Prueba de HEGY en R:
Una guía.

Julio César Alonso
Paul Semaán

No. 23
Junio de 2010

Apuntes de Economía

ISSN 1794-029X
No. 23, Junio de 2010

Editor
Julio César Alonso
jcalonso@icesi.edu.co

Gestión Editorial
Departamento de Economía - Universidad Icesi

www.icesi.edu.co
Tel: 5552334 ext: 8398. Fax: 5551441
Calle 18 # 122-135 Cali, Valle del Cauca, Colombia

Prueba de HEGY en R: Una guía.

Julio César Alonso
Paul Semaán *
CIENFI - Departamento de Economía - Universidad Icesi
Cali - Colombia

29 de septiembre de 2010

Resumen

Este documento es una guía práctica cuyo fin es el de brindarle ayuda a la persona que trabaje con modelos de series de tiempo que necesite emplear la prueba de HEGY. Es común encontrar discusiones acerca de las implicaciones de la presencia de raíces unitarias en las series, sin embargo, recientemente ha cobrado importancia el estudio de las raíces unitarias estacionales en los datos de frecuencia trimestral o mensual. En esta guía identificaremos los distintos tipos de raíces unitarias estacionales, la aplicación de la prueba HEGY para su identificación y la manera de solucionar el problema mediante una adecuada diferenciación de las series. También explicaremos paso a paso como realizar estos procedimientos en el software estadístico *R*. Por la forma didáctica como está escrito el documento, puede ser usado en un curso de series de tiempo en pregrado o de maestría o por profesionales que desean aplicar esta prueba.

Palabras claves: Raíz unitaria, estacionalidad, estacionariedad, series de tiempo, HEGY.

*Grupo de Investigación Economía, Métodos Cuantitativos y Políticas Públicas

> *"The mechanical application of the seasonal difference filter is likely to produce serious misspecification in many instances. The evidence presented here indicates that unit roots are often absent at some or all of the seasonal frequencies, so empirical researchers should check for their presence (using procedures such as the one discussed above) rather than imposing them at all seasonal frequencies a priori".*
>
> *Beaulieu, Miron (1993)*

1. Introducción

En los estudios aplicados con series macroeconómicas es común emplear series de tiempo trimestrales o mensuales. Los libros de texto normalmente discuten las implicaciones de la presencia de una raíz unitaria en series anuales, tanto en el comportamiento de las series como en el famoso problema de la regresión espuria. No obstante, es poco común encontrar textos de nivel introductorio que discutan las implicaciones de la presencia de una o varias raíz unitarias en series trimestrales o mensuales y cómo emplear la prueba de Hylleberg et al. (1990) (de aquí en adelante HEGY) para encontrar la correspondiente serie estacionaria.

Este documento tiene como objetivos discutir las implicaciones en el comportamiento de series con raíces unitarias estacionales, tanto para datos trimestrales como mensuales y presentar cómo aplicar con R[1] la prueba HEGY tanto para datos mensuales como trimestrales.

En la siguiente sección del documento se definirá el concepto de estacionalidad e integración estacional. También, se explicarán los diferentes métodos para trabajar con datos estacionales dependiendo de su Data Generating Process (DGP). En la tercera y cuarta sección se identificarán los diferentes tipos de raíces estacionales presentes en datos mensuales y trimestrales. Así como, el tratamiento necesario para obtener variables estacionarias y su relación con la prueba de HEGY. Finalmente, en la quinta sección, se ilustra el uso de la prueba HEGY empleando el Software *R*, tanto para datos mensuales como trimestrales.

2. Consideraciones especiales para los datos estacionales

Cuando se trabaja con datos con una periodicidad más alta que la anual (por ejemplo datos trimestrales o mensuales), es común encontrar un comportamiento recurrente en iguales periodos de los años; es decir, se presenta estacionalidad.

La estacionalidad comúnmente es tratada o modelada empleando:

1. variables dummy

[1] Este software puede ser descargado de manera gratuita del link http://www.r-project.org. Además en línea se pueden encontrar de manera gratuita diferentes manuales introductorios al uso de este software.

2. modelos ARMA estacionales[2]

3. empleando integración estacional[3]

Cada uno de estos tres diferentes tratamientos de las series será correcto si se le aplica a un determinado comportamiento del proceso generador de los datos (DGP de aquí en adelante por su sigla en inglés del término Data Generating process) real. Es decir, si empleamos variables dummy para tratar la estacionalidad cuando lo correcto era diferenciar, entonces estaremos frente a un modelo mal especificado que nos podrá llevar a conclusiones erradas.

Si el DGP implica que el comportamiento estacional es puramente determinístico, entonces la mejor aproximación será emplear variables dummy (Ver el modelo 1 de la Figura 1). Por ejemplo, si se cuenta con una serie trimestral, esto implicaría el siguiente DGP:

$$y_t = \mu + \sum_{i=1}^{3} \alpha_i D_{i,t} + \varepsilon_t \tag{1}$$

donde ε_t es el error del modelo[4] y $D_{i,t}$ corresponde a una variable dummy que toma el valor de uno si la observación corresponde al trimestre i y cero en caso contrario. O el DGP puede incluir una tendencia determinística[5] (modelo 2 de la Figura 1) como por ejemplo:

$$y_t = \mu + \sum_{i=1}^{3} \alpha_i D_{i,t} + \sum_{i=1}^{s} \beta_i \cdot D_{i,t} \cdot t + \varepsilon_t \tag{2}$$

Si por el contrario, el DGP real corresponde a un proceso estacionario (en su covarianza) estacional (modelo 3 de la Figura 1), entonces la mejor opción será estimar un modelo SARMA. En este caso el DGP, se puede expresar como:

$$y_t = \mu + \theta y_{t-4} + \varepsilon_t \tag{3}$$

o el proceso puede ser estacional estacionario alrededor de una tendencia (modelo 4 de la Figura 1) como por ejemplo:

$$y_t = \mu + \theta y_{t-4} + \beta \cdot t + \varepsilon_t \tag{4}$$

donde $|\theta| < 1$.

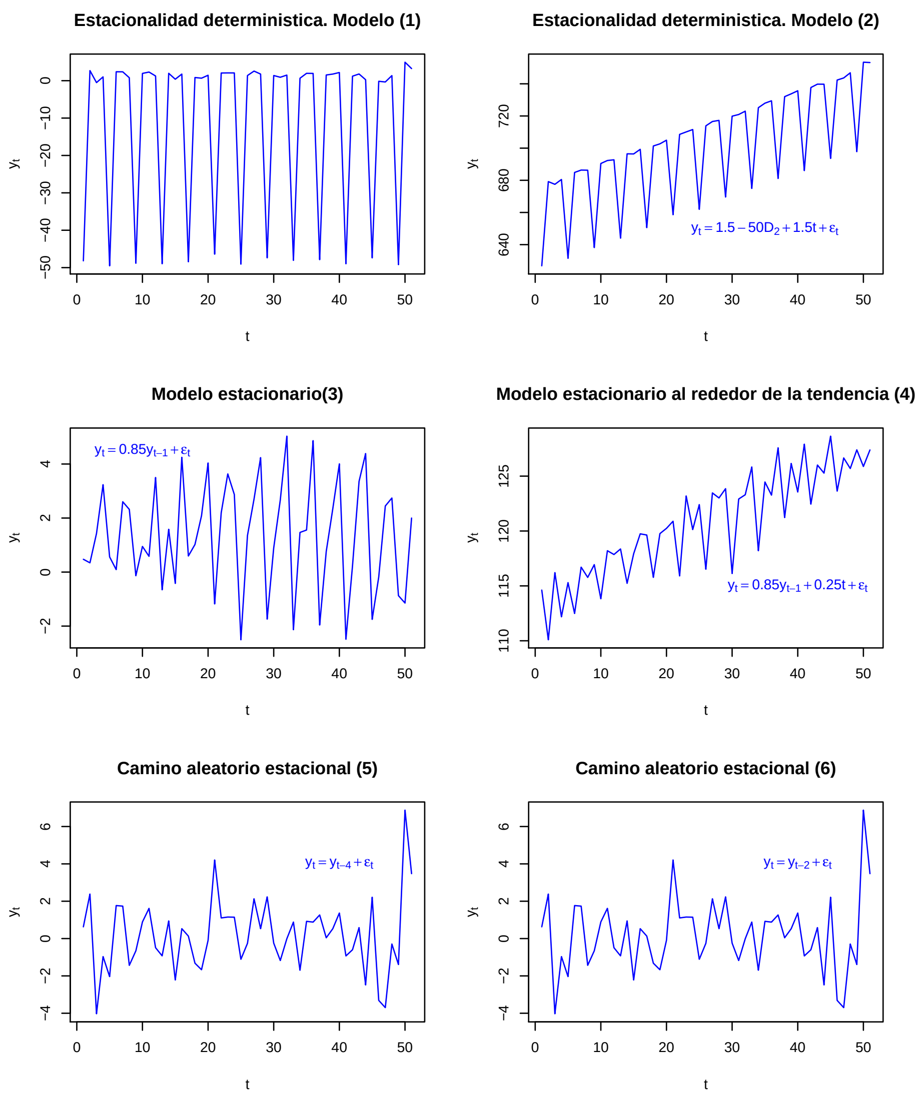
[2]O también conocidos como modelos SARMA. La S en el término SARMA proviene del término en inglés para la estacionalidad (seasonal).

