

Anàlisi de Dades Financeres

UNIVERSITAT AUTÒNOMA DE BARCELONA

ENTREGA 1

Informe

Abril Pérez Martí - 1600601

Arnau Perich Iglesias - 1603567

Eric Jiménez Barril - 1599092

Joan González Martínez - 1597201

Laia Escursell Rof - 1600578

13 d'octubre del 2023

Exercici 1:

a) Descriu els signes de no estacionarietat que es veuen a les sèries temporals i als diagrames ACF.

Diem que un procés estocàstic és estacionari quan aquest té un comportament invariant envers el temps. En particular, en aquesta pràctica treballarem amb el que es coneixen com processos estacionaris febles. Ho definim: sigui $\{Y_n\}$ un procés estocàstic, diem que $\{Y_n\}$ és estacionari feble si tenim que

$$\mathbb{E}(Y_i) = \mu \quad \text{Var}(Y_i) = \sigma^2 \quad \text{Corr}(Y_i, Y_j) = \rho(|i - j|)$$

per tot i, j i per alguna funció ρ .

A continuació, descriurem, de manera qualitativa (veure Figura ??), algunes característiques de les nostres dades que ens portaran a pensar que no són estacionàries. En primer lloc, observem que les dades corresponents a r (“the 91-day treasury bill rate”) no poden ser estacionàries ja que no tenen ni mitjana ni variància constant. Fins als anys 80 mostren una tendència creixent i dels 80 en endavant una decreixent. D’altre banda, les dades corresponents a y (“the log of real GDP”) tenen una tendència creixent durant tots els anys; i per tant, la mitjana no és constant, fet que ens permet concloure que tampoc són estacionàries. Per últim, les dades corresponents a π (“the inflation rate”) no tenen una variància constant en el temps i llavors, tampoc poden ser estacionàries.

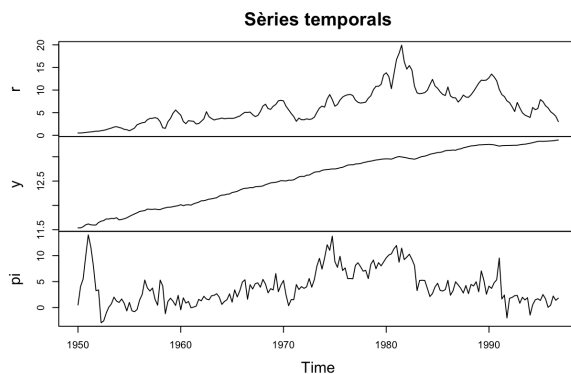


Figura 1: Representació de les sèries temporals r , y i π .

Per estudiar l’estacionarietat amb més detall i precisió podem usar els gràfics ACF d’aquestes sèries (veure Figura 2). El gràfic ACF d’una sèrie estacionària es caracteritza per decaure a zero de manera ràpida. Observem que els tres gràfics tenen una tendència decreixent cap a zero, però de manera lenta i no s’acaba d’assolir mai aquest valor; fet que reafirma el que ja pensàvem: aquestes sèries no son estacionaries.

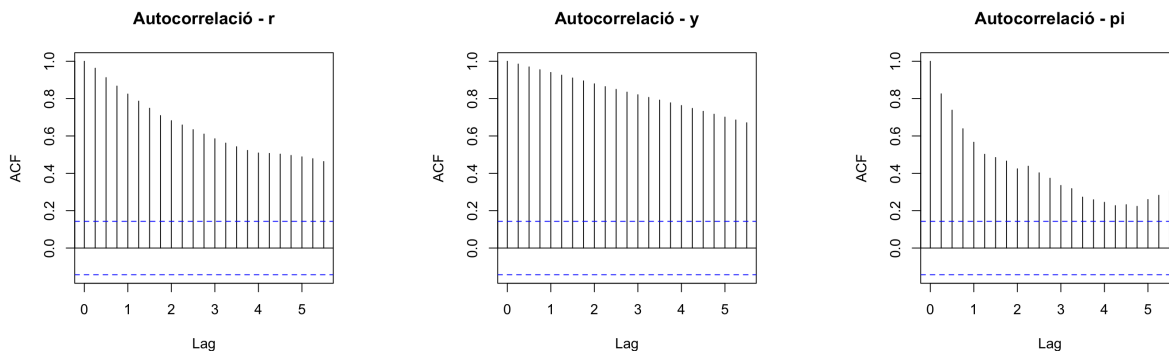


Figura 2: Autocorrelació de les sèries temporals r , y i π .

b) Utilitzeu les proves de Dickey-Fuller augmentades per decidir quines de les sèries no són estacionàries. Les proves corroboren el conclusions de les gràfics de la sèrie temporal ACF?

