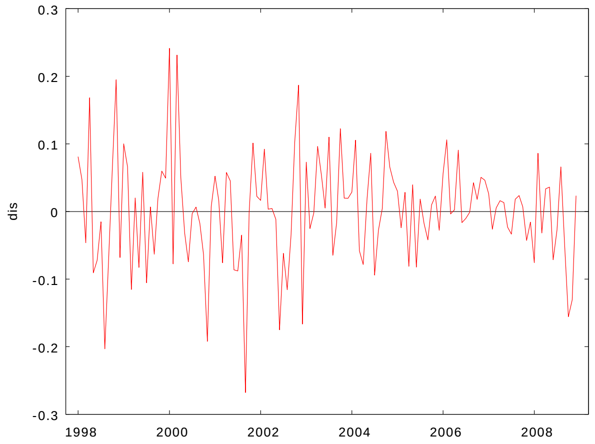
- Introducción
- Correlación espuria
- Transformaciones de datos
- Cointegración
- Modelos deterministas
- Autocorrelación
- Ideas principales

Transformaciones de datos (I.a): Estacionariedad

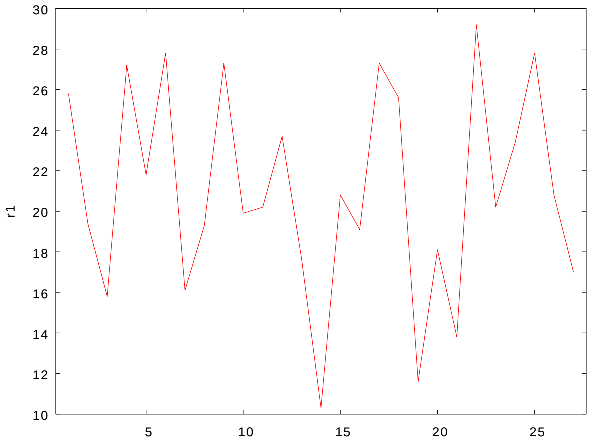
- Un enfoque habitual para estabilizar los momentos de una serie temporal consiste en transformar los datos para estabilizar su media y su varianza
- Una serie temporal con una media constante se llama “estacionaria en la media” o simplemente “estacionaria”
- Una serie temporal es estacionaria en varianza cuando es estacionaria en media y, además, su evolución presenta una dispersión constante alrededor de su nivel medio.
- En este apartado presentaremos:
  - Una transformación de datos que permite estabilizar la varianza (a veces)
  - Una familia de transformaciones que permiten estabilizar los cambios de media debidos a la tendencia o a la estacionalidad
  - La interpretación de las variables que resultan de las combinaciones más recuentes de las transformaciones previas

Transformaciones de datos (I.b): Ejemplos

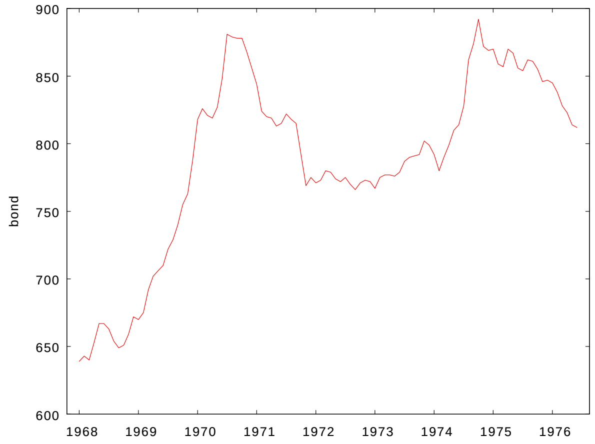
Retorno mensual de las acciones de Disney
Dataset: POE-CAPM4



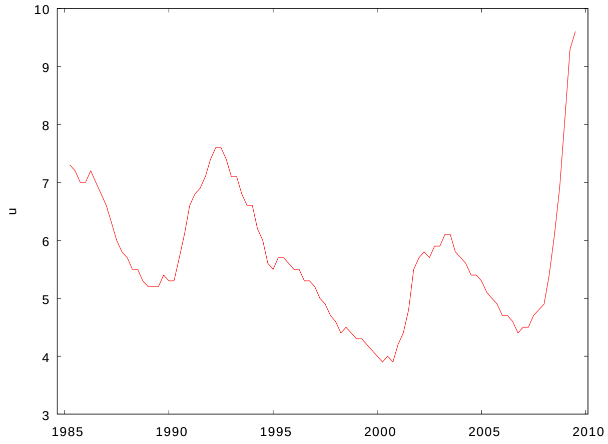
Niveles anuales de lluvia en Midwest
Dataset: POE-cattle