[3]En otras palabras diferenciando las series teniendo en cuanta el periodo. En general, en la literatura de series de tiempo integrar es sinónimo de diferenciar o sacar diferencias. Por ejemplo, para datos trimestrales se puede trabajar con una serie que corresponde a la diferencia entre la observación actual y la observación del año anterior para el mismo período. Es decir, $y_t - y_{t-4}$.

[4]En todo el documento supondremos que ε_t es un proceso ruido blanco.

[5]El coeficiente asociado a la tendencia puede o no cambiar con la estación.

Figura 1: Simulaciones de los diferentes modelos con estacionalidad



Pero, si por el contrario el DGP real implica la presencia de un proceso no estacionario estacional (raíces unitarias estacionales) entonces la aproximación correcta es emplear la integración (o diferenciación) estacional (modelo 5 de la Figura 1). En este caso el DGP sería por ejemplo:

$$y_t = y_{t-4} + \varepsilon_t \tag{5}$$

Este proceso se conoce como un camino aleatorio estacional (en inglés seasonal random walk). O el DGP también puede ser (modelo 6 de la Figura 1):

$$y_t = y_{t-2} + \varepsilon_t \tag{6}$$

La gran diferencia entre estas tres posibles formas de estacionalidad es que en el caso de la estacionalidad determinística [6] y la estacionalidad estacionaria[7] los choques al modelo desaparecen en el largo plazo (se retorna a la media o a la tendencia según sea el caso). Por otro lado, en los modelos no-estacionarios estacionales los choques tienen efecto permanente y se incorporan a la serie.

De hecho, los procesos no-estacionarios estacionales tienen propiedades similares a las series integradas ordinarias (anuales o también conocidas como de frecuencia cero, $s = 0$). Por ejemplo, presentan memoria larga de tal manera que los choques pueden cambiar los comportamientos estacionales permanentemente, tienen varianzas que crecen linealmente y asintóticamente no está correlacionados con otras raíces de otras frecuencias.

Definición de integración estacional

En general se dirá que una serie de tiempo y_t es integrada de orden d ($y_t \sim I(d)$) si es necesario diferenciar(integrar) a la serie original d veces. Por ejemplo, si z_t es integrada de orden uno (z_t ∼I(1)), $entonces \Delta z_t = z_t - z_{t-1}$ será una serie estacionaria.
Por otro lado, cuando consideramos series de tiempo con una frecuencia mayor a un año, entonces aparece la necesidad de definir el termino integración estacional. Engle et al. (1989) propone la siguiente definición:

> una serie y_t, que tiene una frecuencia de s periodos en un año, será integrada de orden d_0 y d_s (y_t ∼SI(d_o, d_s)) si $y_t = \Delta^{d_0} \left[S\left(L\right)\right]^{d_s} y_t$ es estacionario, donde el polinomio $S\left(L\right)$ se define como $S\left(L\right) = 1 + L + L^2 + \ldots + L^{s-1}$ y L representa el polinomio de rezagos.

Es importante anotar que $\Delta_s = 1 - L^s$ lo cual equivale a $\Delta_s = (1 - L)\left(1 + L + L^2 + \ldots + L^{s-1} = \Delta S\left(L\right)\right.$.

3. La prueba de HEGY en muestras trimestrales

Como se discutió anteriormente, es común que una serie trimestral muestre comportamientos estacionales. Cómo modelar esa estacionalidad de manera adecuada, dependerá de poder detectar cuál de los tres tipos de estacionalidad se presenta.

En especial, determinar el orden de integración de las series trimestrales permitirá determinar si se necesitará integrar o no y además que tipo de in-

[6]Por ejemplo las expresiones (1) y (2).
[7]Por ejemplo las expresiones (3) y (4).

tegración se necesitará. A continuación se describe como emplear la prueba de HEGY para determinar si una serie trimestral se debe integrar o no; y en caso de ser necesaria la integración, cómo se debe integrar.

Para entender la prueba HEGY, partamos del hecho que para para datos trimestrales contamos con una frecuencia de observaciones de 4 por año. Es decir, $s = 4$. En este caso tendremos que el operador de diferencias trimestral $\Delta_4 = y_t - y_{t-4} = \left(1 - L^4\right) y_t$ se puede escribir como:

$$\Delta_4 y_t = y_t - y_{t-4} = \left(1 - L^4\right) y_t \tag{7}$$

factorizando, tenemos que:

$$\left(1 - L^4\right) = (1 - L)(1 + L)\left(1 + L^2\right) \tag{8}$$

Si $\Delta_4 y_t$ es estacionaria, esto implica que al menos una de las siguientes condiciones aplica:

Condición 1: $(1 - L) = 0$ y por lo tanto tendremos que $y_t - y_{t-1}$ será estacionario. Es decir, existe una *raíz unitaria no estacional.* Esto implica que $y_t \approx y_{t-1}$.

Condición 2: $(1 + L) = 0$, esto implica que $y_t + y_{t-1}$ será estacionario. Esto implica que $y_t \approx -y_{t-1} \approx y_{t-2}$. Es decir, existe una *raíz semestral*(bianual).

Condición 3: $\left(1 + L^2\right) = 0$ y por lo tanto $y_t + y_{t-2}$ será estacionario. Por lo tanto, $y_t \approx -y_{t-2} \approx y_{t-4}$. Es decir, existe una *raíz unitaria anual.*

Ahora, supongamos que la **Condición 1** no se cumple, entonces tendrá que ser cierto que si estimamos una regresión de y_t en función de y_{t-1} no obtendremos una raíz unitaria. O lo que es equivalente, si se corre la regresión de $(1 - L)\, y_t$ en función de y_{t-1}, la pendiente deberá ser cero. El problema de estimar la regresión

$$(1 - L\ )\, y_t = \mu + \pi y_{t-1} + \varepsilon_t$$

es que podemos estar empleando series no-estacionarias y el problema de la regresión espuria aparece. Para resolver el problema, y asegurar que contamos con series estacionarias al momento de realizar la regresión, podemos multiplicar a ambos lados por $(1 + L)\left(1 + L^2\right)$, de tal manera que obtenemos la regresión de $(1 + L)\left(1 + L^2\right)(1 - L)\, y_t = \Delta_4 y_t$ con respecto a $(1 + L)\left(1 + L^2\right)$. Y si la **Condición 1** no se cumple, entonces la pendiente será igual a cero.

De manera similar, si se desea probar si la **Condición 2** se cumple, podemos correr la regresión de $\Delta_4 y_t$ con respecto a $(1 - L)\ \left(1 + L^2\right)$. Y para probar la **Condición 3**, se puede correr la regresión de $\Delta_4 y_t$ en función de $(1 - L)(1 + L)\, y_t$. Pero esta última regresión solo podría probar los ciclos que tienen picos en los trimestres uno y tres. Para probar los ciclos que tienen pico para el trimestre dos y cuatro debemos emplear como variable explicativa de la regresión a $(1 - L)\ (1 + L)\, y_{t-2}$.

