Mortalidad y matrimonio en Inglaterra 1866–1911

Los datos

Los datos de este ejercicio corresponden a la mortalidad anual y la proporción de matrimonios eclesiásticos en Inglaterra entre 1866 y 1911

Fuente: Ejercicio proporcionado por el Prof. Miguel Jerez

`Std_mortality` Mortalidad anual por cada 1000 personas. Serie estandarizada.

`Proportion_marriages` Proporción de matrimonios eclesiásticos anuales por cada 1000 personas.

`d_Std_mortality` Primera diferencia de `Std_mortality`.

`d_Proportion_marriages` Primera diferencia de `Proportion_marriages`.

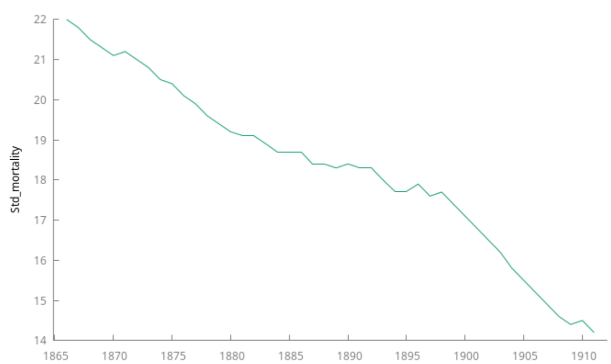
```
open mortality-marriages.gdt
```

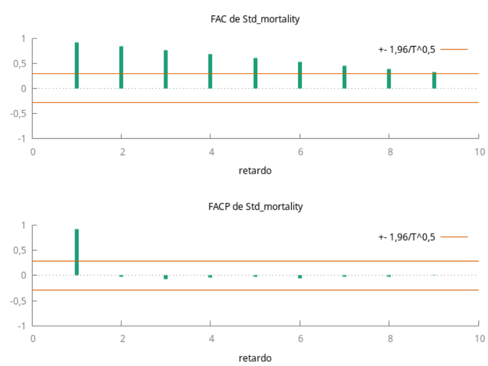
- Ficheros `https://github.com/mbujosab/EconometriaAplicada-SRC/tree/main/Ejercicios`
- Versión en pdf

Datos en nivel de la serie de mortalidad

Gráfico de la serie temporal y su correlograma

```
gnuplot Std_mortality --time-series --with-lines --output="mortality.png"
corrgm Std_mortality 9 --plot="mortalityACF-PACF.png"
```





Estimación de un modelo univariante para la serie de mortalidad

```
arima 1 0 2 ; Std_mortality
```

```
Evaluaciones de la función: 289
Evaluaciones del gradiente: 80

Modelo 2: ARMA, usando las observaciones 1866-1911 (T = 46)
Estimado usando AS 197 (MV exacta)
Variable dependiente: Std_mortality
Desviaciones típicas basadas en el Hessiano

             coeficiente   Desv. típica      z      valor p
  ---------------------------------------------------------
  const       18,0782       3,69647         4,891   1,00e-06 ***
  phi_1        0,996455     0,00501938    198,5     0,0000   ***
  theta_1      0,401166     0,171108        2,345   0,0191   **
  theta_2      0,345176     0,108887        3,170   0,0015   ***

Media de la vble. dep.  18,32174   D.T. de la vble. dep.   2,135615
Media de innovaciones  -0,094657   D.T. innovaciones       0,185241
R-cuadrado              0,994379   R-cuadrado corregido    0,994117
Log-verosimilitud       9,085184   Criterio de Akaike     -8,170368
Criterio de Schwarz     0,972839   Crit. de Hannan-Quinn  -4,745268

                       Real Imaginaria     Módulo Frecuencia
  -----------------------------------------------------------
  AR
   Raíz  1           1,0036     0,0000     1,0036     0,0000
  MA
   Raíz  1          -0,5811    -1,5998     1,7021    -0,3055
   Raíz  2          -0,5811     1,5998     1,7021     0,3055
  -----------------------------------------------------------
```