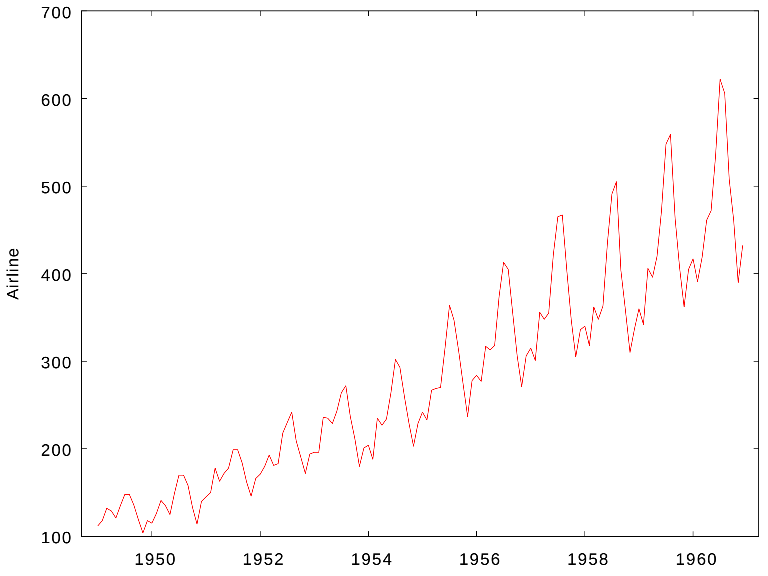
Retorno mensual de los bonos ferroviarios AA
Dataset: POE-bond

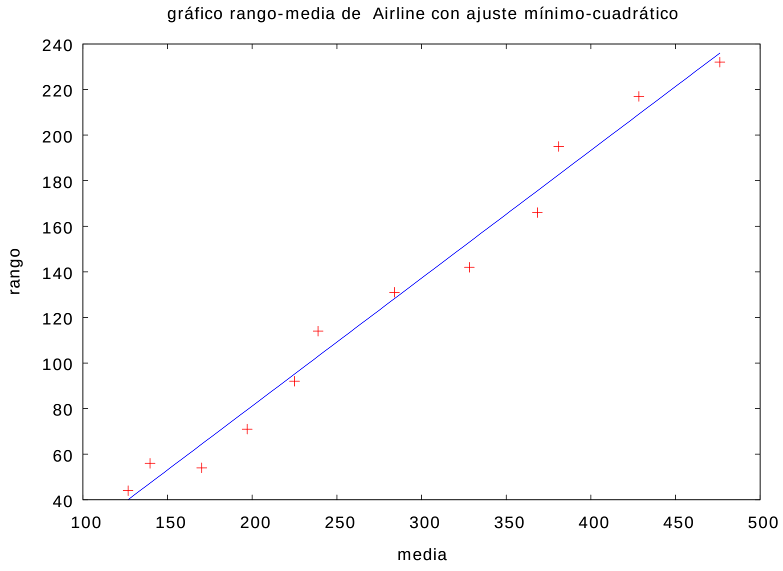


Desempleo civil desestacionalizado en EE.UU.
Dataset: POE-okun



Transformaciones de datos (II.a): Estabilización de la varianza



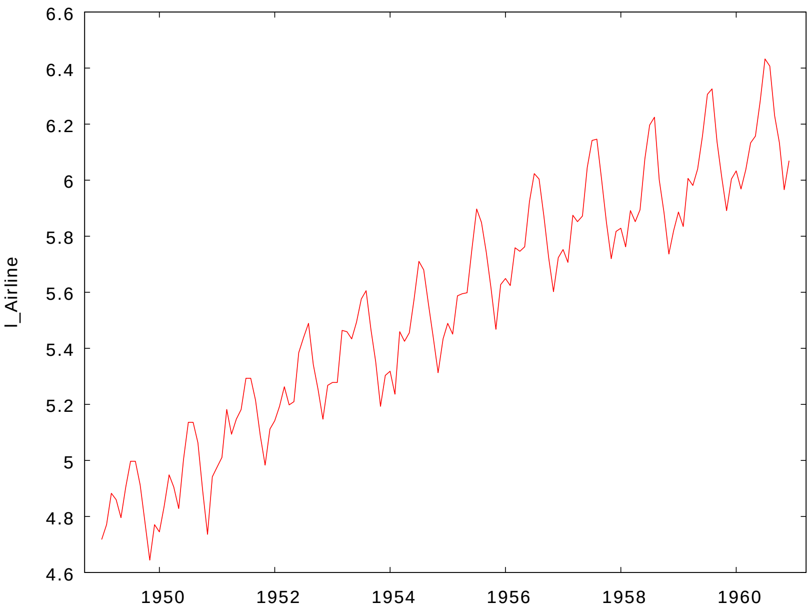


Las series económicas observadas con baja frecuencia a menudo muestran una variabilidad que crece con el nivel de la serie

Esta característica puede detectarse:

- En el perfil de la serie temporal, o bien
- …como una relación aproximadamente lineal entre
  - una medida de nivel local (la media o la mediana calculadas por submuestras) y
  - una medida de dispersión local (p. ej. el rango o la desviación típica calculadas para las mismas submuestras)

Transformaciones de datos (II.b): Estabilización de la varianza



Este tipo de heterocedasticidad a menudo se corrige aplicando a los datos una transformación logarítmica:

