

Anàlisi del consum global d'energia primària als EEUU

Metodologia Box-Jenkins

Pau Mateo i Pau Fernández

Projecte Sèries Temporals



Grau en Ciència i Enginyaria de Dades
Universitat Politècnica de Catalunya
28 d'abril de 2024

Índex

1	Introducció	2
2	Descrició del dataset	2
3	Metodologia	2
3.1	Models ARMA, (S)ARIMA i ARIMAX	3
4	Resultats	5
4.1	Anàlisi preliminar	5
4.2	Tractament de no-estacionarietat	6
4.3	Identificació	6
4.4	Estimació	7
4.5	Validació	7
4.6	Prediccions	10
4.7	Efectes de Calendari	14
4.8	Tractament d'atípics	19
5	Prediccions a posteriori	28
6	Conclusions	29
	Apèndix	30
	Referències	50

1 Introducció

En aquest estudi s'analitza el consum global d'energia primària mensual als Estats Units des de l'any 1990 fins a l'any 2019. L'objectiu d'aquest estudi és proposar un model vàlid que s'ajusti correctament a les dades i ofereixi una bona predicció dels valors futurs de la sèrie, utilitzant la metodologia Box-Jenkins.

2 Descripció del dataset

Les dades corresponen al consum mensual d'energia total primària dels Estats Units, en milers de milions de Btu (British Thermal Units). Per definició, aquesta unitat correspon a l'energia que cal per a fer pujar la temperatura en un grau Fahrenheit a una lliura d'aigua a pressió atmosfèrica constant. Per donar una referència, 1 unitat de Btu equival a 1054 Joules. La sèrie comença al gener de l'any 1990 i acaba al desembre de l'any 2019.

Les dades són proveïdes per la U.S. Energy Information Administration [1], que és una agència que forma part del U.S. Department of Energy, i l'agència principal del U.S. Federal Statistical System, que s'encarrega de recolectar, analitzar i disseminar informació sobre els diferents tipus d'energia consumida i produïda als Estats Units.

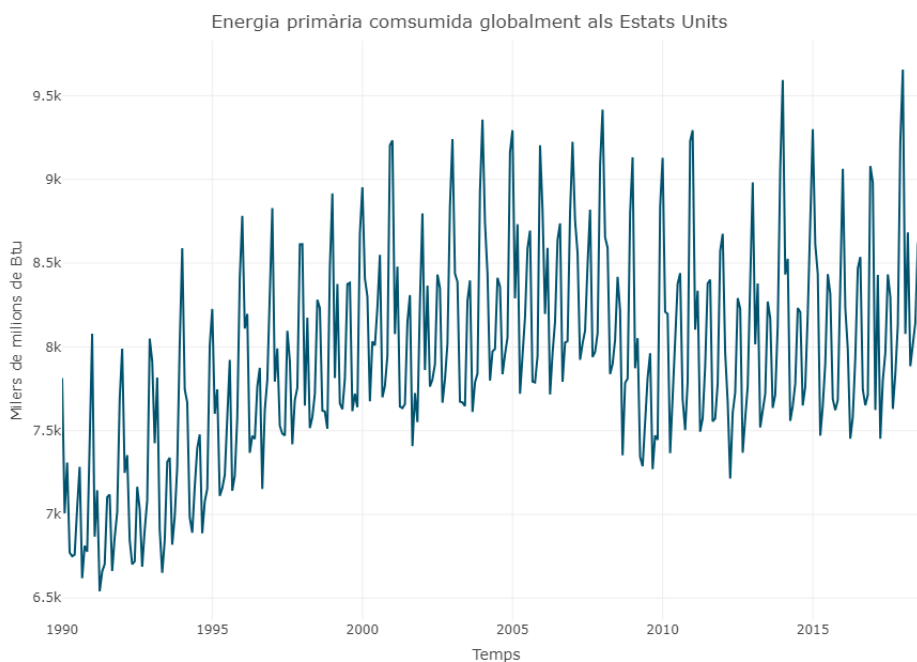


Figura 1: Sèrie temporal EnerUSA

3 Metodologia

Per a l'estudi i tractament de la sèrie temporal i l'ajust de models a la sèrie utilitzarem la metodologia Box-Jenkins, que pren el nom dels estadístics George Box i Gwilym Jenkins, que van introduir aquesta metodologia al 1970 amb la publicació *Time Series Analysis: Forecasting and Control* [2].

La metodologia usa els models ARMA per ajustar un model a les dades de la sèrie i poder obtenir prediccions futures.

Abans d'explicar les metodologies, convé entendre mínimament l'estructura dels models ARMA.

3.1 Models ARMA, (S)ARIMA i ARIMAX

Els models Auto Regressive Moving Average (ARMA) o Auto Regressive Integrated Moving Average (ARIMA) són models estadístics que s'usen per a l'anàlisi de sèries temporals. El terme ARMA es va descriure a la tesi de 1951 de Peter Whittle, *Hypotesis testing in times series analysis* [3], i després es va popularitzar amb la metodologia Box-Jenkins. Aquests models combinen dos models més simples, AR i MA (Auto Regressive i Moving Average, respectivament).

El model AR intenta descriure una observació de la sèrie X_t com a una combinació lineal de p valors anteriors a aquesta:

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

Sent Z_t la diferència entre el valor de X_t i el valor de la combinació de variables mostrada.

El model MA intenta descriure el valor d'una observació X_t com a una combinació lineal de q errors $Z_t - i$ anteriors:

$$X_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q} + Z_t$$

I el model ARMA simplement combina aquests dos models:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q} + Z_t \quad (1)$$

Aquests models però, assumeixen que s'ajusten a una sèrie temporal dèbilment estacionària, ja que sinó hi hauria massa paràmetres per estimar. Una sèrie temporal dèbilment estacionària compleix que la mitjana i variància és constant i que les correlacions entre variables només depèn entre la separació d'aquestes dues (estructura d'autocovariància constant). Per tant, abans de fer res, hem de comprovar que la nostra sèrie és dèbilment estacionària i si no ho és, aplicar transformacions les necessàries per a que ho passi a ser.

A partir d'aquests models se'n deriven els models ARIMA i SARIMA: ((Seasonal) Auto Regressive Integrated Moving Average). Aquests integren també al model les possibles diferenciacions¹, i permeten utilitzar la correlació dels valors de la sèrie de diferents períodes; és a dir, aprofiten l'estacionalitat de la sèrie (en cas que en tingui). El model, en funció de l'operador *backshift*², té la següent forma:

$$\phi_p(B) \cdot \Phi_P(B) \cdot (1 - B)^d (1 - B^s)^D X_t = \theta_q(B) \cdot \Theta_Q(B) \cdot Z_t \quad (2)$$

Amb:

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \dots - \phi_p B^p \\ \Phi_P(B) &= 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps} \end{aligned}$$

¹Una diferenciació d'ordre s és una transformació que s'aplica a una sèrie on substraïem a cada valor de la sèrie X_t el valor de s unitats de temps anterior: $X_t - X_{t-s}$. La metodologia Box-Jenkins aplica diferenciacions per transformar una sèrie temporal a estacionària.

²L'operador *backshift* és una eina molt útil per a la notació dels retards: $B^s X_t = X_{t-s}$

$$\theta_q(B) = 1 + \theta_1 B + \dots + \theta_q B^q$$

$$\Theta_Q(B) = 1 + \Theta_1 B^s + \dots + \Theta_Q B^{Qs}$$

Per tant, a partir de les gràfiques de ACF i PACF de la sèrie, haurem d'identificar els valors de p, q i P, Q .

Per últim, per tractar efectes de calendari i valors atípics utilitzarem els models ARIMAX: models ARIMA amb variables exògenes:

$$\phi_p(B) \cdot \Phi_P(B) \cdot (Y_t - (\beta_0 + \beta_1 X_t)) = \theta_q(B) \cdot \Theta_Q(B) \cdot Z_t \quad (3)$$

On Y_t és la sèrie observada i X_t la sèrie o variable exògena que introduïm. Per suposat, es poden fer servir més d'una variable exògena.

Nota: si es fessin servir altres sèries temporals com a variables exògenes, s'hauria de fer tot un procediment per tal de garantir que no hi ha correlacions espúries entre les dues sèries temporals, però nosaltres només utilitzarem variables indicatives i per tant no ens caldrà aquest pas.

Més endavant parlarem de la sèrie linealitzada: $\tilde{Y}_t = Y_t - \beta_1 X_{1,t} - \beta_2 X_{2,t} \dots$. Aquesta serà la sèrie a la qual li hem tret els efectes dels valors atípics (tractant efectes de calendaris farem el mateix però no li direm sèrie linealitzada). El que farem serà identificar i ajustar un model a la sèrie tractada d'atípics i/o efectes de calendari i després a les prediccions d'aquest afegir-li els efectes que hem de tenir en compte (ja sigui efectes de valors atípics o de calendari).

Tornant a la metodologia Box-Jenkins: aquesta es basa en un procés iteratiu de les següents etapes:

- Tractar no-estacionarietat + Identificació
- Estimació + Validació + Selecció
- Prediccions

A la primera etapa es fan les transformacions necessàries per transformar la sèrie a estacionària (eliminar tendència i estacionalitat, fer constant la variància...) i posteriorment s'identifiquen varies opcions per als possibles valors de p i q , amb els quals després ajustarem els models ARMA a partir de les gràfiques ACF i PACF (Autocorrelation Function i Partial Autocorrelation Function).

A l'etapa d'estimació s'estimen els models amb els diferents valors de p i q identificats prèviament (nosaltres ho farem utilitzant el software R), i després es validen els models (es comprovar que compleixin les hipòtesis de homodeasticitat, independència i normalitat de residus, i la invertibilitat i estacionarietat del model, entre d'altres). Si cap model proposat ens valida, hauríem de tornar a l'etapa d'identificació (i fins i tot a revisar si hem transformat correctament la sèrie).

Finalment, seleccionem un model d'entre els validats en funció de mètriques d'adequació a les dades i de qualitat de les prediccions i calculem les prediccions per a valors futures de la sèrie.

4 Resultats

A continuació mostrem el procediment d'aplicar tota la metodologia Box-Jenkins a la nostra sèrie temporal.

4.1 Anàlisi preliminar

Per entendre el comportament de la sèrie va bé mostrar la serie com a una descomposició de la tendència, la component estacional, i la component irregular (Figura 2)³.

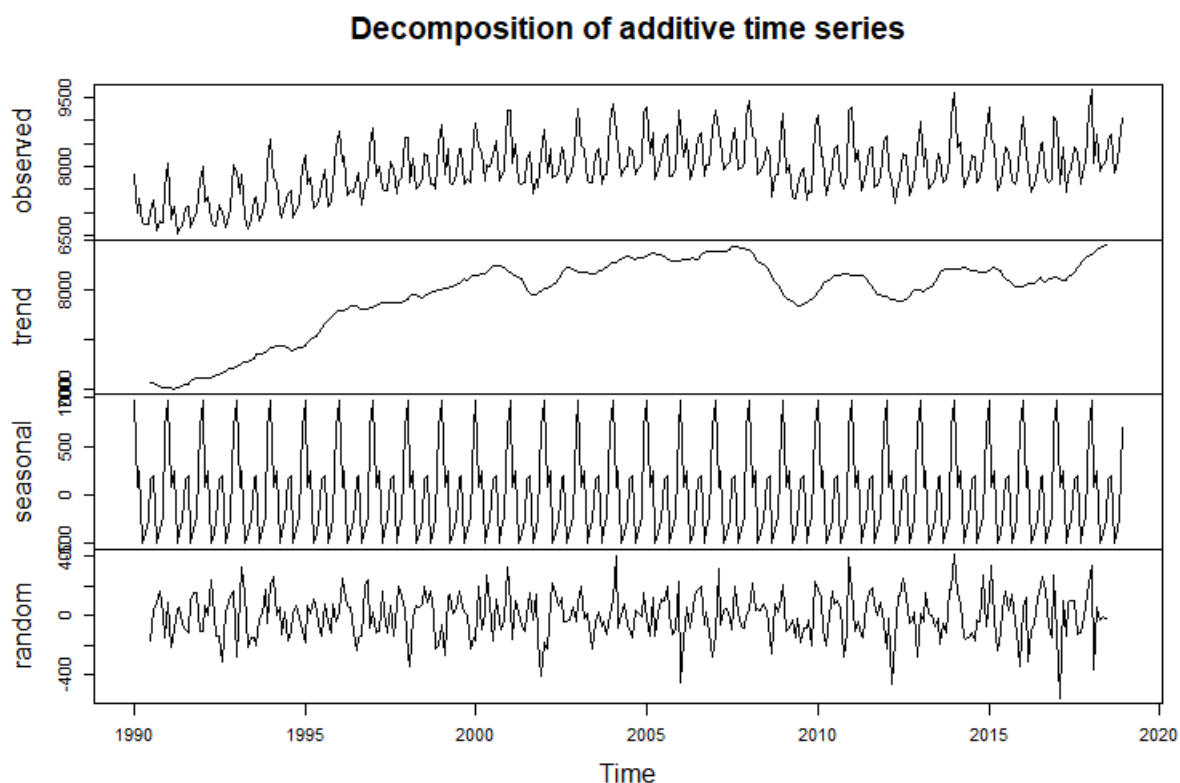


Figura 2: Descomposició de la sèrie

La sèrie mostra un clar patró estacional: pren el valor més alt de l'any a l'hivern, i té un segon pic, més baix, a l'estiu. Intuïtivament, és degut a que a l'hivern s'augmenta molt més el consum perquè fa fred, i a l'estiu perquè fa calor.

Des de l'any 1990 mostra una tendència creixent fins a l'any 2001, on pateix una petita caiguda, posteriorment a l'any 2008 cau encara més avall. I per últim també té una caiguda (no tan forta) a l'any 2011. Les dues primeres caigudes són resultat de les crisis del març de 2001 i del juny de 2008. Aquesta darrera és coneguda com a la Gran Recessió, que va ser detonada per la Crisis de les hipoteques subprime, i va ser la crisi més greu desde la Gran Depressió del 1929. Aquestes crisis van afectar directament l'economia dels Estats Units, i per tant el consum global també en va sortir repercutit.

³Aquesta descomposició de la sèrie l'hem feta amb la funció *decompose* el packet *stats*, del software R, que calcula la descomposició de la serie fent servir mitjanes mòvils.

4.2 Tractament de no-estacionarietat

Per poder aplicar els models ARIMA, hem de transformar la sèrie a estacionària. El primer que s'ha de comprovar de la sèrie és si la variància és constant o si augmenta amb la mitjana. D'entrada no queda gaire clar mirant només el plot de la sèrie (Figura 1). Es pot veure més clarament a la Figura 3. A la Figura 27 de l'Apèndix A es mostra una altra manera de veure la variància creixent.

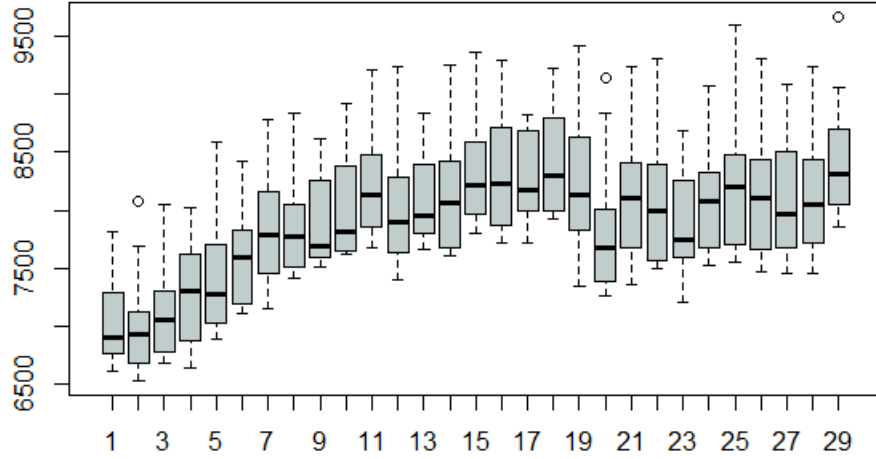


Figura 3: Boxplot de la sèrie EnerUSA

Aquí és veu que la variància creix subtilment amb la mitjana. Per tant, hem aplicat una transformació logarítmica a la sèrie per eliminar l'augment de variància. S'ha de tenir en compte que ara les unitats de la nova sèrie (l'anomenarem *lnserie*) estaran en unitats logarítmiques i que per tant les prediccions que es facin sobre aquesta sèrie també.

Per eliminar el patró estacional mencionat abans, hem d'aplicar una diferenciació estacional d'ordre 12, ja que tenim dades mensuals. I per últim, per obtenir mitjana constant, s'ha d'aplicar tantes diferenciacions regulars (d'ordre 1) sense arribar a sobre diferenciar⁴.

En el nostre cas hem aplicat només una diferenciació regular. Per tant, les transformacions que li apliquem a la sèrie per obtenir una sèrie estacionària són el logaritme, una diferenciació estacional d'ordre 12 i una diferenciació regular.

La Figura 28 de l'Apèndix A mostra la sèrie estacionària i el seu monthplot.

$$W_t = (1 - B)(1 - B^{12})\log(X_t)$$

4.3 Identificació

Per identificar diferents opcions possibles per als valors de p, q, P i Q es fan servir les funcions d'autocorrelació i d'autocorrelació parcial de la sèrie (ACF i PACF, respectivament). A la Figura 4 es mostren aquestes gràfiques de la sèrie. Per a la part estacional hem considerat els models $ARMA_{12}(3, 0)$, $ARMA_{12}(0, 2)$ i $ARMA_{12}(1, 1)$, i per a la part regular, $ARMA(2, 0)$, $ARMA(0, 3)$ i $ARMA(1, 1)$.

⁴Estarem sobre diferenciant quan la variància de la sèrie ens augmenti. Si la variància augmenta insignificantment, ens pot interessar diferenciar igualment. En cas de no fer-ho, si els models no ens valen val la pena reajustar de nou els models amb una diferenciació més a la sèrie.

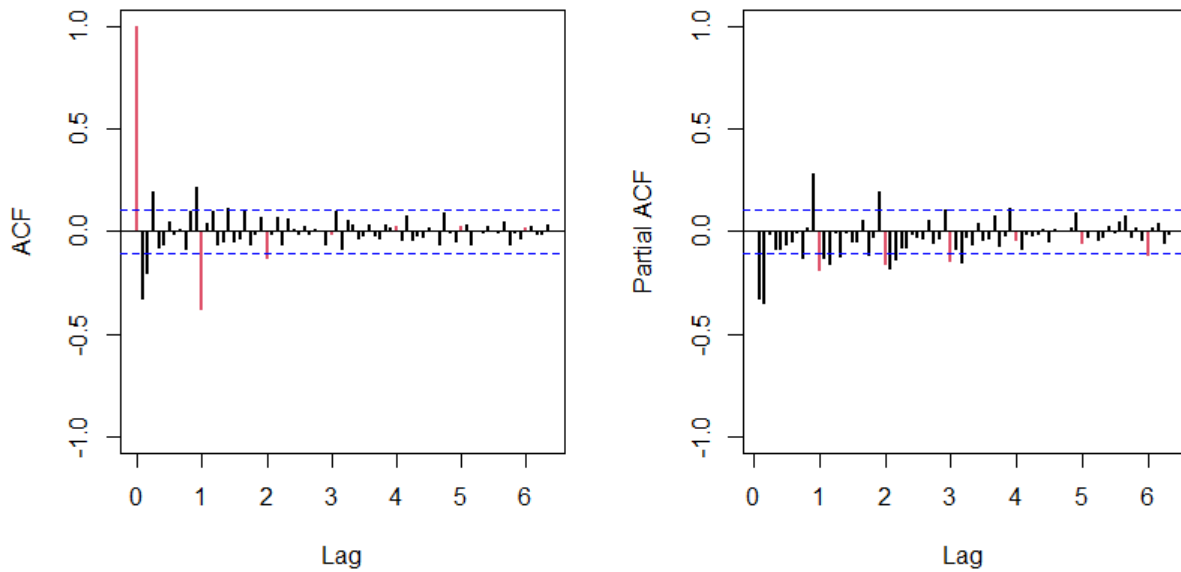


Figura 4: ACF i PACF de la sèrie estacionària

4.4 Estimació

Hem ajustat tots els models possibles amb les diferents combinacions de p, q i P, Q que hem considerat per a les parts regular i estacional, i hem seleccionat els dos que tenien millor AIC⁵. Els anomenarem models 1 i 2, i són, respectivament, $ARIMA(0, 1, 2)(0, 1, 2)_{12}$ i $ARIMA(1, 1, 1)(0, 1, 2)_{12}$. Del model 1 hem descartat el coeficient $ma3$ ja que no era significatiu. Els coeficients estimats són:

Model 1:

	ma1	ma2	sma1	sma2
valor	-0.4866	-0.1909	-0.793	-0.1229
s.e.	0.0563	0.0627	0.065	0.0673

Model 2:

	ar1	ma1	sma1	sma2
valor	0.2930	-0.7918	-0.8006	-0.1159
s.e.	0.0939	0.0654	0.0656	0.0679

Cal notar que el coeficient $sma2$ no és significatiu en cap dels dos casos, però hem considerat no treure'l ja que l'AIC⁵ disminueix sense aquest. Els AIC's són, respectivament, de -1599.56 i -1599.12. Per veure la sortida completa dels models, veure models 1 i 2 de l'Apèndix B

4.5 Validació

Per poder garantir que els models ajustats siguin fiables, cal validar-los i comprovar que compleixen les hipòtesis establertes, com la independència i homodeasticitat de residus. Part de les gràfiques de la validació dels models les adjuntem a l'Apèndix B.

