Consumo de petroleo y frecuencia del nombre Óscar

Datos

Ejemplo obtenido de https://tylervigen.com/spurious/correlation/8118_popularity-of-the-first-name-oscar_correlates-with_petroluem-consumption-in-greece

Datos anuales. Muestra: 1980–2022

Consumo de petroleo en Grecia `ConsumoPetroleo`

Título detallado de la variable Volume of petroluem consumption consumed in Greece in millions of barrels per day

Fuente Energy Information Administration

Popularidad del nombre Óscar en EEUU `FrecuenciaOscar`

Título detallado de la variable Babies of all sexes born in the US named Óscar

Fuente US Social Security Administration

```
open ../../datos/NombreOscarYConsumoDePetroleo.gdt
```

```
gnuplot ConsumoPetroleo FrecuenciaOscar  --time-series --with-lines --output="PetroleoOscar.png"
```

./Examen-NombreOscarYConsumoDePetroleo/PetroleoOscar.png

Ficheros: Versión del ejercicio en pdf; html.

- Datos: NombreOscarYConsumoDePetroleo.gdt
- Guión de gretl: Examen-NombreOscarYConsumoDePetroleo.inp

Datos en nivel del consumo de petroleo en Grecia

Gráfico de la serie temporal y su correlograma

```
gnuplot ConsumoPetroleo --time-series --with-lines --output="consumoPetroleo.png"
corrgm ConsumoPetroleo 9 --plot="consumoPetroleoACF-PACF.png"
```

./Examen-NombreOscarYConsumoDePetroleo/consumoPetroleo.png

./Examen-NombreOscarYConsumoDePetroleo/consumoPetroleoACF-

Estimación de un primer modelo univariante para la serie de consumo de petroleo

```
arima 1 0 1 ; ConsumoPetroleo
```

```
series res1petroleo = $uhat
corrgm res1petroleo
```

Estimación de un segundo modelo univariante para la serie de consumo de petroleo

```
arima 1 1 0 --nc ; ConsumoPetroleo
```

```
series res2petroleo = $uhat
corrgm res2petroleo
```

Datos en nivel de la popularidad del nombre Óscar en EEUU

Gráfico de la serie temporal y su correlograma

```
gnuplot FrecuenciaOscar --time-series --with-lines --output="consumoOscar.png"
corrgm FrecuenciaOscar --plot="consumoOscarACF-PACF.png"
```

./Examen-NombreOscarYConsumoDePetroleo/consumoOscar.png

./Examen-NombreOscarYConsumoDePetroleo/consumoOscarAC

Estimación de un primer modelo univariante para la serie de popularidad del nombre Óscar

```
arima 1 0 1 ; FrecuenciaOscar
```

```
series res1Oscar = $uhat
corrgm res1Oscar
```

Estimación de un segundo modelo univariante para la serie de popularidad del nombre Óscar

```
arima 1 1 0 --nc ; FrecuenciaOscar
```

```
series res2Oscar = $uhat
corrgm res2Oscar
```

Contraste de cointegración

```
coint 2 ConsumoPetroleo FrecuenciaOscar --test-down
```

Regresión del consumo de petroleo sobre la popularidad del nombre Óscar

Primer modelo

Segundo modelo: regresión del consumo de petroleo sobre la popularidad del nombre Óscar con modelo de corrección de error AR1

Regresión en primeras diferencias

Primer modelo

Segundo modelo: Regresión en primeras diferencias con intervención en el año 2020

Dado que hubo una caída muy acusada en el consumo de petroleo del año 20 debido al confinamiento por la Covid19 (*circunstancia que no afectó de manera particular a la popularidad del nombre "Óscar"*), el siguiente modelo introduce una variable ficticia para el año 2020 (se introduce en primeras diferencias como el resto de variables del modelo).

Preguntas

Pregunta 1

Discuta de todas las formas posibles si las series temporales de consumo de petroleo (`ConsumoPetroleo`) y popularidad del nombre Óscar (`FrecuenciaOscar`) son estacionarias en media (i.e., son la realización de procesos estocásticos estacionarios), usando para ello los resultados de los apartados Datos en nivel del consumo de petroleo en Grecia, Datos en nivel de la popularidad del nombre Óscar en EEUU y .

(Respuesta 1)

Pregunta 2

Discuta si las series temporales `ConsumoPetroleo` y `FrecuenciaOscar` están cointegradas, a partir de los resultados del apartado .

(Respuesta 2)

Pregunta 3

¿Contradice la Regresión en primeras diferencias la posibilidad de que están relacionados el consumo de petroleo en Grecia y la popularidad del nombre de pila Oscar en los EEUU?

(Respuesta 3)

Pregunta 4

Los listados de la Regresión del consumo de petroleo sobre la popularidad del nombre Óscar y la 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 (dichos modelos están referidos como "*primeros modelos*").

