Análisis de series de tiempo

Diplomado

Orlando Joaqui-Barandica

Universidad del Valle

2023

Contenido

Contenido

Contenido

Modelos ARIMA

Los modelos ARIMA proporcionan otro enfoque para el pronóstico de series de tiempo. El suavizado exponencial y los modelos ARIMA son los dos enfoques más utilizados para la predicción de series de tiempo, y proporcionan enfoques complementarios al problema.

Mientras que los modelos de suavizado exponencial se basan en una descripción de la tendencia y la estacionalidad en los datos, los modelos ARIMA tienen como objetivo describir las **autocorrelaciones** en los datos.

Antes de presentar los modelos ARIMA, primero debemos analizar el concepto de estacionariedad y la técnica de diferenciar series temporales.

Estacionariedad

Una serie temporal estacionaria es aquella cuyas propiedades no dependen del momento en que se observa la serie.

Por lo tanto, las series temporales con tendencias, o con estacionalidad, no son estacionarias:

- La tendencia y la estacionalidad afectarán el valor de la serie temporal en diferentes momentos.
- Por otro lado, una serie de ruido blanco es estacionaria: no importa cuando la observe, debería verse muy similar en cualquier momento.

Estacionariedad

Una serie temporal estacionaria es aquella cuyas propiedades no dependen del momento en que se observa la serie.

Por lo tanto, las series temporales con tendencias, o con estacionalidad, no son estacionarias:

- La tendencia y la estacionalidad afectarán el valor de la serie temporal en diferentes momentos.
- Por otro lado, una serie de ruido blanco es estacionaria: no importa cuando la observe, debería verse muy similar en cualquier momento.

Algunos casos pueden ser confusos: una serie temporal con comportamiento cíclico (pero sin tendencia ni estacionalidad) es estacionaria.

Estacionariedad

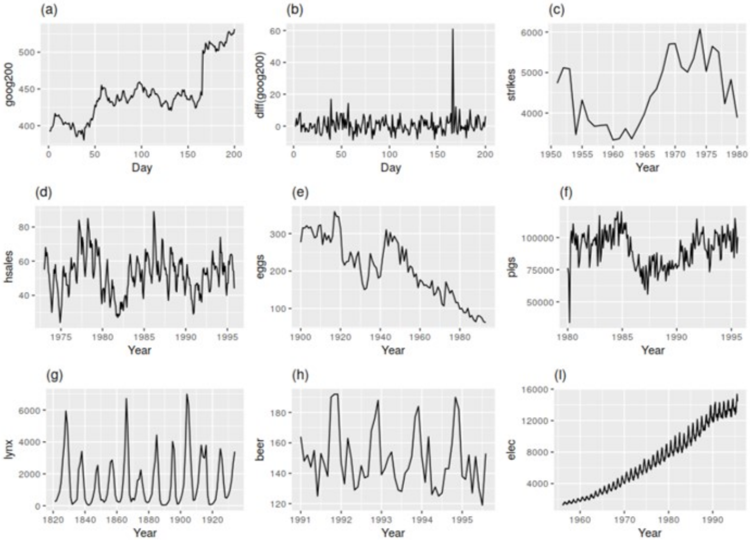
Definición

Si y_t es una serie de tiempo estacionaria, entonces para todos los periodos s, la distribución de $(y_1, \ldots, y_{t+s})$ no depende de t

Una serie estacionaria es:

- Aproximadamente horizontal (aunque es posible algún comportamiento cíclico)
- Presenta varianza constante
- No tendrá patrones predecibles a largo plazo

Estacionarias?



Estacionarias?

- La estacionalidad obvia descarta las series **(d)**, **(h)** e **(i)**.
- Las tendencias y los niveles cambiantes descartan las series **(a)**, **(c)**, **(e)**, **(f)** e **(i)**.
- La varianza creciente también descarta **(i)**.
- Eso deja solo (b) y (g) como series estacionarias.

A primera vista, los ciclos fuertes en la serie (g) pueden parecer no estacionarios. Pero estos ciclos son aperiódicos: A largo plazo, el momento de estos ciclos no es predecible. Por lo tanto, la serie es estacionaria.

Estacionariedad

Las transformaciones ayudan a **estabilizar la varianza**. Para los modelos ARIMA, también se necesita **estabilizar la media**

Identificar series no estacionarias:

- Gráfico de serie de tiempo
- La ACF de datos estacionarios cae a cero relativamente rápido
- La ACF de datos no estacionarios decrece lentamente
- Para los datos no estacionarios el valor de r_1 es a menudo grande y positivo

Contenido

Diferenciación

Definición

Una serie diferenciada es el cambio entre cada observación en la serie original:

$$y'_t = y_t - y_{t-1} \tag{1}$$

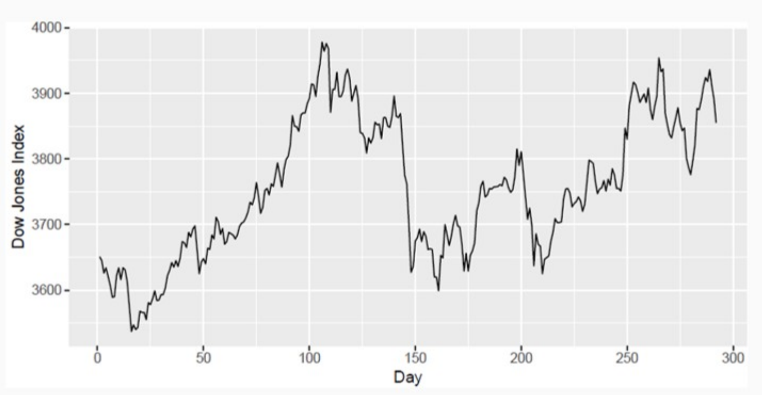
- La diferenciación ayuda a estabilizar la media
- La diferenciación tendrá solo $(T - 1)$ valores ya que no es posible calcular una diferencia y'_1 para la primera observación

