

Grau en Estadística

Títol: Anàlisi i prediccions dels preus de l'electricitat mitjançant sèries temporals

Autor: Judit Domènech Fernández

Director: David Moríña Soler

Departament: Econometria, Estadística i economia aplicada

Convocatòria: Juny 2023



Agraïments

M'agradaria agrair la dedicació del tutor d'aquest treball, Dr. David Moríña Soler. Gràcies pel teu temps i el tracte. La teva visió i coneixements de la matèria han sigut de gran valor per aquest projecte.

També vull agrair el suport de la meva família que sap com ha sigut de complex aquest camí. En especial a tu Jaume, pel teu suport constant i per caminar aquest últim tram amb mi.

Gràcies.

Resum

Davant d'una situació de crisi climàtica i encariment dels recursos energètics, en especial els no renovables, es pren com a objectiu d'aquest treball generar prediccions dels preus de l'electricitat que ajudin a l'usuari a estimar quin impacte econòmic li pot ocasionar. Prenent la informació que posa a disposició la Unió Europea, com registres dels preus del gas i de diferents combustibles fòssils i informació recopilada sobre temperatures ambientals, es busca comprendre com totes les sèries s'interrelacionen i si poden confluïr per predir amb detall el cost de l'electricitat. Es farà ús de sèries temporals, utilitzant mètodes univariats i multivariats, fent una comparació dels resultats extrets i determinant quin permet capturar millor les fluctuacions del preu.

Paraules clau: recursos energètics, sèries temporals, sèries temporals multivariades, *forecasting*, *scraping*, suavitzat de dades, ARIMA, VAR

Classificació AMS: 37M10 - Time series analysis, 62M10 - Time series, auto-correlation, regression

Abstract

Facing a climate crisis and the increasing of energy resources costs (especially non-renewables), the aim of this project is to generate electricity price predictions. Using available data provided by the European Union, that includes records of gas and various fossil fuel prices, plus other collected data such as temperatures, the aim is to understand how all these series are interrelated and if they can converge to accurately predict future electricity prices. Time series analysis will be used, implementing both univariate and multivariate methods. It will be conducted a comparison of the extracted results, determining which method better captures price fluctuations

Key words: energy resources, time series, multivariate time series, forecasting, scraping, data smoothing, ARIMA, VAR

AMS classification: 37M10 - Time series analysis, 62M10 - Time series, auto-correlation, regression

Índex general

I. Introducció	1
1. Indicadors d'estudi	2
1.1. Preus de l'electricitat.....	2
1.2. Preus del gas.....	3
1.3. Preus dels combustibles fòssils	4
1. Temperatures	4
2. Extracció de les dades de recursos energètics	5
3. Extracció de les dades de temperatures	7
II. Metodologia	11
1. Plantejament d'un model ARIMA	12
2. Model autoregressiu	12
3. Model de mitjanes mòbils	13
4. Model multiplicatiu d'estacionalitat	14
5. Procés de modelització	15
6. Anàlisi multivariant	16
6.1. Model vectorial autoregressiu.....	16
6.2. Model vectorial autoregressiu de mitjanes mòbils	17
7. Suavitat de Kalman	17
III. Anàlisis de sèries temporals univariant	19
4. Preus de l'electricitat pel consumidor domèstic	19
1.1. Modelització univariant.....	23
1.2. Anàlisi dels residus	24
1.3. Predicció de valors futurs	25
5. Preus pel consumidor de la indústria	27
2.1. Modelització univariant.....	31
2.2. Anàlisi dels residus	32
2.3. Predicció de valors futurs	33
6. Preus del mercat majorista	35
3.1. Modelització univariant.....	40
3.2. Anàlisi dels residus	41
3.3. Predicció dels valors futurs.....	42
IV. Anàlisi de sèries temporals multivariant	44
1. Anàlisi descriptiu de la resta de variables	44
2. Selecció de la base mostral definitiva	46

3.	Modelització multivariant	49
4.	Causalitat de Granger	52
5.	Anàlisi dels residus	53
6.	Predicció de valors futurs	56
7.	Comparació amb grup control	59
V.	<i>Discussió del rendiment dels models dels preus d'electricitat</i>	64
VI.	<i>Conclusions</i>	66
VII.	<i>Bibliografia</i>	68
VIII.	<i>Annex A</i>	70
IX.	<i>Annex B</i>	75

Índex de figures

Figura I-1 Sèries de temperatures màximes, mínimes i mitjanes d'Alemanya	10
Figura III-1 Sèrie mensual dels preus domèstics d'electricitat.....	19
Figura III-2 Dades suavitzades de la sèrie temporal dels preus domèstics d'electricitat.....	20
Figura III-3 Diferencials de les dades suavitzades de la sèrie temporal dels preus domèstics d'electricitat.....	21
Figura III-4 Descomposició additiva de la sèrie temporal dels preus domèstics d'electricitat	22
Figura III-5 Autocorrelació dels preus domèstics d'electricitat.....	22
Figura III-6 Autocorrelació parcial dels preus domèstics d'electricitat.....	22
Figura III-7 Autocorrelació dels residus de la sèrie de preus domèstics d'electricitat amb el Model1.2.....	24
Figura III-8 Autocorrelació parcial dels residus de la sèrie de preus domèstics d'electricitat amb el Model1.2.....	24
Figura III-9 QQPLOT dels residus de la sèrie de preus domèstics d'electricitat del Model1.2.....	25
Figura III-10 Forecasts del model ARIMA(1,1,0)(0,1,1)[12] dels preus domèstics d'electricitat amb els intervals de confiança (95%) en blau.....	26
Figura III-11 Comparació amb test de prediccions dels preus domèstics d'electricitat.....	26
Figura III-12 Sèrie mensual dels preus per a la indústria d'electricitat	28
Figura III-13 Dades suavitzades de la sèrie temporal dels preus per a la indústria d'electricitat	29
Figura III-14 Diferencials de les dades suavitzades de la sèrie temporal dels preus per a la indústria d'electricitat.....	29
Figura III-15 Descomposició additiva de la sèrie temporal dels preus per a la indústria d'electricitat	30
Figura III-16 Autocorrelació parcial dels preus per a la indústria d'electricitat	30
Figura III-17 Autocorrelació dels preus per a la indústria d'electricitat.....	30
Figura III-20 Autocorrelació parcial dels residus dels preus per a la indústria d'electricitat amb el Model1.....	32
Figura III-21 Autocorrelació dels residus dels preus per a la indústria d'electricitat amb el Model1	32
Figura III-22 QQPLOT dels residus dels preus per a la indústria d'electricitat amb el Model1	32
Figura III-23 Forecasts del model ARIMA(1,1,0) dels preus d'electricitat de la indústria amb intervals de confiança (de fins al 95%) en blau.....	33
Figura III-24 Comparació amb test de les prediccions dels preus d'electricitat de la indústria...	34
Figura III-25 Sèrie mensual dels preus d'electricitat del mercat majorista.....	35
Figura III-26 Dades suavitzades de la sèrie temporal dels preus d'electricitat del mercat majorista	36
Figura III-27 Diferencials de les dades suavitzades de la sèrie temporal dels preus d'electricitat del mercat majorista.....	36
Figura III-29 Descomposició additiva de la sèrie de preus d'electricitat del mercat majorista ...	37

Figura III-30 Autocorrelació parcial dels preus d'electricitat del mercat majorista	37
Figura III-31 Autocorrelació dels preus d'electricitat del mercat majorista.....	37
Figura III-32 Descomposició additiva de la sèrie de diferències de preus d'electricitat del mercat majorista	38
Figura III-33 Autocorrelació parcial de diferències dels preus d'electricitat del mercat majorista	39
Figura III-34 Autocorrelació de diferències dels preus d'electricitat del mercat majorista	39
Figura III-35 Autocorrelació parcial dels residus dels preus d'electricitat del mercat majorista amb el Model1.....	41
Figura III-36 Autocorrelació dels residus dels preus d'electricitat del mercat majorista amb el Model1.....	41
Figura III-37 QQPLOT dels residus dels preus d'electricitat del mercat majorista amb el Model1	41
Figura III-38 Forecasts del model ARIMA(0,1,1) dels preus d'electricitat del mercat majorista amb intervals de confiança (de fins el 95%) en blau.....	42
Figura III-39 Comparació amb test de les prediccions dels preus d'electricitat del mercat majorista	43
Figura IV-1 Sèries temporals dels preus del gas, combustibles fòssils i temperatura	44
Figura IV-2 Correlograma del conjunt de sèries temporals	46
Figura IV-3 Conjunt de sèries temporals suavitzades	48
Figura IV-4 Conjunt de sèries temporals.....	48
Figura IV-5 Conjunt de diferencials de les sèries temporals.....	49
Figura IV-6 Conjunt de diferencials de les sèries temporals suavitzades	49
Figura IV-7 Valors AIC dels models VARMA(p,q).....	51
Figura IV-8 Sèrie de valors AIC en funció del número de retards al model VAR.....	52
Figura IV-9 Sèrie dels residus del model VAR(12)	53
Figura IV-10 Autocorrelacions dels residus del model VAR(12).....	54
Figura IV-11 Autocorrelacions parcials dels residus del model VAR(12)	55
Figura IV-12 Representació de les prediccions en blau amb intervals de confiança (del 95%) en vermell	57
Figura IV-13 Representació de les prediccions contra el grup control de preus d'electricitat pel consumidor domèstic	59
Figura IV-14 Representació de les prediccions contra el grup control de preus d'electricitat per a la indústria	60
Figura IV-15 Representació de les prediccions contra el grup control de preus d'electricitat del mercat majorista.....	62

Índex de taules

Taula III-1 Estadístics de la sèrie temporal dels preus domèstics d'electricitat.....	20
Taula III-2 Resultats de models de la part regular de la sèrie temporal de preus d'electricitat domèstics.....	23
Taula III-3 Resultats de models de la part estacional de la sèrie temporal de preus d'electricitat domèstics.....	23
Taula III-4 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat pel consumidor domèstic	27
Taula III-5 Estadístics de la sèrie temporal dels preus per a la indústria d'electricitat	28
Taula III-6 Resultats de models de la part regular de la sèrie temporal de preus per a la indústria d'electricitat.....	31
Taula III-7 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat per a la indústria	35
Taula III-8 Estadístics de la sèrie temporal dels preus d'electricitat del mercat majorista.....	36
Taula III-9 Resultats de models de la part regular de la sèrie temporal de preus d'electricitat del mercat majorista.....	40
Taula III-10 Resultats de models de la part estacional de la sèrie temporal de preus d'electricitat del mercat majorista.....	40
Taula III-11 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat pel mercat majorista.....	43
Taula IV-1 Estadístics de sèries temporals del gas, combustibles fòssils i temperatura.....	45
Taula IV-2 Models univariants de les sèries temporals de tots els recursos energètics	50
Taula IV-3 Comparació de test, prediccions i intervals de confiança dels preus domèstics d'electricitat amb model multivariant	59
Taula IV-4 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat per a la indústria amb model multivariant.....	61
Taula IV-5 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat majorista amb model mutlivariant	62

Capítol 1

I. INTRODUCCIÓ

El preu de l'electricitat és (o es considera) un factor crucial a l'economia de qualsevol país, afectant de forma directa al consumidor final. En els últims anys, s'ha observat una tendència a l'alça en els preus de l'electricitat, la qual ha estat àmpliament comentada en els mitjans de comunicació.