Resuma y comente los resultados de estimación y diagnosis que le parezcan más relevantes de esos dos primeros modelos en niveles y en diferencias.

Si detecta alguna desviación del cumplimiento de las hipótesis habituales, discuta sus consecuencias sobre las propiedades del estimador MCO y sugiera alguna forma de tratarla.

(Respuesta 4)

Pregunta 5

Tanto en el caso de las regresiones en niveles como en el caso de las regresiones en primeras diferencias, también se muestra los resultados de un segundo modelo de regresión.

Explique en cada caso si ese segundo modelo responde a algún posible tratamiento que haya indicado en la pregunta anterior y por qué (o si dicho tratamiento no tiene nada que ver con lo que usted dijo). En cualquier caso, señale (en cada caso) si considera que ese segundo modelo es mejor o peor que el primero, y en qué aspectos.

(Respuesta 5)

Pregunta 6

En la Sección Datos en nivel del consumo de petroleo en Grecia aparecen dos modelos univariantes. Compare los resultados he indique si alguno de ellos es preferible y por qué.

(Respuesta 6)

Pregunta 7

En la Sección Datos en nivel de la popularidad del nombre Óscar en EEUU aparecen dos modelos univariantes. Compare los resultados he indique si alguno de ellos es preferible y por qué.

(Pregunta 7)

Pregunta 8

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

(Respuesta 8)

Respuestas

Respuesta 1

Ambas series (`ConsumoPetroleo` y `FrecuenciaOscar`) parecen ser NO estacionarias en media,

- Sus gráficos muestran una clara evolución de su nivel a lo largo de la muestra (los primeros años ascendente y desde 2005 descendente).
- Ambas funciones de autocorrelación (FAC) muestran persistencia (sus coeficientes decrecen despacio y a un ritmo aproximadamente lineal); y el primer coeficiente de la PACF está próximo a uno en ambos casos.
- Estimación de un primer modelo univariante para la serie de consumo de petroleo: El modelo univariante estimado tiene una raíz AR aproximadamente igual a 1.
- Estimación de un primer modelo univariante para la serie de popularidad del nombre Óscar: El modelo univariante estimado tiene una raíz AR aproximadamente igual a 1.
- : Los test ADF calculados en las etapas 1 y 2 no rechazan la hipótesis (raíz unitaria) con p-valores superiores al 0.4

(Pregunta 1)

Respuesta 2

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

Etapa 1 El test ADF no rechaza que la serie `ConsumoPetroleo` sea I(1) para niveles de significación inferiores al 40 % (p-valor asintótico `0,4672`).

Etapa 2 El test ADF no rechaza que la serie `FrecuenciaOscar` sea I(1) para niveles de significación inferiores al 40 % (p-valor asintótico `0,4218`).

Etapa 3 En la regresión (cointegrante) de mortalidad sobre la proporción de matrimonios eclesiásticos ambos parámetros (constante y pendiente) resultan ser muy significativos, y el R^2 está próximo a 1.

Etapa 4 El test ADF rechaza que los residuos de la regresión cointegrante sean I(1) tanto al 10 % como al 5 % de significación (p-valor asintótico `0,03258`)

Consecuentemente, el test NO rechaza la cointegración de ambas series (*en contra de lo que sugiere el sentido común*).

(Pregunta 2)

Respuesta 3

La relación NO se desvanece al diferenciar los datos para lograr la estacionariedad; que es precisamente lo que cabe esperar cuando la relación existe, pues si

$$\boldsymbol{y} = \beta_1 \mathbf{1} + \beta_2 \boldsymbol{x} + \boldsymbol{u}$$

Entonces también debe ser cierto que

$$\nabla \boldsymbol{y} = \beta_2 \nabla \boldsymbol{x} + \nabla \boldsymbol{u}$$

Sorprendentemente, en la Regresión en primeras diferencias la constante es NO significativa, la pendiente es muy significativa y el R^2 no es, en absoluto, despreciable (`R-cuadrado 0,327929`). Es decir, la Regresión en primeras diferencias no contradice la posibilidad de que ambas variables estén relacionadas.

Comentario y moraleja: Pese a los resultados estadísticos, la relación entre `ConsumoPetroleo` y `FrecuenciaOscar` es evidentemente espuria (es imposible argumentar con algún fundamento que la frecuencia del nombre Óscar en EEUU tenga ninguna influencia sobre el consumo de petroleo en Grecia… o viceversa). ¡Ojo con interpretar los resultados estadísticos sin un mínimo espíritu crítico!

(Pregunta 3)

Respuesta 4

Primer modelo para datos en nivel (Regresión del consumo de petroleo sobre la popularidad del nombre Óscar): 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 la hipótesis nula de normalidad, pero si rechazan la hipótesis de homocedasticidad y de ausencia de autocorrelación.

