

## CAPÍTULO 10

---

### Raíces Unitarias Estacionales y Estacionalidad Estocástica (SARIMA)

---

#### 10.1. Modelos SARIMA

En esta sección se utilizará el operador rezago estacional,  $L^s$ , donde  $s > 1$ , entero, es el período de una dinámica estacional, que se asume está presente en la serie  $Y_t$ . Se define mediante las operaciones

$$L^s(Y_t) = Y_{t-s}, \quad (10.1a)$$

$$(I - L^s)Y_t = \Delta_s Y_t = Y_t - Y_{t-s}, \quad (10.1b)$$

$$(I - L^s)^2 Y_t = \Delta_s^2 Y_t = (I - 2L^s + L^{2s})Y_t. \quad (10.1c)$$

Si la serie  $Y_t$  tiene una componente estacional con período  $s$  es posible eliminarla diferenciando una ó dos veces, es decir, transformando  $Y_t$  en  $W_t$ :

$$W_t = \Delta_s^D Y_t, \quad D = 1, 2. \quad (10.2)$$

En R la instrucción para calcular (10.2) es `diff(y, s, D)`. Se espera que la serie  $W_t$  sea estacionaria en covarianza, con una posible estructura *SARMA*. En caso

de existir una componente de tendencia y otra de estacionalidad en la serie  $Y_t$ , es posible eliminar ambas mediante diferenciaciones ordinaria y estacional

$$W_t = (1 - L^s)^D (1 - L)^d Y_t = \Delta_s^D \Delta^d Y_t \quad (10.3)$$

En R la transformación (10.3) se puede realizar con la instrucción siguiente

```
Wt = diff(diff(Yt, 1, d), s, D)
```

La transformación (filtro lineal)  $X_t = \Delta^d \Delta_s^D Y_t$  elimina la tendencia y la estacionalidad de  $Y_t$ , dejando una posible estructura  $SARMA(p, q)(p_s, q_s)[s]$ , definida en (8.44), pag. 199.

**Ejemplo 10.1.1.** *Retomando el Ejemplo (7.3.2) en la pag. 141, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_t(t) + \varepsilon_t.$$

*En el Ejemplo 8.2.5, en la pag. 169, se ajustó un MA(6) a los residuos estimados  $\hat{\varepsilon}_t$ . En el Ejemplo 8.5.5, en la pag. 208, se estimó un modelo alterno SARMA(2,1)(2,1)[4] para los residuos. En la Figura 10.1 se muestra el efecto de la diferenciación ordinaria,  $U_t = \text{diff}(Y_t, 1, d)$ , en el panel superior derecho. Luego el de la diferenciación estacional  $W_t = \text{diff}(U_t, s, D)$ , en el panel inferior derecho. El resultado es que la serie  $W_t$  muestra una dinámica que puede corresponder a un  $ARIMA(3, 0, 2)(0, 0, 1)[4]$  with zero mean.*

### Código R 10.1.1.

```
#-----Ejemplo diferencias serie cemento
E = read.table("cementq.dat", header = TRUE)
attach(E)
y = ts(y, frequency=4, start=c(1956, 1), end=c(1994, 3))

#----- con stl()
m1 = stl(y, s.window = 'per')
plot(m1)
```

```

dyos = diff(diff(y, 4, 1), 1, 1)
dys = diff(y, 4, 1)
dyo = diff(y, 1, 1)

par(mfrow = c(2, 2))
ts.plot(y, main=" (A) ")
ts.plot(dyo, main=" (B) ")
ts.plot(dys, main=" (C) ")
ts.plot(dyos, main=" (D) ")

require(TSA)
par(mfrow = c(1, 2))
acf(dyos, 16, drop.lag.0=TRUE, ci.type="ma")
pacf(dyos, 16)

dyos = ts(dyos, frequency=4)
require(forecast)
auto.arima(dyos)
ARIMA(3, 0, 2)(0, 0, 1)[4] with zero mean

```

**Definición 10.1.1.** *Se dice que la serie  $(Y_t, t \in \mathbb{Z})$ , sigue un modelo SARIMA( $p, d, q$ )( $p_s, D, q_s$ )[ $s$ ] si cumple que*

$$\Phi_{p_s}(L^s)\Phi_p(L)\Delta_s^D\Delta^d Y_t = \Theta_{q_s}(L^s)\Theta_q(L)\varepsilon_t, \quad (10.4)$$

donde  $d, D = 1, 2$ , y  $\varepsilon_t \sim RB(0, \sigma^2)$ .  $Y$

$$\begin{aligned} \Phi_p(L) &= 1 - \sum_{j=1}^p \varphi_j L^j, \\ \Theta_q(L) &= 1 + \sum_{j=1}^q \theta_j L^j, \\ \Phi_{p_s}(L^s) &= 1 - \sum_{j=1}^{p_s} \varphi_{s,j} L^{js}, \\ \Theta_{q_s}(L^s) &= 1 + \sum_{j=1}^{q_s} \theta_{s,j} L^{js}. \end{aligned}$$

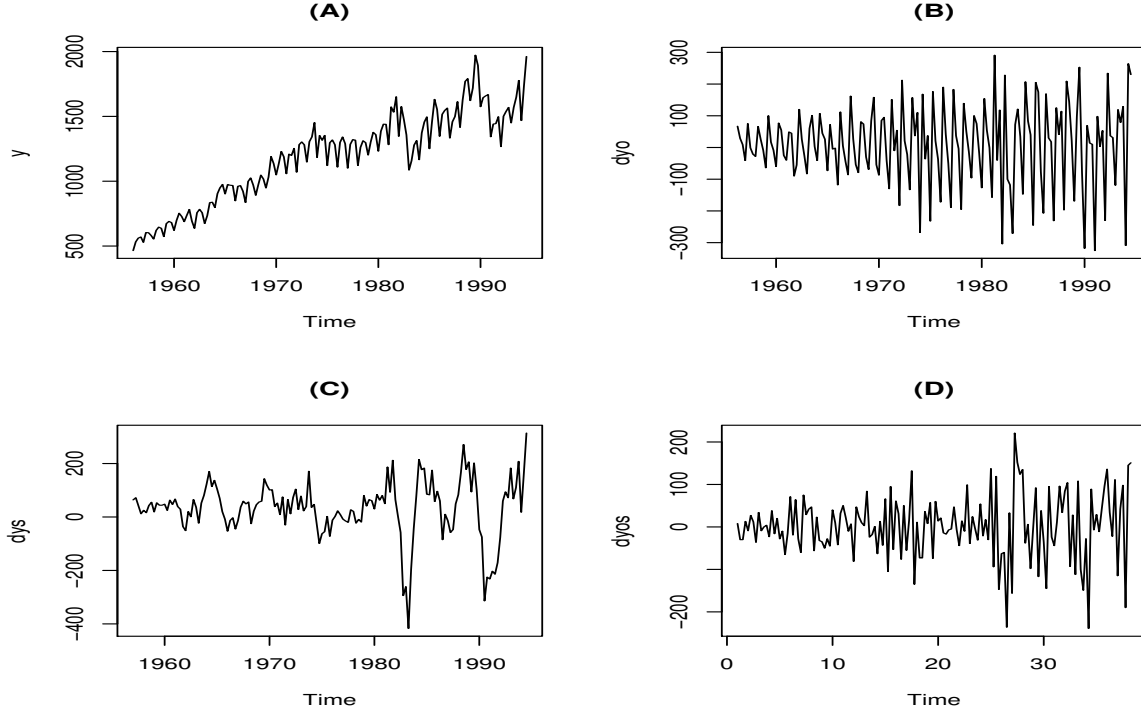


Figura 10.1: Aplicación de diferencia simple y luego diferencia estacional a la serie de producción de cemento