En el panorama de crisi energètica en la qual es troba el món actualment, factors externs com també els preus a l'alça dels combustibles fòssils i del gas, actuen conjuntament en l'encariment del cost de la vida dels ciutadans.

Davant aquesta problemàtica, es pren com a objecte d'estudi un conjunt de dades sobre recursos energètics d'**Alemanya**¹ proporcionades per la Unió Europea. Aquestes dades contenen informació històrica sobre les principals fonts energètiques utilitzades als països de la Unió Europea i els seus preus i es presenta com els seus preus evolucionen en el temps.

Els objectius d'aquest treball s'enfocaran en els següents punts:

- En primer lloc, s'analitzarà l'evolució dels preus de l'electricitat, el gas natural i els combustibles d'origen fòssil. La finalitat és poder comprendre millor l'impacte que aquests recursos energètics tenen en l'economia i, sobretot, a la vida quotidiana dels consumidors.
- En segon lloc, es pretén comprendre com aquests recursos energètics estan relacionats entre si. S'investigarà si existeix una correlació entre els preus de l'electricitat, el gas natural i els combustibles d'origen fòssil, i com aquesta correlació pot afectar el preu final que els consumidors han de pagar.

¹ Es seleccionen dades d'Alemanys degut a l'històric de la mostra i que s'ha pogut trobar altra informació adicional sobre el país que complementarà el model de series temporals, com les temperatures.

- Projecció d'aquests preus mitjançant sèries temporals i poder obtenir prediccions a futur que permeti al consumidor comprendre millor l'impacte que pot tenir en la seva economia.
- Finalment, i amb l'objectiu de poder avaluar les eines estadístiques utilitzades, es farà una comparació de totes les metodologies emprades, posant èmfasis a la utilització de mètodes multivariants.

Concretament, l'anàlisi i les prediccions dels preus giraran principalment entorn dels preus de l'electricitat, on la comprensió dels factors claus que afecten aquests preus serà elemental.

1. Indicadors d'estudi

1.1. *Preus de l'electricitat*

La fixació del preu de l'electricitat a la comunitat europea depèn de diversos factors. Principalment, es determinen l'oferta i la demanda en què es trobi el mercat en aquell moment. També influeix el cost de les matèries primeres usades per la generació de l'energia elèctrica. Les polítiques que els governs de cada país apliquin en matèria d'impostos o subvencions també afecta el preu final de l'electricitat.

Cal destacar que a part d'aquests factors que conflueixen a l'hora de fixar el preu del mercat, l'electricitat no es pot emmagatzemar, cosa que obliga que el mercat majorista generalment hagi de negociar l'electricitat a curt termini.

El conjunt de tot això provoca una alta volatilitat dels preus de l'electricitat en un *pool* energètic on el preu varia diàriament.

En el mercat de l'electricitat hi ha tres tipus de preus depenent de qui és el comprador:

- 1- El preu que s'estableix al mercat majorista. Aquest preu és el que les distribuïdores assumeixen per comprar el recurs directament als generadors. Aquest preu és variable i es determina diàriament.

- 2- El preu dirigit a petites empreses i a llar, el qual pot ser regulat mitjançant polítiques governamentals i depèn de la tarifa contractada, en conseqüència els preus són més estables que els anteriors.
- 3- El preu dirigit a grans indústries. Són preus que estan pactats entre la companyia energètica i el client i en conseqüència està més subjecte a negociacions. Són preus més baixos que el dels consumidors de la llar i petites empreses, ja que el consum de l'electricitat és més elevat en aquests casos.

Aquests tres preus seran objecte d'estudi i seran processats perquè, malgrat al mercat es recullin diàriament, la sèrie tingui una periodicitat mensual. Els preus seran agregats mitjançant mitjanes.

1.2. Preus del gas

A Alemanya hi ha un sistema parcialment regulat del mercat del gas. És a dir, els preus a l'engròs estan marcats essencialment per l'oferta i la demanda dels proveïdors i els compradors mentre que, de cara a la comercialització del recurs energètic a consumidors residencials i petites empreses, les tarifes dels preus poden estar regulades per l'agència Federal anomenada *Bundesnetzagentur*.

Aquesta agència és una autoritat reguladora independent que supervisa el sector energètic del país, la competència del mercat i les telecomunicacions.

La regularització governamental dels preus i el fet que el comerç del gas estigui influenciat per una economia global, on són diversos països del món que exporten i participen en aquest intercanvi, fa que la variabilitat dels preus sigui menys voluble en el temps, el que afectarà la sèrie temporal.

Igual que en els preus de l'electricitat, pel gas es recull el preu de compra del gas com a matèria primera al mercat majorista, el preu dirigit a petites empreses i consumidors de la llar i el preu de les grans indústries.

1.3. Preus dels combustibles fòssils

El mercat dels combustibles fòssils també està regit principalment a les lleis de l'oferta i la demanda, però també hi ha altres factors que cal tenir en compte pel que fa a la fixació dels seus preus. Sobretot en el mercat dels combustibles d'origen fòssil, influeixen significativament les polítiques governamentals, com impostos sobre els combustibles o regulacions mediambientals que s'apliquen amb l'objectiu de fomentar la transició cap a fonts d'energia renovable. Això fa que els preus dels combustibles fòssils puguin ser més volàtils i subjectes a canvis bruscs en funció de les polítiques dels diferents governs.

Els preus d'estudi recollits dels combustibles fòssils són els del consumidor i s'estudia els següents tipus de combustibles: combustible líquid, *diesel*, petroli i el petroli sense processar (o altrament dit "cru").

El petroli sense processar té diversos indicadors de preu que es prenen com a referència a l'hora de comercialitzar la matèria primera, els quals són incorporats també a la sèrie temporal:

- *Brent*: fa referència al petroli produït a Europa, Àfrica i Orient Mig.
- *West Texas Intermediate (WTI)*: preu que s'estableix del petroli produït a els Estats Units.
- *Dubai*: preu de referència del petroli produït a Orient Mig i Àsia.

1. Temperatures

Amb motiu de poder enriquir la projecció de les dades al futur i poder fer prediccions emprant models de sèries temporals multivariants, s'ha inclòs en la mostra d'estudi les temperatures mitjanes de diferents ciutats d'Alemanya.

Les dades són recollides de la web meteorològica *Wetterzentrale*, web que recopila informació de nombroses estacions meteorològiques del país i posa a disposició diferents *forecasts* d'indicadors meteorològics.

2. Extracció de les dades de recursos energètics

La informació dels costos de l'electricitat, combustibles fòssils i el gas que posa a disposició la web de les Comissions Europees. La comissió està organitzada per diferents departaments, cada un enfocant a certes problemàtiques socials, legislatives o propostes polítiques. Un d'ells és el departament d'energia, que recurrentment posa a disposició dades sobre la taxació de preus del sector energètic.

La forma d'obtenir aquestes dades és a partir del següent *Dashboard*:

https://energy.ec.europa.eu/data-and-analysis/energy-prices-and-costs-europe/dashboard-energy-prices-eu-and-main-trading-partners_en

Aquest, primerament deixa escollir quin tipus de recurs energètic es vol observar:

How to use the dashboard

First, select a dimension. Then, the available indicators will appear. On the right hand side a menu will appear, which allows you to download, print or change data visualisation.



Dimension

- Electricity prices
- Natural gas prices
- Petroleum product prices

Indicator

Il·lustració I-1 Recursos energètics disponibles

Seguidament, dona l'opció de seleccionar el preu dependent del comprador:

Dimension

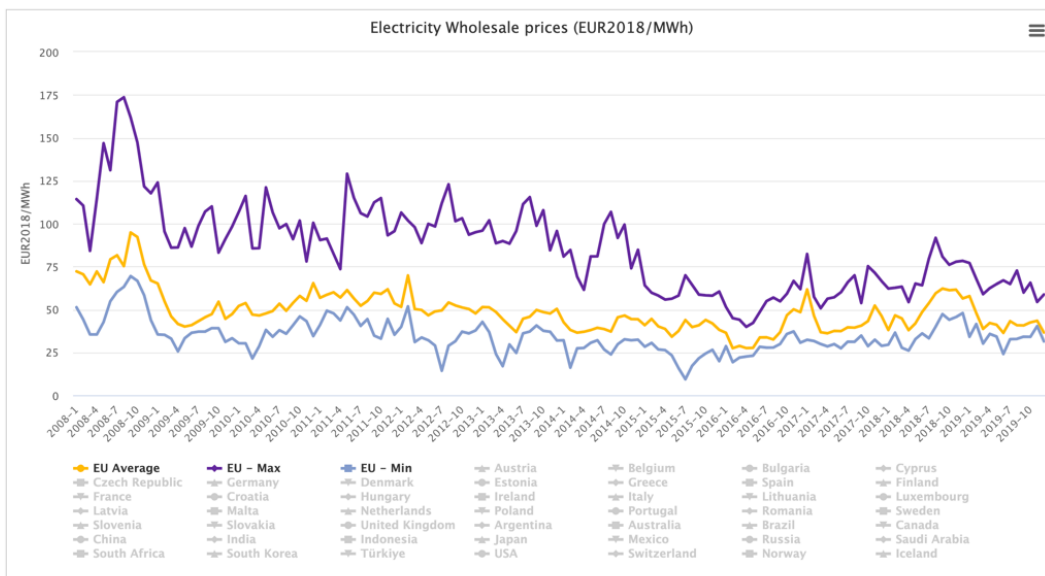
- Electricity prices
- Natural gas prices
- Petroleum product prices

