

SERIES TEMPORALES

UNIVERSITAT AUTÒNOMA DE BARCELONA

*Apuntes basados en los documentos i las clases de Alejandra Cabaña Nigro
y las prácticas de Anna López Ratera.*

Eric Jiménez Barril

18 de enero del 2024

Índice

1. INTRODUCCIÓN AL ANÁLISIS DE SERIES TEMPORALES	3
1.1. Descomposición clásica de modelos de series temporales	7
1.2. Estimación de la tendencia	8
1.2.1. Método de mínimos cuadrados para tendencia lineal	8
1.3. Suavidazo o Filtrado	9
1.3.1. Otros filtros	10
1.3.2. Estimación y eliminación tanto de tendencia como estacionalidad con MAs	10
1.4. Operación diferencia	11
2. SERIES TEMPORALES COMO PROCESO ESTOCÁSTICO	12
2.1. Series temporales como proceso estocástico	12
2.2. Introducción a la Estacionariedad	12
2.3. Distribuciones marginales	12
2.4. Propiedades de segundo orden	13
2.5. Función de AutoCovarianza (ACVF) y AutoCorrelación (ACF)	14
2.6. Función de AutoCovarianza (ACVF) y AutoCorrelación (ACF) Empíricas	15
3. PROCESOS LINEALES	16
3.1. Procesos lineales	16
3.2. Moving Average of order q (Media Móvil de orden q (MA(q))	17
3.3. AutoRegressive Model (Modelo AutorRegresivo) (AR(p))	18
3.3.1. Resultados generales de un proceso AR(p)	19
3.3.2. FUNCIÓN DE AUTOCORRELACIÓN PARCIAL PACF	19
3.3.3. PROCESO AR(1)	20
3.3.4. PROCESO AR(2)	21
3.4. PROCESOS LINEALES: ARMA(p,q)	26
3.4.1. Polinomio Característico	26
3.4.2. ARMA(1,1)	26
3.4.3. Representación ARMA(1,1) como MA(∞)	28
3.4.4. CAUSALIDAD E INVERTIBILIDAD	29
3.4.5. Ecuaciones de Yule-Walker de un proceso ARMA	31
3.4.6. Identificación de modelos	33
4. DIAGNOSIS	35
4.1. Datos Color	35
4.1.1. Ajuste del modelo	35
4.1.2. Comportamiento de los residuos	35
4.1.3. Ajuste de otro modelo (AR(2))	37
4.1.4. Forecast para AR(1)	38
4.2. Número de liebres atrapadas cerca del río Hudson entre 1905 y 1931	39
4.2.1. Ajuste del modelo	39
4.2.2. Estudio de los residuos	40
4.2.3. Forecast	40
4.3. Ejemplo Tsay sobre US GNP worth eate 1947 to 1991	41
5. FORECASTING	43
5.1. El "mejor" predictor	43
5.1.1. El mejor predictor es la esperanza condicional	43
5.1.2. Caso Gaussiano	44
5.2. Predictores Lineales	44
5.2.1. Encontremos el Mejor Predictor Lineal (BLP)	45
5.3. Distribuciones Condicionales de variables Gaussianas	46
5.4. Predicción a la práctica	47
5.5. Predicción para MA(1)	48
5.5.1. Predicción a un paso	48
5.5.2. Predicción a dos o más pasos	48
5.5.3. Estimando el ruido para MA(1)	48
5.6. Predicción en un ARMA (1,1)	49
5.6.1. Predicciones	49
5.7. Otra forma de estimar en el ARMA (1,1)	49
5.8. El <i>Innovation Algorithm</i>	50
5.8.1. Algoritmo	50
5.9. Ejercicios	51

6. ESTIMACIÓN DE PARÁMETROS EN MODELOS ARMA	53
6.1. Las ecuaciones de Yule-Walker	53
6.2. MAs y ARMAs	53
6.3. La varianza del ruido	54
6.4. Estimación por mínimos cuadrados	54
6.5. Estimación de máxima verosimilitud	56
6.6. Procedimiento de estimación simple para ARMA(p, q)	57
6.6.1. Métodos basados en la verosimilitud	58
6.7. Propiedades distributivas de los estimadores	59
6.8. Matriz de información	59
7. MODELOS para PROCESOS NO ESTACIONARIOS (ARIMA)	60
7.1. Procesos no estacionarios	60
7.2. Procesos Integrados	60
7.3. Paseo Aleatorio	63
7.4. El papel de las constantes en los procesos no estacionarios	63
7.5. Test de raíz unitaria: <i>Dickey-Fuller</i>	64
7.6. ¿Cómo funciona auto.arima?	68
7.7. Modelos ARIMA estacionales	68
7.7.1. Diferenciación Estacional	69
7.7.2. Diferenciación para la tendencia y la estacionalidad	69
7.8. El Modelo ARIMA Estacional	70
7.8.1. Identificación de modelos Estacionales	70
8. OTROS MODELOS PARA SERIES NO ESTACIONARIAS: MODELOS ARCH Y GARCH	72
8.1. Motivación: Series Temporales Financieras	72
8.1.1. Rentabilidad o beneficio (<i>returnos</i>)	72
8.2. Lectura de una tabla de precios de activos financieros	73
8.3. De la <i>log</i> -Normal a la Normal	75
8.3.1. AutoCovarianza y AutoCorrelación	75
8.4. “ <i>Stylized facts</i> ”	76
8.5. Volatilidad	76
8.6. Modelos para la Volatilidad	76
8.7. Modelo ARCH (Engle, 1982)	77
8.7.1. Efecto ARCH	78
8.7.2. Propiedades de los ARCH	78
8.7.3. Momentos condicionales de un ARCH(1)	79
8.7.4. Momentos incondicionales de un ARCH(1)	79
8.7.5. Estimación de un modelo ARCH(1)	79
8.7.6. Estimador insesgado de la varianza	80
8.7.7. ACF para el ARCH(1)	80
8.8. GARCH (Generalized Autoregressive Conditional Heterocedasticity)	83
8.8.1. GARCH(1,1)	84
8.9. ARIMA(p, d, q)/GARCH(P, Q)	84
9. BONDAD DE AJUSTE Y SELECCIÓN DE MODELOS	86
9.1. Ejemplo de modelo predicción	86
9.2. Glosario	91

1. INTRODUCCIÓN AL ANÁLISIS DE SERIES TEMPORALES

Definición: Serie Temporal

Una serie temporal

$$X_{t_1}, X_{t_2}, \dots$$

es una secuencia de puntos de datos provenientes de un fenómeno que evoluciona a lo largo del tiempo (del espacio o otra magnitud ordenada) medido en intervalos (a menudo uniformes).

Tiene aplicaciones en diferentes campos como

1. Economía: Datos de empleo mensual, admisiones en un hospital, ...
2. Finanzas: Cambios diarios en los intereses, precios de acciones y beneficios.
3. Medicina: actividad de las ondas cerebrales, concentración de oxígeno (o medicamentos) en sangre...

Observación 1.1. Técnicamente hablando:

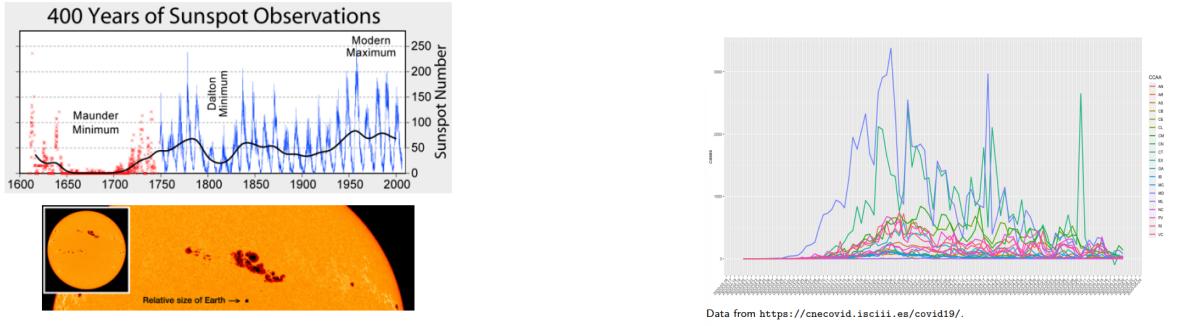
- Por tiempo queremos decir que las observaciones se registran secuencialmente, pero puede ser espacial (1^a planta seguida, 2^a planta seguida...) o en profundidad (1 cm hacia abajo, 2 cm hacia abajo,...) o cualquier otra variable independiente ordenada .
- Puede haber observaciones **univariadas** (se centran en una sola variable) o **multivariadas** (se centran en el análisis de dos o más variables simultáneamente), **numéricas** (números reales o enteros) o **categóricas**.
- Las observaciones se pueden tomar en un **conjunto discreto** t_1, t_2, t_3, \dots o **continuamente durante algún intervalo**.
- Los tiempos de registro pueden ser **regulares** (cada hora, mensual,... . .) o **irregulares** debido a valores faltantes o la naturaleza de los datos.
- Trataremos en la mayor parte de las bibliografías tratando con las series temporales clásicas donde:

$$X_{t_i} = X(t_i)$$

son *variables aleatorias continuas normalmente distribuidas*, pero también pueden ser *variables aleatorias discretas o categóricas*.

- En este curso nos centraremos en *series de tiempo clásicas* estudiadas a través de la metodología de *Box & Jenkins*
- Los modelos de *cadenas de Markov* se utilizan frecuentemente para describir series de tiempo categóricas (las variables aleatorias toman valores en un conjunto contable de estados posibles) esto se verá en el curso de procesos estocásticos.

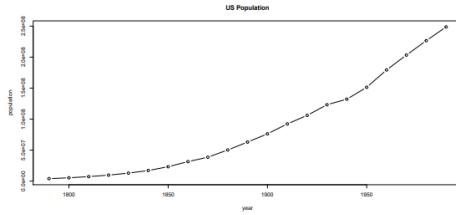
Ejemplos 1.2. Veamos tres ejemplos de series temporales:



(a) Números de manchas solares (*Número de Wolf*)

(b) Casos de PCR+ en España durante la primera ola de COVID-19

Figura 1: Ejemplos de series temporales.



```
usa=read.table("uspopulation.txt")
t=seq(1790,1990,10); usa=zoo(usa,order.by=t)
plot(t,usa,type="b", main="US Population", xlab="year",
ylab="population")
```

Figura 2: Trazar una serie temporal es un paso importante en su análisis.

En la Figura 2 notamos que

- Hay una clara tendencia al alza.
- Hay un cambio de forma alrededor del *baby boom*.
- El comportamiento NO es lineal (tal vez es cuadrático).

Observación: Plot de una Serie Temporal

En general, un *plot* puede revelarnos

1. **Tendencia:** movimientos alcistas o bajistas que pueden extrapolarse al futuro.
2. **Cambios de nivel:** Variaciones abruptas o discontinuidades en el valor medio de la serie a lo largo del tiempo.
3. **Estacionalidad:** Comportamientos periódicos con un periodo reconocido (por ejemplo 12 meses o 4 semanas).
4. **Patrones cíclicos:** Los datos muestran aumentos o caídas que no son de periodo fijo (por ejemplo los ciclos económicos).
5. **Heterocedasticidad:** varianza cambiante (la varianza no es constante en todas las observaciones realizadas), se refiere a la falta de homogeneidad en la varianza de los errores. En cambio, decimos que hay **homocedasticidad** si la varianza es constante (normalmente hablando de los errores de un modelo).
6. **Dependencia:** positiva (las observaciones sucesivas son similares) o negativa (las observaciones sucesivas son diferentes, los valores tienden a mostrar una dependencia inversa o negativa entre sí).
7. Datos faltantes (*missing data*, valores atípicos *outliers*, rupturas estructurales etc.

Ejemplos 1.3. Veamos el ejemplo clásico de los datos de la aerolínea de *Box & Jenkins*: El número total de pasajeros internacionales de una aerolinea entre 1949 y 1969.

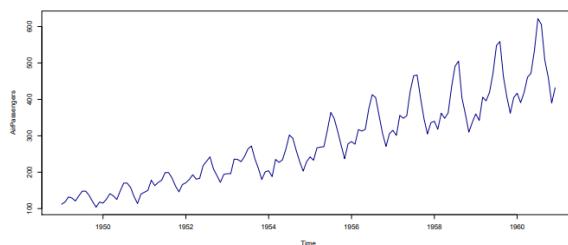


Figura 3: Número mensual de pasajeros de una aerolínea entre 1949-1969

1. Tendencia fuertemente alcista.
2. Estacionalidad en un intervalo de 12 meses.
3. Variabilidad creciente (heterocedasticidad).

Ejemplos 1.4. Venta mensuales en litro de vino tinto en Australia desde enero de 1980 hasta octubre de 1991.

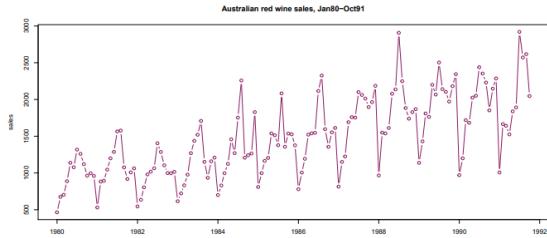
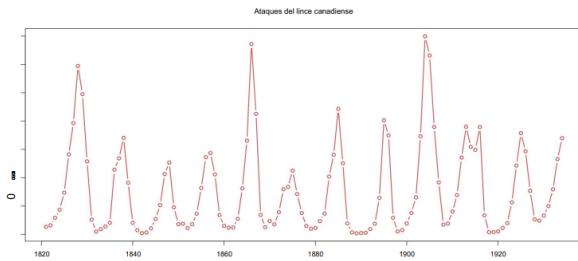


Figura 4: El gráfico muestra una tendencia alcista y un patrón de estacionalidad anual con un pico en julio y el nivel más bajo en enero.

Ejemplos 1.5. Número actual de lince canadienses atrapados cerca del río *McKenzie*



1. No muestra una tendencia aparente.
2. No muestra estacionalidad de periodo conocido.
3. Muestra un ciclo a largo plazo.

Ejemplos 1.6. Registro de 197 lecturas de concentración de procesos químicos tomados cada 2h. Es una serie difícil de estudiar. Muestra una ligera **dependencia negativa**.

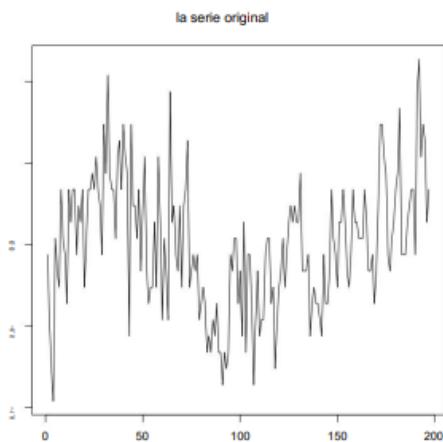


Figura 5: *Box-Jenkins-Reinsel Serie A*

1.1. Descomposición clásica de modelos de series temporales.

En series temporales:

1. Las *variables aleatorias* suelen ser no independientes.
2. Las **medias** se ven afectadas por la tendencia y la estacionalidad.
3. La **varianza puede variar significativamente**.
4. Las variables aleatorias suelen no estar idénticamente distribuidas.

Lo que trataremos de hacer es estimar por separado la tendencia, estacionalidad y eliminarlas de la serie junto a la heterocedasticidad para posteriormente hacer un modelo para los residuos (modelo de la parte no determinista).

Definición: Descomposición Clásica de una Serie Temporal

Método sencillo para describir una serie con descomposición clásica

La idea es descomponer la serie en 3 elementos:

1. **Tendencia** m_t : Movimientos a largo plazo de la media.
2. **Movimientos estacionales** s_t : Fluctuaciones cíclicas relacionadas con el calendario o ciclos económicos.
3. **Ruido** Y_t : Fluctuaciones sistemáticas o aleatorias.

La idea es crear un modelo para cada uno de estos modelos por separado y luego combinarlos, ya sea:

- **Forma aditiva**, si la variación estacional parece constante (no cambia cuando el valor de la serie temporal aumenta):

$$X_t = m_t + s_t + Y_t$$

- **Forma multiplicativa**, si a medida que la serie aumenta en magnitud, la variación estacional aumenta/disminuye:

$$X_t = m_t \times s_t \times Y_t$$

Ejemplos 1.7. Ejemplos de descomposición de modelos multiplicativos, notemos que cuando la serie temporal aumenta en magnitud (sobretodo en la segunda), la variación aumenta, por lo que un modelo multiplicativo se ajustará mejor

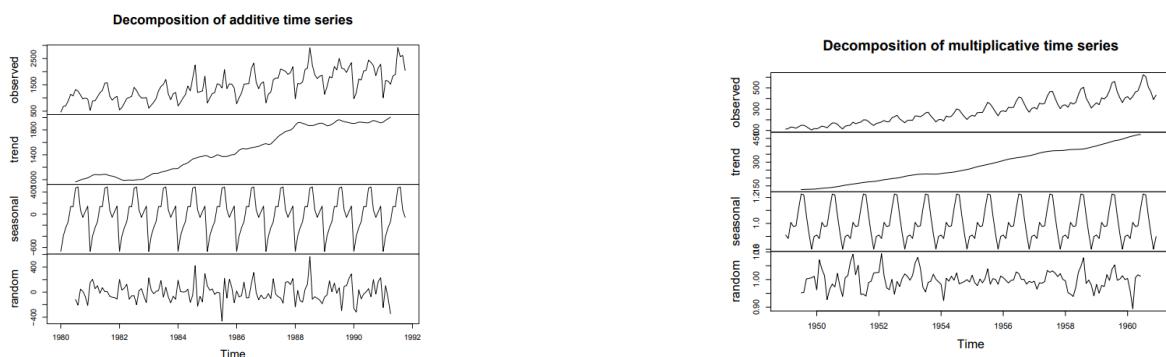


Figura 6: Ejemplos de descomposición clásica de series temporales.

1.2. Estimación de la tendencia

1.2.1. Método de mínimos cuadrados para tendencia lineal

Supongamos que la parte estacional está ausente. Una estructura de tendencia temporal lineal simple

$$m_t = a + bt$$

o cualquier otra tendencia polinómica se puede estimar mediante mínimos cuadrados. Esto tiene dos principales inconvenientes:

1. No puede hacer frente a cambios de tendencia a largo plazo.
2. Las formas lineales pueden ser demasiado restrictivas.

Ejemplos 1.8. Veamos el ejemplo de la población de USA entre 1970 y 1990. Para estimar:

$$m_t = a_0 + a_1 t + a_2 t^2$$

para los datos de la población, reetiquetamos los ejes para que $t = 1$ corresponda a 1790 y $t = 21$ a 1990 y ajustamos un modelo lineal.

```
> t=1:21
> usa.lm=lm(usa~t+I(t^2))
> summary(usa.lm)

Call:
lm(formula = usa ~ t + I(t^2))

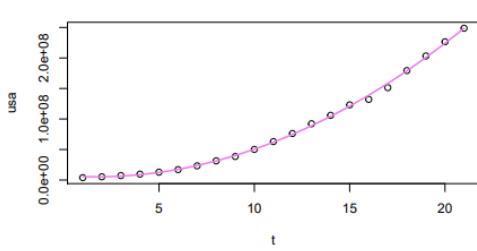
Residuals:
    Min      1Q  Median      3Q     Max 
-6947521 -358167  436285 1481410 3391761 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 6957920   1998526   3.482  0.00266 **  
t           -2159870    418437  -5.162 6.55e-05 ***  
I(t^2)       650634    18472  35.223 < 2e-16 ***  
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

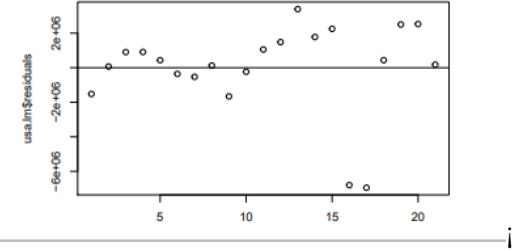
Residual standard error: 2767000 on 18 degrees of freedom
Multiple R-squared:  0.9989,   Adjusted R-squared:  0.9988 
F-statistic:  8050 on 2 and 18 DF,  p-value: < 2.2e-16
```

Figura 7: Ajuste de un modelo lineal

Obtenemos de dicho modelo las siguientes estimaciones:



(a) Estimación de la tendencia



(b) Estimación de los residuos del modelo lineal

La estimación de la tendencia:

$$\hat{m}_t = 69579 - 2159870t + 650634t^2$$

es un predictor natural de valores futuros de X_t , sin embargo los residuos Y_t están altamente correlacionados como podemos ver a la derecha por lo que podríamos usar una mejor estimación para X_t .

1.3. Suavidazo o Filtrado

Definición: Filtrado lineal con pesos

Dada una serie temporal X_t y un conjunto de números reales $a_{-r}, a_{-r+1}, \dots, a_s$ tal que $r, s > 0$ y $r + s + 1 \leq n$, la transofmración lineal

$$X_t^* = \sum_{u=-r}^s a_u X_{t-u} \quad t = s + 1, \dots, n - r$$

se conoce como **filtro lineal con pesos** $a_{-r}, a_{-r+1}, \dots, a_s$.

Observación: Filtrados lineales con pesos

- Observe que hay menos datos de salida que de entrada.
- Usualmente $r=s$ (**centrado**) y $a_{-k} = a_k$ (**simétrico**).
- **Medias Móviles:** $\sum_k a_k = 1$
- **Medias Móviles Simétricas (impar, MA(k) con $k = 2s + 1$):**

$$a_u = \frac{1}{2s + 1}, \quad u = -s, \dots, s$$

- **Medias Móviles Simétricas (par con $k = 2s$):**

$$a_s = a_{-s} = \frac{1}{4s}; \quad a_u = \frac{1}{2s}, \quad u = -s + 1, \dots, s - 1$$

Ejemplos 1.9. Ejemplo de precios de *Yahoo Finance*

The chart was obtained from Yahoo! Finance, it represents the closing price of Netflix Inc. shares, together with a smoothing wit a Moving average of order 5 and a Moving average of order 15.



Observación 1.10. Por qué funcionan las medias móviles

Consideramos un modelo, sin necesariamente estacionalidad, $X_t = m_t + Y_t$ con $\mathbb{E}(Y_t) = 0$.

- Si $m_t = c_0 + c_1 t$ entonces para cualquier *media móvil simétrica*,

$$a_j = \frac{1}{2q + 1}, \quad -q \leq j \leq q, \quad \sum_{j=-q}^q a_j m_{t-j} = m_t$$

- Para q grande, la varianza de MA es pequeña y por tanto

$$\hat{m}_t = \sum_{j=-q}^q a_j X_{t-j}$$

es un buen estimador de m_t

Por tanto la secuencia de ruido la podemos estimar como

$$\hat{Y}_t = X_t - \hat{m}_t$$

1.3.1. Otros filtros

Existen también otros filtros para eliminar tendencias polinómicas o estacionales como: Medias móviles ponderadas, Suavizadores polinomiales locales, Suavizadores de pepitas, Suavizadores exponenciales (e.g *Holt-Winters*).

Todos ellos herramientas descriptivas.

1.3.2. Estimación y eliminación tanto de tendencia como estacionalidad con MAs

1. Identifica el periodo d de la componente estacional.
2. Estima la tendencia \hat{m}_t con una *media móvil simétrica* par o impar dependiendo de d .
3. Estima la componente estacional. Para cada $k = 1, \dots, d$ calcula el promedio w_k de las desviaciones:

$$\{(x_{k+jd} - \hat{m}_{k+jd}), q < k + jd \leq n - q\}$$

Dado que las desviaciones promedio no necesariamente suman 0, estima la componente estacional como:

$$\hat{s}_k = w_k - \frac{1}{d} \sum_{i=1}^d w_i, \quad k = 1, \dots, d, \quad \hat{s}_k = \hat{s}_{k-d} \quad \text{per } k > d$$

4. Para los datos no estacionales $d_t = x_t - \hat{s}_t$ estima de nuevo la tendencia con mínimos cuadrados si quieras un modelo paramétrico.

1.4. Operacion diferencia

Otro método para estudiar la tendencia es tomando sucesivas diferencias.

Definición: Back-Shift Operator

$$B(X_t) = X_{t-1}$$

Definición: Operador diferencia de lag-1 ∇

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

Observación: Propiedades

Se definen las potencias de B y ∇ de manera obvia:

$$\begin{cases} B^j X_t = X_{t-j} \\ \nabla^n X_t = \nabla(\nabla^{n-1} X_t) = (1 - B)^n X_t \\ \nabla^2 X_t = (1 - B)^2 X_t = X_t - 2X_{t-1} + X_{t-2} \end{cases}$$

Observación 1.11. Si aplicamos el operador ∇ a una tendencia lineal $m_t = c_0 + c_1 t$ obtenemos la constante c_1 .

Definición: Operador diferencia de lag-d ∇_d

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t$$

Observación 1.12. Si aplicamos ∇_d a $X_t = m_t + s_t + Y_t$ obtenemos

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$$

(en caso de que la componente estacional sea de periodo d), lo que da una descomposición de $\nabla_d X_t$ en una tendencia ($m_t - m_{t-d}$) y un ruido ($Y_t - Y_{t-d}$).

Ahora el ruido puede eliminarse aplicando alguno de los métodos dichos, como por ejemplo aplicando una alguna potencia de ∇ .

2. SERIES TEMPORALES COMO PROCESO ESTOCÁSTICO

2.1. Series temporales como proceso estocástico

Adoptamos el enfoque de que los datos son una realización de variables aleatorias. Sin embargo, está claro que estas observaciones no son *IID* y más bien interpretamos la serie temporal como una realización de un **proceso estocástico** $X(\omega, t)$ definido en $\omega \times [a, b]$.

- Para una t fija observamos una variable aleatoria.
- Para un ω_f fijo observamos un camino de muestra (una trayectoria) o más bien, las evaluaciones de $X(\omega_f, t_i)$ para una colección de puntos de tiempo t_i .

Este tipo de procesos X_1, X_2, \dots están determinados por la especificación de todas las posibles distribuciones conjuntas de los vectores aleatorios X_1, X_2, \dots, X_n para $n = 1, 2, \dots$, o equivalentemente las probabilidades

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \quad -\infty < x_1, \dots, x_n < \infty, \quad t_n = 1, 2, \dots$$

2.2. Introducción a la Estacionariedad

Estacionariedad

Los fenómenos dinámicos que observamos en una serie temporal pueden mostrar:

- Valores estables en el tiempo alrededor de un nivel constante, sin mostrar una tendencia creciente o decreciente a largo plazo. Esos procesos se llaman **procesos estacionarios** (Precipitaciones anuales, temperaturas medias anuales, proporción de nacimientos de niñas...).
- Los procesos no estacionarios cuyas características probabilísticas varían en el tiempo: pueden mostrar tendencia, estacionalidad y otros efectos evolutivos en el tiempo (ingresos de un país, número de PCR+ de una región...).

Observación 2.1. En la práctica, la clasificación de una serie como estacionaria o no, depende del período de observación, ya que la serie puede ser estable en un período corto y no estacionaria en uno más largo.

2.3. Distribuciones marginales

Mean Function

La *mean function* del proceso se refiere a una función del tiempo que representa el valor esperado de las distribuciones (marginales) de X_t para cada t :

$$\mathbb{E}(X_t) = \mu_t$$

Cuando todas las variables tienen la misma media, la función media es una constante. Las realizaciones del proceso no muestran tendencia y decimos que el proceso es *estable en la media*.

Si, por el contrario, las medias cambian con el tiempo, las observaciones en diferentes momentos revelarán ese cambio.

En muchas ocasiones tenemos **una sola realización** del proceso estocástico y de ahí tenemos que deducir si la función media del proceso es, o no, constante en el tiempo.

2.4. Propiedades de segundo orden

Si hay alguna esperanza de predecir valores futuros de una serie temporal, debemos suponer que hay “algo” que no varía en el tiempo. Dado que en la mayoría de los casos de series temporales las distribuciones conjuntas son muy complicadas (a menos que los datos provengan de un mecanismo muy simple, como observaciones iid), un procedimiento habitual es especificar sólo los **momentos de primer y segundo orden de las distribuciones conjuntas**.

Propiedades de primer y segundo orden

$$\begin{cases} \mu_X(t) := \mathbb{E}(X_t) \\ \gamma_X(t+h, t) := \mathbb{E}((X_{t+h} - \mu_{t+h})(X_t - \mu_t)) \quad t = 1, 2, \dots, \quad h = 0, 1, \dots \end{cases}$$

centrándose en propiedades que dependen únicamente de estas. Estas propiedades se conocen como **propiedades de segundo orden**.

Observación 2.2. En el caso particular donde las observaciones son gaussianas, las propiedades de primer y segundo orden determinan completamente las distribuciones conjuntas.

En general, perderemos información al observar series de tiempo “a través de gafas de segundo orden”. Sin embargo, como estudiaremos más adelante, los predictores lineales de mínimos cuadrados dependen sólo de propiedades de segundo orden, por lo tanto, la caracterización del tiempo de segundo orden. La serie es interesante y menos limitada de lo que parece a primera vista.

Procesos estrictamente estacionarios y débilmente estacionarios

- Un proceso estocástico $X_t, t \in T$ es **estRICTAMENTE ESTACIONARIO** si sus distribuciones conjuntas no dependen de t , es decir:

$$\mathcal{L}(X_1, \dots, X_n) = \mathcal{L}(X_{1+h}, \dots, X_{n+h}) \quad \forall n \in T, \quad \forall h \in T.$$

- Un proceso $X_t, t \in T$ es **dÉBILMENTE ESTACIONARIO** si

1. La media $\mu_X(t) := \mathbb{E}(X_t) = \mu_X < \infty$ no depende de t .
2. El momento de segundo orden

$$\gamma_X(t+h, t) := \mathbb{E}((X_{t+h} - \mu_X(t+h))(X_t - \mu_X(t)))$$

no depende de t , pero solo en el *lag* h , es decir,

$$\gamma_X(t+h, t) = \gamma_X(h, 0) := \gamma_X(h) < \infty$$

Observación 2.3. ■ Si $X_t, t \in T$ es **estRICTAMENTE ESTACIONARIO** y $\mathbb{E}(X_t^2) < \infty$, entonces es **dÉBILMENTE ESTACIONARIO**.

- A partir de ahora, para nosotros un proceso estacionario será débilmente estacionario.

2.5. Función de AutoCovarianza (ACVF) y AutoCorrelación (ACF)

Consideramos el proceso **estacionario** $\{X_t, t \in T\}$.

Definición: Función de AutoCovariancia (ACVF)

Definimos la función de AutoCovariancia en el *lag* h como:

$$\gamma_X(h) := \gamma_X(h, 0) = \gamma_X(t + h, t)$$

Propiedades:

1. $\gamma(0) \geq 0$
2. $\gamma(0) \geq \gamma(h) \quad \forall h$
3. $\gamma(h) = \gamma(-h) \quad \forall h \quad (\text{es par})$

Definición: Función de AutoCorrelación (ACF)

Definimos la función de AutoCorrelación en el *lag* h como:

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = \text{Corr}(X_{t+h}, X_t)$$

Propiedades:

1. $\rho(0) = 1$
2. $\rho(h) \leq 1 \quad \forall h$
3. $\rho(h) = \rho(-h) \quad \forall h \quad (\text{es par})$

Ejemplos 2.4. Veamos 3 ejemplos:

1. **IID Noise (Ruido IID):** X_t son variables aleatorias *iid* con $\mathbb{E}(X_t) = 0$ y $\text{Var}(X_t) = \sigma^2$.