El *Augmented Dickey-Fuller Test* és un test d'hipòtesi per decidir si una sèrie temporal és estacionària. Formalment, es prova la hipòtesi nul·la:

$$H_0 = \text{Les dades són no estacionàries}$$

en contra l'alternativa

$$H_1 = \text{Les dades són estacionaries}$$

amb un nivell de significació del 5%. Realitzem el test descrit i presentem el resultats obtinguts pels diferents p-valors a la Taula 1. Observem que a les tres proves obtenim un p-valor superior a 0.05. I per tant no tenim evidències significatives, amb un nivell de significació del 5%, per rebutjar la hipòtesi nul·la. És a dir, les dades són no estacionàries, tal i com hem vist al primer apartat.

Test 1: r	Test 2: y	Test 3: pi
0.6075	0.9873	0.1788

Taula 1: p-valors obtinguts al *Augmented Dickey-Fuller Test* per les sèries temporals r , y i pi .

c) La sèrie de diferències apareix estacionària segons la prova augmentada de Dickey-Fuller?

L'operador diferencial Δ sobre una sèrie temporal $\{Y_n\}$ es defineix com l'operador:

$$\Delta Y_t = Y_t - Y_{t-1}$$

A continuació, considerem la sèrie diferenciada per cada una de les tres sèries anteriors i tornem a realitzar un *Augmented Dickey-Fuller Test* per a les noves dades. Obtenim tots els p-valors inferiors a 0.01. Per tant, tenim evidències significatives per rebutjar la hipòtesi nul·la. És a dir que, en aquest cas, cadascuna de les tres sèries diferenciades són estacionàries.

De manera visual, observem que els tres gràfics (Figura 3) tenen unes característiques similars. Primerament, notem que les dades tenen una esperança fixa, és a dir, oscil·len respecte un valor fix. A més aquesta variació en les dades és similar al llarg del temps. Per tant, podem pensar que la variància també és constant. Per últim, si mirem els gràfics ACF (veure Figura 4), podem observar com la autocorrelació decreix ràpidament cap al 0. Justament, aquestes tres característiques són pròpies d'una sèrie temporal estacionària. Aquest resultat corrobora el que hem vist amb el *Augmented Dickey-Fuller Test*.

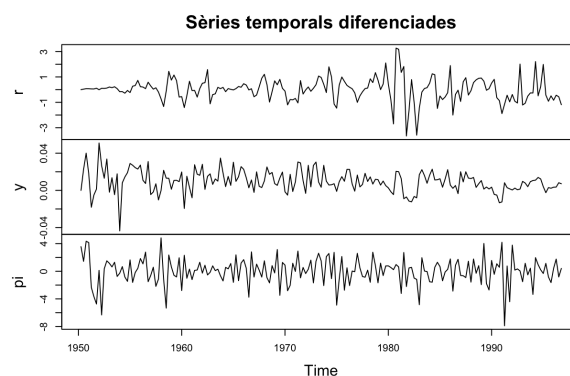


Figura 3: Representació de les sèries temporals diferenciades r , y i pi .

d) Veieu proves d'autocorrelació a les sèries diferenciades? Si és així, descriu aquestes correlacions.