Indicator

- Wholesale prices
- Retail prices – Households
- Retail prices – Industry

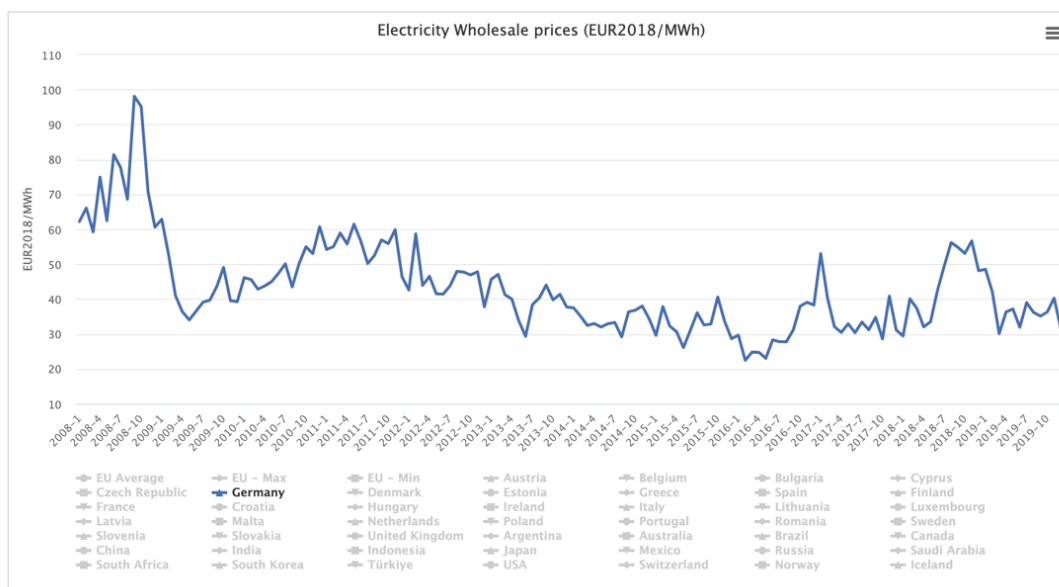
Il·lustració I-2 Tipus de preus disponibles

Les dades comprenen informació del gener de 2008 fins a l'octubre de 2019, amb una periodicitat mensual



Il·lustració I-3 Preus de compra d'electricitat de la Unió Europea

De forma interactiva, es poden escollir quins països visualitzar. En el cas d'aquest treball es selecciona Alemanya.



Il·lustració I-4 Preus de compra d'electricitat a Alemanya

3. Extracció de les dades de temperatures

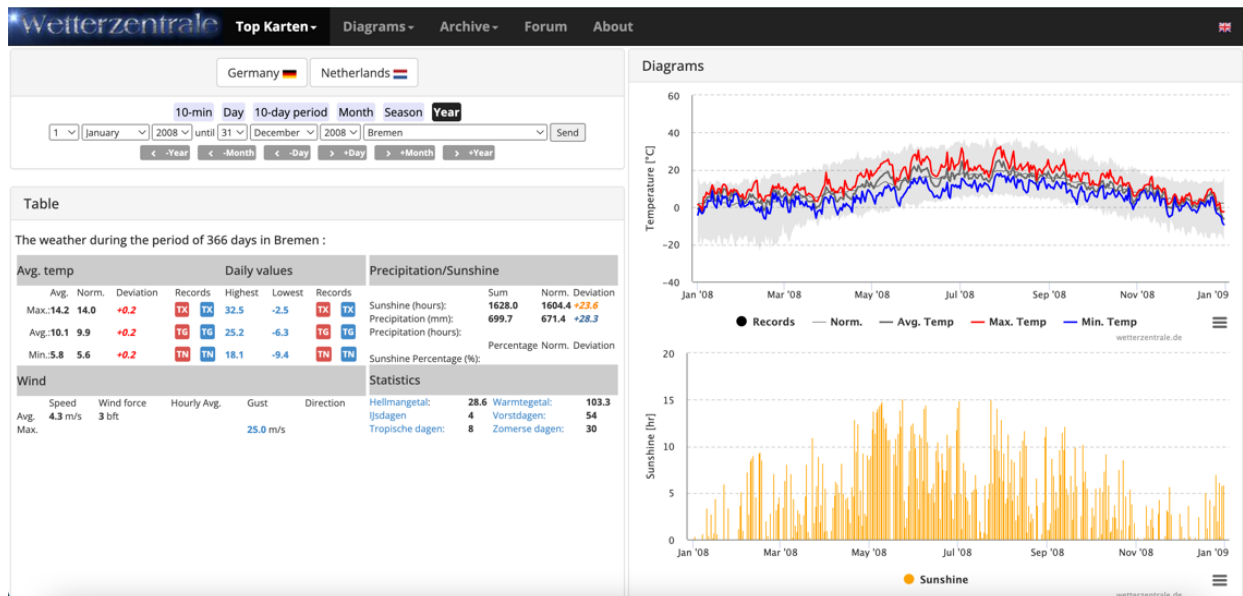
Per poder implementar les metodologies de sèries temporals plantejades, és necessari un set de temperatures d'Alemanya agrupades mensualment i amb un històric que compregui totes les observacions del gener de 2008 fins a l'octubre de 2019. Sovint trobar dades que compleixin certs requisits dificulta la cerca de la mostra que s'estudiarà. En aquest cas s'ha hagut d'emprar tècniques d'*scraping* de la web *wetterzentrale* per poder recuperar les dades que presenten visualment amb el format correcte.

L'*scraping* és una tècnica d'obtenció d'informació d'una pàgina web de forma automatitzada i sistemàtica mitjançant un *script*. Sovint aquesta pràctica pot no ser lícita si es recopila la informació sense consentiment, comproment així els termes de servei del lloc *web*, o fent cas omís dels drets d'autor.

En aquest cas s'ha demanat el permís de desenvolupadors de la *wetterzentrale* i un cop amb el seu consentiment s'ha procedit a dur a terme l'*scraping* de les dades de temperatures diàries del 2008 al 2019.

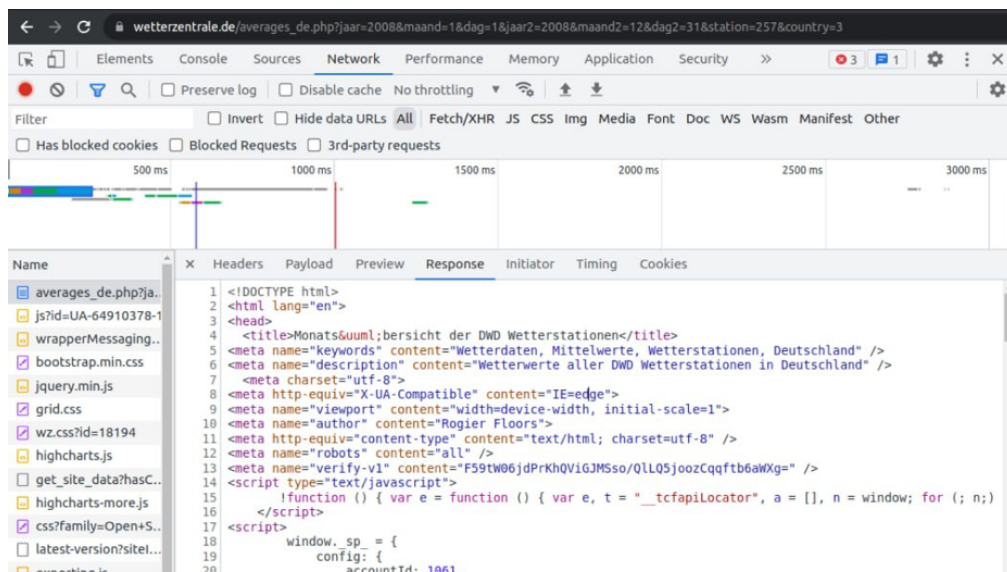
Per fer-ho primer s'ha hagut d'entendre la composició de la pàgina *web*:

- Aquesta consta d'un tauler dinàmic per poder filtrar la informació per país (donant com a opció Alemanya o Països Baixos) i per any, setmana o mes que es vol observar.
- En un lateral esquerre hi ha una visualització d'estadístics com el màxim, el mínim o la mitjana de les temperatures observades a la zona i el període seleccionat, així com estadístics que fan referència a les precipitacions, velocitat del vent o altres.
- I finalment, al lateral dret hi ha una representació gràfica de les temperatures mínimes, màximes i mitjanes i un diagrama de barres de mesura de les hores de sol diàries de tot el període.



Il·lustració I-5 Pàgina web de la font de dades meteorològiques

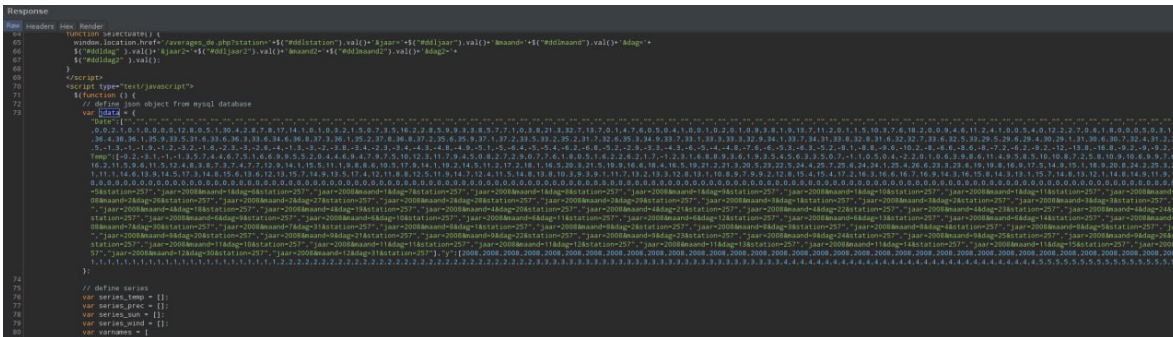
Tenint en compte la visualització dels diferents paràmetres que componen la pàgina web, es localitza la informació d'estudi. En aquest cas es voldrà informació sobre temperatures únicament d'Alemanya, amb un període que contemplici de 2008 fins a 2019 i l'objecte de dades que es recollirà són les observacions de temperatures que conformen el gràfic situat a la part superior del lateral dret. Aquesta informació es troba en el document *HTML* de la *web*, així com tota l'estructura i contingut d'ella.