Comencem amb la validació del model 1:

⁵ Akaike Information Criterion: és una mesura de la bondat d'ajust del model a les dades, que penalitza al model amb la complexitat. Com a norma general per comparar models, l'AIC indica que el model amb AIC més baix s'ajusta millor a les dades tenint en compte la seva complexitat.

Figura 5: Residuos del model 1

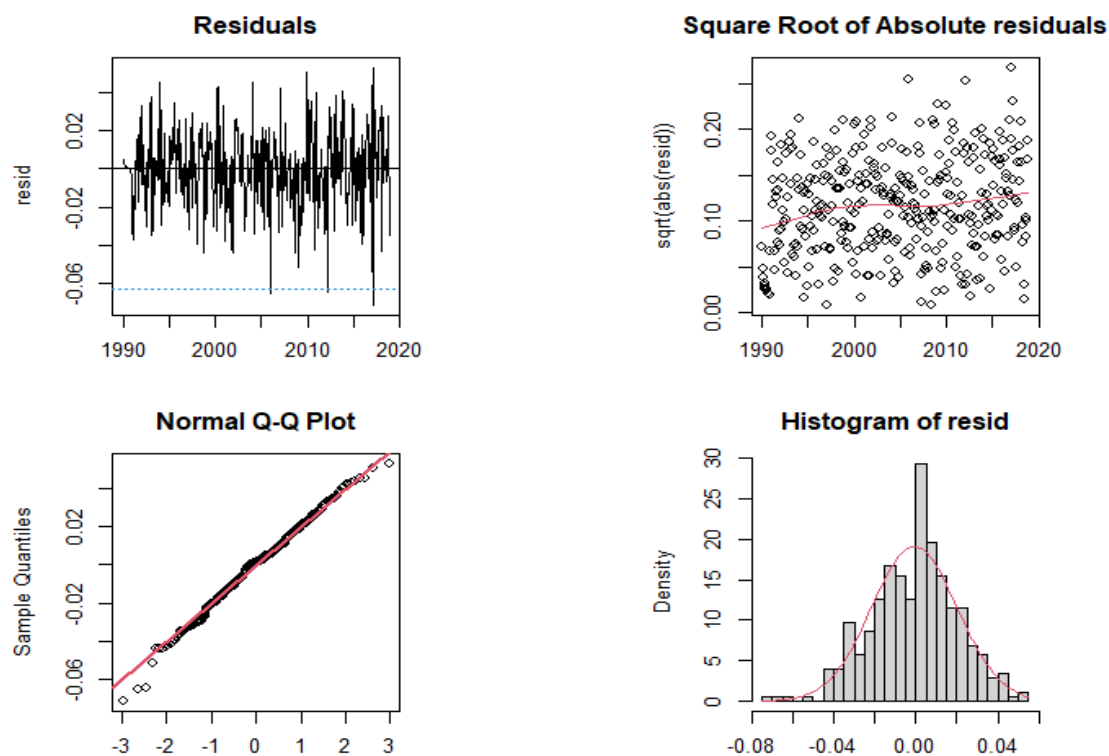
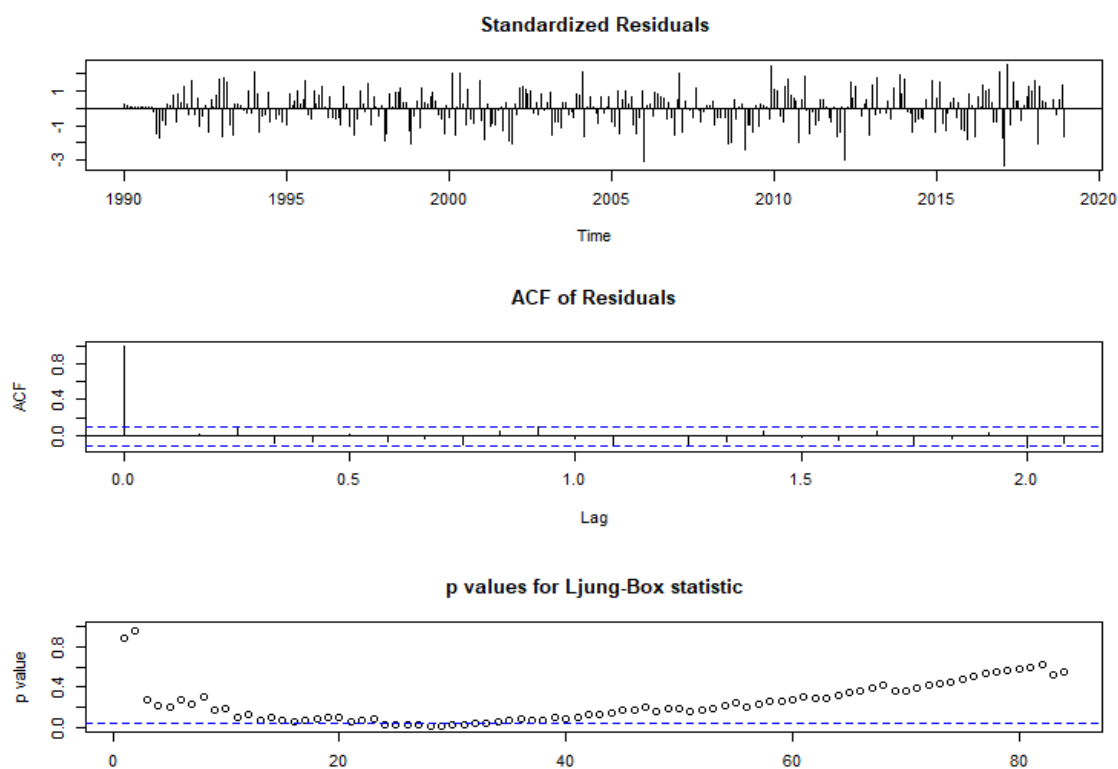


Figura 6: Test de Ljung-Box del model 1

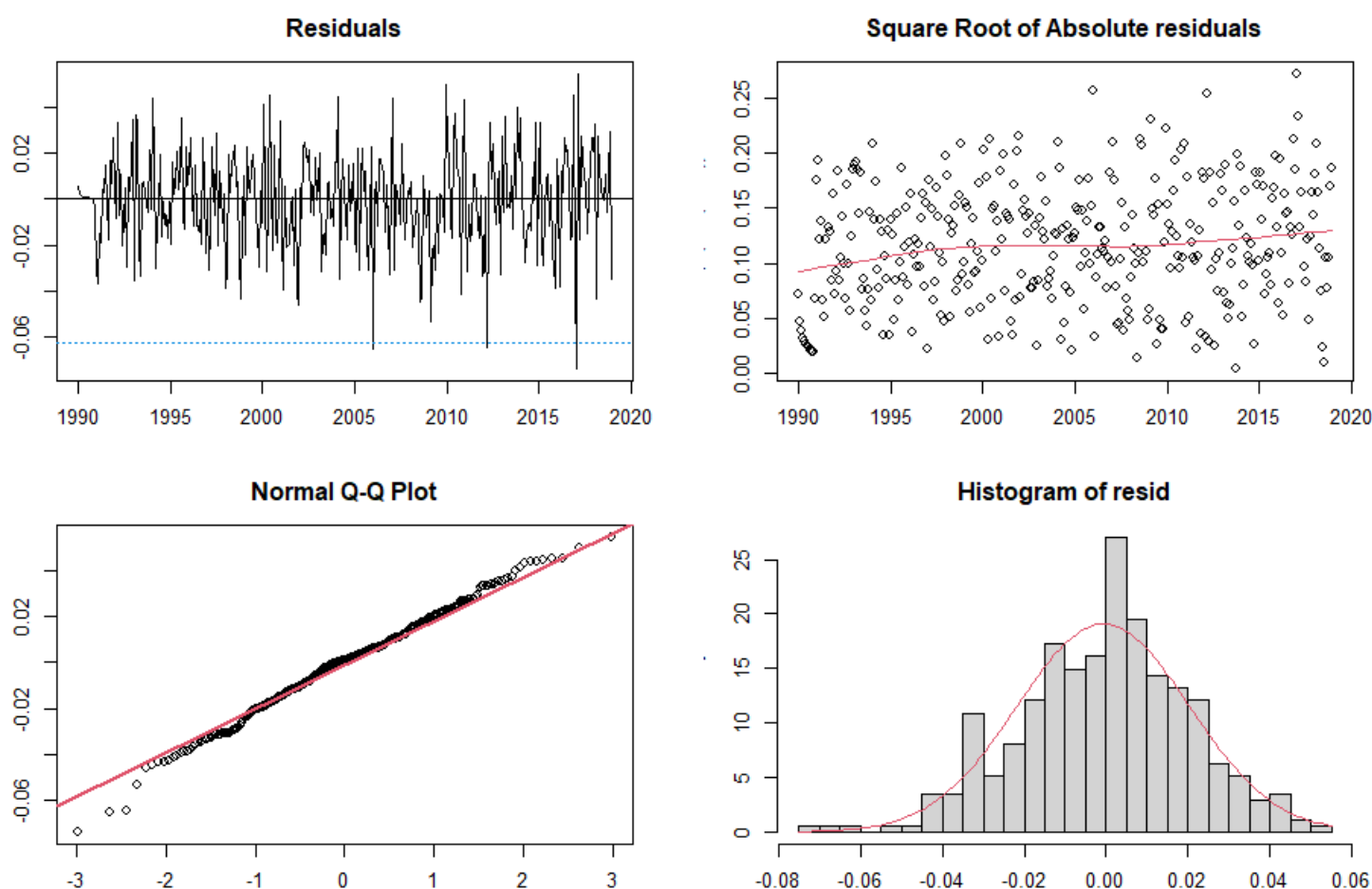


La variància residual mostra una lleugera tendència creixent, però podem assumir que és constant, ja que segurament és deguda a valors atípics. Pel que fa a la normalitat, el qqplot i l'histograma dels residus encaixen amb una distribució normal. A la cua inferior del qqplot veiem de nou alguns possibles valors atípics, però no ens fan rebutjar l'hipòtesi de normalitat. Per donar més solidesa al compliment d'aquestes premisses també es poden fer tests d'hipòtesi. Hem realitzat diferents testos (veure Apèndix B), i garanteixen també normalitat i homodeasticitat.

Pel que fa a la independència residual, els resultats ja no són tan bons. Observant els p-valors del test de Ljung-Box⁶ veiem que es mantenen tots per sobre del llindar 0.05, però a partir del retard 24 passen a sota del llindar. Això que complica la validació del model, però segurament amb tractament de valors atípics i efectes de calendari aconseguiríem validar aquest model. Per últim, el model és causal i invertible ja que totes les arrels tenen mòdul superior a 1 (veure Apèndix B).

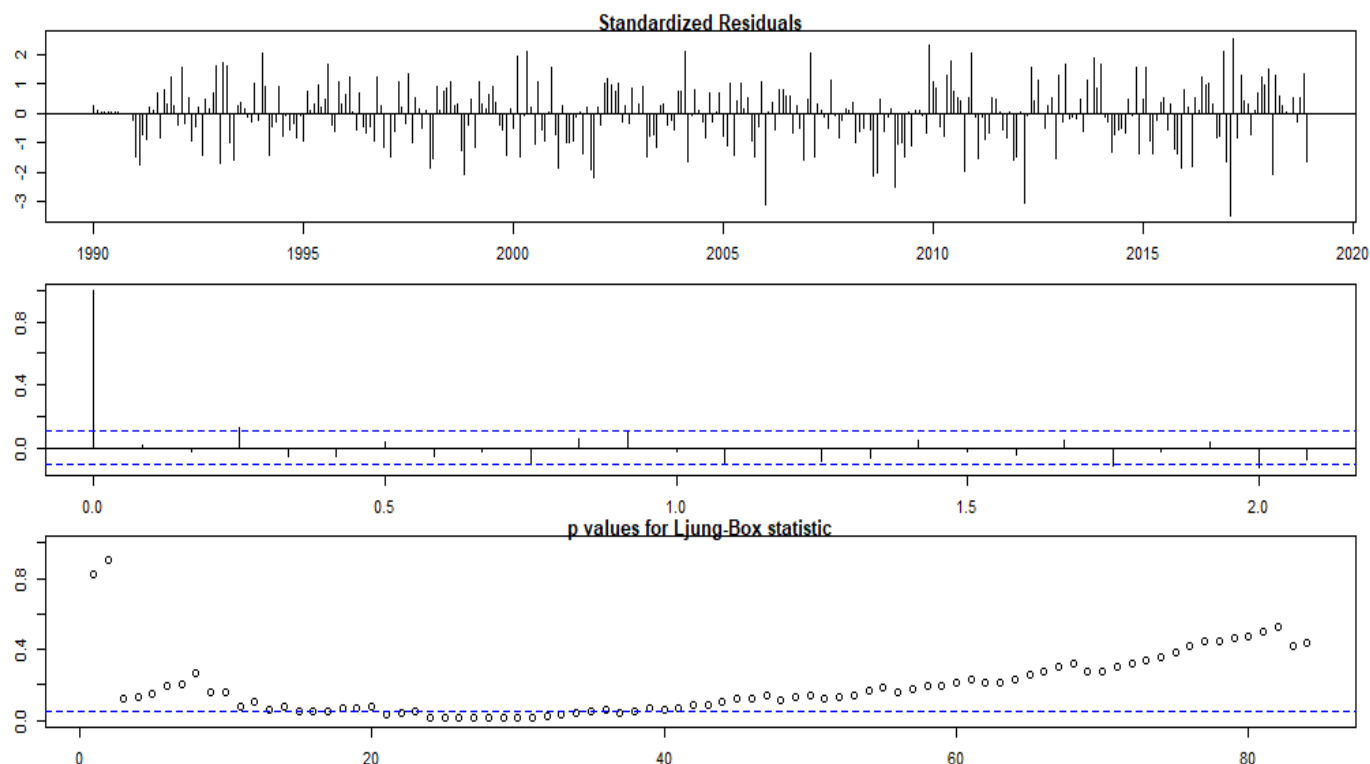
Validació model 2:

Figura 7: Residus del model 2



⁶Test de Ljung-Box: test estadístic que prova si un grup d'autocorrelacions d'una sèrie temporal és diferent o no de zero. En aquest cas, l'apliquem a les correlacions dels residus del model per veure si aquests són independents.

Figura 8: Test de Ljung-Box del model 2



Els residus del model 2 són molt similars als del model 1. Es veu un lleuger creixement de la variància, però podem considerar-la constant. També veiem uns possibles valors atípics a la mateixa franga que abans, i la normalitat es compleix bastant bé, amb algun valor atípic a la cua inferior. El model també és invertible i causal. Però la independència de residus en aquest cas és encara pitjor; a la Figura 8 es veu que el p-valor de la prova de Ljung-Box ja baixa encara més ràpid que amb el model 1, passant per sota de llindar $p = 0.05$ abans del retard 20, i des del retard 22 al 37 pràcticament. Aquí sí que no podem garantir independència residual. Tot i així, igual que hem comentat abans, amb tractament d'efectes de calendari i valors atípics segurament aconseguiríem millorar la validació del model.

4.6 Prediccions

En aquesta secció procedirem a veure com de bones són les capacitats predictores dels models. Així doncs el primer que farem serà fer un enfilestrat de la sèrie, la qual anomenarem *lnserie2* i ajustem amb cadascun dels models escollits per comprovar que són estables⁷. Anomenarem model 1a al que ajustem a la sèrie original *lnserie* i model 1b al que ajustem a la nova sèrie enfilestrada *lnserie2* (ídem pel model 2). Tot seguit, un cop comprovada la seva estabilitat, calcularem els paràmetres necessaris per fer les

⁷Estabilitat: considerarem que un model és estable quan els coeficients del model ajustat a la sèrie enfilestrada són tots del mateix ordre, signe i significació que els del model original. És necessari tenir estabilitat per poder generalitzar les mesures de predicció obtingudes amb el model ajustat a la sèrie enfilestrada al model original.

prediccions així com les seves mètriques d'exactitud i precisió (rmse⁸, mae⁹, rmspe¹⁰, mape¹¹, etc.).

Comencem amb el model 1:
Després d'ajustar per a cadascuna de les sèries amb el primer model, aquests són els seus coeficients:

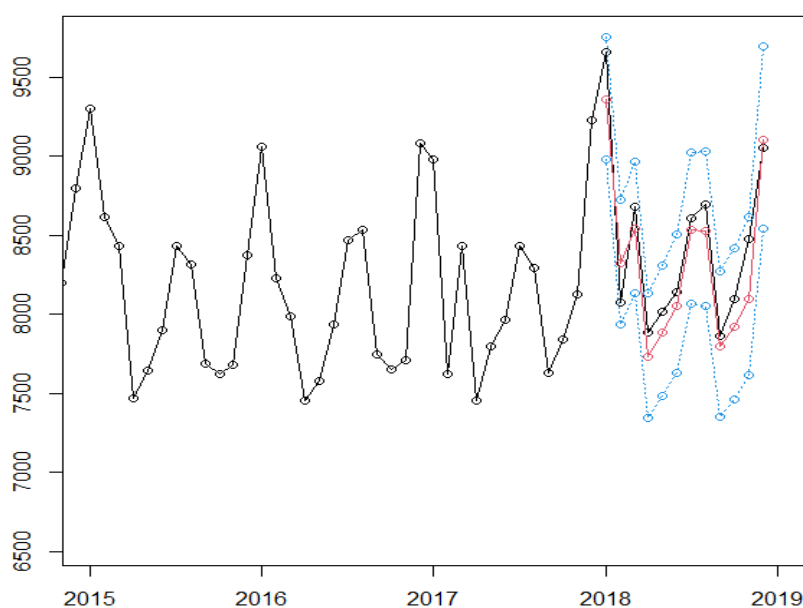
Model 1a:

	ma1	ma2	sma1	sma2
valor	-0.4866	-0.1909	-0.793	-0.1229
s.e.	0.0563	0.0627	0.065	0.0673

Model 1b:

	ma1	ma2	sma1	sma2
valor	-0.4694	-0.2082	-0.8286	-0.0798
s.e.	0.0569	0.0637	0.0686	0.0718

Figura 9: Prediccions del model 1



⁸ Root Mean Square Error: És una mesura de la diferència entre els valors predits per un model i els valors observats. Com més petit sigui el RMSE, millor serà l'ajust del model a les dades.

⁹ Mean Absolute Error: És similar al RMSE, però en lloc de prendre l'arrel quadrada de les diferències, simplement calcula la mitjana de les diferències absolutes entre els valors predits i els valors reals.

¹⁰ Root Mean Square Percentage Error: És una mesura d'error relatiu utilitzada comúment en problemes de pronòstic. Es calcula com l'arrel quadrada de la mitjana dels quadrats dels errors percentuals entre els valors predits i els valors reals. És especialment útil quan es treballa amb dades que varien en magnituds significatives.

¹¹ Mean Absolute Percentage Error: Similar al RMSPE, però en lloc de prendre la mitjana de les diferències absolutes, calcula la mitjana dels errors percentuals absoluts entre els valors predits i els valors reals. Avalua la precisió d'un model en termes de percentatge d'error absolut mitjà.

Mètriques d'exactitud i precisió:

```
rmse: 188.7221
mae: 162.9916
rmspe: 2.208195 %
mape: 1.922612 %
mean: 903.1193
```

Per veure més contingut dels resultats, veure Apèndix C.

Com podem veure, el model 1 seria estable si no fos pel coeficient sma2. De fet, aquest coeficient ja no era significatiu per molt poc però no l'havíem eliminat perquè l'AIC del model millorava al conservar aquest coeficient. Tot i això, ara veiem que per poder garantir l'estabilitat del model hauríem d'eliminar aquest coeficient i tornar a mirar el model 1b per corroborar que és estable i tenir unes prediccions més fiables. De moment però, no l'eliminarem ja que la validació del model tampoc era gaire bona. Per tant, passarem a tractar els possibles efectes de calendari que hi hagi, a veure si se'ns soluciona aquest problema de validació i estabilitat, i en cas que no, llavors ja mirariem de forçar a treure aquest paràmetre insignificant. Tant la gràfica de les prediccions com les mètriques d'exactitud i precisió, són bastant bones.