A continuació, busquem indicis d'autocorrelació en les nostres dades. Per fer això, analitzem els gràfics ACF (Figura 4). Podem observar com Y_t i Y_{t-1} estan lleugerament correlacionades (la correl·lació està

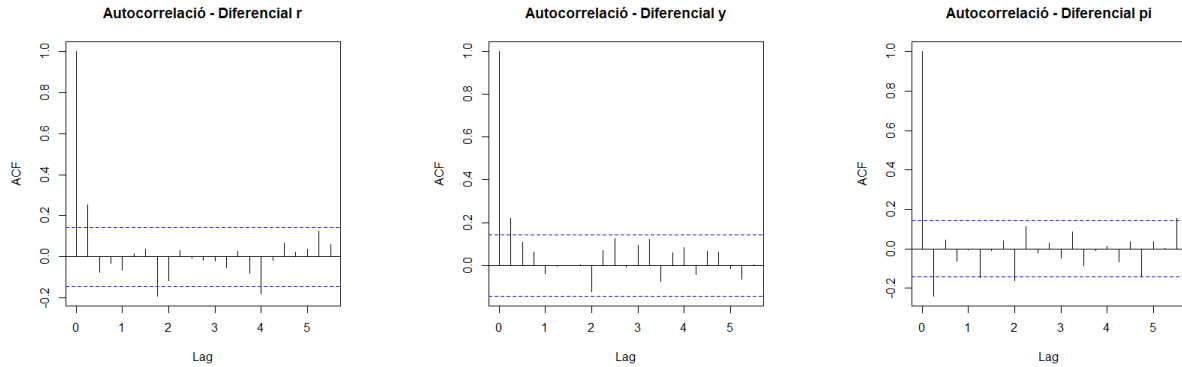


Figura 4: Autocorrelació de les sèries temporals diferenciades r , y i π .

fora de l'interval de confiança del zero), mentre que, a partir de Y_{t-2} la correlació és zero (el valor sí que està dins l'interval). És a dir, l'autocorrelació decau bruscament cap al 0, però tot i així, no podem negligir la correlació amb el terme anterior; aquest comportament es característic dels models $MA(1)$.

NOTA: A partir d'ara, només treballarem amb les dades corresponents a les lletres del tresor: r .

e) Quin ordre de diferenciació s'escull a `auto.arima`? Aquest resultat coincideix amb el vostre anterior conclusions?

Tot seguit, cerquem, mitjançant la funció '`auto.arima()`', el model que millor s'ajusta a les nostres dades segons el criteri AIC. Recordem que aquesta funció ajusta les nostres dades a un model $ARIMA(p, d, q)$.

Diem que un procés estocàstic $\{Y_n\}$ és un $ARIMA(p, d, q)$ si $\Delta^d Y_t$ és un $ARMA(p, q)$, on Δ es correspon amb l'operador diferencial. D'altra banda, per $\{\varepsilon_i\}_i$ un $WN(0, \sigma_\varepsilon^2)$ ¹, diem que un procés estocàstic $\{Y_n\}_n$ és un $ARMA(p, q)$ si per certs paràmetres constants μ , $\{\phi_i\}_{i=1}^p$ i $\{\theta_j\}_{j=1}^q$, es compleix que:

$$Y_t - \mu = \sum_{i=1}^p [\phi_i (Y_{t-i} - \mu)] + \varepsilon_t + \sum_{j=1}^q [\theta_j \varepsilon_{t-j}] \quad (1)$$

En particular, aquest model escull $d = 1$ com a ordre de diferenciació, tal i com esperàvem, doncs hem vist que la sèrie original no és estacionària però si la diferenciem un cop sí que ho és.

f) Quin model va escollir l'AIC?

L'AIC va escollir el model $ARIMA(0, 1, 1)$, és a dir, podem ajustar la nostra sèrie $\{Y_n\}_n$ segons un model de la forma:

$$\Delta Y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$$

on μ i θ son paràmetres constants i $\{\varepsilon_i\}_i$ un és $WN(0, \sigma_\varepsilon^2)$.

g) Quin criteri de bondat d'ajust s'utilitza aquí?

Tal i com hem mencionat als apartats anteriors, el criteri utilitzat ha estat l'AIC, pel qual hem obtingut un valor de 495.3.

h) Canvia el criteri a BIC. Aleshores, canvia el model més adequat?

Si fem servir el criteri BIC per construir el nostre model mitjançant la funció '`auto.arima()`' obtenim el mateix resultat: el model que millor s'ajusta a les nostres dades és un $ARIMA(0, 1, 1)$. En aquest cas, el valor donat per BIC ha estat de 501.76.