Las notaciones en la Definición anterior no son las utilizadas en la mayoría de los textos, están tomadas del libro Prado and West [2010]. La idea del modelo SARIMA es que la serie diferenciada  $W_t$  que satisface

$$\varphi_p(L)\Phi_P(L^s)W_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2).$$

tiene un papel equivalente al residuo estructural  $\varepsilon_t$  en el modelo de Componentes

$$Y_t = \sum_{j=0}^k \beta_j t^j + \sum_{j=1}^{s-1} \delta_j I_j(t) + \varepsilon_t. \quad (10.5)$$

En (10.4), al diferenciar la serie  $Y_t$ , se eliminan la tendencia y estacionalidad aleatorias. Pero diferenciar en (10.5) para eliminar las respectivas componentes determinísticas puede ser incorrecto ya que estas componentes son básicas para calcular los pronósticos. Este hecho llama la atención sobre la necesidad de realizar pruebas de hipótesis que permitan decidir entre ambos modelos cuando hay evidencia de estacionalidad.

**Ejemplo 10.1.2.** Suponga un modelo  $SARIMA(1, 1, 0)(1, 1, 0)[12]$ , es decir,  $s = 12$ ,  $p = 1, q = 0, p_s = 1, q_s = 0, d = D = 1$ , dado por

$$(1 - \varphi_1 L)(1 - \varphi_{12} L^{12})(1 - L^{12})(1 - L)Y_t = \varepsilon_t,$$

Colocando  $X_t = (1 - L)(1 - L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$  se obtiene, de manera equivalente,  $(1 - \varphi_1 L)(1 - \varphi_{12} L^{12})X_t = \varepsilon_t$ . Cuando se desarrollan los polinomios autoregresivos del modelo se obtiene

$$(1 - \varphi_1 L - \varphi_{12} L^{12} + \varphi_1 \varphi_{12} L^{13})X_t = \varepsilon_t,$$

ó también,  $X_t = \varphi_1 X_{t-1} - \varphi_{12} X_{t-12} + \varphi_1 \varphi_{12} X_{t-13} + \varepsilon_t$ .

Note que este modelo es un  $X_n \sim AR(13)$  con restricciones sobre ciertos coeficientes que se asumen cero. Se podría indicar por  $AR([1, 12, 13])$ . El parámetro  $\varphi_1 \varphi_{12}$  se estima como un solo valor  $\varphi_{13}$  y puede ser no significativo. En el caso que sea no significativo el modelo se dice sin interacción, y se expresa como

$$(1 - \varphi_1 L - \varphi_{12} L^{12})X_t = \varepsilon_t.$$

Considere los parámetros  $\varphi_1 = 0.8$ ,  $\varphi_{12} = -0.137$ ,  $\sigma^2 = 2$ . El siguiente código en R utiliza la función `sarima.Sim()` de la librería *CombMSC*, ver Smith [2019], para simular trayectorias de modelos SARIMA y así visualizar este modelo.

```
# simulacion sarima(1,1,0)(1,1,0)[12] con la libreria CombMSC
library(CombMSC)
library(forecast)
library(TSA)

fi=0.8; fi12=0.737;
sigma= sqrt(2)
y = sarima.Sim(n = 20, period =12,
model=list(order=c(1,1,0),ar = fi,
ma = NULL, sd = sigma),
seasonal=list(order=c(1,1,0),ar=fi12,ma = NULL),
rand.Gen.Fun = rnorm, rand.Gen.Seas = rnorm)

#-----note: longitud de la serie: 20*12=240
```

```

dyls = diff(diff(y,12,1),1,1)
dys = diff(y,12,1)
dyl = diff(y,1,1)

```

```

par(mfrow = c(2,2))
ts.plot(y,main="(A) ")
ts.plot(dyl,main="(B) ")
ts.plot(dys,main="(C) ")
ts.plot(dyls,main="(D) ")

```

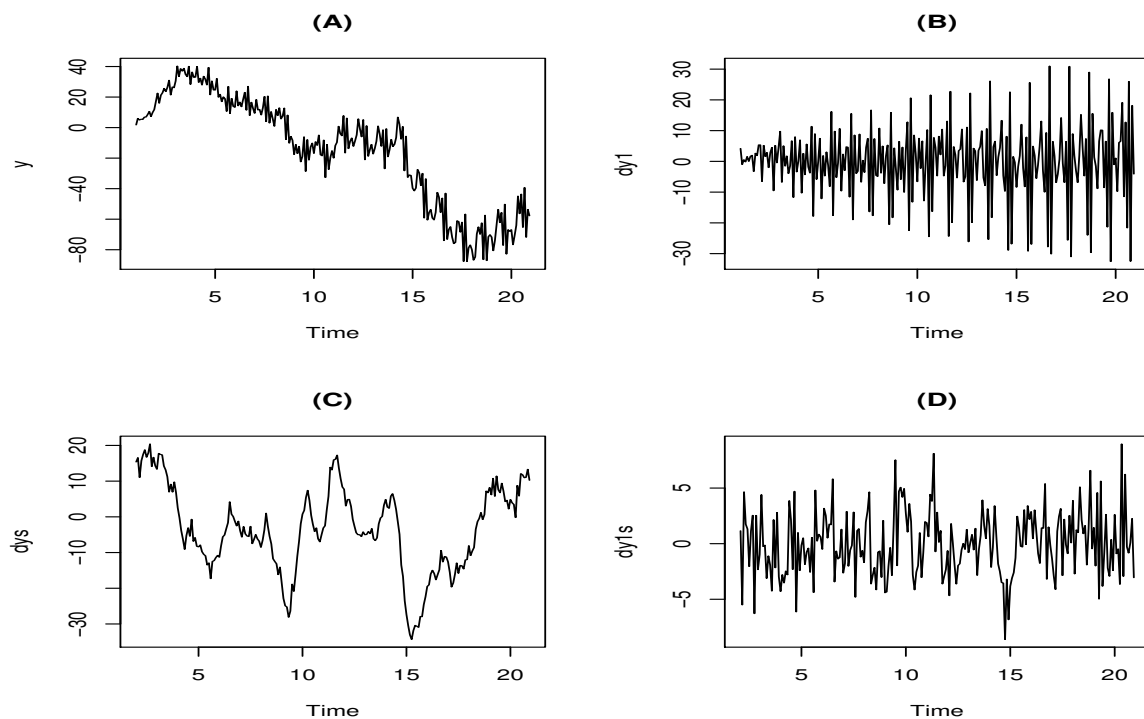


Figura 10.2: (A) Trayectoria de un modelo SARIMA(1,1,0)(1,1,0)[12], del Ejemplo 10.1.2, con (B) diferencia ordinaria, (C) diferencia estacional, (D) ambas diferencias

**Ejemplo 10.1.3. Modelo “Air-passengers”** Se define como un modelo multiplicativo de la forma SARIMA(0,1,1)(0,1,1)[12] dado por

$$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1)(1 + \theta_{12}L^{12})\varepsilon_t, \quad |\theta_1| < 1, |\theta_{12}| < 1. \quad (10.6)$$