Mirem ara el model 2:

Els coeficients són els següents:

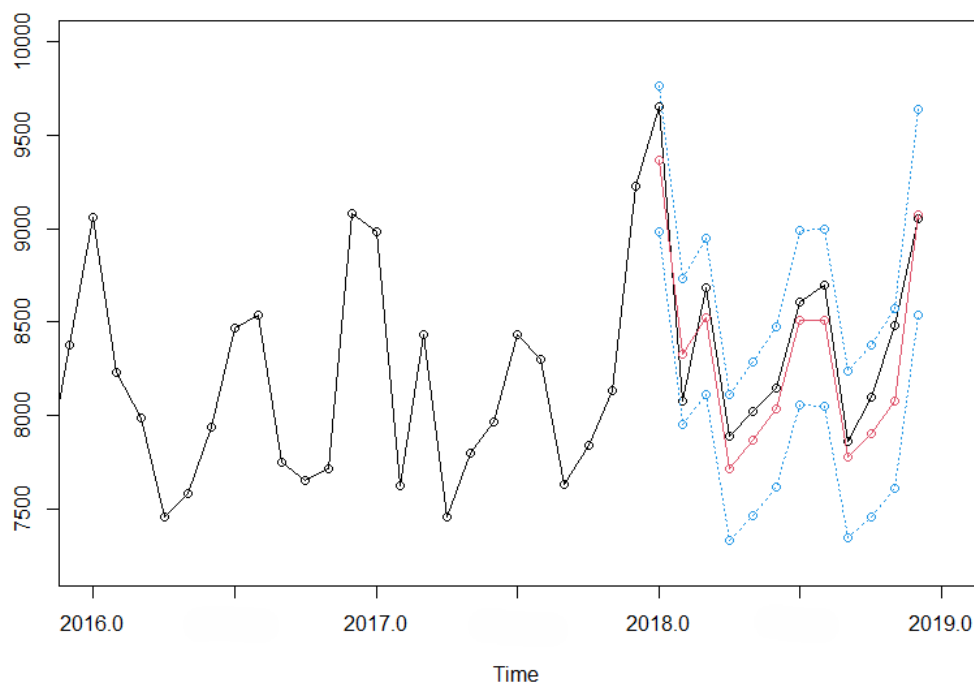
Model 2a:

	ar1	ma1	sma1	sma2
valor	0.2930	-0.7918	-0.8006	-0.1159
s.e.	0.0939	0.0654	0.0656	0.0679

Model 2b:

	ar1	ma1	sma1	sma2
valor	0.3314	-0.8112	-0.8408	-0.0657
s.e.	0.0940	0.0647	0.0696	0.0730

Figura 10: Prediccions del model 2



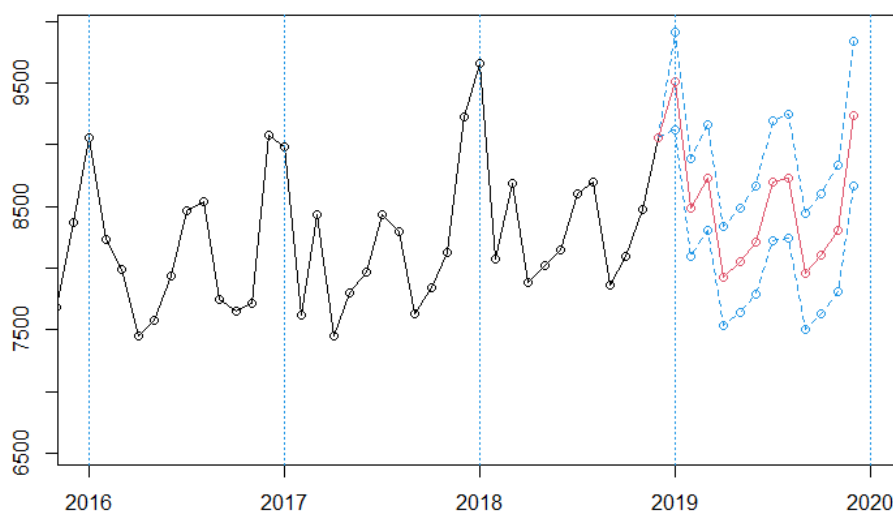
Mètriques d'exactitud i precisió:

```
rmse: 202.3824
mae: 177.1128
rmspe: 2.381435 %
mape: 2.097393 %
mean: 885.9577
```

Com es pot veure, el model 2 és bastant semblant a l'1 en tots els aspectes, tenim el mateix problema amb el coeficient sma2 i tant la gràfica com les mètriques de predicció són molt similars. L'únic lleuger canvi és que els errors són un pèl majors al segon model (això el fa menys exacte) però la mitjana de les bandes de confiança és menor, cosa que el fa lleugerament més precís. Així doncs, en aquest cas hem decidit quedar-nos amb el primer model ja que, en quant a validació, té una millor independència de residus. Les prediccions a llarg termini són les següents:

	Jan	Feb	Mar	Apr	May	Jun
2019	9.160217	9.046259	9.073968	8.978192	8.994208	9.013830
	Jul	Aug	Sep	Oct	Nov	Dec
2019	9.070876	9.075049	8.982350	9.000257	9.025198	9.130541

Figura 11: Prediccions a llarg termini del model 1



A continuació intentarem millorar la validació i les mètriques de predicció dels nostres models amb tractament d'efectes de calendari.

4.7 Efectes de Calendari

Al treballar amb sèries temporals mensuals, el fet que cada any canviï la distribució del calendari pot afectar a la sèrie. Per exemple, un mes pot tenir més o menys dies laborables, o coincidir amb la setmana santa o no. Com que això són dades que són conegudes, podem aprofitar aquest coneixement a l'hora d'ajustar els models, i d'aquesta manera aconseguirem explicar millor les dades i obtenir una millor validació i predicció del model.

El càlcul de les variables exògenes que utilitzarem a aquest apartat l'hem realitzat amb les funcions *Wtrad* i *Weaster*, que les podeu trobar l'Apèndix F.

Anàlisi d'Intervenció

Primer cal veure quins efectes de calendari afecten a la sèrie i quins no. Ajustarem el model escollit fins ara (model 1) i hi inclourem les diferents combinacions de variables exògenes. Considerarem només els efectes de calendari de Setmana Santa i Trading Days. Tractarem un efecte de calendari com a significatiu si el coeficient de la variable exògena és significatiu.

Hem obtingut que només és significatiu l'efecte de calendari de Trading Days; és a dir, no podem considerar que la Setmana Santa té un efecte significatiu sobre la sèrie. Veure Apèndix D per més detall en l'ajust de models per anàlisi d'intervenció.

El model 1 ajustat amb Trading Days té els següents coeficients:

Coefficients :				
	ma1	ma2	sma1	wTradDays
	-0.4685	-0.2040	-0.8850	0.0011
s.e.	0.0544	0.0589	0.0376	0.0003

Cal notar que hem tret el coeficient *sma2* ja que ara sí que passava a ser molt poc significatiu. Procedim a calcular la sèrie tractada d'efectes de calendari, que l'anomenarem *serieTD*:

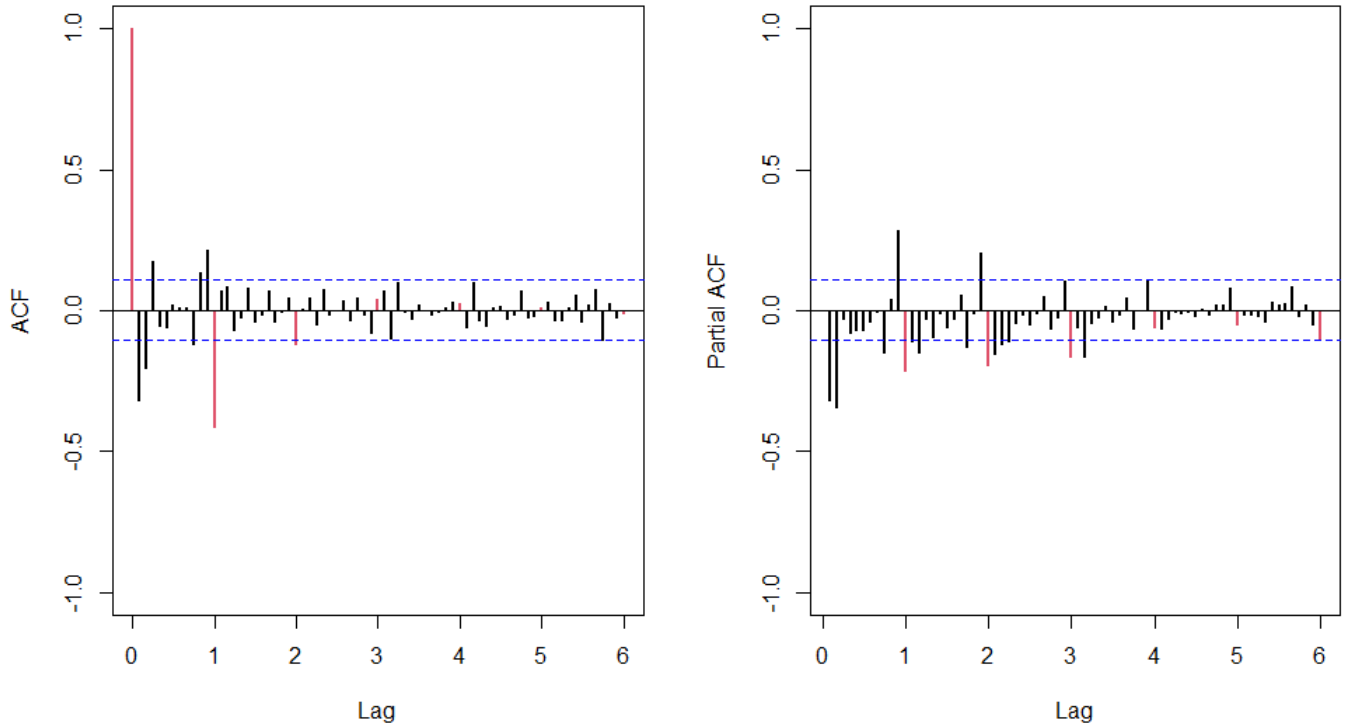
$$serie_{TD} = \exp(\log(serie) - 0.0011 \cdot X_{TD})$$

Hem de tornar a tractar la sèrie per a que sigui estacionària. En aquest cas, la variable TD no indueix massa canvi a la sèrie, així que hem aplicat les mateixes transformacions per a obtenir estacionarietat que abans. La sèrie estacionària és:

$$(1 - B)(1 - B^{12})\log(serie_{TD})$$

Amb aquesta identifiquem de nou possibles models.

Figura 12: ACF i PACF de la sèrie estacionària tractada d'efectes de calendari



Per a la part regular identifiquem un $MA(3)$, un $AR(2)$ i un $ARMA(1,1)$, i per la part estacional un $AR(3)_{12}$, un $MA(2)_{12}$ tot i que segurament el coeficient $sar2$ no serà significatiu i l'haurem de treure, i un $ARMA(1,1)_{12}$. Els models amb millor AIC tornen a ser els equivalents al model 1 i model 2 que havíem seleccionat abans, però sense coeficient $sar2$, ja que amb la introducció de la variable exògena perd molta significació. Els models per tant són $ARIMA(0,1,2)(0,1,1)_{12}$ i $ARIMA(1,1,1)(0,1,1)_{12}$, i els anomenarem model1.TD i model2.TD, respectivament. Els coeficients estimats són:

Model 1.TD

	ma1	ma2	sma1
coef	-0.4685	-0.2040	-0.8850
s.e.	0.0544	0.0589	0.0376

Model 2.TD

	ar1	ma1	sma1
coef	0.3361	-0.8082	-0.8887
s.e.	0.0880	0.0590	0.0372

Els AIC's són molt similars: -1612.02 del model 1.TD i -1612.41 del segon. Passem a mirar les validacions dels models per escollir amb quin seguim (veure Apèndix D per més gràfics).

Figura 13: Residus del model 1.TD

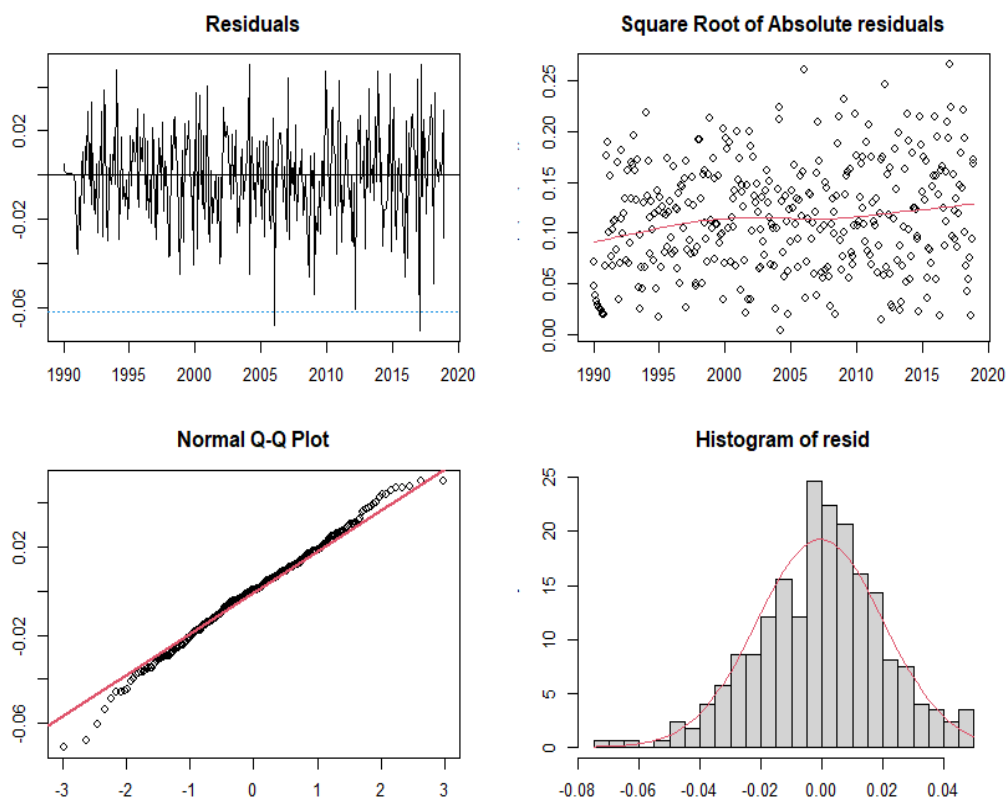


Figura 14: Test de Ljung-Box del model 1.TD

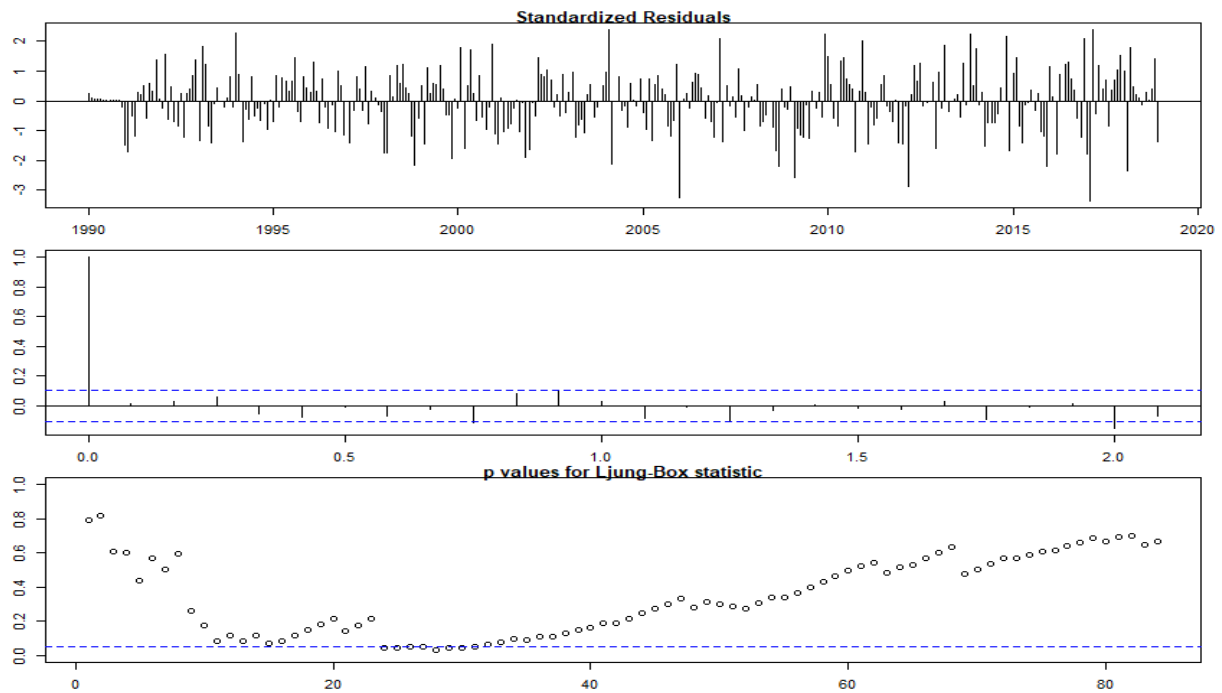


Figura 15: Residus del model 2.TD

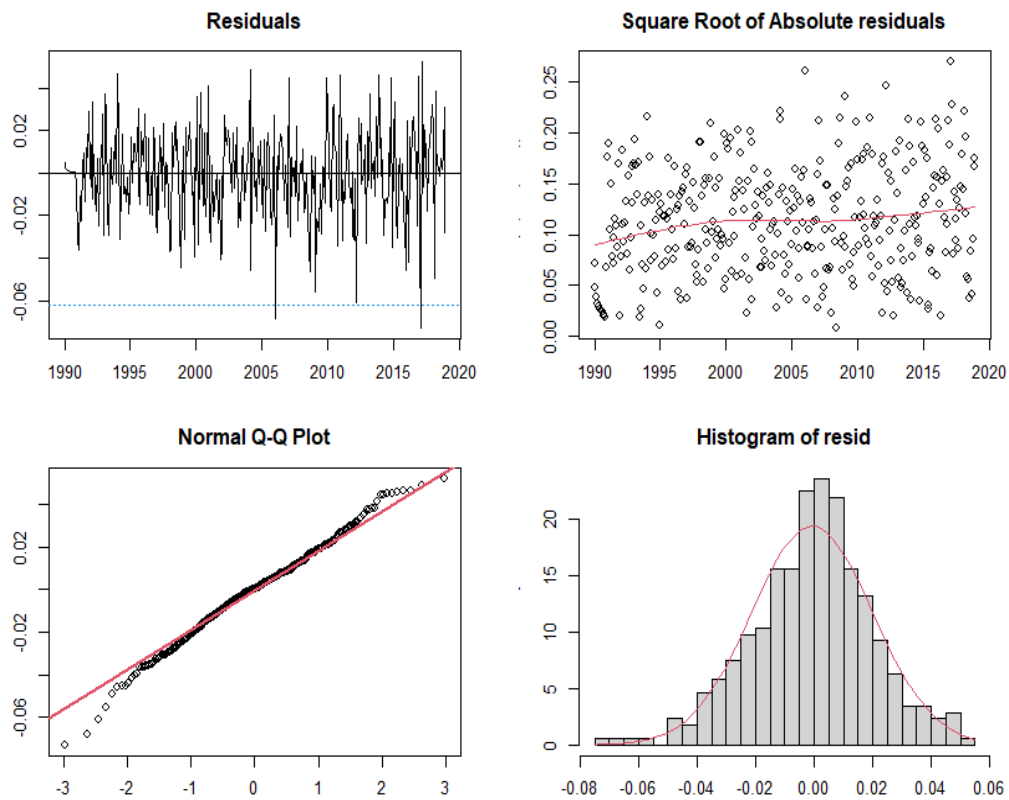
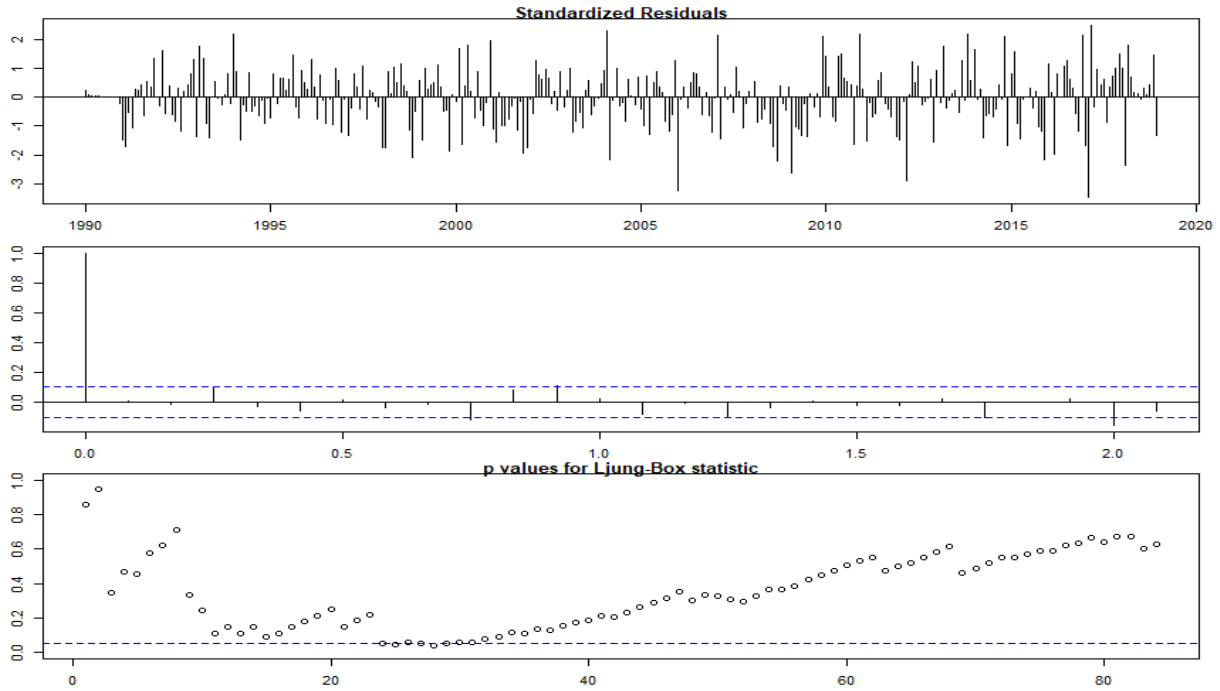