Ejemplo: Índice Dow-Jones

```
autoplot(dj)
```

Ejemplo: Índice Dow-Jones

```
autoplot(dj)
```

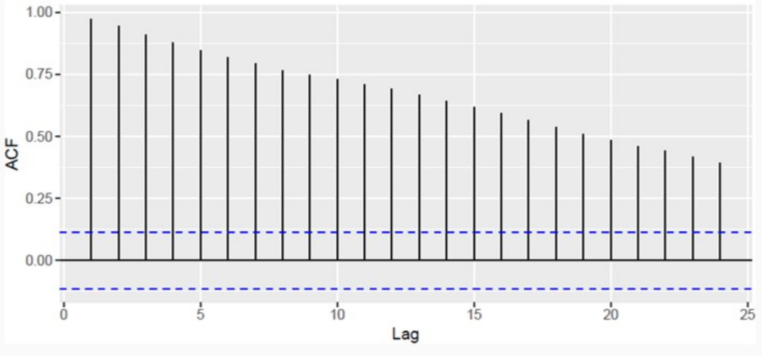


Ejemplo: Índice Dow-Jones

```
ggAcf(dj)
```

Ejemplo: Índice Dow-Jones

```
ggAcf(dj)
```

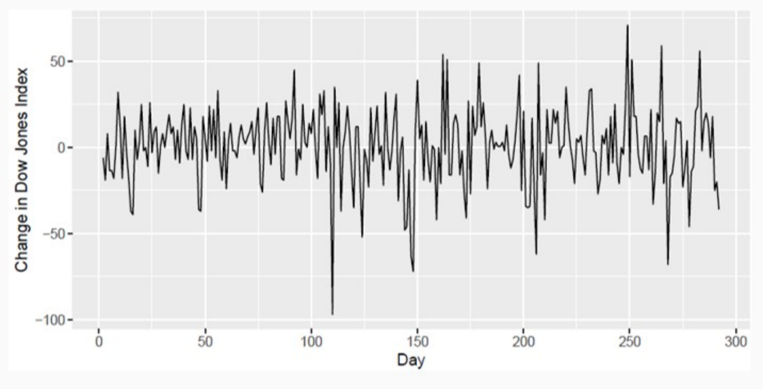


Ejemplo: Índice Dow-Jones

```
diff(dj)
```

Ejemplo: Índice Dow-Jones

```
diff(dj)
```

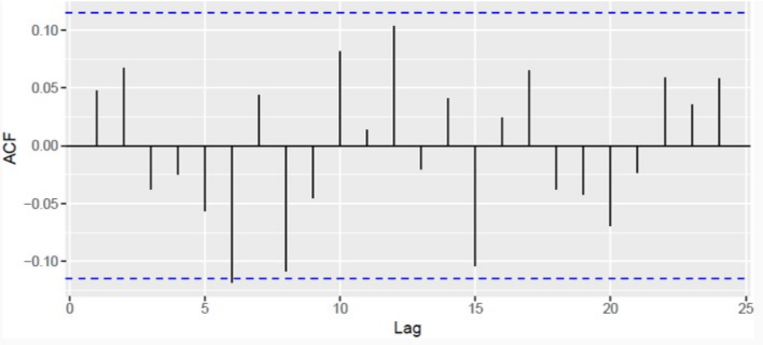


Ejemplo: Índice Dow-Jones

```
ggAcf(diff(dj))
```

Ejemplo: Índice Dow-Jones

```
ggAcf(diff(dj))
```



Ejercicio

Evalúe la estacionariedad para la serie goog200. Posteriormente realice la diferenciación y evalúe la estacionariedad sobre los cambios en la serie goog200.

Realice la prueba de Ljung-Box para cada series (Original y diferenciada) para determinar si existe correlación entre las observaciones. Utilice lag $= 10$.

¿Que puede concluir?

Ejercicio

Evalúe la estacionariedad para la serie goog200. Posteriormente realice la diferenciación y evalúe la estacionariedad sobre los cambios en la serie goog200.

Realice la prueba de Ljung-Box para cada series (Original y diferenciada) para determinar si existe correlación entre las observaciones. Utilice lag $= 10$.

¿Que puede concluir?

```
Box.test(goog200, lag=10, type="Ljung-Box")
Box.test(diff(goog200), lag=10, type="Ljung-Box")
```

Contenido

Diferenciación de segundo orden

Ocasionalmente, los datos diferenciados no parecerán estacionarios y puede ser necesario diferenciar los datos por segunda vez para obtener una serie estacionaria:

$$\begin{aligned} y_t'' &= y_t' - y_{t-1}' && (2) \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) && (3) \\ &= y_t - 2y_{t-1} + y_{t-2} && (4) \end{aligned}$$

- En este caso, y_t'' tendrá $T-2$ valores
- En la práctica, casi nunca es necesario ir más allá de las diferencias de segundo orden.

Contenido

Diferenciación Estacional

Una diferencia estacional es la diferencia entre una observación y la observación previa de la misma estación. Entonces

$$y'_t = y_t - y_{t-m} \tag{5}$$

- m = número de estaciones
- m = 12 (Para datos mensuales)
- m = 4 (Para datos trimestrales)

Ejemplo

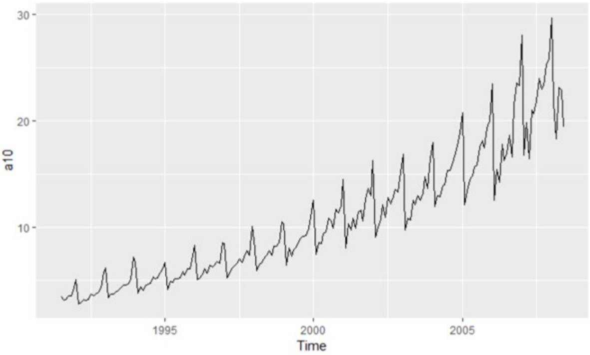
Analicemos la serie a10 (Venta de medicamentos antidiabéticos). Es Estacionaria la serie?

```
autoplot(a10)
```

Ejemplo

Analicemos la serie a10 (Venta de medicamentos antidiabéticos). Es Estacionaria la serie?

```
autoplot(a10)
```



Ejemplo

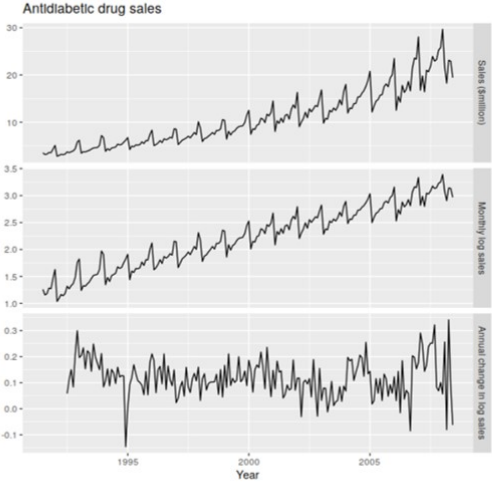
```
cbind("Sales ($million)" = a10,
     "Monthly log sales" = log(a10),
     "Annual change in log sales" = diff(log(a10),12)) %>%
 autoplot(facets=TRUE) +
  xlab("Year") + ylab("") +
  ggtitle("Antidiabetic drug sales")
```