¹Un $WN(\mu, \sigma_\varepsilon^2)$ o '*White Noise*' és un procés estacionari feble tal que:

$$\mathbb{E}(Y_i) = \mu \quad \text{Var}(Y_i) = \sigma^2 \quad \text{Corr}(Y_i, Y_j) = \delta_{ij}$$

on δ_{ij} correspon a la *Delta de Kronecker*.

i) Creus que hi ha autocorrelació residual? Si és així, descriu aquesta autocorrelació i suggereix un model adequat per a la sèrie "T-bill".

Per saber si existeix autocorrelació farem servir el *Ljung-Box-test*, que és una prova que verifica si una sèrie temporal té autocorrelació. És a dir, prova la hipòtesi nul·la

$$H_0 : \text{Els residus tenen correlació } 0.$$

en contra de l'alternativa

$$H_1 : \text{Els residus tenen correlació}$$

A continuació, realitzem aquest test amb un nivell de significació del 5% pels residus del model escollit $\text{ARIMA}(0,1,1)$ i obtenim un p-valor de 0.1618. Com que el p-valor > 0.05 , no tenim evidències significatives per rebutjar la hipòtesi nul·la. És a dir, podem concloure que les dades són independents i per tant no presenten autocorrelació.

Analitzant el gràfic ACF (Figura 5) podem observar indicis d'aquesta independència en els residus: la correlació entre Y_t i Y_{t-i} és pràcticament 0, ja que aquests valors estan dins l'interval de confiança del zero. Per tant, corrobora el resultat obtingut amb el *Ljung-Box-test*.

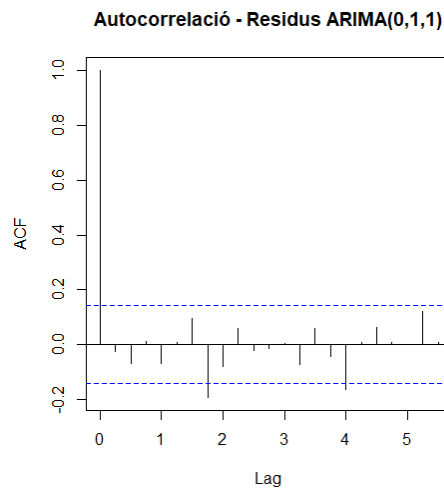


Figura 5: Autocorrelació de la sèrie temporal dels residus del model $\text{ARIMA}(0,1,1)$

Exercici 2:

Cosidereu el model AR:

$$Y_t = 5 - 0.55Y_{t-1} + \varepsilon_t$$

i assumiu $\sigma_\varepsilon^2 = 1.2$.

a) És el procés estacionari? Per què?

Per inspecció, podem afirmar que tenim un procés $\text{AR}(1)$. Sabem que aquest tipus de procés és estacionari feble si i només si $|\phi| < 1$, és a dir, $-1 < \phi < 1$. En aquest cas tenim que $\phi = -0.55$, que compleix la condició. Per tant, podem afirmar que el procés és estacionari feble.

b) Quina és la mitjana del procés?

L'equació d'un procés $\text{AR}(1)$ és de la forma:

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t.$$

Aquesta equació la podem reescriure com:

$$Y_t = \mu(1 - \phi) + \phi Y_{t-1} + \varepsilon_t,$$

d'on obtenim que:

$$\mu(1 - \phi) = 5 \iff \mu = \frac{5}{1 - \phi} \approx 3.2258$$

c) Quina és la variància del procés?

Sabem que, pels processos estacionaris febles, es compleix la següent equació:

$$\sigma_Y^2 = \phi^2 \sigma_Y^2 + \sigma_\epsilon^2,$$

on σ_Y^2 és la variància de Y_t . Aïllant-la obtenim:

$$\sigma_Y^2 = \frac{\sigma_\epsilon^2}{1 - \phi^2}$$

i substituint els valors que ja coneixem obtenim:

$$\sigma_Y^2 = \frac{1.2}{1 - (-0.55)^2} \approx 1.7204$$

d) Quina és la funció de covariància del procés?