Figura 16: Test de Ljung-Box del model 2.TD

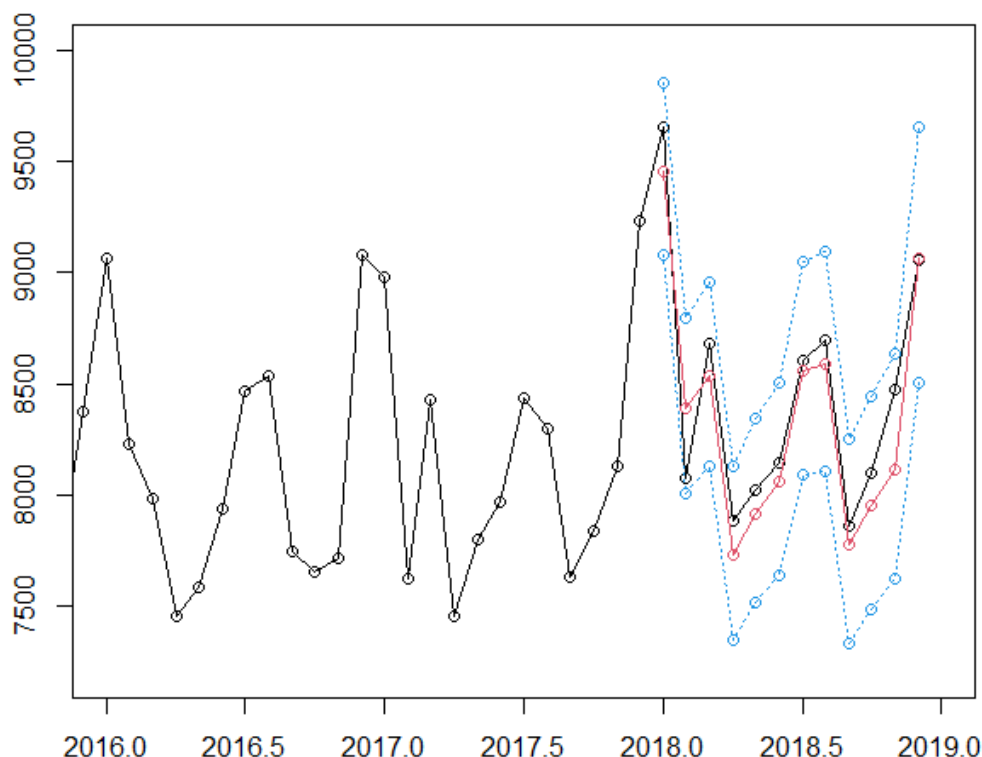


Els dos models tenen una validació també molt similar i molt millor que abans d'ajustar efectes de calendari. Tenen una mica de desviació de les cues inferior i superior del qq-plot (sobretot inferior), però la normalitat es compleix, i també la homodeasticitat. Seguim observant residus amb valors grans a partir de l'any 2005; no s'ha arreglat amb el tractament d'efectes de calendari. La independència de residus també ha millorat, sobretot del model 2. En els dos casos però, els p-valors del test de Ljung-Box baixen per sota el llindar entre els retards 24-30 aproximadament. També observem que el ACF i PACF del model 1.TD ens encaixen una mica més amb els teòrics que els del model 2.TD (veure figures 35 i 37 de l'Apèndix D). Els dos models són estables (veure figures 34 i 36 de l'Apèndix D).

També hem mirat les capacitats predictives del model, i el model 1.TD sembla tenir unes precisions una mica més exactes:

	Model 1	Model 2
rmspe	2.118581	2.208195
mape	1.756752	1.922612
mean	903.5756	903.1193

Figura 17: Prediccions de l'últim any del model 1.TD



Hem decidit continuar amb el model 1.TD ja que té millors mesures de predicció, tot i que els dos models són molts iguals.

Per tant, fins aquest punt ens decantaríem pel model 4, però podem millorar encara més les qualitats d'aquests model amb tractament de valors atípics.

4.8 Tractament d'atípics

Com hem esmentat anteriorment el fet de tractar el possibles valors atípics pot fer millorar les qualitats del nostre model, així doncs, en aquest apartat en disposem a fer aquest procediment. Per la detecció de valors atípics hem utilitzat una funció d'implementació pròpia.

Els atípics detectats amb regió crítica igual a 3 i $LS^{12} = \text{False}$ són els següents:

Obs	type_detected	W_coeff	ABS_L_Ratio	Fecha	perc.Obs
5 170	AO	0.05083500	3.176316	Feb 2004	0.05083500+
3 193	TC	-0.06388348	3.647580	Ene 2006	-0.06388348
4 267	AO	-0.05277061	3.231989	Mar 2012	-0.05277061
1 326	AO	-0.07069718	4.045425	Feb 2017	-0.07069718
2 338	AO	-0.06198804	3.587062	Feb 2018	-0.06198804

Hem obtingut 5 valors atípics, 4 dels quals són AO¹³ i l'altre és un TC¹⁴. Veiem que el model aconsegueix explicar les crisis mencionades a l'anàlisi preliminar de la sèrie i és

¹² Level Shift: És un tipus d'atípic, són canvis sobtats i grans en les dades que no segueixen el patró general de la sèrie temporal o la distribució de les dades. La sèrie pateix un canvi de nivell en la distribució de valors a partir d'aquell punt.

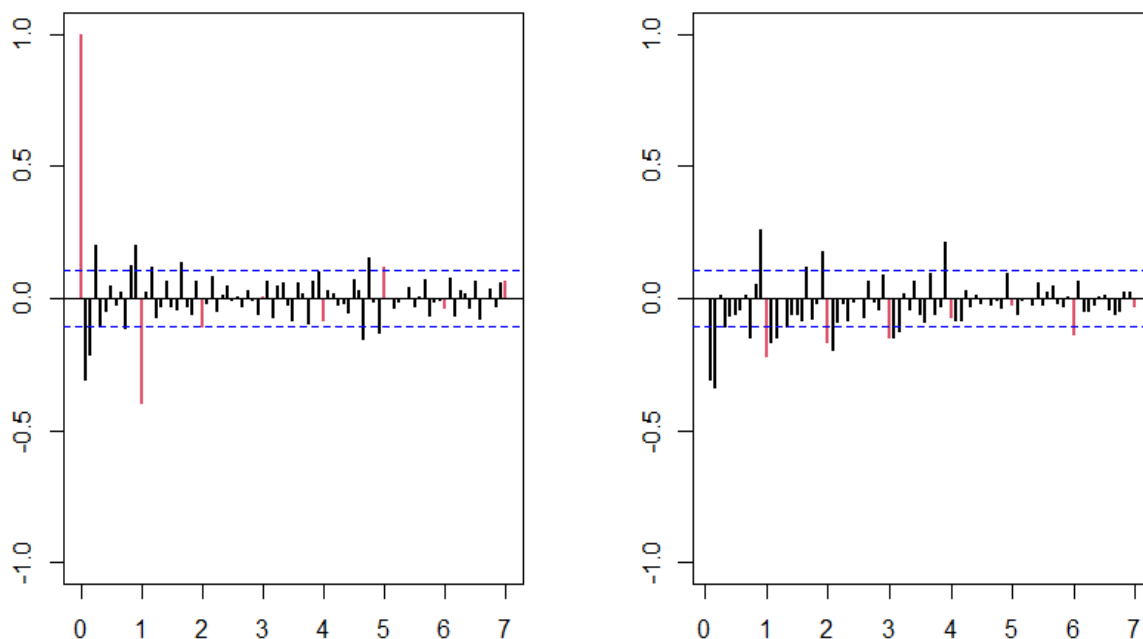
¹³ Additive Outlier: Valors atípics que no afecten als següents valors de la sèrie. Poden ser el resultat d'errors de mesura, entrades incorrectes o esdeveniments rars i poc comuns.

¹⁴ Transitory Change: són atípics que provoquen un desajust al patró de la sèrie i ocasionen canvis momentanis en la sèrie que poc a poc va recuperant, exponencialment, el seu patró original. EN aquest

per això que no s'identifica cap outlier de tipus TC que ens podríem esperar trobar al 2001 i 2008 com a efecte d'aquestes crisis.

A continuació, ens disposem a linealitzar la nostra sèrie prescindint d'aquests valors atípics (Veure Apèndix E) i apliquem les diferenciacions necessàries per transformar-la a estacionària.

Figura 18: ACF i PACF de la sèrie linealitzada i estacionària



Les dues gràfiques són bastant similars a les que havíem obtingut amb la sèrie original i, per tant, distingim els mateixos possibles models: per a la part estacional, els models $ARMA_{12}(3, 0)$, $ARMA_{12}(0, 2)$ i $ARMA_{12}(1, 1)$, i per a la part regular, $ARMA(2, 0)$, $ARMA(0, 3)$ i $ARMA(1, 1)$.

Els dos millors models (basant-nos en el seu AIC) de totes les possibles combinacions que podem fer són un $ARIMA(0, 1, 2)(0, 1, 1)_{12}$ i un $ARIMA(1, 1, 1)(0, 1, 1)_{12}$, els anomenarem respectivament model1.lin i model2.lin. El coeficient ma3 com ja sabíem d'abans és no significatiu i com es pot veure a continuació el sma2 tampoc dona significatiu:

$ARIMA(0, 1, 2)(0, 1, 2)_{12}$

Coefficients :

	ma1	ma2	sma1	sma2	wTradDays
	-0.4279	-0.2038	-0.8493	-0.0373	0.0011
s.e.	0.0541	0.0586	0.0622	0.0645	0.0003

sigma² estimated as 0.0003665: log likelihood = 840.56, aic = -1669.11

$ARIMA(0, 1, 2)(0, 1, 1)_{12}$

Coefficients :

cas no hem permès la detecció d'aquest tipus d'atípics perquè ja d'entrada no es veia cap canvi de nivell, i sinó la funció considerava dos LV consecutius i contraris.

	ma1	ma2	sma1	wTradDays
	-0.4299	-0.2013	-0.8793	0.0011
s.e.	0.0538	0.0580	0.0344	0.0003

sigma^2 estimated as 0.0003672: log likelihood = 840.39, aic = -1670.78

ARIMA(1,1,1)(0,1,2)₁₂

Coefficients:

	ar1	ma1	sma1	sma2	wTradDays
	0.3771	-0.8037	-0.8664	-0.0183	0.0011
s.e.	0.0913	0.0620	0.0626	0.0648	0.0003

sigma^2 estimated as 0.0003651: log likelihood = 841.11, aic = -1670.22

ARIMA(1,1,1)(0,1,1)₁₂

Coefficients:

	ar1	ma1	sma1	wTradDays
	0.3773	-0.8036	-0.8813	0.0011
s.e.	0.0905	0.0613	0.0338	0.0003

sigma^2 estimated as 0.0003653: log likelihood = 841.07, aic = -1672.14

Ara veiem que, després de linealitzar i treure el coeficient sma2, el model 2 és el que té un millor AIC. Anem a mirar però com validen els dos models i les seves mètriques de predicció per determinar amb quin dels dos models ens quedem.

Comencem amb el model 1:

Figura 19: Residus del model1.lin

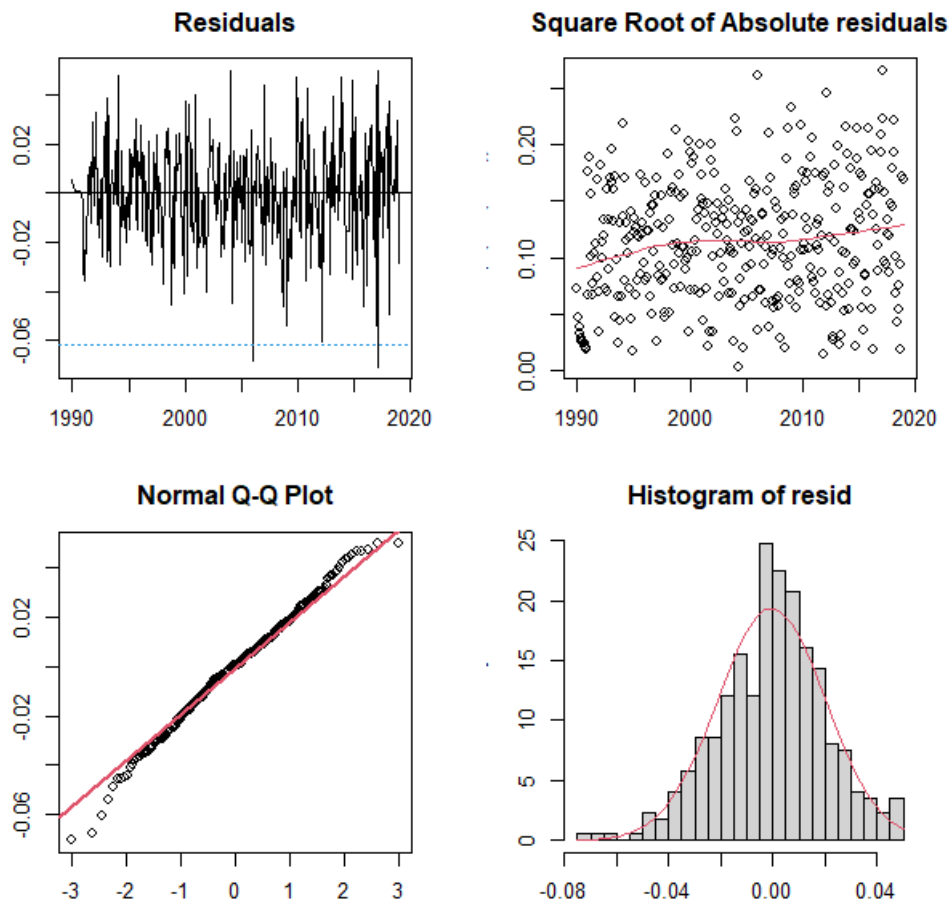
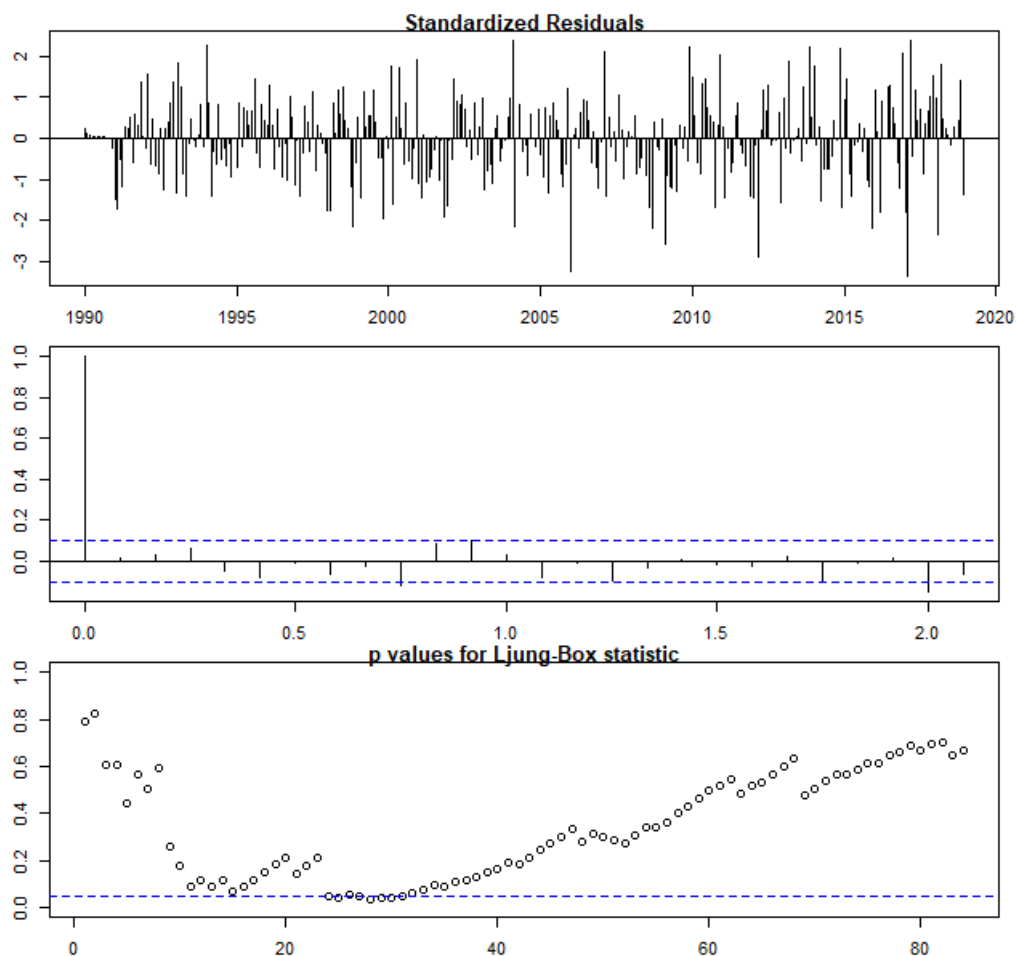


Figura 20: Test de Ljung-Box del model1.lin



Es pot veure que el model 1 valida un pèl millor que abans, s'han reduït una mica les cues pesants i la normalitat dels residus és més considerable. Al Ljung-box gairebé tots els valors estan per sobre del 0.05 excepte algun que cau per sota per molt poc, i a més cau tard i es recupera ràpid.

Passem a veure com de bé prediu el model, primer fem l'enfinestrat de la sèrie linealitzada i amb els efectes de calendari (veure Apèndix E) per comprovar la seva estabilitat:

Model enfinestrat:

Coefficients :

	ma1	ma2	sma1	wTradDays.2
	-0.4303	-0.2051	-0.8743	0.0011
s.e.	0.0549	0.0593	0.0354	0.0003

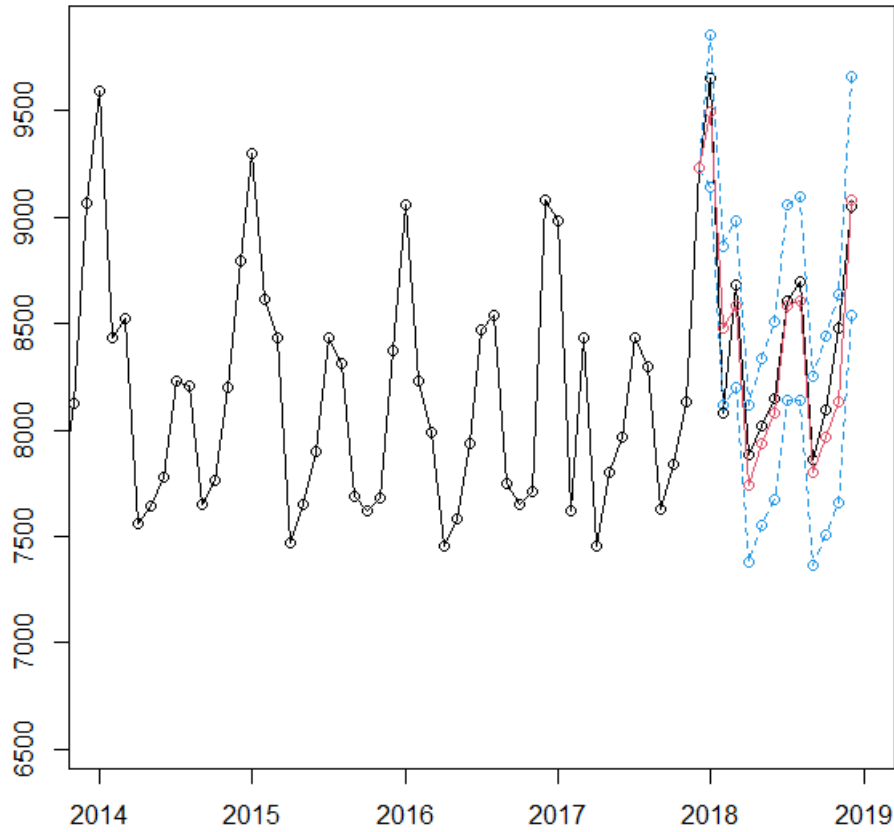
σ^2 estimated as 0.0003743: log likelihood = 807.07, aic = -1604.15

Comparativa dels coeficients:

	ma1	ma2	sma1	wTradDays
modell1.lin	-0.4298823	-0.2012795	-0.8792689	0.001112973
modell1.lin.enfinestrat	-0.4303304	-0.2050631	-0.8743013	0.001055126

El model és estable.