*Es un modelo utilizado con frecuencia para modelar series con tendencia lineal y componente estacional. Por ejemplo la serie “airline”, del texto de Box y Jenkins: “Monthly totals of international airline passengers, 1949 to 1960”. En R se puede visualizar con*

```
data("AirPassengers")
AP = AirPassengers
ts.plot(AP)
```

### Nota sobre la identificación de este modelo

*Al utilizar  $W_t = (1 - L)(1 - L^{12})Y_t$  y examinar la fac de  $W_t$  se observa que  $W_t = (1 + \theta_1 L)(1 + \theta_{12} L^{12})\varepsilon_t$  es un  $MA(13)$ . Se puede comprobar que la fac de  $W_t$  tiene valores diferentes de cero sólo en los rezagos 1,12,13. La función de autocorrelación de  $W_t$  es*

$$\begin{aligned}\sigma^2 &= (1 + \theta_1^2)(1 + \theta_{12}^2) \\ \rho_1 &= \theta_1(1 + \theta_{12}^2) \\ \rho_{12} &= \theta_{12}(1 + \theta_1^2) \\ \rho_{13} &= \theta_1\theta_{12}\end{aligned}$$

*Suponga el modelo air-passengers con parámetros  $\theta_1 = -0.377$ ,  $\theta_{12} = -0.572$ ,  $\sigma^2 = 0.0014$ , el siguiente código simula una trayectoria de este modelo. El resultado está en la Figura 10.3. Nótese que  $\rho_1, \rho_{12} < 0$  y  $\rho_{13} > 0$ .*

### Código R 10.1.2.

```
#----- simular el modelo airline
ti=-0.377; ti12=-0.572;
sigma= sqrt(0.0014)
library(CombMSC)
y = sarima.Sim(n = 20, period =12,
model=list(order=c(0,1,1),ma = ti,
ar = NULL, sd = sigma),
seasonal=list(order=c(0,1,1),ma=ti12,ar = NULL),
rand.Gen.Fun = rnorm, rand.Gen.Seas = rnorm)
```

```

Wt = diff(diff(y,12,1),1,1)
par(mfrow=c(2,2))
ts.plot(y)
ts.plot(Wt)
acf(Wt,48,drop.lag.0=TRUE,ci.type="ma")
pacf(Wt,48)

# identificar
auto.arima(dy)
ARIMA(0,0,1)(0,0,1)[12] with zero mean

# identificador armasubset() libreria TSA
res=armasubsets(y=dy,
nar=14,nma=14,
y.name='dy',
ar.method='ols')
layout(1:1)
plot(res)

# estimar
m1 = arima(y,order=c(0,1,1),
seasonal=list(order=c(0,1,1),period=12))
summary(m1)

# pronosticar
pr = predict(m1,n.ahead=12)$pred

```



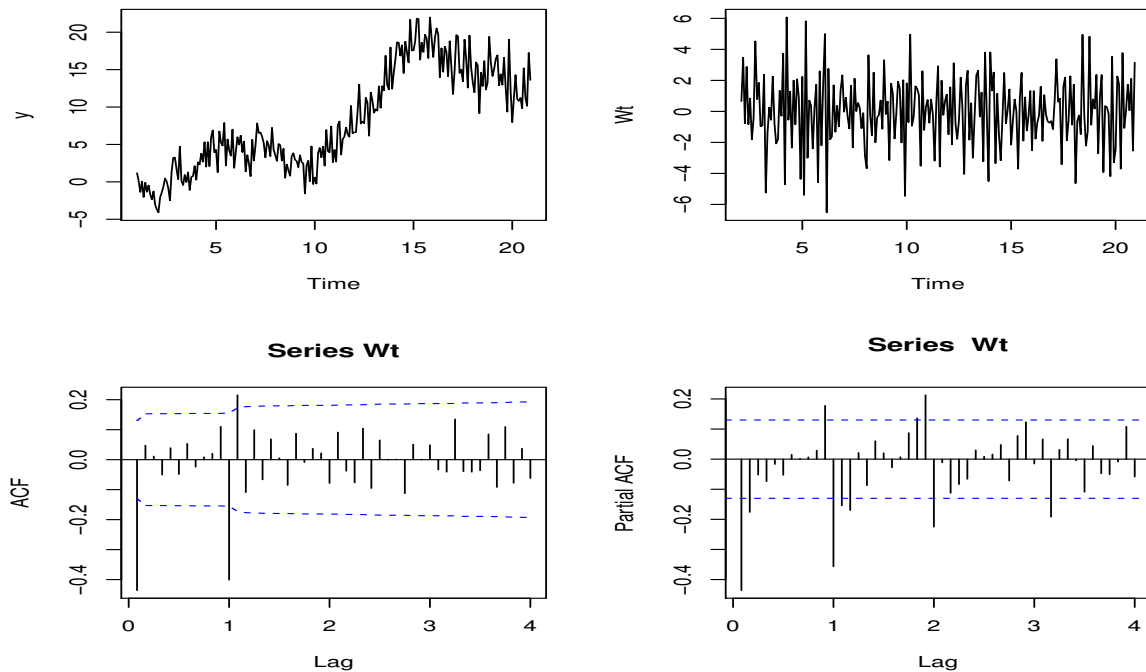


Figura 10.3: Trayectoria de un modelo SARIMA(0,1,1)(0,1,1)[12], serie doblemente diferenciada y su acf y pacf. Nótese que las autocorrelaciones en los rezagos 1,12,13 son significativas, con el signo correcto

**Ejercicio 10.1.1.** Utilice el código de los ejemplos anteriores para simular y estimar un modelo SARIMA  $SARIMA(1, 1, 1)(1, 1, 1)[4]$  definido por

$$(1 - \varphi_1 L)(1 - \varphi_4 L^4)(1 - L)(1 - L^4)Y_t = (1 - \theta_1 L)(1 - \theta_4 L^4)\varepsilon_t.$$

Utilice el código de los ejemplos anteriores para simular y estimar un modelo SARIMA con la serie  $Y_t$ . Es equivalente a estimar un modelo ARMA(1,1)(1,1) con la serie diferenciada,  $W_t = (1 - L)(1 - L^4)Y_t$ ? Use los siguientes valores para los parámetros.

```
fi=-0.8; fi4=-0.37; ti=-0.64; ti4=-0.513; sigma= sqrt(0.014)
```

## 10.2. Pruebas de Raíz Unitaria Estacional

Para introducir la idea de las pruebas de raíz unitaria estacional hay que recordar que una serie  $Y_t \sim ARMA(p, q)$  se dice integrada, ó con una raíz unitaria si se cumple que el polinomio autorregresivo  $\Phi_p(z)$  tiene una raíz igual a 1. Es decir, si se cumple que  $\Phi_p(1) = 0$ .

**Definición 10.2.1.** Una serie de tiempo  $Y_t \sim SARMA(p, q)(p_s, q_s)[s]$  se dice integrada estacional, de período  $s > 1$ , si se cumple que algunas ó todas las raíces de  $z^s = 1$  son raíces de la ecuación del polinomio autorregresivo conjunto  $\Phi_p(z)\Phi_{p_s}(z) = 0$ .