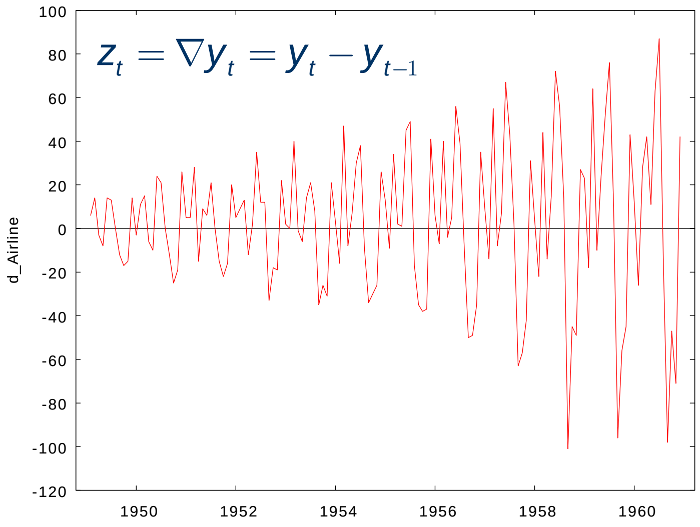
$$z_t = \ln y_t$$

La transformación log tiene otras ventajas, ya que ayuda a inducir en los datos normalidad y linealidad en las relaciones

Estas propiedades se deben a la forma de la función log que:

- Aumenta la simetría de las distribuciones de valores positivos
- Reduce las diferencias de escala de la variable transformada

Transformaciones de datos (III): Estabilización de la media

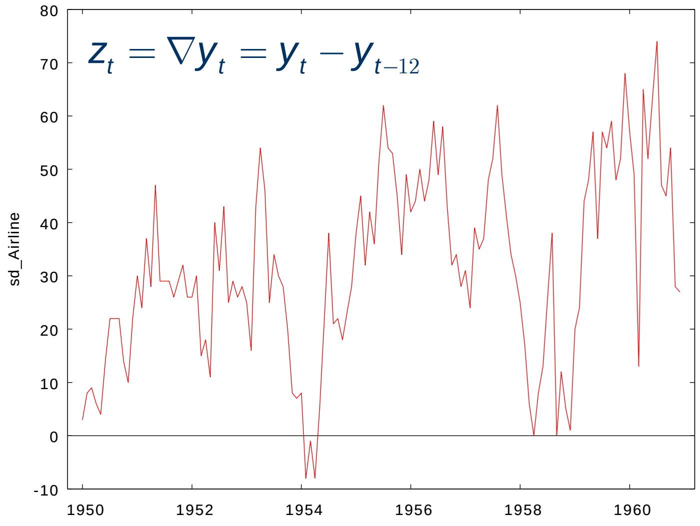


La estacionariedad en media se consigue a menudo diferenciando las series, por ejemplo, mediante una diferencia regular:

$$z_t = y_t - y_{t-1} = \nabla y_t$$

…o una diferencia estacional:

$$z_t = y_t - y_{t-s} = \nabla_s y_t$$



Para conseguir estabilidad en la media uy en la varianza a menudo necesitaremos combinar la transformación log con la diferenciación:

$$z_t = \ln y_t - \ln y_{t-1} = \nabla \ln y_t$$

Como veremos, la variable resultante es una tasa logarítmica (o “tasa log”) de variación, alternativa a la tasa porcentual

Transformaciones de datos (IV): La tasa log

La tasa log de variación se define como:

$$z_t = \ln y_t - \ln y_{t-1} = \nabla \ln y_t \quad \text{or} \quad z_t = 100 \times \left(\ln y_t - \ln y_{t-1} \right) = 100 \times \nabla \ln y_t$$

…y se puede interpretar como una aproximación a la tasa porcentual porque:

$$z_t = \ln y_t - \ln y_{t-1} = \ln \left(\frac{y_t}{y_{t-1}} \right) = \ln \left(1 + \frac{\pi_t}{100} \right)$$

…en donde π_t es la tasa de variación porcentual de y_t

Ejemplo

t	y_t	Inc. %	$\ln y_t$	$(\ln y_t - \ln y_{t-1}) \times 100$
1	100.00	--	4.605170	--
2	101.00	1.00%	4.615121	0.995033
3	103.02	2.00%	4.634923	1.980263

Transformaciones de datos (III): Ejemplos

Transformación	Observaciones
$z_t = \ln y_t$	Logaritmo neperiano de y_t. A veces independiza la volatilidad del nivel e induce normalidad y linealidad
$z_t = y_t - y_{t-1} = \nabla y_t$	Cambio en y_t. Es un indicador de crecimiento absoluto
$z_t = \ln y_t - \ln y_{t-1} = \nabla \ln y_t$	Tasa log de crecimiento. En un indicador de crecimiento relativo, en términos porcentuales si se multiplica por 100
$w_t = \ln y_t - \ln y_{t-1} = \nabla \ln y_t$ $z_t = w_t - w_{t-1} = \nabla w_t = \nabla^2 \ln y_t$	Cambio en la tasa log de crecimiento. Es un indicador de “aceleración” en el crecimiento relativo
$z_t = \ln y_t - \ln y_{t-S} = \nabla_S \ln y_t$	Tasa de crecimiento acumulada en una estación (*S* períodos). Cuando el período estacional es de un año, se conoce como “tasa anual” o “tasa interanual”
$w_t = \ln y_t - \ln y_{t-S} = \nabla_S \ln y_t$ $z_t = w_t - w_{t-1} = \nabla \nabla_S \ln y_t$	Cambio en la tasa de crecimiento acumulada en una estación. Es un indicador de aceleración en el crecimiento acumulado en un ciclo estacional