Il·lustració I-6 Visualització dels primers paràmetres del document html

Sabent que la informació d'interès es troba dins d'aquest document, es busca poder fer una petició *HTTP*² per poder interceptar el codi *HTML* en la visualització de la pàgina web on albergui les temperatures d'Alemanya dels anys d'interès.

Es fa una primera aproximació d'una sol·licitud de la informació de només l'any 2008 i es localitza en el document *HTML* les dades de temperatures. Aquestes estan preparats per ser emmagatzemats en un fitxer *JSON*.



Il·lustració I-7 Resposta de la petició HTTP per les dades de l'any 2008

Un cop localitzat la informació que es requereix, es fa una *script* que itera per tots els anys que es necessiten. En cada iteració el procediment és el següent:

1. Es guarda tot el document *HTML*.
2. Es processa la informació concreta on es troba el que serà el fitxer *JSON*, però que encara està en una cadena de caràcters.
3. La cadena de caràcters es passa a format *JSON*.
4. Es recopilen tots els anys i s'insereixen al mateix fitxer, per posteriorment ser processat i analitzat per R.

² El missatge *Hypertext Transfer Protocol* s'envia pel client a un servidor web sol·licitant una acció específica o un recurs.

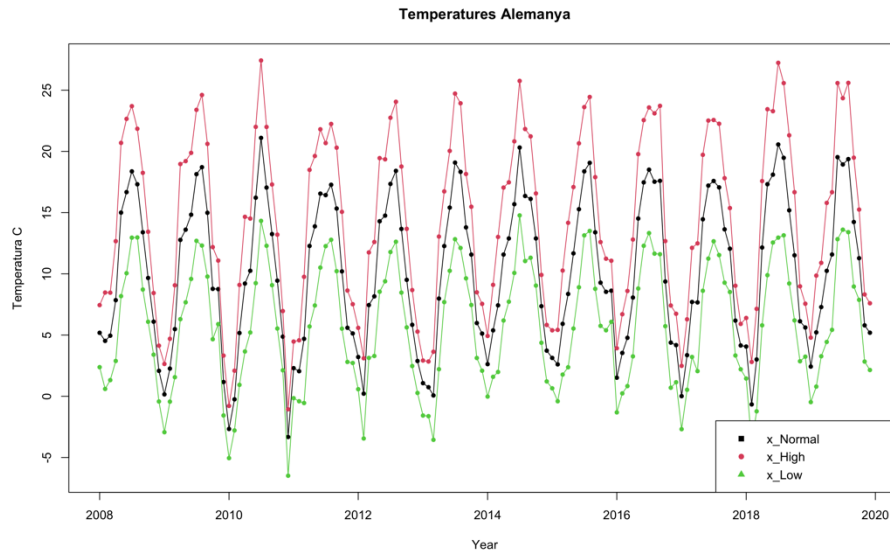


Figura I-1 Sèries de temperatures màximes, mínimes i mitjanes d'Alemanya

Un cop visualitzades les sèries temporals que s'han aconseguit, es pren la decisió d'aplicar una d'elles a l'anàlisi sobre els recursos energètics. En aquest cas es selecciona la variable x_High , que representa les temperatures màximes mensuals registrades.

Capítol 2

II. METODOLOGIA

Una sèrie temporal és una successió de dades que segueixen un ordre cronològic. És a dir, es considera com una seqüència de punts que es denoten com a $x_1, x_2, x_3, \dots, x_T$, sent T el període de temps complet que la variable X contempla.

L'anàlisi de sèries temporals permet entendre el comportament d'aquests punts i la correlació d'una variable usant la seva pròpia història per, posteriorment, obtenir prediccions adequades mitjançant models univariants.

Quins avantatges pot tenir implementar un model de sèries temporals en comptes d'un model estocàstic determinista?

En una mostra de dades com les que s'han seleccionat per aquest projecte, s'ha d'assumir que poden haver-hi altres factors desconeguts que afectin el comportament de la mostra. La mostra, sent una recopilació de mesures mensuals, inevitablement seran dependents de factors com el temps. Malgrat que això compliciti el càlcul de valors futurs, sí que es pot calcular amb una determinada probabilitat aquests valors entre uns intervals de confiança.

En aquest projecte es plantejarà un model univariant de sèries temporals per cada una de les variables dels preus de. En últim lloc, es plantejarà un model multivariant mitjançant tot un conjunt de sèries temporals d'altres recursos energètics (de gas i diferents combustibles fòssils) juntament amb les temperatures enregistrades, amb l'objectiu de poder fer una comparació dels resultats i poder fer una discussió sobre quines prediccions són més ajustades a la realitat.

Amb l'objectiu de poder fer el *forecasting* de les dades recopilades i tractades, es farà un anàlisi de sèries temporals i s'aplicaran els models adients.

1. Plantejament d'un model ARIMA

És important que a l'hora de plantejar un model de sèrie temporal, el conjunt de dades, es comportin de forma estacionària. Altrament dit, que la seva mitjana i la seva variància sigui constant en el temps.

Malgrat que això ens asseguri que les prediccions que aporti el model molt probablement no estaran esbiaixades per factors probabilístics intrínsecs a les dades desconeguts per nosaltres, sovint qualsevol conjunt de dades que conformin una sèrie temporal no tindrà una mitjana constant, de la mateixa manera que pot presentar heteroscedasticitat.

En aquests casos és quan aplicar un model Autoregressiu Integrat de Mitjanes Mòbils pot salvar aquestes limitacions. Aquest s'expressa com a $ARIMA(p, d, q)$.

El model ARIMA es constitueix per 3 elements:

- La primera, que representa la part **AR**. Contempla l'autoregressió de la sèrie.
- La segona, que representa la part **I**. És la integradora que indica quants cops s'ha de transformar la sèrie amb l'objectiu d'aconseguir estacionarietat de les dades. Es pot fer per mitjà de diferències de cada valor amb el seu valor anterior.

$$x_t = x_t - x_{t-d}$$

- La tercera, que representa la part **MA**. Fa referència a les mitjanes mòbils, que passants

2. Model autoregressiu

El model $AR(p)$ assumeix que existeix certa dependència lineal entre valors històrics dins la sèrie. Aquest es defineix amb la següent formulació:

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + w_t$$

On x_t és estacionari, per tant, es parteix de la base que es compleixen les condicions:

$$E(x_t) = \mu; E[(x_t - \mu)^2] = \sigma_x^2; Cov(x_t, x_{t-k}) = \gamma(k), t \neq k$$

$\phi_1, \phi_2, \dots, \phi_p$ són constants ($\phi_i \neq 0$) i w_t són residus distribuïts com a $w_t \sim i.i.d. N(0, \sigma^2)$ denominat també com a soroll blanc on es compleix:

$$E(w_t) = 0; E(w_t^2) = \sigma^2; Cov(w_t, w_{t-k}) = 0, t \neq k$$

En el cas d'implementar un model autoregressiu de primer ordre $AR(1)$ es representaria com a

$$x_t = \phi x_{t-1} + w_t$$

A tenir en compte que quan $\phi = 1$ el model es tracta d'un camí aleatori.

3. Model de mitjanes mòbils

En el cas d'un model de mitjanes mòbils $MA(q)$ es pren com a premissa que el soroll blanc ε_t pot ser utilitzat combinat linealment pot explicar un valor de la sèrie. És a dir:

$$x_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i} + w_t$$

On partim que x_t és estacionari, $\theta_1, \theta_2, \dots, \theta_q$ són els paràmetres del model i w_t és soroll blanc.

En molts casos, combinant ambdós models es pot aconseguir unes millors prediccions perquè es basarà fent servir una estructura estocàstica més flexible i complexa. És així com s'empraria un model $ARMA(p, q)$, combinant els termes autoregressius p i els termes de mitjanes mòbils q que s'adeqüi a la mostra.

La combinació dels dos models quedaria definida així:

$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + w_t$$

4. Model multiplicatiu d'estacionalitat

El model multiplicatiu de mitjanes mòbils integrada autoregressiva estacional, altrament conegut com a SARIMA, es un model que també contempla el component $ARIMA(p, d, q)$ i que afegeix el component multiplicatiu dels efectes estacionals. La seva denotació és $ARIMA(p, d, q) \times (P, D, Q)_s$, que es defineix com a:

$$\Phi_P(B^s)\phi(B)\nabla_s^D\nabla^d x_t = \Theta_Q(B^s)\theta(B)w_t$$

Les diferents parts de la formula s'expressen de la següent forma:

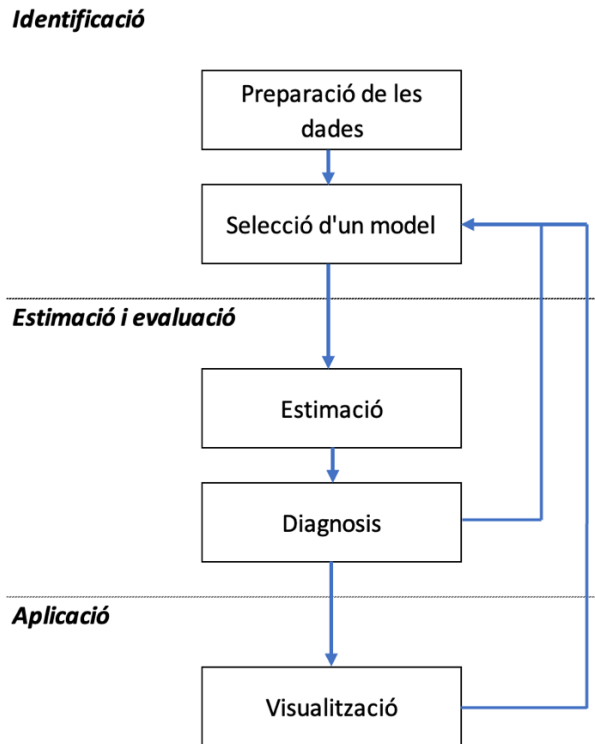
$$\begin{aligned}\phi(B) &= 1 - \sum_{i=1}^p \phi_i B^i \\ \Phi_P(B^s) &= 1 - \sum_{i=1}^p \Phi_i B^{is} \\ \theta(B) &= 1 + \sum_{i=1}^q \theta_i B^i \\ \Theta_Q(B^s) &= 1 + \sum_{i=1}^q \Theta_i B^{is}\end{aligned}$$

I els components de diferències estacionals:

$$\begin{aligned}\nabla^d &= (1 - B)^d \\ \nabla_s^D &= (1 - B^s)^D\end{aligned}$$

5. Procés de modelització

Per dur a terme el procés de selecció i aplicació d'un model per a cada una de les variables d'estudi, s'utilitzarà l'esquema lògic de la metodologia *Box-Jenkins*³.