Contraste de cointegración

```
coint 9 Std_mortality Proportion_marriages --test-down
```

```
Etapa 1: contrastando la existencia de una raíz unitaria en Std_mortality

Contraste aumentado de Dickey-Fuller para Std_mortality
contrastar hacia abajo desde 9 retardos, con el criterio AIC
tamaño muestral 45
la hipótesis nula de raíz unitaria es: [a = 1]

  contraste con constante
  incluyendo 0 retardos de (1-L)Std_mortality
  modelo: (1-L)y = b0 + (a-1)*y(-1) + e
  valor estimado de (a - 1): 0,00678121
  estadístico de contraste: tau_c(1) = 0,615887
  valor p asintótico 0,9902
  Coef. de autocorrelación de primer orden de e: 0,085

Etapa 2: contrastando la existencia de una raíz unitaria en Proportion_marriages

Contraste aumentado de Dickey-Fuller para Proportion_marriages
contrastar hacia abajo desde 9 retardos, con el criterio AIC
```

```
tamaño muestral 39
la hipótesis nula de raíz unitaria es: [a = 1]

  contraste con constante
  incluyendo 6 retardos de (1-L)Proportion_marriages
  modelo: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  valor estimado de (a - 1): 0,0831149
  estadístico de contraste: tau_c(1) = 1,04236
  valor p asintótico 0,9971
  Coef. de autocorrelación de primer orden de e: -0,068
  diferencias retardadas: F(6, 31) = 3,197 [0,0147]

Etapa 3: regresión cointegrante

Regresión cointegrante -
MCO, usando las observaciones 1866-1911 (T = 46)
Variable dependiente: Std_mortality

                      coeficiente  Desv. típica  Estadístico t  valor p
  ---------------------------------------------------------------------
  const              -10,8466      1,42447        -7,614      1,45e-09 ***
  Proportion_marri~    0,418536    0,0203914      20,53       3,67e-24 ***

Media de la vble. dep.  18,32174   D.T. de la vble. dep.   2,135615
Suma de cuad. residuos  19,40865   D.T. de la regresión    0,664158
R-cuadrado               0,905434  R-cuadrado corregido    0,903284
Log-verosimilitud      -45,42395   Criterio de Akaike      94,84790
Criterio de Schwarz     98,50518   Crit. de Hannan-Quinn   96,21794
rho                      0,228283  Durbin-Watson           1,535570

Etapa 4: contrastando la existencia de una raíz unitaria en uhat

Contraste aumentado de Dickey-Fuller para uhat
contrastar hacia abajo desde 9 retardos, con el criterio AIC
tamaño muestral 45
la hipótesis nula de raíz unitaria es: [a = 1]

  contraste sin constante
  incluyendo 0 retardos de (1-L)uhat
  modelo: (1-L)y = (a-1)*y(-1) + e
  valor estimado de (a - 1): -0,771717
  estadístico de contraste: tau_c(2) = -5,22784
  valor p asintótico 5,236e-05
  Coef. de autocorrelación de primer orden de e: 0,023

Hay evidencia de una relación cointegrante si:
(a) La hipótesis de existencia de raíz unitaria no se rechaza para las variables individuales y
(b) La hipótesis de existencia de raíz unitaria se rechaza para los residuos (uhat) de la regresión coi
```

Regresión de la mortalidad sobre la proporción de matrimonios eclesiásticos

```
ols Std_mortality 0 Proportion_marriages
modtest --normality --quiet
modtest --white --quiet
modtest --autocorr 5 --quiet
```

```
Modelo 6: MCO, usando las observaciones 1866-1911 (T = 46)
Variable dependiente: Std_mortality

                     coeficiente  Desv. típica  Estadístico t  valor p
  --------------------------------------------------------------------
  const              -10,8466      1,42447         -7,614      1,45e-09 ***
  Proportion_marri~    0,418536    0,0203914       20,53       3,67e-24 ***

Media de la vble. dep.  18,32174   D.T. de la vble. dep.   2,135615
Suma de cuad. residuos  19,40865   D.T. de la regresión    0,664158
R-cuadrado              0,905434   R-cuadrado corregido    0,903284
F(1, 44)                421,2815   Valor p (de F)          3,67e-24
Log-verosimilitud      -45,42395   Criterio de Akaike      94,84790
Criterio de Schwarz     98,50518   Crit. de Hannan-Quinn   96,21794
rho                     0,228283   Durbin-Watson           1,535570


Contraste de la hipótesis nula de distribución Normal:
Chi-cuadrado(2) = 0,260 con valor p 0,87796


Contraste de heterocedasticidad de White

Estadístico de contraste: TR^2 = 1,729996,
con valor p = P(Chi-cuadrado(2) > 1,729996) = 0,421052


Contraste de Breusch-Godfrey para autocorrelación hasta el orden 5

Estadístico de contraste: LMF = 1,947454,
con valor p = P(F(5,39) > 1,94745) = 0,108

Estadístico alternativo: TR^2 = 9,190388,
con valor p = P(Chi-cuadrado(5) > 9,19039) = 0,102

Ljung-Box Q' = 9,05845,
con valor p = P(Chi-cuadrado(5) > 9,05845) = 0,107
```