Índice

- Introducción
- Correlación espuria
- Transformaciones de datos
- Cointegración
- Modelos deterministas
- Autocorrelación
- Ideas principales

Cointegración (I): Concepto

- Una serie temporal es integrada de orden *d*, o I(*d*), si *d* es el mínimo número de diferencias requeridas para inducir estacionariedad en media
- A menudo, un conjunto de series temporales son individualmente I(*d*), pero existe una combinación lineal que está integrada de orden inferior, I(*d*-*m*). En este caso se dice que la serie están cointegradas
- Un procedimiento simple para estimar la relación de cointegración consiste en ajustar una regresión lineal para las variables potencialmente cointegrados, y luego evaluar la estacionariedad de los residuos
- La situación más habitual es que las dos series sean I(1) y su diferencia (exacta o aproximada) sea I(0)
- La cointegración implica que:
  - ...existe un componente de tendencia común a ambas series, por lo que una combinación ponderada adecuadamente de ellos cancela este componente
  - ...existe un equilibrio a largo plazo entre ambas series, por lo que las desviaciones del equilibrio tienden a desaparecer en plazos relativamente cortos

Cointegración (II.a): Ejemplo

La figura muestra el perfil del:

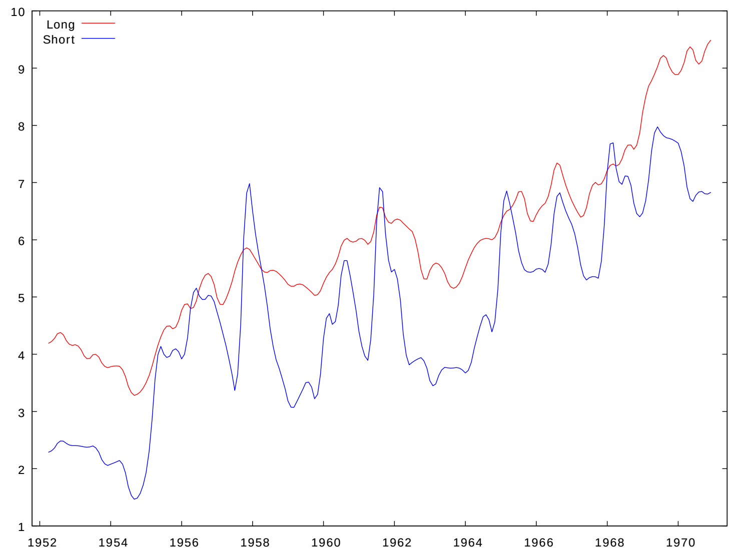
- ...rendimiento porcentual a 20 años de los bonos soberanos del Reino Unido (Long) por ciento y
- ...el rendimiento de las letras del tesoro a 91-días (Short),

...entre el segundo trimestre de 1952-segundo trimestre y el cuarto de 1979

Ambas series son no estacionarias

Para probar si la relación es espuria, ajustamos una regresión entre las primeras diferencias

El coeficiente de regresión es muy significativo, por lo que la correlación no es espuria



Modelo 2: MCO, usando las observaciones 1952:3-1970:4 (T = 74)
Variable dependiente: d_Short

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	-0.0273645	0.0773056	-0.3540	0.7244	
d_Long	1.26015	0.280995	4.485	2.70e-05	***

Media de la vble. dep.	0.060676	D.T. de la vble. dep.	0.722515
Suma de cuad. residuos	29.78754	D.T. de la regresión	0.643207
R-cuadrado	0.218340	R-cuadrado corregido	0.207484
F(1, 72)	20.11170	Valor p (de F)	0.000027
Log-verosimilitud	-71.33238	Criterio de Akaike	146.6648
Criterio de Schwarz	151.2729	Crit. de Hannan-Quinn	148.5030
rho	-0.085839	Durbin-Watson	2.166905

Cointegración (II.b): Ejemplo

El listado muestra los resultados de una regresión entre el nivel de ambas variables

Esta relación se acerca a un "spread" o "prima de plazo"

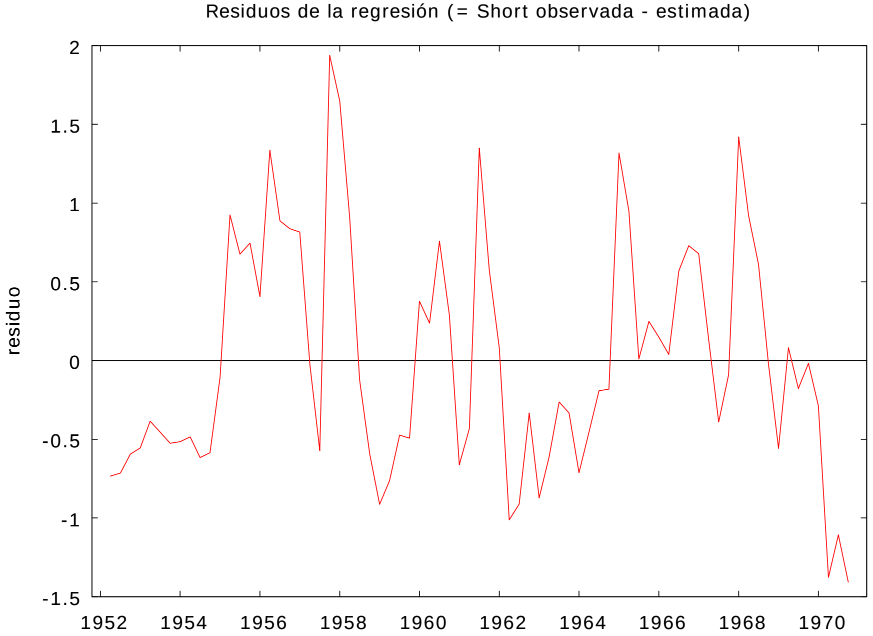
Los residuos de este modelo parecen estacionarios, por lo que las series podrían estar cointegradas