En este caso

$$\gamma_X(h) = \mathbb{E}(X_t X_{t+h}) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}$$

2. **Ruido Blanco:** X_t son variables aleatorias *incorelacionadas* con $\mathbb{E}(X_t) = 0$ y $\text{Var}(X_t) = \sigma^2$. En este caso de nuevo

$$\gamma_X(h) = \mathbb{E}(X_t X_{t+h}) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}$$

3. **Paseo Aleatorio:** Dadas X_t variables aleatorias con $\mathbb{E}(X_t) = 0$ y $\text{Var}(X_t) = \sigma^2$ y se define

$$\begin{cases} S_0 = 0 \\ S_1 = X_1 \\ S_k = S_{k-1} + X_k = \sum_{j=1}^k X_j \end{cases}$$

de donde $\forall k \in T$

$$\begin{cases} \mathbb{E}(S_k) = 0 \\ \text{Var}(S_k) = k\sigma^2 \\ \mathbb{E}(S_k S_{k+h}) = k\sigma^2 \end{cases}$$

por lo que el **Paseo Aleatorio no es estacionario**

4. (ejercicio ver libreta): Si $Z_n \sim N(0, 1)$, iid, $n = 1, 2, \dots$ y

$$X_t := \begin{cases} Z_t & t \text{ par} \\ \frac{Z_{t-1}^2 - 1}{\sqrt{2}} & t \text{ impar} \end{cases}$$

entonces X_t es **Ruido Blanco** ($WN(0, 1)$) pero las variables NO son iid.

2.6. Función de AutoCovarianza (ACVF) y AutoCorrelación (ACF) Empíricas

En los problemas reales no tendremos modelos, si no que tendremos datos. Las siguientes funciones empíricas nos ayudaran a elegir modelo

Definición 2.5. Media muestral

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

Definición 2.6. Función de AutoCovariancia muestral (ACVF) Definimos la función de AutoCovariancia en el *lag* h como:

$$\hat{\gamma}(h) := \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x})$$

Definición 2.7. Función de AutoCorrelación (ACF) muestral Definimos la función de AutoCorrelación en el *lag* h como:

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

La ACVF y ACF pueden calcularse para cualquier serie, y si no es *estacionaria*, se notarán algunas características, por ejemplo, para series estacionales, los patrones del ACF se repetirán periódicamente.

3. PROCESOS LINEALES

3.1. Procesos lineales

Definición: Proceso Lineal

Una serie $X_t, t \in T$ es un **proceso lineal** si admite una representación de la forma

$$X_t = \mu + \sum_{r=-\infty}^{\infty} \psi_r \epsilon_{t-r}$$

para ϵ_t incorrelacionados (o *i.i.d.*), con $\mathbb{E}(\epsilon_t) = 0$, $\text{Var}(\epsilon_t) = \sigma^2$ y μ es la media el proceso.

Definición: Proceso Estacionario

Decimos que un proceso lineal es **estacionario** si satisface que $\sum_r |\psi_r| < \infty$, para $\mathbb{E}(|X_t|) < \infty$ y $\sum_r \psi_r^2 < \infty$ para que $\text{Var}(X_t) < \infty$.

Asumiremos que los procesos lineales que estudiemos son estacionarios.

Definición: Proceso Causal

- Un proceso $\{X_t\}_{t \in T}$ es **causal** si está influenciado por valores y condiciones pasadas, i.e no depende del futuro (en el tiempo t no depende de valores futuros de t).
- Un **proceso lineal** que satisface $\psi_r = 0$ para $r < 0$, diremos que es un **proceso causal**, es decir, el proceso en el tiempo t no depende de los valores futuros de t , es decir, de los valores no observados de ϵ_t

Observación 3.1. La estructura de la ACFV de un proceso lineal es fácil de encontrar.

Teorema: *Descomposición de Wold*

Si un proceso estocástico $\{X_t\}_{t \in T}$ es estacionario y no contiene componente determinista ^a, entonces admite una representación del tipo

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

donde

1. $\psi_0 = 1$ y $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

2. $Z_t \sim WN(0, \sigma^2)$

y las secuencias Z_t , ψ_t son únicas y pueden escribirse explícitamente.

^aUn proceso es determinista si puede predecirse con precisión arbitraria, esto es, que podemos determinarlo con un *MSE* arbitrariamente pequeño, por ejemplo la componente estacional o la tendencia.

Observación 3.2. El resultado anterior garantiza que:

- Cualquier proceso estacionario es un proceso lineal, o la suma de un proceso lineal y un componente determinista.
- La clase de procesos lineales proporciona un marco general para estudiar procesos estacionales.

3.2. Moving Average of order q (Media Móvil de orden q (MA(q)))

Definición: MA(q) Media Móvil de orden q .

Si sólo un número finito q de coeficientes ψ_j son diferentes de cero, para $j \geq 0$, llamamos al proceso resultante **media móvil de orden q** (moving average de orden q):

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

Propiedades

- Es **q -dependiente**: X_t y X_s son independientes si $|t - s| > q$ (si fueran iid serían 0-dependientes).
- Es **q -correlacionada**, es decir, $\gamma(h) = 0$ si $|h| > q$.

- **Polinomio Característico** El proceso MA(q) puede escribirse en términos del polinomio característico

$$\Theta(z) = 1 + \sum_{j=1}^q \theta_j z^j$$

- **Definición con el Backshift operator B**

$$X_t = \Theta(B)\epsilon_t$$

- **ACF**

1. Proceso MA(1)

$$X_t = \epsilon_t + \theta \epsilon_{t-1}$$

Podemos calcular

$$\gamma_0 = \text{Var}(X_t) = \mathbb{E}(\epsilon_t^2) + \mathbb{E}(2\epsilon_t \epsilon_{t-1} + \mathbb{E}(\theta^2 \epsilon_{t-1}^2)) = \sigma^2 + \theta^2 \sigma^2 = (1 + \theta^2) \sigma^2$$

i per $h = 1$ tenim:

$$\gamma_1 = \mathbb{E}[(\epsilon_t + \theta \epsilon_{t-1})(\epsilon_{t-1} + \theta \epsilon_{t-2})] = \theta \sigma^2$$

per tant obtenim:

$$ACVF := \gamma_h = \begin{cases} (1 + \theta^2) \sigma^2 & h = 0 \\ \theta \sigma^2 & h = 1 \\ 0 & h \geq 2 \end{cases}$$

Por tanto tenemos que

$$ACF := \rho_h = \begin{cases} 1 & h = 0 \\ \frac{\theta}{1 + \theta^2} & h = 1 \\ 0 & h \geq 2 \end{cases}$$

Observación 3.3. Notamos que en un proceso MA(1) con $\bar{\theta} = \frac{1}{\theta}$ obtenemos la misma ACF, por tanto tenemos un problema de **identificabilidad**. Para solucionar esto y poder identificar los proceso MA, pedimos que las raíces de su polinomio característico estén *fueras del círculo unidad*.

2. Proceso MA(q)

$$ACF := \rho_h = \begin{cases} 1 & h = 0 \\ \frac{\theta_j \theta_{j+h}}{1 + \theta_1^2 + \dots + \theta_q^2} & 0 \leq h \leq q \\ 0 & h > q \end{cases}$$

3.3. AutoRegresivo Model (Modelo AutorRegresivo) (AR(p))

Definición: Proceso AutoRegresivo AR(p)

Un proceso AutoRegresivo (AR) es un modelo estadístico que describe la dependencia temporal de la serie en tiempo t en función de sus p valores pasados. Es decir, un modelo AR(p) es un proceso AutoRegresivo de orden “p”. Se puede describir analíticamente como

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t$$

con $\epsilon_t \sim WN(0, \sigma^2)$ siempre que $\mathbb{E}(X_t) = 0$.

- Si $\mathbb{E}(X_t) = \mu$ entonces

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \epsilon_t$$

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t$$

$$\alpha = \mu(1 - \phi_1 - \dots - \phi_p).$$

- Suponiendo que $\mathbb{E}(\epsilon_t) = 0$, $\text{Var}(\epsilon_t) = \sigma^2$ y $\mathbb{E}(X_t) = 0$

El proces ocentrado puede escribirse como

$$\Phi(B)X_t = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)X_t = \epsilon_t$$

- Polinomi Característico

$$\Phi(z) = 1 - \sum_{j=1}^p \phi_j z^j$$

- Definición con el Backshift operator B

$$\Phi(B)X_t = \epsilon_t \quad X_t = \Phi^{-1}(B)\epsilon_t$$

donde el proceso AR(1) es estacionario y causal si las raíces del poliomio característico están fuera del círculo unidad.

- Estacionariedad

Condición Estacionariedad Modelos AR

Un proceso AR(p), $\Phi(B)X_t = \epsilon_t$ es estacionario si las raíces de la ecuacion $\Phi(B) = 0$ estan fuera del círculo unidad. **Una condicion necesaria (NO suficiente) es**

$$\phi_1 + \dots + \phi_p < 1 \quad \&\& \quad |\phi_p| < 1$$

Observación 3.4. IMPORTANTE: Si $|\phi| > 1$, no hay solución estacionaria al modelo AR(1).

- Ecuaciones de Yule-Walker y ACF La varianza de X_t satisface:

$$\gamma_0 = \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + \sigma_\epsilon^2$$

Las **Ecuaciones de Yule-Walker** son ¹

$$\rho_h = \phi_1 \rho_{h-1} + \dots + \phi_p \rho_{h-p}$$

¹Notemos que $\rho_h = \frac{\gamma_h}{\gamma_0}$, pero para $h > 0$ tenemos que $\gamma_h = \mathbb{E}(X_{t+h}, X_t) = \phi_1 \gamma_{h-1} + \dots + \phi_p \gamma_{h-p} + \mathbb{E}(X_t \epsilon_{t+h})$ per tant l'últim terme val 0 doncs són independents, d'aquí les equacions de Yule-Walker

y se pueden resolver recursivamente con

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & \ddots & \rho_1 & \rho_{p-2} \\ \vdots & \rho_1 & \ddots & \vdots \\ \rho_{p-1} & \cdots & \ddots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

Se pueden escribir en términos de covarianzas y no de correlaciones, lo que lleva a

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} 1 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \ddots & \gamma_1 & \gamma_{p-2} \\ \vdots & \gamma_1 & \ddots & \vdots \\ \gamma_{p-1} & \cdots & \ddots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

Estimación de momentos de los coeficientes y la varianza

Es decir

$$\gamma = \mathbf{\Gamma}\phi$$

por lo que un estimador de momentos de ϕ es

$$\hat{\phi} = \hat{\mathbf{\Gamma}}^{-1}\hat{\gamma}$$

y

$$\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{\phi}^T\hat{\gamma}$$

Al ser un estimador de momentos, funciona independientemente de la distribución de la secuencia de ruido.

3.3.1. Resultados generales de un proceso AR(p)

1. Si $1 - \hat{\phi}_1 z - \dots - \hat{\phi}_p z^p \neq 0$ per $|z| < 1$, entonces el modelo ajustado $\hat{\Phi}(B)X_t = \epsilon_t$ es casual.
2. Las covariancias del modelo ajustado son $\hat{\gamma}_k$.
3. Para medida n, $\hat{\phi}$ es aproximadamente $N(\phi, n^{-1}\mathbf{\Gamma}^{-1}\sigma^2)$ ²

3.3.2. FUNCIÓN DE AUTOCORRELACIÓN PARCIAL PACF

Un AR(p) tiene efectos directos en las observaciones separadas por $1, 2, \dots, p$ lags y los efectos en las observaciones separadas por lags mayores que p son

$$\begin{aligned} X_t &= \phi_{1,1}X_{t-1} + \epsilon_{1,t} \\ X_t &= \phi_{1,2}X_{t-1} + \phi_{2,2}X_{t-2} + \epsilon_{2,t} \\ X_t &= \phi_{1,3}X_{t-1} + \phi_{2,3}X_{t-2} + \phi_{3,3}X_{t-3} + \epsilon_{3,t} \\ &\vdots \\ X_t &= \phi_{1,k}X_{t-1} + \phi_{2,k}X_{t-2} + \phi_{3,k}X_{t-3} + \dots + \phi_{k,k}X_{t-k} + \epsilon_{k,t} \end{aligned}$$

- Estos modelos pueden estimarse con OLS y pueden compararse a través del F-test.

²Ojo, no son lo mismo que los del punto 1, mientras que el apartado .1 se refiere al polinomio con coeficientes estimados ($\hat{\Phi}$), este se refiere al vector de los coeficientes $\hat{\phi}$

- Los estimadores $\hat{\phi}_{j,j}$ son los **coeficientes de autocorrelación parcial** y representan la contribución de X_{t-j} al anterior modelo AR(j-1).
- Por construcción, si $j > p$ entonces el coeficiente PACF es 0. Por lo que el orden de p puede identificarse a través del último *lag* para el cual la PACF es *estadísticamente* diferente de 0.

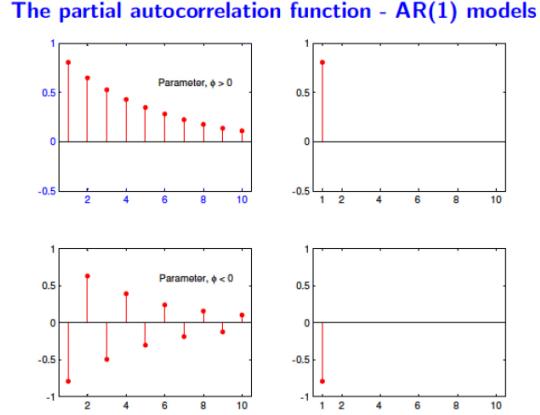


Figura 9: ACF decrece rápidamente en valor absoluto y la PACF se corta después del *lag* 1

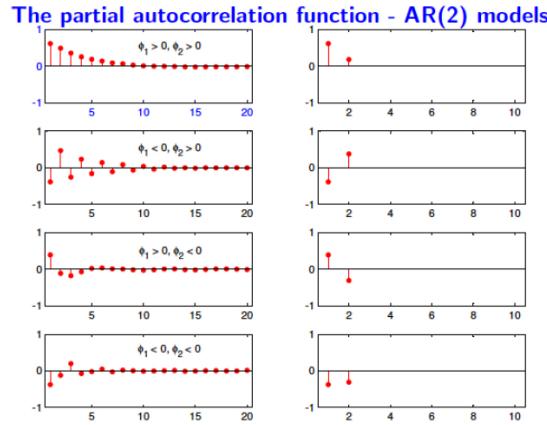


Figura 10: La ACF decrece rápidamente en valor absoluto y la PACF se corta después del *lag* 2

3.3.3. PROCESO AR(1)

Es un modelo con ecuación

$$X_t = \theta X_{t-1} + \epsilon_t$$

- Proceso AR(1) como proceso lineal**

$$X_t = \theta X_{t-1} + \epsilon_t = \epsilon_t + \theta(\epsilon_{t-1} + \theta X_{t-2}) = \epsilon_t + \theta\epsilon_{t-1} + \theta^2\epsilon_{t-2} = \epsilon_t + \theta\epsilon_{t-1} + \theta^2\epsilon_{t-2} + \theta^3\epsilon_{t-3} + \dots$$

donde para que X_t tenga variancia finita, $\sum_{j=0}^{\infty} \theta^{2j} < \infty$, es decir, $|\theta| < 1$.

En este caso, tiene sentido escribir $\Phi(z) = 1 - \phi z$ y por tanto:

- Proceso AR(1) como serie**

$$X_t = \Theta^{-1}\epsilon_t = \frac{1}{1 - \theta B}\epsilon_t = \left(\sum_{j=0}^{\infty} \theta^j B^j \right) \epsilon_t$$

con la condición de que X_t es *estacionario* y *causal* si las raíces de $\Theta(z)$ están fuera de $[-1, 1]$

■ ACF Proceso AR(1)

Observación 3.5. Veamos en primer lugar dos aclaraciones para calcular la ACF del modelo.

- Notamos en primer lugar que

$$\begin{aligned}\mathbb{E}(X_t) &= \phi\mathbb{E}(X_{t-1}) + \mathbb{E}(\epsilon_t) \\ &= \mathbb{E}\left(\sum_{i=0}^{\infty} (\theta^i \epsilon_{t-i})\right) \\ &= 0\end{aligned}$$

pues $\epsilon_t \sim N(0, \sigma^2)$

- Recordemos que si $\mathbb{E}(X_t) = \mu$, restamos μ en ambos lados.

Ahora, en primer lugar calculamos la función de AutoCovarianza

1. Para $h = 0$

$$\begin{aligned}\gamma_0 &= \text{Var}(X_t) = \text{Var}(\phi X_{t-1} + \epsilon_t) = \sigma^2 + \phi^2 \text{Var}(X_{t-1}) = \sigma^2 + \phi^2 \gamma_0 \\ \gamma_0 &= \text{Var}(X_t) = \frac{\sigma^2}{1 - \phi^2}\end{aligned}$$

2. Para $h \geq 1$

$$\begin{aligned}\gamma_h &= \mathbb{E}[(X_t)(X_{t+h})] = \mathbb{E}[(X_t)(\epsilon_{t+h} + \phi\epsilon_{t+h-1} + \phi^2\epsilon_{t+h-2} + \dots + \phi^{h-1}\epsilon_{t-1} + \phi^h X_t)] = \mathbb{E}(\phi^h X_t^2) = \phi^h \text{Var}(X_t) \\ &\Rightarrow \gamma_h = \phi^h \frac{\sigma^2}{1 - \phi^2}\end{aligned}$$

Por tanto tenemos

$$ACVF := \gamma_h = \begin{cases} \frac{\sigma^2}{1 - \phi^2} & h = 0 \\ \phi^h \gamma_0 & h \geq 1 \end{cases}$$

y de aquí se deriva que

$$ACF := \rho_h = \phi^h \quad h \geq 0$$

3.3.4. PROCESO AR(2)

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$$

$\epsilon_t \sim N(0, \sigma^2)$ white noise.

Queremos un modelo estacionario por lo que la media es constante, de donde

$$\mu = c + \mu(\phi_1 + \phi_2) \longrightarrow \mu = \frac{c}{1 - \phi_1 - \phi_2}$$

Por lo que el proceso tiene media finita si $1 - \phi_1 - \phi_2 \neq 0$. (Este razonamiento vale para cualquier proceso $AR(p)$ lo que en todos hemos considerado el proceso reemplazando X_t por $X_t - \mu$ y por tanto $c = 0$.

■ AR(2) Polinomio Caracteristico

$$1 - \phi_1 z - \phi_2 z^2$$

Tiene raíces

$$z_i = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} \quad i = 1, 2$$

El proceso es estacionario si y solo si las raíces están fuera del círculo unidad, si y solo si

$$\phi_1 + \phi_2 < 1 \quad \&\& \quad \phi_2 - \phi_1 < 1 \quad \&\& \quad |\phi_2| < 1$$

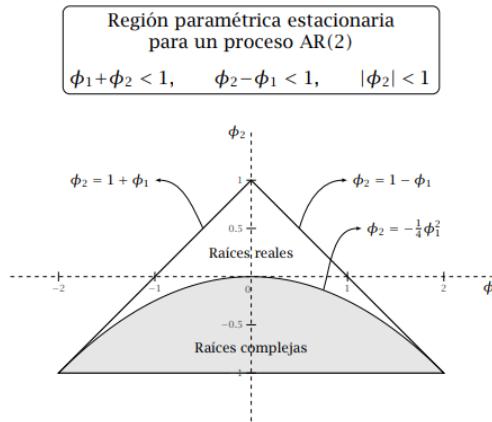


Figura 11: Raíces del proceso AR(2)

■ ACF Proceso AR(2)

Por el cambio que hemos hecho de X_t por $X_t - \mu$, tenemos $\mathbb{E}(X_t) = 0$. Observamos que

$$\mathbb{E}(X_t X_{t-k}) = \mathbb{E}(\phi_1 X_{t-1} X_{t-k} + \phi_2 X_{t-2} X_{t-k} + \epsilon_t X_{t-k})$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \quad k = 1, 2, \dots$$

dividiendo por γ_0

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

Ahora $\rho_k = \rho_{-k}$ de donde, con las **Ecuaciones de Yule-Walker** resueltas iterativamente con $\rho_0 = 1$

- $h = 1$:

$$\begin{aligned} \rho_1 &= \phi_1 \rho_0 + \phi_2 \rho_{-1} = \phi_1 + \phi_2 \rho_1 \\ \rho_1 &= \frac{\phi_1}{1 - \phi_2} \end{aligned}$$

- $h = 2$:

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0 = \phi_1 \frac{\phi_1}{1 - \phi_2} + \phi_2 = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1 - \phi_2^2}$$

Obtenemos pues que los tres primeros valores de la ACF de un AR(2) son

$$\rho_h = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\phi_1}{1 - \phi_2} & \text{si } h = 1 \\ \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1 - \phi_2} & \text{si } h = 2 \end{cases}$$

y el resto de ellos los podemos encontrar con

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad k \geq 2$$

Observación 3.6. Observemos que se puede resolver el sistema de ecuaciones:

$$\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \quad (1)$$

Se puede resolver para ϕ y eso nos da sus estimadores ³, a través de las estimaciones de ρ_i , $i = 1, 2$.

■ Varianza del proceso AR(2)

$$\mathbb{E}(X_t^2) = \mathbb{E}(\phi_1^2 X_{t-1}^2) + \mathbb{E}(\phi_2^2 X_{t-2}^2) + \mathbb{E}(2\phi_1\phi_2 X_{t-1}X_{t-2}) + \mathbb{E}(\epsilon_t^2)$$

El resto de combinaciones tienen esperanza 0 debido a que $\epsilon \sim WN$.

$$\gamma_0 = \phi_1^2 \gamma_0 + \phi_2^2 \gamma_0 + 2\phi_1\phi_2 \gamma_1 + \sigma^2$$

Como $\gamma_1 = \rho_1 \gamma_0$

$$\sigma_X^2 = \gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 - \phi_2)(1 - \phi_1^2 - \phi_2^2) - 2\phi_1^2\phi_2}$$

Que es finita bajo las mismas hipótesis de la media finita.

■ Proceso AR(2) como serie

Denotamos por

$$w_i = \frac{1}{z_i}$$

z_i raíces del polinomio característico. Entonces

$$(1 - w_1 B)(1 - w_2 B)X_t = \epsilon_t$$

Por tanto podemos escribir **AR(2) como MA(∞)**

$$X_t = (1 - w_1 B)^{-1}(1 - w_2 B)^{-1}\epsilon_t = \left(\sum_{j=0}^{\infty} w_1^j B^j \right) \left(\sum_{j=0}^{\infty} w_2^j B^j \right) \epsilon_t$$

$$1 + (w_1 + w_2)\epsilon_{t-1} + (w_1^2 + w_2^2 + w_1 w_2)\epsilon_{t-2} + \dots$$

Que es real, ya que en caso de tener raíces complejas, w_1 y w_2 serían conjugados.

■ Recursion de las Ecuaciones de Yule-Walker

Se puede demostrar que la solución de las **Ecuaciones de Yule-Walker de una proceso AR(2)** son de la forma

$$\rho_k = A_1 w_1^k + A_2 w_2^k$$

Como

$$\rho_0 = 1 \quad \rho_1 = \frac{\phi_1}{1 - \phi_2}$$

³Esto es solo útil para estimar los coeficientes AR, no funciona para MA ni para ARMA

Podemos encontrar A_1 y A_2 , ya que conocemos w_1 y w_2 (son las inversas de las raíces del polinomio característico).

Veamos ahora algunos ejemplos de representaciones de modelos AR(2).

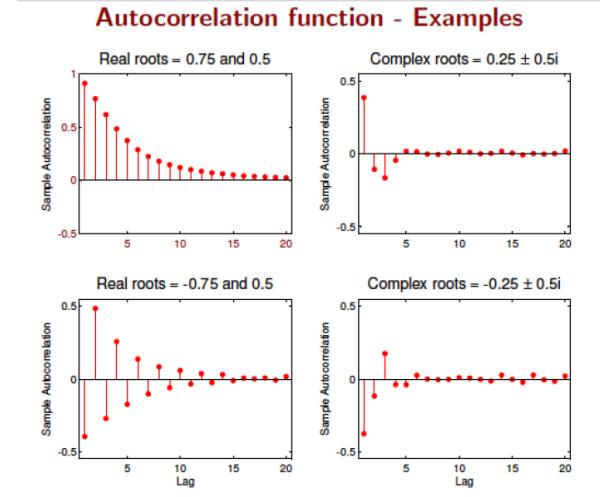
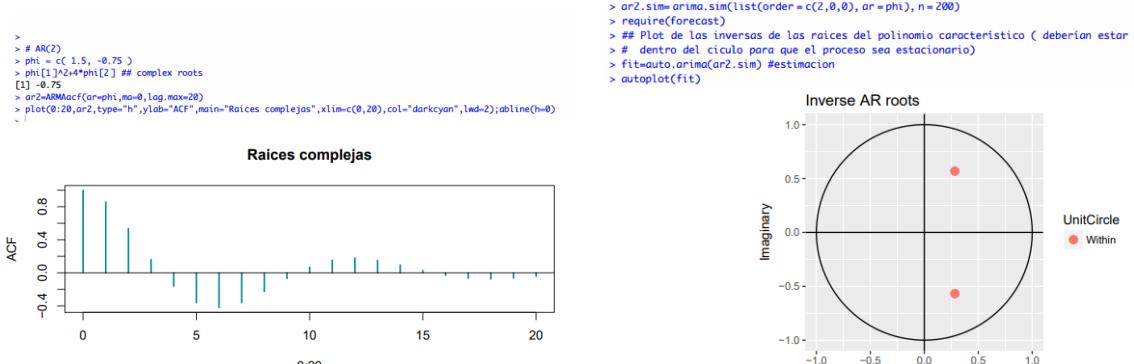


Figura 12: Ejemplos Funciones de AutoCorrelación modelos AR(2)



(a) Ejemplo ACF de un AR(2)

(b) Inversas de las raíces de un AR(2)

Ejemplos 3.7. Veamos un par de ejemplos

1. Escribamos la función ACF de un proceso AR(2):

$$X_t = 1,2X_{t-1} - 0,32X_{t-2} + \epsilon_t$$

Tiene ecuación característica:

$$\Phi(z) = 0,32z^2 - 1,2z + 1 = 0$$

por tanto tiene raíces $z_1 = 2,5, z_2 = 1,25$ por lo tanto el proceso es estacionario, y las inversas son $w_1 = 0,4$ y $w_2 = 0,8$. Por tanto, la ecuación característica la podemos poner como:

$$\Phi(z) = 0,32X^2 - 1,2X + 1 = (1 - 0,4X)(1 - 0,8X)$$

Entonces, el proceso es estacionario, con raíces reales y función de autocorrelación

$$\rho_k = A_1 0,4^k + A_2 0,8^k$$

Sabemos además que $\rho_0 = 1$ y $\rho_1 = \frac{\phi:1}{1-\phi_2} = \frac{1,2}{1,32} = 0,91$ por tanto

$$\begin{cases} A_1 + A_2 = 1 \\ 0,4A_1 + 0,8A_2 = 0,91 \end{cases}$$

por tanto $A_1 = -\frac{0,11}{0,4}$ y $A_2 = \frac{0,51}{0,4}$

2. Obtengamos los parametros de un proceso AR(3) cuyas primeras autocorrelaciones son $\rho_1 = 0,9$, $\rho_2 = 0,8$, $\rho_3 = 0,5$. Es estacionario?

► *The Yule-Walker equation system is:*

$$\begin{bmatrix} 0.9 \\ 0.8 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 1 & 0.9 & 0.8 \\ 0.9 & 1 & 0.9 \\ 0.8 & 0.9 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix}$$

whose solution is:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} = \begin{bmatrix} 5.28 & -5 & 0.28 \\ -5 & 10 & -5 \\ 0.28 & -5 & 5.28 \end{bmatrix} \begin{bmatrix} 0.9 \\ 0.8 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.89 \\ 1 \\ -1.11 \end{bmatrix}$$

Figura 14: Enter Caption

El proceso AR(3) con la correlación especificada es

$$X_t = 0,89X_{t-1} + X_{t-2} - 1,11X_{t-3} + \epsilon_t$$

Veamos las raíces del polinomio característico:

$$\Phi(z) = 1 - 0,89z - z^2 + 1,11z^3 = 0$$

tiene raíces $0,9243664 + 0,3098916i$, $-0,9478319$, $0,9243664 - 0,3098916i$, con normas $0,9749287$ y $0,9478319$ por lo tanto no es estacionario, ni causal.

3.4. PROCESOS LINEALES: ARMA(p,q)

Definición: ARMA(p,q)

En un proceso ARMA(p, q), se combinan ambas componentes autorregresiva y de media móvil para modelar la serie temporal. La idea detrás de este modelo es que los valores pasados de la serie (componente AR) y los errores pasados (componente MA) influyen en el valor presente.

Son procesos lineales con expresión

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

donde $\epsilon_t \sim WN(0, \sigma^2)$

3.4.1. Polinomio Característico

Un proceso ARMA(p,q) puede escribirse como

$$\Phi(B)X_t = \Theta(B)\epsilon_t$$

donde

$$\Phi(z) = 1 - \sum_{j=1}^p \phi_j z^j \quad \Theta(z) = 1 + \sum_{j=1}^q \theta_j z^j$$

Asumiendo que la serie tiene media cero ⁴, tenemos:

- Los procesos MA(q) son débilmente estacionarios para cualquier valor de ϕ_1, \dots, ϕ_q , pero para que el modelo sea identificable agregamos la condición de que las raíces del polinomio característico $\Phi(z) = 0$ se encuentren fuera del círculo unitario.
- Los procesos AR(p) y ARMA(p,q) no necesariamente definen series temporales estacionarias de segundo orden.

Observación: Raíces Comunes

En general, Φ y Θ no deben tener raíces comunes.

3.4.2. ARMA(1,1)

Definición modelo ARMA(1,1)

En proceso ARMA(1,1) puede escribirse como

$$X_t - \phi X_{t-1} = \epsilon_t + \theta \epsilon_{t-1}$$

es decir,

$$\Phi(B)X_t = \Theta(B)\epsilon_t$$

con

$$\Phi(z) = 1 - \phi z \quad \Theta(z) = 1 + \theta z$$

donde pediremos que $\phi + \theta \neq 0$ para que $\Phi(z) \neq \Theta(z)$.

2.4.2.1 Estacionariedad de un proceso ARMA(1,1)

En general, que necesitamos para que $X_t - \phi X_{t-1} = \epsilon_t + \theta \epsilon_{t-1}$ tenga solución estacionaria?

⁴Si no, eliminamos la media de cada observación, es decir, escribimos $X_t - \mu = a_1(X_{t-1} - \mu) + \dots + a_p(X_{t-p} - \mu) + \epsilon_t + \phi_1 \epsilon_{t-1} + \dots + \phi_q \epsilon_{t-q}$

- **Caso 1:** $|\phi| < 1$. En este caso

$$\chi(z) = \sum_{j=0}^{\infty} \phi^j z^j = \frac{1}{1 - \phi z} = \Phi^{-1}(z)$$

Como $\chi(z)\Phi(z) = 1$, entonces

$$\Psi(B)\epsilon_t := \chi(B)\Theta(B)\epsilon_t = \chi(B)\Phi(B)X_t$$

donde

$$\Psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = (1 + \phi B + \phi^2 B^2 + \dots)(1 + \theta B)$$

por lo tanto

$$\begin{aligned} B^0 : \psi_0 &= 1 \\ B^j : \psi_j &= \phi^{j-1}\theta + \phi^j = \phi^{j-1}(1 + \phi) \end{aligned}$$

por tanto, X_t puede ser escrito como un proceso MA(∞)

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \epsilon_t + (\phi + \theta) \sum_{j=1}^{\infty} \phi^{j-1} \epsilon_{t-j}$$

- **Caso 2:** $|\phi| \geq 1$. En este caso no es *causal* y no puede escribirse por tanto como un proceso MA(∞) que sea estacionario. No obstante puede encontrarse alguna solución *no causal* (que dependa del futuro) y que sea estacionaria.