Tenim que la covariància del procés ve donada per:

$$\gamma(h) = Cov\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}, \sum_{j=0}^{\infty} \phi^j \epsilon_{t-h-j}\right) = \frac{\sigma_\epsilon^2}{1 - \phi^2} \phi^{|h|}$$

on $h \in \mathbb{Z}$. Substituint obtenim:

$$\gamma(h) = \frac{1.2}{1 - 0.55^2} (-0.55)^{|h|}$$

Exercici 3:

Un model AR(3) s'ha ajustat a una sèrie temporal. Les estimacions són

- $\hat{\mu} = 104$
- $\hat{\phi}_1 = 0.4, \hat{\phi}_2 = 0.25$ i $\hat{\phi}_3 = 0.1$

Les últimes quatre observacions han sigut $Y_{n-3} = 105, Y_{n-2} = 102, Y_{n-1} = 103$ i $Y_n = 99$.

Prediu Y_{n+1} i Y_{n+2} .

Tenim un AR(3), que ve donat per l'equació següent:

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \phi_3(Y_{t-3} - \mu) + \epsilon_t$$

Volem obtenir el valor de Y_t per $t = n + 1$; per tant, ens queda:

$$Y_{n+1} = \mu + \phi_1(Y_n - \mu) + \phi_2(Y_{n-1} - \mu) + \phi_3(Y_{n-2} - \mu) + \epsilon_{n+1}$$

Però com que ϵ_{n+1} és un soroll blanc (independent del seu passat i de tots els Y_t), la millor predicció per ϵ_{n+1} serà la seva mitjana, *i.e.*, 0. Conseqüentment, obtenim que:

$$\hat{Y}_{n+1} = \hat{\mu} + \hat{\phi}_1(Y_n - \hat{\mu}) + \hat{\phi}_2(Y_{n-1} - \hat{\mu}) + \hat{\phi}_3(Y_{n-2} - \hat{\mu}),$$

i substituint els valors donats per l'enunciat trobem el valor de $Y_n + 1$:

$$\hat{Y}_{n+1} = 104 + 0.4(99 - 104) + 0.25(103 - 104) + 0.1(102 - 104) = 101.55$$

Utilitzem el mateix procediment per predir el valor de Y_{n+2} :

$$\begin{aligned} \hat{Y}_{n+2} &= \hat{\mu} + \hat{\phi}_1(Y_{n+1} - \hat{\mu}) + \hat{\phi}_2(Y_n - \hat{\mu}) + \hat{\phi}_3(Y_{n-1} - \hat{\mu}) \\ \hat{Y}_{n+2} &= 104 + 0.4(101.55 - 104) + 0.25(99 - 104) + 0.1(103 - 104) = 101.67 \end{aligned}$$

Exercici 4:

Considerem el conjunt de dades *Mishkin*, en particular la taxa mensual "T-bill"

a) Utilitzeu les sèries temporals i els diagrames ACF per determinar la quantitat de diferenciació necessària per obtenir una sèrie estacionària.

Comencem observant les gràfiques de la Figura 6, on podem observar l'evolució de la sèrie temporal a estudiar i l'autocorrelació d'aquesta. Notem que observant les dues gràfiques i fent servir els criteris esmentats prèviament per afirmar si una sèrie és estacionària, podem concloure que no ho és. En primer lloc, la gràfica de la sèrie té una tendència creixent, com a mínim fins als anys 80 i per tant la mitjana no és constant (veure línia vermella a la Figura 6). D'altra banda, podem observar que la gràfica té pics que ressalten de la mitjana i per tant la variància tampoc és constant en el temps. Per últim, veiem que la funció d'autocorrelació no decreix a 0 ràpidament. Per tant no pot ser un procés estacionari.