Dos series cointegradas: (a) comparten una tendencia común y (b) mantienen una relación de equilibrio a largo plazo

```
Modelo 1: MCO, usando las observaciones 1952:2-1970:4 (T = 75)
Variable dependiente: Short

             Coeficiente   Desv. Típica   Estadístico t    Valor p
  ---------------------------------------------------------------
  const       -1.16917      0.350071          -3.340       0.0013    ***
  Long         0.998553     0.0573798         17.40        1.09e-027 ***

Media de la vble. dep.   4.738000   D.T. de la vble. dep.   1.670707
Suma de cuad. residuos  40.11837    D.T. de la regresión    0.741328
R-cuadrado               0.805772   R-cuadrado corregido    0.803112
F(1, 73)               302.8478     Valor p (de F)          1.09e-27
Log-verosimilitud      -82.95837    Criterio de Akaike    169.9167
Criterio de Schwarz    174.5517     Crit. de Hannan-Quinn 171.7674
rho                      0.623582   Durbin-Watson           0.751581
```



Índice

- Introducción
- Correlación espuria
- Transformaciones de datos
- Cointegración
- Modelos deterministas
- Autocorrelación
- Ideas principales

Modelos deterministas (I): Descomposición estructural

Ahora discutiremos cómo capturar la tendencia y estructura de temporada utilizando variables deterministas

El punto de partida será la "descomposición estructural" de una serie histórica:

$$y_t = T_t + S_t + C_t + \varepsilon_t \quad \left(t = 1, 2, \ldots, n\right)$$

…en donde los sumandos en el lado derecho de la igualdad son:

Tendencia (T_t), que capta los cambios persistentes en la media,

Componente estacional (S_t), que capta las fluctuaciones periódicas en la media.

Componente irregular (ε_t), con las propiedades habituales de un término de error: media cero, homoscedástico y no autocorrelado

Ciclo (C_t), que incluye todo lo que no fue captado por los componentes anteriores

Asimismo, se puede considerar un componente de "efectos exógenos," si hay variables explicativas que afectan a la endógena (por ejemplo, efectos calendario)

Por último conviene tener en cuenta que, si la serie se ha transformado logarítmicamente, la descomposición de la variable original será multiplicativa:

$$e^{y_t} = e^{T_t} \cdot e^{S_t} \cdot e^{C_t} \cdot e^{\varepsilon_t}$$

Modelos deterministas (II): Tendencia

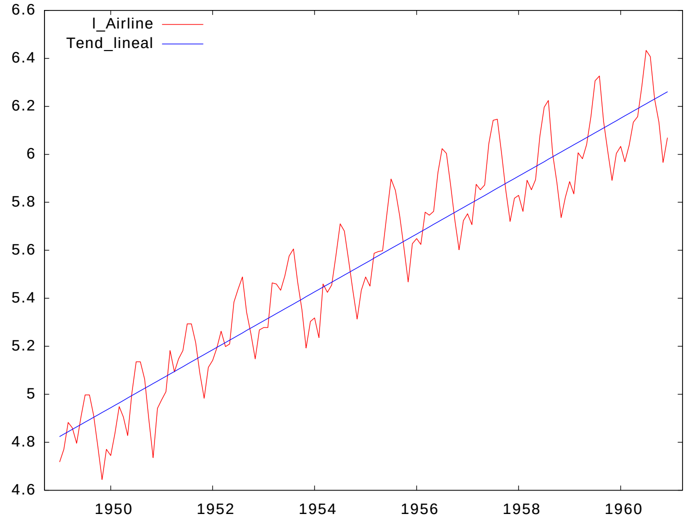
Los modelos más simples para la tendencia son regresiones de la variable endógena sobre una variable determinista “tiempo”, tales como:

- Tendencia lineal

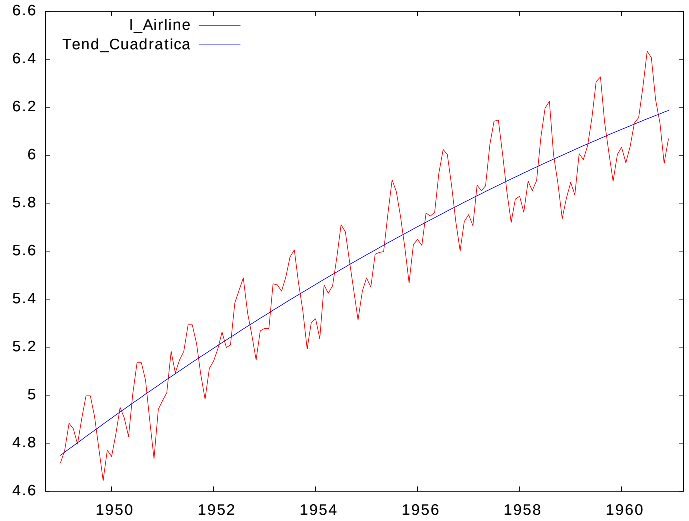
$$T_t = \alpha_0 + \alpha_1 t \quad \left(t = 1, 2, \ldots, n\right)$$