2.4.2.2 Cálculo de la función ACF para un ARMA(1,1)

La ecuación del proceso ARIMA(1,1) es

$$X_t - \phi X_{t-1} = \epsilon_t + \theta \epsilon_{t-1}$$

Con tal de calcular γ_0 , multiplicamos en ambos lados por X_t y calculamos las esperanzas

$$\mathbb{E}(X_t^2) = \phi\gamma_1 + \mathbb{E}(X_t(\epsilon_t + \theta\epsilon_{t-1}))$$

$$\gamma_0 = \phi\gamma_1 + \mathbb{E}(X_t\epsilon_t) + \theta\mathbb{E}(X_t\epsilon_{t-1})$$

Ahora las covarianzas de X_t con ϵ_{t-1} NO son cero (X_t depende del pasado)

$$\mathbb{E}(X_t\epsilon_t) = \mathbb{E}(\phi X_{t-1}\epsilon_t) = 0 + \mathbb{E}(\epsilon_t^2) = \sigma^2 + \mathbb{E}(\theta\epsilon_{t-1}\epsilon_t) = 0 = \sigma^2$$

ya que

$$\begin{cases} \mathbb{E}(X_{t-1}\epsilon_t) = 0 \\ \mathbb{E}(\epsilon_t^2) = \sigma^2 \\ \mathbb{E}(\epsilon_{t-1}\epsilon_t) = 0 \end{cases}$$

pues los errores son incorrelacionados y X_{t-1} no depende del futuro.

$$\mathbb{E}(X_t\epsilon_{t-1}) = \mathbb{E}(\phi X_{t-1}\epsilon_{t-1}) + \mathbb{E}(\epsilon_t\epsilon_{t-1}) + \mathbb{E}(\phi\epsilon_{t-1}^2) = \phi\sigma^2 + 0 + \theta\sigma^2 = (\phi + \theta)\sigma^2$$

Por lo tanto

$$\gamma_0 = \phi\gamma_1 + \sigma^2(1 + \theta\phi + \theta^2)$$

y para $k = 1$

$$\begin{aligned}\mathbb{E}(X_t X_{t-1}) &= \mathbb{E}(X_{t-1}(\phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1})) \\ \gamma_1 &= \phi \gamma_0 + \theta \sigma^2\end{aligned}$$

y resolviendo el sistema

$$\gamma_0 = \frac{\sigma^2(1 + 2\theta\phi + \theta^2)}{1 - \phi^2} \quad \gamma_1 = \frac{\sigma^2(1 + \phi\theta)(\phi + \theta)}{1 - \phi^2}$$

Para $k > 2$

$$\begin{aligned}\gamma_k &= \mathbb{E}(X_t X_{t-k}) - \mathbb{E}(X_{t-1} X_{t-k}) = \mathbb{E}(\epsilon_t X_{t-k}) + \mathbb{E}(\theta \epsilon_{t-1} X_{t-k}) = 0 \quad k \geq 2 \\ \Rightarrow \gamma_k &= \phi \gamma_{k-1} = \phi^2 \gamma_{k-2} = \dots = \phi^{k-1} \gamma_1\end{aligned}$$

3.4.3. Representación ARMA(1,1) como MA(∞)

ARMA(1,1) com MA(∞)

Sea $\{X_t\}_{t \in T}$ un proceso ARMA(1,1) causal (con $\phi < 1$), entonces

$$\begin{aligned}X_t &= \phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \iff X_t(1 - \phi B) = \epsilon_t(1 + \theta B) \iff \\ X_t &= (1 - \phi B)^{-1} \epsilon_t(1 + \theta B) = \sum_{j=0}^{\infty} (\phi^j B^j)(1 + \theta B) \epsilon_t \\ X_t &= \epsilon_t + (\phi + \theta) \sum_{j=1}^{\infty} (\phi^{j-1} \epsilon_{t-j})\end{aligned}$$

Podemos ver otra forma de calcular la ACF a través de la representación como MA(∞) del proceso ARIMA(1,1).

$$\begin{aligned}\gamma_0 &= \mathbb{E}(X_t^2) = \mathbb{E}((\epsilon_t + (\phi + \theta) \sum_{j=1}^{\infty} (\phi^{j-1} \epsilon_{t-j}))^2) \\ &= \mathbb{E}(\epsilon_t^2) = \sigma^2 + 2\mathbb{E}(\epsilon_t(\phi + \theta) \sum_{j=1}^{\infty} (\phi^{j-1} \epsilon_{t-j})) = 0 + (\phi + \theta)^2 \mathbb{E}((\sum_{j=1}^{\infty} (\phi^{j-1} \epsilon_{t-j}))^2)_{\text{diag}} \\ &= \sigma^2 + (\phi + \theta)^2 \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \sigma^2 \frac{1 + \theta^2 + 2\phi\theta}{1 - \phi^2}\end{aligned}$$

Obtenemos el mismo resultado.

3.4.4. CAUSALIDAD E INVERTIBILIDAD

2.4.3.1 Causalidad: ARMA como MA(∞)

ARMA com MA(∞)

Veamos cuando es posible escribir un proceso ARMA como MA(∞).

$$X_t = \Phi^{-1}(B)\Theta(B)\epsilon_t$$

En primer lugar, escribimos

$$\Phi(z) = (1 - w_1z)(1 - w_2z) \cdots (1 - w_pz)$$

donde w_i son las inversas de las raíces de la ecuación característica $\Phi(z) = 0$. Por tanto

$$\Phi^{-1}(z) = \frac{1}{(1 - w_1z)} \frac{1}{(1 - w_2z)} \cdots \frac{1}{(1 - w_pz)}$$

Observación 3.8. Nótese que el producto $w_1 \times w_2 \times \dots \times w_p = 1$ pues el término independiente siempre es uno y por tanto $z_1 \times z_2 \times \dots \times z_p = 1$.

Así, tenemos el siguiente teorema

Teorema: Causalidad

Sea un proceso ARMA(p, q) con polinomio AR $\Phi(z)$ y polinomio MA $\Theta(z)$ sin raíces en común. Entonces el proceso ARMA es **causal** si y solo si las raíces de $\Phi(z)$ están fuera del círculo unidad. En ese caso:

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

donde los coeficientes se pueden obtener resolviendo

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\Theta(z)}{\Phi(z)} \quad |z| \leq 1$$

2.4.3.2 Invertibilidad:ARMA como AR(∞)

ARMA com AR(∞)

Un proceso ARMA(p, q) es **invertible** si puede escribirse como un proceso AR(∞):

$$\Pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = \epsilon_t$$

con $\Pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, $\sum_{j=0}^{\infty} |\pi_j| < \infty$ i $\pi_0 = 1$ (es el término independiente).

- Un proceso ARMA(p,q) es invertible si y solo si $\Theta(z) \neq 0$ para $|z| \leq 1$ y

$$\Pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\Phi(z)}{\Theta(z)} \quad |z| < 1$$

- En la representación AR(∞) de un proceso MA, el error más reciente se puede escribir como una función lineal de las observaciones actuales y pasadas

$$\epsilon_t = \sum_{j=0}^{\infty} (-\theta)^j X_{t-j}$$

- 1. Para un proceso invertible, $|\theta| < 1$ y, por tanto, las observaciones más recientes tienen mayor peso que las observaciones del pasado más distante.
- 2. Pero cuando $|\theta| > 1$, los pesos aumentan a medida que aumentan los *lags*, por lo que cuanto más distantes sean las observaciones, mayor será su influencia en el error actual.
- 3. Cuando $|\theta| = 1$, los pesos son de tamaño constante y las observaciones distantes tienen la misma influencia que las observaciones recientes.

Ninguna de estas últimas situaciones tiene sentido, por lo que preferimos la inveribilidad en procesos. De aquí, sacamos el siguiente teorema.

Teorema: Teorema: Invertibilidad

Sea un proceso ARMA(p,q) con polinomio AR $\Phi(z)$ y polinomio MA $\Theta(z)$ sin raíces en común. Entonces el proceso ARMA es **invertible** si y solo si las raíces de $\Theta(z)$ están fuera del círculo unidad. En ese caso:

$$\sum_{j=0}^{\infty} \pi_j X_{t-j} = \epsilon_t$$

donde los coeficientes π_j se pueden obtener resolviendo

$$\Pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\Phi(z)}{\Theta(z)} \quad |z| \leq 1$$

Si $\Phi(z)$ y $\Theta(z)$ tienen raíces comunes

Hay dos posibilidades

1. Ninguno de los ceros comunes se encuentra en el círculo unitario, en cuyo caso X_t es la única *solución estacionaria* de la ecuación ARMA sin ceros comunes, obtenida cancelando los factores comunes de $\Phi(z)$ y $\Theta(z)$.
2. Al menos uno de los ceros comunes se encuentra en el círculo unidad, en cuyo caso la ecuación ARMA tiene más de una solución estacionaria.

Observación 3.9. Si $\Phi(z)$ y $\Theta(z)$ no tienen factores en común y $\Phi(z)$ tiene un cero en el círculo unidad (raíz unitaria), entonces no hay ninguna solución estacionaria para las ecuaciones del proceso ARMA.

3.4.5. Ecuaciones de Yule-Walker de un proceso ARMA

Es posible escribir algo análogo a las ecuaciones de Yule-Walker para ARMA. De hecho, ya hicimos esto para ARMA(1,1). El sistema de ecuaciones nos permite obtener recursivamente los valores del ACF a partir de los coeficientes, pero desafortunadamente, no será posible (como en el caso AR) obtener estimadores para los coeficientes a partir del ACF empírico.

Cuidado con el matiz!!!!

- Un modelo AR (o ARMA) **CAUSAL** es **ESTACIONARIO**.
- Un modelo AR (o ARMA) *puede tener una solución ESTACIONARIA y ser NO CAUSAL* ver ejercicio 4 examen 2020
- En Series Temporales, no tiene mucho sentido trabajar con procesos NO causales.
- **NO PUEDE HABER UN PROCESO CAUSAL CON LAS RAÍCES DENTRO DEL CÍRCULO UNIDAD**, entonces si vemos que la naturaleza de la expresión no depende del futuro pero tiene una raíz en el círculo unidad, será **no estacionario**, en cambio si la expresión **depende del futuro** podríamos encontrar una solución estacionaria (ver examen 2020).

Ejemplo Examen 2020

Consideramos el proceso $Y_t = 3Y_{t-1} + \epsilon_t$ con $\epsilon_t \sim WN(0, \sigma^2)$

(a) Muestra que el proceso Y_t es un *AR(1)* no estacionario.

Notemos que una condición **necesaria** para que un proceso *AR(p)* sea estacionario es que las raíces del polinomio que definen la ecuación del modelo estén fuera del círculo unidad.

En este caso:

$$Y_t = 3Y_{t-1} + \epsilon_t \iff Y_t(1 - 3B) = \epsilon_t$$

Por lo que tiene raíz $z_1 = \frac{1}{3}$, por tanto no es estacionario, ya que

$$Y_t = (1 - 3B)^{-1}\epsilon_t = \sum_{j=0}^{\infty} (3^j B^j) \epsilon_t = \sum_{j=0}^{\infty} (3^j \epsilon_{t-j})$$

por lo que

$$Var(Y_t) = \sigma^2 \sum_{j=0}^{\infty} (3^{2j}) = \infty$$

por tanto el modelo no es estacionario.

Ejemplo Examen 2020

(b) Muestra que el proceso definido a través de la serie

$$X_t = - \sum_{j=1}^{\infty} \left(\frac{1}{3}\right)^j \epsilon_{t+j}$$

satisface la ecuación

$$X_{t+1} = 3X_t + \epsilon_{t+1}$$

Simplemente hay que calcularlo y se da

(c) Muestra que el proceso X_t definido en (b) es estacionario (débilmente)

Cabe notar que

$$\mathbb{E}(X_t) = 0$$

$$Var(X_t) = -\sigma^2 \sum_{j=1}^{\infty} \left(\frac{1}{3}\right)^{2j} = \sigma^2 \frac{1}{1-\frac{1}{9}} = \frac{9\sigma^2}{8} < \infty$$

Como el proceso es lineal, satisface las dos condiciones que ha de cumplir, por tanto es estacionario.

(d) En qué sentido es esta solución insatisfactoria?

Esta solución a la ecuación AR(1) es insatisfactoria en el sentido que depende del futuro, por tanto no es causal. En general en ST, los modelos han de ser causales.

(e) Esto demuestra que si una ecuación de un modelo AR (o la ecuación de la parte AR del modelo ARMA) tiene raíces fuera del círculo unidad, entonces **no tiene soluciones (estacionarias) causales**.

Ejemplo Examen 2019

Supon un modelo $\{X_t\}_{\{t=\dots,-2,-1,0,1,2,\dots\}}$ es un AR(1) causal y estacionario.

(a) Escribe el proceso ARMA Y_t que se obtiene diferenciando X_t , es decir

$$Y_t = X_t - X_{t-1}$$

Notemos que

$$\begin{aligned} X_t - X_{t-1} &= \phi(X_{t-1} - X_{t-2}) + \epsilon_t - \epsilon_{t-1} \\ \Rightarrow Y_t &= \phi Y_{t-1} + \epsilon_t - \epsilon_{t-1} \end{aligned}$$

Que es la ecuación de un proceso ARMA(1,1).

(b) Es estacionario? E invertible?

Notamos que es estacionario pues $|\phi| < 1$ y por tanto

$$Y_t(1 - \phi B) = (1 - B)\epsilon_t$$

por lo que

$$Y_t = \sum_{j=0}^{\infty} (\phi^j B^j)(1 - B)\epsilon_t = \sum_{j=0}^{\infty} (\phi^j (\epsilon_{t-j} - \epsilon_{t-j-1}))$$

Y por tanto tiene varianza y esperanza finita, por lo que es estacionario.

Por otro lado, la raíz del polinomio $\Theta(z) = 1 - z$ es 1, por lo que tiene una raíz unitaria y por tanto NO es invertible.

Ejemplo Examen 2019

(c) Calcula la varianza y la ACF

$$\begin{aligned}\text{Var}(Y_t) &= \text{Var}(X_t - X_{t-1}) = \mathbb{E}((X_t - X_{t-1})^2) = \mathbb{E}(X_t^2) - 2\mathbb{E}(X_t X_{t-1}) + \mathbb{E}(X_{t-1}^2) = \\ &= 2\text{Var}(X_t) - 2\text{Cov}(X_t, X_{t-1})\end{aligned}$$

Ahora bien

$$\begin{aligned}\text{Var}(X_t) &= \mathbb{E}((\phi^2)X_{t-1}^2) + 2\phi\mathbb{E}(X_{t-1}\epsilon_t) = 0 + \mathbb{E}(\epsilon_t^2) \\ \text{Var}(X_t) &= \frac{\sigma^2}{1 - \phi^2}\end{aligned}$$

$$\text{Cov}(X_t, X_{t-1}) = \phi\mathbb{E}(X_{t-1}^2) + \mathbb{E}(\epsilon_t X_t) = 0 = \phi \frac{\sigma^2}{1 - \phi^2}$$

Por tanto

$$\begin{aligned}\text{Var}(Y_t) &= 2\frac{\sigma^2}{1 - \phi^2} - 2\phi\frac{\sigma^2}{1 - \phi^2} = 2\sigma^2 \frac{1 - \phi}{1 - \phi^2} \\ &= 2\sigma^2 \frac{1}{1 + \sigma}\end{aligned}$$

Ahora queremos calcular la covarianza

$$\begin{aligned}\mathbb{E}(Y_t Y_{t-h}) &= \mathbb{E}((X_t - X_{t-1})(X_{t-h} - X_{t-h-1})) = \\ &= \mathbb{E}(X_t X_{t-h}) - \mathbb{E}(X_t X_{t-h-1}) - \mathbb{E}(X_{t-1} X_{t-h}) + \mathbb{E}(X_{t-1} X_{t-h-1}) \\ &= \gamma_X(h) - \gamma_X(h+1) - \gamma_X(h-1) + \gamma_X(h) \\ &= 2\gamma_X(h) - \gamma_X(h+1) - \gamma_X(h-1) = \gamma_X(0)\phi^{h-1}(2\phi - 1 - \phi^2) = \\ &= -\frac{\sigma^2}{1 - \phi^2}\phi^{h-1}(\phi - 1)^2 = -\sigma^2\phi^{h-1}\frac{1 - \phi}{1 + \phi}\end{aligned}$$

3.4.6. Identificación de modelos

Las funciones ACF y PACF son herramientas útiles para identificar AR(p) y MA(q) puros.

- Para un modelo AR(p), la verdadera ACF decaerá hacia cero.
- Para un modelo AR(p), la verdadera PACF será cero después del *lag* p.
- Para un modelo MA(q), la verdadera ACF será cero después del *lag* q.
- Para un modelo MA(q), la verdadera PACF decaerá hacia cero.

Si proponemos un modelo AR(p) o un modelo MA(q) para una serie de tiempo observada, podríamos examinar el ACF o PACF de muestra para ver si se aproximan a lo que se vería el verdadero ACF o PACF para esta propuesta modelo.

Extended Auto Correlation Function (EACF)

La idea detrás del uso de la **Función de AutoCorrelación Extendida (EACF)** para identificar procesos ARMA es la siguiente:

- Si conociera la parte AR del proceso, se puede filtrar y lo que queda es un MA.
- Dado que la parte AR no se conoce, se puede estimar mediante OLS.
- Los residuos $w_{t,k,j}$ (asumiendo que el modelo es un ARMA(k, j)) deben tener ACF nulo para $lags > j$, siempre que k sea el orden correcto para la parte AR.
- Como se desconocen los órdenes, los ajustes se realizan de forma recursiva para una tabla de valores de p y q .

Veamos ejemplos de esto

AR/MA	0	1	2	3	4	5	6	7	8	9	10	AR/MA	0	1	2	3	4	5	6	7	8	9	10
0	x	x	x	x	x	x	x	x	x	x	x	0	x	x	x	x	x	x	x	x	x	x	
1	x	0	0	0	0	0	0	0	0	0	0	1	x	x	x	x	x	x	x	x	x	x	
2	x	x	0	0	0	0	0	0	0	0	0	2	x	x	x	0	0	0	0	0	0	0	0
3	x	x	x	0	0	0	0	0	0	0	0	3	x	x	x	x	0	0	0	0	0	0	0
4	x	x	x	x	0	0	0	0	0	0	0	4	x	x	x	x	x	0	0	0	0	0	0
5	x	x	x	x	x	0	0	0	0	0	0	5	x	x	x	x	x	x	0	0	0	0	0
6	x	x	x	x	x	x	0	0	0	0	0	6	x	x	x	x	x	x	x	0	0	0	0
7	x	x	x	x	x	x	x	0	0	0	0	7	x	x	x	x	x	x	x	x	0	0	0

Figura 15: EACF teórico de un ARMA(1,1)

Figura 16: EACF teórico de un ARMA(2,3)

Veamos en el ejemplo del ARMA(1,1), como sería simulado

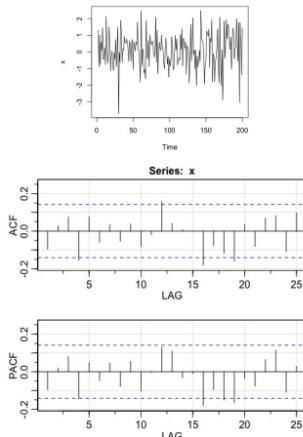


Figura 17: ACF y PACF de un ARMA(1,1) simulado

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	o	o	x	o	o	o	o	o	o	o	x	o	o
1	x	o	o	o	o	o	o	o	o	o	o	o	o	o
2	x	x	o	o	o	o	o	o	o	o	o	o	o	o
3	x	x	x	o	o	o	o	o	o	o	o	o	o	o
4	x	x	x	x	o	o	o	o	o	o	o	o	o	o
5	x	o	o	x	o	o	o	o	o	o	o	o	o	o
6	x	o	o	o	o	o	o	o	o	o	o	o	o	o
7	x	x	o	o	o	o	o	o	o	o	o	o	o	o

Figura 18: EACF de un ARMA(1,1) simulado

Figura 19: Proceso ARMA(1,1) simulado con $\phi = 0,6$ y $\theta = -0,5$

4. DIAGNOSIS

4.1. Datos Color

En primer lugar, veamos la gráfica de los datos, su función ACF y PACF.

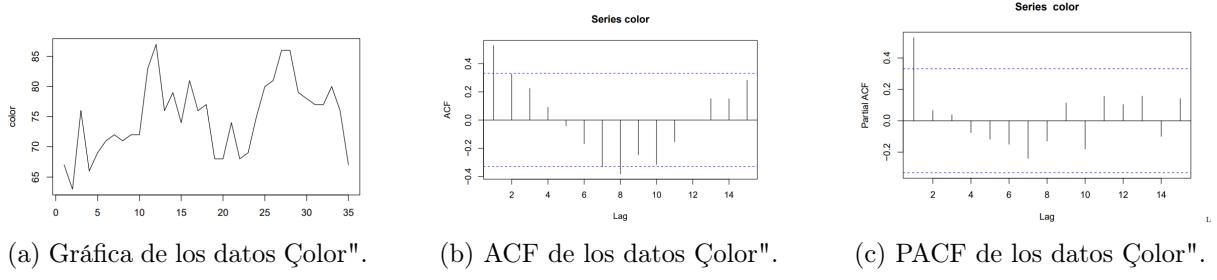


Figura 20: ACF decrece rápidamente y la PACF se anula a partir del lag 1, así que son compatibles con un AR(1).

4.1.1. Ajuste del modelo

Estimamos con la función `arima` un modelo AR(1).

```
m1.color=arima(color,order=c(1,0,0)); m1.color
##
## Call:
## arima(x = color, order = c(1, 0, 0))
##
## Coefficients:
##             ar1  intercept
##             0.5705    74.3293
## s.e.   0.1435   1.9151
##
## sigma^2 estimated as 24.83:  log likelihood = -106.07,  aic = 216.15
```

(a) Código R del ajuste del modelo

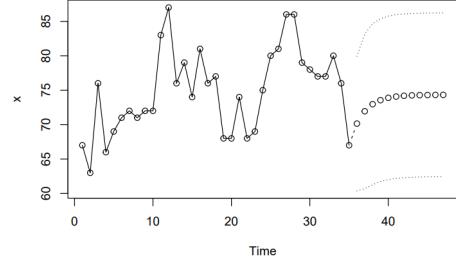


Figura 21: Ajuste del modelo AR(1)

El modelo ajustado es

$$(1 - 0.5705B)(X_t - 74.3293) = \epsilon_t \quad \epsilon \sim N(0, 24.83)$$

Observación 4.1. El intercepto aquí es la estimación de la media del proceso

4.1.2. Comportamiento de los residuos

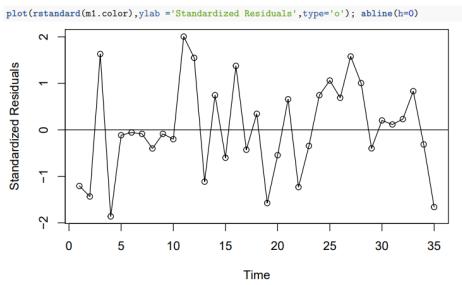
Para un modelo AR, los residuos se definen de la manera obvia, supongamos que la serie tiene media 0,

$$\hat{\epsilon}_t = X_t - \hat{\phi}_1 X_{t-1} - \dots - \hat{\phi}_p X_{t-p}$$

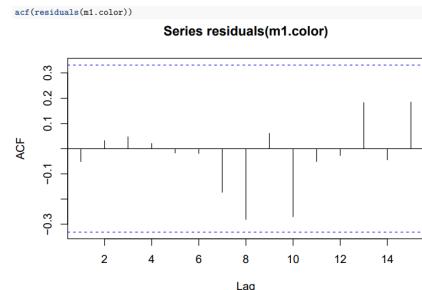
Para un modelo ARMA más general, causal e invertible, los tenemos que definir a partir de la representación como un AR(∞),

$$\epsilon_t = \Theta^{-1}(B)\Phi(B)X_t = \Pi(B)X_t$$

Observación 4.2. En la práctica, los residuos se obtienen como un sub-producto de la predicción: para cada t , el residuo se obtiene como la diferencia entre el valor observado y la predicción.



(a) Plot de los residuos estandarizados



(b) ACF de los residuos.

Figura 22: Tanto el aspecto del plot como la ACF, parecen compatibles con un ruido blanco

Si los residuos fueran *iid* y tuvieran varianza finita, las ACF serían:

1. Normales.
2. Con media 0
3. Varianza $\frac{1}{n}$

Por eso los plots de las ACF marcan el intervalo $\pm 1,96 \frac{1}{\sqrt{n}}$.

Además de mirar los correlogramas (ACF) para saber si los residuos son ruido, podemos hacer un test. El test de *Ljung-Box* tiene como hipótesis nula que los residuos son ruido blanco estricto (es decir, independientes):

$$H_0 : \epsilon_t \sim SWN(0, \sigma^2)$$

El estadístico de contraste es

$$Q_{LB} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}^2(j)}{n-j}$$

cuya distribución asintótica bajo H_0 es χ_h^2 .

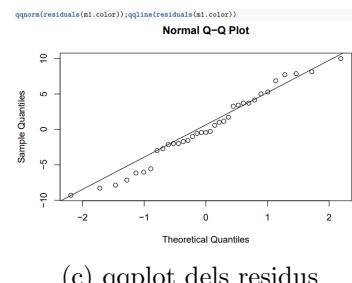
Si los residuos no son un ruido blanco estricto (es decir, si no son independientes), la aproximación no necesariamente es buena.

Hay otros tests similares, por ejemplo, el de *Mc Leod -Li*, que usa la ACF de los cuadrados de los datos.

```
shapiro.test(residuals(m1.color))
## 
## Shapiro-Wilk normality test
## data: residuals(m1.color)
## W = 0.97536, p-value = 0.6057
## X-squared = 0.10056, df = 1, p-value = 0.7512
```

(b) textit{Shapiro-Wilk}

(a) *Ljung-Box* test amb *pva*-normality test amb *pval*=0.6057



(c) qqplot dels residus

Figura 23: Estudio de la normalidad e independencia de los residuos.

El *p-valor* del test de independencia de *Ljung-Box test* (Portmanteau) para los residuos, no nos permite rechazar la hipótesis de independencia. Además el plot de normalidad y el textit{Shapiro-Wilk normality test} indican que la distribución de los residuos es compatibles con una distribución normal.

Otra forma de obtener el plot de diagnóstico

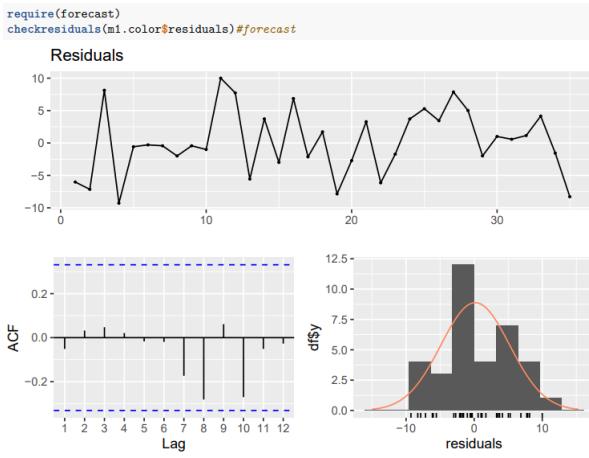


Figura 24: Plot de diagnóstico y estudio de los residuos

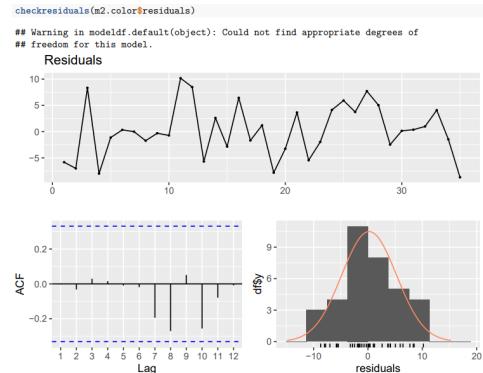
En conclusión, no hay motivo aparente para no aceptar el modelo 1.

4.1.3. Ajuste de otro modelo (AR(2))

Veamos qué ocurriría si ajustamos un modelo AR(2).

```
m2.color=arima(color,order=c(2,0,0)); m2.color
##
## Call:
## arima(x = color, order = c(2, 0, 0))
##
## Coefficients:
##      ar1     ar2  intercept 
##  0.5173  0.1005   74.1551 
##  s.e.  0.1717  0.1815   2.1463 
## 
## sigma^2 estimated as 24.6:  log likelihood = -105.92,  aic = 217.84
```

(a) Ajuste modelo AR(2)



(b) Estudio de los residuos del modelo AR(2).

El AIC es mas grande! y ϕ_1 es parecido en ambos casos. Los residuos son parecidos a los del AR(1). No hay razón para adoptar el modelo más complicado.

Pero qué es el *AIC (Akaike Information Criterion)*?

Definición: AIC (Akaike Information Criterion)

Es un criterio basado en un equilibrio entre la verosimilitud (evaluada en el MLE), penalizada por el número de parámetros del modelo (k).

$$AIC = -2\log(L) + 2k$$

Hay otros criterios que penalizan la complejidad del modelo de otra forma, por ejemplo el BIC (Bayesian Information Criterion)

Definición: BIC (Bayesian Information Criterion)

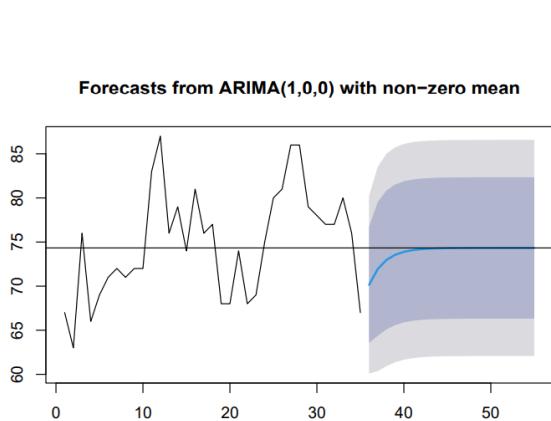
$$BIC = -2\log(L) + \log(n)k$$

Para elegir entre dos modelos, preferiremos el que tenga menor AIC o BIC (y no necesariamente los criterios coinciden).

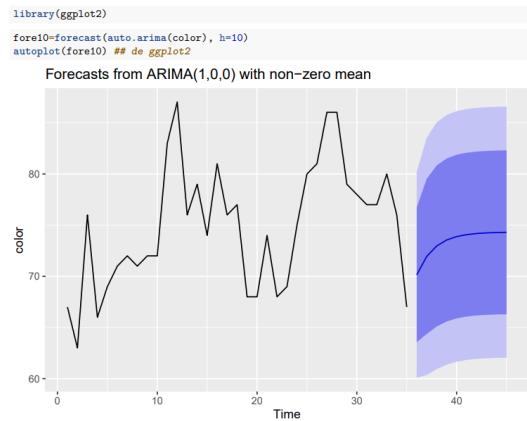
4.1.4. Forecast para AR(1)

Lo podemos hacer tanto con `plot` como con `ggplot`

```
plot(forecast(auto.arima(color), h=20))
abline(h=coef(m1.color)[names(coef(m1.color))=='intercept'])
```



(a) Forecast con `plot`



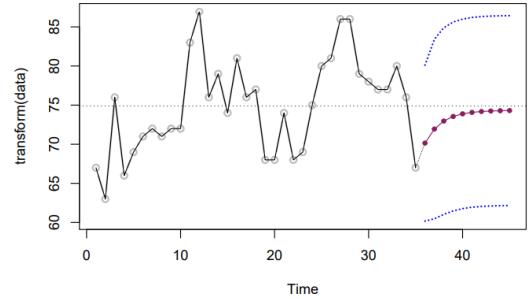
(b) Forecast con `ggplot`

El siguiente código es para una función que sirve para hacer los plots de las predicciones sin preocuparse de cuántas y cuáles librerías tengamos cargadas y podemos cambiarle el aspecto a las gráficas cambiando color, tipo de línea, nombre en los ejes, etc.