Il·lustració II-1 Diagrama de metodologia Box-Jenkins

Aquesta metodologia es centra en ser un procés iteratiu amb 3 fases principals ben definides:

- En una primera fase **d'identificació**, es preparen les dades per poder elaborar un anàlisi de sèries temporals, en aquest pas es fa el *preprocessing* que sigui adequat a més d'un anàlisi descriptiu exhaustiu de les dades. Això permetrà proposar un primer model apte.
- Un cop definit el model es pot fer una **estimació** dels paràmetres i una **avaluació**. Es passa a modelitzar i estudiar els resultats. En la diagnòs cal estudiar els errors del model, s'espera que aquests siguin estacionaris i que es puguin interpretar com soroll blanc.

³ George Box i Gwilym Jenkins van ser dos matemàtics que en el 1970 van proposar la metodologia Box-Jenkins. Durant les seves carreres es van centrar en la modelització de sèries temporals. Actualment les seves aportacions són considerades punt neuràlgic en matèria d'anàlisi de sèries temporals.

També en aquest punt es pot analitzar les prediccions i fer una comparació amb el grup control seleccionat.

Aquest primer anàlisi permet decidir si el model suggerit sembla ser adient, en cas contrari cal modificar el model de partida i fer els canvis que semblin pertinents.

- Finalment, es porta a cap una sèrie de visualitzacions gràfiques, que ajudin a comprendre el comportament de les dades i dels resultats. Amb el criteri de l'interpret, en aquest també es pot tornar al punt de selecció d'un model, i si es creu necessari implementar un altre diferent.

6. Anàlisi multivariant

6.1. Model vectorial autoregressiu

Per últim, s'implementarà un model multivariant al *set* de dades d'aquest projecte. Finalment, es podrà fer una comparació dels models utilitzats i extreure una conclusió dels seus rendiments i resultats.

Per poder implementar un model multivariant que tingui en compte diverses sèries temporals ahora s'emprarà un model de Vector Autoregressiu, $VAR(p)$.

El model $VAR(p)$ interpreta el conjunt de sèries temporals com a variables endògenes. És a dir, la variable del *forecast* no només és explicada per les variables explicatives de la mostra sinó que aquest fet també passa la inversa, on les variables explicatives també són explicades per la variable per predir. Tot mitjançant un sistema d'equacions lineals, una per cada variable.

Posant com a exemple un model de vector autoregressiu d'ordre 1 amb dues variables, el sistema d'equacions seria el següent:

$$\begin{aligned}y_{1,t} &= c_1 + \phi_{11,1}y_{1,t-1} + \phi_{12,1}y_{2,t-1} + w_{1,t} \\y_{2,t} &= c_2 + \phi_{21,1}y_{1,t-1} + \phi_{22,1}y_{2,t-1} + w_{2,t}\end{aligned}$$

Aquesta modelització no només permet predir valors futurs sinó que també, amb els paràmetres ϕ_{12} , ϕ_{21} es podrà comprovar la relació que tenen, en aquest cas, la sèrie y_1 i y_2 .

En el cas d'aquest cas d'estudi, l'objectiu d'aquest projecte és poder utilitzar l'anàlisi multivariant de sèries temporals per investigar com els diferents preus dels recursos energètics seleccionats s'influeixen mútuament i determinen, més concretament, al preu de l'electricitat.

6.2. Model vectorial autoregressiu de mitjanes mòbils

Si el model $VAR(p)$, es construeix sobre la base del nombre màxim de retards per fer prediccions sobre la regressió de cada sèrie que contempla; el model $VARMA(p, q)$ té en compte aquesta primera part regressora especificada en el model $VAR(p)$, afegint una estructura basada en mitjanes mòbils $MA(q)$.

Així doncs, en un conjunt de sèries temporals localitzades en format de matriu on cada vector que la conforma es denota com a y_t , sent aquesta estacionària; l'equació del model és la següent:

$$y_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + w_t + \Theta_1 w_{t-1} + \dots + \Theta_q w_{t-q}$$

7. Suavitzat de Kalman

El suavitzat de *Kalman* és un mètode emprat per estimar i suavitzar estats ocults en un model d'espai d'estats⁴. Aquest mètode és especialment usat en anàlisis de sèries temporals doncs permet obtenir estimacions de valors passats més precisos sense presència d'excés de soroll.

El mètode del suavitzat de *Kalman* es basa en el filtre de *Kalman*, que és un procés recursiu fet servir per estimar l'estat actual d'un sistema en funció de les observacions passades i les relacions dinàmiques del sistema.

Com a mètode recursiu que és, si s'inicialitza amb els paràmetres $x_0^0 = \mu_0$ i $P_0^0 = \Sigma_0$, per a $t = 1, \dots, n$; el sistema d'equacions del model d'estats en el que es basa el filtratge de *Kalman* és el següent:

⁴ Un model d'estats es una representació matemàtica en la que es relacionen variables d'entrada, sortida i estats mitjançant un sistema d'equacions.

$$x_t^{t-1} = \phi x_{t-1}^{t-1} + \Upsilon u_t$$

$$P_t^{t-1} = \phi P_{t-1}^{t-1} \phi' + Q$$

$$x_t^t = x_t^{t-1} + K_t(y_t - A_t x_t^{t-1} - A_t x_t^{t-1})$$

$$P_t^t = [I - K_t A_t] P_t^{t-1}$$

On cal especificar que:

$$K_t = P_t^{t-1} A_t' [A_t P_t^{t-1} A_t' + R]^{-1}$$

$$w_t = y_t - E(y_t | y_{1:t-1}) = y_t - A_t x_t^{t-1} - A_t x_t^{t-1}$$

$$\Sigma_0 := \text{var}(w_t) = \text{var}[A_t(x_t - x_t^{t-1}) + v_t] = A_t P_t^{t-1} A_t' + R$$

Un cop obtinguts els estimadors filtrats, es poden obtenir els estimador x_t^n , on $t = 1, \dots, n$. Aquests seran en denominats valors suavitzats de la sèrie. La seva construcció depèn de les equacions:

$$x_{t-1}^n = x_{t-1}^{t-1} + J_{t-1}(x_t^n - x_t^{t-1})$$

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_n^{t-1})J_{t-1}'$$

On:

$$J_{t-1} = P_{t-1}^{t-1} \phi' - [P_t^{t-1}]^{-1}$$

Capítol 3

III. ANÀLISIS DE SÈRIES TEMPORALS UNIVARIANT

En aquest apartat es farà una anàlisi descriptiu dels principals conceptes entorn les sèries temporals per a cada una dels indicadors dels preus de l'electricitat recollits. L'objectiu és comprendre l'estructura i el comportament de les sèries a través de mesures i gràfics, com la mitjana, la variància, la tendència, l'estacionalitat i l'autocorrelació.

4. Preus de l'electricitat pel consumidor domèstic

Pel que fa a els preus dirigits al consumidor del recurs energètic com a ús domèstic, s'observa una tendència clarament alcista fins a l'any 2013, on finalment s'estabilitza i disminueix la seva variabilitat.

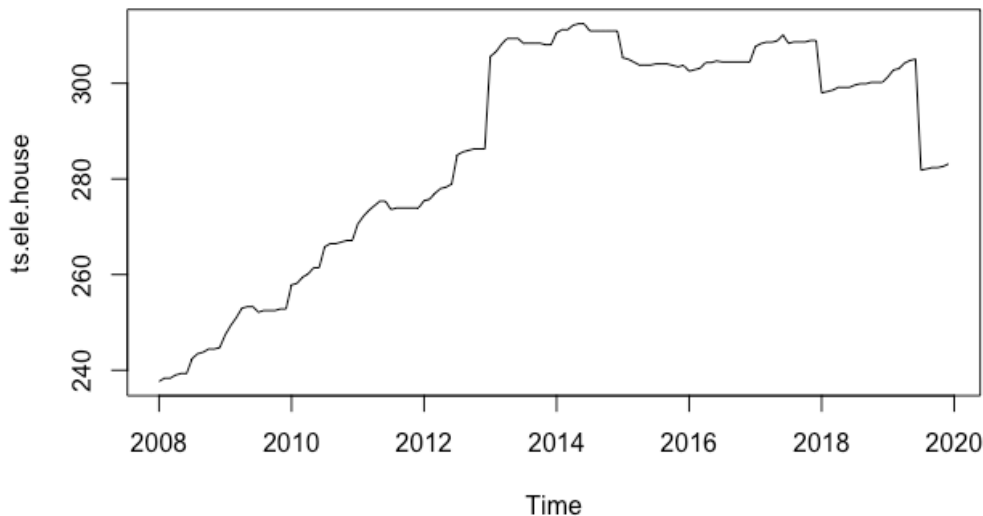


Figura III-1 Sèrie mensual dels preus domèstics d'electricitat

Aquest fenomen provoca un cert biaix de la mitjana del preu cap a la cua de la dreta, com es nota a la **figura III-1**. També es nota com la mitjana del preu domèstic passa a ser de 286,7€, influenciat pels valors pròxims al 300€ de a partir de l'any 2013; també reflectit en una mediana del preu (299,4€) major a la mitjana.

Mínim	1r quartil	Mediana	Mitjana	3r quartil	Màxim
237.7€	269.7€	299.4€	286.7€	305.4€	312.4€

Taula III-1 Estadístics de la sèrie temporal dels preus domèstics d'electricitat

Caldrà transformar les dades per tal de poder parametritzar la sèrie temporal. Es proposa per tal de poder tractar l'heteroscedasticitat suavitzar les dades amb el mètode Kalman.