- Tendencia cuadrática

$$T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 \quad \left(t = 1, 2, \ldots, n\right)$$



Las figuras muestran el ajuste de tendencias lineales y cuadráticas al (log) de pasajeros de líneas aéreas



Estos modelos son útiles para descomponer una serie pero suelen prever bastante mal a corto plazo, ya que le dan el mismo peso a todas las observaciones y, por tanto, resultan poco “adaptativos”

Modelos deterministas (III): Estacionalidad

La estacionalidad es un cambio en la media que recorre un ciclo a lo largo de un número fijo de estaciones *S*. Por ejemplo, para datos trimestrales o mensuales, este número sería $S = 4$ y $S = 12$

Un modelo determinista simple para la estacionalidad es una regresión de la variable endógena cobre variables 0-1 estacionales

La Figura muestra el resultado de ajustar un modelo de tendencia cuadrática más estacionalidad para el (log) de pasajeros de líneas aéreas:

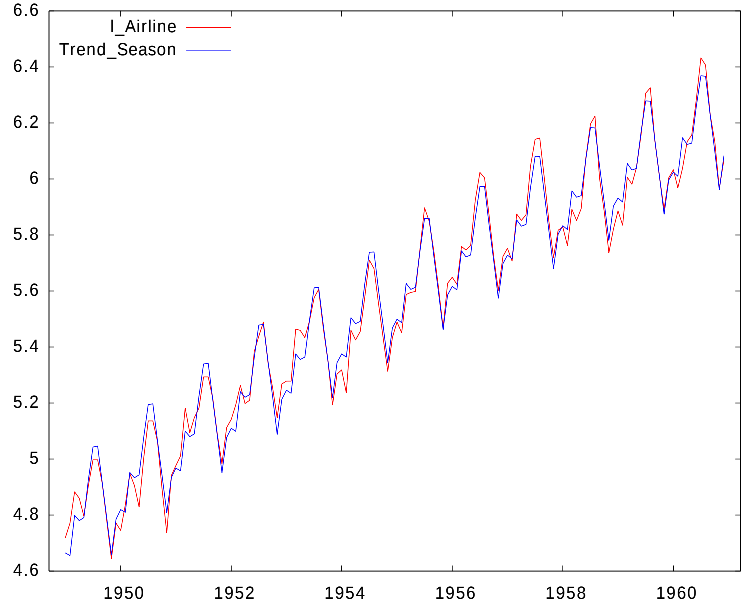
$$\ln(A_t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta_1 S_t^1 + \beta_2 S_t^2 + \ldots \beta_{11} S_t^{11} + \varepsilon_t$$

```
Model 1: OLS, using observations 1949:01-1960:12 (T = 144)
Dependent variable: l_Airline

             coefficient    std. error    t-ratio     p-value
  -----------------------------------------------------------
  const        4.65138      0.0178617     260.4      1.45e-178 ***
  time         0.0131837    0.000389158    33.88     2.27e-066 ***
  sq_time     -2.14819e-05  2.59920e-06    -8.265    1.41e-013 ***
  dm2         -0.0222696    0.0196778      -1.132    0.2598
  dm3          0.107786     0.0196786       5.477    2.15e-07  ***
  dm4          0.0763879    0.0196798       3.882    0.0002    ***
  dm5          0.0739293    0.0196815       3.756    0.0003    ***
  dm6          0.196033     0.0196837       9.959    1.04e-017 ***
  dm7          0.299975     0.0196863      15.24     1.03e-030 ***
  dm8          0.290723     0.0196894      14.77     1.40e-029 ***
  dm9          0.146174     0.0196930       7.423    1.33e-011 ***
  dm10         0.00814498   0.0196970       0.4135   0.6799
  dm11        -0.135401     0.0197015      -6.873    2.36e-010 ***
  dm12        -0.0213211    0.0197065      -1.082    0.2813

Mean dependent var   5.542176   S.D. dependent var   0.441456
Sum squared resid    0.302022   S.E. of regression   0.048200
R-squared            0.989163   Adjusted R-squared   0.988079
Log-likelihood       239.7018   Akaike criterion    -451.4036
Schwarz criterion   -409.8262   Hannan-Quinn        -434.5089
rho                  0.671344   Durbin-Watson        0.647915

Excluding the constant, p-value was highest for variable 16 (dm10)
```