`Forecastplot`

```
#####
# Forecastplot
forecastplot <- function (data, fit, n.ahead = 1, coef = 2,
                           showmean= FALSE, transform = function(x) x, ...)
{
  p = predict(fit, n.ahead = n.ahead)
  pred = p$pred
  pse = p$se
  xlim <- range(time(data), time(p$pred))
  ylim <- range(transform(data),
                transform(pred - coef * pse),
                transform(pred + coef * pse))
  plot(transform(data),type="b",col="grey", lwd=2, xlim = xlim, ylim = ylim, ...)
  if (showmean)
    abline(h = transform(mean(data)), lty=dotted)
  if (all(start(lag(pred, 1)) == end(data))) {
    jseg = ts(c(window(data, end(data)),
               window(pred, end = start(pred))),
              start = end(data), frequency = frequency(data))
    lines(jseg, lty = "11", lwd = 1)
  }
  lines(transform(data))
  # lines(transform(p$pred), lty = "11", lwd = 2)
  # points(transform(p$pred), pch = 20, col = "black")
  lines(transform(p$pred, type = "o", pch = 20, col="maroon4"))
  lines(transform(p$pred - coef*p$se), lty = "11", lwd = 2, col="blue")
  lines(transform(p$pred + coef*p$se), lty = "11", lwd = 2, col="blue")
}
forecastplot(color,m1.color,n.ahead=10,showmean=TRUE)
```

(a) Código para obtener siguiente gráfica



(b) Forecast plot

4.2. Número de liebres atrapadas cerca del río Hudson entre 1905 y 1931

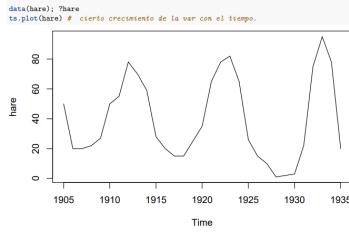
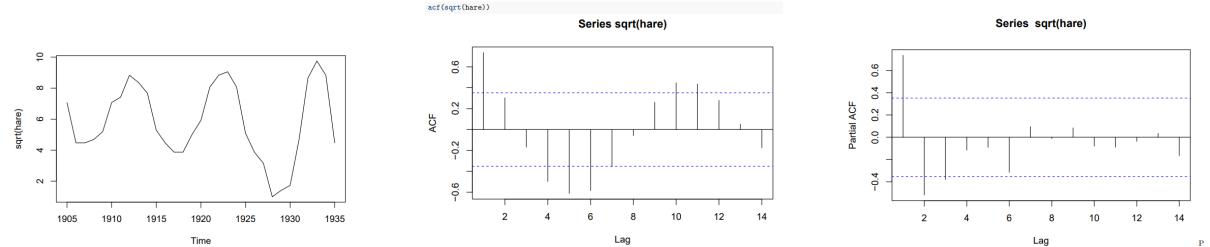


Figura 28: Gráfico de nuestros nuevos datos



(a) Gráfica de la raíz de los datos "Liebres".
(b) ACF de la raíz de los datos "Liebres".
(c) PACF de la raíz de los datos "Liebres".

Figura 29: Para estabilizar la varianza de la serie original, usamos la raíz de los datos. La ACF y PACF parecen de un modelo AR(3). Es interesante observar los signos de la correlación.

4.2.1. Ajuste del modelo

```
m1.hare=arima(sqrt(hare),order=c(3,0,0)); m1.hare
## 
## Call:
## arima(x = sqrt(hare), order = c(3, 0, 0))
## 
## Coefficients:
##          ar1      ar2      ar3  intercept 
##        1.0519 -0.2292 -0.3931   5.6923 
##  s.e.  0.1877  0.2942  0.1915   0.3371 
## 
## sigma^2 estimated as 1.066: log likelihood = -46.54,  aic = 101.08

(a) Código R del ajuste del modelo AR(3). Observamos que  $\phi_2$  es muy pequeño y mirando los errores estándares podríamos pensar que es estadísticamente nulo. Lo fijamos con 0.
```

```
m2.hare=arima(sqrt(hare),order=c(3,0,0),fixed=c(NA,0,NA,NA))
# mas bajo el AIC haciendo phi_2=0
m2.hare
## 
## Call:
## arima(x = sqrt(hare), order = c(3, 0, 0), fixed = c(NA, 0, NA, NA))
## 
## Coefficients:
##          ar1      ar2      ar3  intercept 
##        0.9190     0   -0.5313   5.6889
```

(b) Código R del ajuste del modelo AR(3) con $\phi_2 = 0$ obteniendo $\sigma^2 = 1,088$, $\log_{\text{likelihood}}=-46.85$ y $\text{AIC}=99.69$

Figura 30: Ajuste del modelo AR(1):

$$Y_t = 0,919Y_{t-1} - 0,5314Y_{t-3} + \epsilon_t$$

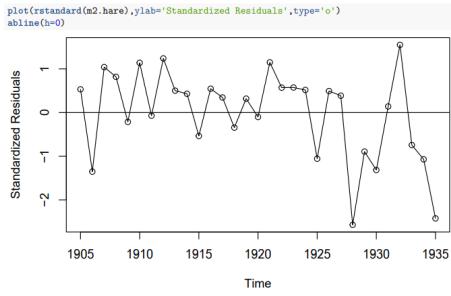
Recordemos que el término de intercepto nos da la media, así que

$$Y_t = \sqrt{hare} - 5,69$$

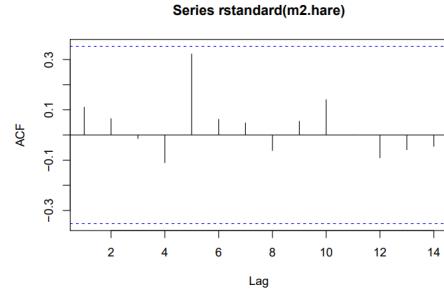
con término independiente

$$5,6898(1 - 0,919 + 5,5313) = 3,483.$$

4.2.2. Estudio de los residuos



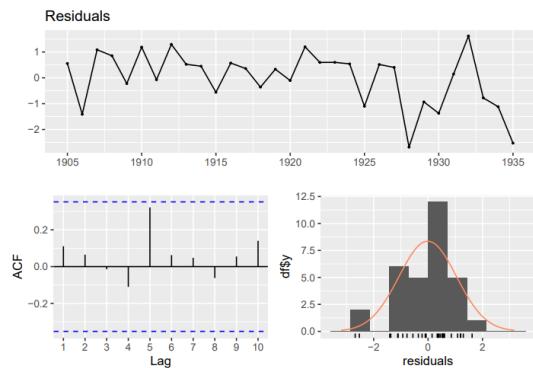
(a) Plot de los residuos estandarizados para el segundo modelo



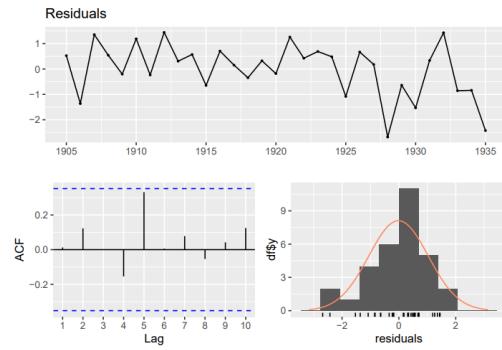
(b) ACF de los residuos para el segundo modelo.

Figura 31: Tanto el aspecto del plot com la ACF, parecen compatibles con un ruido blanco

Veamos el checkresiduals para cada uno de los modelos



(a) Estudio de los residuos del modelo 2



(b) Estudio de los residuos del modelo 1

4.2.3. Forecast

Haciendo:

```
forehare=forecast(sqrt(hare),model=arima(sqrt(hare),order=c(3,0,0),fixed=c(NA,0,NA,NA)),h=10)
```

obtenemos con un `autoplot(forehare)` (a la izquierda).

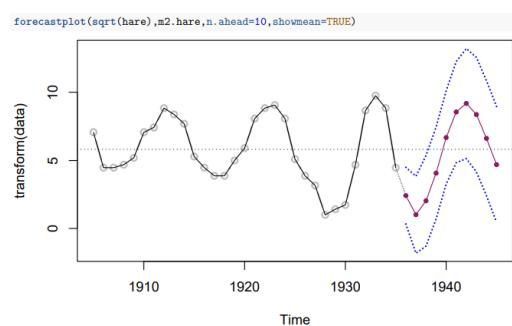
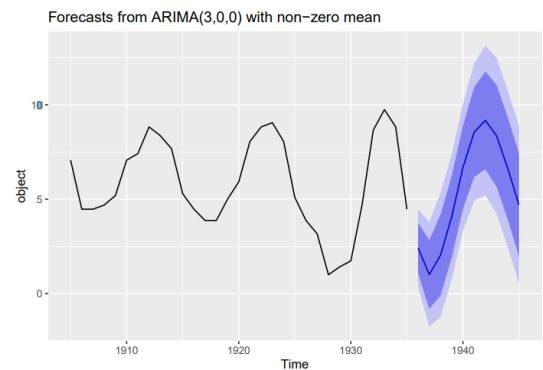
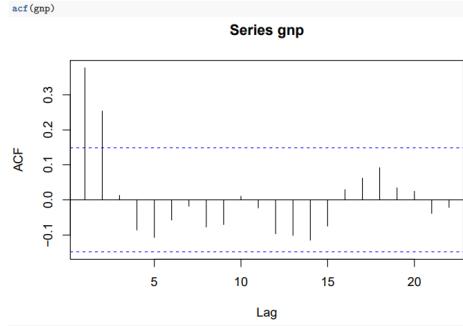


Figura 33: Dos predicciones con diferentes métodos.

4.3. Ejemplo Tsay sobre US GNP worth eate 1947 to 1991

```
da=read.table("q-gnp4791.txt"
gpn=da$V1
```



(a) ACF de los datos.

```
eacf(gnp) # ARMA(2,2)?
## AR/MA
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x o o o o o o o o o o o o
## 1 x x o o o o o o o o o o o o
## 2 x x o o o o o o o o o o o o
## 3 x x x o o o o o o o o o o o
## 4 x x x o o o o o o o o o o o
## 5 x x o o o o o o o o o o o o
## 6 x x o x x o o o o o o o o o o
## 7 o x x x o o o o o o o o o o o
```

(b) EACF de los datos.

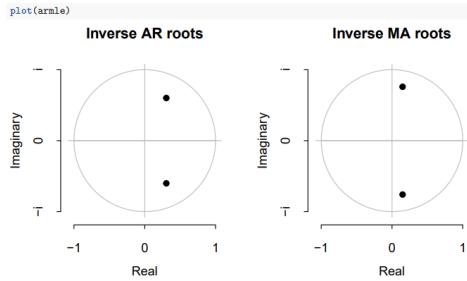
Figura 34: Parece claro que la componente MA será 2 y la AR irá entre 0 y 3

Veamos que nos recomienda la función `auto.arima`⁵

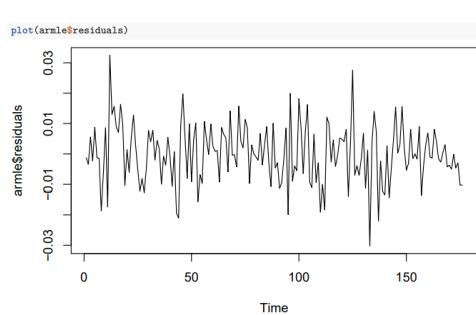
```
armle=auto.arima(gnp);armle
## Series: gnp
## ARIMA(2,0,2) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ma1      ma2      mean
##          0.6090  -0.4541  -0.2988  0.5991  0.0077
##  s.e.   0.1623   0.1684   0.1379  0.1712  0.0011
##
## sigma^2 = 9.513e-05: log likelihood = 567.5
## AIC=-1122.99  AICc=-1122.5  BIC=-1103.97
```

Figura 35: Auto.Arima a nuestros datos gnp. Nos recomienda un modelo ARMA(2,2)

Podemos ver graficadas las raíces y los residuos del modelo.



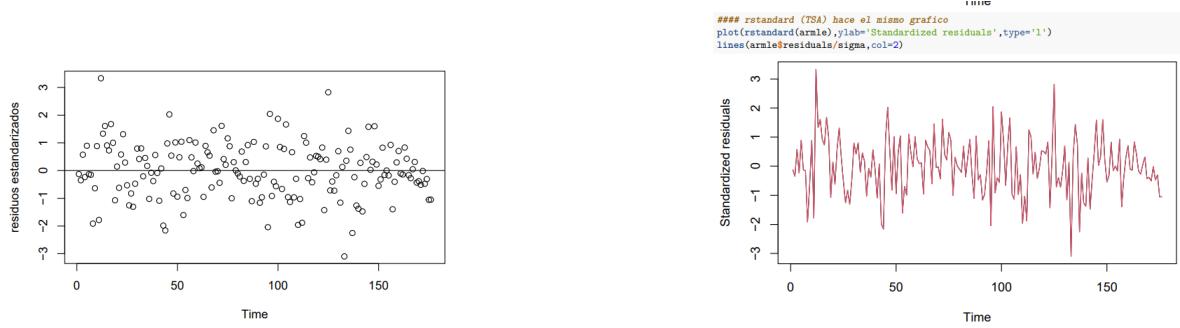
(a) Inversas de las raíces de las ecuaciones características



(b) Residuos del modelo ARMA(2,2)

⁵Obsérvese que en esta función ya aparece *mean* y no *intercept*.

Podemos ver una representación de los residuos como serie temporal.



(a) Representación de los residuos junto a la media

(b) Representación de los residuos como Serie Temporal.

Por último, veamos si los residuos son o no normales.

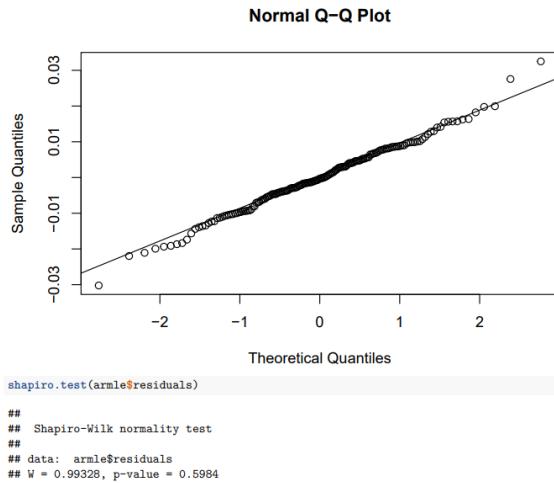


Figura 38: Como se ve en el gráfico y el no poder rechazar la hipótesis nula del test, concluimos que los residuos son normales.

5. FORECASTING

5.1. El "mejor" predictor

Definición: Mejor predictor

Dada una serie estacionaria ^a X_t , con media 0 y covarianzas γ_h , queremos predecir X_{t+h} basándonos en cierta información conocida \mathbf{Z} (generalmente hasta el instante t), $\mathbf{Z} = \{X_t, \dots, X_{t-p}\}$.

Dado el \mathbf{Z} , un **predictor** (o **estimador**) \hat{X}_{t+h} será una función de \mathbf{Z} :

$$\hat{X}_{t+h} = F(\mathbf{Z})$$

La propuesta de Box & Jenkins es usar aquella F que minimice el error cuadrático medio (ECM)

$$\operatorname{argmin}\{\mathbb{E}((X_{t+h} - F(\mathbf{Z}))^2)\}$$

^aNo tiene por qué ser estacionaria.

5.1.1. El mejor predictor es la esperanza condicional

Teorema: El mejor predictor es la esperanza condicional

Si X y Z son dos variables aleatorias con $\mathbb{E}(X^2) < \infty$ y densidad conjunta $f_{X,Z}$, entonces la función F que minimiza el ECM, $\mathbb{E}((X - F(Z))^2)$ es $F(Z) = \mathbb{E}(X|Z)$, donde

$$\mathbb{E}(X|Z = z) = \int_{-\infty}^{\infty} x f_{X|Z=z}(x) dx$$

Demostración. Supongamos que el mínimo se alcanza en otra función $G(Z)$, entonces

$$\begin{aligned} \mathbb{E}((X - G(Z))^2) &= \mathbb{E}((X - \mathbb{E}(X|Z) + \mathbb{E}(X|Z) - G(Z))^2) \\ &= \mathbb{E}((X - \mathbb{E}(X|Z))^2) + \mathbb{E}((\mathbb{E}(X|Z) - G(Z))^2) + \\ &\quad 2\mathbb{E}((X - \mathbb{E}(X|Z))(\mathbb{E}(X|Z) - G(Z))) \\ &\geq \mathbb{E}((X - \mathbb{E}(X|Z))^2) \end{aligned}$$

porque el segundo sumando es positivo y el tercero es nulo ya que la esperanza de $((X - \mathbb{E}(X|Z))(\mathbb{E}(X|Z) - G(Z)))$ se calcula condicionando en Z y usando que

$$\mathbb{E}(\mathbb{E}(Y|Z)) = \mathbb{E}(Y)$$

tenemos que

$$\mathbb{E}((X - \mathbb{E}(X|Z))(\mathbb{E}(X|Z) - G(Z))) = \mathbb{E}((\mathbb{E}((X - \mathbb{E}(X|Z))(\mathbb{E}(X|Z) - G(Z)))|Z))$$

Pero dado Z , tenemos que

$$\mathbb{E}(X|Z) - G(Z)$$

es constante, por lo que

$$\mathbb{E}((\mathbb{E}((X - \mathbb{E}(X|Z))(\mathbb{E}(X|Z) - G(Z)))|Z)) = \mathbb{E}((\mathbb{E}(X|Z) - G(Z))(\mathbb{E}((X - \mathbb{E}(X|Z))|Z)))$$

Pero dado que la esperanza condicional también es lineal, tenemos que

$$\mathbb{E}((X - \mathbb{E}(X|Z))|Z) = \mathbb{E}(X|Z) - \mathbb{E}(\mathbb{E}(X|Z)|Z) = \mathbb{E}(X|Z) - \mathbb{E}(X|Z) = 0$$

Por tanto en la primera ecuación, obtenemos que

$$\mathbb{E}((\mathbb{E}(X|Z) - \mathbb{E}(X|Z))(\mathbb{E}(X|Z) - G(Z))|Z)) = 0$$

Por lo tanto concluimos que

$$\mathbb{E}((X - G(X))^2) \geq E((X - E(X|Z))^2)$$

es decir, el mejor predictor de futuro (respecto del error cuadrático medio) es la esperanza condicional (respecto del pasado). \square

5.1.2. Caso Gaussiano

Lema: Caso Gaussiano

Si $X_{t+h}, X_t, \dots, X_{t-p}$ tienen distribución conjunta Normal (multivariante), entonces

$$\mathbb{E}(X_{t+h}|X_t, \dots, X_{t-p}) = \sum_{j=0}^p \alpha_j X_{t-j} = \Sigma_{XY} \Sigma_{YY}^{-1} Y$$

donde $Y = \{X_t, \dots, X_{t-p}\}$.

Por lo tanto si el proceso es Gaussiano, el mejor predictor es combinación lineal (función lineal) del pasado reciente.

Nota 5.1. Si el proceso no es Gaussiano, calcular la esperanza condicional puede ser muy complicado, así que se buscará el **mejor predictor lineal**.

Observación 5.2. En el caso Gaussiano es lineal, veamos que coinciden (el mejor predictor lineal y el mejor predictor).

5.2. Predictores Lineales

Sea $\{X_t, t \in T\}$ una serie temporal estacionaria con media $\mathbb{E}(X_t) = \mu$ y *función de autocovarianza* $\gamma_t = \gamma(t)$.

Nota 5.3. El objetivo es predecir X_{n+h} para $h = 1, 2, \dots$ basándonos en las observaciones

$$X_1, X_2, \dots, X_n$$

Buscamos el **mejor predictor lineal** $\mathbf{P}_n(\mathbf{X}_{n+h})$ en el sentido de que

1. Es combinación lineal de las observaciones

$$P_n(X_{n+h}) = a_0 + a_1 X_n + a_2 X_{n-1} + \dots + a_n X_1$$

2. Es óptimo (minimiza el *mean square error (MSE)*)

$$P_n(X_{n+h}) = \operatorname{argmin}\{\mathbb{E}((X_{n+h} - a_0 - a_1 X_n - a_2 X_{n-1} - \dots - a_n X_1)^2)\}$$

5.2.1. Encontremos el Mejor Predictor Lineal (BLP)

Lo que queremos hacer es encontrar los coeficientes a_0, \dots, a_n tal que se minimice

$$S(a_0, \dots, a_n) = \mathbb{E}((X_{n+h} - a_0 - a_1 X_n - a_2 X_{n-1} - \dots - a_n X_1)^2)$$

donde S es un polinomio de grado 2, siempre positivo por lo que tiene al menos un mínimo y puede encontrarse resolviendo el sistema de ecuaciones

$$\frac{\partial S(a_0, \dots, a_n)}{\partial a_j} \quad j = 0, \dots, n$$

Esto es,

$$\frac{\partial S(a_0, \dots, a_n)}{\partial a_0} = -2\mathbb{E}((X_{n+h} - a_0 - a_1 X_n - a_2 X_{n-1} - \dots - a_n X_1)) = 0$$

\vdots

$$\frac{\partial S(a_0, \dots, a_n)}{\partial a_j} = -2\mathbb{E}[(X_{n+h} - a_0 - a_1 X_n - a_2 X_{n-1} - \dots - a_n X_1)X_{n+1-j}] = 0$$

for $j = 1, \dots, n$.

De la primera ecuación

$$-2\frac{\partial S(a_0, \dots, a_n)}{\partial a_0} = \mathbb{E}((X_{n+h} - a_0 - a_1 X_n - a_2 X_{n-1} - \dots - a_n X_1)) = 0$$

obtenemos que

$$\mu - a_0 - a_1 \mu - \dots - a_n \mu = 0 \iff a_0 = \mu \left(1 - \sum_{i=1}^n a_i \right)$$

Para las otras ecuaciones ⁶ tenemos

$$\begin{array}{ll} \text{for } j = 1, & \gamma_h = a_1 \gamma_0 + a_2 \gamma_1 + \dots + a_n \gamma_{n-1} \\ \text{for } j = 2, & \gamma_{h+1} = a_1 \gamma_1 + a_2 \gamma_0 + \dots + a_n \gamma_{n-2} \end{array}$$

tenemos

$$\begin{bmatrix} \gamma_h \\ \gamma_{h+1} \\ \vdots \\ \gamma_{n+h-1} \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \ddots & \gamma_1 & \gamma_{n-1} \\ \vdots & \gamma_1 & \ddots & \vdots \\ \gamma_{n-1} & \cdots & \ddots & \gamma_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad (2)$$

Observación 5.4. Dividiendo por γ_0 obtenemos el sistema en términos de la ACF.

Por tanto

$$P_n(X_{n+h}) = \mu + \sum_{i=1}^n a_i (X_{n+1-i} - \mu)$$

y la esperanza del **error predicho** es

$$\mathbb{E}(P_n(X_{n+h}) - X_{n+h}) = 0$$

y su varianza es

$$E((P_n(X_{n+h}) - X_{n+h})^2) = \gamma_0 - 2 \sum_{i=1}^n a_i \gamma_{h+i-1} + \sum_{i=1}^n \sum_{j=1}^n a_i \gamma_{i-j} a_j = \gamma_0 - a^t \Gamma a \quad ^7$$

Observa que solo necesitamos los momentos de segundo orden para predecir, por lo que no es tan restrictivo trabajar solo con modelos (débilmente) estacionarios.

⁶Recuerde que para calcular el ACVF podemos olvidar el término constante

⁷Observese que Γ es la matriz de covarianzas de la Ecuación 2

Ejemplos 5.5. Encontremos el predictor de 1-paso adelante de una serie temporal AR(1),

$$X_t = \phi X_{t-1} + \epsilon_t \quad \epsilon_t \sim WN(0, \sigma^2)$$

1. Usando el mejor predictor lineal:

Ya sabemos que

$$P_n(X_{n+1}) = \sum_{i=1}^n a_i X_{n+1-i}$$

donde los coeficientes a_i son solución de

$$\begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \vdots \\ \phi^n \end{bmatrix} = \begin{bmatrix} 1 & \phi^2 & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & 1 & \phi \\ \phi^{n-2} & \vdots & \vdots & \phi & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad (3)$$

Per tant, tenim que amb

$$\begin{aligned} a_1 &= \phi \\ a_h &= 0 \quad 2 \leq h \leq n \end{aligned}$$

tenemos solución, por lo que el mejor predictor lineal de X_{n+1} basado en el pasado hasta X_n es

$$P_n(X_{n+1}) = \phi X_n$$

y tiene MSE

$$\gamma_0 - a^t(\gamma_1, \gamma_2, \dots, \gamma_n)^t = \frac{\sigma^2}{1 - \phi^2} \phi \gamma_1 = \sigma^2$$

2. Usando la esperanza condicionada:

Notamos que $\mathbf{Z} = X_t, \dots, X_1, \epsilon_t, \dots, \epsilon_1$.

$$\mathbb{E}(X_{t+1}|Z) = \mathbb{E}(X_{t+1}|X_t, \dots, X_1, \epsilon_t, \dots, \epsilon_1) = \phi \mathbb{E}(X_t|Z) + \mathbb{E}(\epsilon_{t+1}|Z)_{\text{son indep}} = \phi X_t$$

Podemos calcular también el error de predicción:

$$\begin{aligned} e_{t+1} &= X_{t+1} - \hat{X}_{t+1} \\ &= \epsilon_{t+1} \end{aligned}$$

y por tanto

$$\text{Var}(e_{t+1}, t) = \text{Var}(\epsilon_t) = \sigma^2$$

Observación 5.6. Con tal de crear predicciones, es suficiente conocer los momentos de primer y segundo orden de las series.

5.3. Distribuciones Condicionales de variables Gaussianas

Queremos ver que el mejor predictor (esperanza condicional de X_{t+h} condicionado al pasado hasta t) coincide con una función lineal del pasado, es decir, el mejor predictor es el mejor predictor lineal (solo depende de las autocovarianas.) En general, si

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim \text{Normal} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right)$$

Buscamos una transformación lineal (esperanza) \mathbf{AY} de manera que se satisfaga que $\mathbf{X} - \mathbf{AY}$ sea independiente de \mathbf{AY} . Queremos ver como es $\mathbb{E}(X|Y)$.

La descomposición

$$\mathbf{X} = \mathbf{AY} + \mathbf{X} - \mathbf{AY}$$

tendrá componentes independientes si elegimos A de manera que ⁸

$$\mathbb{E}((\mathbf{X} - \mathbf{AY})\mathbf{Y}^T \mathbf{A}^T) = 0 \iff \mathbb{E}((\mathbf{XY}^T \mathbf{A}^T)) = \mathbf{A}\mathbb{E}((\mathbf{YY}^T)^T \mathbf{A}^T) \iff \mathbf{A} = \Sigma_{\mathbf{XY}} \Sigma_{\mathbf{YY}}^{-1}$$

Por lo tanto, la ley condicional de \mathbf{X} dado \mathbf{Y} es una MVNormal con esperanza \mathbf{AY} y varianza $\mathbb{E}((\mathbf{X} - \mathbf{AY})(\mathbf{X} - \mathbf{AY})^T)$

$$\mathcal{L}_{X|Y} = \text{Normal}(\Sigma_{\mathbf{XY}} \Sigma_{\mathbf{YY}}^{-1} \mathbf{Y}, \Sigma_{\mathbf{XX}} - \Sigma_{\mathbf{XY}} \Sigma_{\mathbf{YY}}^{-1} \Sigma_{\mathbf{YX}})$$

Ejemplos 5.7. Si ponemos $X := X_{t+h}$, $Y = X_t$, entonces tenemos que

$$\Sigma_{XX} = \gamma_0 \quad \Sigma_{YY} = \gamma_0 \quad \Sigma_{XY} = \gamma_0 \rho_h$$

Por lo tanto, tenemos que

$$A = \Sigma_{XY} \Sigma_{YY}^{-1} = \frac{\gamma_0 \rho_h}{\gamma_0} = \rho_h$$

la esperanza condicional será Normal, con media (predicción)

$$\mu + \rho_h(X_t - \mu)$$

y varianza (error de predicción)

$$\gamma_0 - \gamma_0 \rho_h \frac{1}{\gamma_0} \gamma_0 \rho_h = \gamma_0(1 - \rho_h^2)$$

Así, por ejemplo, para un AR(1), $X_t = \phi X_{t-1} + \epsilon_t$, el mejor predictor de X_{t+1} dado el pasado hasta X_t es

$$\hat{X}_{t+1} = \phi X_t$$

como ya habíamos visto.

Ejemplos 5.8. Veamos otra manera de calcular el mejor predictor para un proceso AR(1) de 1-paso. Tenemos el proceso

$$X_t = \phi X_{t-1} + \epsilon_t \quad \epsilon_t \sim WN(0, \sigma^2)$$

El pasado es $Z = X_1, \dots, X_t, \epsilon_1, \dots, \epsilon_t$, por tanto

$$\mathbb{E}(X_{t+1}|Z) = \mathbb{E}(\phi X_t + \epsilon_{t+1}|Z) = \mathbb{E}(\phi X_t|Z) + \mathbb{E}(\epsilon_{t+1}|Z) = \phi X_t$$

pues $X_t \in Z$ y el error ϵ_{t+1} es independiente del pasado.

5.4. Predicción a la práctica

Las ecuaciones de estimación del BLP se suelen resolver recursivamente. En general, se escribe el ARMA (invertible) como un AR(∞), se trunca la serie, y se resuelven las ecuaciones anteriores (que en el caso de un AR son muy sencillas) usando algún tipo de recursión (una de las más populares es la de Durbin-Levinson).