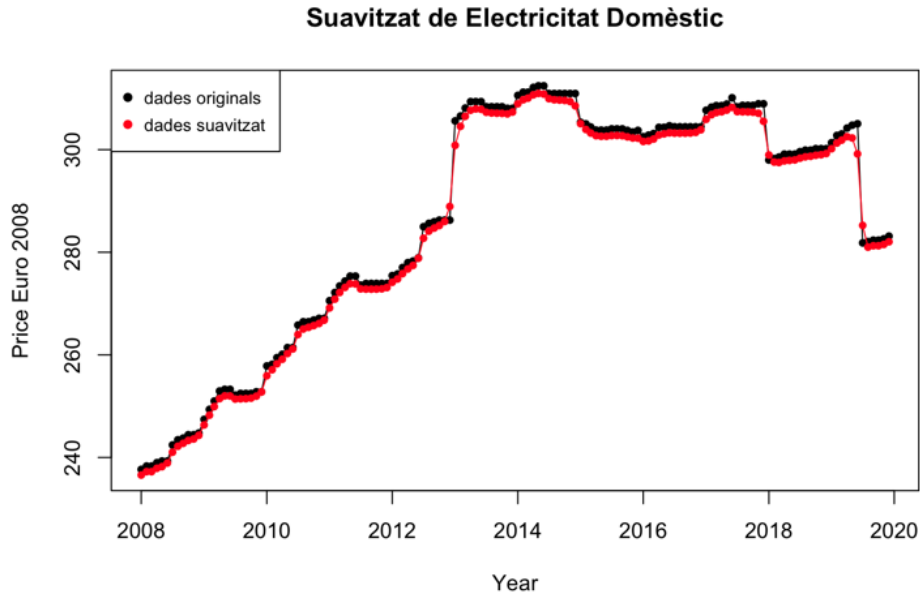


Figura III-2 Dades suavitzades de la sèrie temporal dels preus domèstics d'electricitat

Amb el suavitzat se salva lleugerament salts pronunciats dels preus de l'electricitat domèstica al llarg del temps i a la sèrie transformada amb una diferenciació de les dades, disminueix la presència de possibles *outliers*.

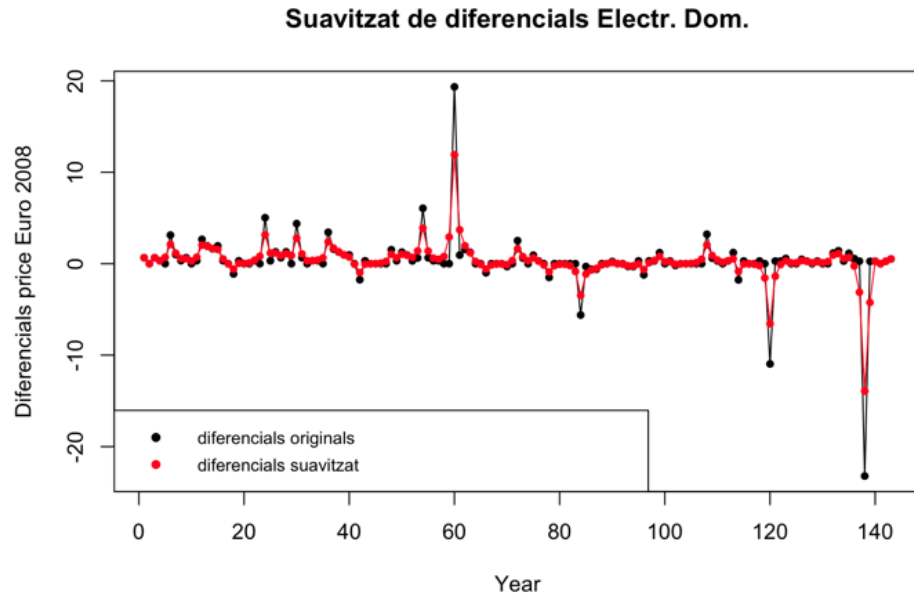


Figura III-3 Diferencials de les dades suavitzades de la sèrie temporal dels preus domèstics d'electricitat

Visualment, s'aprecia com en termes d'estacionarietat la sèrie necessitarà ser tractada mitjançant diferències per ser parametrizada. Amb una prova de Dickey-Fuller s'acaba de determinar amb un 95% de confiança que la sèrie temporal no presenta evidència de tenir una arrel unitària i que, per tant, la suposició és correcta.

A continuació es realitza un anàlisi exploratori de la sèrie del preu de l'electricitat dirigit al consumidor domèstic. En aquest punt es guarda el tram final (comprèn els 12 mesos de l'any 2019) de la sèrie temporal com a grup control per poder, posteriorment fer una comparació dels valors obtinguts del *forecast*.

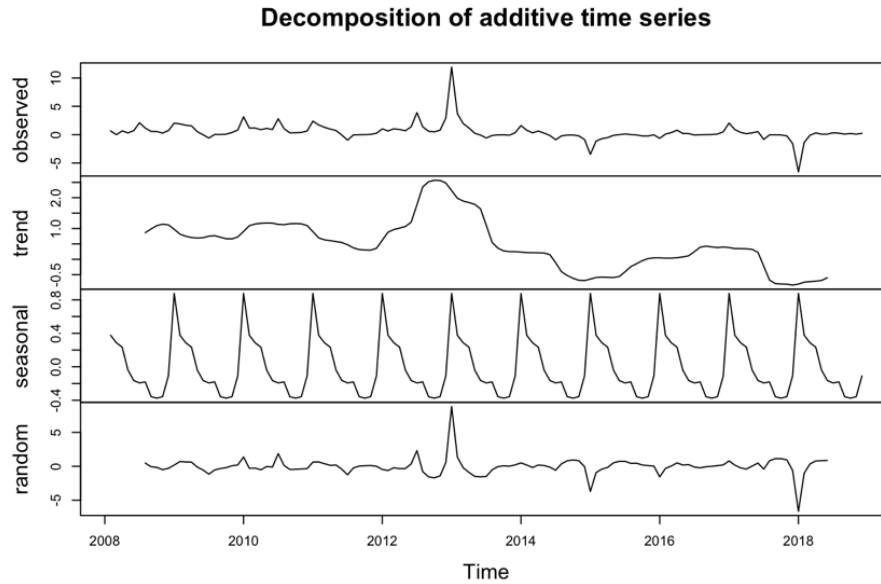


Figura III-4 Descomposició additiva de la sèrie temporal dels preus domèstics d'electricitat

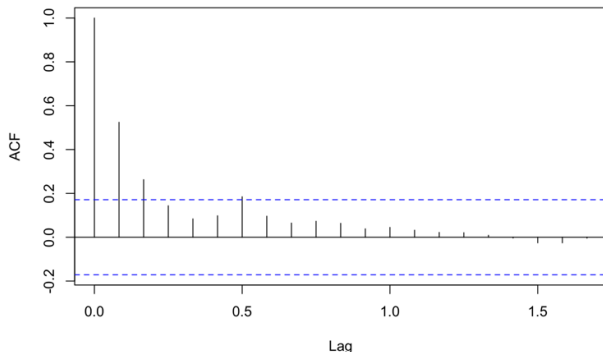


Figura III-5 Autocorrelació dels preus domèstics d'electricitat

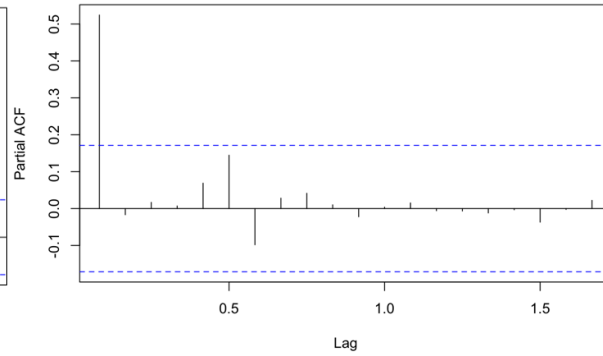


Figura III-6 Autocorrelació parcial dels preus domèstics d'electricitat

A la representació gràfica de l'autocorrelació, es pot observar com la significança dels coeficients d'autocorrelació a mesura que augmenta el nombre de retards va disminuint. El primer coeficient és clarament significatiu, el que indica que hi ha una relació entre una observació de la sèrie i la seva anterior. Fenomen que, menys pronunciadament, però també significatiu, passa amb un segon retard. Per altra banda, a la gràfica d'autocorrelació parcial, únicament s'observa un valor significatiu.

1.1. Modelització univariant

Es farà una estimació dels paràmetres òptims que pot prendre la sèrie temporal per tal que s'ajusti bé i que tingui un poder predictiu acurat. Es triarà la sèrie temporal idònia amb el criteri AIC. Així doncs, es començarà fent una estimació dels coeficients per la part regular de la sèrie en funció de l'observat a les gràfiques d'autocorrelació i autocorrelació:

Model	AIC	Estimació dels coeficients	$P(> z)$
Model1: ARI(1,1)	446.4922	ar1 = 0.563508	ar1 = 3.076e-15 ***
Model2: ARIMA(1,1,1)	448.4717	ar1 = 0.580000 ma1 = -0.024331	ar1 = 1.22e-05 *** ma1 = 0.885

Taula III-2 Resultats de models de la part regular de la sèrie temporal de preus d'electricitat domèstics

El model emprat òptim entre els 2 proposats és un AR(1). S'observa com l'ARMA(1,1), no només té un valor AIC major, sinó que el coeficient de la mitjana mòbil no resulta ser significatiu.

Un cop identificada la part regular de la sèrie, es contempla la parametrització per tractar una possible estacionalitat de la sèrie. Si es torna a la **figura III-5** de l'autocorrelació s'observa un patró d'augment puntual al retard número 6, que està al llindar de significació. Com aquest patró no sembla ser molt pronunciat i que cada valor de la sèrie no sembla tenir molta memòria estacional, les propostes de parametrització que es duen a terme a continuació seran: un model SARIMA(1,1,0)(1,1,0)[12] i un model SARIMA(1,1,0)(0,1,1)[12]. No s'aplicarà un model que combini un regressor i un regressor parcial estacional a la vegada.

