

---

## Raíces Unitarias Estacionales y Estacionalidad Estocástica (SARIMA)

---

### 8.1. Modelos SARIMA

Si la serie  $Y_t$  tiene una componente con período  $s$  es posible eliminarla diferenciando con un rezago de orden  $s$ , es decir, transformando  $Y_t$  a

$$W_t = (1 - L^s)^D Y_t = \Delta_s^D Y_t, \quad D = 0, 1, 2. \quad (8.1)$$

Y buscando una estructura *ARMA* para  $W_t$ .

#### 8.1.1. Modelo Multiplicativo SARIMA

Sin embargo, si hay estacionalidad, pueden existir estructuras *ARIMA* *intra* período y *entre* los  $s$  períodos. Las *ARIMA* *intra* se refieren a modelos de la forma

$$\varphi_p(L) \Delta_s^d Y_t = \theta_q(L) \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2). \quad (8.2)$$

Los *ARIMA* *entre* se refieren a modelos  $ARIMA(P, D, Q)_s$  de la forma

$$\Phi_P(L^s) \Delta_s^D Y_t = \Theta_Q(L^s) \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2). \quad (8.3)$$

Y se define el modelo  $SARIMA(p, d, q)(P, D, Q)_s$  como

$$\varphi_p(L)\Phi_P(L^s)\Delta^d\Delta_s^D Y_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t, \quad (8.4)$$

Donde  $X_t = \Delta^d\Delta_s^D Y_t = (1-L)^d(1-L^s)^D Y_t$  es la transformación (filtro lineal) que elimina la tendencia y la estacionalidad dejando una estructura  $ARMA(p+Ps, q+Qs)$ .

La idea es que el modelo

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

es similar o equivalente al modelo

$$Y_t = a + bt + \sum_{j=1}^{s-1} \delta_j I_j(t) + X_t \quad (8.5)$$

En ambos la operación inicial de calcular la serie diferenciada  $X_t$  es similar a eliminar una tendencia y estacionalidad aleatorias en (8.4) y determinísticas en (8.5), y luego ajustar un modelo ARMA a los residuos estructurales.

**Ejemplo 8.1.1.** Suponga un modelo  $SARIMA(1, 1, 0)(1, 1, 0)_{12}$ , es decir,  $p = 1, d = 1, q = 0, P = 1, D = 1, Q = 0, s = 12$ .

$$\begin{aligned} (1 - \varphi_1 L)(1 - \varphi_{12} L^{12})(1 - L)(1 - L^{12})(Y_t - \mu) &= \varepsilon_t \\ (1 - \varphi_1 L)(1 - \varphi_{12} L^{12})X_t &= \varepsilon_t \end{aligned}$$

con

$$X_t = (1 - L)(1 - L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}.$$

Cuando se desarrollan los polinomios autoregresivos del modelo se obtiene

$$\begin{aligned} (1 - \varphi_1 L - \varphi_{12} L^{12} + \varphi_1 \varphi_{12} L^{13})X_t &= \varepsilon_t \\ X_t &= \varphi_1 X_{t-1} - \varphi_{12} X_{t-12} + \varphi_1 \varphi_{12} X_{t-13} + \varepsilon_t. \end{aligned}$$

Note que este modelo se puede considerar como un  $AR(13)$  con restricciones sobre los coeficientes. 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, es decir

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

**Ejemplo 8.1.2.** Sea el modelo sin interacción del ejemplo anterior, con parámetros  $\varphi_1 = 0.8, \varphi_{12} = -0.137, \sigma^2 = 2$

```

library(CombMSC)
library(fArma)
library(forecast)

Mod(polyroot(c(1, -0.8, rep(0, 10))))

Y = sarima.Sim(n = 144, period = 12, model=list(ar = 0.8,
  sd = sqrt(2)), seasonal=list(ar = -0.137),
  rand.Gen.Fun = rnorm, rand.Gen.Seas = rnorm)

ts.plot(Y)
modelo = auto.arima(Y)
summary(modelo)

```

### Modelo “Air-passengers”

Se define como un modelo multiplicativo de la forma  $SARIMA(0, 1, 1)(0, 1, 1)_{12}$  dado por

$$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1)(1 + \theta_{12}L^{12})\varepsilon_t, \quad |\theta_1| < 1, |\theta_{12}| < 1. \quad (8.6)$$

Es un modelo utilizado con frecuencia. Por ejemplo la series (benchmark) log(No. de pasajeros aéreos internacionales).