La prueba de HEGY explota estos resultados y emplea estadísticos que permiten comprobar las 3 condiciones conjuntamente y de manera individual. Para lograrlo, los autores definen las siguientes variables[8]:
$y_{1t} \equiv (1+L)\left(1+L^2\right) y_t = y_t + y_{t-1} + y_{t-2} + y_{t-3}$
$y_{2t} \equiv -(1-L)\left(1+L^2\right) y_t = -\left(y_t - y_{t-1} + y_{t-2} - y_{t-3}\right)$
$y_{3t} \equiv (1-L)(1+L) y_t = y_t - y_{t-2}$
$y_{4t} \equiv \Delta_4 y_t = y_t - y_{t-4}$.
Con estas variables, y teniendo en cuenta la posibilidad de que exista una tendencia determinística y una estacionalidad determinística la prueba implica correr la siguiente regresión:

$$y_{4t} = \mu_t + \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{4,t-1} + \sum_{i=1}^{p} y_{4t-i} + \varepsilon_t \tag{9}$$

donde $\mu_t = \delta + \beta t + \sum_{i=1}^{3} \alpha_i D_{i,t}$. Los rezagos de y_{4t} se incluyen para eliminar cualquier posible autocorrelación en el error; es decir, para garantizar que el error estimado sea ruido blanco[9].

A partir de (9) podemos comprobar las siguientes hipótesis nulas:

- H_A: $\pi_1 = 0$ (presencia de una *raíz unitaria no estacional*).
- H_B: $\pi_2 = 0$ (presencia de una *raíz semestral*(bianual)).
- H_C: $\pi_3 = \pi_4 = 0$ (presencia de una*raíz unitaria anual).*

Las H_A y H_B pueden ser probadas por medio de pruebas tipo t de una sola cola (la hipótesis alterna será $\pi_i < 0$). Para probar H_C se deberá emplear una prueba tipo F. Hylleberg et al. (1990) demostraron que la distribución de estos estadísticos de prueba (tanto el individual como el de la prueba conjunta) no siguen una distribución convencional y sus valores críticos son además función de la presencia o no de las variables dummy o la tendencia en (9).

Los resultados de esta prueba nos permitirá entonces determinar si es necesario realizar diferenciaciones y determinar cuáles serían esas diferenciaciones. En el cuadro 1 se presenta un resumen de como emplear los resultados de esta prueba.

[8]Es importante resaltar que el signo negativo que se adiciona en y_{2t} es para asegurar la dirección usual de la prueba de hipótesis correspondiente.
[9]De manera similar a la prueba ADF para datos anuales.

Cuadro 1: Interpretación de los resultados de la prueba HEGY trimestral

Hipótesis que *NO* se rechazan	**Hipótesis que se rechazan**	**Tipo de raíz unitaria presente**	**Transformación para obtener variable estacionaria**
H_A, H_B, H_C	-	no estacional semestral anual	$\Delta_4 y_t = y_t - y_{t-4}$
H_A, H_B	H_C	no estacional semestral	$\Delta_2 y_t = y_t - y_{t-2}$
H_A, H_C	H_B	no estacional anual	$(1-L)\left(1+L^2\right) y_t$ $y_t - y_{t-1} + y_{t-2} - y_{t-3}$
H_B, H_C	H_A	semestral anual	$(1+L)\left(1+L^2\right) y_t$ $y_t + y_{t-1} + y_{t-2} + y_{t-3}$
H_A	H_B, H_C	no estacional	$\Delta_1 y_t = y_t - y_{t-1}$
H_B	H_A, H_C	semestral	$(1+L)\, y_t = y_t + y_{t-1}$
H_C	H_A, H_B	anual	$\left(1+L^2\right) y_t = y_t + y_{t-2}$
-	H_A, H_B, H_C	no hay	ninguna

4. La prueba de HEGY y muestras mensuales

Franses (1991a) y Franses (1991b) extiende la prueba de HEGY a datos mensuales. Al igual que la prueba para datos trimestrales, podemos partir del hecho que para datos mensuales observamos los datos con una frecuencia de 12 periodos al interior de un año; es decir, $s = 12$.

El operador de diferencia para $s = 12$ $\left(\Delta_{12} y_t = y_t - y_{t-12} = \left(1 - L^{12}\right) y_t\right)$ se puede factorizar de la siguiente manera:

$$
\begin{aligned}
\left(1-L^{12}\right) &= (1-L)(1+L)(1-i\cdot L)(1+i\cdot L) \\
&\times \left[1+\frac{\left(1+\sqrt{3}i\right)}{2}L\right]\left[1+\frac{\left(1-\sqrt{3}i\right)}{2}L\right] \\
&\times \left[1-\frac{\left(1-\sqrt{3}i\right)}{2}L\right]\left[1-\frac{\left(1+\sqrt{3}i\right)}{2}L\right] \\
&\times \left[1+\frac{\left(\sqrt{3}+i\right)}{2}L\right]\left[1+\frac{\left(\sqrt{3}-i\right)}{2}L\right] \\
&\times \left[1-\frac{\left(\sqrt{3}+i\right)}{2}L\right]\left[1-\frac{\left(\sqrt{3}-i\right)}{2}L\right]
\end{aligned}
\tag{10}
$$

Así, el hecho que $\Delta_{12} y_t = y_t - y_{t-12} = \left(1 - L^{12}\right) y_t$ sea estacionario implicará necesariamente las siguientes condiciones:

Condición 1: $(1 - L) = 0$. De tal manera que $y_t \approx y_{t-1}$ Esto implica que $y_t - y_{t-1}$ sería estacionario. Es decir, existe una *raíz unitaria no estacional*. A esta raíz se le denomina *raíz de ciclo cero* o *raíz a la frecuencia cero*.

Condición 2: $(1 + L)$=0. En este caso, $y_t + y_{t-1}$ será estacionario y por tanto $y_t \approx -y_{t-1} \approx y_{t-2}$. Así, esta condición implica una *raíz unitaria bimensual*, es decir 6 ciclos por año. A esta raíz se le denomina *raíz de ciclo* π.

Condición 3: $(1 - i \cdot L)(1 + i \cdot L) = \left(1 + L^2\right) = 0$. Esto implica que $y_t + y_{t-2}$ será estacionario y que $y_t \approx -y_{t-2} \approx y_{t-4}$. Es decir, existirá una *raíz unitaria que corresponde a un ciclo de 3 veces por año.* Es decir un camino aleatorio entre observaciones separadas cuatro períodos. Estas dos raíces [10] corresponden a la *frecuencia* $\pi/2$.

Condición 4: $\left[1 + \left(1 + \sqrt{3}i\right) L/2\right]\left[1 + \left(1 - \sqrt{3}i\right) L/2\right] = 0$. Esta condición implica que $1 + L + L^2 = 0$. Es decir, $y_t + y_{t-1} + y_{t-2}$ será estacionario. Por tanto tendremos que[11] $y_t \approx y_{t-3}$. Es decir, existirá una *raíz unitaria trimestral*,es decir 4 ciclos anuales por año. A estas dos raíces se les conoce también como *raíces de la frecuencia* $2\pi/3$.

Condición 5: $\left[1 - \left(1 + \sqrt{3}i\right) L/2\right]\left[1 - \left(1 - \sqrt{3}i\right) L/2\right] = 0$. Esto equivale a $1 - L + L^2 = 0$; en otras palabras, $y_t - y_{t-1} + y_{t-2}$ será estacionario. Esto implicará que[12] $y_t \approx y_{t-6}$. Esto implica que existirá una *raíz unitaria semestral*,es decir 2 ciclos anuales por año. Estas raíces son conocidas como *raíces de la frecuencia* $\pi/3$.

Condición 6: $\left[1 + \left(\sqrt{3} + i\right) L/2\right]\left[1 + \left(\sqrt{3} - i\right) L/2\right] = 0$. Esto es equivalente a $1 + \sqrt{3}L + L^2 = 0$. Esto implica que $y_t - \sqrt{3}y_{t-1} + y_{t-2}$ será estacionario. Así, tenemos que dos raíces corresponden a las *raíces de la frecuencia* $5\pi/6$.

Condición 7: $\left[1 - \left(\sqrt{3} + i\right) L/2\right]\left[1 - \left(\sqrt{3} - i\right) L/2\right]$ =0 esta condición implica que $1 - \sqrt{3}L + L^2 = 0$. En este caso[13] al igual que en el caso anterior, tendremos una *raíz unitaria anual.* Estas dos raíces corresponden a las *raíces de la frecuencia* $\pi/6$.