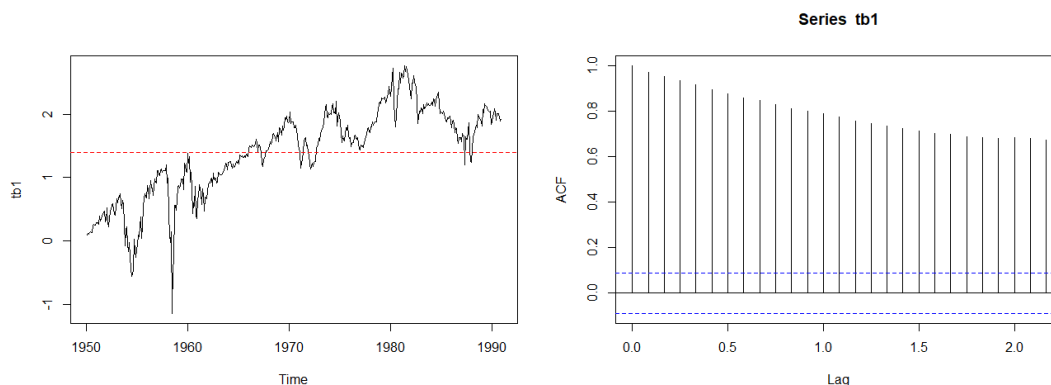


Figura 6: Gràfica de la taxa mensual i funció ACF.

A continuació, estudiem com és la sèrie diferenciada. Representem l'evolució de la nova sèrie i la seva funció d'autocorrelació a la Figura 7. Observem que la gràfica de la sèrie diferenciada té la mitjana constant, com es pot observar amb la línia vermella del primer gràfic, i centrada a 0 i que la variància sembla constant (podem pensar que el pic que hi ha entre els anys 1958-1959 és causat per un esdeveniment aleatori i de gran magnitud, impossible de predir amb els coneixements que tenim actualment²). D'altra banda, veiem com ara l'autocorrelació decau ràpidament cap al 0. Tot això ens porta a pensar que aquesta sí que és una sèrie estacionària. Per assegurar-nos, podem portar a terme un *Augmented Dickey-Fuller Test*. El fem i obtenim un *p*-valor de 0.01. Com és menor a 0.05, ens permet rebutjar la hipòtesi nul·la, que la sèrie sigui no estacionària.

Si fem un estudi més en profunditat del gràfic ACF, observem que la funció correlació és semblant a:

$$\rho_j = \begin{cases} 1 & j = 0 \\ < 0 & j = 1 \\ \approx 0 & j > 1 \end{cases}$$

Això ens porta a pensar que la sèrie diferenciada segueix un model MA(1).

b) A continuació, utilitzeu `auto.arima` per determinar els models ARIMA no estacionals més adequats. Utilitzeu tant AIC com BIC i compareu els resultats.

En primer lloc, la funció "`auto.arima`", basant-se en el criteri del coeficient AIC, ens aconsella aproximar la sèrie per un model ARIMA(3, 1, 5) amb desplaçament. Notem, però, que fent un '*z-test*'³ dels coeficients (veure Taula 2), obtenim que les dues últimes components de les variables corresponents a la

²Posteriorment aprendrem a fer models (GARCH) que siguin capaços de predir variacions en la variança en certs períodes de temps.

³El *z-test* és un test per a cada coeficient del model que prova la hipòtesi nul·la que el coeficient és 0.

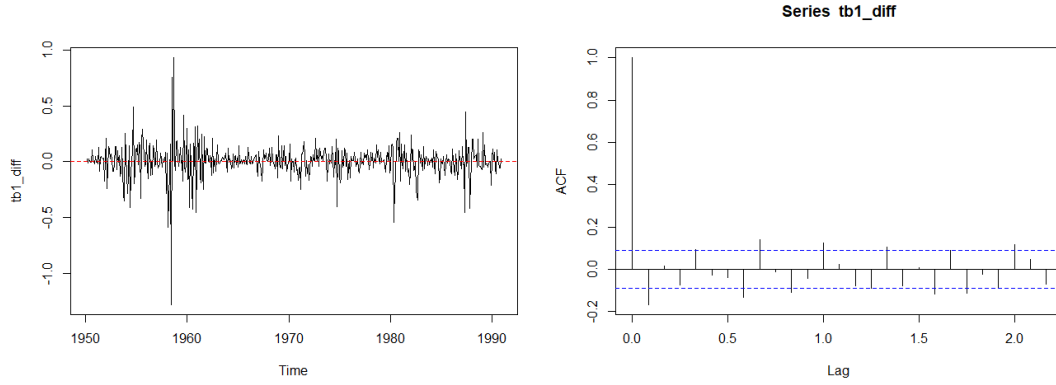


Figura 7: Gràfica de la taxa mensual diferenciada i funció ACF.

mitjana mòbil i el desplaçament no són significatives, i llavors, podem considerar que tenim un model $\text{ARIMA}(3,1,3)$.