Modelos deterministas (IV): Extracción de señal

- La descomposición de series temporales puede realizarse aplicando unos procedimientos más sofisticados que se denominan "…de extracción de señal":
  - En EE.UU. el procedimiento oficial de descomposición y ajuste estacional se llama **X-13-ARIMA** (http://www.census.gov/srd/www/x13as/)
  - En Europa se usa un procedimiento alternativo: **Tramo-Seats** (www.bde.es/bde/en/secciones/servicios/Profesionales/Programas_estadi/Programas_estad_d9fa7f3710fd821.html)
- En el menú "Filtros", Gretl ofrece muchos otros algoritmos de descomposición / extracción de señal.
- Por su uso en Macroeconomía, cabe destacar el filtro de: Hodrick y Prescott

Índice

- Introducción
- Correlación espuria
- Transformaciones de datos
- Cointegración
- Modelos deterministas
- Autocorrelación
- Ideas principales

Autocorrelación (I): El enfoque clásico

- La Econometría clásica considera que la autocorrelación es un problema del modelo de regresión, análogo al de heterocedasticidad:
  - La autocorrelación de errores no afecta a la insesgadez y consistencia de MCO, pero se pierden la eficiencia y los resultados estándar de inferencia
  - Consecuentemente, el enfoque clásico se centran en detectar el problema y, si es necesario, en recuperar eficiencia o ajustar los contrastes de hipótesis
- Un test clásico de autocorrelación es el de Breusch-Godfrey, que contrasta la nula de no autocorrelación frente a la alternativa de que los errores siguen un proceso de orden autorregresivo de orden *p*:

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \ldots + \phi_p \varepsilon_{t-p} + a_t$$

- El estadístico de Breusch-Godfrey se basa en una regresión auxiliar de los residuos MCO sobre las variables explicativas del modelo y *p* retardos de los residuos:

$$\hat{\varepsilon}_t = \boldsymbol{x}_t^T \boldsymbol{\delta} + \phi_1 \hat{\varepsilon}_{t-1} + \phi_2 \hat{\varepsilon}_{t-2} + \ldots + \phi_p \hat{\varepsilon}_{t-p} + a_t$$

- El R^2 de esta regresión mide la influencia de los residuos pasados en el actual. La hipótesis de que esta influencia no existe, puede contrastarse con:

$$BG = nR^2 \underset{H_0}{\sim} \chi_p^2$$

Autocorrelación (II): Errores estándar robustificados

Un tratamiento simple de autocorrelación consiste en usar errores estándar “robustificados” para hacer inferencia. Se trata de un enfoque análogo al que usamos en el tema de Heterocedasticidad

El listado muestra algunos resultados de una regresión CAPM de los retornos diarios del IBEX35 sobre las de las acciones de BBVA (CAPM conjunto de datos)

Las estimaciones de los errores estándar son consistentes con errores heterocedásticos y autocorrelados (HAC), como los inicialmente propuestos por Newey y West

Cuando la muestra es grande, esta es una buena manera sin recurrir a estimadores complejos (MCG).

```
Model 1: OLS, using observations 1-4409
Dependent variable: BBVA
HAC standard errors, bandwidth 12 (Bartlett kernel)

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       0.0264451    0.0155302     1.703     0.0887    *
  IBEX_35     1.13034      0.0539466    20.95      5.27e-093 ***

Mean dependent var    0.076712   S.D. dependent var   1.878959
Sum squared resid     5889.964   S.E. of regression   1.156072
R-squared             0.621526   Adjusted R-squared   0.621440
F(1, 4407)            439.0285   P-value(F)           5.27e-93
Log-likelihood       -6894.527   Akaike criterion     13793.05
Schwarz criterion     13805.84   Hannan-Quinn         13797.56
rho                  -0.029938   Durbin-Watson        2.059492

Breusch-Godfrey test for autocorrelation up to order 10
OLS, using observations 1-4409
Dependent variable: uhat

             coefficient    std. error    t-ratio    p-value
  ----------------------------------------------------------
  const      -9.54218e-05   0.0173669    -0.005494   0.9956
  IBEX_35     0.00176775    0.0132680     0.1332     0.8940
  uhat_1     -0.0329541     0.0150826    -2.185      0.0289   **
  uhat_2     -0.0307853     0.0150850    -2.041      0.0413   **
  uhat_3     -0.0201082     0.0150943    -1.332      0.1829
  uhat_4     -0.0274434     0.0150613    -1.822      0.0685   *
  uhat_5     -0.0262328     0.0150706    -1.741      0.0818   *
  uhat_6      6.14750e-05   0.0150721     0.004079   0.9967
  uhat_7     -0.0651592     0.0150613    -4.326      1.55e-05 ***
  uhat_8     -0.0106493     0.0150913    -0.7057     0.4804
  uhat_9     -0.00229477    0.0150848    -0.1521     0.8791
  uhat_10     0.0282851     0.0150776     1.876      0.0607   *

  Unadjusted R-squared = 0.008425

Test statistic: LMF = 3.736112,
with p-value = P(F(10,4397) > 3.73611) = 5.15e-005

Alternative statistic: TR^2 = 37.147443,
with p-value = P(Chi-square(10) > 37.1474) = 5.34e-005
```

Autocorrelación (III): Modelo del error

La Econometría moderna contempla los errores autocorrelados como un elemento más del modelo: el "modelo del error"