Ejemplo

```
cbind("Sales ($million)" = a10,
    "Monthly log sales" = log(a10),
    "Annual change in log sales" = diff(log(a10),12)) %>%
 autoplot(facets=TRUE) +
  xlab("Year") + ylab("") +
  ggtitle("Antidiabetic drug sales")
```

- Para distinguir las diferencias estacionales de las diferencias ordinarias, a veces nos referimos a las diferencias ordinarias como **"primeras diferencias"**, lo que significa diferencias en el rezago 1.
- La transformación y la diferenciación harán que la serie parezca relativamente estacionaria.

Ejemplo

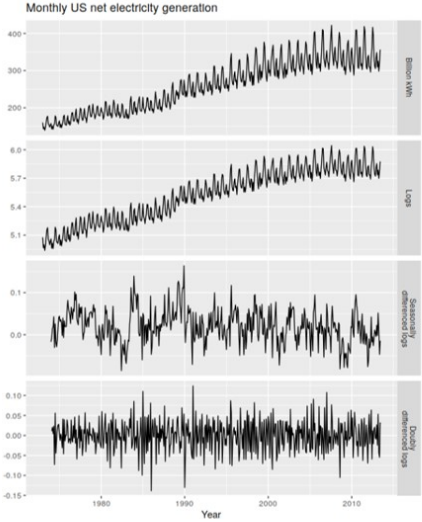

Antidiabetic drug sales
30
20
10
Sales ($million)
3.5
3.0
2.5
2.0
1.5
1.0
Monthly log sales
0.3
0.2
0.1
0.0
-0.1
Annual change in log sales
1995
2000
2005
Year

Ejercicio

A veces es necesario tomar una diferencia estacional y una primera diferencia para obtener datos estacionarios.

De tal manera, diferencie estacionalmente la serie usmelec (generación de electricidad). Es decir, como el ejemplo anterior realice la transformación, diferencia estacional y posteriormente diferencie la diferenciación estacional.

Ejercicio


Monthly US net electricity generation
400
300
200
Billion kWh
6.0
5.7
5.4
5.1
Logs
0.1
0.0
Seasonally differenced logs
0.10
0.05
0.00
-0.05
-0.10
-0.15
Doubly differenced logs
1980
1990
2000
2010
Year

Ejercicio

- Hay un grado de subjetividad en la selección de las diferencias a aplicar.
- Los datos diferenciados estacionalmente en a10 no muestran un comportamiento sustancialmente diferente de los datos diferenciados estacionalmente en usmelec.
- En el último caso, podríamos haber decidido detenernos con los datos diferenciados estacionalmente, y no haber hecho una ronda adicional de diferenciación.
- En el primer caso, podríamos haber decidido que los datos no eran lo suficientemente estacionarios y haber tomado una ronda adicional de diferenciación.

Siempre hay que tomar algunas decisiones en el proceso de modelado, y diferentes analistas pueden tomar diferentes decisiones.

Ejercicio

Código usmelec.

```
cbind("Billion kWh" = usmelec,
     "Logs" = log(usmelec),
     "Seasonally\n differenced logs" =
       diff(log(usmelec),12),
     "Doubly\n differenced logs" =
       diff(diff(log(usmelec),12),1)) %>%
  autoplot(facets=TRUE) +
    xlab("Year") + ylab("") +
    ggtitle("Monthly US net electricity generation")
```

Diferenciación Estacional

Cuando se aplican tanto las diferencias estacionales como las primeras, no importa qué se haga primero: el resultado será el mismo.

- Sin embargo, si los datos tienen un patrón estacional fuerte, ***se recomienda*** hacer primero la diferenciación estacional, porque las series resultantes a veces serán estacionarias y no habrá necesidad de una primera diferencia adicional.
- Si la primera diferenciación se realiza primero, aún habrá estacionalidad presente.

Diferenciación Estacional

Cuando se aplican tanto las diferencias estacionales como las primeras, no importa qué se haga primero: el resultado será el mismo.

- Sin embargo, si los datos tienen un patrón estacional fuerte, ***se recomienda*** hacer primero la diferenciación estacional, porque las series resultantes a veces serán estacionarias y no habrá necesidad de una primera diferencia adicional.
- Si la primera diferenciación se realiza primero, aún habrá estacionalidad presente.

Ojo con las interpretaciones de las diferenciaciones

Las primeras diferencias son el cambio entre una observación y la siguiente. Las diferencias estacionales son el cambio entre un año y el siguiente. Es poco probable que otros retrasos tengan mucho sentido interpretable y deben evitarse.

Contenido

Test de raíz unitaria

Estas son pruebas de hipótesis estadísticas de estacionariedad diseñadas para determinar si se requiere diferenciación.

Hay disponibles varias pruebas de raíz unitaria, que se basan en supuestos diferentes y pueden dar lugar a respuestas conflictivas.

1. Test aumentado de Dickey Fuller: Hipótesis nula es que los datos son no estacionarios y no estacionales.
2. Test Kwiatkowski-Phillips-Schmidt-Shin (KPSS): Hipótesis nula es que los datos son estacionarios y no estacionales.
3. Otros test disponibles para datos estacionales.

Test de raíz unitaria

Test KPSS

```
library(urca)
summary(ur.kpss(goog))
```

Test de raíz unitaria

Test KPSS

```
library(urca)
summary(ur.kpss(goog))
```

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 7 lags.

Value of test-statistic is: 10.7223

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Test de raíz unitaria

Test KPSS

```
library(urca)
summary(ur.kpss(goog))
```

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 7 lags.

Value of test-statistic is: 10.7223

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Interpretación

El estadístico de prueba **(10.72)** es mucho mayor que el valor crítico del 1 % (0.739) , lo que indica que la hipótesis nula es rechazada. Es decir, los datos no son estacionarios. Podemos diferenciar los datos y aplicar la prueba nuevamente.