Entonces, similarmente al caso de los datos trimestrales, Franses (1991a) extendió la prueba HEGY a datos mensuales. Para realizar esta prueba se necesitan crear variables que permitan probar cada una de las condiciones descritas

[10] Noten que esta condición se deriva de combinar dos raíces: $-i$ y $+i$.
[11] En el Anexo se presenta una demostración de esta afirmación.
[12] En el Anexo se presenta una demostración de esta afirmación.
[13] En el anexo, al final, se demuestra esta afirmación.

anteriormente. Por ejemplo, para probar la **Condición 1** se deberá crear la variable[14]:
$y_{1t} \equiv (1+L)\left(1+L^2\right)(1+L^4+L^8)y_t = \left(1+L+L^2+L^3+L^4+\ldots L^{11}\right)y_t$
donde la parte en azul ha sido adicionada para garantizar la estacionariedad al momento de estimar una regresión y así evitar una regresión espuria. Para probar la **Condición 2** se requiere crear la siguiente variable:
$y_{1t} \equiv -(1-L)\left(1+L^2\right)(1+L^4+L^8)y_t = -(1-L+L^2-L^3+L^4-L^5\cdots-L^{11})y_t$.
nuevamente, la parte en azul ha sido adicionada para garantizar estacionariedad. Para probar la **Condición 3** se deberán crear las siguientes dos variables[15]:
$y_{3t} \equiv -(L-L^3+L^5-L^7+L^9-L^{11})y_t$
$y_{4t} \equiv -(1-L^2+L^4-L^6+L^8-L^{10})y_t$.

Para probar las **Condición 4 a 7** se deberá crear, de manera similar, las siguientes variables[16]:
$y_{5t} \equiv -\frac{1}{2}(1+L-2L^2+L^3+L^4-2L^5+L^6+L^7-2L^8+L^9+L^{10}-2L^{11})y_t$
$y_{6t} \equiv \frac{\sqrt{3}}{2}(1-L+L^3-L^4+L^6-L^7+L^9-L^{10})y_t$
$y_{7t} \equiv \frac{1}{2}(1-L-2L^2-L^3+L^4+2L^5+L^6-L^7-2L^8-L^9+L^{10}+2L^{11})y_t$
$y_{8t} \equiv -\frac{\sqrt{3}}{2}(1+L-L^3-L^4+L^6+L^7-L^9-L^{10})y_t$
$y_{9t} \equiv -\frac{1}{2}(1-L-2L^2-L^3+L^4+2L^5+L^6-L^7-2L^8-L^9+L^{10}+2L^{11})y_t$
$y_{10t} \equiv \frac{1}{2}(1-\sqrt{3}L+2L^2-\sqrt{3}L^3+L^4-L^6+\sqrt{3}L^7-2L^8+\sqrt{3}L^9-L^{10})y_t$
$y_{11t} \equiv \frac{1}{2}(\sqrt{3}+L-L^3-\sqrt{3}L^4-2L^5-\sqrt{3}L^6-L^7+L^9-\sqrt{3}L^{10}+2L^{11})y_t$
$y_{12t} \equiv -\frac{1}{2}(1+\sqrt{3}L+2L^2+\sqrt{3}L^3+L^4-L^6-\sqrt{3}L^7-2L^8-\sqrt{3}L^9-L^{10})y_t$

Entonces, la prueba de HEGY para datos mensuales implicará estimar el siguiente modelo:

$$y_{13t} = \mu_t + \sum_{j=1}^{12} \pi_1 y_{j,t-1} + \sum_{i=1}^{p} y_{13t-i} + \varepsilon_t \tag{11}$$

donde $y_{13t} \equiv \Delta_{12} y_t = y_t - y_{t-12}$ y $\mu_t = \delta + \beta t + \sum_{i=1}^{1} 1\alpha_i D_{i,t}$. Es importante recordar que los rezagos de y_{13t-i} se incluyen para eliminar cualquier posible autocorrelación en el error; es decir, para garantizar que el error estimado sea ruido blanco. Además, se incluye μ_t para permitir la posibilidad de que exista una tendencia determinística y/o una estacionalidad determinística.

Partiendo de (11) podemos comprobar las siguientes hipótesis nulas:

- H_a: $\pi_1 = 0$ (*raíz unitaria no estacional*)
- H_b: $\pi_2 = 0$ (*raíz bimensual*)

[14]Es importante anotar que $\left[1+\frac{(\sqrt{3}+i)}{2}L\right]\left[1+\frac{(\sqrt{3}-i)}{2}L\right]\left[1-\frac{(\sqrt{3}+i)}{2}L\right]\left[1-\frac{(\sqrt{3}-i)}{2}L\right]$ es igual a $(1+L^4+L^8)$.

[15]La **Condición 3** implica dos raíces y por eso se necesita crear dos variables

[16]y_{5t} y y_{6t} permiten probar la **Condición 4**, y_{7t} y y_{8t} permiten probar la **Condición 6** y así sucesivamente.

- H_c: $\pi_3 = \pi_4 = 0$ (*raíz unitaria para períodos de cuatro meses)*
- H_d: $\pi_5 = \pi_6 = 0$ (*raíz unitaria trimestral)*
- H_e: $\pi_7 = \pi_8 = 0$ (*raíz unitaria semestral)*
- H_f: $\pi_9 = \pi_{10} = 0$ (*raíz una la frecuencia* $5\pi/6$ *)*
- H_g: $\pi_{11} = \pi_{12} = 0$ (*raíz unitaria anual*)

Las hipótesis H_a y H_b pueden ser probadas por medio de pruebas tipo t de una sola cola (la hipótesis alterna será $\pi_i < 0$). Para probar H_c a H_f se deberá emplear una prueba tipo F. (Franses (1991a)) demostró que la distribución de estos estadísticos de prueba (tanto el individual como el de la prueba conjunta) no siguen una distribución convencional y sus valore críticos dependen de la presencia o no de las variables dummy o la tendencia en (11). En el cuadro 2 se presenta un resumen de los resultados hasta aquí expuestos.

Cuadro 2: Posibles raíces unitarias presentes en datos mensuales

Condición	**Polinómio**	**Raíz**	**nombre de la raíz**	**Hipótesis nula**	**Frecuencia en término de π**	**Número de ciclos por año**
1	$(1-L)$	$+1$	no estacional	H_a	0	0
2	$(1+L)$	-1	bimensual	H_b	π	6
3	$(1+L^2)$	$\pm i$	cuatrimestral (cada 4 meses)	H_c	$\pm\frac{\pi}{2}$	3
4	$(1+L+L^2)$	$-\frac{1}{2}(1 \pm \sqrt{3}i)$	trimestral	H_d	$\mp\frac{2\pi}{3}$	4
5	$(1-L+L^2)$	$\frac{1}{2}(1 \pm \sqrt{3}i)$	semestral	H_e	$\pm\frac{\pi}{3}$	2
6	$(1+\sqrt{3}L+L^2)$	$-\frac{1}{2}(\sqrt{3} \pm i)$	-	H_f	$\mp\frac{5\pi}{6}$	5
7	$(1-\sqrt{3}L+L^2)$	$\frac{1}{2}(\sqrt{3} \pm i)$	anual	H_g	$\pm\frac{\pi}{6}$	1

Noten que además de las pruebas anteriores se puede probar las siguientes hipótesis nulas:

- H_h: $\pi_2 = \pi_3 = \ldots = 0$ (están presentes todas las raíces unitarias estacionales)
- H_i: $\pi_1 = \pi_3 = \ldots = 0$ (están presentes todas las raíces unitarias tanto la no estacional como las estacionales)

Los resultados de esta prueba nos permitirá entonces determinar si es necesario realizar diferenciaciones y determinar cuáles serían esas diferenciaciones. En los cuadro 3 a 5 se presentan algunas de las 127 posibilidades que aparecen al aplicar esta prueba.