Figura 21: Prediccions del model1.lin



Mètriques d'exactitud i precisió:

rmse: 178.2166
mae: 136.6682
rmspe: 0.742133%
mape: 1.634467%
mean: 1.101084

Es pot veure clarament com totes les mètriques s'han reduït respecte a com havíem vist abans de tractar efectes de calendari i atípics. Això fa el nostre model més exacte i més precís, i per tant, més fiable alhora de fer prediccions.

Anem a veure ara el model 2:

Figura 22: Residus del model2.lin

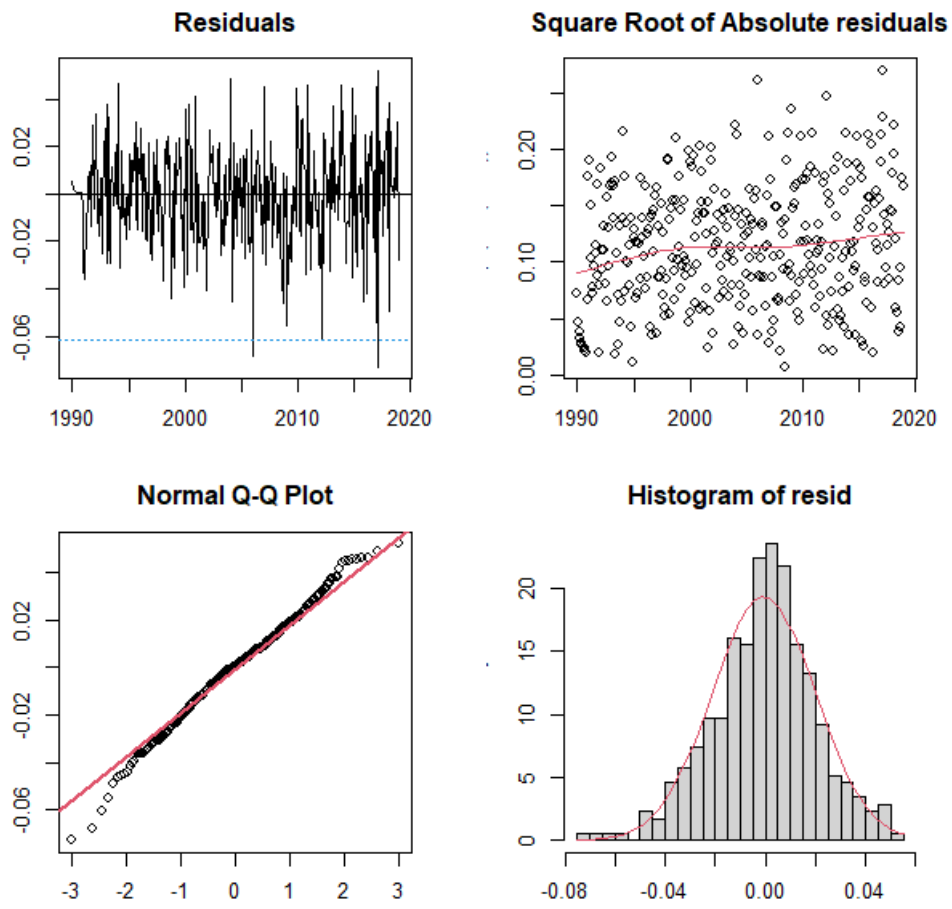
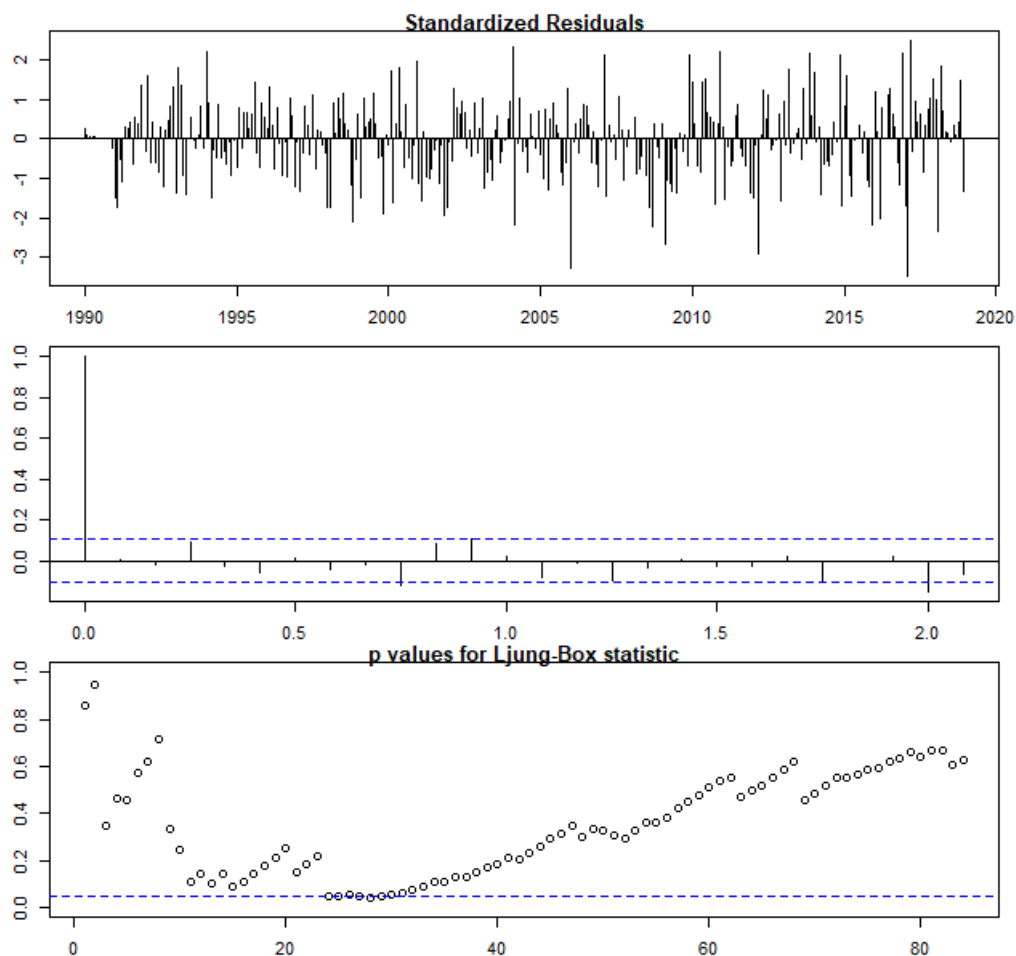


Figura 23: Test de Ljung-Box del model2.lin



Com amb el model 1 es veu com el model 2 també valida millor que abans, reduint les cues pesants i presentant normalitat de residus. I el Ljung-box és fins i tot millor que el del model 1, tenint així independència de residus. Per tant, es podria dir que aquest model valida lleugerament millor que el model 1.

Passem a veure com de bé prediu el model, fem l'enfinestrat de la sèrie linealitzada i amb els efectes de calendari (veure Apèndix E) per comprovar la seva estabilitat:

Model enfinestrat:

Coefficients :

	ar1	ma1	sma1	wTradDays.2
	0.3877	-0.8135	-0.8758	0.0011
s.e.	0.0920	0.0624	0.0346	0.0003

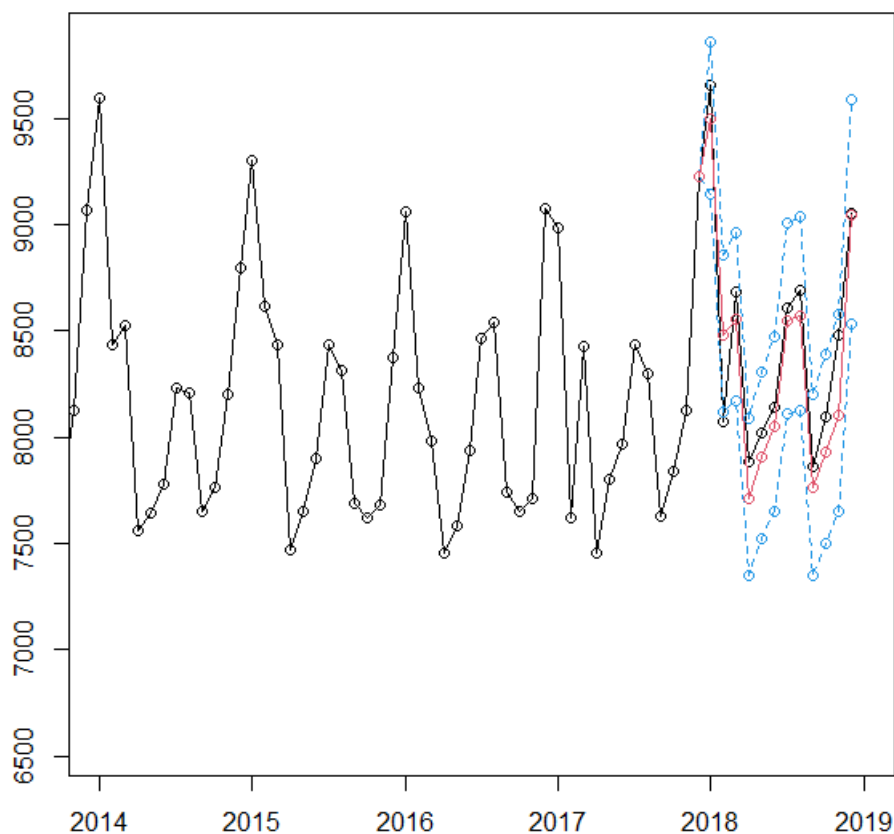
σ^2 estimated as 0.0003723: log likelihood = 807.82, aic = -1605.64

Comparativa dels coeficients:

	ar1	ma1	sma1	wTradDays
model2.lin	0.3773326	-0.8036095	-0.8813052	0.001138107
model2.lin.enfinestrat	0.3877447	-0.8135276	-0.8757732	0.001080245

El model és estable.

Figura 24: Prediccions del model2.lin



Mètriques d'exactitud i precisió:

rmse: 194.1247
mae: 157.8703
rmspe: 1.058507%
mape: 1.893588%
mean: 1.098543

Es pot veure clarament com totes les mètriques s'han reduït respecte a com havíem vist abans de tractar efectes de calendari i atípics dle model 2. Tot i així, les mètriques del model1.lin eren millors, ja que ens sortien uns errors més baixos i la mitjana de les bandes és gairebé igual, difereix de 0.003 només.

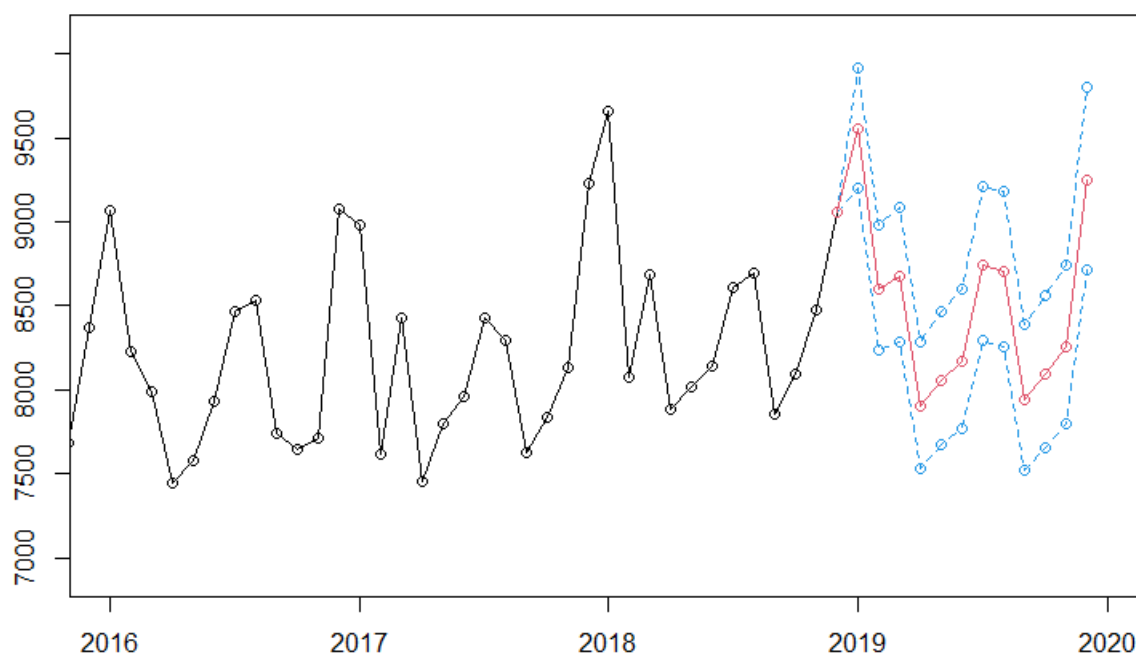
Tenint tot això en compte, hem decidit quedar-nos amb el model1.lin ja que, tot i que no valida tant bé com el model2.lin (però la diferència és mínima) i l'AIC és pitjor (per 2 unitats només), les mètriques de predicció són molt millors al model1.lin. Per tant, hem cregut que, de cara a voler predir valors futurs de l'empresa, ens interessa tenir més precisió i exactitud al nostre model; per aquests motius ens quedem amb el model1.lin. Tot i que, tant un com l'altre són bons models, es podria justificar també per termes d'AIC i validació d'agafar el segon model.

5 Prediccions a posteriori

Un cop escollit el millor model per a aquesta sèrie, hem obtingut les prediccions d'un any a posteriori de l'última dada que teníem de la sèrie.

	Jan	Feb	Mar	Apr	May	Jun
2019	9.100631	8.995567	9.004593	8.910738	8.930793	8.944456
	Jul	Aug	Sep	Oct	Nov	Dec
2019	9.011987	9.007942	8.916449	8.935313	8.955168	9.067825

Figura 25: Prediccions a un any posteriori amb el model final



Ara, aquestes prediccions són bastant similars a les primeres que havíem fet, veiem com es manté la tendència de creixement que hi havia a la sèrie original i es segueixen mantenint els patrons estacionals. Tot i que, amb aquest model final les prediccions són més precises. Segons aquest gràfic sembla que les prediccions determinin que s'hagi de mantenir bastant constant el consum d'energia (amb un lleuger creixement potser) del 2019 al 2020.

6 Conclusions

La metodologia de Box-Jenkins és una eina molt útil (i quasi imprescindible) per a analitzar sèries temporals i ajustar-hi models ARIMA. És molt important tractar correctament la sèrie i fer un anàlisi de la significació dels efectes de calendari, així com analitzar els possibles valors atípics que un model no és capaç d'explicar.

Finalment, hem obtingut un model que podem validar, que s'ajusta bé a les dades i que té bones mètriques de predicció. El model en concret és un $ARIMA(0, 1, 2)(0, 1, 1)_{12}$ amb una diferenciació regular i estacional amb un AIC de -1670.78 i les següents mètriques de predicció:

```
rmse: 178.2166
mae: 136.6682
rmspe: 0.742133%
mape: 1.634467%
mean: 1.101084
```

El model descartat (model2) era un $ARIMA(1, 1, 1)(0, 1, 1)_{12}$ amb les següents mètriques:

```
rmse: 194.1247
mae: 157.8703
rmspe: 1.058507%
mape: 1.893588%
mean: 1.098543
```

És també un bon model i podria ser un possible candidat a escollir depenent de les nostres preferències ja que té un millor AIC (-1672.14) i una millor validació del model, però l'hem deixat de banda perquè hem preferit tenir millors mètriques d'exactitud i precisió a l'hora de fer prediccions a futur.

En quant a les prediccions finals a llarg termini hem vist que ara són més precises al considerar efectes de calendari i després de fer el tractament d'atípics de la sèrie. Les prediccions conserven el patró estacional de la sèrie original i una lleugera tendència creixent (Veure Prediccions finals).

Apèndix A: Gràfiques i sortides addicionals de les sèries

Taula 1: Taula de la sèrie EnerUSA

Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
1990	7812.968	7007.057	7309.419	6771.440	6750.377	6759.778	7036.448	7283.533	6618.050	6812.415	6776.439	7546.646
1991	8078.662	6866.739	7144.067	6540.725	6659.215	6702.511	7102.869	7117.702	6662.226	6862.826	7011.115	7688.592
1992	7988.957	7247.706	7353.584	6845.998	6702.404	6719.323	7164.274	7033.412	6686.895	6906.883	7083.548	8049.211
1993	7912.457	7424.599	7816.593	6907.478	6650.484	6847.313	7313.064	7337.387	6819.006	6988.180	7285.728	8022.314
1994	8589.798	7746.402	7668.422	6983.323	6890.509	7166.434	7395.025	7477.232	6886.162	7075.224	7152.791	8008.858
1995	8226.194	7600.223	7746.002	7110.693	7155.401	7235.885	7630.435	7922.338	7140.891	7229.697	7577.393	8415.664
1996	8781.983	8109.507	8197.537	7367.170	7467.870	7450.956	7763.061	7874.805	7151.972	7622.802	7801.332	8411.306
1997	8829.389	7792.360	7990.107	7528.446	7481.767	7470.768	8096.188	7914.989	7417.611	7682.070	7754.952	8612.466
1998	8613.351	7650.206	8174.688	7515.051	7579.490	7719.460	8280.964	8230.512	7619.270	7613.627	7511.037	8473.948
1999	8916.763	7813.919	8374.959	7661.384	7627.872	7809.610	8374.703	8382.843	7616.189	7719.811	7639.151	8677.387
2000	8952.554	8408.735	8299.976	7675.245	8031.614	8008.683	8236.273	8548.829	7698.595	7766.958	7945.919	9202.867
2001	9232.646	8076.542	8478.756	7644.101	7634.334	7657.921	8149.907	8309.182	7406.412	7723.424	7550.218	8265.424
2002	8797.528	7862.338	8366.017	7763.633	7805.747	7895.255	8431.417	8351.021	7666.391	7807.864	8019.295	8838.279
2003	9240.030	8436.906	8388.090	7672.877	7668.140	7646.589	8276.561	8396.660	7611.488	7787.898	7840.975	8931.792
2004	9356.765	8739.592	8436.531	7798.779	7972.423	7987.734	8411.715	8360.396	7836.502	7954.676	8056.208	9161.797
2005	9293.639	8289.373	8732.282	7720.135	7944.437	8177.881	8587.429	8694.082	7792.942	7786.970	7944.991	9203.622
2006	8821.073	8196.253	8590.982	7715.810	7962.594	8151.441	8638.209	8737.107	7791.269	8025.726	8034.403	8799.535
2007	9224.113	8756.161	8560.982	7922.623	8020.243	8096.728	8489.795	8819.246	7940.026	7971.090	8081.360	9088.570
2008	9416.832	8655.721	8591.320	7838.374	7897.176	8043.083	8418.173	8217.800	7352.797	7785.016	7810.328	8798.727
2009	9131.508	7872.212	8051.971	7341.651	7287.596	7537.066	7814.128	7962.170	7270.279	7471.085	7445.247	8838.365
2010	9129.452	8208.891	8197.307	7365.336	7672.732	8004.568	8376.046	8440.346	7690.109	7502.370	7793.089	9228.253
2011	9294.125	8102.879	8336.086	7493.759	7574.199	7887.591	8378.721	8401.595	7554.621	7572.396	7780.887	8573.242
2012	8675.482	7962.263	7673.051	7214.299	7607.309	7728.856	8290.760	8228.099	7367.968	7572.483	7765.637	8393.366
2013	8982.147	8014.712	8379.170	7519.205	7616.574	7719.693	8270.491	8170.581	7635.619	7714.237	8127.669	9068.347
2014	9593.910	8431.769	8525.757	7558.169	7646.405	7779.348	8231.033	8207.966	7649.603	7762.415	8198.035	8797.340
2015	9300.080	8612.739	8435.045	7470.291	7647.966	7902.508	8433.824	8316.314	7689.250	7621.466	7681.646	8373.332
2016	9062.724	8231.447	7985.772	7452.048	7581.689	7934.814	8469.121	8537.499	7745.433	7651.249	7713.210	9080.218
2017	8982.378	7623.145	8430.199	7452.221	7799.991	7964.436	8432.669	8297.670	7629.796	7838.268	8129.521	9228.817
2018	9655.100	8076.522	8683.815	7884.303	8019.308	8143.949	8607.324	8694.376	7859.887	8094.910	8477.193	9054.374