Las raíces de  $z^s = 1$  se denominan las raíces de la unidad, y son  $s$  números complejos. Con  $s = 4, 12$ , por ejemplo.

Para el caso  $s = 4$  se tiene que las cuatro raíces de la unidad son  $1, -1, i, -i$ , donde  $i = \sqrt{-1}$  es la unidad imaginaria.

Para el caso  $s = 12$ , las raíces son

$$\pm 1, \pm i, -(1 \pm i\sqrt{3})/2, (1 \pm i\sqrt{3})/2, -(\sqrt{3} \pm i)/2, (\sqrt{3} \pm i)/2.$$

Con la representación  $z = re^{i\theta}$  y la fórmula de Euler  $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ , donde  $r = |z|$  y  $\theta$  es el ángulo que forma  $|z|$  con el eje X, se pueden identificar las raíces unitarias con los ángulos  $\theta = 2\pi j/s, j = 0, 1, \dots, s-1$ , ya que se tiene  $r = 1$  en todos los casos, pues las raíces unitarias tienen módulo 1. Estos ángulos se denominan las “frecuencias estacionales”. Por ejemplo, para el caso  $s = 12$  se identifican con los ángulos  $\theta_j = 2\pi j/12, j = 0, 1, \dots, 11$

Se han desarrollado pruebas para detectar raíces unitarias estacionales, como las pruebas HEGY, ver Hylleberg et al. [1990], y la prueba Canova-Hansen, indicada por CH, ver Canova and Hansen [1995], que se exponen a continuación,

## 10.3. La prueba Canova-Hansen

Los modelos de componentes estructurales con componente estacional y (posibles) errores SARMA y los SARIMA pueden considerarse, inicialmente, equivalentes. Pero, como señalan Canova and Hansen [1995, pag. 238],

“Es difícil saber apriori cuál posibilidad [modelo] produce la mejor descripción de la serie. Algunas series muestran cambios en los patrones estacionales, por ejemplo, el consumo de energía eléctrica, las series de consumo e ingresos, la serie del producto interno bruto”.

En Ghysels and Osborn [2001, pag. 29] se anota:

"[...] for the sample sizes often observed in practice, it may be difficult to discriminate between deterministic seasonality and a seasonal unit root process."

La prueba CH se basa en esta observación. La inestabilidad estructural de la componente estacional se toma como una característica equivalente a la existencia de una raíz unitaria estacional, opuesta a Componente estacional más residuo estacionario estacional (SARMA).

Ghysels and Osborn [2001, pag. 42]:

"It is now well known that a series generated by a unit root process can wander widely and smoothly over time without any inherent tendency to return to its underlying mean value [...]. In the seasonal context, there are  $S$  unit root processes, none of which has an inherent tendency to return to a deterministic pattern. As a result, the values for the seasons can wander widely and smoothly in relation to each other [...]"

### Descripción de la Prueba

El modelo que asume la prueba es un modelo sin tendencia y con componente estacional, descrita mediante funciones trigonométricas, con el primer rezago de  $y$  como variable exógena. Concretamente, la ecuación (29) de la página 243 en el artículo de Canova and Hansen [1995], es

$$Y_t = bY_{t-1} + \sum_{j=1}^{s/2} \gamma_j \cos\left(\frac{2\pi jt}{s}\right) + \gamma_j^* \sin\left(\frac{2\pi jt}{s}\right), \quad (10.7)$$

donde se asume  $s$  es par,  $t = 1, 2, \dots$ . Nótese que cuando  $j = s/2$  entonces  $\sin\left(\frac{2\pi jt}{s}\right) = \sin(\pi t) \equiv 0$ . Luego, el número de parámetros en el modelo (10.7) es

$s - 1$ , el mismo del modelo con variables indicadoras. Los coeficientes se colocan en vectores

$$\underline{\gamma} = (\gamma_1, \dots, \gamma_{s/2})',$$

$$\underline{\gamma}^* = (\gamma_1^*, \dots, \gamma_{s/2-1}^*)'.$$

La hipótesis nula en la prueba CH es que los coeficientes  $\underline{\gamma}, \underline{\gamma}^*$  no cambian con  $t$  versus que varían con  $t$ , según modelos de marcha aleatoria:

$$\underline{\gamma}_t = \underline{\gamma}_{t-1} + \underline{u}_t, \quad (10.8)$$

$$\underline{\gamma}_t^* = \underline{\gamma}_{t-1}^* + \underline{v}_t, \quad (10.9)$$

donde  $\underline{u}_t, \underline{v}_t$  son vectores iid, independientes de media cero y varianzas no nulas  $\sigma^2, \sigma_*^2$ . Esta hipótesis alterna es la forma en la que se establece la presencia de raíces unitarias estacionales, equivalente a inestabilidad estructural en la componente estacional. Las hipótesis de la prueba se pueden escribir como sigue.

$$H_0 : \quad \underline{\gamma}_t \equiv \underline{\gamma}, \underline{\gamma}_t^* \equiv \underline{\gamma}^*,$$

$$H_a : \quad \underline{\gamma}_t = \underline{\gamma}_{t-1} + \underline{u}_t, \underline{\gamma}_t^* = \underline{\gamma}_{t-1}^* + \underline{v}_t.$$

En Canova and Hansen [1995, pag. 240, sec. 2.2], se re-escribe la prueba con base en cierto parámetro  $\tau^2$ , colocando  $H_0 : \tau^2 = 0$ , versus  $H_a : \tau^2 > 0$ . El estadístico de la prueba es del tipo Multiplicador de Lagrange, de la forma

$$LM = \frac{1}{T^2} tr \left( (A' \Omega A)^{-1} A' \left( \sum_{j=1}^T \underline{F}_j \underline{F}_j' \right) A \right), \quad (10.10)$$

donde  $A, \Omega, F_j$  son matrices,  $tr()$  es la traza de una matriz,  $T$  es la longitud de la serie, pero las definiciones específicas no se considerarán aquí y se remite al artículo original. Además, en el artículo se proveen los valores críticos para varios niveles de significación. La nula se rechaza si el valor observado del estadístico es mayor que el valor crítico

$$LM_{obs} > LM_{\alpha}.$$

## Implementación en R

La prueba Canova-Hansen está implementada en la librería `uroot`, con la función `CH.test`, que tiene los parámetros siguientes

```
ch.test(x, type = c("dummy", "trigonometric"),
lag1 = FALSE, NW.order = NULL, sid = NULL,
xreg = NULL, pvalue = c("RS", "raw"),
rs.nobsreg = 13),
```

la serie debe estar en la variable  $x$ , y debe haberse declarado como objeto “ts”, con `frequency=4, 12`, solamente.

El parámetro `type = c("dummy", "trigonometric")` indica si se utilizan variables trigonométricas ó indicadoras para modelar la estacionalidad.

El parámetro `lag1` es variable lógica. Si `TRUE` se incluye el primer rezago de la serie como variable explicativa en la regresión. Por defecto se coloca `FALSE`.

El parámetro `NW.order` es un entero que determina cuántos rezagos se utilizan para la estimación de la matriz de covarianza en la fórmula de Newey y West.

El parámetro `sid` es un vector numérico optional que especifica cuáles frecuencias deberían incluirse en la prueba. Por ejemplo, para  $s = 4$  hay 4 frecuencias que corresponden a los ángulos  $0, \pi/2, \pi, 3\pi/2$ . Por defecto se escogen todas las frecuencias.