Cuadro 3: Interpretación de los resultados de la prueba HEGY mensual

Hipótesis que no se rechaza	**Hipótesis que se rechaza**	**tipo de raíz unitaria presente**	**Transformación para obtener variable estacionaria**
H_i	-	no estacional bimensual cuatrimestral trimestral semestral frecuencia $\frac{5\pi}{6}$ anual	$\Delta_{12}y_t = y_t - y_{t-12}$
H_a, H_b, H_c, H_d, H_e, H_f, H_g	-	no estacional bimensual cuatrimestral trimestral semestral frecuencia $\frac{5\pi}{6}$ anual	$\Delta_{12}y_t = y_t - y_{t-12}$
H_b, H_c, H_d, H_e, H_f, H_g	H_a	bimensual cuatrimestral trimestral semestral frecuencia $\frac{5\pi}{6}$ anual	$(1 - L + L^2 - L^3 + L^4 - L^5 + L^6$ $-L^7 + L^8 - L^9 + L^{10} - L^{11})y_t$ $y_t - y_{t-1} + y_{t-2} - y_{t-3} + y_{t-4}$ $-y_{t-5} + \ldots + y_{t-10} - y_{t-11}$
H_c, H_d, H_e, H_f, H_g	H_a, H_b	cuatrimestral trimestral semestral frecuencia $\frac{5\pi}{6}$ anual	$\left(1 + L^2 + L^4 + L^6 + L^8 + L^{10}\right) y_t$ $y_t + y_{t-2} + y_{t-4} + y_{t-6}$ $+y_{t-8} + y_{t-10}$
H_d, H_e, H_f, H_g	H_a, H_b, H_c	trimestral semestral frecuencia $\frac{5\pi}{6}$ anual	$\left(1 + L^4 + L^8\right) y_t$ $y_t + y_{t-4} + y_{t-8}$
H_e, H_f, H_g	H_a, H_b, H_c, H_d	semestral frecuencia $\frac{5\pi}{6}$ anual	$\left(1 - L + L^3 - L^5 + L^6\right) y_t$ $y_t - y_{t-1} + y_{t-3} - y_{t-5} + y_{t-6}$
H_f, H_g	H_a, H_b, H_c, H_d, H_e	frecuencia $\frac{5\pi}{6}$ anual	$\left(1 - L^2 + L^4\right) y_t$ $y_t - y_{t-2} + y_{t-4}$
H_g	H_a, H_b, H_c, H_d, H_e, H_f	anual	$\left(1 - \sqrt{3}L + L^2\right) y_t$ $y_t - \sqrt{3}y_{t-1} + y_{t-2}$
H_a, H_c, H_d, H_e, H_f, H_g	H_b	no estacional cuatrimestral trimestral semestral frecuencia $\frac{5\pi}{6}$ anual	$(1 - L + L^2 - L^3 + L^4 - L^5 + L^6$ $-L^7 + L^8 - L^9 + L^{10} - L^{11})y_t$ $y_t - y_{t-1} + y_{t-2} - y_{t-3} + y_{t-4}$ $-y_{t-5} + \ldots + y_{t-10} - y_{t-11}$

Cuadro 4: Interpretación de los resultados de la prueba HEGY mensual. Cont.

Hipótesis que no se rechaza	**Hipótesis que se rechaza**	**tipo de raíz unitaria presente**	**Transformación para obtener variable estacionaria**
H_a, H_b, H_d, H_e, H_f, H_g	H_c	no estacional bimensual trimestral semestral frecuencia $\frac{5\pi}{6}$ anual	$((1-L^2+L^4-L^6$ $+L^8-L^{10})y_t$ $y_t-y_{t-2}+y_{t-4}-y_{t-6}$ $+y_{t-8}+y_{t-10}$
H_a, H_b, H_c, H_e, H_f, H_g	H_d	no estacional bimensual cuatrimestral semestral frecuencia $\frac{5\pi}{6}$	$(1-L+L^3-L^4+L^6$ $-L^7+L^9-L^{10})y_t$ $y_t-y_{t-1}+y_{t-3}-y_{t-4}$ $+y_{t-6}-y_{t-7}+y_{t-9}-y_{t-10}$
H_a, H_b, H_c, H_d, H_f, H_g	H_e	no estacional bimensual cuatrimestral trimestral frecuencia $\frac{5\pi}{6}$ anual	$(1+L-L^3-L^4+L^6$ $+L^7-L^9-L^{10})y_t$ $y_t+y_{t-1}-y_{t-3}-y_{t-4}$ $+y_{t-6}+y_{t-7}-y_{t-9}-y_{t-10}$
H_a, H_b, H_c, H_d, H_e, H_g	H_f	no estacional bimensual cuatrimestral trimestral semestral	$(1-\sqrt{3}L+2L^2-\sqrt{3}L^3$ $+L^4-L^6+\sqrt{3}L^7-2L^8$ $+\sqrt{3}L^9-L^{10})y_t$ $y_t-\sqrt{3}y_{t-1}+2y_{t-2}-\sqrt{3}y_{t-3}$ $+y_{t-4}-y_{t-6}-\sqrt{3}y_{t-7}-2y_{t-8}$ $+\sqrt{3}y_{t-9}-y_{t-10}$
H_a, H_b, H_c, H_d, H_e, H_f	H_g	anual no estacional bimensual cuatrimestral trimestral semestral frecuencia $\frac{5\pi}{6}$	$(1+\sqrt{3}L+2L^2+\sqrt{3}L^3+L^4-L^6$ $-\sqrt{3}L^7-2L^8-\sqrt{3}L^9-L^{10})y_t$ $y_t+\sqrt{3}y_{t-1}+2y_{t-2}+\sqrt{3}y_{t-3}$ $+y_{t-4}-y_{t-6}-\sqrt{3}y_{t-7}$ $-2y_{t-8}-\sqrt{3}y_{t-9}-y_{t-10}$
H_a, H_b	H_c, H_d, H_e, H_f, H_g	no estacional bimensual	$(1-L)(1+L)\,y_t$ $\left(1-L^2\right)y_t$ y_t-y_{t-2}
H_a	H_b, H_c, H_d, H_e, H_f, H_g	no estacional	$\Delta_1 y_t = y_t - y_{t-1}$
H_b	H_a, H_c, H_d, H_e, H_f, H_g	bimensual	$(1+L)\,y_t$ y_t+y_{t-1}

Cuadro 5: Interpretación de los resultados de la prueba HEGY mensual. Cont.

Hipótesis que no se rechaza	**Hipótesis que se rechaza**	**tipo de raíz unitaria presente**	**Transformación para obtener variable estacionaria**
H_c	H_a, H_b, H_d, H_e, H_f, H_g	cuatrimestral	$\left(1+L^2\right)y_t$ y_t+y_{t-2}
H_d	H_a, H_b, H_c, H_e, H_f, H_g	trimestral	$\left(1+L+L^2\right)y_t$ $y_t+y_{t-1}+y_{t-2}$
H_e	H_a, H_b, H_c, H_d, H_f, H_g	semestral	$\left(1-L+L^2\right)y_t$ $y_t-y_{t-1}+y_{t-2}$
H_f	H_a, H_b, H_c, H_d, H_e, H_g	frecuencia $\frac{5\pi}{6}$	$\left(1+\sqrt{3}L+L^2\right)y_t$ $y_t+\sqrt{3}y_{t-1}+y_{t-2}$
H_g	H_a, H_b, H_c, H_d, H_e, H_f	anual	$\left(1-\sqrt{3}L+L^2\right)y_t$ $y_t-\sqrt{3}y_{t-1}+y_{t-2}$
H_a	H_h	no estacional	$\Delta_1 y_t = y_t - y_{t-1}$
-	H_a, H_b, H_c, H_d, H_e, H_f, H_g	no hay	ninguna
-	H_i	no hay	ninguna
H_a, H_c, H_f	H_b, H_d, H_e, H_g	no estacional cuatrimestral frecuencia $\frac{5\pi}{6}$	$(1+\sqrt{3}L+L^2-L^4-\sqrt{3}L^5$ $-L^6)y_t$ $y_t+\sqrt{3}y_{t-1}+y_{t-2}$ $-y_{t-4}-\sqrt{3}y_{t-5}-y_{t-6}$

En las siguientes secciones se describirá como efectuar la prueba de HEGY tanto con datos mensuales como trimestrales empleando R.