Test de raíz unitaria

Test aumentado de Dickey Fuller

```
library(tseries)
adf.test(goog)
```

Test de raíz unitaria

Test aumentado de Dickey Fuller

```
library(tseries)
adf.test(goog)
```

```
Augmented Dickey-Fuller Test

data:  goog
Dickey-Fuller = -2.5417, Lag order = 9, p-value = 0.349
alternative hypothesis: stationary
```

Test de raíz unitaria

Test aumentado de Dickey Fuller

```
library(tseries)
adf.test(goog)
```

```
Augmented Dickey-Fuller Test

data:  goog
Dickey-Fuller = -2.5417, Lag order = 9, p-value = 0.349
alternative hypothesis: stationary
```

Interpretación

No rechaza la hipótesis nula. Por lo tanto los datos de goog son no estacionarios

Test de raíz unitaria

```
library(urca)
summary(ur.kpss(diff(goog)))
```

Test de raíz unitaria

```r
library(urca)
summary(ur.kpss(diff(goog)))
```

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 7 lags.

Value of test-statistic is: 0.0324

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Test de raíz unitaria

```
library(urca)
summary(ur.kpss(diff(goog)))
```

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 7 lags.

Value of test-statistic is: 0.0324

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

Interpretación

Esta vez, el estadístico de prueba **(0.0324)** es pequeño y está dentro del rango que esperaríamos para los datos estacionarios. Entonces podemos concluir que los datos diferenciados son estacionarios.

Test de raíz unitaria

La función lleva a cabo este proceso de usar una secuencia de pruebas KPSS para determinar el número apropiado de primeras diferencias

```
ndiffs(goog)
```

Test de raíz unitaria

La función lleva a cabo este proceso de usar una secuencia de pruebas KPSS para determinar el número apropiado de primeras diferencias

```
ndiffs(goog)
```

```
[1] 1
```

Como vimos en las pruebas de KPSS anteriores, se requiere una diferencia para que los datos de goog sean estacionarios.

Test de raíz unitaria

La función lleva a cabo este proceso de usar una secuencia de pruebas KPSS para determinar el número apropiado de primeras diferencias

```
ndiffs(goog)
```

```
[1] 1
```

Como vimos en las pruebas de KPSS anteriores, se requiere una diferencia para que los datos de goog sean estacionarios.

Una función similar para determinar si se requiere diferenciación estacional es la nsdiffs() para determinar el número apropiado de diferencias estacionales requeridas.

Descomposición STL: $y_t = T_t + S_t + R_t$
Fuerza estacional: $F_s = max(0, 1 - \frac{Var(R_t)}{Var(S_t + R_t)})$
Si $F_s > 0{,}64$ hacer una diferenciación estacional.

Test de raíz unitaria

Podemos aplicar **nsdiffs()** a los datos de electricidad mensuales registrados en los EE. UU.

```
usmelec %>% log() %>% nsdiffs()
```

```
[1] 1
```

```
usmelec %>% log() %>% diff(lag=12) %>% ndiffs()
```

```
[1] 1
```

- Debido a que nsdiffs( )devuelve 1 (lo que indica que se requiere una diferencia estacional), aplicamos la función ndiffs() a los datos diferenciados estacionalmente.
- Estas funciones sugieren que deberíamos hacer una diferencia estacional y una primera diferencia.

Test de raíz unitaria

Test aumentado de Dickey Fuller

```
library(tseries)
adf.test(diff(goog))
```

Test de raíz unitaria

Test aumentado de Dickey Fuller

```
library(tseries)
adf.test(diff(goog))
```

```
Augmented Dickey-Fuller Test

data:  diff(goog)
Dickey-Fuller = -11.157, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary
```

Test de raíz unitaria

Test aumentado de Dickey Fuller

```
library(tseries)
adf.test(diff(goog))
```

```
Augmented Dickey-Fuller Test

data:  diff(goog)
Dickey-Fuller = -11.157, Lag order = 9, p-value = 0.01
alternative hypothesis: stationary
```

Interpretación

Rechaza la hipótesis nula. Por lo tanto los datos de goog son estacionarios

Contenido

Modelos Autorregresivos

En un modelo de regresión múltiple, pronosticamos la variable de interés utilizando una combinación lineal de predictores.

En un modelo autorregresivo, pronosticamos la variable de interés utilizando una combinación lineal de valores pasados de la variable. El término auto regresión indica que es una regresión de la variable contra sí misma.

Modelo autorregresivo de **orden p**

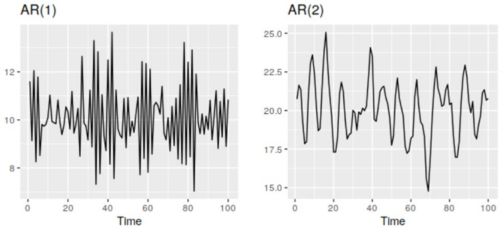
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t \quad (6)$$

dónde ε_t es ruido blanco.

Nos referimos como un AR(p) a un modelo autorregresivo de orden p.

Modelos Autorregresivos

Los modelos autorregresivos son notablemente flexibles para manejar una amplia gama de patrones de series temporales diferentes.



Cambiar los parámetros $\phi_1, ..., \phi_p$ da como resultado diferentes patrones de series de tiempo. La varianza del término de error ε_t sólo cambiara la escala de la serie, no los patrones.

Modelos Autorregresivos

Para un modelo **AR(1)**:

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t \tag{7}$$

- Cuando $\phi_1 = 0$, y_t es equivalente al ruido blanco.
- Cuando $\phi_1 = 1$ y $C = 0$, y_t es equivalente a una caminata aleatoria.
- Cuando $\phi_1 = 1$ y $C \neq 0$, y_t es equivalente a una caminata aleatoria con deriva.
- Cuando $\phi_1 < 0$, y_t tiende a oscilar alrededor de la media.

Modelos Autorregresivos

Normalmente restringimos los modelos autorregresivos a datos estacionarios, en cuyo caso se requieren algunas restricciones en los valores de los parámetros.

- Para un modelo AR(1): $-1 < \phi_1 < 1$
- Para un modelo AR(2): $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$
- Para un modelo AR(3): Las condiciones son más complicadas. El software se encarga de estas restricciones al estimar el modelo.