### 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, 11, 12, 13.

La función de autocovarianza de  $W_t$  es

$$\begin{aligned}
 \gamma_0 &= (1 + \theta_1^2)(1 + \theta_{12}^2)\sigma^2 \\
 \gamma_1 &= \theta_1(1 + \theta_{12}^2)\sigma^2 \\
 \gamma_{11} &= \theta_1\theta_{12}\sigma^2 \\
 \gamma_{12} &= \theta_{12}(1 + \theta_1^2)\sigma^2 \\
 \gamma_{13} &= \theta_1\theta_{12}\sigma^2
 \end{aligned}$$

**Ejemplo 8.1.3.** Suponga el modelo air-passengers con parámetros  $\theta_1 = -0.377$ ,  $\theta_{12} = -0.572$ ,  $\sigma^2 = 0.0014$

```

Y = sarima.Sim(n = 132, period =12, model=list(order=c(0,1,1),ma =
  -0.377, sd = sqrt(0.0014)), seasonal=list(order=c(0,1,1),ma = -0.572))

ts.plot(Y)

W = diff(diff(Y,1,1),1,1)
acf(W)

modelo = auto.arima(Y)
summary(modelo)

arima(Y,order=c(0,1,1),seasonal=list(order=c(0,1,1),period=12))

```

**Ejemplo 8.1.4.** Sea un proceso  $SARIMA(1,1,1)(1,1,1)_4$  definido por

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

```

library(CombMSC)
library(forecast)

fi=-0.8; fi4=-0.37; ti=-0.64; ti4=-0.513; sigma= sqrt(0.014)
Mod(polyroot(c(1,-fi,0,0,-fi4)))
Mod(polyroot(c(1,-ti,0,0,-ti4)))

Y = sarima.Sim(n = 60, period =4, model=list(order=c(1,1,1),ar = fi,
  ma = ti, sd = sigma), seasonal=list(order=c(1,1,1),ar=fi4,ma = ti4))

ts.plot(Y)
W = diff(diff(Y,1,1),4,1)
acf(W)
pacf(W)

arima(Y,order=c(1,1,1),seasonal=list(order=c(1,1,1),period=4))

auto.arima(Y)

```

**Ejemplo 8.1.5.** Sea un proceso  $SARIMA(1,1,1)(0,1,0)_{12}$  dado por

$$\Delta_{12}\Delta Y_t = \left( \frac{1 + \theta_1 L}{1 - \varphi_1 L} \right) \varepsilon_t,$$

$$X_t = (1 - L)(1 - L^{12}),$$

por tanto,  $X_t \sim ARMA(1, 1)$ , con  $|\varphi_1| < 1$ ,  $|\theta_1| < 1$ .

- *Cómo es una trayectoria de  $Y_t$ ?*
- *Cómo es la fac de  $X_t$ ?*

**Ejemplo 8.1.6.** Para la serie  $m^2$  de vivienda construida (total).

- *Cuál prodría ser un modelo  $SARIMA(p, d, q)(P, D, Q)_s$ ?*
- *Cómo identificarlo?*

Se propuso un modelo

$$Y_t = a \sin\left(\frac{2\pi t}{s}\right) + \sum_{j=1}^{12} \delta_j I_j(t) + \varepsilon_t, \quad (8.7)$$

con  $\varepsilon_t \sim ARMA(p, q)$ , o también un modelo  $SARIMA(p, 0, q)(0, 1, 0)_s$

**Ejemplo 8.1.7.** Serie de empleo en Canada.

## 8.2. Pruebas de Raíz Unitaria Estacional

A partir de la definición de los modelos SARIMA hay dos posibles modelos para una serie con componente estacional:

1. Modelos con componente determinística, es decir, el modelo de descomposición.
2. Modelos no estacionarios, integrados estacionalmente, es decir,  $SARIMA(p, d, q)(P, D, Q)[s]$ , con  $d \neq 0$  y  $D \neq 0$ .