⁸recordar que en el caso gausiano, independencia y ortogonalidad son equivalentes

Durbin-Levinson Algorithm: $\{X_t\}$ is a zero-mean stationary time series with

ACF $\gamma(h)$ and write

$$\hat{X}_{t+1} = P_{sp(1, X_1, \dots, X_t)} X_{t+1} = \phi_{tt} X_t + \dots + \phi_n X_1$$

Then the coefficients ϕ_{tt}, \dots, ϕ_n and prediction errors $v_{t,t}$ can be computed recursively from the equations,

$$\begin{aligned}\phi_n &= \left[\gamma(t) - \sum_{j=1}^{t-1} \phi_{t-1,j} \gamma(t-j) \right] v_{t-1}^{-1}, \\ \begin{bmatrix} \phi_{tt} \\ \vdots \\ \phi_{t,t-1} \end{bmatrix} &= \begin{bmatrix} \phi_{t-1,1} \\ \vdots \\ \phi_{t-1,t-1} \end{bmatrix} - \phi_n \begin{bmatrix} \phi_{t-1,t-1} \\ \vdots \\ \phi_{t-1,1} \end{bmatrix},\end{aligned}$$

and

$$v_t = v_{t-1}(1 - \phi_n^2).$$

5.5. Predicción para MA(1)

5.5.1. Predicción a un paso

$$\begin{aligned}\hat{X}_t(1) &= \mathbb{E}(X_{t+1}|Z_t) = \mathbb{E}(\epsilon_{t+1} + \theta\epsilon_t|X_1, \dots, X_t, \epsilon_1, \dots, \epsilon_t) \\ &= \mathbb{E}(\epsilon_{t+1}|Z_t)_{\text{indepts}} + \theta\mathbb{E}(\epsilon_t|Z_t)_{\epsilon_t} = \theta\epsilon_t = \theta(X_t - \theta\epsilon_{t-1}) = \dots = \\ &= \sum_{i=0}^{t-1} (-1)^i \theta^{i+1} X_{t-i}\end{aligned}$$

y el error de predicción es

$$X_{t+1} - \hat{X}_t(1) = \epsilon_{t+1} + \phi\epsilon_t - \theta\epsilon_t = \epsilon_{t+1}$$

que tiene esperanza 0 y varianza σ^2

Observación 5.9. Notese que estimar este con el método de BLP se complica más pues hay que resolver el sistema $t \times t$ que no tiene solución trivial.

5.5.2. Predicción a dos o más pasos

$$\begin{aligned}\hat{X}_t(2) &= \mathbb{E}(X_{t+2}|Z_t) = \mathbb{E}(\epsilon_{t+2} + \theta\epsilon_{t+1}|X_1, \dots, X_t, \epsilon_1, \dots, \epsilon_t) \\ &= \mathbb{E}(\epsilon_{t+2}|Z_t)_{\text{indepts}} + \theta\mathbb{E}(\epsilon_{t+1}|Z_t)_{\text{indepts}} \\ &= \mathbb{E}(\epsilon_{t+2}) + \theta\mathbb{E}(\epsilon_{t+1}) = 0\end{aligned}$$

Observación 5.10. Observa que para un proceso MA(q) la predicción de paso k con $k > q$ será 0.

5.5.3. Estimando el ruido para MA(1)

El problema con las predicciones anteriores es que necesitamos los ruidos que no son observables,

$$X_t = \epsilon_t + \theta\epsilon_{t-1} \quad \epsilon_t \sim WN(0, \sigma^2)$$

El pasado de la serie, hasta el instante t es X_1, \dots, X_t y los ruidos que se estiman recursivamente son

$$\epsilon_1 = X_1 - \theta\epsilon_0 \quad \text{que suponemos igual a 0}$$

$$\epsilon_2 = X_2 - \theta\epsilon_1 = X_2 - \phi X_1$$

\vdots

$$\epsilon_t = X_t - \theta\epsilon_{t-1}$$

de modo que $\mathbf{Z}_t = \{X_1, \dots, X_t, \epsilon_1, \dots, \epsilon_t\}$ y obtenemos que podemos expresar ϵ_{t+1} como hacemos en la Ecuación de la predicción a un paso.

Otra alternativa es usar que $\gamma_0 = \sigma^2(1 + \theta^2)$, $\gamma_1 = \sigma^2\theta$, $\gamma_2 = \gamma_3 = \dots = 0$ y calcular el *BLP*. Es difícil de resolver pero veamos que imponiendo la primera condición van saliendo los resultados, queremos ver que si

$$P_t(X_{t+1}) = a_1 X_t + a_2 X_{t-1} + \dots + a_n X_1 \quad (4)$$

entonces

$$a_j = (-1)^{j-1} \theta^j$$

Sabiendo que

$$\begin{bmatrix} \rho_1 = \frac{\theta}{1+\theta^2} \\ \rho_2 = 0 \\ \vdots \\ \vdots \\ \rho_t = 0 \end{bmatrix} = \begin{bmatrix} 1 & \frac{\theta}{1+\theta^2} & 0 & \cdots & 0 \\ \frac{\theta}{1+\theta^2} & 1 & \frac{\theta}{1+\theta^2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \vdots & \vdots & 1 & \frac{\theta}{1+\theta^2} \\ 0 & \phi^{n-2} & \cdots & \frac{\theta}{1+\theta^2} & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

si sustituimos se cumple la igualdad.

5.6. Predicción en un ARMA (1, 1)

$$\begin{aligned} X_t &= \phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \\ X_{t+1} &= \phi X_t + \epsilon_{t+1} + \theta \epsilon_t \\ X_{t+2} &= \phi X_{t+1} + \epsilon_{t+2} + \theta \epsilon_{t+1} \\ &= \phi(\phi X_t + \epsilon_{t+1} + \theta \epsilon_t) + \theta \epsilon_{t+1} + \epsilon_{t+2} \\ &= \phi^2 X_t + (\phi + \theta) \epsilon_{t+1} + \phi \theta \epsilon_t + \epsilon_{t+2} \end{aligned}$$

5.6.1. Predicciones

Con $Z_t = \{X_1, \dots, X_t, \epsilon_1, \dots, \epsilon_t\}$.

$$\begin{aligned} \hat{X}_t(1) &= \mathbf{E}(\phi X_t + \epsilon_{t+1} + \theta \epsilon_t | Z_t) = \phi X_t + \theta \epsilon_t \\ \hat{X}_t(2) &= \mathbf{E}(\phi^2 X_t + (\phi + \theta) \epsilon_{t+1} + \phi \theta \epsilon_t + \epsilon_{t+2} | Z_t) = \phi^2 X_t + \phi \theta \epsilon_t \end{aligned}$$

Otra vez haremos "back-substitution" para estimar ϵ_t , como en el ejemplo del MA(1). Suponiendo que $X_0 = \epsilon_0 = 0$, tenemos

$$\begin{aligned} \epsilon_1 &= X_1 \\ \epsilon_2 &= X_2 - \phi X_1 - \theta \epsilon_1 = X_2 - (\phi + \theta) X_1 \\ \epsilon_3 &= X_3 - \phi X_2 - \theta \epsilon_2 = X_3 - \phi X_2 - \theta(X_2 - (\phi + \theta) X_1) \\ &= X_3 - (\phi + \theta) X_2 + \theta(\phi + \theta) X_1 \end{aligned}$$

5.7. Otra forma de estimar en el ARMA (1, 1)

Si el ARMA(1, 1) es invertible, se puede usar su expresión como un AR(∞) para hacer predicciones en forma recursiva.

$$\frac{1 - \phi z}{1 + \theta z} = (1 - \phi z) \sum_{j=0}^{\infty} (-1)^j \theta^j z^j = 1 - (\phi + \theta)(z - \theta z^2 + \theta^2 z^3 - \dots)$$

por lo tanto $X_t = \epsilon_t + (\phi + \theta)X_{t-1} - \theta(\phi + \theta)X_{t-2} + \theta^2(\phi + \theta)X_{t-3} - \dots$ y con esto podemos predecir (truncando la serie)

$$\hat{X}_t(1) = (\hat{\phi} + \hat{\theta}) \sum_{j=1}^{\infty} (-\hat{\theta})^{j-1} X_{t+1-j}$$

5.8. El *Innovation Algorithm*

Este algoritmo sirve para estimar y predecir en cualquier serie de tiempo con segundo momento finito, sea o no estacionaria, pero en casos particulares (**MA**, **ARMA**) resulta sencillo.

5.8.1. Algoritmo

1. Llamamos $\mathbb{E}(X_i X_j) = \kappa(i, j)$ al núcleo de la covarianza, \hat{X}_n al mejor predictor a un paso basado en el pasado hasta X_{n-1} , y $\nu_n = \mathbb{E}(X_n - \hat{X}_n)^2$.
2. Introducimos las *innovaciones* (errores de predicción a un paso):

$$U_n = X_n - \hat{X}_n$$

Observación 5.11. Se puede ver que si el proceso es estacionario, las predicciones a un paso se pueden escribir como combinaciones lineales de U_n , es decir

$$\hat{X}_{n+1} = \sum_{j=1}^n \theta_{n,j} (X_{n+1-j} - \hat{X}_{n+1-j}) \quad n = 1, 2, \dots$$

entonces una vez determinados los coeficientes, podemos calcular las predicciones recursivamente (ver Figura 39⁹).

The Innovations Algorithm:

The coefficients $\theta_{n1}, \dots, \theta_{nn}$ can be computed recursively from the equations

$$v_0 = \kappa(1, 1),$$

$$\theta_{n,n-k} = v_k^{-1} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} v_j \right), \quad 0 \leq k < n,$$

and

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j.$$

(It is a trivial matter to solve first for v_0 , then successively for $\theta_{11}, v_1; \theta_{22}, \theta_{21}, v_2; \theta_{33}, \theta_{32}, \theta_{31}, v_3; \dots$)

Figura 39: Calculo de los coeficientes con el *Innovation Algorithm*

⁹Ver detalles en el ITSM de Brockwell y Davies, capítulos 2 y 5

5.9. Ejercicios

- Verificar que si $X_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$, entonces $\hat{X}_t(1) = \mu + \theta\epsilon_t$

Si $\mathbf{Z} = \{X_t, X_{t-1}, \dots, X_1, \epsilon_t, \dots, \epsilon_1\}$. Entonces

$$\hat{X}_t(1) = \mathbb{E}(X_{t+1}|Z) = \mu + \theta\epsilon_t$$

como queríamos.

- Usar las ecuaciones del BLP para encontrar el predictor $\hat{X}_t(1)$.

- Calcular el BLP para el MA(1).

Ver razonamiento de las ecuaciones 4.

- Escribir la predicción a 1, 2 y 3 pasos y la varianza del error para un MA(2).

Sea el modelo

$$X_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$$

y $\mathbf{Z} = \{X_t, X_{t-1}, \dots, X_1, \epsilon_t, \dots, \epsilon_1\}$, entonces

$$\hat{X}_t(1) = \mathbb{E}(X_{t+1}|Z) = \theta_1\epsilon_t + \theta_2\epsilon_{t-1}$$

$$\text{Var}(X_{t+1} - \hat{X}_t(1)) = \sigma^2$$

$$\hat{X}_t(2) = \mathbb{E}(X_{t+2}|Z) = \theta_2\epsilon_t$$

$$\text{Var}(X_{t+2} - \hat{X}_t(2)) = \text{Var}(\epsilon_{t+2} + \theta_1\epsilon_{t+1}) = \sigma^2(1 + \theta_1^2)$$

$$\hat{X}_t(3) = \mathbb{E}(X_{t+3}|Z) = 0$$

$$\text{Var}(X_{t+3} - \hat{X}_t(3)) = \text{Var}(X_{t+3}) = \sigma^2(1 + \theta_1^2 + \theta_2^2)$$

- Observar que para un MA(q) sólo podemos hacer predicciones a $1, 2, \dots, q$ pasos.

Sea el modelo

$$X_t = \epsilon_t + \theta_1\epsilon_{t-1} + \dots + \theta_{t-q}\epsilon_{t-q}$$

y $\mathbf{Z} = \{X_t, \dots, X_{t-q}, \epsilon_t, \epsilon_{t-q}\}$. Entonces

$$\hat{X}_t(1) = \mathbb{E}(X_{t+1}|Z) = \mathbb{E}(\epsilon_{t+1} + \theta_1\epsilon_t + \dots + \theta_q\epsilon_{t+1-q}|Z) = \theta_1\epsilon_t + \dots + \theta_q\epsilon_{t+1-q}$$

$$\hat{X}_t(2) = \mathbb{E}(X_{t+2}|Z) = \mathbb{E}(\epsilon_{t+2} + \theta_1\epsilon_{t+1} + \dots + \theta_q\epsilon_{t+2-q}|Z) = \theta_2\epsilon_t + \dots + \theta_q\epsilon_{t+2-q}$$

\vdots

$$\hat{X}_t(q) = \mathbb{E}(X_{t+q}|Z) = \mathbb{E}(\epsilon_{t+q} + \theta_1\epsilon_{t+q-1} + \dots + \theta_q\epsilon_t|Z) = \theta_q\epsilon_t$$

$$\hat{X}_t(j) = \mathbb{E}(X_{t+j}|Z) = \mathbb{E}(\epsilon_{t+j} + \theta_1\epsilon_{t+j-1} + \dots + \theta_q\epsilon_{t+j-q}) = 0 \quad j > q$$

- Escribir las predicciones a 1, 2, 3 pasos para un AR(2).

- Calcular la varianza de los errores de predicción anteriores.

En primer lugar escribimos las ecuaciones

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \\ X_{t+1} &= \phi_1 X_t + \phi_2 X_{t-1} + \epsilon_{t+1} \\ X_{t+2} &= (\phi_1^2 + \phi_2) X_t + \phi_1 \phi_2 X_{t-1} + \epsilon_{t+2} + \phi_1 \epsilon_{t+1} \\ X_{t+3} &= \phi_1 (\phi_1^2 + 2\phi_2) X_t + \phi_2 (\phi_1^2 + \phi_2) X_{t-2} + \epsilon_{t+3} + \phi_1 \epsilon_{t+2} (\phi_1^2 + \phi_2) \epsilon_{t+1} \end{aligned}$$

Con esto y sabiendo que $\mathbf{Z} = \{X_t, \dots, X_1, \epsilon_t, \dots, \epsilon_1\}$ entonces

$$\begin{aligned} \hat{X}_t(1) &= \phi_1 X_t + \phi_2 X_{t-1} \\ \hat{X}_t(2) &= (\phi_1^2 + \phi_2) X_t + \phi_1 \phi_2 X_{t-1} \\ \hat{X}_t(3) &= \phi_1 (\phi_1^2 + 2\phi_2) X_t + \phi_2 (\phi_1^2 + \phi_2) X_{t-2} \end{aligned}$$

y tenemos que los errores son

$$\begin{aligned} \hat{\epsilon}_t(1) &= \epsilon_{t+1} \\ \hat{\epsilon}_t(2) &= \epsilon_{t+2} + \phi_1 \epsilon_{t+1} \\ \hat{\epsilon}_t(3) &= \epsilon_{t+3} + \phi_1 \epsilon_{t+2} (\phi_1^2 + \phi_2) \epsilon_{t+1} \end{aligned}$$

con varianzas

$$\begin{aligned} \text{Var}(\hat{\epsilon}_t(1)) &= \sigma^2 \\ \text{Var}(\hat{\epsilon}_t(2)) &= \sigma^2(1 + \phi_1^2) \\ \text{Var}(\hat{\epsilon}_t(3)) &= \sigma^2(1 + \phi_1^2 + (\phi_1^2 + \phi_2)^2) \end{aligned}$$

8. Supongamos que $\phi = \theta = 1/2$, y $X_1 = 3, X_2 = 4$. Predecir los valores de X_3 y X_4 . simplemente sustituir en las ecuaciones anteriores. Da (suponiendo $\epsilon_0 = 0$)

$$\begin{aligned} \epsilon_1 &= X_1 &= 3 \\ \epsilon_2 &= X_2 - \theta \epsilon_1 &= \frac{5}{2} \\ \hat{X}_3 &= \mathbb{E}(X_3 | \{X_1, X_2, \epsilon_1, \epsilon_2\}) &= \frac{13}{5} \\ \hat{X}_4 &= \mathbb{E}\left(\frac{1}{4}X_2 + \epsilon_4 + \epsilon_3 + \frac{1}{2}\epsilon_2\right) | \{X_1, X_2, \epsilon_1, \epsilon_2\} &= \frac{13}{8} \end{aligned}$$

6. ESTIMACIÓN DE PARÁMETROS EN MODELOS ARMA

6.1. Las ecuaciones de Yule-Walker

Ya sabemos que para un proceso AR(p) con media cero

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t \quad \epsilon_t \sim WN(0, \sigma^2)$$

las primeras p ecuaciones de Yule-Walker son

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p} \quad k = 1, 2, \dots, p$$

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & \ddots & \rho_1 & \rho_{p-2} \\ \vdots & \rho_1 & \ddots & \vdots \\ \rho_{p-1} & \dots & \ddots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

proporcionan estimadores para ϕ_1, \dots, ϕ_p .

Ejemplos 6.1. AR(1) y AR(2)

Para un AR(1), tenemos que $\rho_1 = \phi$ por lo que estimamos ϕ mediante

$$\hat{\phi} = \hat{\rho}_1 = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-1} (x_{t+1} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{n-1} (x_t - \bar{x})^2}$$

Para un AR(2),

$$\begin{aligned} \rho_1 &= \phi_1 + \rho_1 \phi_2 \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \end{aligned}$$

por lo tanto

$$\hat{\phi}_1 = \frac{\hat{\rho}_1(1 - \hat{\rho}_2)}{1 - \hat{\rho}_1^2} \quad \frac{\hat{\rho}_2^2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2}$$

6.2. MAs y ARMAs

Las ecuaciones de Yule-Walker son casi inútiles para estimar parámetros en un proceso MA (y por tanto en un proceso ARMA).

Ejemplos 6.2. Ecuaciones de Yule-Walker para un proceso MA(1)

Recordemos que

$$\begin{aligned} \gamma_0 &= \sigma^2(1 + \theta^2) \\ \gamma_1 &= \theta\sigma^2 \end{aligned}$$

Por lo tanto,

$$\rho_1 = \frac{\theta}{1 + \theta^2}$$

por lo que tenemos una ecuación cuadrática en θ con dos raíces:

$$\theta = \frac{1 \pm \sqrt{1 - 4\rho_1^2}}{2\rho_1}$$

- El producto de las 2 raíces es 1, por lo que sólo una de ellas satisface la condición de invertibilidad.

- Para tener una raíz real, $1 - 4\rho_1^2 > 0$.
- Aunque $|\rho_1| < \frac{1}{2}$, podría ser que $|\hat{\rho}_1| > \frac{1}{2}$ y no podríamos estimar θ .
- Si $|\rho_1| = \frac{1}{2}$, habrá una raíz real, pero el proceso no es invertible.

Ejemplos 6.3. Modelo ARMA(1,1)

Para un proceso ARMA(1,1) la situación es aún peor. Si consideramos un modelo con:

$$\rho_k = \frac{(1 + \phi\theta)(\phi + \theta)}{1 + 2\phi\theta + \theta^2} \phi^{k-1} \quad k \geq 1$$

implica que

$$\frac{\rho_2}{\rho_1} = \theta \quad \text{por lo que} \quad \hat{\phi} = \frac{\hat{\rho}_2}{\hat{\rho}_1}$$

y reemplazando ϕ por $\hat{\phi}$ tenemos de nuevo una ecuación cuadrática para θ que podría no tener solución real

$$\rho_1 = \frac{(1 + \hat{\phi}\theta)(\hat{\phi} + \theta)}{1 + 2\hat{\phi}\theta + \theta^2} \phi^{k-1}$$

e incluso teniendo soluciones reales, solo una de ellas satisface la condición de invertibilidad.

6.3. La varianza del ruido

La varianza marginal γ_0 de la serie se puede estimar mediante

$$\hat{\gamma}_0 = s^2 = \frac{1}{n-1} \sum_{t=1}^n (X_t - \bar{X})^2$$

y a partir de esta estimación, la varianza del ruido σ_ϵ^2 se puede estimar teniendo en cuenta la relación entre γ_0 y σ_ϵ^2 (que puede ser diferente para cada modelo).

$$\begin{aligned} \sigma_\epsilon^2 &= \gamma_0(1 - \phi_1\rho_1 - \phi_2\rho_2 - \dots - \phi_p\rho_p) && \text{en un AR(p)} \\ \sigma_\epsilon^2 &= \frac{\gamma_0}{1 - \sum_{i=1}^q \theta_i^2} && \text{en un MA(q)} \end{aligned}$$

6.4. Estimación por mínimos cuadrados

Los modelos AR(p) son regresión lineal de la serie con respecto a su propio lagged pasado, por lo que encontrar LSE (*Least Squares Estimation*) para los coeficientes ϕ_j y la media μ es directo.

Ejemplos 6.4. AR(1): $X_t - \mu = \phi(X_{t-1} - \mu) + \epsilon_t$

Necesitamos encontrar μ y θ tal que minimicen la suma de las diferencias al cuadrado.

Nos fijamos en la *suma condicional de cuadrados* $S_c(\mu, \phi)$

$$S_c(\mu, \phi) := \sum_{t=2}^n ((X_t - \mu) - \phi(X_{t-1} - \mu))^2$$

EL mínimo, puede encontrarse explícitamente resolviendo

$$\frac{\partial S_c(\mu, \phi)}{\partial \mu} = 0 \quad \frac{\partial S_c(\mu, \phi)}{\partial \phi} = 0$$

$$\begin{aligned}\frac{\partial S_c(\mu, \phi)}{\partial \mu} &= 2(\phi - 1) \left(\sum_{t=2}^n X_t - \phi \sum_{t=2}^n X_{t-1} \right) - (n-1)\mu(1-\phi) = 0 \\ \Rightarrow \hat{\mu} &= \frac{1}{(n-1)(1-\phi)} \left(\sum_{t=2}^n X_t - \phi \sum_{t=2}^n X_{t-1} \right)\end{aligned}$$

Observe que ambos $\sum_{t=2}^n X_t/(n-1)$ y $\sum_{t=2}^n X_{t-1}/(n-1)$ son aproximadamente \bar{X} , por lo que, independientemente del valor de ϕ ,

$$\begin{aligned}\hat{\mu}(1-\phi) &\approx \bar{X} - \phi\bar{X} = (1-\phi)\bar{X} \\ \Rightarrow \hat{\mu} &\approx \bar{X}\end{aligned}$$

Por otro lado, tenemos que

$$\begin{aligned}\frac{\partial S_c(\mu, \phi)}{\partial \phi} &= 2 \left(\sum_{t=2}^n (X_t - \bar{X}) - \phi(X_{t-1} - \bar{X}) \right) (X_{t-1} - \bar{X}) = 0 \\ \Rightarrow \hat{\phi} &= \frac{\sum_{t=2}^n (X_t - \bar{X})(X_{t-1} - \bar{X})}{\sum_{t=2}^n (X_{t-1} - \bar{X})^2} \approx \hat{\rho}_1\end{aligned}$$

excepto que la suma en el denominador de $\hat{\phi}_1$ tiene la suma de $t=1$, de modo que, para n grande, este es aproximadamente el mismo resultado obtenido de las ecuaciones YW.

La situación para un AR(p) es similar, pero estas sumas empiezan por $t=p+1$.

Con tal de calcular LSE para procesos ARMA invertibles, utilizaos la represenación AR(∞) (los ϵ_t no son observables).

Ejemplos 6.5. Modelo MA(1)

$$X_t = (1 + \theta B)\epsilon_t$$

Se puede escribir como un AR(∞) como

$$\epsilon_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \dots$$

entonces

$$S_c(\theta) = \sum_{t=1}^n \epsilon_t^2 = \sum_{t=1}^{\infty} (X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \dots)^2$$

Pero esto presenta varios problemas:

- No es lineal.
- El lado derecho es una serie (suma infinita).
- El lado izquierdo no es observable.

La minimización se realiza mediante la búsqueda en red, elige una red de valores de θ , calcula S_c y elige el valor que produce el mínimo.

Para estimar la secuencia de ruido t para θ fijo, supongamos que

$$\epsilon_0 = \mathbb{E}(\epsilon_t) = 0$$

Entonces

$$\begin{aligned}\epsilon_1 &= X_1 \\ \epsilon_2 &= X_2 + \theta\epsilon_1 = X_2 + \theta X_1 \\ \epsilon_3 &= X_3 + \theta\epsilon_2 = X_3 + \theta X_2 + \theta^2 X_1 \\ &\vdots \\ \epsilon_n &= X_n + \theta\epsilon_{n-1} = \dots\end{aligned}$$

por lo que, condicionando a $\epsilon_0 = 0$, para cada valor θ fijo, podemos calcular $\sum_{t=1}^n$ y mediante búsqueda en red, escogemos $\hat{\theta}$

6.5. Estimación de máxima verosimilitud

Suponiendo un modelo paramétrico para el ruido, tenemos un modelo paramétrico para las densidades de las observaciones X_t .

Consideramos un proceso ARMA con media cero y ruidos *Gaussianos* $\epsilon_t \sim N(0, \sigma^2)$ y parámetros $\phi = (\phi_1, \dots, \phi_p)$ y $\theta = (\theta_1, \dots, \theta_q)$, dadas las observaciones $\mathbf{x} = (x_1, \dots, x_n)$, la verosimilitud puede escribirse como

$$L(\phi, \theta, \sigma^2, \mathbf{x}) = \frac{\exp\{-x' \Gamma_n^{-1} \frac{\mathbf{x}}{2\sigma^2}\}}{|\Gamma_n|^{\frac{1}{2}} (2\pi\sigma^2)^{\frac{n}{2}}}$$

donde $\sigma^2 \Gamma_n$ es la matriz de covarianza de las observaciones y solo depende de ϕ y θ .

Esta covarianza debe escribirse de manera inteligente para facilitar la minimización de la log-verosimilitud, lo cual es difícil y requiere un esfuerzo computacional.

Ejemplos 6.6. Modelo AR(1)

En este caso, ya que $\gamma(h) = \sigma^2 \phi^h \frac{1}{1-\phi^2}$, la matriz *Gamma*_n es

$$\Gamma_n = \frac{1}{1-\phi^2} \begin{bmatrix} 1 & \phi & \dots & \phi^{n-1} \\ \phi & \ddots & \dots & \phi n - 2 \\ \vdots & \phi & \ddots & \vdots \\ \phi^{n-1} & \dots & \ddots & 1 \end{bmatrix}$$

El determinante $|\Gamma_n| = \frac{1}{1-\phi^2}$ y

$$\Gamma_n^{-1} = \begin{bmatrix} 1 & -\phi & 0 & 0 & \dots & \phi^{n-1} \\ -\phi & 1 + \phi^2 & -\phi & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & 0 \\ 0 & \dots & \dots & -\phi & 1 + \phi^2 & -\phi \\ 0 & 0 & 0 & \dots & -\phi & 1 \end{bmatrix}$$

Otra derivación para las ecuaciones MLE para un AR(1)

$$X_t - \mu = \phi(X_{t-1} - \mu) + \epsilon_t$$

La verosimilitud puede escribirse como

$$L(X_1, \dots, x_n) = f(x_2, \dots, x_n | x_1) f(x_1)$$

Para calcular la densidad dde (x_2, \dots, x_n) dado x_1 , observamos que la densidad conjunta de la secuencia de ruidos $\epsilon_2, \dots, \epsilon_n$ es

$$\frac{1}{(\sqrt{2\pi\sigma^2})^{n-1}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=2}^n \epsilon_t^2\right\}$$

Condicionando sobre x_1 , ya que la relación entre x_2, \dots, x_n y el ruido $\epsilon_2, \dots, \epsilon_n$ es lineal y el Jacobiano es 1, la densidad condicional conjunta es

$$f(x_2, \dots, x_n | x_1) = (2\pi\sigma^2)^{\frac{n-1}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=2}^n ((x_t - \mu) - \phi(x_{t-1} - \mu))^2\right\}$$

Con tal de calcular $f(x_1)$, recordemos que $\mathbb{E}(X_1) = \mu$ y $\text{Var}(X_1) = \frac{\sigma^2}{1-\phi^2}$ por lo que

$$L(\phi, \mu, \sigma^2; x_1, \dots, x_n) = (2\pi\sigma^2)^{\frac{n}{2}}(1-\phi^2)\exp\left\{-\frac{1}{2\sigma^2}S(\phi, \mu)\right\}$$

donde

$$S(\phi, \mu) = \sum_{t=2}^n ((x_t - \mu) - \phi(x_{t-1} - \mu))^2 + (1 - \phi^2)(x_1 - \mu)^2$$

Por lo tanto

$$l = \log(L) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}S(\phi, \mu)$$

De aquí, tenemos que

$$\hat{\sigma}^2 = \frac{S(\phi, \mu)}{n}$$

aunque a veces se utiliza denominador $n - 2$ para obtener un estimador insesgado.

El máximo de ϕ y μ se obtiene mediante algoritmos iterativos.

Observe que la suma incondicional de cuadrados $S(\phi, \mu)$ está relacionada con la suma condicional de cuadrados por

$$S(\phi, \mu) = S_c(\phi, \mu) + (1 + \phi^2)(x_1 - \mu)^2$$

El problema de estimación MLE está directamente relacionado con el problema de pronóstico y puede resolverse con algoritmos iterativos:

- Algoritmo de *Durbin-Levinson*.
- Algoritmo de Innovaciones. ¹⁰

6.6. Procedimiento de estimación simple para ARMA(p,q)

- Asumamos que $X_t \sim \text{ARMA}(p, q)$ estacionario, causal e invertible:

$$X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

donde $\epsilon_t \sim \text{WN}(0, \sigma^2)$, $\Phi(z)$ y $\Theta(z)$ no tienen raíces comunes y $\phi_p, \theta_q \neq 0$.

- Problema con el enfoque OLS (Ordinary Least Squares): X_t depende no solo de p si no también de los q términos de error pasado que no son observables.

Observación 6.7. La idea es reemplazar $\epsilon_t, \epsilon_{t-1}, \dots$ por alguna estimación.

- Para los ARMA(p,q) tenemos la representación como AR(∞). Dividimos el procedimiento en dos etapas:

1. Paso 1: Aproximar el ARMA(p,q) como un proceso AR(r) para un r de gran orden ($r \geq \max(p, q)$). Estimar este AR(r) usando el enfoque OLS. Las estimaciones de $\epsilon_1, \dots, \epsilon_T$ son los residuos OLS, es decir, $\hat{\epsilon}_1, \dots, \hat{\epsilon}_T$
2. Paso 2: Estimar los parámetros del modelo mediante OLS aplicados a:

$$X_t = c + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \hat{\epsilon}_t + \theta_1 \hat{\epsilon}_{t-1} + \dots + \theta_q \hat{\epsilon}_{t-q}$$

- Estimadores consistentes de $c, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \dots, \theta_q$
- Posibles generalizaciones: Inclusión de variables ficticias estacionales y/o tendencias temporales lineales (en ambos pasos)

¹⁰Mirar secciones 2.5.1 y 2.5.2 y capítulo 5 de Brockwell y Davies ITSF o capítulo 7 de Cryer y Chan

6.6.1. Métodos basados en la verosimilitud

- Necesitamos una suposición extra: $\epsilon_t \sim \text{GWN}(0, \sigma^2)$ ¹¹.
- La verosimilitud como para el caso AR, pero ahora Γ_n depende también de $\theta_1, \dots, \theta_q$:
- X_t tiene representación MA(∞):

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

donde

1. $\mu = \frac{c}{\phi(1)}$
2. $\Psi(z) = \Phi^{-1}(z)\Theta(z)$
3. $\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \phi_{j+h}$ con $h = 0, 1, \dots$

- $\gamma(h)$ es una función de σ^2 y $\xi = (\phi_1, \dots, \phi_p, \Theta_1, \dots, \Theta_q)^T$:
 - $\Gamma_n = V(x) = ((\gamma(i-j)))_{1 \leq i,j \leq n} =: \sigma^2 \Omega_T$
 - $\det(\Gamma_n) = \sigma^2 \det(\Omega_T)$

La función de verosimilitud con $\beta = (\mu, \xi^T)^T$:

$$\begin{aligned} I(\mu, \xi, \sigma^2 | x) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) \\ &\quad - \frac{1}{2} \log(\det \Omega_n) - \frac{1}{2\sigma^2} (x - \nu)^T \Omega_n^{-1} (x - \nu) \end{aligned}$$

- El MLE de $\zeta = (\mu, \xi, \sigma^2)^T$ y $\hat{\zeta} = \text{argmax}_{\zeta} I(\zeta | x)$
- Es un problema de optimización no lineal, requiere métodos numéricos.
- Para un proceso ARMA estacionario, causal e invertible, sin raíces comunes en $\Phi(z)$ y $\Theta(z)$, el MLE de los parámetros es consistente, asintóticamente normal distribuido y asintóticamente eficiente.
- **MLE condicional:** Maximiza la función de verosimilitud condicional dada una muestra de valores previos de X_t, ϵ_t .
 - Si el proceso ARMA incluye términos MA, esta maximización es un problema no lineal y requiere métodos numéricos con valores iniciales obtenidos por ejemplo mediante el procedimiento simple basado en OLS.
 - Bajo algunas condiciones, el estimador de OLS de dos pasos, el MLE exacto y MLE condicional son asintóticamente equivalentes.
- Para datos no-Gaussianos, el quasi-ML se puede aplicar como el caso AR descrito.
- Para un modelo MA(q) puro, se pueden estimar los parámetros estableciendo la parte AR en 0.