5. Prueba de HEGY empleando R.

El software estadístico R cuenta con un paquete llamado "uroot" que permite realizar diversas pruebas de raíces unitarias, entre ellas la prueba de HEGY. Este paquete se puede instalar desde la consola de R mediante el siguiente comando:

```
> install.packages("uroot", repos="http://R-Forge.R-project.org")
```

Luego de haber sido instalado se debe cargar el paquete:

```
> library(uroot)
```

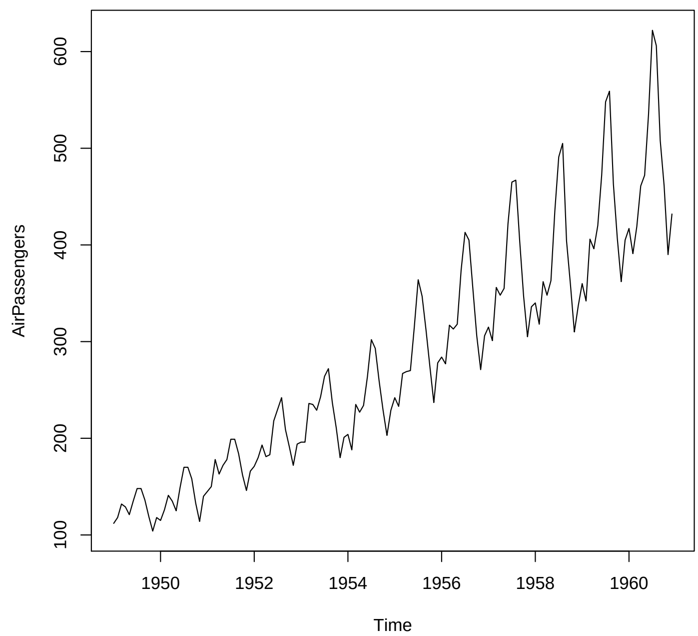
5.1. Datos mensuales

El primer paso para realizar la prueba de HEGY, es determinar que versión del modelo (11) se debe emplear. En otras palabras, se debe decidir si se necesita emplear una constante $\mu_t = \delta$, una tendencia $\mu_t = \beta t$, variables dummy para los meses $\mu_t = \sum_{i=1}^{11} \alpha_i D_{i,t}$ o los tres componentes $\mu_t = \delta + \beta t + \sum_{i=1}^{11} \alpha_i D_{i,t}$.

Lo mas frecuente al trabajar con series de tiempo es desconocer el proceso generador de los datos (DGP). Por lo tanto, es imposible conocer con seguridad cuál es la mejor opción (incluir intercepto, tendencia y/o variables dummy) para estimar la prueba de HEGY.

Primero, es buena idea graficar la serie con el fin de identificar la necesidad de modelar una posible presencia de una tendencia y/o estacionalidades no estocásticas que serían tratadas con variables dummy. Por ejemplo, en la Figura 4 se aprecia que la serie no crece ni presenta estacionalidades no estocásticas como se aprecia en el modelo 2 de la Figura 1 o en la serie clásica utilizada por Box & Jenkins del total mensual de pasajeros aéreos (ver Figura 2).

Figura 2: Serie mensual de pasajeros aéreos



Así, en el caso de la serie de la Figura 4 no será necesario incluir tendencia y variables dummy ($\mu_t = \delta$). Por otro lado, para las series correspondientes al modelo 2 de la Figura 1 y Figura 2, será necesario además incluir la tendencia y las variables dummy ($\mu_t = \delta + \beta t + \sum_{i=1}^{11} \alpha_i D_{i,t}$).

El segundo paso para realizar la prueba es determinar el número de rezagos óptimos p que se deben emplear al momento de estimar la prueba. Lo convencional, para determinar este número de rezagos es emplear criterios de información. Los criterios de información más empleados son:

- **Criterio de Akaike (aic):**

$$AIC = 2k + n\left[\ln\left(RSS\right)\right] \tag{12}$$

  donde k son el número de parámetros y RSS corresponde a la suma de los residuos al cuadrado.

- **Criterio de Schwarz (bic):**

$$BIC = n \times \ln(\hat{\sigma}_{\varepsilon}^{2}) + k \times \ln(n) \tag{13}$$

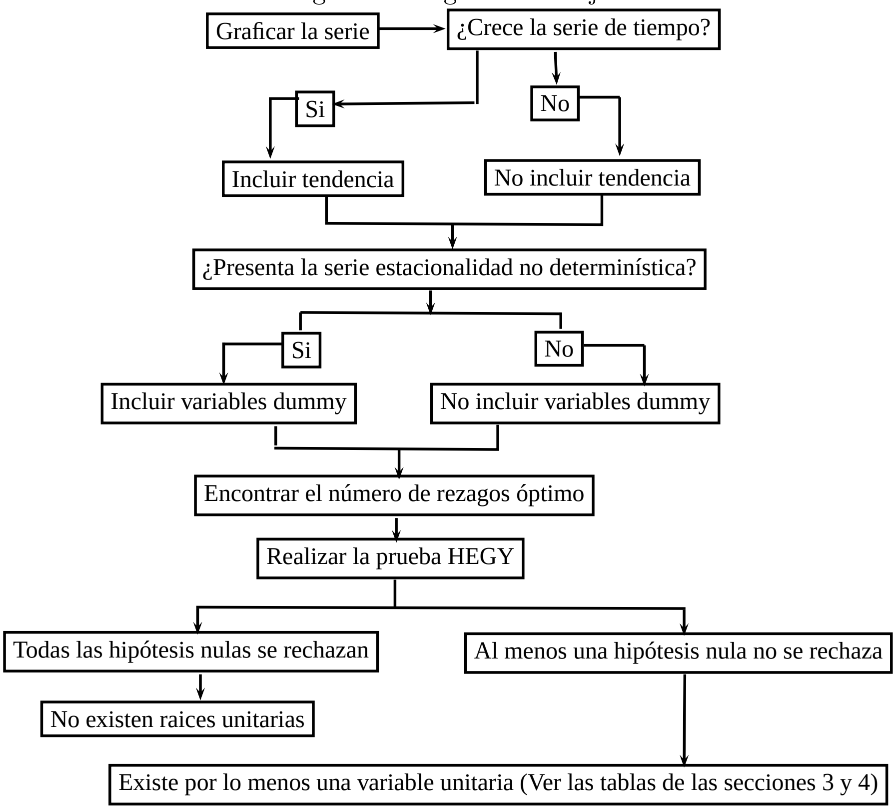
  donde k es el número de parámetros, n la cantidad de datos y $\hat{\sigma}_{\varepsilon}^{2}$ representa la varianza del error.

Cuando se emplean estos criterios de información, se selecciona el número de rezagos que minimiza el correspondiente criterio de información.

Otra opción para seleccionar el número de rezagos óptimo es ir removiendo los rezagos que no sean significativos (por ejemplo a un nivel del 10 % de significancia).

Los pasos para el análisis de raíces unitarias por medio de la prueba HEGY los podemos resumir mediante el siguiente diagrama de flujo (Ver Figura 3).

Figura 3: Diagrama de flujo



5.1.1. Prueba HEGY en *R*.

La función "HEGY.test" del paquete "uroot" permite elegir distintas opciones en varios argumentos. En general la función implica los siguientes argumentos:

```
\begin{center}
HEGY.test (wts, itsd, selectlags=list(mode="signf", Pmax=NULL))
\end{center}
```

donde:

- **wts**: Es la serie de tiempo a la que vamos a aplicar la prueba. Esta función detecta automáticamente si se trata de una serie mensual o trimestral
- **itsd**: Corresponde a una lista de tres elementos. Por ejemplo, "itsd=c(1,0,0)". Esta lista corresponde a los elementos determinísticos que se incluirán en el modelo. En otras palabras, corresponde a los elementos de μ_t. El primer y segundo parámetro determinan la inclusión o no de intercepto y tendencia en el modelo. El tercer elemento corresponde a las variables dummy. Para

la inclusión de uno de estos tres elementos se utiliza el "1" y "0" para la no inclusión. Por ejemplo "itsd=c(1,0,0)" implica la inclusión de intercepto pero no de tendencia ni de variables dummy estacionales ($\mu_t = \delta + \beta t$) , mientras que si se utiliza "itsd=c(1,0,c(1:11))" el modelo será estimado con intercepto, sin tendencia y con las once variables dummy mensuales ($\mu_t = \delta + \sum_{i=1}^{11} \alpha_i D_{i,t}$).

- **selectlags**: Es el método seleccionado para elegir el número de rezagos óptimo.Las opciones son "aic", "bic" y "signf" que representan el criterio de información de Akaike, el criterio de información de Schwarz y el criterio de significancia, respectivamente.