Sin embargo, un modelo de la forma  $SARIMA(p, 0, q)(P, 0, Q)[s]$ , equivalente a un  $ARMA(p+P, q+Q)$ , podría ser tanto estacionario como mostrar características estacionales. Por tanto, un tercer modelo es el anterior, que se denomina un modelo “estacionario estacional”.

El modelo con componente determinística no es estacionario, pero si se elimina la componente de tendencia, el proceso filtrado resulta la suma de un proceso determinístico periódico y otro estacionario en covarianza. Concretamente,  $Y_t - \hat{T}_t = \hat{S}_t + \hat{\varepsilon}_t$ . Este modelo también se denominará “estacionario estacional”.

Se necesitan técnicas para discriminar cuál modelo aplicar. Aunque en el curso se han implementado criterios tales como la validación cruzada y la comparación del MAPE para escoger el mejor modelo para pronosticar, en la teoría de series de tiempo se ha dado mucho desarrollo en las pruebas de raíces unitarias, por ejemplo, las pruebas para detectar raíces unitarias estacionales, como las pruebas de Hylleberg y otros, indicada por HEGY, y la prueba Canova-Hansen, indicada por CH, que se exponen a continuación, con el propósito de utilizarlas como herramientas para escoger el modelo más adecuado.

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  $\varphi_p(z)$  tiene una raíz igual a 1. Es decir, si se cumple que  $\varphi_p(1) = 0$ . En este caso el polinomio se puede factorizar como  $\varphi_p(z) = (1 - z)\varphi_{p-1}(z)$ .

**Definición 8.2.1.** Una serie de tiempo  $Y_t \sim ARMA(p, q)$  se dice integrada estacional, de período  $s$  con  $s$  par, si se cumple que algunas ó todas las raíces de  $z^s = 1$  son raíces del polinomio autorregresivo  $\varphi_p(z)$ .

Las raíces de  $z^s = 1$ , para por ejemplo  $s = 4, 12$ , se denominan las raíces de la unidad, y son  $s$  números complejos. 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  $s = 12$  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  $z = re^{i\theta}$  donde  $r = |z|$  y  $\theta$  es el ángulo que forma  $|z|$  con el eje X, se pueden identificar las cuatro raíces unitarias con los ángulos  $\theta = 2\pi j/4, j = 0, 1, 2, 3$ , ya que  $r = 1$  en todos los casos, pues las raíces unitarias tienen módulo 1. Estos ángulos se denominan las “frecuencias estacionales”.

### 8.2.1. La prueba HEGY

La prueba HEGY 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 se concluirá que la serie en cuestión está integrada estacionalmente integrada.

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. Lo que sigue está tomado del artículo de Hylleberg et al. [1990].

Inicialmente se asume que la serie  $Y_t$  está generada por un proceso AR(p), es decir, se cumple

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

para  $\varepsilon_t$  ruido blanco. Se asume que están dados p números complejos  $\theta_j \in \mathbb{C}$ ,  $j = 1, \dots, p$ , en los cuales  $\varphi(\theta_j)$  es un número finito. Estos p números son las p raíces estacionales. El valor p es el período, indicado anteriormente por s. Adicionalmente se definen las cantidades auxiliares:

$$\begin{aligned} \delta_k(z) &= 1 - z/\theta_k, \\ \Delta(z) &= \prod_{k=1}^p \delta_k(z), \\ \lambda_k &= \varphi(\theta_k) / \prod_{j \neq k} \delta_j(\theta_k). \end{aligned} \quad (8.8)$$

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

$$\varphi(z) = \sum_{k=1}^p \frac{\lambda_k \Delta(z)(1 + \delta_k(z))}{\delta_k(z)} + \Delta(z)\varphi^*(z), \quad (8.9)$$

donde  $\varphi^*(z)$  es un polinomio posiblemente infinito ó racional. Nótese que, por la definición (8.8) se cumple

$$\varphi(\theta_k) = 0 \Leftrightarrow \lambda_k = 0 \quad (8.10)$$

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 (8.9). 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  $\varphi(z) = 1 - z^4$ , donde  $\theta_1 = 1$ ,  $\theta_2 = -1$ ,  $\theta_3 = i$ ,  $\theta_4 = -i$ , la identidad (8.9) se transforma en

$$\begin{aligned} \varphi(z) &= \lambda_1 z(1+z)(1+z^2) + \lambda_2 (-z)(1-z)(1+z^2) \\ &+ \lambda_3 (-iz)(1-z)(1+z)(1-iz) + \lambda_4 (iz)(1-z)(1+z)(1+iz) \\ &+ \varphi^*(z)(1-z^4). \end{aligned} \quad (8.11)$$