¹¹Gaussian white noise

6.7. Propiedades distributivas de los estimadores

En el caso general de los modelos ARMA(p, q) causales e invertibles, la estimación de máxima verosimilitud y la estimación de mínimos cuadrados condicional e incondicional (y la estimación de Yule-Walker en el caso de los modelos AR) conducen a estimadores óptimos¹².

Denote los parámetros de los coeficientes del ARMA por $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^t$.

Los estimadores de σ_ϵ^2 son consistentes y la distribución asintótica de $\hat{\beta}$ es asintóticamente normal, en particular cuando $n \rightarrow \infty$

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow_d N(0, \sigma_\epsilon^2 \Gamma_{-1,p,q})$$

La matriz asintótica de varianza-covarianza del estimador $\hat{\beta}$ es la inversa de la matriz de información.

6.8. Matriz de información

En particular, la matriz $(p+q)(p+q)$, $\Gamma_{p,q}$ tiene la forma

$$\Gamma_{p,q} = \begin{bmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta} \end{bmatrix}$$

donde

- $\Gamma_{\phi\phi}$ es la matriz de autocovarianzas de un proceso AR(p) con $\Phi(B)X_t = \epsilon_t$
- $\Gamma_{\theta\theta}$ es la matriz de autocovarianzas de un proceso AR(q) con $\Phi(B)Y_t = \epsilon_t$
- Las matrices $\Gamma_{\phi\theta}$ y $\Gamma_{\theta\phi}$ son las matrices de covarianza cruzada de los dos procesos AR.

Ejemplos 6.8. Modelo AR(p) La matriz de covarianza asintótica en este caso es la misma que para las estimaciones de Yule-Walker dadas por

$$V(\phi) = \sigma^2 \Gamma_{-1,p}$$

En los casos especiales de $p = 1$ y $p = 2$, tenemos

$$\begin{aligned} AR(1) : V(\phi) &= (1 - \phi_1^2) \\ AR(2) : V(\phi) &= \begin{bmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{bmatrix} \end{aligned}$$

Ejemplos 6.9. Modelo MA(q)

Let Γ_q^* be the covariance matrix of Y_1, \dots, Y_q , where $\{Y_t\}$ is the autoregressive process with autoregressive polynomial $\theta(z)$, i.e.,

$$Y_t + \theta_1 Y_{t-1} + \dots + \theta_q Y_{t-q} = Z_t, \quad \{Z_t\} \sim WN(0, 1).$$

Then it can be shown that

$$V(\theta) = \Gamma_q^{*-1}.$$

Inspection of the results of Example 5.2.1 and replacement of ϕ_i by $-\theta_i$ yields

$$\begin{aligned} MA(1) : V(\theta) &= (1 - \theta_1^2), \\ MA(2) : V(\theta) &= \begin{bmatrix} 1 - \theta_2^2 & \theta_1(1 - \theta_2) \\ \theta_1(1 - \theta_2) & 1 - \theta_2^2 \end{bmatrix}. \end{aligned} \quad \square$$

Ejemplos 6.10. Modelo ARMA(1,1)

For a causal and invertible ARMA(1,1) process with coefficients ϕ and θ .

$$V(\phi, \theta) = \frac{1 + \phi\theta}{(\phi + \theta)^2} \begin{bmatrix} (1 - \phi^2)(1 + \phi\theta) & -(1 - \theta^2)(1 - \phi^2) \\ -(1 - \theta^2)(1 - \phi^2) & (1 - \theta^2)(1 + \phi\theta) \end{bmatrix}$$

¹²La demostración del resultado general está en el Brockwell y Davis

7. MODELOS para PROCESOS NO ESTACIONARIOS (ARIMA)

7.1. Procesos no estacionarios

Definición: Proceso No Estacionario

Un proceso puede ser **no estacionario** porque la media, la varianza, las autocorrelaciones u otras características de la distribución de variables cambian con el tiempo.

- Cuando el nivel de la serie no es estable en el tiempo, en particular mostrando tendencias crecientes o decrecientes, decimos que *la serie no es estable en la media*.
 - Cuando la variabilidad o autocorrelaciones cambian con el tiempo, decimos que *la serie no es estacionaria en la varianza o autocovarianza*.
 - Finalmente, si la distribución de la variable en cada momento varía con el tiempo, decimos que *la serie no tiene una distribución estacionaria*.
1. Una clase de procesos no estacionarios fáciles de tratar son los **procesos integrados**, los cuales tienen la propiedad de que al diferenciarlos obtenemos procesos estacionarios.

Por lo general, la **diferenciación elimina tendencias**, es decir, es útil para la no estacionariedad de la media. Las estacionalidades también pueden eliminarse tomando las diferencias adecuadas.

7.2. Procesos Integrados

Una propiedad importante que distingue los **procesos integrados** de los estacionarios es la forma en que la dependencia desaparece con el tiempo.

- En los procesos estacionarios ARMA las autocorrelaciones disminuyen geométricamente, y prácticamente llegan a cero en pocos *lags*.
- En los procesos integrados, las autocorrelaciones disminuyen linealmente con el tiempo y es posible encontrar coeficientes de autocorrelación diferentes de cero incluso para *lags* muy altos.
- Además de su interés teórico, estos procesos pueden逼近 mucho al comportamiento observado en series temporales climatológicas o financieras.
- Hay una clase de procesos estacionarios donde las autocorrelaciones decaen mucho más lentamente con el tiempo que en el caso de los procesos ARMA o en los procesos integrados. Estos se conocen como procesos de memoria larga.

Ejemplos 7.1. A continuación podemos un ejemplo de serie no estacionaria (Figura 40) y la serie diferenciada (Figura 41)

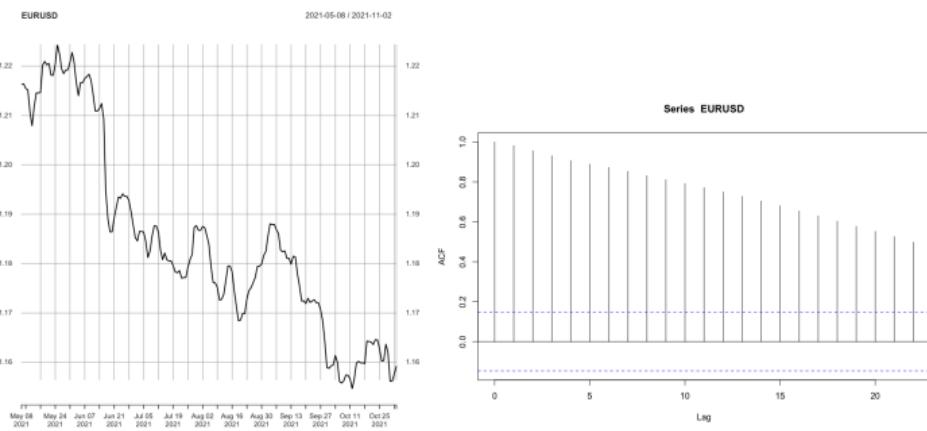


Figura 40: EUR/USD exchange rate

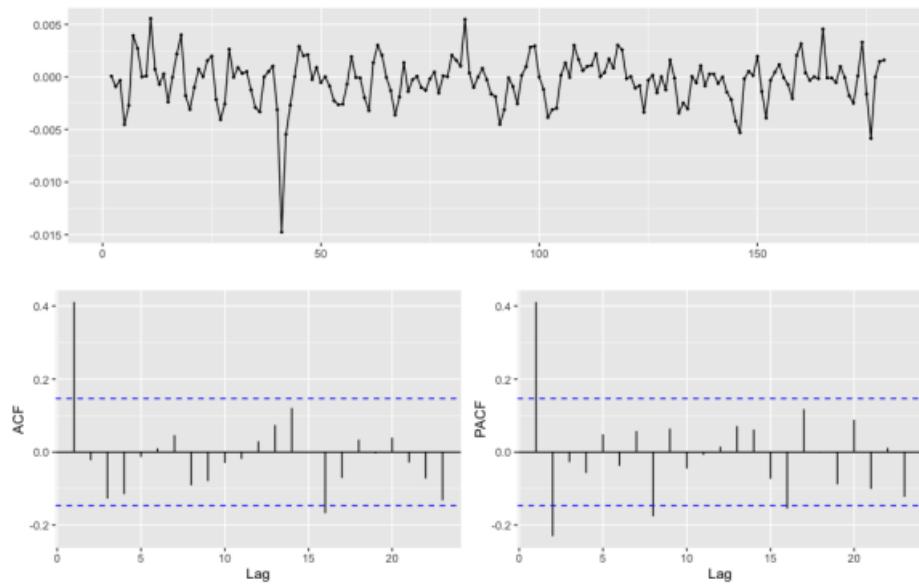


Figura 41: 1^a diferencia de EUR/USD exchange rate

Propiedades Procesos Integrados

- A veces es necesario diferenciar más de una vez para obtener un proceso estacionario.

Ejemplos 7.2. Por ejemplo, si X_t denota la serie original, y sus variaciones (o crecimiento) medidas por $Y_t = \nabla X_t = X_t - X_{t-1}$ no son estacionarias, pero las variaciones de crecimiento, medidas por $\nabla^2 X_t$, sí lo son. Eso es el serie:

$$Z_t = \nabla Y_t = \nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2}$$

que representa el crecimiento de la serie Y_t es estacionaria.

- La serie Z_t se llama segunda diferencia de la serie original, X_t , y decimos que X_t es integrada de orden 2, o con incrementos estacionarios de segundo orden.
- Generalizando, decimos que un proceso es integrado de orden $d > 0$, si al diferenciarlo d veces se obtiene un proceso estacionario.
- Por lo tanto, un proceso estacionario es siempre $I(0)$, i.e, integrado de orden 0.

ARIMA(p;d;q): the Box-Jenkins family (Casos)

Cuando la serie X_t es no estacionaria (por ejemplo, debido a que tiene tendencia), a veces, tomar diferencias produce una secuencia estacionaria.

- Si la tendencia es *lineal*, sabemos que $Y_t = X_t - X_{t-1}$ no tiene tendencia, y con suerte, será un proceso estacionario. Si ese es el caso, podemos intentar modelar Y_t con ARMA(p, q).

Observación 7.3. En este caso, X_t es un ARIMA(p, 1, q).

- Si tenemos que diferenciar la serie d veces (por ejemplo para eliminar una tendencia polinomial), tenemos que el modelo será ARIMA(p, d, q)
- Si diferenciamos para eliminar la estacionalidad, tenemos un modelo SARIMA.

Observación 7.4. Estos procesos son muy flexibles, pero normalmente su identificación es muy empírica.

7.3. Paseo Aleatorio

El paseo aleatorio

X_t es un **paseo aleatorio** si puede representarse como

$$X_t = c + X_{t-1} + \epsilon_t \quad \epsilon_t \sim WN(0, \sigma^2)$$

Observación 7.5. Si $c \neq 0$, decimos que es un **paseo aleatorio con drift** c

Observación 7.6. Observe que esto es un AR(1) con raíz unidad. Escribimos por tanto

$$\begin{aligned} X_t &= c + X_{t-1} + \epsilon_t \\ &= 2c + X_{t-2} + \epsilon_{t-1} + \epsilon_t \\ &= \dots \\ &= ct + \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \dots + \epsilon_1 \end{aligned}$$

es facil ver que, si $X_0 = 0$, $\mathbb{E}(X_t) = ct$, $\text{Var}(X_t) = t\sigma^2$ y $\text{Cov}(X_t, X_{t-k}) = (t - k)\sigma^2$

En particular, el *error de predicción de lag h adelante* es

$$e_t(h) = e_{t+1} + \dots + e_{t+h}$$

y tiene varianza $h\sigma^2$, por lo tanto, es impredecible

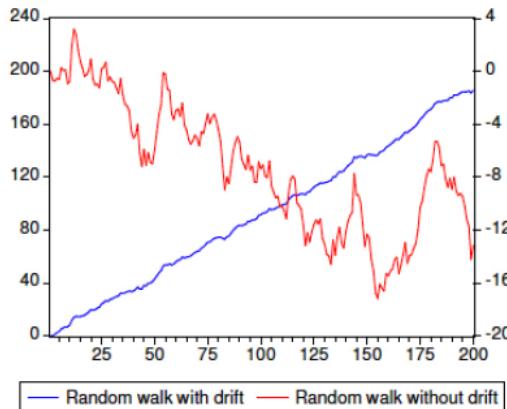


Figura 42: Gráfico de un paseo aleatorio con y sin *drift*

7.4. El papel de las constantes en los procesos no estacionarios

Una característica importante que distingue los procesos estacionarios de los no estacionarios es el papel de las constantes ¹³:

- En un proceso estacionario la constante no es importante, y hemos restado su media a las observaciones y hemos trabajado con procesos de media cero. Tanto la forma del proceso como sus propiedades básicas son las mismas ya sea que la media sea cero o diferente de cero.
- Sin embargo, en un proceso no estacionario las constantes, si existen, son muy importantes y representan alguna propiedad permanente del proceso.

¹³Se puede apreciar por ejemplo en la Figura 42

7.5. Test de raíz unitaria: *Dickey-Fuller*

Estas pruebas nos dicen si tenemos que tomar una diferencia adicional en una serie para hacerla estacionaria. Supongamos que deseamos decidir entre el proceso no estacionario:

$$\nabla X_t = \alpha + \beta t + \epsilon_t$$

y el proceso estacionario:

$$(1 - \phi B)X_t = \alpha + \epsilon_t$$

o un proceso NO estacionario:

$$(1 - \phi B)X_t = \alpha + \beta t + \epsilon_t$$

La hipótesis nula del test

$$H_0 : \phi = 1$$

es que la serie no es estacionaria y es necesario tomar diferencias.

Este modelo, puede reescribirse como

$$\nabla X_t = X_t - X_{t-1} = \alpha + \beta t + \gamma X_{t-1} + \epsilon_t$$

donde $\gamma = \phi - 1$ y la hipótesis nula es que $\gamma = 0$.

La prueba se realiza haciendo una regresión de ∇X_t en t y X_{t-1} . El estadístico del test es

$$\frac{\hat{\gamma}}{\sqrt{\text{Var}(\gamma)}}$$

El proceso será estacionario si $-2 < \gamma < 0 \iff -1 < \phi < 1$.

El test de *Dickey Fuller* se generaliza a procesos AR(p) con $H_0 : \gamma = 0$ en

$$\nabla X_t = \alpha + \beta t + \gamma X_{t-1} + \nabla_1 \nabla X_{t-1} + \cdots + \nabla X_{t-p} + \epsilon_t$$

La distribución del estadístico de prueba es un funcional del *proceso de Wiener*, y es diferente si c es cero o no, y ambos están tabulados.

En R, la prueba la realiza *adf.test* en la biblioteca **tseries** y *adfTest* en la biblioteca **fUnitRoots**.

Table: Critical values for the Dickey-Fuller unit-root test

T	without constant				with constant			
	.01	.025	.05	.1	.01	.025	.05	.1
25	-2.66	-2.26	-1.95	-1.60	-3.75	-3.33	-3.00	-2.63
50	-2.62	-2.25	-1.95	-1.61	-3.58	-3.22	-2.93	-2.60
100	-2.60	-2.24	-1.95	-1.61	-3.51	-3.17	-2.89	-2.58
250	-2.58	-2.23	-1.95	-1.62	-3.46	-3.14	-2.88	-2.57
500	-2.58	-2.23	-1.95	-1.62	-3.44	-3.13	-2.87	-2.57
∞	-2.58	-2.23	-1.95	-1.62	-3.43	-3.12	-2.86	-2.57

Observación 7.7. Para las consecuencias negativas de la subdiferenciación, utilice un nivel de significancia bajo, para evitar el riesgo de rechazar que el proceso no es estacionario cuando en realidad lo es.

Ejemplos 7.8. log-NASDAQ Index

Si Z_t es una serie del índice bursátil general, entonces $Y_t = \nabla \log(Z_t)$ es

$$Y_t = \nabla \log(Z_t) = \log\left(\frac{Z_t}{Z_{t-1}}\right) = \log\left(1 + \frac{Z_t - Z_{t-1}}{Z_{t-1}}\right) \approx \frac{\nabla \log(Z_t)}{Z_{t-1}}$$

La serie **lnas** es $Y_t \text{ auto.arima}(lnas)$

```

ar1      mean
-0.1206 5e-04
s.e.  0.0162 2e-04
sigma^2 estimated as  0.000191: log likelihood=10741.32
AIC=-21476.64 AICc=-21476.64 BIC=-21457.95

```

El modelo ajustado es

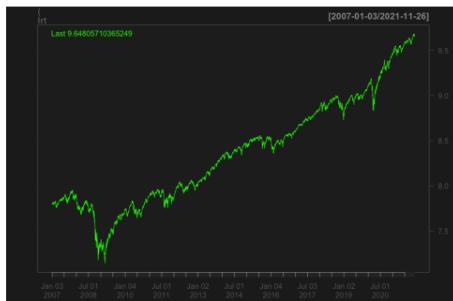
$$Y_t - 0,0005 = -1,206(Y_{t-1} - 0,0005) + \epsilon_t \quad \sigma_\epsilon^2 = 0,000191$$

ESTO ESTAMAL MIRAR CODIGO

```

getSymbols("^IXIC")
nas=IXIC$IXIC.Adjusted
lnas=log(nas)
chartSeries(lnas)

```

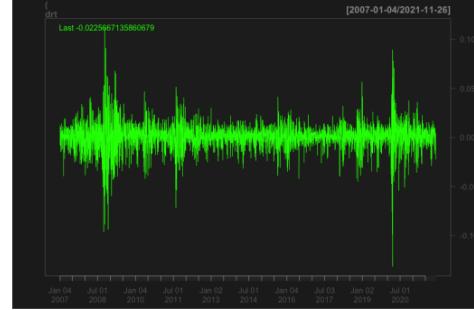


(a) log -NASDAQ Index

```

lnas=na.omit(diff(log(nas)))
chartSeries(lnas)

```



(b) 1st differences of log-NASDAQ Index (log-returns)

Ejemplos 7.9. Tenemos datos del 8 de mayo de 2019 al 2 de noviembre de 2019.¹⁴



Figura 44: EUR/USD tasa de cambio

¹⁴Ver Ver el código en UnitRoot.R

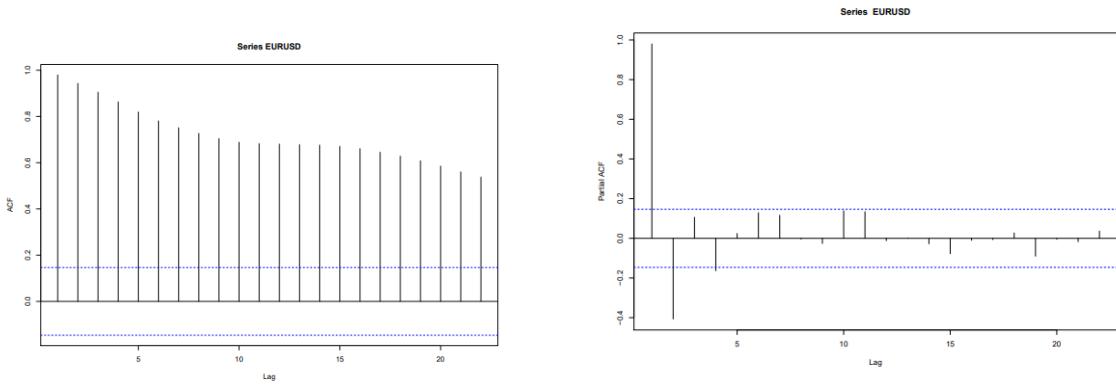
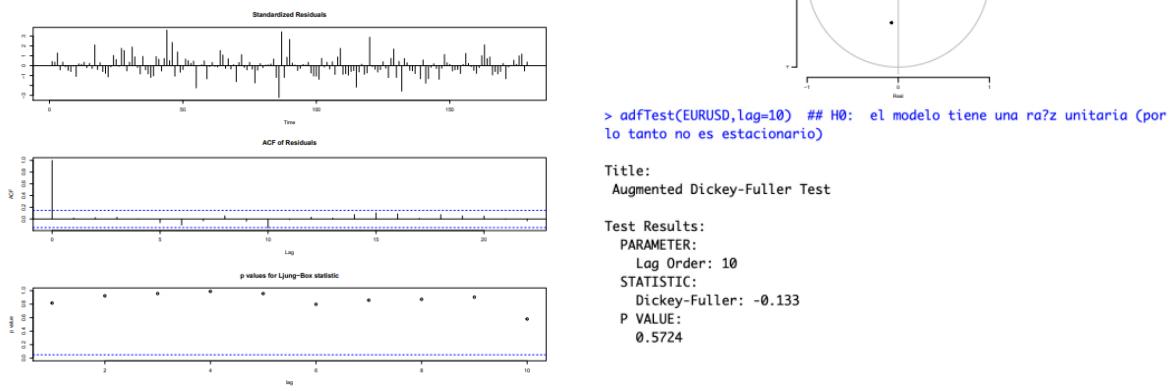


Figura 45: Funciones ACF y PACF de la Figura 44

Los residuos después de ajustar un modelo AR(4) para que los datos parezcan ruido blanco, pero son bastante grandes:



(a) Funciones de los residuos del modelo AR(4)

(b) Test de *Dickey-Fuller* para los residuos

Figura 46: Estudio de los residuos.

Lo que nos dice la Figura 46 es que al haber una raíz muy cercana al círculo unidad, estamos estimando los datos con un modelo con raíz unidad que por tanto no sería estacionario. Por tanto, descartamos este modelo. Como parece que la serie no estacionaria, podemos estudiar la serie diferenciada:

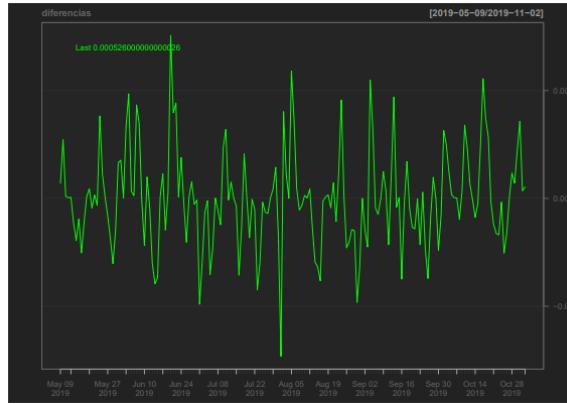


Figura 47: Plot de la serie diferenciada, que a priori parece mejor (en términos de estacionariedad)

Podemos estimar un modelo para la serie diferenciada

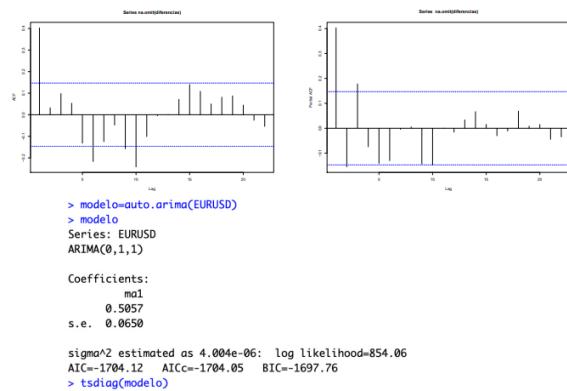


Figura 48: Funciones *ACF* y *PACF* junto a la estimación de un modelo

Podemos ver que los residuos de la estimación del modelo MA(1) para la serie diferenciada parecen ser ruido blanco:

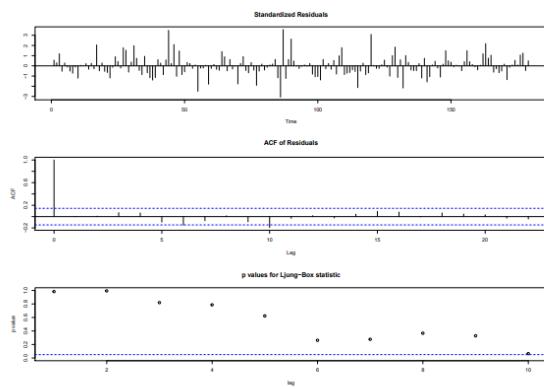


Figura 49: *Plots* de los residuos del modelo

7.6. ¿Cómo funciona `auto.arima`?

La función `auto.arima` de R usa una variación del algoritmo de *Hyndman-Khandakar*¹⁵, que combina tests de la raíz unidad, la minimización de AICc y MLE para obtener el modelo ARIMA

Hyndman-Khandakar algorithm for automatic ARIMA modelling	
1.	The number of differences $0 \leq d \leq 2$ is determined using repeated KPSS tests.
2.	The values of p and q are then chosen by minimising the AICc after differencing the data d times. Rather than considering every possible combination of p and q , the algorithm uses a stepwise search to traverse the model space.
a.	Four initial models are fitted: <ul style="list-style-type: none">◦ ARIMA(0, d, 0),◦ ARIMA(2, d, 2),◦ ARIMA(1, d, 0),◦ ARIMA(0, d, 1). A constant is included unless $d = 2$. If $d \leq 1$, an additional model is also fitted: <ul style="list-style-type: none">◦ ARIMA(0, d, 0) without a constant.
b.	The best model (with the smallest AICc value) fitted in step (a) is set to be the “current model.”
c.	Variations on the current model are considered: <ul style="list-style-type: none">◦ vary p and/or q from the current model by ± 1;◦ include/exclude c from the current model. The best model considered so far (either the current model or one of these variations) becomes the new current model.
d.	Repeat Step 2(c) until no lower AICc can be found.

Figura 50: Algoritmo de *Hyndman-Khandakar* para la modelización de ARIMA automática.

En el futuro, volveremos al modelado de series temporales no estacionarias y definiremos las familias ARCH y GARCH para abordar la no estacionariedad en el diferenciado.

7.7. Modelos ARIMA estacionales

Recordemos primero la definición de estacionalidad.

Definición 7.10. La *estacionalidad* en una serie temporal es un patrón regular de cambios que se repite durante s períodos de tiempo, donde s define el número de períodos de tiempo hasta que el patrón se repite nuevamente.

Ejemplos 7.11. Por ejemplo, existe una estacionalidad en los datos mensuales para los cuales los valores altos siempre tienden a ocurrir en algunos meses en particular y los valores bajos siempre tienden a ocurrir en otros meses en particular. En este caso, $s = 12$ (meses por año) es el lapso del comportamiento estacional periódico.

Para datos trimestrales, $s = 4$ períodos temporales por año. En la Figura 51 podemos ver una serie temporal con una clara estacionalidad trimestral. Además entre el inicio de la serie y el momento del COVID, hay una clara tendencia creciente lineal.

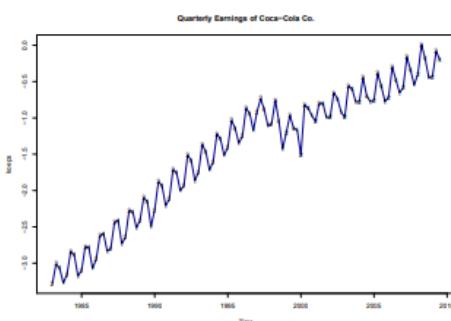


Figura 51: Serie temporal con estacionalidad trimestral.

¹⁵ Automatic time series forecasting: The forecast package for R. Journal of Statistical Software, 27(1), 1-22, <https://doi.org/10.18637/jss.v027.i03>

En un modelo ARIMA estacional, los términos AR y MA estacionales predicen X_t utilizando valores de datos y errores en momentos con *lags* que son múltiplos de s (el *lag* de la estacionalidad). Con datos mensuales (y $s = 12$), un modelo AR(1) estacional usaría X_{t-12} para predecir X_t . Un AR(2) estacional usaría X_{t-12} y X_{t-24} para predecir X_t , y así sucesivamente.

La estacionalidad generalmente hace que la serie no sea estacionaria porque los valores promedio en algunos momentos particulares dentro del lapso estacional (meses, por ejemplo) pueden ser diferentes de los valores promedio en otros momentos. Por ejemplo, las ventas de ventiladores de refrigeración, limonada o cerveza siempre serán mayores en los meses de verano.

7.7.1. Diferenciación Estacional

Definición 7.12. Diferenciación estacional

La diferenciación estacional se define como una diferencia entre un valor y un valor con un *lag* múltiplo de s .

Diferenciaciones más comunes.

- Con $s = 12$, lo que puede ocurrir con datos mensuales, una diferencia estacional es $(1 - B_{12})X_t = X_t - X_{t-12}$.

Las diferencias (con respecto al año anterior) pueden ser aproximadamente las mismas para cada mes del año, lo que nos da una serie estacionaria.

- Con $s = 4$, lo que puede ocurrir con datos trimestrales, una diferencia estacional es $(1 - B^4)X_t = X_t - X_{t-4}$.

La diferenciación estacional elimina la tendencia estacional y también puede eliminar un tipo de no estacionariedad de paseo aleatorio estacional.

7.7.2. Diferenciación para la tendencia y la estacionalidad

- Si la tendencia está presente en los datos (como en las ganancias trimestrales de CocaCola ¹⁶), es posible que también necesitemos diferencias no estacionales.

A menudo (no siempre) una primera diferencia (no estacional) “eliminará la tendencia” de los datos. Es decir, utilizamos

$$(1 - B)X_t = X_t - X_{t-1}$$

en presencia de tendencia.

- Cuando están presentes tanto la tendencia como la estacionalidad, es posible que necesitemos aplicar tanto una primera diferencia no estacional como una diferencia estacional. Es decir, es posible que necesitemos examinar el ACF y el PACF de

$$(1 - B^S)(1 - B)X_t = (X_t - X_{t-1}) - (X_{t-S} - X_{t-(S+1)})$$

¹⁶Ver sarima.R

7.8. El Modelo ARIMA Estacional

Modelo ARIMA Estacional (SARIMA)

El modelo ARIMA estacional incorpora ambos factores, estacional y no estacional, en un modelo multiplicativo. Una forma de escribir el modelo es

$$ARIMA(p, d, q) \times (P, D, Q)_s$$

o

$$ARIMA(p, d, q) \times (P, D, Q)[s]$$

donde p = orden del AR no estacional, d = diferenciación no estacional, q = orden del MA no estacional, P = orden del AR estacional, D = diferenciación estacional, Q = orden del MA estacional y s es el lapso de tiempo en el que se repite un patrón estacional. Sin operadores diferencia, escribimos el modelo como

$$\Phi(B^s)\Phi(B)(X_t - \mu) = \Theta(B^s)\Theta(B)\epsilon_t$$

La componente no estacional es

- AR: $1 - \phi_1 B - \cdots - \phi_p B^p$
- MA: $1 + \theta_1 B + \cdots + \theta_q B^q$

La componente estacional es

- AR: $1 - \phi_{1,s} B^s - \cdots - \phi_{P,s} B^{Ps}$
- MA: $1 + \theta_{1,s} B^s + \cdots + \theta_{Q,s} B^{Qs}$

7.8.1. Identificación de modelos Estacionales

1. Haga un gráfico de serie temporal de los datos.

Examínelo en busca de características como tendencia y estacionalidad. Sabrá que ha recopilado datos estacionales (meses, trimestres, etc.), así que observe el patrón en esas unidades de tiempo (meses, etc.) para ver si realmente existe un patrón estacional.