También es posible determinar manualmente el número máximo de rezagos a ser considerados para acotar la búsqueda de diferentes modelos en la prueba "Pmax". Por defecto esta cantidad corresponderá a: $10 \times log10(n)$, donde n es la cantidad de observaciones.

Para realizar un ejemplo de cómo opera esta función, generemos los datos de tal manera que la serie resultante presente alguno de los procesos no estacionarios que hemos descrito con anterioridad.

Con las siguientes líneas de código generamos la serie no estacionaria:

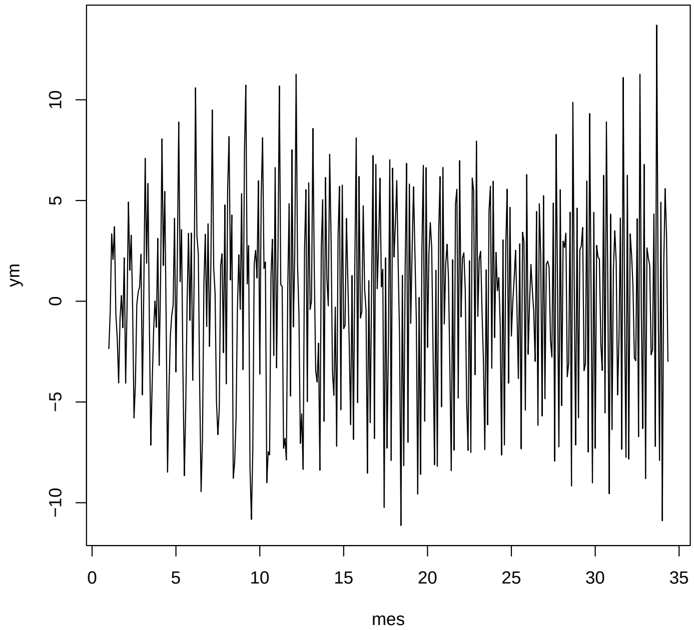
$$y_t = -y_{t-4} - y_{t-8} + \epsilon_t \tag{14}$$

esta serie tiene una raíz trimestral, otra semestral, una a la frecuencia $\frac{5\pi}{6}$, y una anual (Ver Cuadro 5).

```
> set.seed(1234567)
> e<-rnorm(500)
> ym<-matrix(0, 500,1)
> for (i in 13:500){
+ ym[i] <- -ym[i-4]-y1[i-8] + (e[i])  }
> plot.ts(y1t, ylab="ym", xlab=''mes'')
```

En la Figura 4 podemos ver el comportamiento de la serie.

Figura 4: Serie $ym_t = -ym_{t-4} - ym_{t-8} + \epsilon_t$



Para estimar la prueba de HEGY con los datos ym_t, debemos determinar primero si es necesario incluir una tendencia y variables dummy. En este caso de la Figura 4 podemos concluir que no son necesarias ni la tendencia ni las variables dummy [17]; en otras palabras, (itsd=c(1,0,0). Así, para estimar la prueba HEGY podemos emplear las siguientes líneas de código:

```
> ym<-ts(ym[100:500], frequency=12, start=1)
> HEGY.test(ym, itsd=c(1,0,0),
+ selectlags=list(mode="aic", Pmax=NULL))
```

Para seleccionar la cantidad de rezagos óptimo estamos usando el criterio de Akaike y finalmente dejamos por default la cantidad máxima de rezagos en que la prueba inicia el análisis de la serie (selectlags=list(mode=”aic”, Pmax=NULL)).

Ahora bien, al analizar los resultados de la prueba, concluimos que al 1 % de significancia las hipótesis nulas que se rechazan son las de raíz unitaria no estacional (H_a), bimensual (H_b) y cuatrimestral (H_c), correspondiendo al quinto caso del Cuadro 3, mientras que las otras hipótesis no se rechazan. En otras palabras, la prueba de HEGY permite concluir que esta serie tiene una raíz trimestral, otra semestral, una a la frecuencia $\frac{5\pi}{6}$, y una anual (Ver Cuadro 5).

[17]Naturalmente, en este caso como conocemos el DGP de los datos sabemos que estamos en lo correcto, pero en general esto no se podrá saber.

```
   ---- ----
  HEGY test
  ---- ----

  Null hypothesis: Unit root.
  Alternative hypothesis: Stationarity.

----
  HEGY statistics:

          Stat. p-value
tpi_1      -8.602   0.010
tpi_2     -10.397   0.010
Fpi_3:4    90.809   0.010
Fpi_5:6     0.036   0.100
Fpi_7:8     0.695   0.100
Fpi_9:10    1.829   0.100
Fpi_11:12   2.677   0.076
Fpi_2:12   26.945      NA
Fpi_1:12   32.208      NA

  Lag orders: 11 13 15 16 19 20 25
  Number of available observations: 364
```

Por lo tanto, la diferenciación necesaria para obtener una variable estacionaria es:

$$yms_t = ym_t + ym_{t-4} + ym_{t-8} \tag{15}$$

Luego de obtener la nueva variable se efectúa la prueba de HEGY sobre la variable yms_t.

```
> yms<-ym+lag(ym,-4)+lag(ym,-8)
> HEGY.test(yms, itsd=c(1,0,0),
+ selectlags=list(mode="aic", Pmax=NULL))
```

Los resultados de la prueba indican que la serie no presenta raíces unitarias estacionales ni tampoco no estacionales.

```
   ---- ----
  HEGY test
  ---- ----

  Null hypothesis: Unit root.
  Alternative hypothesis: Stationarity.

----
  HEGY statistics:
```

```
          Stat. p-value
tpi_1     -5.690    0.01
tpi_2     -4.578    0.01
Fpi_3:4   31.599    0.01
Fpi_5:6   23.948    0.01
Fpi_7:8   29.467    0.01
Fpi_9:10  25.680    0.01
Fpi_11:12 23.191    0.01
Fpi_2:12  13.911      NA
Fpi_1:12  12.948      NA

  Lag orders: 5 11 12 24 26
  Number of available observations: 355
```

5.2. Datos trimestrales

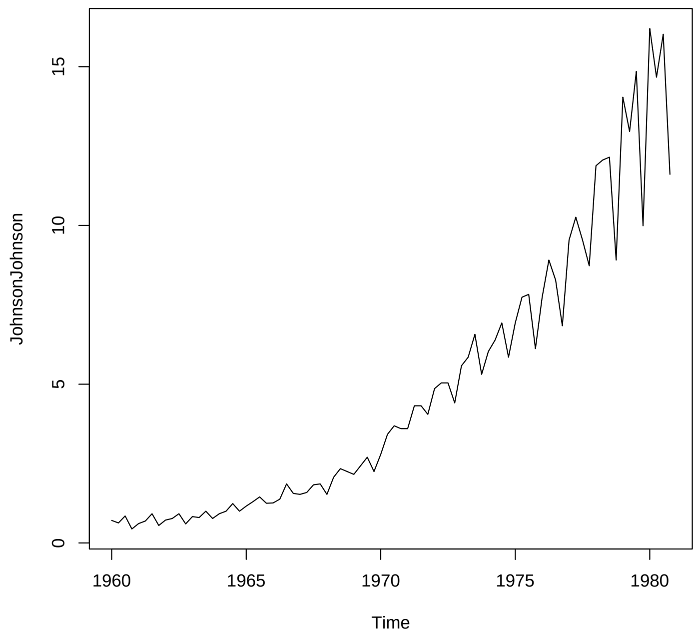
La aplicación de esta prueba para datos trimestrales es muy similar a lo explicado para datos mensuales. En este caso, usemos datos reales correspondientes a las ganancias trimestrales (en dólares) por acción de Johnson & Johnson desde 1960 hasta 1980.

Para emplear estos datos, debemos instalar el paquete “datasets” y posteriormente cargarlo junto con los datos “JohnsonJohnson”. Esto lo haremos con las siguientes líneas.

```
> install.packages("datasets", repos="http://R-Forge.R-project.org")
> library(datasets)
> data(JohnsonJohnson)
> plot(JohnsonJohnson)
```

En la Figura 5 podemos ver el comportamiento de la serie.