Regresión en primeras diferencias

```
diff Std_mortality Proportion_marriages
ols d_Std_mortality 0 d_Proportion_marriages
modtest --normality --quiet
modtest --white --quiet
modtest --autocorr 5 --quiet
```

```
Modelo 8: MCO, usando las observaciones 1867-1911 (T = 45)
```

```
Variable dependiente: d_Std_mortality

                      coeficiente  Desv. típica  Estadístico t  valor p
  ---------------------------------------------------------------------
  const               -0,172792     0,0230316      -7,502      2,43e-09 ***
  d_Proportion_mar~    0,00142536   0,0117781       0,1210     0,9042

Media de la vble. dep. -0,173333   D.T. de la vble. dep.   0,149848
Suma de cuad. residuos  0,987664   D.T. de la regresión    0,151555
R-cuadrado              0,000340   R-cuadrado corregido   -0,022907
F(1, 43)                0,014645   Valor p (de F)          0,904241
Log-verosimilitud       22,07697   Criterio de Akaike     -40,15393
Criterio de Schwarz    -36,54061   Crit. de Hannan-Quinn  -38,80692
rho                     0,089193   Durbin-Watson           1,806988


Contraste de la hipótesis nula de distribución Normal:
Chi-cuadrado(2) = 14,808 con valor p 0,00061


Contraste de heterocedasticidad de White

Estadístico de contraste: TR^2 = 2,149006,
con valor p = P(Chi-cuadrado(2) > 2,149006) = 0,341467


Contraste de Breusch-Godfrey para autocorrelación hasta el orden 5

Estadístico de contraste: LMF = 0,589588,
con valor p = P(F(5,38) > 0,589588) = 0,708

Estadístico alternativo: TR^2 = 3,239657,
con valor p = P(Chi-cuadrado(5) > 3,23966) = 0,663

Ljung-Box Q' = 4,0454,
con valor p = P(Chi-cuadrado(5) > 4,0454) = 0,543
```

Preguntas

Pregunta 1

Discuta de todas las formas posibles si la serie temporal de mortalidad (`Std_mortality`) es estacionaria en media (i.e., la realización de un proceso estocástico estacionario), usando para ello los resultados de los apartados Datos en nivel de la serie de mortalidad y Contraste de cointegración.

(Respuesta 1)

Pregunta 2

Discuta si las series de mortalidad y proporción de matrimonios eclesiásticos están cointegradas, a partir de los resultados del apartado Contraste de cointegración.

(Respuesta 2)

Pregunta 3

Sin embargo, £qué sugieren los resultados de las secciones Regresión de la mortalidad sobre la proporción de matrimonios eclesiásticos y Regresión en primeras diferencias respecto a la relación entre `Std_mortality` y `Proportion_marriages`?

(Respuesta 3)

Pregunta 4

Los listados en Regresión de la mortalidad sobre la proporción de matrimonios eclesiásticos y Regresión en primeras diferencias muestran los principales resultados obtenidos al estimar por MCO dos modelos de regresión que relacionan las dos variables consideradas en este ejercicio. Resuma y comente los resultados de estimación y diagnosis que le parezcan más relevantes. Si detecta alguna desviación del cumplimiento de las hipótesis habituales, discuta sus consecuencias sobre las propiedades del estimador MCO y sugiera una forma de tratarla.

(Respuesta 4)

Pregunta 5

Interprete la pendiente de la regresión cointegrante estimada en la Etapa 3 del Contraste de cointegración.

(Respuesta 5)

Pregunta 6

Indique cuáles de las siguientes expresiones representan el modelo de la sección Estimación de un modelo univariante para la serie de mortalidad, con un redondeo a tres decimales