Contenido

Modelo de media móvil

En lugar de usar valores pasados de la variable de pronóstico en una regresión, un modelo de promedio móvil usa errores de pronóstico pasados en un modelo similar a la regresión.

Modelo de medias móviles de **orden q**

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{8}$$

dónde ε_t es ruido blanco.

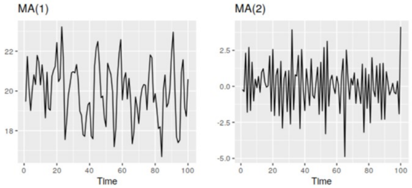
Por supuesto, no observamos los valores de ε_t por lo que no es realmente una regresión en el sentido habitual.

Nos referimos como un MA(q) a un modelo de medias móviles de orden q.

Modelo de media móvil

Cada valor de y_t puede considerarse como un promedio móvil ponderado de los últimos errores de pronóstico. Sin embargo, los modelos de promedio móvil no deben confundirse con el suavizado del promedio móvil.

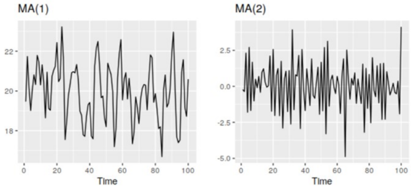
Se usa un modelo de promedio móvil para pronosticar valores futuros, mientras que el suavizado de promedio móvil se usa para estimar el ciclo de tendencia de valores pasados.



Modelo de media móvil

Cada valor de y_t puede considerarse como un promedio móvil ponderado de los últimos errores de pronóstico. Sin embargo, los modelos de promedio móvil no deben confundirse con el suavizado del promedio móvil.

Se usa un modelo de promedio móvil para pronosticar valores futuros, mientras que el suavizado de promedio móvil se usa para estimar el ciclo de tendencia de valores pasados.



Cambiar los parámetros $\theta_1, \ldots, \theta_q$ da como resultado diferentes patrones de series de tiempo. La varianza del término de error ε_t sólo cambiara la escala de la serie, no los patrones.

Modelos MA(∞)

Es posible escribir cualquier modelo estacionario AR(p) como un modelo MA(∞). Por ejemplo, usando la sustitución repetida, podemos demostrar esto para un modelo AR(1):

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \varepsilon_t \\ &= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \phi_1^3 y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\ &\dots \end{aligned}$$

- Previsto $-1 < \phi_1 < 1$:

$$y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \dots \tag{9}$$

Invertibilidad

- Cualquier proceso $MA(q)$ puede escribirse como un proceso $AR(\infty)$ si imponemos algunas restricciones en los parámetros MA
- Entonces el modelo MA es llamado "Invertible"
- Los modelos invertibles tienen algunas propiedades matemáticas que los hace más fácil de usar en la práctica.

Invertibilidad

Considere un proceso $MA(1)$, $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$. En su representación $AR(\infty)$, el error más reciente puede ser escrito como una función lineal de las observaciones actuales y pasadas.

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\theta)^j y_{t-j} \tag{10}$$

- Requerimos que $|\theta| < 1$, por lo que las observaciones más recientes tienen mayor peso que las observaciones del pasado más lejano. Por lo tanto el proceso es invertible cuando $|\theta| < 1$

Invertibilidad

Condición general de invertibilidad

Las raíces complejas de $1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$ se encuentran fuera del circulo unitario en el plano complejo.

Las restricciones de invertibilidad para otros modelos son similares a las restricciones de estacionariedad.

- Para un modelo MA(1): $-1 < \theta_1 < 1$
- Para un modelo MA(2):
  $-1 < \theta_2 < 1 \qquad \theta_2 + \theta_1 > -1 \qquad \theta_1 - \theta_2 < 1.$
- Para un modelo MA(3): Las condiciones son más complicadas. El software se encarga de estas restricciones al estimar el modelo.

Contenido

ARIMA

Si combinamos la diferenciación con autorregresión y un modelo de media móvil, obtenemos un modelo ARIMA no estacional. ARIMA es un acrónimo de AutoRegressive Integrated Moving Average (en este contexto, “integración” es lo contrario de la diferenciación).

ARMA:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

- Los predictores incluyen valores rezagados de y_t y errores rezagados.
- Las condiciones sobre los coeficientes aseguran la estacionariedad.
- Las condiciones sobre los coeficientes aseguran la invertibilidad.

ARIMA

AutoRegressive Integrated Moving Average (ARIMA)

ARIMA(p, d, q)

AR: $p =$ orden de la parte autorrgresiva
I: $d =$ grado de la primera diferenciación
MA: $q =$ orden de la parte de medias móviles

- Modelo ruido blanco: ARIMA(0,0,0)
- Caminata aleatoria: ARIMA(0,1,0) sin constante
- Caminata aleatoria con deriva: ARIMA(0,1,0) con constante
- AR(p): ARIMA(p,0,0)
- MA(q): ARIMA(0,0,q)

Ejemplo

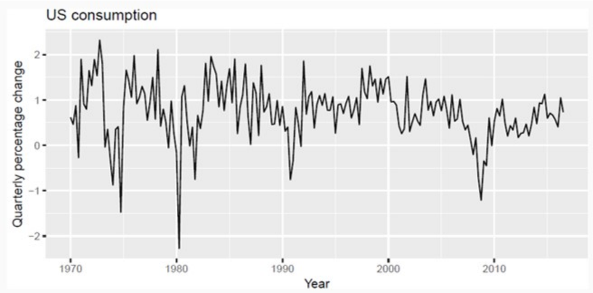
La serie (*uschange*[, "*Consumption*"]) muestra los cambios porcentuales trimestrales en el gasto de consumo de los Estados Unidos. Aunque es una serie trimestral, no parece haber un patrón estacional, por lo que ajustaremos un **modelo ARIMA no estacional**.

```
autoplot(uschange[,"Consumption"]) +
  xlab("Year") + ylab("Quarterly percentage change")
```

Ejemplo

La serie ($uschange[, ``Consumption'']$) muestra los cambios porcentuales trimestrales en el gasto de consumo de los Estados Unidos. Aunque es una serie trimestral, no parece haber un patrón estacional, por lo que ajustaremos un **modelo ARIMA no estacional**.

```
autoplot(uschange[,"Consumption"]) +
  xlab("Year") + ylab("Quarterly percentage change")
```



Ejemplo

El siguiente código R se utilizó para seleccionar un modelo automáticamente.

```
fit <- auto.arima(uschange[,"Consumption"], seasonal=FALSE)
```