Figura 26: Variància en funció de la mitjana de la sèrie EnerUSA

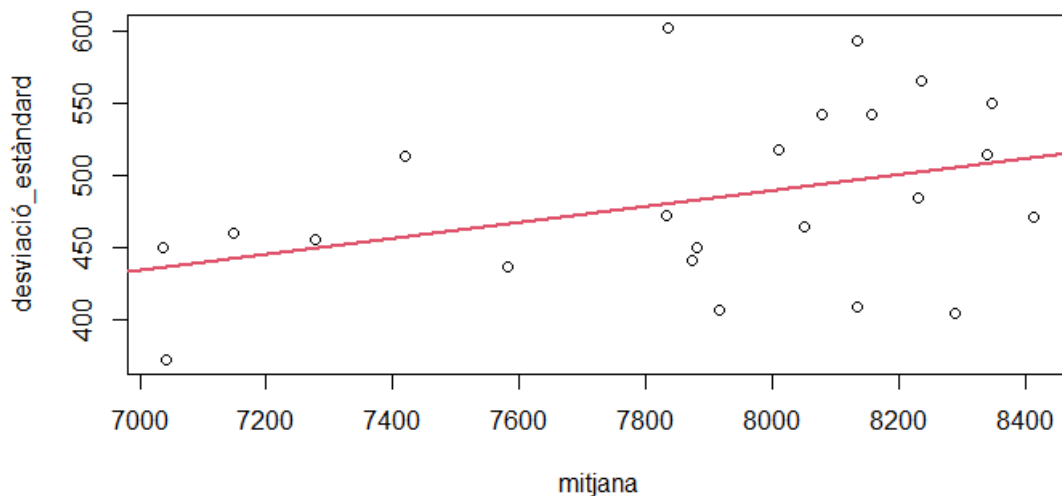


Figura 27: Variància en funció de la mitjana i boxplot de $\log(\text{serie})$

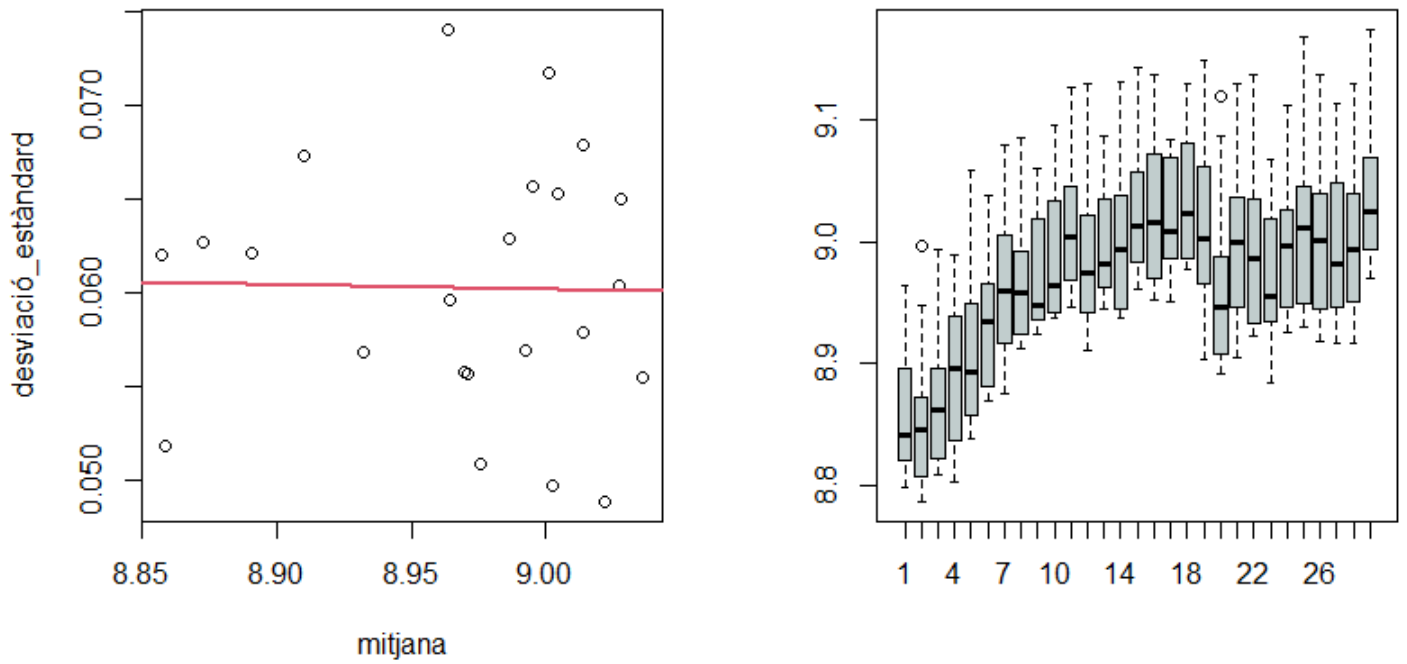
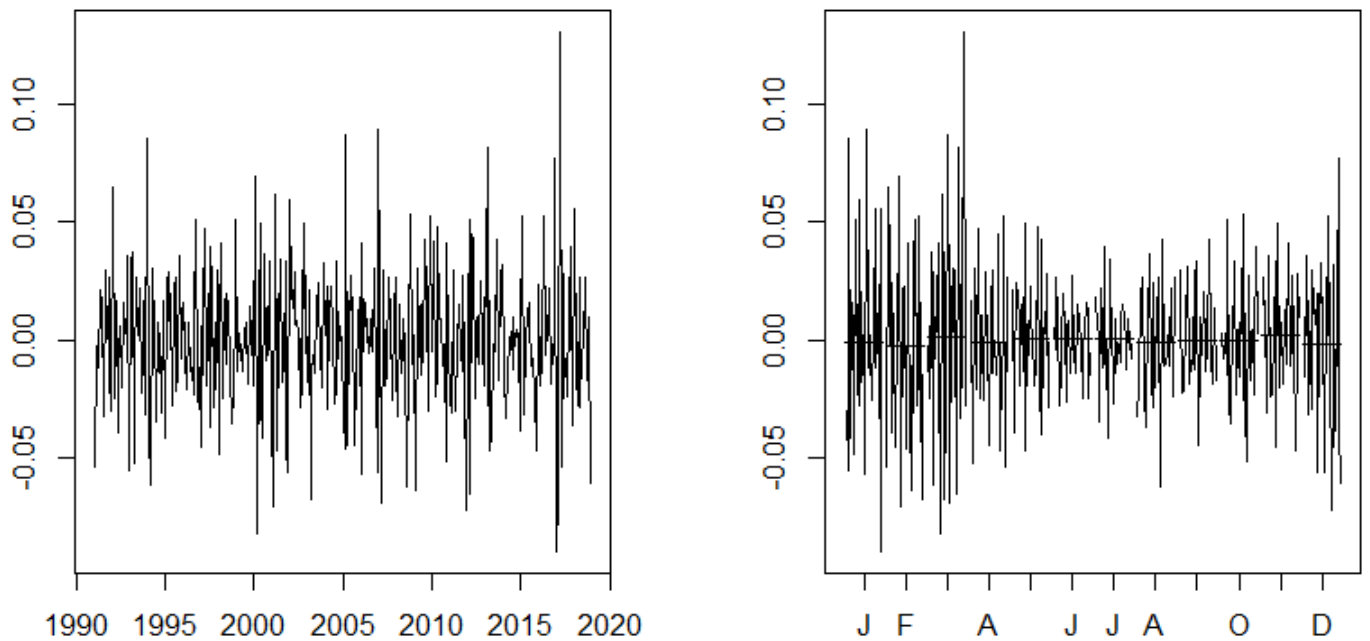


Figura 28: Sèrie estacionària: $(1 - B)(1 - B^{12})X_t$ i monthplot de la sèrie estacionària



Apèndix B: Estimació i validació dels primers models

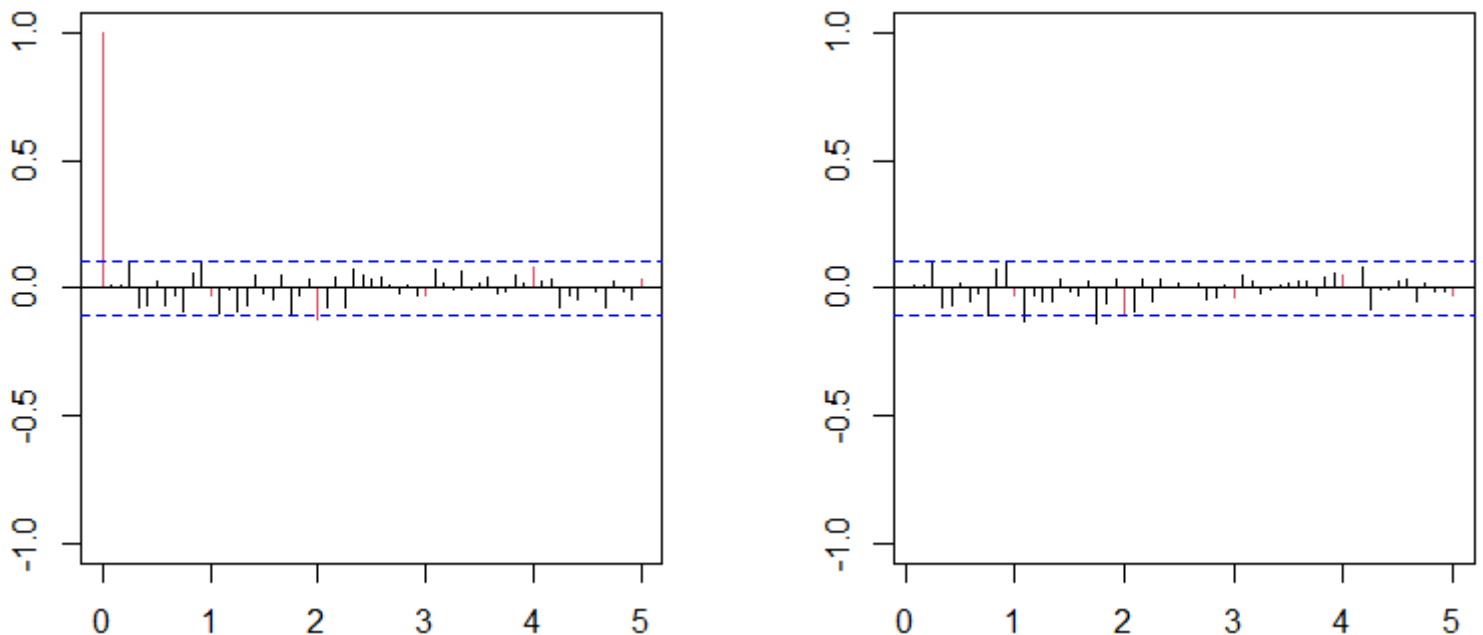
Aquí presentem amb més detall els models ajustats i els gràfics de validació d'aquests. Els models els hem ajustat amb els models *airma* del packet *stats* del llenguatge R, i els gràfics de validació els hem fet també amb el llenguatge R, fent servir funcions ja definides, com *qqplot*, *tsdiag*, *plot (forecast)*, entre d'altres.

Model 1

```
Call: arima(x = lnserie, order = c(0, 1, 3),
            seasonal = list(order = c(0, 1, 2), period = 12))
Coefficients:
            mal            ma2            ma3            smal            sma2
      -0.4852    -0.2049    0.0267    -0.7871    -0.1263
s.e.      0.0551    0.0671    0.0542    0.0659    0.0673
sigma^2 estimated as 0.0004504:  log likelihood = 804.9,  aic = -1597.8
##### ma3 no significatiu!

Call: arima(x = lnserie, order = c(0, 1, 2),
            seasonal = list(order = c(0, 1, 2), period = 12))
Coefficients:
            mal            ma2            smal            sma2
      -0.4866    -0.1909    -0.793    -0.1229
s.e.      0.0563    0.0627    0.065    0.0673
sigma^2 estimated as 0.0004502:  log likelihood = 804.78,  aic = -1599.56
```

Figura 29: ACF i PACF dels residus del model 1



Tests de validació del model 1:

Normality Tests

Shapiro–Wilk normality test
data: resid(model)
W = 0.99379, p-value = 0.1652

Anderson–Darling normality test
data: resid(model)
A = 0.60686, p-value = 0.114

Jarque Bera Test
data: resid(model)
X-squared = 3.0773, p-value = 0.2147

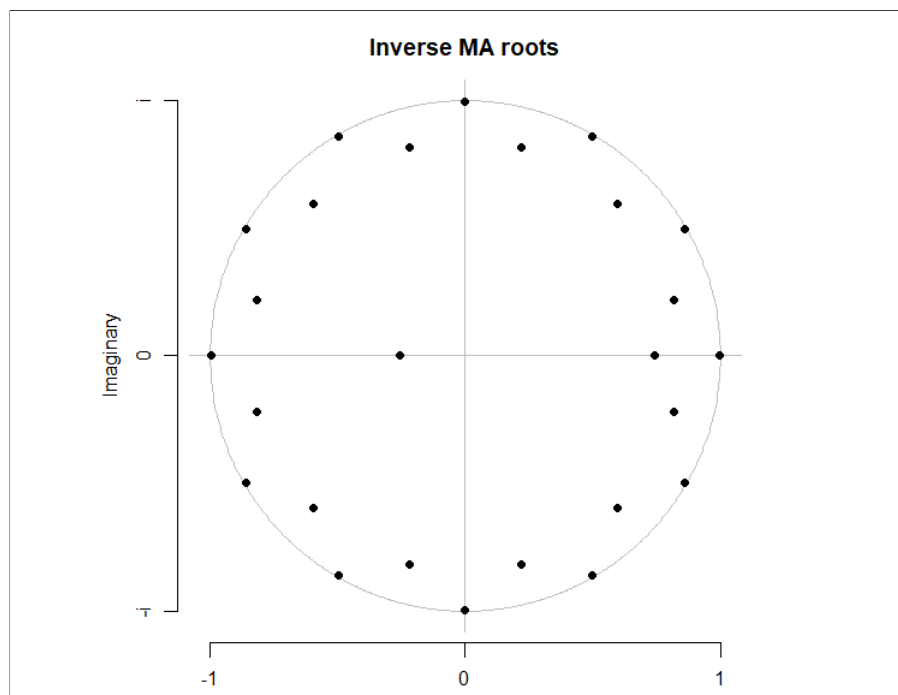
Homoscedasticity Test

studentized Breusch–Pagan test
data: resid(model)
BP = 16.091, df = 1,
p-value = 6.036e-05

Independence Tests

Durbin–Watson test
data: resid(model) ~ I(1:length(resid(model)))
DW = 1.976, p-value = 0.3903
alternative hypothesis: true autocorrelation
is greater than 0

Figura 30: Arrels inverses del model 1



Model 2

Call:
arima(x = lnserie, order = c(1, 1, 1), seasonal = list(order = c(0, 1, 2),
period = 12))

Coefficients:

	ar1	ma1	sma1	sma2
	0.2930	-0.7918	-0.8006	-0.1159
s.e.	0.0939	0.0654	0.0656	0.0679

sigma^2 estimated as 0.0004505: log likelihood = 804.56,
aic = -1599.12

Figura 31: ACF i PACF dels residus del model 2

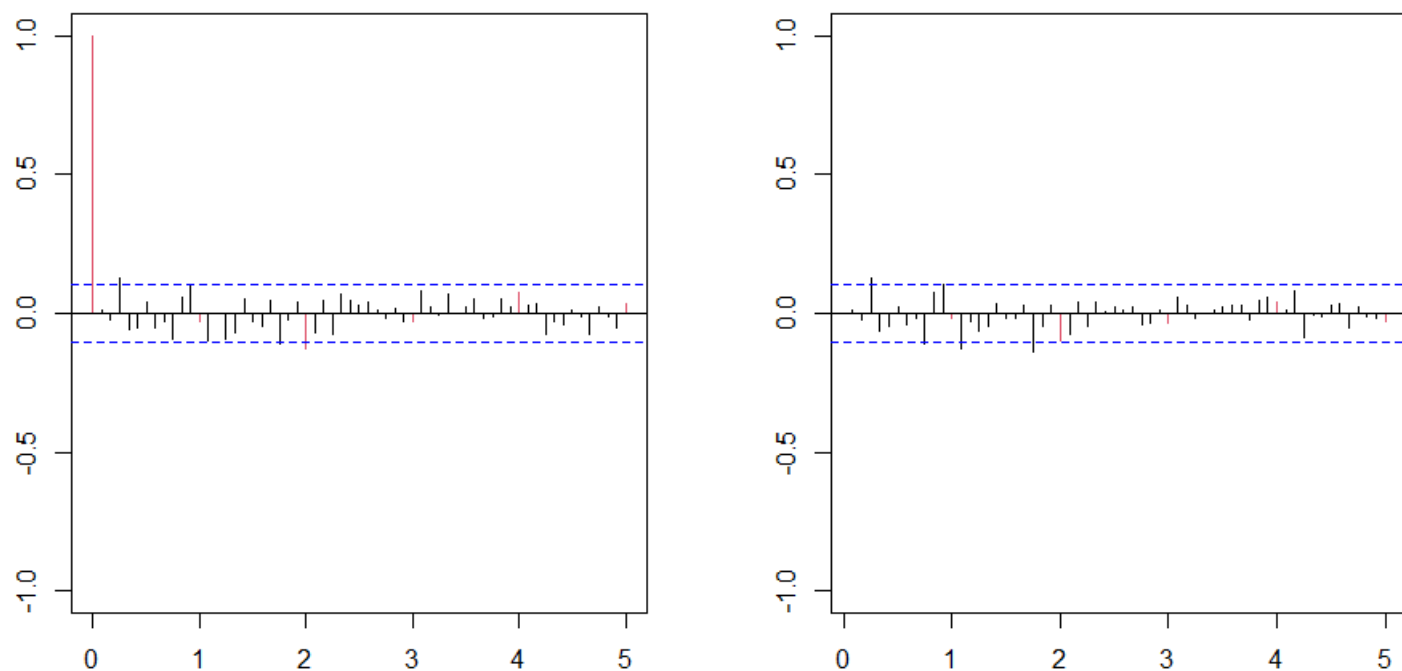
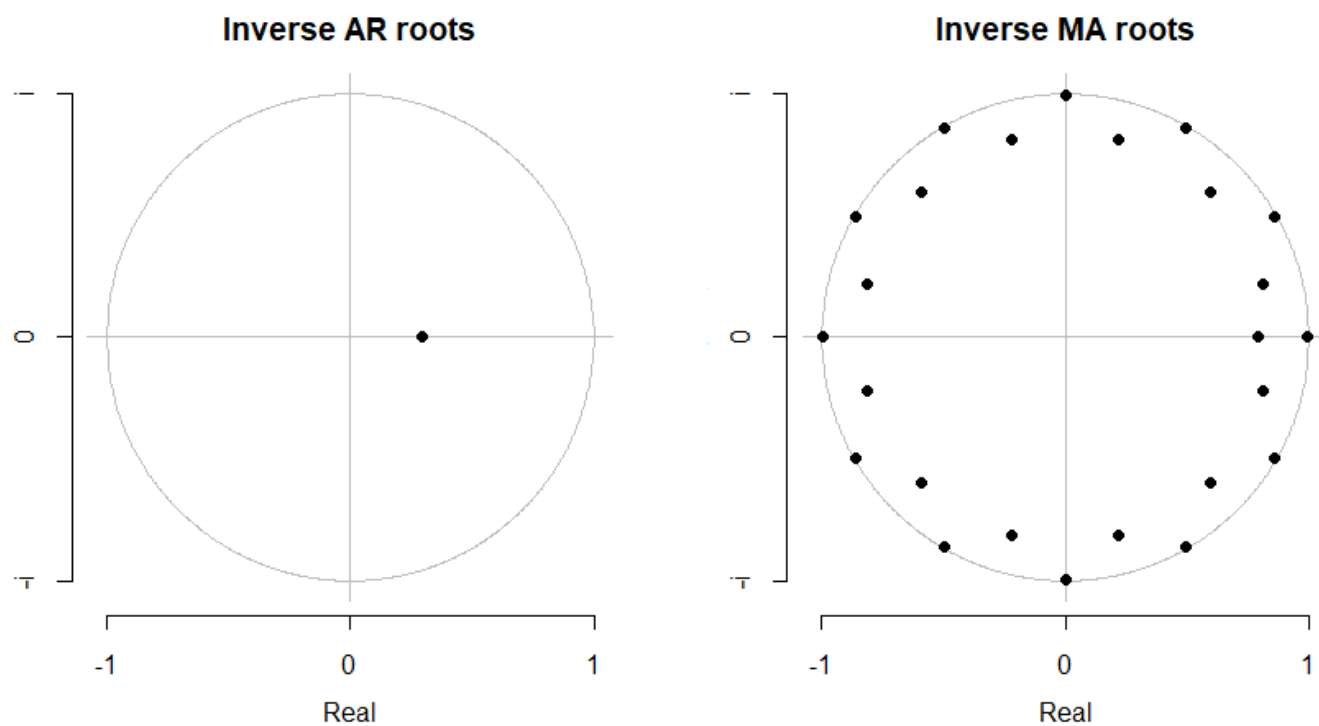