1. $(1-0{,}997\,\mathsf{B})\;(X_t-18{,}078)=\left(1+0{,}401\,\mathsf{B}+0{,}345\,\mathsf{B}^2\right)\hat{U}_t$
2. $(1-0{,}997\,\mathsf{B})\;(X_t-18{,}078)=\left(1-0{,}401\,\mathsf{B}-0{,}345\,\mathsf{B}^2\right)\hat{U}_t$
3. $(1+0{,}997\,\mathsf{B})\;(X_t-18{,}078)=\left(1+0{,}401\,\mathsf{B}+0{,}345\,\mathsf{B}^2\right)\hat{U}_t$
4. $X_t=18{,}078+\frac{1+0{,}401\,\mathsf{B}+0{,}345\,\mathsf{B}^2}{1-0{,}997\,\mathsf{B}}\hat{U}_t$
5. $X_t=-18{,}078+\frac{1+0{,}401\,\mathsf{B}+0{,}345\,\mathsf{B}^2}{1-0{,}997\,\mathsf{B}}\hat{U}_t$
6. $X_t=18{,}078+\frac{1-0{,}401\,\mathsf{B}-0{,}345\,\mathsf{B}^2}{1-0{,}997\,\mathsf{B}}\hat{U}_t$
7. $X_t=18{,}078+\frac{1+0{,}401\,\mathsf{B}+0{,}345\,\mathsf{B}^2}{1+0{,}997\,\mathsf{B}}\hat{U}_t$
8. $\frac{1-0{,}997\,\mathsf{B}}{1+0{,}401\,\mathsf{B}+0{,}345\,\mathsf{B}^2}\;(X_t-18{,}078)=\hat{U}_t$
9. $\frac{1-0{,}997\,\mathsf{B}}{1+0{,}401\,\mathsf{B}+0{,}345\,\mathsf{B}^2}\,X_t=18{,}078+\hat{U}_t$
10. $\frac{1-0{,}997\,\mathsf{B}}{1-0{,}401\,\mathsf{B}-0{,}345\,\mathsf{B}^2}\;(X_t-18{,}078)=\hat{U}_t$

(Respuesta 6)

Pregunta 7

A la luz de la Estimación de un modelo univariante para la serie de mortalidad, si tuviera que clasificar el proceso estocástico subyacente del que la serie temporal es una realización £diría que es invertible? £O que no lo es? £diría que es estacionario? £O que no lo es? Explique su respuesta.

(Respuesta 7)

Pregunta 8

£Cuáles de los modelos de más arriba considera aceptables? £O qué mejoras sugeriría para ellos?
(Respuesta 8)

Respuestas

Respuesta 1

La serie temporal `Std_mortality` NO es estacionaria en media, como se aprecia en las secciones:

- Gráfico de la serie temporal y su correlograma.
  - El gráfico de la serie muestra una tendencia decreciente.
  - La FAC muestra mucha persistencia, los coeficientes decrecen a un ritmo aproximadamente lineal; y el primer coeficiente de la PACF está próximo a uno.
- Estimación de un modelo univariante para la serie de mortalidad: El modelo univariante estimado tiene una raíz AR aproximadamente igual a 1.
- Contraste de cointegración: El test ADF calculado en la Etapa 1 no rechaza la hipótesis (raíz unitaria) con un p-valor de `0.9902`

(Pregunta 1)

Respuesta 2

Las conclusiones de las distintas etapas del test de cointegración son los siguientes:

Etapa 1 El test ADF no rechaza que la serie de mortalidad sea I(1). `(valor p asintótico 0,9902)`

Etapa 2 El test ADF no rechaza que la serie de proporción de matrimonios eclesiásticos sea I(1). `(valor p asintótico 0,9971)`

Etapa 3 La regresión (cointegrante) de mortalidad sobre la proporción de matrimonios eclesiásticos es significativa (parámetros significativos y elevado R^2 `(0,905434)`.

Etapa 4 El test ADF rechaza contundentemente que los residuos de la regresión cointegrante sean I(1). `(valor p asintótico 5,236e-05)`

Consecuentemente, el test indica que ambas series están cointegradas (*pero, como sugiere tanto el sentido común como la Regresión en primeras diferencias la relación es espuria*, véase la pregunta 3).
(Pregunta 2)

Respuesta 3

Aunque el modelo de Regresión de la mortalidad sobre la proporción de matrimonios eclesiásticos muestra un buen ajuste (un elevado R^2) y los parámetros estimados son muy significativos, la relación entre ambas variables se desvanece al diferenciar los datos para lograr la estacionariedad. Ello sugiere, al igual que el sentido común, que la relación es espuria.
(Pregunta 3)

Respuesta 4

Modelo de regresión MCO para datos en nivel (Regresión de la mortalidad sobre la proporción de matrimonios eclesiásticos): Todos los coeficientes son muy significativos. El ajuste del modelo, medido por el valor del R^2 es muy elevado. Los contrastes sobre los residuos no rechazan (ni al 1 %, ni al 5 % ni al 10 % de significación) las hipótesis nulas de normalidad, homoscedasticidad y ausencia de autocorrelación. Es decir, de la salida de Gretl no se puede inferir que haya ningún problema con este modelo.