Ejemplo

El siguiente código R se utilizó para seleccionar un modelo automáticamente.

```
fit <- auto.arima(uschange[,"Consumption"], seasonal=FALSE)
```

```
Series: uschange[, "Consumption"]
ARIMA(1,0,3) with non-zero mean

Coefficients:
         ar1      ma1     ma2     ma3    mean
      0.5885  -0.3528  0.0846  0.1739  0.7454
s.e.  0.1541   0.1658  0.0818  0.0843  0.0930

sigma^2 estimated as 0.3499:  log likelihood=-164.81
AIC=341.61   AICc=342.08   BIC=361
```

Ejemplo

```
Series: uschange[, "Consumption"]
ARIMA(1,0,3) with non-zero mean

Coefficients:
         ar1      ma1     ma2     ma3    mean
      0.5885  -0.3528  0.0846  0.1739  0.7454
s.e.  0.1541   0.1658  0.0818  0.0843  0.0930

sigma^2 estimated as 0.3499:  log likelihood=-164.81
AIC=341.61   AICc=342.08   BIC=361
```

¿Cómo escribir el modelo?

Ejemplo

```
Series: uschange[, "Consumption"]
ARIMA(1,0,3) with non-zero mean

Coefficients:
         ar1      ma1     ma2     ma3    mean
      0.5885  -0.3528  0.0846  0.1739  0.7454
s.e.  0.1541   0.1658  0.0818  0.0843  0.0930

sigma^2 estimated as 0.3499:  log likelihood=-164.81
AIC=341.61   AICc=342.08   BIC=361
```

¿Cómo escribir el modelo?

ARIMA(1,0,3)

$$y_t = C + 0{,}589 y_{t-1} - 0{,}353\varepsilon_{t-1} + 0{,}084\varepsilon_{t-2} + 0{,}174\varepsilon_{t-3} + \varepsilon_t \tag{11}$$

dónde $C = 0{,}745 \times (1 - 0{,}589) = 0{,}307$ y ε_t es un ruido blanco con una desviación estándar de $\sqrt{0{,}3499} = 0{,}592$.

Ejemplo

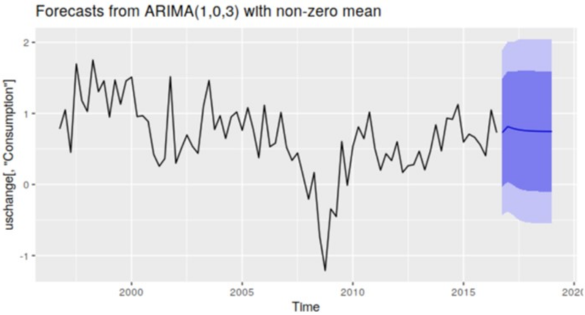
Pronósticos

```
fit %>% forecast(h=10) %>% autoplot(include=80)
```

Ejemplo

Pronósticos

```
fit %>% forecast(h=10) %>% autoplot(include=80)
```



auto.arima() –> Ojo!

La función auto.arima() es útil, pero cualquier cosa automatizada puede ser un poco peligrosa, y vale la pena entender algo sobre el comportamiento de los modelos, incluso cuando confía en un procedimiento automático para elegir el modelo.

Contenido

4 Modelos ARIMA
- Estimación y selección de orden

ACF y PACF

Por lo general, no es posible saber, simplemente a partir de un diagrama de tiempo, qué valores de p y q son apropiados para los datos. Sin embargo, a veces es posible usar el gráfico **ACF**, y el gráfico **PACF** estrechamente relacionado, para determinar los valores apropiados para p y q.

Recuerde:

- Un gráfico ACF muestra las autocorrelaciones que miden la relación entre y_t y y_{t-k} para diferentes valores de k.
  - Si y_t y y_{t-1} están correlacionados, entonces y_{t-1} y y_{t-2} también deben de estar correlacionados.
  - Entonces y_t y y_{t-2} podrían estar correlacionados, simplemente porque ambos están conectados a y_{t-1}, en lugar de cualquier información contenida en y_{t-2} que podría usarse para pronosticar y_t

ACF y PACF

Para superar ese problema, podemos usar las **autocorrelaciones parciales (PACF)**.

Estos miden la relación entre y_t y y_{t-k} después de eliminar los efectos de los rezagos $1, 2, 3, ..., k-1$.

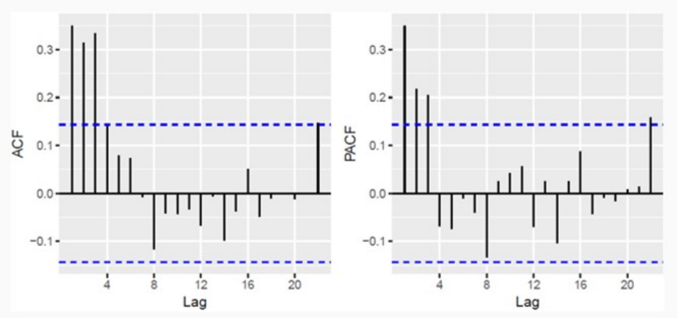
- α_k k-ésima autocorrelación parcial
- $\alpha_k = \rho_1$: La primera autocorrelación parcial es idéntica a la primera autocorrelación, porque no hay nada entre ellas para eliminar.
- Cada autocorrelación parcial se puede estimar como el último coeficiente en un modelo autorregresivo.
- α_k es igual a estimar ϕ_k en un $AR(k)$

ACF y PACF

```
ggAcf(uschange[,"Consumption"])
ggPacf(uschange[,"Consumption"])
ggtsdisplay(uschange[,"Consumption"])
```

ACF y PACF

```
ggAcf(uschange[,"Consumption"])
ggPacf(uschange[,"Consumption"])
ggtsdisplay(uschange[,"Consumption"])
```



ACF y PACF

Los datos pueden seguir un **ARIMA(p,d,0)** si los gráficos ACF y PACF de los datos diferenciados muestran los siguientes patrones:

- El ACF es exponencialmente decadente o sinusoidal.
- Hay un aumento significativo en el rezago p en el PACF, pero ninguno más allá del rezago p.