Pero se tiene que  $\lambda_3, \lambda_4 \in \mathbb{C}$ , y además,  $\lambda_4 = \bar{\lambda}_3$ . 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 (8.11) queda

$$\begin{aligned}\varphi(z) = & -\pi_1 z(1 + z + z^2 + z^3) - \pi_2(-z)(1 - z + z^2 - z^3) \\ & -(\pi_4 + \pi_3 z)(-z)(1 - z^2) + \varphi^*(z)(1 - z^4).\end{aligned}\quad (8.12)$$

Finalmente, a partir de  $\varphi(L)Y_t = \varepsilon_t$ , se reemplaza  $\varphi(L)$  por la correspondiente expresión obtenida al reemplazar la expresión para  $\varphi(z)$  en (8.12). Si se definen las variables

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

se obtiene la ecuación

$$\varphi^*(L)X_{4,t} = \pi_1 X_{1,t} + \pi_2 X_{2,t} + \pi_3 X_{3,t} + \varepsilon_t. \quad (8.13)$$

Colocando  $\varphi^*(L) = 1$ , la ecuación (8.13) es una regresión lineal múltiple. A partir de esta ecuación se plantea la prueba de hipótesis HEGY para el caso de período  $s = 4$ .

### Descripción de la Prueba

La hipótesis en la prueba HEGY, para el caso  $s = 4$ , se compone de tres hipótesis, una para cada raíz unitaria estacional:  $-1, \pm i$ . Nótese que la raíz  $z = 1$  corresponde a una raíz unitaria para tendencia aleatoria.

$$\begin{aligned}H_0 : \varphi(1) = 0 &\Leftrightarrow \pi_1 = 0, \quad H_a : \varphi(1) > 0 \Leftrightarrow \pi_1 < 0, \\ H_0 : \varphi(-1) = 0 &\Leftrightarrow \pi_2 = 0, \quad H_a : \varphi(2) > 0 \Leftrightarrow \pi_2 < 0, \\ H_0 : |\varphi(i)| = 0 &\Leftrightarrow \pi_3 = \pi_4 = 0, \quad H_a : no(H_0).\end{aligned}$$

En el caso  $s = 12$ , hay 7 hipótesis, dos para  $z = 1, -1$ , y una para cada pareja de raíces unitarias estacionales conjugadas.

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.

Nótese que no rechazar una de las hipótesis nulas equivale a aceptar que es una raíz unitaria estacional. En este caso hay que tener en cuenta la potencia de la prueba. Recordando que la potencia se refiere a la probabilidad:  $\text{Prob}(\text{rechazar } H_0 \mid H_a \text{ es cierta})$ , 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.

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 cierto ARMA con órdenes altos, ó también como un modelo de descomposición, como se vió al comienzo del curso.

## Implementación en R

La prueba HEGY está implementada en la función `HEGY.test`, que tiene los parámetros siguientes. Al colocar `res = HEGY.test(y, itsd, regvar, selectlags)`, la serie debe estar en la variable `y`, y debe haberse declarado como objeto “ts”. El parámetros `itsd` se refiere a si se desea incluir en el modelo un intercepto, tendencia lineal y componente estacional con variables indicadoras. El modelo se define en el caso afirmativo como

$$(1 - L^4)(Y_t - (\beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_j(t))) = \varepsilon_t. \quad (8.14)$$

Colocando `itsd=c(1,1,c(1,2))` se especifica este modelo. En caso de que no se desee incluir estas variables se coloca `itsd=c(0,0,c(0))`. El parámetro `regvar` se coloca `regvar=0` si no se incluyen variables exógenas en el modelo. En caso contrario se incluye una matriz con los valores de las variables. Por ejemplo, se podría incluir el primer rezago de `y`,  $Y_{t-1}$ . El parámetro `selectlags` se coloca por defecto con los valores `selectlags = list(mode = "bic", Pmax = 12)`.