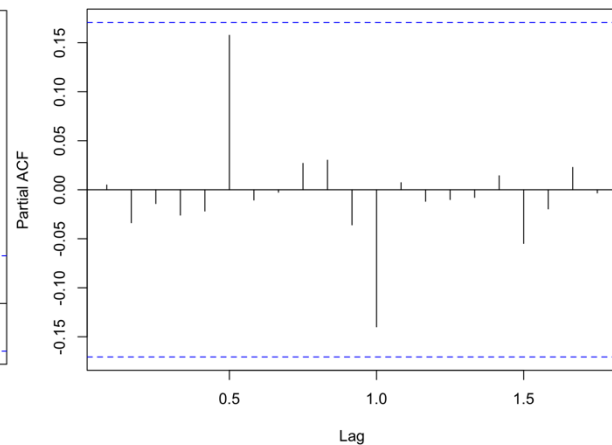
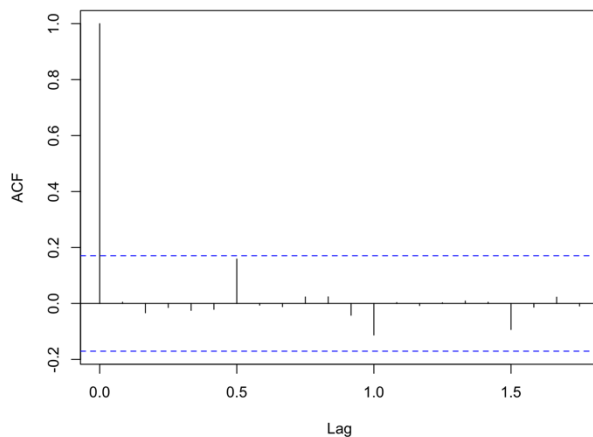
Model	AIC	Estimació dels coeficients	$P(> z)$
Model1.1: SARIMA(1,1,0)(1,1,0)[12]	458.3291	ar1 = 0.508064 sar1 = -0.443688	ar1 = 8.934e-11 *** sar1 = 1.734e-07 ***
Model1.2: SARIMA(1,1,0)(0,1,1)[12]	435.8136	ar1 = 0.580000 ma1 = -0.024331	ar1 = 4.480e-11 *** sma1 = .841e-15 ***

Taula III-3 Resultats de models de la part estacional de la sèrie temporal de preus d'electricitat domèstics

Malgrat ambdues sèries amb components estacionals tinguin uns coeficients estimats significatius, basant-nos en el criteri d'AIC s'escull el SARIMA(1,1,0)(0,1,1)[12]. Aquest ha disminuït respecte al valor AIC que s'obtenia amb un model ARIMA on no es té en compte cap component estacional.

1.2. Anàlisi dels residus

Un cop escollit el model de sèrie temporal que determina el preu de l'electricitat dirigit al consumidor domèstic, es fa un anàlisi dels residus. Es busca que la distribució residual sigui estacionària i que no presenti cap ordre de dependència de les seves observacions. Per poder-ho avaluar, doncs, es farà una representació de l'autocorrelació i d'autocorrelació parcial de la sèrie residual. Es busca que aquesta sèrie sigui el més semblant a soroll blanc possible i que, per tant, les observacions estiguin distribuïdes amb homoscedasticitat.



No s'observa cap nombre de retards que demostrï que hi ha autocorrelació ni autocorrelació dels residus de la sèrie temporal. Per acabar de determinar si es tracta de soroll blanc es veurà si es distribueix de forma normal.

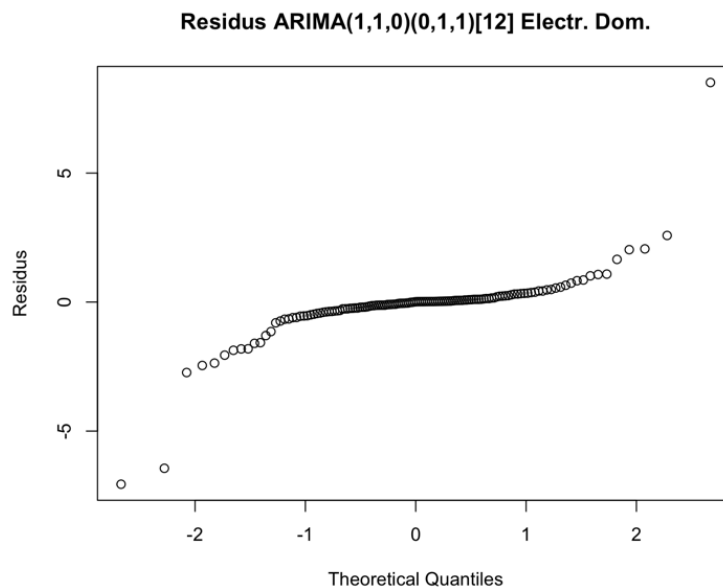


Figura III-9 QQPLOT dels residus de la sèrie de preus domèstics d'electricitat del Model1.2

S'observa com amb el *qqplot* hi ha un gran pes de les observacions que es distribueixen de forma normal, però com les cues són arrossegades en diferents direccions a causa de valors atípics.

Aquests valors, es sospita basant-nos en la distribució dels preus en forma d'altiplans de la **figura III-1**, són deguts a regulacions externes als preus marcats de l'electricitat. Malgrat que amb el suavitzat previ s'ha aconseguit que els valors atípics s'acostin a la resta d'observacions, fent que aquests no superin el rang del valor interquartílic que defineix quan un valor es *outlier*. S'observa com de totes formes afecta els resultats de la sèrie temporal implementada.

1.3. Predicció de valors futurs

Un cop es determina que els paràmetres que millor s'adapten a la sèrie temporal dels preus de l'electricitat per a consumidors domèstics, es fa el *forecasting* per a les dades de l'any 2019. Així mateix, les dades reals de 2019 han sigut reservades per ser comparat amb els resultats de la predicció.

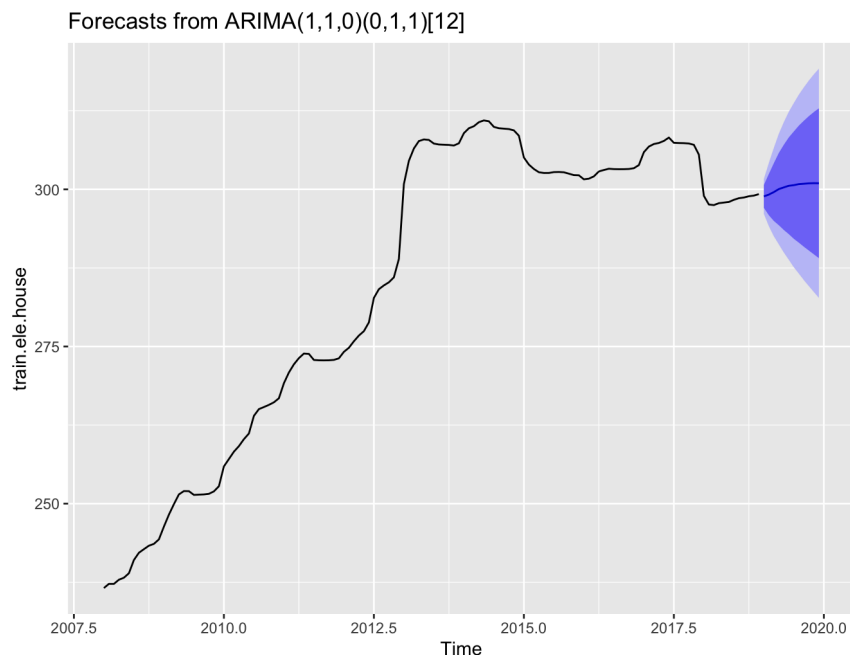


Figura III-10 Forecasts del model $ARIMA(1,1,0)(0,1,1)[12]$ dels preus domèstics d'electricitat amb els intervals de confiança (95%) en blau

A continuació la representació gràfica es focalitza en les observacions de l'any 2019 i es compara amb el grup de test.

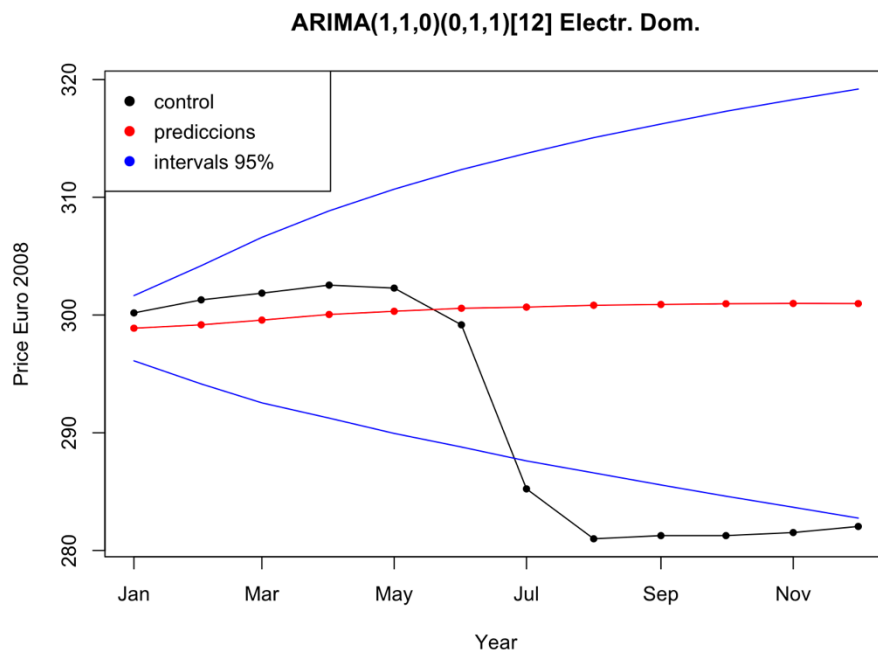


Figura III-11 Comparació amb test de prediccions dels preus domèstics d'electricitat

A la representació gràfica anterior es pot observar com les prediccions del primer semestre de 2019 s'assemblen a les observacions reals. A partir de juliol de 2019, les dades reals dels preus van caure sobtadament, és a partir d'aquí on les prediccions deixen d'acostar-se als valors reals.

TEST	PREDICT	Lower95%	Upper95%
300.1782	298.8790	296.1111	301.6468
301.2868	299.1683	294.1473	304.1893
301.8557	299.5689	292.5393	306.5985
302.5402	300.0460	291.2452	308.8468
302.2856	300.3170	289.9451	310.6888
299.1675	300.5716	288.7906	312.3526
285.2353	300.6703	287.6096	313.7310
280.9941	300.8297	286.5936	315.0658
281.2641	300.8940	285.5674	316.2206
281.2641	300.9586	284.6118	317.3054
281.5241	300.9829	283.6747	318.2910
282.0441	300.9713	282.7519	319.1906

Taula III-4 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat pel consumidor domèstic

A la taula anterior es marquen amb color verd aquells intervals que ha aportat el model que amb una confiança del 95% que tenen en compte el valor real observat. En taronja es marquen aquelles observacions on l'interval de confiança establert pel model no aconsegueix capturar la dada real. S'observa fàcilment com, malgrat que els primers valors predits són prou ajustats a la realitat, els últims valors no són adequats per descriure els preus de l'electricitat del consumidor domèstic.