El parámetro `xreg` es un vector ó matrix opcional de variables exógenas con el mismo número de filas que la serie  $x$ .

Nótese, sin embargo, que para aplicar la prueba CH es necesario eliminar la tendencia lineal de la serie  $Y_t$ , en caso de existir. Para aplicar la prueba se prefirió filtrar previamente esta tendencia, colocando  $Y_t - \hat{T}_t = \hat{S}_t + \hat{\varepsilon}_t$ , donde  $\hat{T}_t$  se estimó mediante el filtro `stl()` ó Loess.

**Ejemplo 10.3.1.** *Retomando el Ejemplo (7.3.2) en la pag. 141, para modelar de la serie de producción de cemento Portland, trimestral, y decidir entre los modelos de tendencia lineal con estacionalidad, con errores MA(6) y un posible SARIMA, se aplica la prueba Canova-Hansen para determinar la posibilidad de una raíz unitaria estacional que apoye un posible modelo SARIMA.*

*La prueba se implementa con los comandos siguientes. Nótese que se utilizó el procedimiento `stl()` para extraer la tendencia de la serie.*

**Código R 10.3.1.**

```

#-----Ejemplo prueba Canova-Hansen
E = read.table("cementq.dat", header = TRUE)
attach(E)
y = ts(y,frequency=4,start=c(1956,1),end=c(1994,3))
ts.plot(y)
#-----eliminar la tendencia con stl()
m1 = stl(y, s.window = 'per')
s1 = m1$time.series[,1]; e1 = m1$time.series[,3];
y1 = s1+e1
#-----implementacion de la prueba Canova-Hansen
require(uroot)
res = ch.test(y1,
type = "trigonometric", pvalue='raw')
res
#-----resultado
data:  y1
      statistic pvalue
pi/2      2.4867   0.01 **
pi         1.0541   0.01 **
joint      2.761   0.01 **
Test type: seasonal cycles
NW covariance matrix lag order: 4
First order lag: no
Other regressors: no
P-values: interpolation in original tables

```

Como se detecta una raíz unitaria estacional al nivel de 5 % en dos de las frecuencias de un total de dos (período 4), se rechaza la nula y existe evidencia de inestabilidad estructural en la componente estacional. Por tanto, el modelo SARIMA sería recomendable para modelar la serie.

## 10.4. La prueba HEGY

La prueba HEGY (Hylleberg, Engle, Granger, Yoo), en Hylleberg et al. [1990], es una generalización de la prueba aumentada de Dickey-Fuller. Permite decidir cuáles de las frecuencias estacionales corresponden a raíces estacionales significativas. En caso de detectarse al menos una, exceptuando la frecuencia que corresponde a cero, se concluirá que la serie en cuestión está integrada estacionalmente, que es la hipótesis nula.

Esta sección describe la idea básica de la prueba, que utiliza una regresión lineal múltiple. Los estadísticos para detectar las raíces unitarias estacionales son tipo t-Student y F, correspondientes a los estadísticos para pruebas de hipótesis sobre los parámetros de la regresión. Esta sección es con base en el artículo de Hylleberg et al. [1990].

Inicialmente se asume que la serie  $Y_t$  tiene una componente estacional de período  $s$  par, y se puede aproximar por un proceso AR( $s$ ), tal que

$$\Phi_s(L)Y_t = \varepsilon_t,$$

para  $\varepsilon_t$  ruido blanco. Se asume las  $s$  raíces estacionales como los números complejos  $\theta_j \in \mathbb{C}$ ,  $j = 1, \dots, s$ , en los cuales  $\Phi_s(\theta_j)$  es un número finito. Adicionalmente se definen las cantidades auxiliares:

$$\begin{aligned} \delta_k(z) &= 1 - z/\theta_k, \\ \Delta(z) &= \prod_{k=1}^s \delta_k(z), \\ \lambda_k &= \Phi_s(\theta_k) / \prod_{j \neq k} \delta_j(\theta_k). \end{aligned} \quad (10.11)$$

Entonces se cumple, por una identidad debida a Lagrange, que

$$\Phi_s(z) = \sum_{k=1}^s \frac{\lambda_k \Delta(z)(1 + \delta_k(z))}{\delta_k(z)} + \Delta(z)\Phi_s^*(z), \quad (10.12)$$

donde  $\Phi_s^*(z)$  es un polinomio posiblemente infinito ó racional. Nótese que, por la definición (10.11) se cumple

$$\Phi_s(\theta_k) = 0 \Leftrightarrow \lambda_k = 0 \quad (10.13)$$

por lo tanto, para determinar si  $\theta_k$  es raíz estacional se debe comprobar  $\lambda_k = 0$ . Pero estas  $\lambda_k$  se pueden asimilar a los coeficientes en una regresión lineal múltiple que se define a partir de la identidad (10.12). Y por tanto, probar  $\lambda_k = 0$  se puede hacer, en principio, mediante una prueba t-Student de significación de parámetros, ó con una prueba F.

Por ejemplo, para el caso  $\Phi_s(z) = 1 - z^4$ , donde  $\theta_1 = 1, \theta_2 = -1, \theta_3 = i, \theta_4 = -i$ , la identidad (10.12) se transforma en

$$\begin{aligned}\Phi_s(z) &= \lambda_1 z(1+z)(1+z^2) + \lambda_2 (-z)(1-z)(1+z^2) \\ &+ \lambda_3 (-iz)(1-z)(1+z)(1-iz) + \lambda_4 (iz)(1-z)(1+z)(1+iz) \\ &+ \Phi_s^*(z)(1-z^4).\end{aligned}\quad (10.14)$$

Si se definen nuevas constantes  $\pi_1, \pi_2, \pi_3, \pi_4$  mediante las relaciones:

$$\pi_1 = -\lambda_1, \pi_2 = -\lambda_2, 2\lambda_3 = -\pi_3 + i\pi_4, 2\lambda_4 = -\pi_3 - i\pi_4$$

entonces la identidad (10.14) queda

$$\begin{aligned}\Phi_s(z) &= -\pi_1 z(1+z+z^2+z^3) - \pi_2 (-z)(1-z+z^2-z^3) \\ &- (\pi_4 + \pi_3 z)(-z)(1-z^2) + \Phi_s^*(z)(1-z^4).\end{aligned}\quad (10.15)$$

Finalmente, en  $\Phi_s(L)Y_t = \varepsilon_t$  se reemplaza  $\Phi_s(L)$  por la correspondiente expresión obtenida en (10.15). Si se definen las variables

$$\begin{aligned}X_{1,t} &= (1 + L + L^2 + L^3)Y_t, \\ X_{2,t} &= -(1 - L + L^2 - L^3)Y_t, \\ X_{3,t} &= -(1 - L^2)Y_t, \\ X_{4,t} &= (1 - L^4)Y_t,\end{aligned}$$

y se reemplaza  $\Phi_s^*(L) = 1$ , se obtiene la ecuación

$$X_{4,t} = \pi_1 X_{1,t} + \pi_2 X_{2,t} + \pi_3 X_{3,t} + \varepsilon_t. \quad (10.16)$$

La ecuación (10.16) es una regresión lineal múltiple. A partir de esta ecuación se plantea la prueba de hipótesis HEGY para el caso de período  $s = 4$ . La hipótesis en la prueba HEGY, para el caso  $s = 4$ , se compone de tres hipótesis, para las raíces unitarias estacionales:  $\pm 1, \pm i$ , definidas de la manera siguiente.

$$H_0 : \Phi_s(1) = 0 \Leftrightarrow \pi_1 = 0, \quad H_a : \Phi_s(1) > 0 \Leftrightarrow \pi_1 < 0,$$



$$H_0 : \Phi_s(-1) = 0 \Leftrightarrow \pi_2 = 0, \quad H_a : \Phi_s(2) > 0 \Leftrightarrow \pi_2 < 0,$$

$$H_0 : |\Phi_s(i)| = 0 \Leftrightarrow \pi_3 = \pi_4 = 0, \quad H_a : no(H_0).$$

Si alguna hipótesis nula no se rechaza se asume que existe raíz unitaria estacional, excepto la raíz  $z = 1$  que corresponde a una raíz unitaria para tendencia aleatoria y es equivalente a la prueba Dickey-Fuller.