	Estimate	Std. Error	z value	Pr(> z)
ar1	-0.9712282	0.0308073	-31.5259	$<2 \times 10^{-16}$
ar2	-0.9642562	0.0284972	-33.8369	$<2 \times 10^{-16}$
ar3	-0.9505994	0.0314635	-30.2128	$<2 \times 10^{-16}$
ma1	0.8191751	0.0545202	15.0252	$<2 \times 10^{-16}$
ma2	0.8711478	0.0657453	13.2504	$<2 \times 10^{-16}$
ma3	0.8006854	0.0759202	10.5464	$<2 \times 10^{-16}$
ma4	-0.0715773	0.0635778	-1.1258	0.2602
ma5	0.0741432	0.0509553	1.4551	0.1457
Desplaçament	0.0036850	0.0057459	0.6413	0.5213

Taula 2: z -test pels coeficients del model $\text{ARIMA}(3,1,5)$

Ara, seguint el criteri BIC, obtenim que el model que millor s'ajusta a les nostres dades és un $\text{ARIMA}(0,1,1)$. De nou, fent el ' z -test' obtenim els resultats que apareixen a la Taula 3, on podem apreciar que el coeficient que acompanya a la variable de la mitjana mòbil és significatiu, i per tant, afirmem que tenim un procés $\text{MA}(1)$

	Estimate	Std. Error	z value	Pr(> z)
ma1	-0.169209	0.044789	-3.7779	0.0001581

Taula 3: z -test pels coeficients del model $\text{ARIMA}(0,1,1)$

Finalment, per escollir amb quin dels models ens quedem, fem una comparativa dels coeficients AIC i BIC obtinguts per cadascun. Podem veure'ls a la Taula 4. Observem que, a priori, l'única conclusió segura que podem extreure és que el nivell de diferenciació és 1. D'altra banda, farem servir el Principi de Parsimònia, per descartar el $\text{ARIMA}(3,1,5)$ (ja que dos dels coeficients del model no eren significatius) i quedar-nos només amb l' $\text{ARIMA}(0,1,1)$ i l' $\text{ARIMA}(3,1,3)$. Finalment, amb la informació que tenim actualment, no podem escollir entre aquests dos models, doncs tot i obtenir valors pels criteris AIC i BIC que són, en conjunt, lleugerament millors pel model $\text{ARIMA}(3,1,3)$, el model $\text{ARIMA}(0,1,1)$ és molt més simple, i per tant, és menys probable que es produeixi *over-fitting* si el triem. Per tant, fem un anàlisi dels residus.

c) Examineu l'ACF dels residus del model que heu seleccionat. Veus algun problema?

Comencem estudiant els residus pel model $\text{ARIMA}(0,1,1)$. Observem, a la Figura 8, que els residus semblen tenir una lleugera correlació entre ells partir d'un cert *lag*. Tot i així, veiem que la mitjana d'aquests és pràcticament 0; com és d'esperar per un bon model.