2. Haga cualquier diferenciación necesaria

- Si **hay estacionalidad y no hay tendencia**, entonces se toma una diferencia de *lag s*. Por ejemplo, tomemos una diferencia número 12 para los datos mensuales con estacionalidad. La estacionalidad aparecerá en el ACF disminuyendo lentamente en múltiplos de s .
- Si **hay una tendencia lineal y no hay una estacionalidad** obvia, entonces tome una primera diferencia. Si hay una tendencia curva, considere una transformación de los datos antes de diferenciar.
- Si **hay tendencia y estacionalidad**, aplique una diferencia estacional a los datos y luego reevalúe la tendencia. Si la tendencia persiste, tomemos primero las diferencias.
- Si **no hay una tendencia obvia ni una estacionalidad**, no hagas diferencias.

3. Estamos utilizando esta información para **determinar posibles modelos**. Esto puede ser complicado si implica algunas conjeturas (educadas). Algunas orientaciones básicas:

- **Términos no estacionales:** los picos en el ACF (con retrasos bajos) indican términos de MA no estacionales. Los picos en el PACF (con *lags* bajos) indican posibles términos AR no estacionales.
 - **Términos estacionales:** examine los patrones entre *lags* que son múltiplos de s . Juzgue el ACF y el PACF en los desfases estacionales de la misma manera que lo hace para los desfases anteriores.
4. Estime los modelos que podrían ser razonables sobre la base del paso 3. No olvide incluir cualquier diferenciación que haya hecho antes de consultar el ACF y el PACF.
 5. Examina los residuos (con ACF, *Ljung-Box test* o otras medidas para ver que el modelo es bueno. Si piensa que no es bueno, vuelve al paso 3, incluso tal vez al 2.

8. OTROS MODELOS PARA SERIES NO ESTACIONARIAS: MODELOS ARCH Y GARCH

8.1. Motivación: Series Temporales Financieras

El precio de un activo como función del tiempo es una serie temporal financiera.

Pero el precio de un activo razonable suele crecer de manera exponencial, como se puede ver por ejemplo en la Figura 53a, por lo tanto la serie resultante es difícil de estudiar. La mayor parte de la teoría que conocemos está hecha para funciones lineales.

El **valor medio** de una serie que crece en forma exponencial no nos da ninguna información. La **derivada** de una exponencial es exponencial, así que los cambios intra-diarios en el precio siguen teniendo los mismos inconvenientes.

Estudiaremos entonces la serie de los beneficios, que como veremos, tiene mejor comportamiento.

8.1.1. Rentabilidad o beneficio (*returns*)

Rentabilidad simple o neta (*simple return*)

Se define la **rentabilidad simple o neta** de una acción como la ganancia o pérdida que se obtiene como resultado de la inversión en acciones, expresada como un porcentaje del capital invertido. Esta medida proporciona una indicación de la eficacia de una inversión en acciones y se calcula utilizando la Fórmula 5 (expresada en tanto por 1).

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (5)$$

Nota 8.1. El beneficio es un resumen completo y libre de escalas para la inversión.

Rentabilidad bruta (*gross return*)

Se define la **rentabilidad bruta** como el rendimiento total de una inversión o activo financiero a lo largo del tiempo antes de tener en cuenta los costos y gastos asociados. En el caso de series temporales financieras, la rentabilidad bruta se calcula a través de la Fórmula 6 (expresada en tanto por 1).

$$R_t + 1 = \frac{P_t}{P_{t-1}} \quad (6)$$

Rentabilidad continua (*log-returns*)

Se define la **rentabilidad continua** describe el rendimiento de una inversión o activo en términos de tasas de interés o rendimientos compuestos de forma continua a lo largo del tiempo.

La rentabilidad continua se calcula a través de la Fórmula 7 (expresada en tanto por 1).

$$r_t = \log(1 + R_t) = \log(P_t) - \log(P_{t-1}) \approx R_t \quad (7)$$

Observación 8.2. Como en general los R_t de un período suelen ser pequeños, se utiliza que

$$\log(1 + R_t) \approx R_t$$

8.2. Lectura de una tabla de precios de activos financieros

Estudiemos la tabla de precios de las acciones de *General Electric Company (GE)*¹⁷

General Electric Company (GE) - NYSE						
22.71 +0.02(0.09%) Sep 28, 4:00PM EDT After Hours : 22.70 -\$0.01 (0.04%) Sep 28,						
Prices						
Date	Open	High	Low	Close	Volume	Adj Close*
Sep 28, 2012	22.77	22.96	22.62	22.71	70,926,100	22.71
Sep 27, 2012	22.24	22.86	22.13	22.73	67,300,400	22.73
Sep 26, 2012	22.16	22.25	22.07	22.10	41,123,300	22.10
Sep 25, 2012	22.39	22.67	22.30	22.31	47,039,200	22.31
Sep 24, 2012	22.40	22.45	22.30	22.36	36,799,800	22.36
Sep 21, 2012	22.55	22.69	22.46	22.53	66,551,700	22.53
Sep 20, 2012	22.19	22.47	22.12	22.43	43,677,000	22.43
Sep 20, 2012				0.17 Dividend		
Sep 19, 2012	22.30	22.49	22.25	22.43	40,462,600	22.26
Sep 18, 2012	21.99	22.24	21.96	22.24	38,831,200	22.07
Sep 17, 2012	21.93	22.05	21.90	22.05	79,283,200	21.88
Sep 14, 2012	22.20	22.37	21.98	22.11	100,148,400	21.94

Figura 52: Precios históricos de GE (NYSE)

Por lo general, las tablas que presentan información sobre acciones suelen contener la información de la Caja 8.2

Información en tablas de acciones

- **Ticker:** La identificación de la compañía en el mercado (e.g. para General Electric Co. es GE).
- **Open:** el precio de la acción en el momento de la apertura de la sesión del mercado.
- **Close:** Precio actual de la acción si el mercado está operando, o el último precio negociado cuando cierra el mercado.
- **High (o Max):** Precio máximo alcanzado por el valor entre el Open y el Close.
- **Low (o Min):** Precio mínimo alcanzado por el valor entre el Open y el Close.
- **Volume:** Número de acciones negociadas entre los tiempos del Open y Close.
- **Adjusted Close:** Precio de cierre ajustado para incluir cualquier dividendo, u otra acción corporativa que afecte el precio de la acción (e.g. splits, o derechos).

Definición 8.3. Precios ajustados al cierre

Los ajustes dependen del intervalo histórico que estemos considerando, y pretenden “desconectar” los efectos resultantes de acciones programadas, tales como el pago de dividendos y los splits.

Close	Adjusted
45,00	$=(1/2)*(1-(0,08/25)) * 45 = 22,43$
48,00	$=(1/2)*(1-(0,08/25)) * 48 = 23,92$
25,00 (split 2:1)	$=(1-(0,08/25)) * 25 = 24,92$
24,00 (dividend 0,08)	$=24 = 24,00$

¹⁷Podemos conseguir este tipo de información en Yahoo (<http://finance.yahoo.com/>), Google (<http://www.google.com/finance>), o MSN Moneycentral (<http://moneycentral.msn.com/>), entre otros.

Ejemplos 8.4. Veamos ejemplos reales de algunos de los conceptos que acabamos de definir.

1. En la Figura 53a podemos ver como el precio de las acciones de Apple tiene un comportamiento exponencial no lineal, por lo que modelarlo es más complicado a priori. En cambio, en la Figura 53b podemos ver el beneficio de las acciones de Apple, una serie que a diferencia de la anterior, parece que tiene media constante, no obstante no parece tener varianza constante. Estudiemos pues como podríamos modelarla.

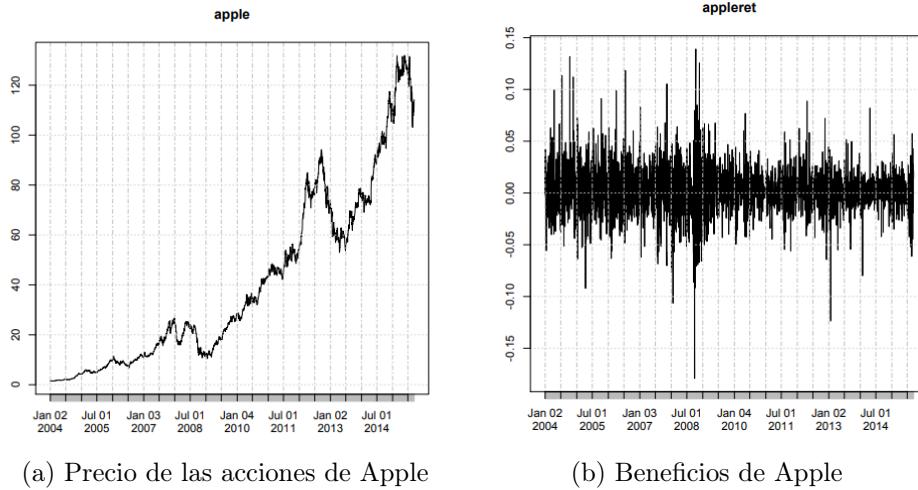


Figura 53: Análisis de las acciones de Apple

2. En la Figura 54 podemos ver los *returns* y *log-returns* de la empresa IBM.

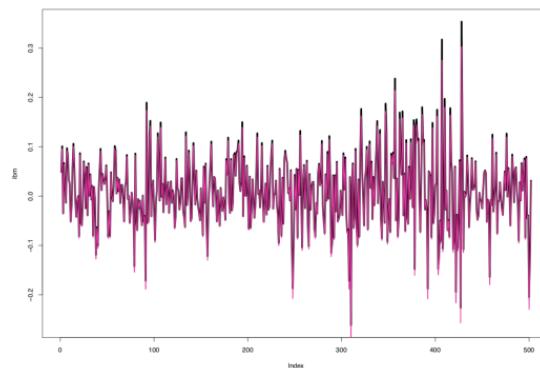


Figura 54: *Returns* (en negro) y *Log-Returns* (en maroon) de la empresa IBM

3. En la Figura 55 podemos ver la gráfica de la densidad de los *returns* diarios del Banco Santander entre mayo del 2012 y noviembre del 2012. Usualmente los $\log \text{returns} \approx R_t$ tienen una distribución normal con media μ y varianza σ^2 .

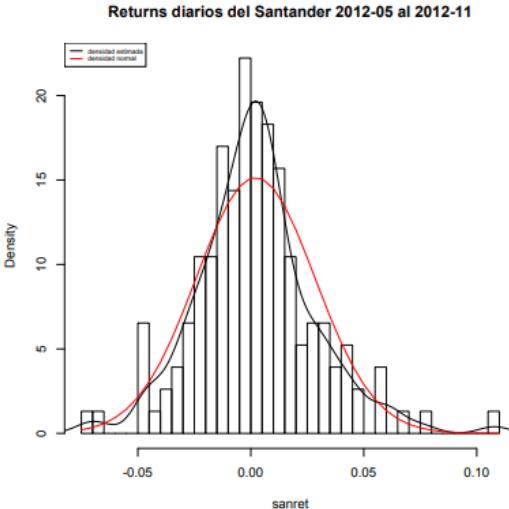


Figura 55: Densidad de los *returns* diarios del Banco Santander entre mayo de 2012 y noviembre de 2012.

8.3. De la *log-Normal* a la Normal

Si $Y \sim \log - N(\mu_y, \sigma_y^2)$, entonces $X = \log(Y)$ es normal con media y varianza dadas por

$$\mathbb{E}(X) = \log \left(\frac{\mu_y}{\sqrt{1 + \frac{\sigma_y^2}{\mu_y^2}}} \right) \quad \text{Var}(X) = \log \left(1 + \frac{\sigma_y^2}{\mu_y^2} \right)$$

ya que

$$\mathbb{E}(Y) = \exp \mu + \frac{\sigma^2}{2} \quad \text{Var} = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

Si el *log-ret* r_t de un activo tiene distribución $r_t \sim N(0,0119, (0,0663)^2)$, la esperanza y varianza del *beneficio simple* $R_t = \exp r_t - 1$ se calcula como:

$$\begin{aligned} \mathbb{E}(Y) &= \mathbb{E}(r_t) = \exp 0,0119 + \frac{0,0663^2}{2} = 1,014 \\ \text{Var}(Y) &= \exp(2 \times 0,0119 + 0,0663^2) \times (\exp(0,0663^2) - 1) = 0,00454 \end{aligned}$$

entonces, como $\mathbb{E}(R_t) = \mathbb{E}(Y) - 1 = 0,014$ y $\text{Var}(R_t) = \text{Var}(Y)$.

8.3.1. AutoCovarianza y AutoCorrelación

Para la serie de log-*returns* $\{r_t : 1, \dots, T\}$, interesa calcular la *lag-l autocovarianza* y la autocorrelación ACF:

$$\gamma_l = \mathbf{Cov}(r_t, r_{t-l}) \quad \rho_l = \frac{\mathbf{Cov}(r_t, r_{t-l})}{\mathbf{Var}(r_t)} = \frac{\gamma_l}{\gamma_0}$$

Lema 8.5. $\{r_t\}$ es débilmente estacionaria

La media y las lag-l autocovarianzas de r_t son invariantes respecto de t .

Esta suposición suele observarse empíricamente y dice que los beneficios fluctúan con variancia constante alrededor de un nivel constante μ .

8.4. “Stylized facts”

La mayoría de las series diarias r_t o R_t de beneficios de activos financieros (acciones, índices, tasas de cambio, *commodities*, ...) presentan ciertas características comunes, que a menudo se observan también en otras escalas (semanal, mensual, *intra-dia*). Algunas de estas características son

1. No son *i.i.d.*, aunque puedan tener poca correlación.
2. Las series de valores absolutos o cuadrados tienen correlación fuerte.
3. Su esperanza condicional (a la información disponible del pasado) es cercana a 0.
4. Su desviación estándar condicional (volatilidad) varía con el tiempo.
5. Tienen colas pesadas.
6. Los extremos (y a menudo la volatilidad) aparecen en *clusters* ¹⁸

8.5. Volatilidad

Volatilidad

Definición 8.6. Definimos la volatilidad como el grado de fluctuación de los precios de los activos financieros.

Nota 8.7. La volatilidad no es directamente observable: Se pueden observar fluctuaciones en la serie de precios pero no se puede especificar la tendencia a producir valores extremos.

Por lo tanto, la volatilidad se estima. Lo más común es estimarla con la desviación estándar observada.

Para la serie de log-returns $\{r_1, \dots, r_T\}$, la *volatilidad histórica* ^a es

$$\hat{\sigma} = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_t - \hat{\mu})^2}$$

^aSi se aceptan modelos como el de Black-Scholes para ponerle precio a las opciones se puede deducir una ecuación para la volatilidad (“*implied volatility*”)

8.6. Modelos para la Volatilidad

Para estudiar la volatilidad diaria, trabajaremos con la serie de beneficios (*returns*):

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

- Supongamos que la media observada es $M_r = 0$, entonces la **estimación básica de la volatilidad el día m es**

$$\sigma_m^2 = \frac{1}{n} \sum_{i=1}^n R_{m-i}^2$$

¹⁸ "Los grandes cambios tienden a estar seguidos por grandes cambios, de cualquier signo, y los pequeños cambios tienden a ser seguidos por pequeños cambios" -Mandelbrot, 1963.

- Los returns más recientes, deberían tener más influencia que los muy lejanos en el tiempo, de modo que podríamos asignar diferentes pesos a cada uno de ellos:

$$\sigma_m^2 = \sum_{i=1}^n \alpha_i R_{m-i}^2$$

con $\alpha_i > \alpha_{i+1}$ y $\sum_i \alpha_i = 1$.

- La maneras más sencilla de asignar los pesos con estas propiedades es poner

$$\alpha_i = \lambda^{i-1}(1 - \lambda) \quad 0 < \lambda < 1$$

(EWMA) Exponentially Weighted Moving Average

Definimos la **Media Móvil Ponderada Exponencialmente (EWMA)** para la volatilidad como

$$\sigma_m^2 = (1 - \lambda) \sum_{i=1}^n \lambda^{i-1} R_{m-i}^2$$

$0 < \lambda < 1$.

A la práctica, $\lambda = 0,94$ ^a.

^aCalculado por J.P. Morgan en RiskMetrics database

8.7. Modelo ARCH (Engle, 1982)

El primer modelo que considera varianzas dependientes del tiempo es el **ARCH(p)**. ¹⁹

La idea es que la varianza condicional es una función del cuadrado de los ruidos pesados. Si r_t es la sucesión de *log-returns*, entonces

$$\begin{aligned} r_t &= \mu_t + a_t \\ a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 a_{t-1}^2 + \dots + \alpha_p a_{t-p}^2 \end{aligned}$$

donde $\{\epsilon_t\}$ es una sucesión de v.a. *iid*, con $\epsilon_t \sim WN(0, 1)$, $\alpha_i \geq 0$.

Con este modelo hay que tener precauciones:

Observación 8.8. σ_t^2 tiene que ser positivo, así que $\omega > 0$ y $\alpha_i \geq 0$.

Por tanto, podemos reescribir el modelo anterior definiendo $W_t = a_t^2 - \sigma_t^2$, obtenemos

$$a_t^2 = \omega + \sum_{i=1}^2 \alpha_i a_{t-i}^2 + W_t$$

es decir, a_t^2 es un **AR(p)**.

Por lo tanto el proceso será débilmente estacionario si

$$|\alpha_1 + \dots + \alpha_p| < 1$$

y en ese caso, la varianza (incondicional) es

$$\sigma^2 = \text{Var}(a_t) = \frac{\omega}{1 - \sum_{i=1}^p \alpha_i} \quad (8)$$

Observación 8.9. Para ajustar un modelo **ARCH(p)**, determinamos el orden p mirando la PACF de los residuos de la ecuación para la media y estimamos $\omega, \alpha_1, \dots, \alpha_m$ (MLE). ²⁰

¹⁹Robert Engle (1982), Autoregressive Conditional heterocedasticity with estimates of the variance of UK inflation, Econometrica 50, 987-1008

²⁰En R lo podemos hacer con **fGarch**.

8.7.1. Efecto ARCH

Para saber si estamos en presencia de heterocedasticidad, podemos hacer lo siguiente: Miramos los residuos²¹

$$\hat{a}_t = \hat{r}_t - \hat{\mu}_t$$

y tenemos dos posibilidades

1. **Test de Ljung – Box** para \hat{a}_t^2 ²². Si rechazamos el test para los primeros p lags, entonces hay algún grado de dependencia serial.
2. **Test de Engle**, que consiste esencialmente en hacer la regresión

$$a_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i a_{t-i}^2$$

y hacer un F -test para $H_0 : \alpha_1 = \dots = \alpha_p = 0$.

3. **Test de los multiplicadores de Lagrange**. El test que queremos llevar a cabo es

$$H_0 : \alpha_1 = \dots = \alpha_m = 0 \quad (\text{homocedasticidad condicional})$$

El test de los multiplicadores de Lagrange (LM)

$$\text{LM}_{H_0} \sim \chi_m^2 \quad (\text{asimtóticamente } T \rightarrow \infty)$$

- Rechazaremos H_0 si $\text{LM} > \chi_m^{2,1-\alpha}$.
- Estadístico:

$$\text{LM} = T \times R^2$$

donde R^2 es el coeficiente de determinación al ajustar el modelo AR(m) a a_t^2 :

$$a_t^2 = \alpha_0 + \sum_{j=1}^m \alpha_j a_{t-j}^2 + v_t$$

- En las aplicaciones, e.g. modelo ARMA(p,q) con errores ARCH(m), $\{a_t\}$ se reemplaza por los residuos.

8.7.2. Propiedades de los ARCH

Podemos dividirlas en dos grupos:

- Ventajas:
 1. Simplicidad.
 2. Genera *Volatility Clusters*.
 3. Colas Pesadas.
- Desventajas:
 1. Simetría (los returns positivos y negativos tienen igual influencia).
 2. Suelen resultar modelos de orden (p) alto.

²¹en el mejor caso $\mu_t = 0$

²²Se conoce como “McLeod-Li-test”

8.7.3. Momentos condicionales de un ARCH(1)

- Usando que ϵ_t es independiente de Ω_{t-1} obtenemos:

$$\mathbb{E}(a_t|\Omega_{t-1}) = \mathbb{E}(\sigma_t\epsilon_t|\Omega_{t-1}) = \sigma_t\mathbb{E}(\epsilon_t|\Omega_{t-1})_{=\mathbb{E}(\epsilon)=0} = 0$$

Observación 8.10. $\{a_t\}$ es una sucesión no correlacionada si los segundos momentos son finitos.

- Como $\mathbb{E}(a_t|\Omega_{t-1}) = 0$, resulta que:

$$\mathbf{Var}(a_t|\Omega_{t-1}) = \mathbb{E}(a_t^2|\Omega_{t-1}) = \sigma_t^2\mathbb{E}(\epsilon_t^2|\Omega_{t-1})_{=\mathbb{E}(\epsilon_t^2=1)} = \sigma_t^2$$

Así que, o bien $a_t|\Omega_{t-1} \sim \text{Normal}(0, \sigma_t^2)$, o bien, $a_t|a_{t-1} \sim \text{Normal}(0, \sigma_t^2)$

8.7.4. Momentos incondicionales de un ARCH(1)

- En primer lugar, $\mathbb{E}(a_t) = 0$.
- Usando la independencia de σ_t y ϵ_t :

$$\mathbf{Var}(a_t) = \mathbb{E}(a_t^2) = \mathbb{E}(\sigma_t^2\epsilon_t^2) = \mathbb{E}(\sigma_t^2)\mathbb{E}(\epsilon_t^2) = \mathbb{E}(\sigma_t^2)$$

- A pesar de que los ϵ_t son normales, $\{a_t\}$ es *No-Gaussiano* porque no es un proceso lineal de $\{\epsilon_t\}$.
- En particular, la kurtosis de a_t es más grande que la de una normal (i.e, 3), Suponiendo que $\mathbb{E}(a_t^4) < \infty$ tenemos:

$$\begin{aligned}\mathbb{E}(a_t^4) &= \mathbb{E}(\sigma_t^4\epsilon_t^4) = \mathbb{E}(\sigma_t^4)\mathbb{E}(\epsilon_t^4) = (\mathbf{Var}(\sigma_t^2) + \mathbb{E}(\sigma_t^2)^2)\mathbb{E}(\epsilon_t^4) \\ &\geq (\mathbb{E}(a_t^2))^2\mathbb{E}(\epsilon_t^4)\end{aligned}$$

- La kurtosis de a_t :

$$\frac{\mathbb{E}(a_t^4)}{(\mathbb{E}(a_t^2))^2} \geq \mathbb{E}(\epsilon_t^4) = 3$$

Serian iguales si $\mathbf{Var}(\sigma_t^2) = 0$ (Condición de homocedasticidad).

8.7.5. Estimación de un modelo ARCH(1)

Usaremos el MLE condicional. Si $f_{\alpha_0, \alpha_1}(a_t|a_{t-1})$ es la densidad de la $\text{Normal}(0, \alpha_0 + \alpha_1 a_{t-1}^2)$, la verosimilitud condicional de a_2, \dots, a_T dado a_1 es:

$$\begin{aligned}L(\alpha_0, \alpha_1|a_1) &= \prod_{t=2}^T f_{\alpha_0, \alpha_1}(a_t|a_{t-1}) \\ - \log(L(\alpha_0, \alpha_1|a_1)) &\propto \frac{1}{2} \sum_{j=2}^T \ln(\alpha_0 + \alpha_1 a_{t-1}^2) \\ &= \frac{1}{2} \sum_{t=2}^T \left(\frac{a_t^2}{\alpha_0 + \alpha_1 a_{t-1}^2} \right)\end{aligned}$$

se puede minimizar respecto a α_0 y α_1 numéricamente. La verosimilitud es bastante plana a menos que T sea muy grande.

8.7.6. Estimador insesgado de la varianza

Sea un modelo ARCH(1):

$$\begin{aligned} X_t &= \mu_t + a_t \\ a_t &= \sigma_t \epsilon_t \quad \epsilon_t \sim \text{iid o WN}(0,1) \\ \sigma_t^2 &= \text{Var}(X_t | X_{t-1}, \dots) \\ \sigma_t^2 &= \omega + \alpha a_{t-1}^2 \end{aligned}$$

suponiendo $\mu_t = 0$.

Veamos que a_t^2 es un estimador insesgado de σ_t^2 . Definimos $w_t = a_t^2 - \sigma_t^2$

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha a_{t-1}^2 \\ \Rightarrow a_t^2 &= \omega + \alpha a_{t-1}^2 + w_t \end{aligned}$$

Por lo tanto es una AR(1) en a .

8.7.7. ACF para el ARCH(1)

Observamos en la Figura 56 a partir de la ACF y PACF que un modelo razonable para a^2 es un $AR(1)$.

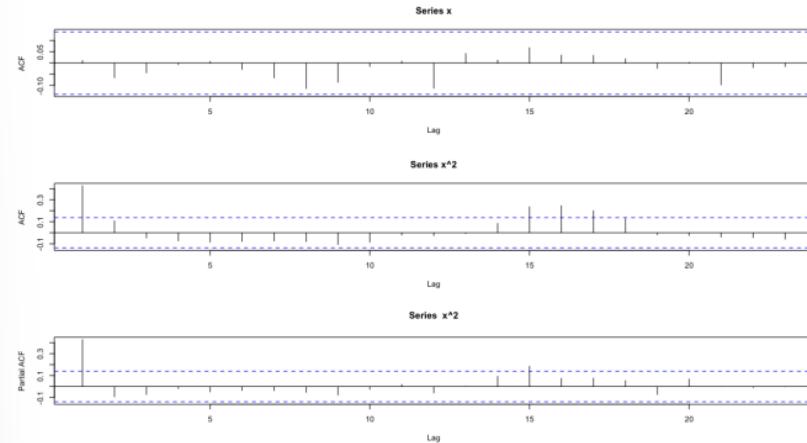


Figura 56: Gráficas de la ACF de una serie que parece ruido blanco, ACF y PACF de la serie al cuadrado.

Ejemplos 8.11. Ejemplo de un ARCH(1) simulado.

```
spec=garchSpec(model=list(alpha=c(0.8),beta=0))
x=garchSim(spec.n=200)
```

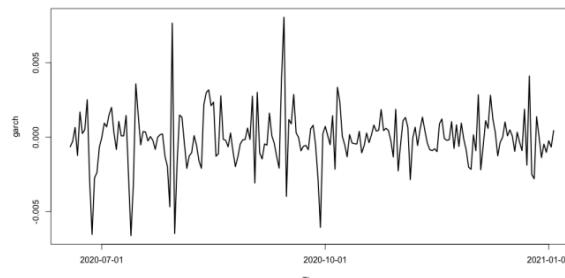


Figura 57: Ejemplo de un modelo ARCH(1) simulado en el que se pueden observar los *clusters* de volatilidad y *clusters* para los extremos.

Ejemplos 8.12. Beneficios para Alliance en 2009 Estudiamos los beneficios de Alliance el año 2009 (alvR) en la Figura 58²³.

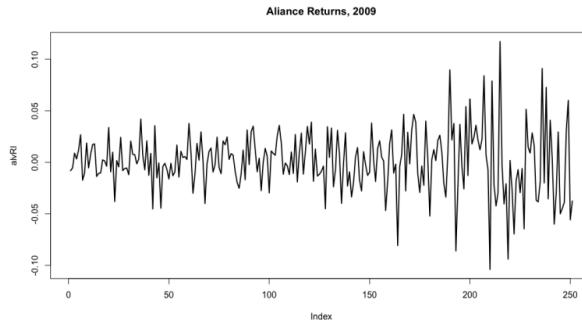


Figura 58: Beneficio del año 2009 de Alliance

Estudiemos como se comporta el ruido de los beneficios en la Figura 59.

```
>t.test(alvR1)
One Sample t-test
data: alvR1
t = 0.47975, df = 250, p-value = 0.6318
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-0.002792031  0.004590275
sample estimates:
mean of x
0.000899122
> auto.arima(alvR1)
Series: alvR1
ARIMA(0,0,0) with zero mean

sigma^2 estimated as 0.0008789: log likelihood=526.97
AIC=-1051.93   AICc=-1051.92   BIC=-1048.41
```

Figura 59: Código de R en el que estudiamos el ruido de los beneficios.

Podemos suponer que la media de estos es 0 (por los resultados obtenidos en el *t.test* y además, que los returns son ruido, por tanto:

$$r_t = a_t$$

Veamos si hay “Efect ARCH” en la Figura 60

```
> Box.test(alvR1^2,type="Ljung",lag=20)
Box-Ljung test
data: alvR1^2
X-squared = 144.43, df = 20, p-value < 2.2e-16
```

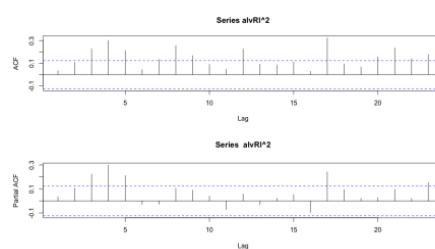


Figura 60: Estudio de la serie al cuadrado

²³En http://computationalfinance.lsi.upc.edu/?page_id=108 aparecen beneficios del año 2009 de diversas compañías que componen el índice DAX

Comprobamos que hay efecto ARCH ya que los residuos de los cuadrados son dependientes. Estimamos el orden P con la ACF y la PACF. En este caso como $r_t = a_t$, no hace falta estudiar los residuos, si no que estudiamos la serie como tal. Probamos estimar un ARCH(5) (Figura 61) para ver si se ajusta bien el modelo ,

```
> garchFit(~garch(5,0),data=alvRl,cond.dist="norm",trace=F,include.mean=F)

Title:
GARCH Modelling

Call:
garchFit(formula = ~garch(5, 0), data = alvRl, cond.dist = "norm",
  include.mean = F, trace = F)

Mean and Variance Equation:
data ~ garch(5, 0)
<environment: 0x123db1000>
[data = alvRl]

Conditional Distribution:
norm

Coefficient(s):
      omega    alpha1    alpha2    alpha3    alpha4    alpha5
0.00026383 0.00000001 0.05062743 0.32960759 0.13879548 0.21052532

Std. Errors:
based on Hessian

Error Analysis:
      Estimate Std. Error t value Pr(>|t|)
omega 2.638e-04 6.632e-05 3.978 6.95e-05 ***
alpha1 1.000e-08 5.761e-02 0.000 1.0000
alpha2 5.063e-02 6.115e-02 0.828 0.4877
alpha3 3.296e-01 1.166e-01 2.827 0.0047 **
alpha4 1.388e-01 7.341e-02 1.891 0.0587 .
alpha5 2.105e-01 1.234e-01 1.706 0.0879 .
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
555.0508   normalized: 2.211358
```

Figura 61: Ajuste de un modelo ARCH(5) para el modelo.

y por último, hacemos un análisis de sus residuos (podemos verlo en la Figura 62)

Standardised Residuals Tests:			
		Statistic	p-Value
Jarque-Bera Test	R	Chi^2	5.262779 0.07197839
Shapiro-Wilk Test	R	W	0.9934684 0.3437112
Ljung-Box Test	R	Q(10)	11.48153 0.3212524
Ljung-Box Test	R	Q(15)	15.92978 0.3867338
Ljung-Box Test	R	Q(20)	29.33704 0.08133188
Ljung-Box Test	R^2	Q(10)	9.396753 0.494909
Ljung-Box Test	R^2	Q(15)	11.29207 0.7316283
Ljung-Box Test	R^2	Q(20)	27.26425 0.1279919
LM Arch Test	R	TR^2	9.532543 0.6568961

Information Criterion Statistics:			
AIC	BIC	SIC	HQIC
-4.374906	-4.290633	-4.376014	-4.340993

Figura 62: summary del modelo estimado.

los *p-valores* son todos grandes (ninguno por debajo de 0.05) por lo que aparentemente las ecuaciones para las medias y las varianzas son adecuadas.