ACF y PACF

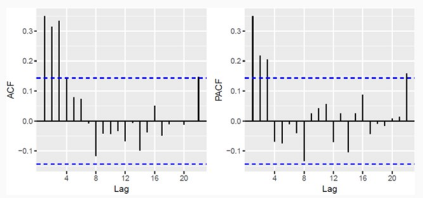
Los datos pueden seguir un **ARIMA(p,d,0)** si los gráficos ACF y PACF de los datos diferenciados muestran los siguientes patrones:

- El ACF es exponencialmente decadente o sinusoidal.
- Hay un aumento significativo en el rezago p en el PACF, pero ninguno más allá del rezago p.

Los datos pueden seguir un **ARIMA(0,d,q)** si los gráficos ACF y PACF de los datos diferenciados muestran los siguientes patrones:

- El PACF es exponencialmente decadente o sinusoidal
- Hay un aumento significativo en el rezago q en el ACF, pero ninguno más allá del retraso q.

Ejemplo



- Hay 3 picos en el ACF, seguidos por un pico casi significativo en el rezago 4.
- En el PACF, hay 3 picos significativos, y luego no hay picos significativos después (aparte de uno justo fuera de los límites en rezago 22).
- Podemos ignorar un pico significativo en cada gráfico si está fuera de los límites, y no en los primeros rezagos.
- La probabilidad de que un pico sea significativo por casualidad es de aproximadamente $1/20$, y estamos trazando 22 picos en cada gráfico.

El patrón en los primeros 3 picos es lo que esperaríamos de un **ARIMA(3,0,0)**, ya que el PACF tiende a disminuir.

Ejemplo

```
fit2 <- Arima(uschange[,"Consumption"], order=c(3,0,0))
```

Ejemplo

```
fit2 <- Arima(uschange[,"Consumption"], order=c(3,0,0))
```

```
Series: uschange[, "Consumption"]
ARIMA(3,0,0) with non-zero mean

Coefficients:
         ar1     ar2     ar3    mean
      0.2274  0.1604  0.2027  0.7449
s.e.  0.0713  0.0723  0.0712  0.1029

sigma^2 estimated as 0.3494:  log likelihood=-165.17
AIC=340.34   AICc=340.67   BIC=356.5
```

Este modelo es en realidad un poco mejor que el modelo identificado por auto.arima(,con un valor AICc de 340.67 en comparación con 342.08. La auto.arima()función no encontró este modelo porque no considera todos los modelos posibles en su búsqueda.

Ejemplo

```
fit2 <- Arima(uschange[,"Consumption"], order=c(3,0,0))
```

```
Series: uschange[, "Consumption"]
ARIMA(3,0,0) with non-zero mean

Coefficients:
         ar1     ar2     ar3    mean
      0.2274  0.1604  0.2027  0.7449
s.e.  0.0713  0.0723  0.0712  0.1029

sigma^2 estimated as 0.3494:  log likelihood=-165.17
AIC=340.34   AICc=340.67   BIC=356.5
```

Este modelo es en realidad un poco mejor que el modelo identificado por auto.arima(,con un valor AICc de 340.67 en comparación con 342.08. La auto.arima()función no encontró este modelo porque no considera todos los modelos posibles en su búsqueda.

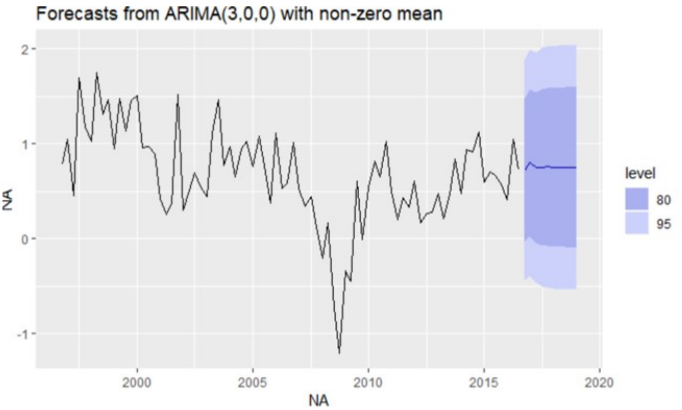
```
fit3 <- auto.arima(uschange[,"Consumption"], seasonal=FALSE,
  stepwise=FALSE, approximation=FALSE)
```

Ejemplo

```
fit2 %>% forecast(h=10) %>% autoplot(include=80)
```

Ejemplo

```
fit2 %>% forecast(h=10) %>% autoplot(include=80)
```



Criterios de información

- AIC
- AICc
- BIC

Buenos modelos son obtenidos minimizando los criterios: AIC, AICc, BIC.

Es importante tener en cuenta que estos criterios de información tienden a no ser buenas guías para seleccionar el orden apropiado de diferenciación (d) de un modelo, sino solo para seleccionar los valores de p y q.

La diferenciación cambia los datos en los que se calcula la probabilidad, lo que hace que los valores de AIC entre modelos con diferentes órdenes de diferenciación no sean comparables.

Procedimiento de modelado

Al ajustar un modelo ARIMA a un conjunto de datos de series temporales (no estacionales), el siguiente procedimiento proporciona un enfoque general.

1. Trace los datos e identifique cualquier observación inusual.
2. Si es necesario, transforme los datos (usando una transformación Box-Cox) para estabilizar la varianza.
3. Si los datos no son estacionarios, tome las primeras diferencias de los datos hasta que los datos sean estacionarios.
4. Examine la ACF / PACF
5. Pruebe los modelos elegidos y use el AICc para buscar un modelo mejor.
6. Verifique los residuos de su modelo elegido trazando el ACF de los residuos y haciendo una prueba de resumen de los residuos. Si no parecen ruido blanco, pruebe con un modelo modificado.
7. Una vez que los residuos se vean como ruido blanco, calcule los pronósticos.