Los estadísticos de las pruebas son t-Student para  $z = 1, -1$ , y F para  $\pm i$ . Las decisiones se toman con base en los valores p correspondientes.

En el caso  $s = 12$ , hay 12 raíces unitarias,  $z_j = \cos(2\pi j/12) + i \sin(2\pi j/12)$ ,  $j = 0, 1, \dots, 11$ .

Ver un ejemplo de código en R para calcular las raíces y graficarlas en: <sup>(1)</sup>.

Hay 7 hipótesis, dos para  $z = 1, -1$ , y una para cada una de las cinco parejas de raíces unitarias estacionales conjugadas.

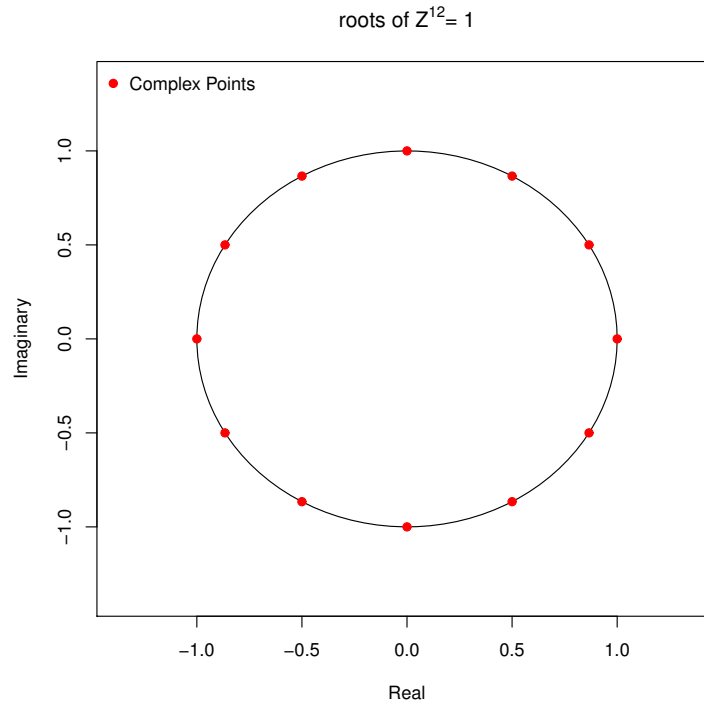


Figura 10.4: Raíces de la unidad,  $z^{12} = 1$

<sup>1</sup><https://codereview.stackexchange.com/questions/132959/roots-of-unity-in-r>

No rechazar alguna de las hipótesis nulas equivale a aceptar que existe una raíz unitaria estacional.

En caso de no rechazar una de las hipótesis nula, cabría esperar que el modelo adecuado para la serie sea de tipo SARIMA integrado, es decir, después de diferenciar la serie se obtiene un proceso ARMA estacionario.

En caso de rechazar todas las hipótesis nulas la serie no tiene raíces unitarias estacionales y entonces se clasifica como una serie “estacionaria estacional”, es decir, se puede modelar como un modelo de tendencia lineal más componentes estacionales con residuos posiblemente SARMA.

Hay que tener en cuenta la potencia de la prueba:  $\text{Prob}(\text{rechazar } H_0 \mid H_1 \text{ es cierta})$ , una baja potencia significa que la prueba no es capaz de detectar la alterna cuando ésta es cierta, o también, que la probabilidad de no rechazar  $H_0$  cuando es falsa, es alta.

### Implementación en R

La prueba HEGY está implementada en la función `hegy.test`, de la librería `uroot`, ver López-de LaCalle [2017]. Ejemplo:

```
hegy.test(x, deterministic = c(1, 0, 0),
lag.method = "fixed", maxlag = 1,
pvalue = "raw")
```

- la serie debe estar en la variable  $x$ , y debe haberse declarado como objeto “ts”.
- El parámetro `deterministic=c(1, 1, 1)` se refiere a si se desea incluir en el modelo un intercepto, tendencia lineal o componente estacional con variables indicadoras. El modelo se define en este caso como

$$\Phi_s(L)[Y_t - (\beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_j(t))] = \varepsilon_t. \quad (10.17)$$

En caso contrario se coloca `deterministic=c(1, 0, 0)`

- El parámetro `pvalue`, según la ayuda:

a character specifying the method employed to compute

p-values: "RS", the default, interpolation based on response surface regressions; "bootstrap", bootstrap; "raw" interpolation in the tables provided in the reference papers.

**Ejemplo 10.4.1.** *Retomando el Ejemplo (10.3.1) en la pag. 255, para modelación de la serie de producción de cemento Portland trimestral, donde se aplicó la prueba Canova-Hansen, se va a aplicar ahora la prueba HEGY con el mismo objetivo de determinar si el modelo SARIMA es adecuado. Los comandos en R son los siguientes.*

```
# prueba hegy con intercepto, tendencia y variables indicadoras.
require(uroot)
hegy.out1 = hegy.test(x=y,
deterministic = c(1,1,1),
lag.method = "fixed", maxlag = 1)
#----resultado
statistic p-value
t_1      -4.153  0.0046 **
t_2      -3.1308 0.0206 *
F_3:4    14.9083      0 ***
F_2:4     12.533      0 ***
F_1:4    15.6967      0 ***
Deterministic terms: constant + trend + seasonal dummies
Lag selection criterion and order: fixed, 1
P-values: based on response surface regressions
```

El resultado anterior muestra que se rechaza la hipótesis nula de raíces unitarias en  $1, -1, \pm i$ , indicadas por  $t_{pi\_1}$ ,  $t_{pi\_2}$ ,  $F_{pi\_3:4}$ . La última corresponde a las raíces en las frecuencias  $\pi/2, 3\pi/2$ , la primera corresponde a la frecuencia cero, y es una raíz unitaria para la tendencia, similar a la prueba Dickey-Fuller. Luego, no sería apropiado ajustar un modelo SARIMA a la serie.