En cuanto a la heterocedasticidad, sería conveniente estimar indicando la opción de desviaciones típicas robustas, pues los p-valores están mal calculados en presencia de heterocedasticidad. Más importante es la presencia de autocorrelación; dado que hay indicios de autocorrelación de orden 1 en los errores de ajuste, sería conveniente estimar el modelo incorporando un modelo AR(1) para el error.

Primer modelo para datos en primeras diferencias (Regresión en primeras diferencias): El único coeficiente significativo es la pendiente (es decir, al diferenciar las series NO ha desaparecido la relación entre ellas), y el ajuste del modelo, medido por el valor del R^2, es superior al 30 %. Los contrastes residuales rechazan la hipótesis nula de normalidad, pero no rechazan las de homocedasticidad 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). En la práctica esto no ocasiona un problema grave en general.

(Pregunta 4)

Respuesta 5

Segundo modelo para datos en nivel (Regresión del consumo de petroleo sobre la popularidad del nombre Óscar): El segundo modelo corresponde a una regresión con modelo AR(1) para el error (tal y como se sugería en la pregunta anterior). La estimación ha convergido en 5 iteraciones, los parámetros son muy significativos y el R^2 ajustado es superior al del primer modelo. Tampoco en este caso se rechaza la hipótesis de normalidad en los residuos del ajuste. Todo ello sugiere que este segundo modelo sería ligeramente superior al primero (*si no fuera porque la relación es evidentemente espuria y, por tanto, ninguno de estos modelos es aceptable*).

Segundo modelo para datos en primeras diferencias (Regresión en primeras diferencias): El segundo modelo incluye un nuevo regresor para captar la caída de consumo de petroleo del año 2020 debida al confinamiento por la Covid19. Por tanto, esta modificación no tiene nada que ver con lo indicado en la pregunta anterior.

No obstante, este modelo parece superior al primero. Los parámetros correspondientes a `d_FrecuenciaOscar` y `d_Covid` son muy significativos, el R^2 ajustado es claramente superior y los criterios de información han mejorado ligeramente (i.e., ahora toman valores más bajos). Además, gracias a la intervención del año atípico 2020, los residuos pasan todos los contrastes (incluido el de normalidad).

(Pregunta 5)

Respuesta 6

El primer modelo es un ARMA$(1,1)$ con media distinta de cero, y los tres parámetros estimados son muy significativos. El mayor inconveniente es que la raíz autorregresiva es prácticamente 1. Dado que hay una fuerte evidencia de que el proceso NO es estacionario en media, es preferible diferenciar la serie e identificar un proceso ARIMA.

El segundo modelo es un ARIMA(1,1,0) con media cero. Su principal ventaja es que el modelo estimado corresponde a un proceso que (una vez diferenciado) es invertible y estacionario (pues no tiene polinomio MA, y el módulo de la raíz AR es `2,9879` > 1).

Pese a que tiene menos parámetros estimados, el ajuste y los criterios de información son ligeramente mejores. Además, los p-valores de los estadísticos Q de Ljung-Box son más elevados en este segundo modelo, por lo que sus residuos tienen una mayor apariencia de ruido blanco". En resumen, este segundo modelo parece mejor que el primero.

(Pregunta 6)

Respuesta 7

Como en el caso anterior, el primer modelo es un ARMA$(1,1)$ con media distinta de cero, y los tres parámetros estimados son muy significativos. De nuevo, el mayor inconveniente es que la raíz autorregresiva es prácticamente 1. Dado que hay una fuerte evidencia de que el proceso NO es estacionario en media, es preferible diferenciar la serie e identificar un proceso ARIMA.

El segundo modelo es un ARIMA(1,1,0) con media cero. Su principal ventaja es que el modelo estimado corresponde a un proceso que (una vez diferenciado) es invertible y estacionario (pues no tiene polinomio MA, y el módulo de la raíz AR es `1,8658` > 1).

Pese a que tiene menos parámetros estimados, el ajuste y los criterios de información son ligeramente mejores. Además, los p-valores de los estadísticos Q de Ljung-Box son más elevados en este segundo modelo, por lo que sus residuos tienen una mayor apariencia de ruido blanco". En resumen, este segundo modelo parece mejor que el primero.

(Pregunta 6)

Respuesta 8

En cuanto a los modelos univariantes Como se ha dicho, para ambas series, el segundo modelo es mejor que el primero. En ambos casos corresponde a un proceso invertible y estacionario, el parámetro estimado es significativo y (según los estadísticos Q de Ljung-Box) los residuos parecen ruido blanco.

En cuanto a los modelos de regresión Los cuatro modelos intentan modelizar una relación evidentemente espuria: nada tiene que ver la popularidad del nombre Óscar en EEUU con el consumo de petroleo en Grecia. Consecuentemente ninguna de estas regresiones ofrece un modelo aceptable o, siquiera, razonable.

(Pregunta 8)