El listado muestra un nuevo modelo de la base de datos de líneas aéreas:

$$\ln(A_t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

$$+ \beta_1 S_t^1 + \beta_2 S_t^2 + \ldots \beta_{11} S_t^{11} + \varepsilon_t$$

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + a_t$$

…que añade a los modelos de la tendencia y la estacionalidad una autorregresión de primer orden para los errores

El comportamiento sistemático descrito por este término puede llamarse "ciclo"

El test de Breusch-Godfrey para los residuos $\hat{a}_t$ no rechaza la nula de ausencia de autocorrelación

```
Model 1: ARMAX, using observations 1949:01-1960:12 (T = 144)
Dependent variable: l_Airline
Standard errors based on Hessian

             coefficient    std. error       z         p-value
  ------------------------------------------------------------
  const        4.65600      0.0194456     239.4       0.0000    ***
  phi_1        0.674472     0.0612648      11.01      3.45e-028 ***
  time         0.0130315    0.000224641    58.01      0.0000    ***
  sq_time     -2.07330e-05  1.30075e-06   -15.94      3.38e-057 ***
  dm2         -0.0218856    0.0106061      -2.064     0.0391    **
  dm3          0.108458     0.0136577       7.941     2.00e-015 ***
  dm4          0.0772910    0.0153130       5.047     4.48e-07  ***
  dm5          0.0750337    0.0162554       4.616     3.91e-06  ***
  dm6          0.197333     0.0167514      11.78      4.94e-032 ***
  dm7          0.301491     0.0169183      17.82      4.91e-071 ***
  dm8          0.292500     0.0167914      17.42      5.86e-068 ***
  dm9          0.148292     0.0163402       9.075     1.13e-019 ***
  dm10         0.0107308    0.0154529       0.6944    0.4874
  dm11        -0.132152     0.0138794      -9.521     1.71e-021 ***
  dm12        -0.0171165    0.0109681      -1.561     0.1186

Mean of innovations -0.000232  S.D. of innovations  0.033792
Log-likelihood       283.1753  Akaike criterion    -534.3506
Schwarz criterion   -486.8336  Hannan-Quinn        -515.0424
```

```
 Breusch-Godfrey test for autocorrelation up to order 12
 OLS, using observations 1949:01-1960:12 (T = 144)
 Dependent variable: uhat

             coefficient    std. error   t-ratio    p-value
   ---------------------------------------------------------
   const      5.71804e-05   0.00279315    0.02047   0.9837
   uhat_1    -0.0214870     0.0861115    -0.2495    0.8033
   uhat_2     0.168697      0.0861436     1.958     0.0523   *
   uhat_3    -0.0351185     0.0874173    -0.4017    0.6885
   uhat_4    -0.190672      0.0871744    -2.187     0.0305   **
   uhat_5     0.0794575     0.0885161     0.8977    0.3710
   uhat_6     0.117895      0.0887872     1.328     0.1865
   uhat_7    -0.111934      0.0892031    -1.255     0.2118
   uhat_8    -0.112162      0.0892419    -1.257     0.2110
   uhat_9     0.128714      0.0911650     1.412     0.1604
   uhat_10   -0.0221900     0.0933591    -0.2377    0.8125
   uhat_11   -0.0161014     0.0923661    -0.1743    0.8619
   uhat_12    0.182390      0.0925201     1.971     0.0508   *

   Unadjusted R-squared = 0.111880

 Test statistic: LMF = 1.375217,
 with p-value = P(F(12,131) > 1.37522) = 0.186

 Alternative statistic: TR^2 = 16.110728,
 with p-value = P(Chi-square(12) > 16.1107) = 0.186
```

Índice

- Introducción
- Correlación espuria
- Transformaciones de datos
- Cointegración
- Modelos deterministas
- Autocorrelación
- Ideas principales

Ideas principales

- La característica distintiva de las series temporales, en comparación con los datos de corte transversal, es que en ellas la muestra tiene un orden natural: del pasado al futuro
- Este hecho determina qué relaciones son posibles y cuáles no
- Las series temporales económicas típicamente son heterocedásticas y muestran tendencia y estacionalidad, que son dos formas de no estacionariedad en media
- Cuando se ponen en relación series no estacionarias se corre el riesgo de modelizar una correlación espuria
- Para evitar este riesgo, la no estacionariedad debe tratarse de alguna de las siguientes formas:
  - ...transformando los datos para estabilizar la media, o
  - ...incluyendo en el modelo variables explicativas cointegradas con la endógena que, por tanto, captan su componente tendencial, o
  - ...incluyendo en el modelo variables explicativas deterministas, diseñadas para captar la tendencia o la estacionalidad
- Los modelos de regresión con series temporales suelen tener errores autocorrelados
- La Econometría tradicional contempla la autocorrelación como una imperfección del modelo, cuyas consecuencias son similares a los de heterocedasticidad
- Una solución fácil de implementar, especialmente adecuada para muestras grandes, consiste en calcular errores estándar robustificados
- Alternativamente se puede tratar la autocorrelación como un elemento más del modelo

Miguel Jerez (mjerez@ccee.ucm.es)

Departamento de Fundamentos del Análisis Económico II
(Economía Cuantitativa)

Facultad de Ciencias Económicas, UCM

Más materiales en:

http://econometriamj.blogspot.com/

http://mjerez-met-cuantitativos.blogspot.com/