Si se repite la prueba con la serie sin tendencia, rechaza nuevamente la hipótesis de raíces unitarias estacionales.

```
# prueba hegy sin tendencia.
```

```

require(uroot)
m1 = stl(y, s.window = 'per', t.window = 50, t.jump = 1)
t1 = m1$time.series[,2]
y1 = y-t1
hegy.out1 = hegy.test(x=y1,
deterministic = c(0,0,1),
lag.method = "fixed", maxlag = 1)
#-----resultado
      statistic p-value
t_1      -5.4194      0 ***
t_2      -3.064    0.0257 *
F_3:4     14.9748      0 ***
F_2:4     12.475      0 ***
F_1:4     19.6446      0 ***
Deterministic terms: constant + seasonal dummies
Lag selection criterion and order: fixed, 1
P-values: based on response surface regressions

```

## 10.5. Prueba OCSB

La prueba Osborn, Chui, Smith, Birchenhall para raíz unitaria estacional, ver Osborn et al. [1988] y Ghysels and Osborn [2001] pag. 66. La prueba se basa en la regresión

$$\Delta \Delta_s Y_t = \beta_1 \Delta_s Y_{t-1} + \beta_2 \Delta Y_{t-s} + \epsilon_t, \quad (10.18)$$

con  $\Delta = I - L$ ,  $\Delta_s = I - L^s$ . Utilizando la identidad

$$I - L^s = (I - L)(I + L + \dots + L^{s-1}),$$

aplicando este operador a  $Y_{t-1}$  se tiene

$$\Delta_s Y_{t-1} = (I + L + \dots + L^{s-1}) \Delta Y_{t-1}$$

luego en (10.18)

$$\begin{aligned} \Delta_s \Delta Y_t &= \beta_1 \Delta_s Y_{t-1} + \beta_2 \Delta Y_{t-s} + \epsilon_t \\ \Leftrightarrow \Delta_s \Delta Y_t &= \beta_1 (I + L + \dots + L^{s-1}) \Delta Y_{t-1} + \beta_2 \Delta Y_{t-s} + \epsilon_t \end{aligned}$$

y todas las variables en la regresión (10.18) incluyen la primera diferencia  $\Delta$ . Por tanto, como afirman Ghysels and Osborn [2001] pag. 66:

An interpretation of the OCSB test, therefore, is that it tests seasonal integration in a process that has been first differenced.

En Rodrigues and Osborn [1999] se propone la regresión sin la primera diferencia  $\Delta$

$$\Delta_s Y_t = \beta_1(I + L + \dots + L^{s-1})Y_{t-1} + \beta_2 Y_{t-s} + \epsilon_t. \quad (10.19)$$

como modelo para la prueba de hipótesis nula de raíz unitaria estacional, incluyendo la raíz unitaria ordinaria. La expresión de la hipótesis nula y la alterna en Osborn et al. [1988], pag. 364 es

$$H_0 : Y_t \sim I(1, 1) \quad (10.20)$$

$$H_1 : Y_t \sim I(0, 0) \vee I(1, 0) \vee I(0, 1) \quad (10.21)$$

donde el símbolo  $I(1, 1)$  significa integrado ordinario y estacional, y los otros  $I(1, 0)$ ,  $I(0, 1)$ ,  $I(0, 0)$  se interpretan de manera similar. El símbolo  $\vee$  es la “o” lógica.

El estadístico de la prueba se define como el estadístico de la prueba ADF,  $t_{\hat{\beta}_2}$ , con respecto a la regresión (10.19). La hipótesis equivale en este caso a  $H_0 : \beta_2 = 0$ , vs.  $H_1 : \beta_2 < 0$ . Se rechaza la nula cuando  $t_{\hat{\beta}_2} < t_\alpha$ .

Además, el estadístico de la prueba ADF,  $t_{\hat{\beta}_1}$  se puede utilizar para la hipótesis de raíz unitaria ordinaria  $H_0 : Y_t \sim I(1)$ . La hipótesis equivale en este caso a  $H_0 : \beta_1 = 0$ , vs.  $H_1 : \beta_1 < 0$ .

La distribución asintótica de estos estadísticos bajo  $H_0$  no se incluyen aquí.

Pero ambas pruebas no pueden equipararse a la prueba Dickey-Fuller, porque, como Ghysels and Osborn [2001] pag. 66 observan, los valores críticos ADF no son estrictamente válidos para éstas, en consecuencia para la prueba OCSB se requieren valores críticos calculados por simulación.

La prueba OCSB está implementada en la librería `forecast`, con la función `ocsb.test`. Los valores críticos en esta implementación se calculan mediante simulación. Esta función calcula solamente  $t_{\hat{\beta}_2}$  y el valor crítico para el nivel de significación de 5 %.

**Ejemplo 10.5.1.** *Retomando los Ejemplos (10.3.1) y (10.4.1) en las pags. 255 y 261, para la modelación de la serie de producción de cemento Portland trimestral, donde se aplicaron las pruebas Canova-Hansen y HEGY. Se va a aplicar ahora la prueba OCSB con el mismo objetivo de determinar si el modelo SARIMA es adecuado.*

*Los comandos en R son los siguientes.*

```
#-----serie con tendencia y estacionalidad
require(forecast)
ocsb.test(y)
Test statistic: -2.092, 5% critical value: -1.8927

#-----decision: -2.092 < -1.8927, rechaza Ho
alternative hypothesis: stationary
#-----eliminar tendencia
m1 = stl(y, s.window = 'per', t.window = 50, t.jump = 1)
s1 = m1$time.series[,1]
t1 = m1$time.series[,2]
e1 = m1$time.series[,3]
y1 = y-t1
ocsb.test(y1)
Test statistic: -2.1177, 5% critical value: -1.8927
```

En el primer caso el estadístico observado es menor que el valor crítico por lo que rechaza la nula  $H_0 : Y_t \sim I(1, 1)$ . Y por tanto acepta una de las alternas  $H_1 : Y_t \sim I(0, 0) \vee I(1, 0) \vee I(0, 1)$ . Además, como se eliminó la tendencia previamente, esto excluye la alternativa  $I(1, 0)$ , y como se rechaza la nula que contiene la integración estacional, la alterna debe ser  $Y_t \sim I(0, 0)$ . Es decir, el modelo SARIMA no es adecuado y se debería utilizar el de Componentes.

## 10.6. Ejemplo de resumen

**Ejemplo 10.6.1.** *Retomando el Ejemplo (7.3.2) en la pag. 141, para modelación de la serie de producción de cemento Portland, trimestral, para decidir entre los*

*modelos de tendencia lineal con estacionalidad, con errores MA(6) y un posible SARIMA.*

$$1) Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_t(t) + \varepsilon_t, \varepsilon_t \sim MA(6), \quad (10.22)$$

$$2) Y_t = SARIMA(3, 1, 2)(1, 1, 2)[4]. \quad (10.23)$$

En la Figura 10.5 se puede ver la serie con y sin tendencia. Se observa que la componente estacional no es constante, por lo que es posible que se tenga un modelo SARIMA (10.23) en lugar del modelo de componentes (10.22). Al examinar la fac de