Figura 32: Arrels inverses del model 2



Tests de validació del model 2:

Normality Tests

Shapiro–Wilk normality test
data: resid(model)
W = 0.99379, p-value = 0.1652

Anderson–Darling normality test
data: resid(model)
A = 0.60686, p-value = 0.114

Jarque Bera Test
data: resid(model)
X-squared = 3.0773,
p-value = 0.2147

Homoscedasticity Test

studentized Breusch–Pagan test
data: resid(model)
BP = 16.091, df = 1,
p-value = 6.036e-05

Independence Tests

Durbin–Watson test
data: resid(model) ~ I(1:length(resid(model)))
DW = 1.976, p-value = 0.3903
alt hip.: true autocorrelation is greater than 0

Apèndix C: Capacitat predictiva dels primers models

Aquí es presenta amb més detall el càlcul de les mètriques de predicció dels diferents models. Les prediccions i les mesures de predicció les hem calculat amb el següent codi (amb el software R), on *serie* representa qualsevol sèrie estacionària.

```
serie2 = window(serie , end=c(2017,12))
obs=window(serie , start=2018)

#### ajustar model.a(serie) i model.b(serie2)
...

pr=predict(model.b, n.ahead=12)
ll=exp(pr$pred-1.96*pr$se)
ul=exp(pr$pred+1.96*pr$se)
pred=exp(pr$pred)

rmse = sqrt(mean((obs-pred)^2))
mae = mean(abs(obs-pred))
rmspe = sqrt(mean(((obs-pred)/obs)^2))*100
mape = mean(abs((obs-pred)/obs))*100
mean = mean(ul-ll)
```

Model 1

```
Call: #modell.a
arima(x = lnserie , order = c(0, 1, 2),
      seasonal = list(order = c(0, 1, 2), period = 12))
```

Coefficients:

	mal	ma2	sma1	sma2
	-0.4866	-0.1909	-0.793	-0.1229
s.e.	0.0563	0.0627	0.065	0.0673

sigma^2 estimated as 0.0004502: log likelihood = 804.78, aic = -1599.56

```
Call: #modell.b
arima(x = lnserie2 , order = c(0, 1, 2), seasonal = list(order = c(0, 1, 2),
  period = 12))
```

Coefficients:

	mal	ma2	sma1	sma2
	-0.4694	-0.2082	-0.8286	-0.0798
s.e.	0.0569	0.0637	0.0686	0.0718

sigma^2 estimated as 0.0004482: log likelihood = 776.59, aic = -1543.17

Els valors obtinguts són:

		ll	pred	ul	serie
Jan	2018	8978.085	9358.709	9755.470	9655.100
Feb	2018	7939.960	8322.070	8722.568	8076.522
Mar	2018	8132.678	8540.005	8967.732	8683.815
Apr	2018	7348.054	7729.981	8131.760	7884.303
May	2018	7484.186	7886.884	8311.250	8019.308
Jun	2018	7631.036	8055.173	8502.885	8143.949
Jul	2018	8070.048	8532.494	9021.440	8607.324
Aug	2018	8055.930	8531.077	9034.249	8694.376
Sep	2018	7353.555	7799.295	8272.054	7859.887
Oct	2018	7460.254	7924.362	8417.342	8094.910
Nov	2018	7614.835	8100.428	8616.987	8477.193
Dec	2018	8545.025	9102.969	9697.343	9054.374

Model 2

```
Call: #model2.a
arima(x = lnserie, order = c(1, 1, 1), seasonal = c(0, 1, 2), period = 12)
Coefficients:
      ar1      mal      smal      sma2
    0.2930  -0.7918  -0.8006  -0.1159
s.e.    0.0939   0.0654   0.0656   0.0679
sigma^2 estimated as 0.0004505:  log likelihood = 804.56,  aic = -1599.12
```

```
Call: #model2.b
arima(x = lnserie2, order = c(1, 1, 1), seasonal = c(0, 1, 2), period = 12)
Coefficients:
      ar1      mal      smal      sma2
    0.3314  -0.8112  -0.8408  -0.0657
s.e.    0.0940   0.0647   0.0696   0.0730
sigma^2 estimated as 0.0004483:  log likelihood = 776.53,  aic = -1543.05
```

Els valors obtinguts són:

		ll	pred	ul	serie
Jan	2018	8985.869	9366.837	9763.958	9655.100
Feb	2018	7949.502	8330.413	8729.576	8076.522
Mar	2018	8111.991	8520.634	8949.863	8683.815
Apr	2018	7328.791	7710.627	8112.356	7884.303
May	2018	7464.013	7864.033	8285.490	8019.308
Jun	2018	7614.581	8033.317	8475.080	8143.949
Jul	2018	8055.992	8509.861	8989.302	8607.324
Aug	2018	8045.800	8509.662	9000.266	8694.376
Sep	2018	7345.829	7778.771	8237.228	7859.887
Oct	2018	7451.448	7899.994	8375.540	8094.910
Nov	2018	7607.096	8074.405	8570.422	8477.193
Dec	2018	8538.905	9073.813	9642.229	9054.374

Apèndix D: Tractament d'Efectes de Calendari

Sortides numèriques i gràfiques del procediment de tractament d'efectes de calendari de la sèrie. Incloem: variables exògenes utilitzades, models ajustats per l'anàlisi d'intervenció i plots i sortides numèriques per a la validació i capacitats predictives del nou model ajustat.

Variables exògenes utilitzades:

Taula 2: Trading Days

Any	Jan	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
1990	3.0	0.0	-0.5	-1.5	3.0	-1.5	-0.5	3.0	-5.0	3.0	2.0	-4.0
1991	3.0	0.0	-4.0	2.0	3.0	-5.0	3.0	-0.5	-1.5	3.0	-1.5	-0.5
1992	3.0	-2.5	-0.5	2.0	-4.0	2.0	3.0	-4.0	2.0	-0.5	-1.5	3.0
1993	-4.0	0.0	3.0	2.0	-4.0	2.0	-0.5	-0.5	2.0	-4.0	2.0	3.0
1994	-4.0	0.0	3.0	-1.5	-0.5	2.0	-4.0	3.0	2.0	-4.0	2.0	-0.5
1995	-0.5	0.0	3.0	-5.0	3.0	2.0	-4.0	3.0	-1.5	-0.5	2.0	-4.0
1996	3.0	1.0	-4.0	2.0	3.0	-5.0	3.0	-0.5	-1.5	3.0	-1.5	-0.5
1997	3.0	0.0	-4.0	2.0	-0.5	-1.5	3.0	-4.0	2.0	3.0	-5.0	3.0
1998	-0.5	0.0	-0.5	2.0	-4.0	2.0	3.0	-4.0	2.0	-0.5	-1.5	3.0
1999	-4.0	0.0	3.0	2.0	-4.0	2.0	-0.5	-0.5	2.0	-4.0	2.0	3.0
2000	-4.0	1.0	3.0	-5.0	3.0	2.0	-4.0	3.0	-1.5	-0.5	2.0	-4.0
2001	3.0	0.0	-0.5	-1.5	3.0	-1.5	-0.5	3.0	-5.0	3.0	2.0	-4.0
2002	3.0	0.0	-4.0	2.0	3.0	-5.0	3.0	-0.5	-1.5	3.0	-1.5	-0.5
2003	3.0	0.0	-4.0	2.0	-0.5	-1.5	3.0	-4.0	2.0	3.0	-5.0	3.0
2004	-0.5	-2.5	3.0	2.0	-4.0	2.0	-0.5	-0.5	2.0	-4.0	2.0	3.0
2005	-4.0	0.0	3.0	-1.5	-0.5	2.0	-4.0	3.0	2.0	-4.0	2.0	-0.5
2006	-0.5	0.0	3.0	-5.0	3.0	2.0	-4.0	3.0	-1.5	-0.5	2.0	-4.0
2007	3.0	0.0	-0.5	-1.5	3.0	-1.5	-0.5	3.0	-5.0	3.0	2.0	-4.0
2008	3.0	1.0	-4.0	2.0	-0.5	-1.5	3.0	-4.0	2.0	3.0	-5.0	3.0
2009	-0.5	0.0	-0.5	2.0	-4.0	2.0	3.0	-4.0	2.0	-0.5	-1.5	3.0
2010	-4.0	0.0	3.0	2.0	-4.0	2.0	-0.5	-0.5	2.0	-4.0	2.0	3.0
2011	-4.0	0.0	3.0	-1.5	-0.5	2.0	-4.0	3.0	2.0	-4.0	2.0	-0.5
2012	-0.5	1.0	-0.5	-1.5	3.0	-1.5	-0.5	3.0	-5.0	3.0	2.0	-4.0
2013	3.0	0.0	-4.0	2.0	3.0	-5.0	3.0	-0.5	-1.5	3.0	-1.5	-0.5
2014	3.0	0.0	-4.0	2.0	-0.5	-1.5	3.0	-4.0	2.0	3.0	-5.0	3.0
2015	-0.5	0.0	-0.5	2.0	-4.0	2.0	3.0	-4.0	2.0	-0.5	-1.5	3.0
2016	-4.0	1.0	3.0	-1.5	-0.5	2.0	-4.0	3.0	2.0	-4.0	2.0	-0.5
2017	-0.5	0.0	3.0	-5.0	3.0	2.0	-4.0	3.0	-1.5	-0.5	2.0	-4.0
2018	3.0	0.0	-0.5	-1.5	3.0	-1.5	-0.5	3.0	-5.0	3.0	2.0	-4.0

Taula 3: Easter

Año	Jan	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
1990	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1991	0.00	0.00	0.50	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1992	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1993	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1994	0.00	0.00	1.67	-1.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1995	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1996	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1997	0.00	0.00	0.50	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1998	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1999	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2000	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2001	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2002	0.00	0.00	0.50	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2003	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2004	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2005	0.00	0.00	0.50	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2006	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2007	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2008	0.00	0.00	0.50	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2009	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2010	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2011	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2012	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2013	0.00	0.00	0.50	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2014	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2015	0.00	0.00	-0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2016	0.00	0.00	0.50	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2017	0.00	0.00	-0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2018	0.00	0.00	0.50	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Models estimats per l'anàlisi d'intervenció:

Call:

```
arima(x = lnserie , order = c(0, 1, 2), seasonal = list(order = c(0, 1, 2),
period = 12), xreg = wTradDays)
```

Coefficients:

```
      ma1      ma2      sma1      sma2  wTradDays
-0.4680 -0.2027 -0.8214 -0.0820      0.0011
s.e.    0.0548  0.0597  0.0651  0.0686      0.0003
sigma^2 estimated as 0.0004361: log likelihood = 810.71, aic = -1609.42
```

sma2 no significatiu !

Call:

```
arima(x = lnserie , order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1),
period = 12), xreg = wTradDays)
```

Coefficients:

```
      ma1      ma2      sma1  wTradDays
-0.4685 -0.2040 -0.8850      0.0011
s.e.    0.0544  0.0589  0.0376      0.0003
sigma^2 estimated as 0.0004394: log likelihood = 810.01, aic = -1610.02
```



```

Call:
arima(x = lnserie, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 2),
                                                         period = 12), xreg = wEast)

Coefficients:
          ma1          ma2          sma1          sma2          wEast
      -0.4870   -0.1903   -0.7928   -0.1230   -0.0003
s.e.    0.0566    0.0632    0.0651    0.0673    0.0044
sigma^2 estimated as 0.0004502:  log likelihood = 804.78,  aic = -1597.56

```

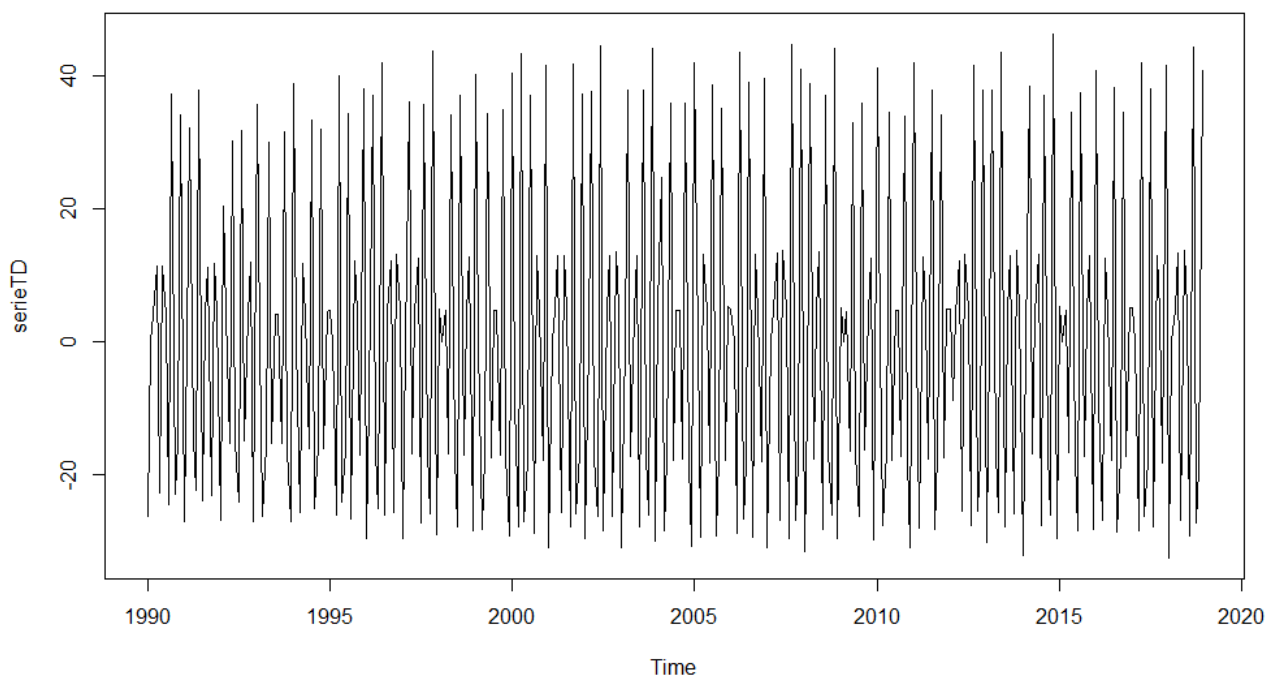
```

Call:
arima(x = lnserie, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 2),
                                                         period = 12), xreg = data.frame(wTradDays, wEast))

Coefficients:
          ma1          ma2          sma1          sma2  wTradDays  wEast
      -0.4656   -0.2058   -0.8227   -0.0812      0.0011  0.0015
s.e.    0.0553    0.0605    0.0652    0.0687      0.0003  0.0043
sigma^2 estimated as 0.0004359:  log likelihood = 810.77,  aic = -1607.54

```

Figura 33: Representació dels efectes de calendari com a serieTD - serie



Validació del model 1.TD

Figura 34: Arrels inverses del model 1.TD

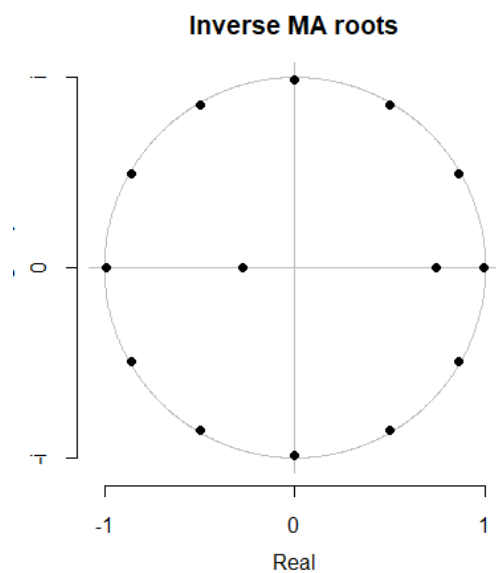
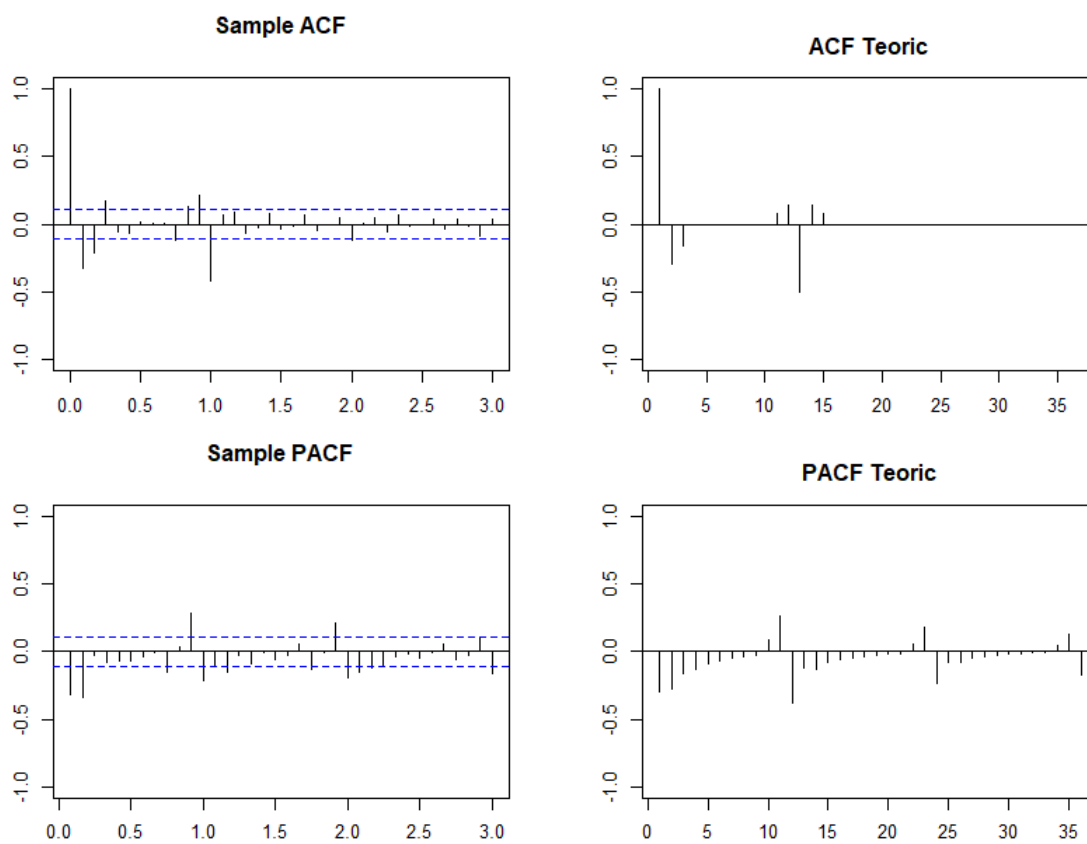


Figura 35: Comparació entre ACF, PACF obtinguts vs teòrics.