Modelo para datos en primeras diferencias (Regresión en primeras diferencias): El único coeficiente significativo es el término constante. El ajuste del modelo, medido por el valor del R^2, es prácticamente nulo. Los contrastes residuales rechazan la hipótesis nula de normalidad, aunque no rechazan las de homoscedasticidad y ausencia de autocorrelación.

Si las perturbaciones no tienen distribución normal las estimaciones no serán eficientes en el sentido máximo-verosímil (aunque sí en el de Gauss-Markov) y la distribución de los estadísticos habituales será distinta de la teórica bajo el supuesto de normalidad de las perturbaciones (por ejemplo, los estadísticos de la t no tendrán exactamente una distribución t de student).

No obstante, dado que la relación entre variables es espuria, ninguno de estos modelos de regresión es válido como explicación de la tasa de mortalidad.

(Pregunta 4)

Respuesta 5

La pendiente de la regresión estimada en la Etapa 3 (que es la misma que la de la sección de la regresión en niveles) indica que un aumento de un uno por mil en la proporción de matrimonios eclesiásticos da lugar a un aumento de un 0.419 por mil en la mortalidad esperada (pero, dado que la relación es espuria, interpretar este resultado carece de sentido).

(Pregunta 5)

Respuesta 6

Recuerde que signo de los parámetros MA en las salidas de Gretl tienen el signo cambiado respecto a convenio habitual en los manuales de series temporales, es decir, para los polinomios AR $(1-\phi_1\mathsf{B}-\cdots-\phi_p\mathsf{B}^p)$, tenemos que `phi_j` es "$\phi_j$"(es decir, al escribir el modelo el signo del parámetro `phi_j` aparece con un menos delante); pero para los MA $(1-\theta_1\mathsf{B}-\cdots-\theta_p\mathsf{B}^p)$, tenemos que `theta_j` es "$-\theta_j$"(es decir, al escribir no cambiamos el signo de parámetro `theta_j` pues ya lleva el "$-$ïncorporado). Además, `const` es la estimación del valor esperado μ del proceso $\boldsymbol{X}$, es decir, que $(X_t-\mu \mid t\in\mathbb{Z})$ es un proceso ARMA de media cero.

Por tanto, las expresiones correctas son:

Expresión 1 modelo ARMA$(1,2)$: $\boldsymbol{\phi}(\mathsf{B})(X_t-\mu)=\boldsymbol{\theta}(\mathsf{B})U_t$

Expresión 4 su representación MA(∞): $(X_t-\mu)=\frac{\boldsymbol{\theta}}{\boldsymbol{\phi}}(\mathsf{B})U_t \;\rightarrow\; X_t=\mu+\frac{\boldsymbol{\theta}}{\boldsymbol{\phi}}(\mathsf{B})U_t$

Expresión 8 su representación AR(∞): $\frac{\boldsymbol{\phi}}{\boldsymbol{\theta}}(\mathsf{B})(X_t-\mu)=U_t$

ąOjo, la cuarta expresión solo es posible porque ϕ_1 no es exactamente 1! Si fuera 1, el polinomio autorregresivo $1-\mathsf{B}$ no tendría una inversa sumable y, por tanto, ni el proceso sería estacionario, ni habría una representación del proceso como media móvil infinita como la Expresión 4.

(Pregunta 6)

Respuesta 7

La raíz AR estimada está muy próxima a 1, por lo que cabe pensar que la serie proviene de un proceso estocástico NO estacionario. Sin embargo, las raíces del polinomio MA tienen un módulo claramente mayor que uno, por lo que el modelo tiene claramente una representación AR(∞), es decir, es invertible.

(Pregunta 7)

Respuesta 8

£Cuáles de los modelos de más arriba considera aceptables? £O qué mejoras sugeriría para ellos?

En cuanto al modelo univariante Probablemente debería incorporar una diferencia ordinaria, en lugar de un término AR(1).

En cuanto a los modelos de regresión En el modelo de las serie en diferencias hay, probablemente, un problema de autocorrelación dado el elevado valor del estadístico Durbin-Watson (es próximo a 2), por lo que quizá debería ser estimado por mínimos cuadrados generalizados asumiendo un modelo autorregresivo AR(1) para el error.

No obstante, el modelo en diferencias (y el sentido común) sugiere que la relación entre ambas variables es espuria. Consecuentemente, ninguna de las dos regresiones (en niveles o en diferencias) arrojará un modelo aceptable ni siquiera con las mejoras sugeridas.

(Pregunta 8)