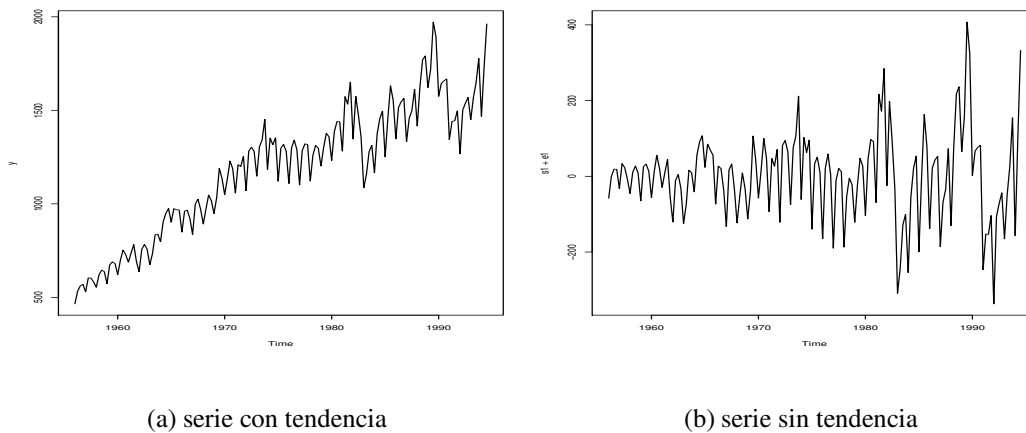


Figura 10.5: Serie de Producción Trimestral de Cemento, Australia

la serie diferenciada  $W_t = (1 - L)(1 - L^4)Y_t$ , para  $Y_t$  la serie excluyendo los últimos 8 datos para realizar validación cruzada, se observa posibles modelos ARMA, por ejemplo, con rezagos 1, 12 y 13 significativos, según la Figura 10.6.

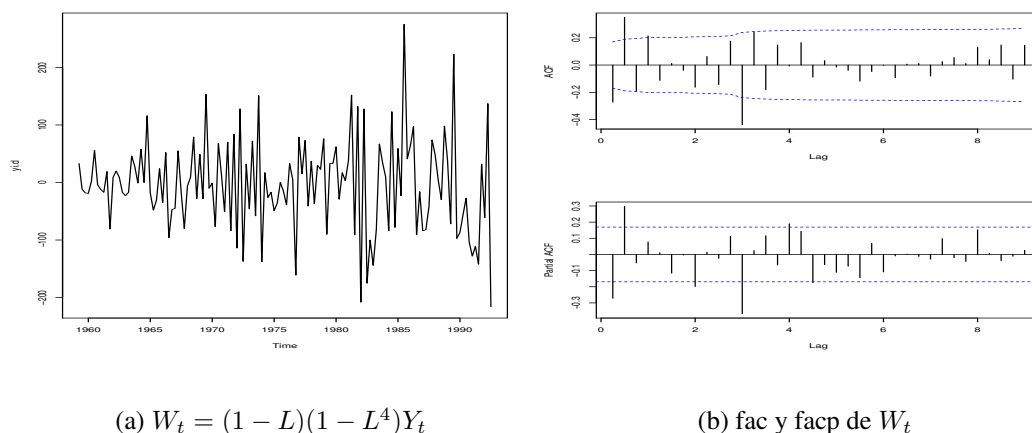


Figura 10.6: Serie Diferenciada de Producción Trimestral de Cemento

Mediante el uso de la función `armasubSets()` de la librería TSA, aplicada a la serie  $W_t$  se llega al modelo  $SARIMA(3, 1, 2)(1, 1, 2)[4]$ . Los parámetros estimados están en la Tabla 10.1.

Tabla 10.1: Parámetros estimados de  $SARIMA(3,1,2)(1,1,2)[4]$

	parametros	sd.dev	est t
ar1	0.55	0.30	1.85
ar2	0.57	0.20	2.82
ar3	-0.31	0.12	-2.60
ma1	-0.73	0.32	-2.28
ma2	-0.23	0.31	-0.73
sar1	0.90	0.25	3.56
sma1	-1.56	0.32	-4.83
sma2	0.67	0.18	3.80

La prueba Ljung-Box para los residuos de este modelo arroja el resultado  $X\text{-squared} = 34.672$ ,  $df = 26$ ,  $p\text{-value} = 0.1189$ , por lo que puede aceptarse este modelo para pronosticar. Los pronósticos con el modelo SARIMA versus los obtenidos con el modelo componentes (10.22) se pueden



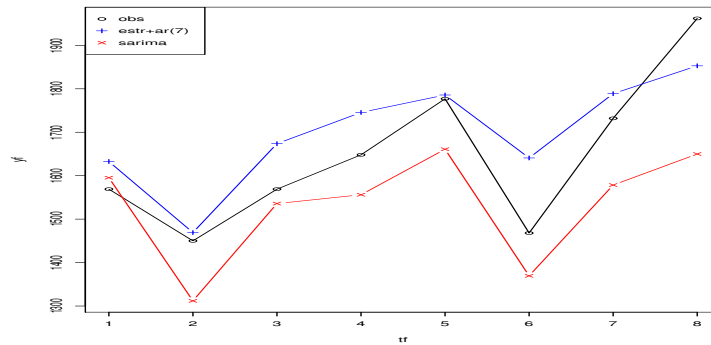


Figura 10.7: Pronósticos a 8 trimestres de la Producción Trimestral de Cemento

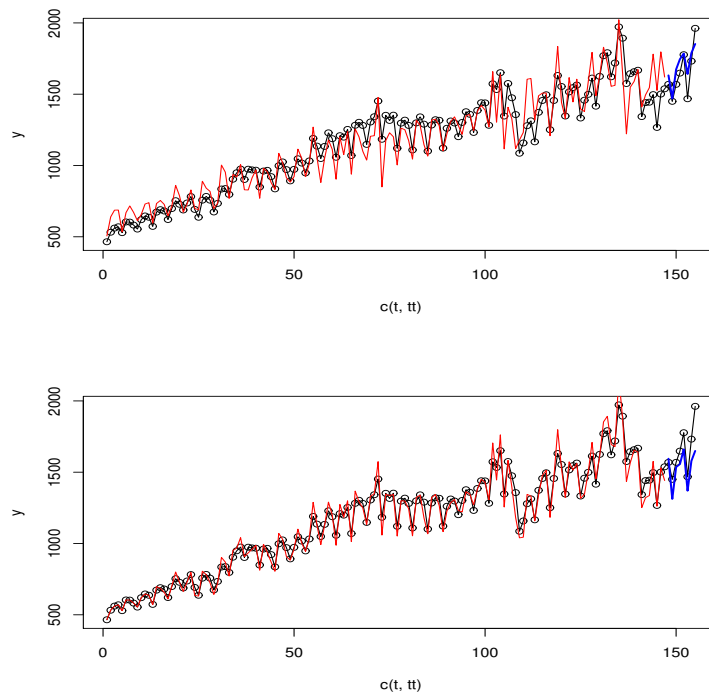


Figura 10.8: Comparación de los Ajustes y Pronósticos. Panel superior: Modelo Descomposicion + AR, Panel inferior: SARIMA

observar en la Figura 10.7. Se llegó a que el SARIMA no se detecta como modelo factible por las prueba CH, HEGY y OCSB. El MAPE de sus pronósticos, a 8 trimestres, es 7.89, mayor que el del modelo de Descomposición con errores MA(6),

que es 4.64. Sin embargo, el modelo SARIMA tiene mejor ajuste dentro de la muestra. En la Figura 10.8 se puede comparar tanto los ajustes dentro de la muestra como los pronósticos.

Cuál modelo escoger?. Desde el punto de vista de la validación cruzada a 8 trimestres, el modelo de Descomposición. Desde el punto de vista del ajuste dentro de la muestra, el modelo SARIMA.