Figura 5: Ganancias por acción de Johnson & Johnson



Dado que desconocemos el proceso generador de los datos (DGP), debemos utilizar inicialmente el análisis gráfico. En la Figura 5 se aprecia que la serie crece en el tiempo, en cuanto a la presencia de estacionalidades no estocásticas no podemos hacer ninguna conclusión con certeza. Por lo tanto, siguiendo el algoritmo presentado en la Figura 3 se debe estimar primero el modelo mas general con variables dummy trimestrales y luego examinar su significancia.

Para estimar la prueba de HEGY a los datos *JohnsonJohnson* se utiliza las siguientes líneas de código:

```
> hegy.out2<-HEGY.test(wts=JohnsonJohnson,
+ itsd=c(1,1,c(1:3)), regvar=0,
+ selectlags=list(mode="aic", Pmax=NULL))
```

Note que en la prueba se ha incluido intercepto y tendencia dado que la serie crece. Además incluimos variables dummy trimestrales (itsd=c(1,1,c(1:3))).Para seleccionar la cantidad de rezagos estamos usando el criterio de AkaikeSin embargo es aconsejable realizar la prueba con los tres métodos y examinar si las conclusiones varían. y estamos dejando por default la cantidad máxima de rezagos en que la prueba inicia el análisis de la serie (selectlags=list(mode="aic", Pmax=NULL)).

En el summary del objeto "hegy.out2" podemos verificar la significancia de la tendencia y de las variables dummy incluidas.

```
> summary(hegy.out2)

  ---- ----
  HEGY test
  ---- ----

----
  Deterministic regressors estimates:

             Estimate Std. Error t value Pr(>|t|)
Intercept     -0.524      0.370  -1.415    0.164
Trend          0.027      0.014   1.977    0.054
SeasDummy.1    0.062      0.184   0.336    0.738
SeasDummy.2    0.139      0.180   0.771    0.445
SeasDummy.3   -0.077      0.182  -0.424    0.673

----
```

Podemos ver que ninguna de las dummy son significativas, por lo tanto reestimemos la prueba sin incluir estas variables.

```
> hegy.out3<-HEGY.test(wts=JohnsonJohnson,
+ itsd=c(1,1,c(0)), regvar=0,
+ selectlags=list(mode="aic", Pmax=NULL))
```

Ahora bien, al analizar los resultados de la prueba, concluimos que al 1 % de significancia no es posible rechazar las hipótesis nulas de raíz no estacional (H_a) y semestral (H_b). Correspondiendo al segundo caso del Cuadro 1.

```
  ---- ----
  HEGY test
  ---- ----

  Null hypothesis: Unit root.
  Alternative hypothesis: Stationarity.

----
  HEGY statistics:

         Stat. p-value
tpi_1    1.384    0.10
tpi_2    4.323    0.10
Fpi_3:4 14.934    0.01
Fpi_2:4 10.971      NA
Fpi_1:4  8.237      NA
```

```
  Lag orders: 1 2 4 5 7 8 9 12 16 17
  Number of available observations: 63
```

Por lo tanto, la diferenciación necesaria para obtener una variable estacionaria es:

$$JohnsonJohnson2_t = JohnsonJohnson_t - JohnsonJohnson_{t-2} \quad (16)$$

Luego de obtener la nueva serie de tiempo diferenciada se efectúa la prueba de HEGY sobre la variable $JohnsonJohnson2_t$.

```
> JohnsonJohnson2<-JohnsonJohnson-lag(JohnsonJohnson,-2)
> HEGY.test(JohnsonJohnson2, itsd=c(1,1,0),
+ selectlags=list(mode="aic", Pmax=NULL))
```

Los resultados de la prueba indican que la serie no presenta raíces unitarias estacionales ni tampoco no estacionales con un nivel de significancia del 5 %.

```
  ---- ----
  HEGY test
  ---- ----

  Null hypothesis: Unit root.
  Alternative hypothesis: Stationarity.

----
  HEGY statistics:

        Stat. p-value
tpi_1   -3.727   0.029
tpi_2   -3.855   0.010
Fpi_3:4  7.454   0.010
Fpi_2:4  9.217      NA
Fpi_1:4  7.186      NA

  Lag orders: 1 2 3 4 6 8 9 10 11 12 13 14 17 19
  Number of available observations: 59
```

6. Comentarios finales

En este documento hemos discutido la intuición de la prueba de HEGY tanto para datos mensuales como trimestrales. Así mismo presentamos de manera didáctica como interpretar los resultados de esta prueba. Finalmente, mostramos como aplicar esta prueba empleando el software estadístico R y el paquete "uroot".

Esta prueba permitirá al estudiante modelar de manera adecuada una serie trimestral o mensual al poder determinar si es necesario algún tipo de diferenciación al momento de construir modelos univariados SARIMA.

Referencias

R. F. Engle, C. W. J. Granger, y J. J. Hallman. Merging short-and long-run forecasts : An application of seasonal cointegration to monthly electricity sales forecasting. *Journal of Econometrics*, 40(1):45–62, 1989.

P. H. Franses. *Model selection and seasonality in time series.* Tinbergen Institute research series. Thesis/Tinbergen Instituut, Amsterdam, 1991a. 92215842 Philip Hans Franses. ill. ; 23 cm. Thesis (doctoral)–Erasmus University Rotterdam, 1991. Summary in Dutch. Includes bibliographical references (p. 169-180).

P. H. Franses. Seasonality, non-stationarity and the forecasting of monthly time series. *International Journal of Forecasting*, 7(2):199–208, 1991b.

S. Hylleberg, R. F. Engle, C. W. J. Granger, y B. S. Yoo. Seasonal integration and cointegration. *Journal of Econometrics*, 44(1-2):215–238, 1990.

Anexo

En este anexo se presentan 3 demostraciones de resultados que se emplean en el texto.

Demostración de resultado empleado en la Condición 4 para datos mensuales.

Si partimos del hecho que bajo esta la **Condición 4** implica que $y_t + y_{t-1} + y_{t-2}$ será una serie estacionaria, entonces tendremos que:

$$y_t \approx -y_{t-1} - y_{t-2} \tag{17}$$

este resultado también debe ser cierto para y_{t-1}, es decir:

$$y_{t-1} \approx -y_{t-2} - y_{t-3}$$

Reemplazando esta última expresión en (17) tenemos que:

$$[y_t \approx y_{t-2} + y_{t-3} - y_{t-2}$$

Es decir,

$$y_t \approx y_{t-3}$$

Demostración de resultado empleado en la Condición 5 para datos mensuales.

La **Condición 5** implica que $y_t - y_{t-1} + y_{t-2}$ será una serie estacionaria, entonces tendremos que:

$$y_t \approx y_{t-1} - y_{t-2} \tag{18}$$

y por lo tanto tendremos que:

$$y_{t-1} \approx y_{t-2} - y_{t-3}$$

remplazando está ultima expresión para y_{t-1} en (18) se obtiene:

$$y_t \approx y_{t-2} - y_{t-3} - y_{t-2}$$

$$y_t \approx -y_{t-3}$$

y por lo tanto tendremos que

$$y_t \approx y_{t-6}$$

Demostración de resultado empleado en la Condición 7 para datos mensuales.

La **Condición 5** implica que $y_t - \sqrt{3}y_{t-1} + y_{t-2}$ será una serie estacionaria, entonces tendremos que:

$$y_t \approx \sqrt{3}y_{t-1} - y_{t-2} \tag{19}$$

y por lo tanto tambien se tiene que:

$$y_{t-1} \approx \sqrt{3}y_{t-2} - y_{t-3}$$

reemplazando en (19) tendremos que:

$$y_t \approx 2y_{t-2} - \sqrt{3}y_{t-3}$$

Ahora, empleando el hecho que $y_{t-2} \approx \sqrt{3}y_{t-3} - y_{t-4}$, tenemos que:

$$y_t \approx \sqrt{3}y_{t-3} - 2y_{t-4}$$

Ahora, reemplazando $y_{t-3} \approx \sqrt{3}y_{t-4} - y_{t-5}$ en la expresión anterior se obtiene:

$$y_t \approx y_{t-4} - \sqrt{3}y_{t-5}$$

Finalmente, dado que $y_{t-4} \approx \sqrt{3}y_{t-5} - y_{t-6}$ tenemos que:

$$y_t \approx \sqrt{3}y_{t-5} - y_{t-6} - \sqrt{3}y_{t-5}$$

$$y_t \approx -y_{t-6}$$

Y por lo tanto tedremos que

$$y_t \approx y_{t-12}$$