Por último, podemos estudiar el modelo que hemos ajustado para los beneficios de Alliance.

Este modelo para los beneficios de Alliance es:

$$r_t = a_t \quad \text{con media 0}$$

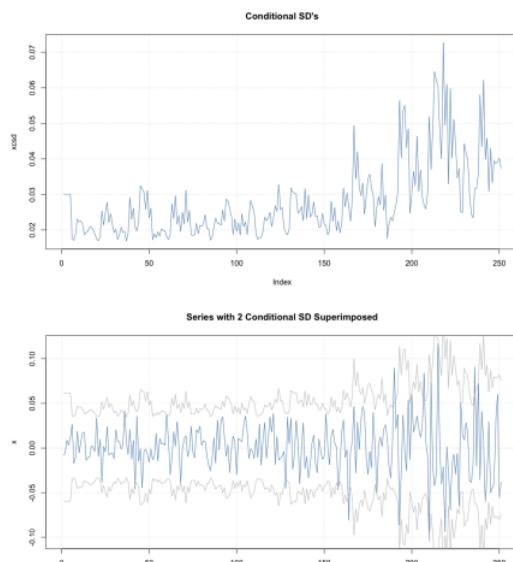
$$a_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \Omega + \sum_i i = 1^5 \alpha_i a_{t-i}^2 = 0,3a_{t-3}^2 + 0,1a_{t-4}^2 + 0,2a_{t-5}^2$$

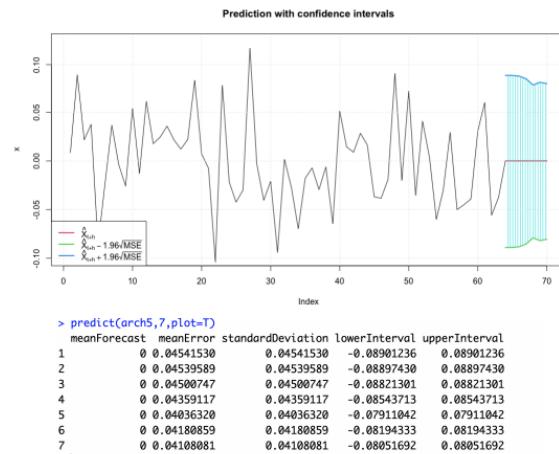
ya que $\omega = 2,6 \times 10^{-4} \approx 0$.

Nota 8.13. `garchFit` no admite poner parámetros fijos (`fixed=c(0,NA,...)`) de modo que sacamos (igualamos a 0) los coeficientes no significativos de manera temeraria -si queremos escribirlo a mano, porque en el R no podemos pedírselo.

Por último, podemos observar en la Figura 63a la desviación estándar condicional y las predicciones en la Figura 63b para el modelo dado.



(a) Desviación estándar condicional del modelo



(b) Predicciones para la serie con el modelo dado

8.8. GARCH (Generalized Autoregressive Conditional Heterocedasticity)

Este modelo ²⁴ supone que la volatilidad actual depende de:

- los p valores más recientes de la volatilidad.
- la magnitud de los q residuos más recientes en la serie observada.

es decir

$$a_{t-i} = r_{t-i} - \mu_{t-i}$$

Entonces, dada una serie de *log-returns* r_t , si $a_t = r_t - \mu_t$ es la innovación en el instante t , entonces a_t es un GARCH(p, q) si:

$$a_t = \sigma_t \epsilon_t \quad \epsilon_t \sim N(0, 1) \quad iid$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i a_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

²⁴Bollerslev (1986) Jr. of Econometrics, 31, 307-327

donde

$$\omega > 0 \quad \alpha_i \geq 0 \quad \beta_j \geq 0 \quad \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$$

Observación 8.14. $\text{GARCH}(m,0) = \text{ARCH}(m)$ y en genera, a_t^2 es un $\text{ARMA}(\max(p,q), q)$.

Considerem el canvi $w_t = a_t^2 - \sigma_t^2$

- Si $q \leq p$.

$$a_t^2 = \omega + \sum_{i=q+1}^p \alpha_i a_{t-i}^2 + \sum_{i=1}^q (\alpha_i + \beta_i) a_{t-i}^2 + \sum_{j=1}^q \beta_j (\sigma_{t-j}^2 - a_{t-j}^2) + \omega_t$$

$$a_t^2 = \omega + \sum_{i=1}^q (\alpha_i + \beta_i) a_{t-i}^2 + \sum_{i=q+1}^p \alpha_i a_{t-i}^2 + \sum_{j=1}^q -\beta_j w_{t-j} + \omega_t \sim \text{ARMA}(p, q)$$

- Si $q > p$:

$$a_t^2 = \omega + \sum_{i=1}^p (\alpha_i + \beta_i) a_{t-i}^2 + \sum_{i=p+1}^q \beta_i a_{t-i}^2 + \sum_{j=1}^q -\beta_j (a_{t-j}^2 - \sigma_{t-j}^2) + \omega_t$$

$$a_t^2 = \omega + \sum_{i=1}^p (\alpha_i + \beta_i) a_{t-i}^2 + \sum_{i=p+1}^q \beta_i a_{t-i}^2 + \sum_{j=1}^q -\beta_j (\omega_{t-j}^2) + \omega_t \sim \text{ARMA}(q, q)$$

Usar modelos ARMA para $\{a_t^2\}$ permite más flexibilidad para las ACFs.

8.8.1. GARCH(1,1)

- El modelo GARCH(1,1) lo escribimos:

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned}$$

donde $\alpha_0, \alpha_1, \beta_1 \geq 0$.

- $\{a_t^2\}$ es estacionario, si y solo si

$$\mathbb{E}(\log(\beta_1 + \alpha_1 \epsilon_t^2)) < 0$$

- Si $\alpha_1 + \beta_1 < 1$, entonces $\{a_t^2\}$ es estacionario y $\mathbb{E}(a_t^2) < \infty$. ²⁵²⁶

8.9. ARIMA(p, d, q)/GARCH(P, Q)

Ahora, basta con combinar todo el proceso anterior que conocemos para crear una familia completa y exhaustiva de modelos:

²⁵Podemos ver un ejemplo de estos modelos en el archivo **ARCH.R**

²⁶En: <https://rpubs.com/Bernlc82/555323> tenemos el ejemplo del ajuste de un modelo ARCH/GARCH

$ARIMA(p, d, q)/GARCH(P, Q)$

Dicimos que $\{Y_t\}_t$ es un proceso $ARIMA(p, d, q)/GARCH(P, Q)$ si cumple la siguiente ecuación:

$$\begin{aligned} a_t &= \sigma_t \epsilon_t \\ \sigma_t &= \sqrt{\omega + \sum_{i=1}^P \alpha_i a_{t-i}^2 + \sum_{j=1}^Q \beta_j \sigma_{t-j}^2} \\ Y_t &= \mu + \sum_{i=1}^p \phi_i (Y_{t-i} - \mu) + a_t + \sum_{j=1}^q \theta_j a_{t-j} \end{aligned}$$

donde $\omega > 0$.

La fórmula anterior describe esencialmente el proceso $ARIMA(p, d, q)/GARCH(P, Q)$ como una superposición de un proceso $ARIMA(p, d, q)$ con ruido $GARCH(P, Q)$.

9. BONDAD DE AJUSTE Y SELECCIÓN DE MODELOS

9.1. Ejemplo de modelo predicción

Carregueu les dades contingudes al fitxer "prac8TS.txt", corresponent a la producció mensual d'electricitat solar tèrmica a Catalunya entre 2013 i 2017.

```
1 prac8TS <- read.csv("C:/Users/Alba/Desktop/prac8TS.txt", sep="")
2 dades<-prac8TS
3 dadesany<-ts(dades, start = c(2013,1), end = c(2017,12), frequency = 12)
4 dades.ts<-dadesany[,3]
```

1.1 Dibuixeu la sèrie. És un procés estacionari? Perquè?

```
1 plot(dades.ts, main = "Producció d'electricitat solar tèrmica a Catalunya",
       ylab = "MWh", xlab = "Temps en anys")
```

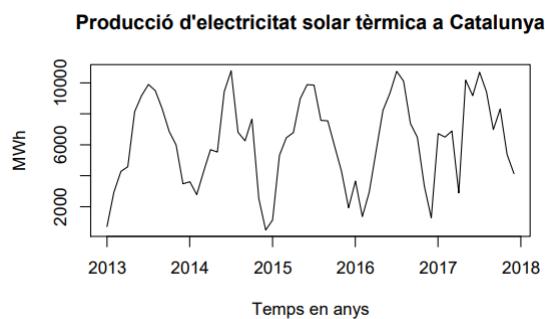


Figura 64: Producció d'electricitat solar tèrmica a Catalunya

Dibuixem la sèrie fent un plot de les dades. La sèrie no sembla que presenti tendència. Sí s'observa certa estacionalitat. Per exemple, podem observar que la producció d'electricitat solar tèrmica a Catalunya sempre serà majors en els mesos d'estiu.

L'estacionalitat sol provocar que la sèrie sigui no estacionària, ja que els valors mitjans en alguns moments particulars dins del període estacional (mesos, per exemple) poden ser diferents als valors mitjans en altres ocasions, i això passa en el nostre cas.

Per estudiar si la sèrie és estacionària també podem utilitzar el test de Dickey-Fuller que busca determinar l'existència o no d'arrels unitàries en una sèrie de temps. La hipòtesi `nul.la` d'aquesta prova és que hi ha una arrel unitària en la sèrie, i per tant, no seria estacionària. I com hipòtesis alternativa la sèrie seria estacionària.

```
1 adf.test(dades.ts)
2 ##
3 ## Augmented Dickey-Fuller Test
4 ##
5 ## data: dades.ts
6 ## Dickey-Fuller = -5.0957, Lag order = 3, p-value = 0.01
7 ## alternative hypothesis: stationary
```

En el cas de la nostra sèrie veiem que ens ha donat un p-valor de 0.01 , per tant més petit que 0.05 (nivell de significació habitual). Amb un 95 % de confiança rebutjaríem la hipòtesis `nul.la` de que la sèrie no es estacionaria.

Però en aquest cas el test de Dickey-Fuller és enganyós. Quan es fa el test de Dickey-Fuller, no arriba a analitzar-ho en el lag 12 (que és la nostra estacionalitat) i per tant ens diu que és un procés estacionari quan en realitat no ho és.

Si fem la funció auto.arima (veiem que ens detecta una estacionalitat de lag = 12.

```

1 auto.arima(dades.ts)
2 ## Series: dades.ts
3 ## ARIMA(1,0,0)(1,1,0) [12]
4 ##
5 ## Coefficients:
6 ##             ar1      sar1
7 ##             0.2822  -0.6373
8 ## s.e.    0.1409   0.1169
9 ##
10 ## sigma^2 estimated as 2688692: log likelihood=-425.57
11 ## AIC=857.13 AICc=857.68 BIC=862.74

```

A més podem observar que demana fer la diferència d'ordre 12 per tal de ser estacionaria. Veiem que en aquest cas la part no estacional (p, d, q) no demana cap diferència, però en canvi la part estacional (P, D, Q) si que demana fer la diferència estacional per tal de que sigui estacionaria.

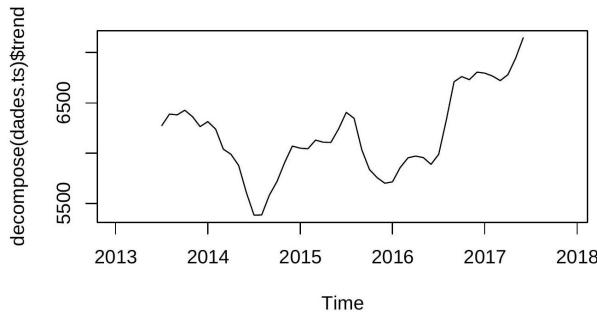
Concloem doncs que és una sèrie no estacionaria.

1.2 Té tendència?

```

1 plot(decompose(dades.ts)$trend)

```



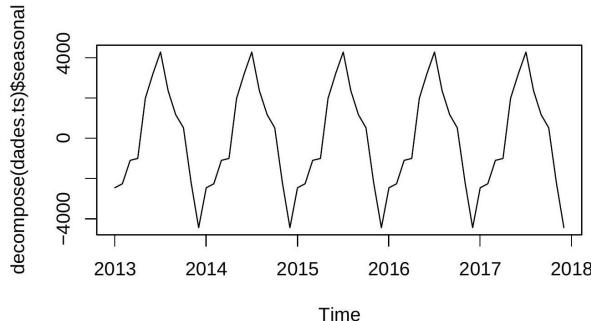
Mirant al primer gràfic ja havíem vist que la sèrie no semblava tenir cap tendència clara. Si fem decompose podem obtenir el gràfic de la tendència *i* tornem a observar que no sembla que segueixi cap model ni lineal ni quadràtic. A més si mirem els resultats obtinguts anteriorment amb autoarima veiem que no detecta drift.

1.3 Té estacionalitat? Si en té, quin et sembla que és el període del cicle?

```

1 plot(decompose(dades.ts)$seasonal)

```



Mirant la gràfica es podria dir que manté els efectes estacionals. Té estacionalitat ja que veiem un patró habitual de canvis que es repeteix durant l'any (o bé 12 mesos)

1.4 Quina és l'estacionalitat que proposa auto.arima () ?

```

1 auto.arima(dades.ts)
2 ## Series: dades.ts
3 ## ARIMA(1,0,0)(1,1,0)[12]
4 ##
5 ## Coefficients:
6 ##          ar1      sari
7 ##      0.2822 -0.6373
8 ##  s.e.  0.1409  0.1169
9 ##
10 ## sigma^2 estimated as 2688692: log likelihood=-425.57
11 ## AIC=857.13 AICc=857.68 BIC=862.74

```

Amb autoarima com ja havíem comentat per sobre anteriorment podem observar que l'estacionalitat proposada és de 12 mesos.

1.5 Quin model proposa auto.arimal per a la sèrie i per a la part estacional? Escriu explícitament el model proposat i comenta els aspectes que et semblin més relevant.

El model proposat per autoarima inclou un terme AR(1) no estacional per a la sèrie i un terme AR(1) estacional i una diferència per la part estacional. El model que mostra és sense termes MA i el període estacional és $S = 12$.

Per la part no estacional

$$\text{AR}(1) : \varphi(B) : 1 - \varphi_1 B = 1 - 0,2822B$$

Per la part estacional

$$\text{SAR}(1) : \phi(B^{12}) : (1 - \phi_{12} B^{12})(1 - B^{12}) = (1 + 0,6373 B^{12})(1 - B^{12})$$

Finalment el model ARIMA $(1,0,0)(1,1,0)_{12}$ queda:

$$(1 - \varphi_1 B)(1 - \phi_{12} B^{12})(1 - B^{12}) X_t = \varepsilon_t$$

$$(1 - 0,2822B)(1 + 0,6373 B^{12})(1 - B^{12}) X_t = \varepsilon_t$$

Per tant hem pogut observar que la part no estacional és estacionaria i en canvi la part estacional no és estacionaria, ja que demana fer la diferència per tal de ser estacionaria fet que és bastant rellevant.

1.6 Calcula les diferències proposades per auto.arima 0 i comprova que la nova sèrie és estacionària. Hi ha alguna diferència entre el model proposat per a la sèrie original i per a la diferenciada? Si n'hi ha, quines son?

Per suavitzar l'estacionalitat es pot diferenciar les dades. L'estacionalitat sol provocar que la sèrie sigui no estacionària, ja que els valors mitjans en alguns moments particulars dins del període estacional poden ser diferents. Per tant examinem les dades diferenciades amb lag 12 concretament per la nostra sèrie i obtenim el següent gràfic, que en aquest cas si que és estacionari.

```

1 dades.ts_diff <- diff(dades.ts, lag = 12)
2 plot(dades.ts_diff, main = "Producció d'electricitat solar tèrmica a
  Catalunya", ylab = "MWh", xlab = "Temps en anys")

```

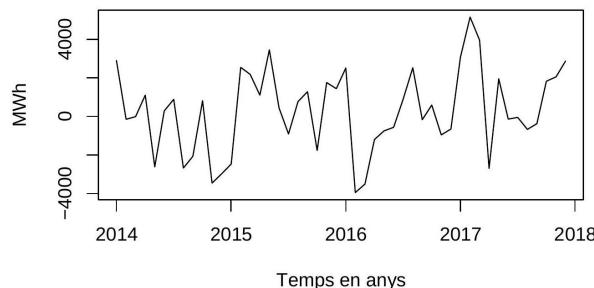


Figura 65: Producció d'electricitat solar tèrmica a Catalunya

```

1 adf.test(dades.ts\_diff)
2 ##
3 ## Augmented Dickey-Fuller Test
4 ##
5 ## data: dades.ts\_diff
6 ## Dickey-Fuller = -3.6411, Lag order=3,
7 p-value = 0.03945
8 ## alternative hypothesis: stationary

```

El test de Dickey - Fuller torna a donar un p-valor més petit que 0.05, per tant rebutjaríem la hipòtesis **nul.la** de que la sèrie no es estacionaria. Abans de fer la diferència també ens havia donat aquest resultat però intuíem que era enganyós. En aquest cas sabem que és cert que és estacionaria.

```

1 auto.arima(dades.ts_diff)
2 ## Series: dades.ts_diff
3 ## ARIMA(1,0,0)(1,0,0)[12] with zero mean
4 ##
5 ## Coefficients:
6 ##             ar1      sar1
7 ##           0.2822   -0.6373
8 ## s.e.    0.1409    0.1169
9 ##
10 ## sigma^2 estimated as 2688692: log likelihood=-425.57
11 ## AIC=857.13 AICc=857.68 BIC=862.74

```

La funció `auto.arima()` ens proposa un model que inclou un terme AR(1) no estacional per a la sèrie i un terme AR(1) estacional per la part estacional. El model que mostra és sense termes MA i el període estacional és $S = 12$. Veiem que en aquest cas ens explica que és amb mitjana 0 i observem que ja no ens demana fer cap diferència per la part estacional. També observem que tots els altres valors són exactament iguals inclosos els valors AIC, AICc i BIC. Així el model després de fer la diferència quedaría:

Per la part no estacional

$$AR(1) : \varphi(B) : 1\varphi_1B = 10,2822B$$

Per la part estacional

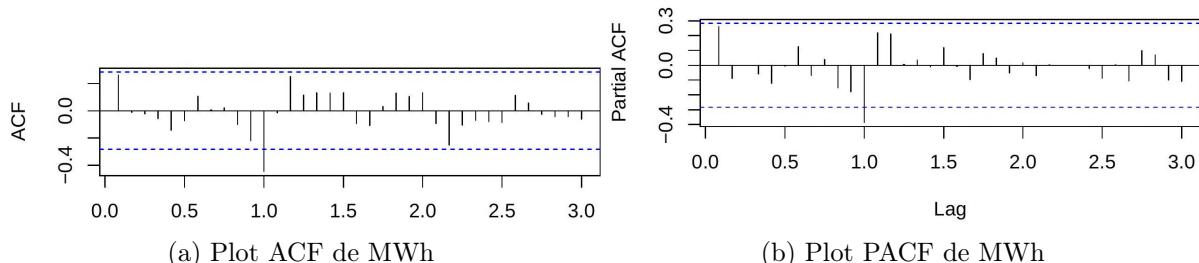
$$SAR(1) : \phi(B^{12}) : (1 - \phi_{12}B^{12}) = (1 + 0,6373B^{12})$$

Finalment el model ARIMA $(1,0,0)(1,0,0)_{12}$ queda: $(1 - \varphi_1B)(1 - \phi_{12}B^{12})X_t = \varepsilon_t (1 - 0,2822B)(1 + 0,6373B^{12})X_t = \varepsilon_t$

```

1 par(mfrow=c(2,1))
2 acf(dades.ts_diff, main = "Plot ACF de MWh", lag.max = 36)
3 pacf(dades.ts_diff, main = "Plot PACF de MWh", lag.max = 36)

```



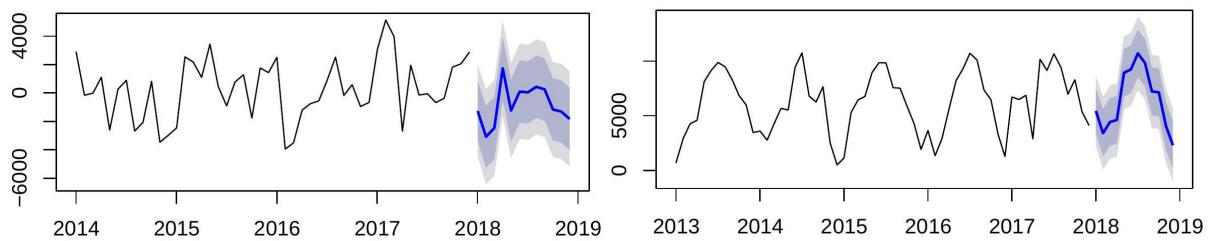
Fent ACF i PACF podem observar el que autoarima ja ens havia dit. Veiem que en el AR(1) trobem un lag signi catiu en el lag 12 ja que és el nostre període.

1.7 Fent servir la funció `forecast()` doneu una estimació per als valors de l'any 2018 i dibuixe la predicció

```

1 par(mfrow=c(2,1))
2 ##### Fent servir la llibreria forecast
3 mod1 <- auto.arima(dades.ts_diff)
4 pred1 <- forecast(dades.ts_diff, h = 12, model = mod1)
5 plot(pred1,main = "Estimació per la sèrie diferenciada")
6 mod2 <- auto.arima(dades.ts)
7 pred2 <- forecast(dades.ts, h = 12, model = mod2)
8 plot(pred2,main = "Estimació per la sèrie original")

```



(a) Estimación para la serie diferenciada

(b) Estimación para la serie original

Figura 67: Comparación de estimaciones para la serie

Com l'enunciat no especifica de quines dades s'havia de fer l'estimació, mostrem aquesta estimació per a la sèrie diferenciada i per a la sèrie original. Això ens ha permès veure que les estimacions són diferents depenent del model que utilitzem.

1.8 Quedeu-vos amb les observacions de 2013 al 2016 i genereu una predicció per als 20 valors restants basada en el model ajustat anteriorment. Dibuixe la sèrie real juntament amb les prediccions. Què diríeu sobre la bondat de l'ajust del model utilitzat? En quants punts els valors reals queden fora de l'interval de confiança del 95 % de les prediccions?

```

1 o1316 <- ts(dades, start = c(2013, 1), end = c(2016, 12), frequency = 12)
2 o1316 <- o1316[,3]
3 pred3 <- forecast(o1316, h = 20, level = 95, model = mod2)
4 plot(pred3, lwd = 5)
5 lines(dades.ts, col = "red", lwd = 1, type = "l")
6 cnt <- 0
7 for(i in 1:12){
8   if(dades.ts[49:60][i]<pred3$lower[i] | dades.ts[49:60][i] > pred3$upper[i]){
9     cnt <- cnt + 1
10  }
11 else{
12   cnt <- cnt
13  }
14 }
15 cnt
16 ## [1] 2

```

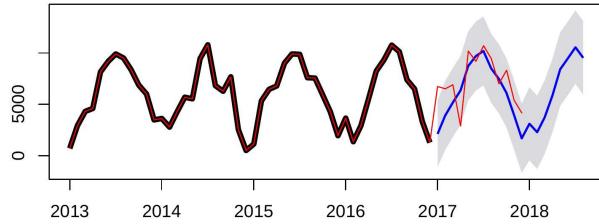


Figura 68: Forecasts from ARIMA(1,0,0)(1,1,0)[12]

Un cop dibuixat el gràfic amb les prediccions i els valors reals veiem que és un model que s'ajustaria prou bé ja que la majoria de valors reals queden dintre de l'interval de confiança del 95 %. A més a més hem creat una funció per tal de comptar quants punts de les dades reals queden fora, i en el cas de la nostra base de dades només dos punts queden fora. Per tant un 10 %. Semblaria un bon ajust, però s'hauria de tenir en compte que estem utilitzant molt poques dades i s'hauria d'observar que passa si agaféssim una base de dades més extensa.

9.2. Glosario

1. Un **buen ajuste a los datos de entrenamiento no siempre** es una indicación de que el modelo **pronosticará bien**.

Ejemplo: En la referencia 8.10 ARIMA vs ETS | Forecasting: Principles and Practice (otexts.com) encontramos la siguiente salida de R; donde se ha realizado una medición de la bondad de ajuste a un serie realizada con dos modelos distintos: Un modelo arima y un modelo ets (suavizado exponencial).

Para realizar el ajuste a la serie se ha separado la muestra real en dos submuestras:

- a) Training set como muestra de entrenamiento de la serie.
- b) Test set como muestra para testear si el ajuste predice bien la serie real.

En la siguiente salida si nos fijamos en los valores que toma el RMSE (raíz cuadrada del MSE donde MSE es el error cuadrático medio calculado como la media de los errores al cuadrado) observamos que:

- Para la muestra Training test el RMSE es menor para el modelo arima que para el modelo ETS.
- Mientras que para la muestra Test set el RMSE es menor en el modelo ETS.

Así concluimos que el modelo arima ajusta mejor la serie de entrenamiento, mientras que la serie ETS ajusta mejor en las predicciones.

```
# Generate forecasts and compare accuracy over the test set
a1 <- fit.arima %>% forecast(h = 4*(2013-2007)+1) %>%
  accuracy(qcement)
a1[,c("RMSE","MAE","MAPE","MASE")]
#>          RMSE      MAE     MAPE    MASE
#> Training set 0.1001 0.07989 4.372 0.5458
#> Test set     0.1996 0.16882 7.719 1.1534
a2 <- fit.ets %>% forecast(h = 4*(2013-2007)+1) %>%
  accuracy(qcement)
a2[,c("RMSE","MAE","MAPE","MASE")]
#>          RMSE      MAE     MAPE    MASE
#> Training set 0.1022 0.07958 4.372 0.5437
#> Test set     0.1839 0.15395 6.986 1.0518
```

2. La función *coeftest* toma por defecto como distribución del estadístico para los coeficientes del modelo la normal ($N(0,1)$).

Puesto que el vector de la variable aleatoria de la serie (los residuos) sigue una distribución normal en el modelo arima, el vector de los coeficientes de la variable también lo harán.

Por lo que $b \rightarrow N(\beta, \text{Var}(\beta))$. Así, dado el siguiente estadístico sigue $\frac{(b-\beta)}{S_b} \rightarrow t_{n-k}$.

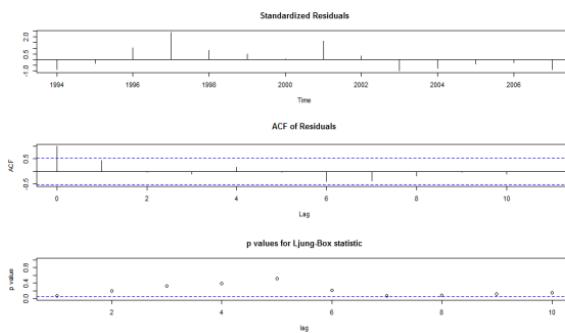
Si n es suficientemente grande, $t_{n-k} \rightarrow N(0, 1)$.

Si en la función `coeftest` se desea que se utilice la t-student para una mejor aproximación cuando la muestra es pequeña, se debe indicar los grados de libertad (df) dentro de la función `coeftest(mod1, df=x)`.

3. En el ejercicio 3 de esta práctica realizamos una regresión donde los residuos siguen un modelo ARIMA. En el siguiente enlace podéis encontrar más información al respecto: *9.2 Regression with ARIMA errors in R / Forecasting: Principles and Practice (otexts.com)*

4. Test de los residuos vistos en la práctica.

La función `tsdiag(mod)` nos ofrece la siguiente salida que incluye la visualización de los residuos estandarizados (*standarized Residuals*), el correlograma de los residuos (*ACF of Residuals*) y los *p*-valores del estadístico *Ljung-Box*.



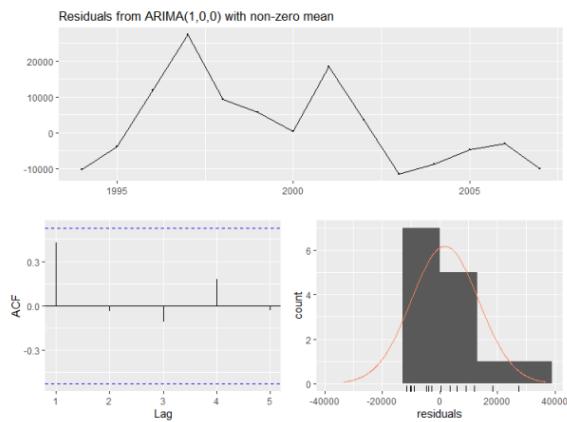
La prueba de Ljung-Box se puede definir de la siguiente manera.

H_0 : Los datos se distribuyen de forma independiente (es decir, las correlaciones en la población de la que se toma la muestra son 0, de modo que cualquier correlación observada en los datos es el resultado de la aleatoriedad del proceso de muestreo).

H_a : Los datos no se distribuyen de forma independiente.

Por lo que el gráfico nos ofrece una prueba de independencia entre los residuos en los diferentes lags. Si todas las bolitas están por encima de la banda, aceptamos la hipótesis de independencia.

5. La función `checkresiduals(mod)` nos ofrece la siguiente salida que incluye la visualización del gráfico de los residuos, el correlograma de los residuos y el histograma con la curva normal ajustada de los residuos.



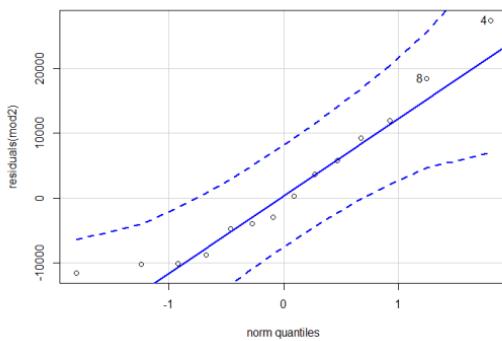
También incluye la prueba de Ljung-Box donde se indica el total de lags utilizados.

```
Ljung-Box test
data: Residuals from ARIMA(1,0,0) with non-zero mean
Q* = 4.1608, df = 3, p-value = 0.2446
Model df: 2. Total lags used: 5
```

6. La función `Box.test()` para realizar el test Box-Pierce de independencia de los residuos de la serie.
7. El gráfico QQPlot estudia la normalidad de los datos:

```
1   qqPlot(residuals(mod), dist="norm", id=list(method="y", n=2, labels=
rownames(DataSet)))
```

Cuánto más cerca estan los puntos a la recta diagonal más parecida es la distribución a una normal. Si todos los puntos se encuentran dentro de las bandas nos indica con una confianza del 95 % que es una normal.



8. El *test de Kolmogorov-Smirnov* nos mide la normalidad de los datos (H_0 : Los datos se distribuyen como una normal) viene dado por la función `ks.test`.