**Ejemplo 8.2.1.** *Retomando el Ejemplo (5.1.1) en la pag. 76, para modelación de la serie de producción de cemento Portland, trimestral se ajustó un modelo de la forma*

$$Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_j(t) + \varepsilon_t, \quad (8.15)$$

$$\varepsilon_t \sim AR(7). \quad (8.16)$$

*En la Figura 8.1 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 en lugar del modelo de componentes (8.15).*

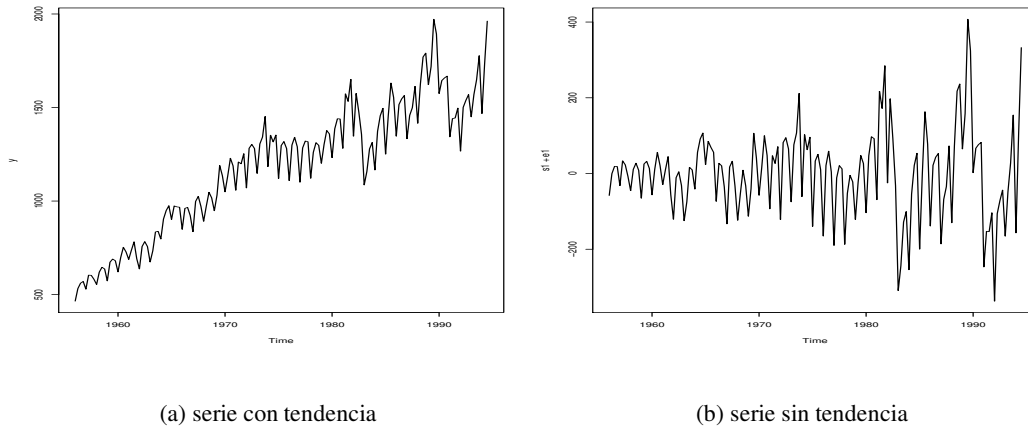


Figura 8.1: Serie de Producción Trimestral de Cemento, Australia

*Con el fin establecer cuál modelo es más adecuado se procede a realizar la prueba HEGY. Los comandos en R son*

```
# prueba hegy
res = HEGY.test(wts=y,
itsd=c(0,0,c(0)),
regvar=0, selectlags=list(mode="signf", Pmax=NULL))
res
-----resultados
Null hypothesis: Unit root.
Alternative hypothesis: Stationarity.
-----

HEGY statistics:
      Stat. p-value
tpi_1    3.245    0.1
tpi_2   -1.020    0.1
Fpi_3:4   2.254    0.1
Fpi_2:4   1.854    NA
Fpi_1:4   4.101    NA
```

*El resultado anterior muestra que no se rechaza la hipótesis nula de raíces unitarias en  $1, -1, \pm i$ , indicadas por  $tpi\_1$ ,  $tpi\_2$ ,  $Fpi\_3:4$ . 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, es apropiado ajustar un modelo SARIMA a la serie. Al examinar la fac de 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 8.2.

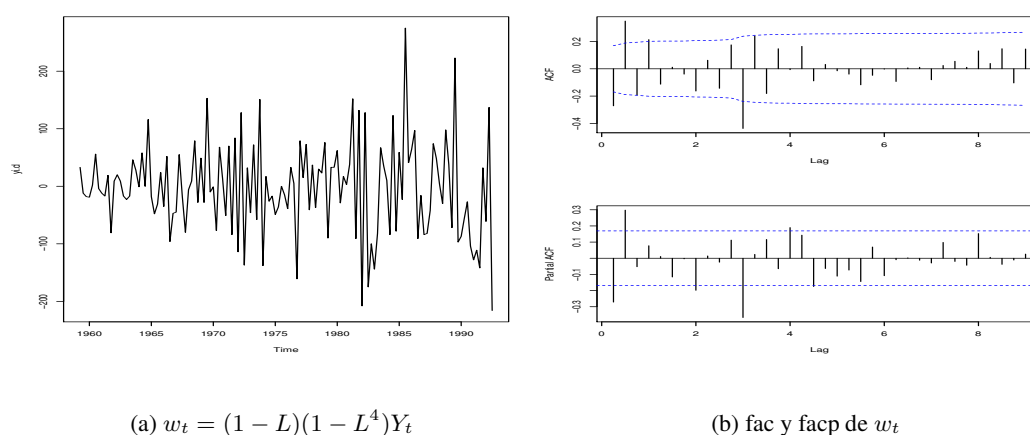


Figura 8.2: Serie Diferenciada de Producción Trimestral de Cemento

Mediante ensayo y error, empezando con un modelo SARIMA  $(1, 1, 1) (1, 1, 1) [4]$ , hasta conseguir residuos ruido blanco se llega al modelo SARIMA  $(3, 1, 2) (1, 1, 2) [4]$ .