	AIC	BIC
ARIMA(0,1,1)	-484.95	-476.57
ARIMA(3,1,3)	-505.73	-476.355
ARIMA(3,1,5)	-505.08	-463.14

Taula 4: Comparació dels coeficients “AIC” i “BIC” entre models

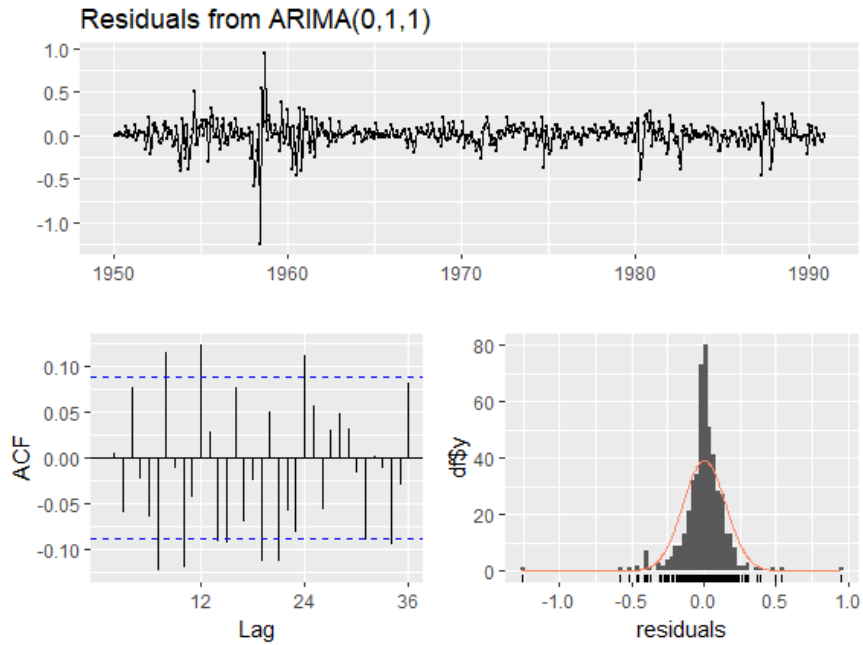


Figura 8: Plot && ACF && Densitat empírica dels residus del model ARIMA(0,1,1)

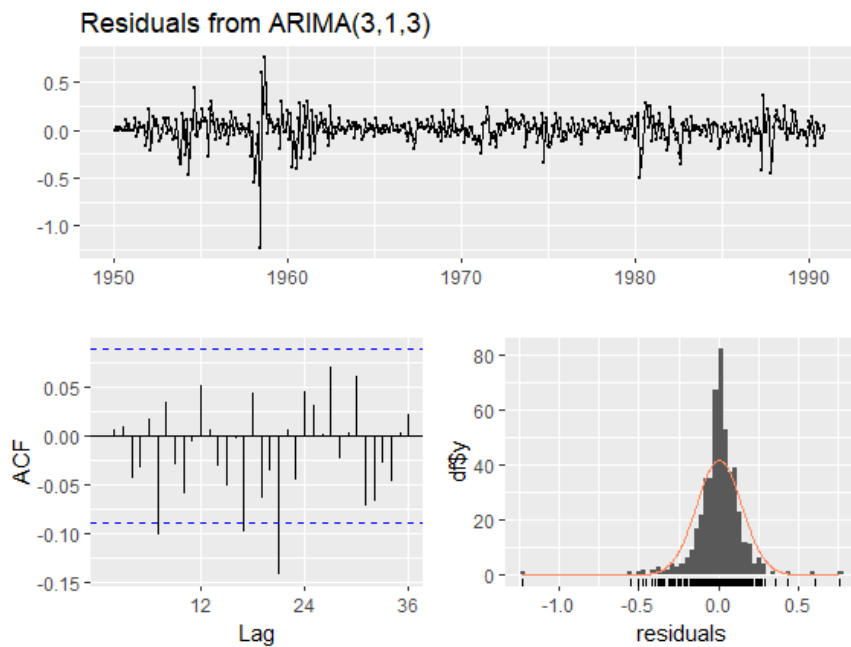


Figura 9: Plot && ACF && Densitat empírica dels residus del model ARIMA(3,1,3)

En canvi, si estudiem els residus pel model ARIMA(3,1,3) (veure Figura 9) veiem com aquests no tenen correlació. A més, la seva mitjana també és aproximadament 0.

A priori, sembla que el model $\text{ARIMA}(3,1,3)$ és millor, doncs no mostra indicis d'autocorrelació als residus, com si ho fa el $\text{ARIMA}(0,1,1)$. Per tal d'assegurar-nos, farem servir un *Ljung-Box-test*. Podem veure els resultats d'aquest test a a Taula 5. Observem que pel model $\text{ARIMA}(0,1,1)$ podem rebutjar la hipòtesi nul·la (ja que el p -valor és menor que 0.05) i per tant els residus no són independents (ja que tenen correlació).

<i>Ljung-Box-Test</i>	p -valor
$\text{ARIMA}(3,1,3)$	0.5654
$\text{ARIMA}(0,1,1)$	5.433×10^{-5}

Taula 5: p -valors obtinguts al *Ljung-box-test*

Aleshores, podem concloure que el model $\text{ARIMA}(3,1,3)$ és el model que millor aproxima la nostra sèrie, tant pels coeficients com per tenir els residus independents.