Validació del model 2.TD

Figura 36: Arrels inverses del model 1.TD

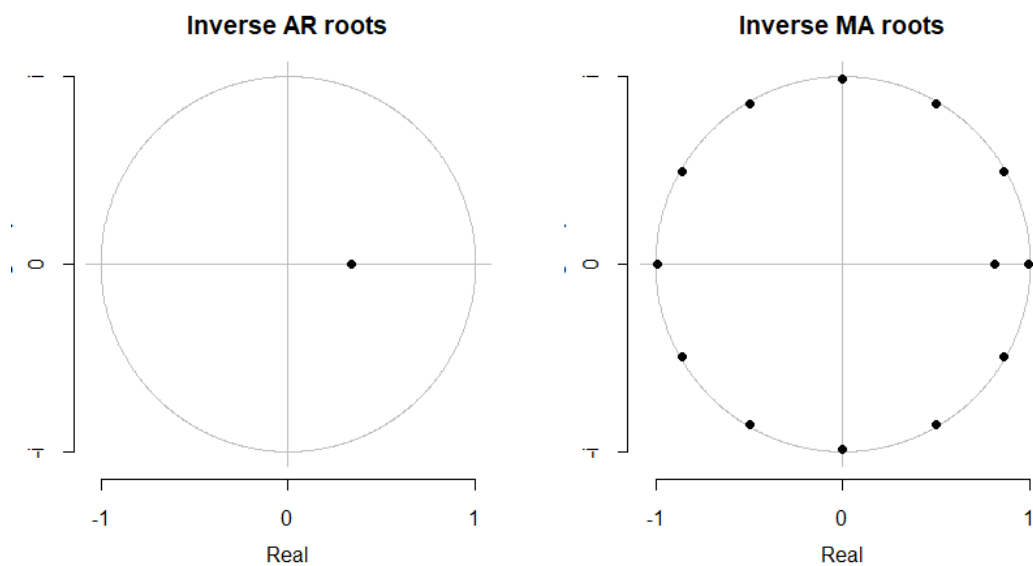
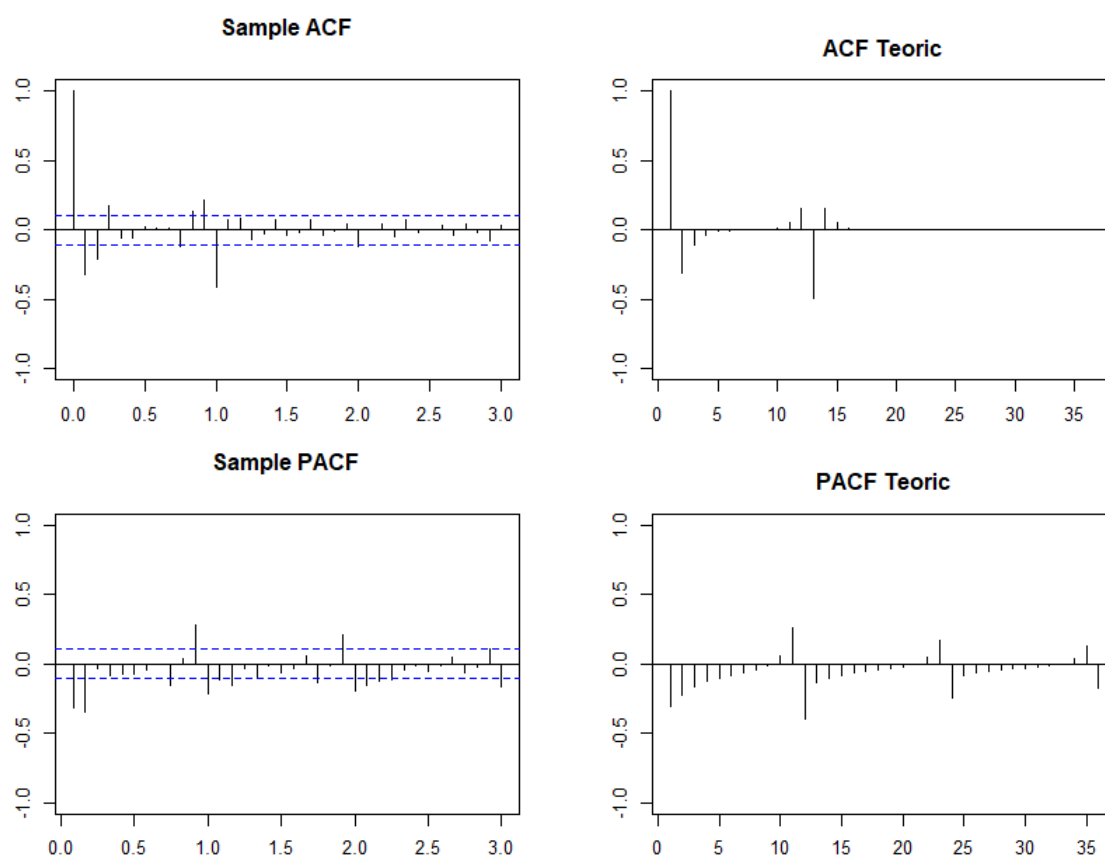


Figura 37: Comparació entre ACF, PACF obtinguts vs teòrics.



Apèndix E: Tractament d'atípics

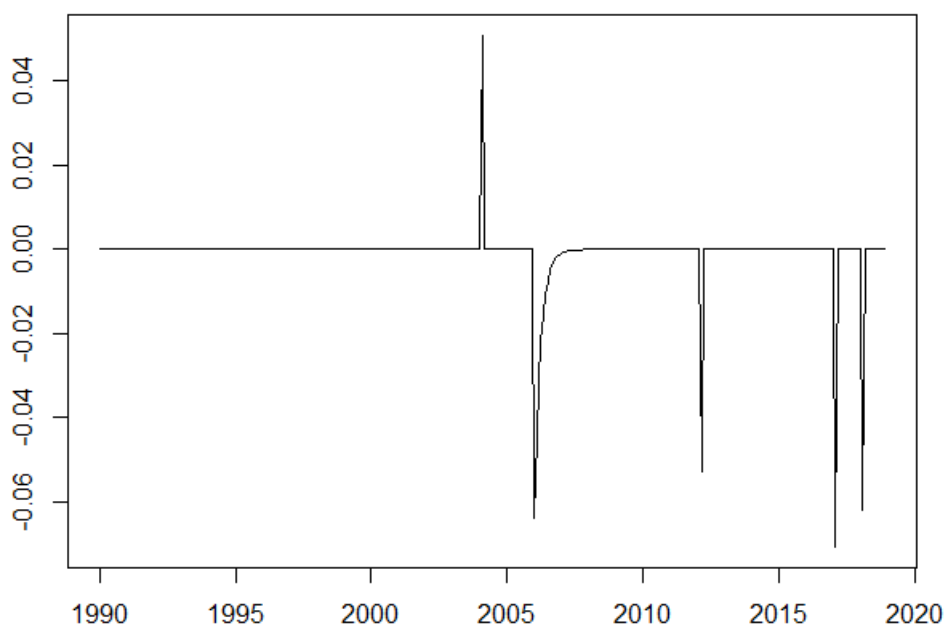
Informació adicional del procediment de tractament d'atípics.

Càlcul d'atípics i linealització de la sèrie

```
source(" atipics2.r")
m1TD.atip=outdetec(m1TD,dif=c(1,12),crit=3.0,LS=F)
atipics=m4TD.atip$atip[order(m4TD.atip$atip[,1]),]
meses=c("Ene","Feb","Mar","Abr","May","Jun","Jul",
        "Ago","Sep","Oct","Nov","Dic")
m1TD.atip$sigma2
data.frame(atipics,Fecha=paste(meses[(atipics[,1]-1)%%12+1],
                                start(serie)[1]+((atipics[,1]-1)%/%12)),
           perc.Obs=atipics[,3])

lnserie.lin = lineal(lnserie , m1TD.atip$atip)
```

Figura 38: Respresentació dels atípics com a $\lnserie - \lnserie.lin$



Enfinestrat del la sèrie linealitzada i ajust de la nova sèrie amb el model1:

```
ultim=c(2017,12)
```

```
lnserie.lin.2=window(lnserie.lin,end=ultim)
```

```
wTradDays.2=window(wTradDays,end=ultim)
```

```
(m1TD.lin)
```

Call:

```
arima(x = lnserie.lin, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1),
period = 12), xreg = wTradDays)
```

Coefficients:

	ma1	ma2	sma1	wTradDays
	-0.4299	-0.2013	-0.8793	0.0011
s.e.	0.0538	0.0580	0.0344	0.0003

sigma^2 estimated as 0.0003672: log likelihood = 840.39, aic = -1670.78

```
(m1TD.lin.2<-arima(lnserie.lin.2,order=c(0,1,2),seasonal=list(order=c(0,1,1),period=12),xreg=wTradDays.2))
```

```
Call:
arima(x = lnserie.lin.2, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1), period = 12), xreg = wTradDays.2)
```

Coefficients:

	ma1	ma2	sma1	wTradDays.2
	-0.4303	-0.2051	-0.8743	0.0011
s.e.	0.0549	0.0593	0.0354	0.0003

sigma^2 estimated as 0.0003743: log likelihood = 807.07, aic = -1604.15

```
(rbind(coef(m1TD.lin),coef(m1TD.lin.2)))
```

	ma1	ma2	sma1	wTradDays
[1,]	-0.4298823	-0.2012795	-0.8792689	0.001112973
[2,]	-0.4303304	-0.2050631	-0.8743013	0.001055126

Enfinestrat del la sèrie linealitzada i ajust de la nova sèrie amb el model2:

```
ultim=c(2017,12)
```

```
lnserie.lin.2=window(lnserie.lin,end=ultim)
wTradDays.2=window(wTradDays,end=ultim)
```

```
(m2TD.lin)
```

Call:

```
arima(x = lnserie.lin, order = c(1, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12), xreg = wTradDays)
```

Coefficients:

	ar1	ma1	sma1	wTradDays
	0.3773	-0.8036	-0.8813	0.0011
s.e.	0.0905	0.0613	0.0338	0.0003

sigma^2 estimated as 0.0003653: log likelihood = 841.07, aic = -1672.14

```
(m2TD.lin.2<-arima(lnserie.lin.2,order=c(1,1,1),seasonal=list(order=c(0,1,1),period=12),xreg=wTradDays.2))
```

Call:

```
arima(x = lnserie.lin.2, order = c(1, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12), xreg = wTradDays.2)
```

Coefficients:

	ar1	ma1	sma1	wTradDays.2
	0.3877	-0.8135	-0.8758	0.0011
s.e.	0.0920	0.0624	0.0346	0.0003

sigma^2 estimated as 0.0003723: log likelihood = 807.82, aic = -1605.64

(rbind(coef(m2TD.lin),coef(m2TD.lin.2)))

	ar1	ma1	sma1	wTradDays
[1,]	0.3773326	-0.8036095	-0.8813052	0.001138107
[2,]	0.3877447	-0.8135276	-0.8757732	0.001080245

Apèndix F: Codis

Codis usats per a la creació de variables exògenes per al tractament d'efectes de calendari, detecció d'atípics i linealització de sèries.

Funció outdetec

```
outdetec<-function(object,dif=c(0,0),crit,LS=T)
{
  residuals<-object$residuals
  m<-length(residuals)

  piweight<-ARMAtoMA(ar=object$model$theta,ma=-object$model$phi,lag.max=m+sum(dif))

  if (dif[1]!=0) for(i in 1:dif[1]) piweight<-c(piweight,0)-c(-1,piweight)
  if (length(dif)>1){
    for (i in 2:length(dif)){
      if (dif[i]>1) piweight<-c(piweight,rep(0,dif[i]))-c(rep(0,dif[i]-1),-1,piweight)
    }
  }
  piweight<-piweight[1:m]

  atip<-NULL

  num<-NULL
  type<-NULL
  wcoeff<-NULL
  LCrit<-NULL
  if (crit<=0) {cat("The Critical value may be positive") }

  va<-mean(residuals^2)

  c<-cumsum(piweight)-1

  d<-rep(0,m)
  delta<-0.7
  d[1]<-piweight[1]-delta
  for (i in 2:m) d[i]<-delta*d[i-1]+piweight[i]

  sum1<-1+sum(piweight*piweight)
  sum2<-1+sum(c*c)
  sum3<-1+sum(d*d)

  maxL<-crit+1

  while (maxL>crit)
  {
    ka1<-sum1
    ks1<-sum2
    kt1<-sum3
    maxL<-0
    for (i in 1:m)
    {
      sum1<-sum(residuals[i:m]*c(1,-piweight[1:(m-i)]))
      sum2<-sum(residuals[i:m]*c(1,-c[1:(m-i)]))
      sum3<-sum(residuals[i:m]*c(1,-d[1:(m-i)]))

      ka1<-ka1 - piweight[m-i+1]*piweight[m-i+1]
      w_ao<-sum1/ka1
      v_ao<-va/ka1
      l_ao<-w_ao/sqrt(v_ao)

      ks1<-ks1 - c[m-i+1]*c[m-i+1]
```

```

w_ls<-suma2/ks1
v_ls<-va/ks1
l_ls<-w_ls/sqrt(v_ls)

kt1<-kt1 - d[m-i+1]*d[m-i+1]
w_tc<-suma3/kt1
v_tc<-va/kt1
l_tc<-w_tc/sqrt(v_tc)

if(abs(l_ao)>maxL & i>sum(dif))
{
  maxL<-abs(l_ao)
  t<-i
  w<-w_ao
  v<-v_ao
  ts<-"AO"
}
if(abs(l_ls)>maxL & LS==T & i!=m & i>sum(dif))
{
  maxL<-abs(l_ls)
  t<-i
  w<-w_ls
  v<-v_ls
  ts<-"LS"
}
if(abs(l_tc)>maxL & i!=m & i>sum(dif))
{
  maxL<-abs(l_tc)
  t<-i
  w<-w_tc
  v<-v_tc
  ts<-"TC"
}
}

if(maxL > crit){
  if(ts=="AO") residuals[t:m]<-residuals[t:m]+w*c(-1,piweight[1:(m-t)])
  if(ts=="LS") residuals[t:m]<-residuals[t:m]+w*c(-1,c[1:(m-t)])
  if(ts=="TC") residuals[t:m]<-residuals[t:m]+w*c(-1,d[1:(m-t)])

  val<-mean(residuals^2)
  l<-w/sqrt(v*val/va)
  va<-val

  num<-c(num,t)
  type<-c(type,ts)
  wcoeff<-c(wcoeff,w)
  LCrit<-c(LCrit,abs(l))

  atip<-data.frame(Obs=num,type_detected=type,W_coeff=wcoeff,ABS_L_Ratio=LCrit)
}
}
return(list(atip=atip,sigma2=va,resid=residuals))
}

```


Funció lineal

```
lineal<-function(serie , atip)
{
  m<-length(serie)
  for(i in 1:nrow(atip))
  {
    t<-atip[i,1]
    ts<-atip[i,2]
    w<-atip[i,3]
    if(ts=="TC")      serie[t:m]<-serie[t:m]-w*c(1,0.7^(1:(m-t)))
    if(ts=="LS")      serie[t:m]<-serie[t:m]-w
    if(ts=="AO")      serie[t]<-serie[t]-w
  }

  return(serie)
}
```

Funció Weaster

```
library(chron)
Weaster=function(data , idur=6){
EasterCalendar = matrix(
  c(1901, 4, 7, 1902, 3, 30, 1903, 4, 12, 1904, 4, 3, 1905, 4, 23, 1906, 4, 15, 1907, 3, 31,
    1908, 4, 19, 1909, 4, 11, 1910, 3, 27, 1911, 4, 16, 1912, 4, 7, 1913, 3, 23, 1914, 4, 12,
    1915, 4, 4, 1916, 4, 23, 1917, 4, 8, 1918, 3, 31, 1919, 4, 20, 1920, 4, 4, 1921, 3, 27,
    1922, 4, 16, 1923, 4, 1, 1924, 4, 20, 1925, 4, 12, 1926, 4, 4, 1927, 4, 17, 1928, 4, 8,
    1929, 3, 31, 1930, 4, 20, 1931, 4, 5, 1932, 3, 27, 1933, 4, 16, 1934, 4, 1, 1935, 4, 21,
    1936, 4, 12, 1937, 3, 28, 1938, 4, 17, 1939, 4, 9, 1940, 3, 24, 1941, 4, 13, 1942, 4, 5,
    1943, 4, 25, 1944, 4, 9, 1945, 4, 1, 1946, 4, 21, 1947, 4, 6, 1948, 3, 28, 1949, 4, 17,
    1950, 4, 9, 1951, 3, 25, 1952, 4, 13, 1953, 4, 5, 1954, 4, 18, 1955, 4, 10, 1956, 4, 1,
    1957, 4, 21, 1958, 4, 6, 1959, 3, 29, 1960, 4, 17, 1961, 4, 2, 1962, 4, 22, 1963, 4, 14,
    1964, 3, 29, 1965, 4, 18, 1966, 4, 10, 1967, 3, 26, 1968, 4, 14, 1969, 4, 6, 1970, 3, 29,
    1971, 4, 11, 1972, 4, 2, 1973, 4, 22, 1974, 4, 14, 1975, 3, 30, 1976, 4, 18, 1977, 4, 10,
    1978, 3, 26, 1979, 4, 15, 1980, 4, 6, 1981, 4, 19, 1982, 4, 11, 1983, 4, 3, 1984, 4, 22,
    1985, 4, 7, 1986, 3, 30, 1987, 4, 19, 1988, 4, 3, 1989, 3, 26, 1990, 4, 15, 1991, 3, 31,
    1992, 4, 19, 1993, 4, 11, 1994, 4, 3, 1995, 4, 16, 1996, 4, 7, 1997, 3, 30, 1998, 4, 12,
    1999, 4, 4, 2000, 4, 23, 2001, 4, 15, 2002, 3, 31, 2003, 4, 20, 2004, 4, 11, 2005, 3, 27,
    2006, 4, 16, 2007, 4, 8, 2008, 3, 23, 2009, 4, 12, 2010, 4, 4, 2011, 4, 24, 2012, 4, 8,
    2013, 3, 31, 2014, 4, 20, 2015, 4, 5, 2016, 3, 27, 2017, 4, 16, 2018, 4, 1, 2019, 4, 21,
    2020, 4, 12, 2021, 4, 4, 2022, 4, 17, 2023, 4, 9, 2024, 3, 31, 2025, 4, 20, 2026, 4, 5,
    2027, 3, 28, 2028, 4, 16, 2029, 4, 1, 2030, 4, 21, 2031, 4, 13, 2032, 3, 28, 2033, 4, 17,
    2034, 4, 9, 2035, 3, 25, 2036, 4, 13, 2037, 4, 5, 2038, 4, 25, 2039, 4, 10, 2040, 4, 1,
    2041, 4, 21, 2042, 4, 6),
  ncol=3,byrow=T,dimnames=list(1901:2050,c("Year","Month","Day")))
Holdays=rep(0,150)
Holdays[EasterCalendar[,2]==4]=ifelse(EasterCalendar[EasterCalendar[,2]==4,3]>idur,idur,
  EasterCalendar[EasterCalendar[,2]==4,3]-1)
Weight=(idur/2-Holdays)/idur
Wcalend=rep(0,150*12)
Wcalend[12*(0:149)+3]=Weight
Wcalend[12*(0:149)+4]=-Weight

window(ts(Wcalend,start=c(1901,1),freq=12),start=data[1:2],
  end=c(data[1]+data[3]%/%12,(data[2]+data[3]-1)%%12))
}
```

Funció Wtrad

```
library(chron)
Wtrad=function(data){
  dmes=c(31,28,31,30,31,30,31,31,30,31,30,31)
  valu=function(par){
    anyo=par[1]
    if (leap.year(anyo)) dmes[2]=29
    mes1=par[2]
    we=sum(table weekdays(dates(paste(mes1,"/1/",anyo,sep="")))+0:(dmes[mes1]-1)))[c(1,7)]
    dmes[mes1]-7*we/2
  }
  mesos=data[3]
  any=rep(data[1],mesos)
  mes=data[2]+0:(mesos-1)
  any=any+(mes-1)%/%12
  mes=(mes-1)%%12+1
  ts(apply(cbind(any,mes),1,valu),start=data[1:2],freq=12)
}
```

Referències

- [1] U.S. Energy Information Administration. *Total primary energy consumption*. <https://www.eia.gov/totalenergy/data/browser/index.php?tbl=T01.01#/?f=M&start=199001&end=201812&charted=4-6-7-14>.
- [2] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. & Ljung, G.M. 2016. *Time Series Analysis: Forecasting and Control (5th ed.)*. New Jersey: Wiley. ISBN 978-1-118-67502-1.
- [3] Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Uppsala: Almqvist & Wiksells Boktryckeri AB.
- [4] Makridakis, S., & Hibon, M. (1997b). *ARMA models and the Box-Jenkins methodology*. *Journal of Forecasting*, 16(3), 147–163. France: Fointeinebleau. [https://doi.org/10.1002/\(sici\)1099-131x\(199705\)16:3](https://doi.org/10.1002/(sici)1099-131x(199705)16:3).
- [5] Witt, A., Kurths, J., & Pikovsky, A. (1998). *Testing stationarity in time series*. *Physical Review E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*. Germany: Potsdam. <https://doi.org/10.1103/physreve.58.1800>.
- [6] Wikipedia contributors. (2024, April 1). *Subprime mortgage crisis*. Wikipedia. https://en.wikipedia.org/wiki/Subprime_mortgage_crisis.
- [7] Wikipedia contributors. (2024b, April 11). *Great recession*. Wikipedia. https://en.wikipedia.org/wiki/Great_Recession.
- [8] Wikipedia contributors. (2024, January). *Autoregressive integrated moving average*. Wikipedia. https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average.