Tabla 8.1: 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 son el modelo SARIMA versus los obtenidos con el modelo componentes (8.15) se pueden observar en la Figura 8.3.

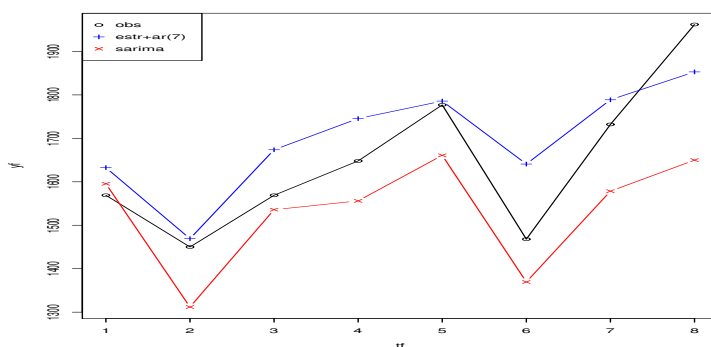


Figura 8.3: Pronósticos a 8 trimestres de la Producción Trimestral de Cemento

Como conclusión, se llega a que el SARIMA se detecta como modelo correcto por la prueba HEGY. Pero el MAPE de sus pronósticos, a 8 trimestres, es 7.89, mayor que el del modelo de Descomposición con errores AR(7), que es 4.64. Sin embargo, el modelo SARIMA tiene mejor ajuste dentro de la muestra. En la Figura 8.4 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.

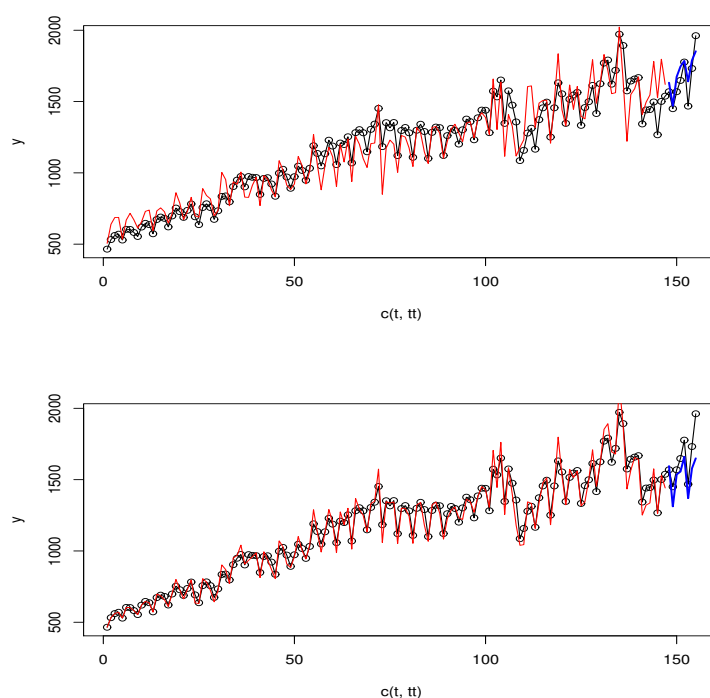


Figura 8.4: Comparación de los Ajustes y Pronósticos. Panel superior: Modelo Descomposicion + AR, Panel inferior: SARIMA

### 8.2.2. La prueba Canova-Hansen

Los modelos de Descomposición con errores ARMA y los SARIMA se consideran, 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”.

La prueba CH se basa en esta idea. La inestabilidad estructural de la componente estacional se toma como una característica equivalente a la existencia de una raíz unitaria estacional y opuesta a estacionario estacional.