Ejemplo

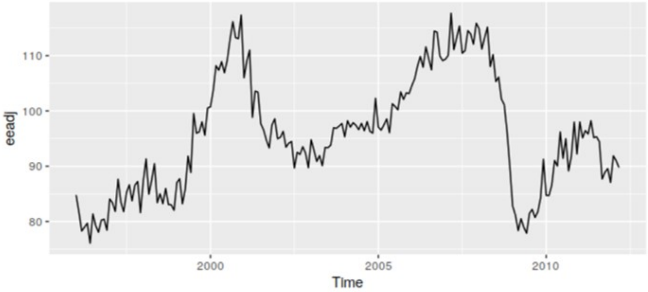
Aplicaremos este procedimiento a los datos de pedidos de equipos eléctricos ajustados estacionalmente.

```
elecequip %>% stl(s.window='periodic') %>% seasadj() -> eeadj
autoplot(eeadj)
```

Ejemplo

Aplicaremos este procedimiento a los datos de pedidos de equipos eléctricos ajustados estacionalmente.

```
elecequip %>% stl(s.window='periodic') %>% seasadj() -> eeadj
autoplot(eeadj)
```



Ejemplo

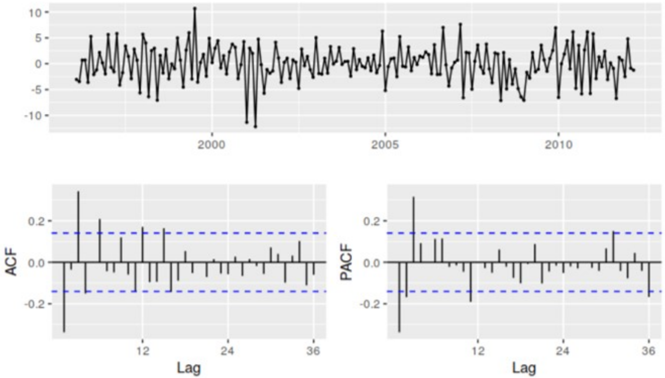
- El diagrama de tiempo muestra algunos cambios repentinos, particularmente la gran caída en 2008/2009. Estos cambios se deben al entorno económico mundial. De lo contrario, no hay nada inusual en el diagrama de tiempo y parece que no hay necesidad de hacer ningún ajuste de datos.
- No hay evidencia de cambios en la varianza, por lo que no haremos una transformación de Box-Cox.
- Los datos son claramente no estacionarios, ya que la serie se desplaza hacia arriba y hacia abajo durante largos períodos.

Ejemplo

```
eeadj %>% diff() %>% ggtsdisplay(main="")
```

Ejemplo

```
eeadj %>% diff() %>% ggtsdisplay(main="")
```



Ejemplo

- El PACF que se muestra sugiere un modelo AR(3). Entonces, un modelo candidato inicial es un ARIMA (3,1,0). No hay otros modelos candidatos obvios.

- Ajustamos un modelo ARIMA(3,1,0) junto con variaciones que incluyen ARIMA(4,1,0), ARIMA(2,1,0), ARIMA(3,1,1), etc. De estos, el ARIMA(3,1,1) tiene un valor AICc ligeramente menor.

Ejemplo

```
fit <- Arima(eeadj, order=c(3,1,1))
```

Ejemplo

```r
fit <- Arima(eeadj, order=c(3,1,1))
```

```
Series: eeadj
ARIMA(3,1,1)

Coefficients:
         ar1     ar2     ar3      ma1
      0.0044  0.0916  0.3698  -0.3921
s.e.  0.2201  0.0984  0.0669   0.2426

sigma^2 estimated as 9.577:  log likelihood=-492.69
AIC=995.38   AICc=995.7   BIC=1011.72
```

Ejemplo

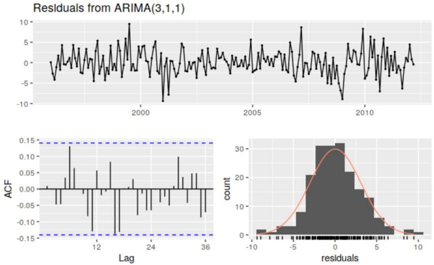
El gráfico ACF de los residuos del modelo ARIMA(3,1,1) muestra que todas las autocorrelaciones están dentro de los límites del umbral, lo que indica que los residuos se comportan como el ruido blanco. Una prueba de portmanteau devuelve un valor p grande, lo que también sugiere que los residuos son ruido blanco.

```
checkresiduals(fit)
```

Ejemplo

El gráfico ACF de los residuos del modelo ARIMA(3,1,1) muestra que todas las autocorrelaciones están dentro de los límites del umbral, lo que indica que los residuos se comportan como el ruido blanco. Una prueba de portmanteau devuelve un valor p grande, lo que también sugiere que los residuos son ruido blanco.

```
checkresiduals(fit)
```



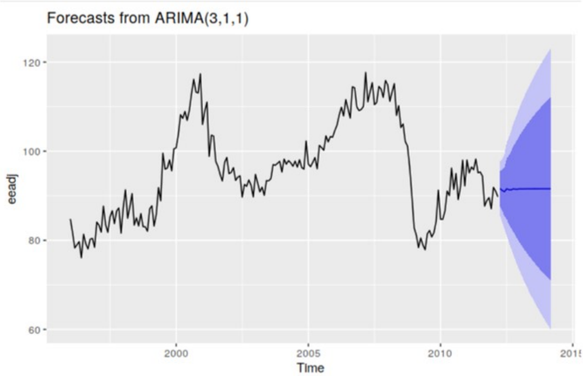

```
#>
#>  Ljung-Box test
#>
#> data:  Residuals from ARIMA(3,1,1)
#> Q* = 24, df = 20, p-value = 0.2
#>
#> Model df: 4.   Total lags used: 24
```

Ejemplo

```
autoplot(forecast(fit))
```

Ejemplo

```
autoplot(forecast(fit))
```

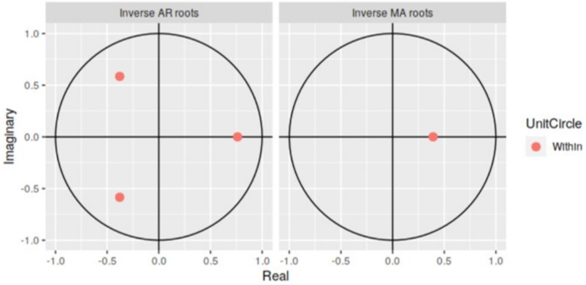


Ejemplo

```
autoplot(fit)
```

Ejemplo

```
autoplot(fit)
```



Los tres puntos rojos en la gráfica de la izquierda corresponden a las raíces de los polinomios $\phi(B)$, mientras que el punto rojo en la gráfica de la derecha corresponde a la raíz de $\theta(B)$. Todos están dentro del círculo unitario, como es de esperar, porque R asegura que el modelo ajustado sea tanto estacionario como invertible.

Preguntas?

Gracias!! ,

Jr.
orlando.joaqui@correounivalle.edu.co