## 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

$$Y_t = bY_{t-1} + \sum_{j=1}^{s/2} \gamma_j \cos\left(\frac{2\pi jt}{s}\right) + \gamma_j^* \sin\left(\frac{2\pi jt}{s}\right), \quad (8.17)$$

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 (8.17) es  $s - 1$ , el mismo del modelo con variables indicadoras. Los coeficientes se colocan en vectores  $\underline{\gamma} = (\gamma_1, \dots, \gamma_{s/2})'$ ,  $\underline{\gamma}^* = (\gamma_1^*, \dots, \gamma_{s/2-1}^*)'$ ,  $\underline{\gamma}^* = (\gamma_1^*, \dots, \gamma_{s/2-1}^*)'$

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 un modelo de marcha aleatoria,  $\underline{\gamma}_t = \underline{\gamma}_{t-1} + \underline{u}_t$ ,  $\underline{\gamma}_t^* = \underline{\gamma}_{t-1}^* + \underline{v}_t$ , donde  $\underline{u}_t, \underline{v}_t$  son vectores iid, independientes de media cero. 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.

$$\begin{aligned} H_0 : \quad & \underline{\gamma}_t \equiv \underline{\gamma}, \underline{\gamma}_t^* \equiv \underline{\gamma}^*, \\ H_a : \quad & \underline{\gamma}_t = \underline{\gamma}_{t-1} + \underline{u}_t, \underline{\gamma}_t^* = \underline{\gamma}_{t-1}^* + \underline{v}_t. \end{aligned}$$

En Canova and Hansen [1995, pag. 240, sec. 2.2], se re-escribe la prueba con base en un parámetro  $\tau^2$ , tal que  $H_0 : \tau^2 = 0$ , versus  $H_a : \tau^2 > 0$ . El estadístico de la prueba  $L$ , requiere resultados avanzados para su definición, por lo que se remite al artículo original. Además en el artículo se proveen los valores críticos para varios niveles de significación.

## Implementación en R

La prueba Canova-Hansen está implementada en en la función `CH.test`, que tiene los parámetros siguientes. Al colocar `res = CH.test(y, freq, f0, DetTr)`, la serie debe estar en la variable `y`, y debe haberse declarado como objeto “ts”. El parámetro `freq` se refiere cuáles frecuencias se prueban para detectar raíces unitarias. Nótese que si  $s = 4$  hay 2 frecuencias:  $-1, \pm i$ , por lo que se coloca `freq=c(1, 1)`. En el caso  $s = 12$  hay 6 frecuencias y se coloca `freq=c(1, 1, 1, 1, 1, 1)`. El parámetro `f0 = 0, 1` se coloca igual a 1 si la variable  $Y_{t-1}$  se incluye como variable explicativa en el modelo, y se coloca igual a

0 si no se incluye. El parámetro DetTr es una variable lógica con valores TRUE, FALSE. La opción TRUE indica que se incluye en el modelo una tendencia lineal. Nótese, sin embargo, que para aplicar la prueba Canova and Hansen [1995] 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()`. En consecuencia colocamos DetTr=FALSE.

**Ejemplo 8.2.2.** *Continuando el Ejemplo 8.2.1, para decidir entre los modelos de Descomposición con errores AR(7) y el SARIMA*

$$1) Y_t = \beta_0 + \beta_1 t + \sum_{j=1}^3 \delta_j I_t(t) + \varepsilon_t, \varepsilon_t \sim AR(7),$$

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

*La prueba se implementa con los comandos*

```
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 = s1+e1
ch.out1 = CH.test(wts=y1, freq=c(1,1,1), f0=0, DetTr=FALSE)
ch.out1
```

-----  
Y los resultados son

```
Canova & Hansen test
----- - -----
```

```
Null hypothesis: Stationarity.
Alternative hypothesis: Unit root.
Frequency of the tested cycles: pi/2 , pi , NA ,
```

```
L-statistic: 2.274
```

```
Critical values:
0.10 0.05 0.025 0.01
1.28 1.47 1.63 1.88
```

*Como el estadístico observado es 2.274, mayor que el valor crítico al nivel de 5 %, se rechaza la hipótesis nula y se detecta una raíz unitaria estacional, que corresponde a inestabilidad estructural en la componente estacional. Por tanto, el modelo SARIMA es preferible para modelar la serie.*