5. Preus pel consumidor de la indústria

Pel que fa al preu de l'electricitat, però ara dirigit a les indústries, s'identifiquen patrons similars a la sèrie temporal anterior. Com s'observa a la **figura III-12**, els preus al llarg del temps es distribueixen de forma alcista, de forma que es pot veure com aquesta sèrie no es pot considerar que sigui estacionària. De forma recurrent en la finestra temporal que comprèn la sèrie, s'arriben

a pics dels preus que cada vegada són majors, segurament degut a factors externs com regulacions dels preus que desconeixem.

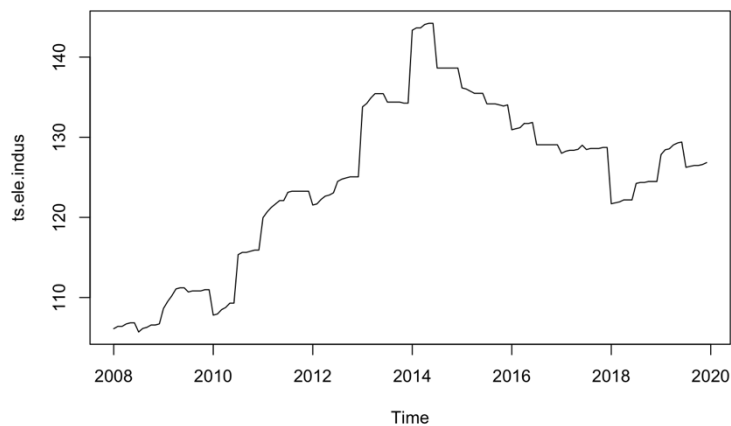


Figura III-12 Sèrie mensual dels preus per a la indústria d'electricitat

A diferència que els preus d'electricitat que anaven dirigits al consumidor domèstic, els preus de la indústria, després d'arribar en un preu màxim a principis de l'any 2014, va disminuir fins que el 2018 va canviar la tendència que es portava observant per entrar en una estabilització dels preus al voltant dels 125€.

Mínim	1r quartil	Mediana	Mitjana	3r quartil	Màxim
105.7€	119.0€	125.7€	124.6€	132.3€	144.2€

Taula III-5 Estadístics de la sèrie temporal dels preus per a la indústria d'electricitat

Visualment, ja es pot comprovar que serà necessari aplicar una part integradora a l'hora de parametritzar un model de sèrie temporal. Variacions brusques dels preus en poc temps de transcurs, farà que a l'hora de tractar amb la sèrie transformada mitjançant diferències, ens trobem amb valors atípics. Fent el mateix tractament de suavitzat de les dades pot facilitar que aquests canvis pronunciats quedin mitigats i que sigui més fàcil trobar el model idoni de sèrie temporal que generi soroll blanc com a residu.

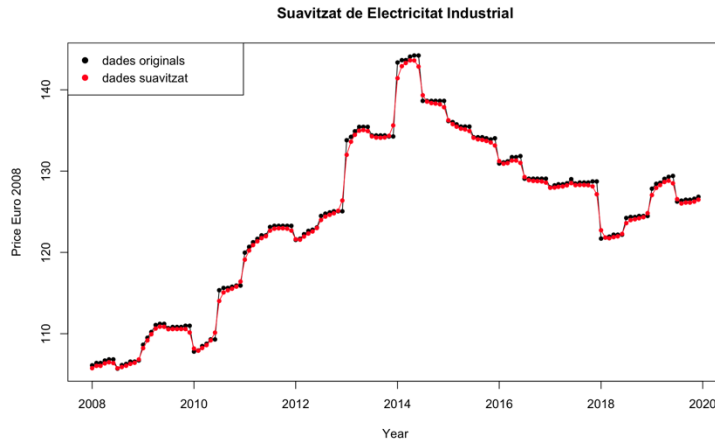


Figura III-13 Dades suavitzades de la sèrie temporal dels preus per a la indústria d'electricitat

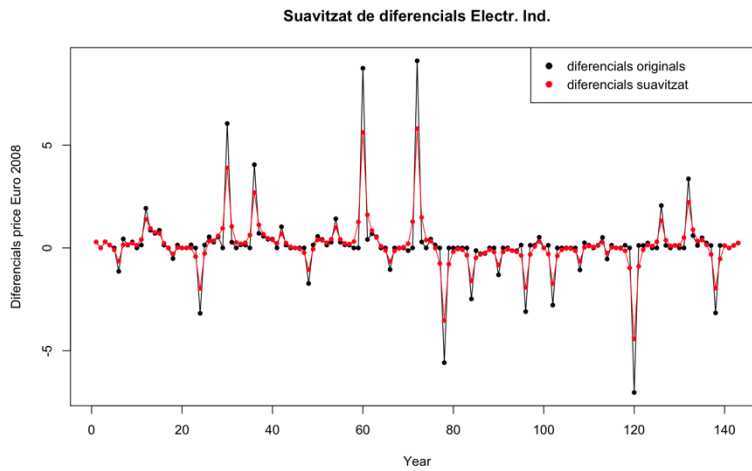


Figura III-14 Diferencials de les dades suavitzades de la sèrie temporal dels preus per a la indústria d'electricitat

Aplicant una prova de *Dickey-Fuller* a la sèrie suavitzada, es comprova les suposicions que s'havien fet prèviament amb el suport gràfic. No es pot considerar que la sèrie tingui arrel

unitària, per tant, s'assumeix que la sèrie no és estacionària. S'ha de tenir en compte que serà necessari emprar un ordre d'integració de la sèrie mitjançant diferenciació.

A continuació es farà una selecció dels valors *train/test* i es fa l'anàlisi dels diferents components de la sèrie temporal suavitzada i amb un ordre d'integració.

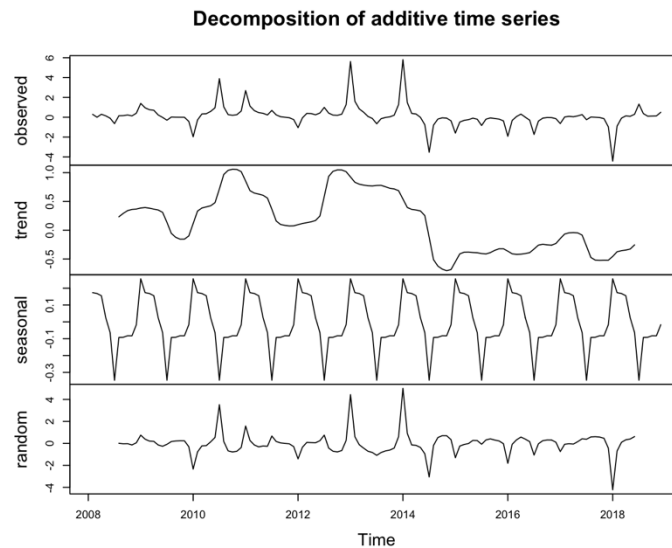


Figura III-15 Descomposició additiva de la sèrie temporal dels preus per a la indústria d'electricitat

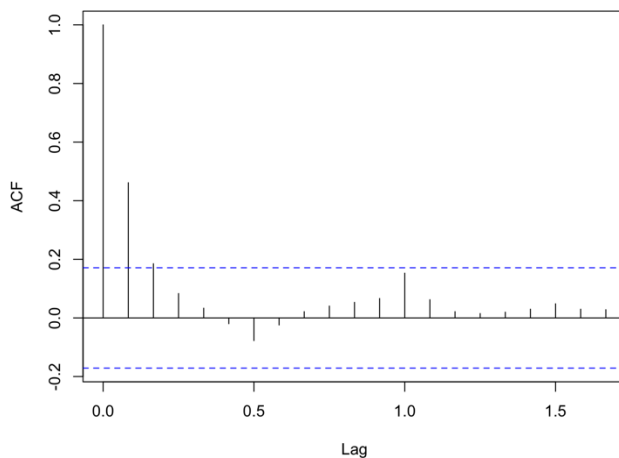


Figura III-17 Autocorrelació dels preus per a la indústria d'electricitat

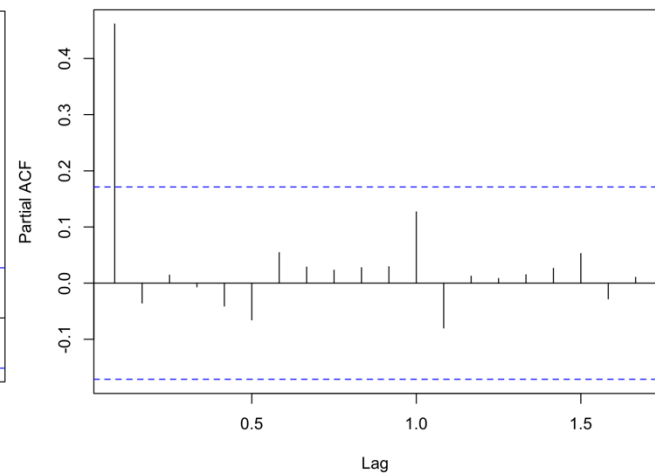


Figura III-16 Autocorrelació parcial dels preus per a la indústria d'electricitat

En el gràfic on es veu representada la descomposició additiva de la sèrie, es reflecteix al gràfic de tendència com la sèrie a partir del 2015 la seva tendència canvia disminuint, respecte als pics que s'observaven el 2011 o 2013. Pel que fa a l'estacionalitat, els pics positius que es donen recurrentment per a cada any, se situen aproximadament a principis de l'any.

Pel que l'autocorrelació, s'observa com el primer retard és clarament significatiu. A partir d'aquest la presència d'altres retards va disminuint successivament, mentre que a partir del retard número 6 es fa un repunt no significatiu. Veient el gràfic d'autocorrelació parcial, destaca un únic retard parcial.

2.1. Modelització univariant

Tenint en compte aquestes consideracions, es planteja un model AR(1) per a la sèrie diferenciada, que serà comparat amb un ARMA(1,1) el qual, amb les conclusions extretes de l'anàlisi d'autocorrelacions probablement no serà un model òptim.

Model	AIC	Estimació dels coeficients	$P(> z)$
Model1: ARI(1,1)	373.0786	ar1 = 0.467290	ar1 = 1.108e-09 ***
Model2: ARIMA(1,1,1)	374.9571	ar1 = 0.41489 ma1 = 0.0672	ar1 = 1.22e-05 *** ma1 = 0.72606

Taula III-6 Resultats de models de la part regular de la sèrie temporal de preus per a la indústria d'electricitat

Veiem com el model ARIMA(1,1,1) no només fa augmentar el seu AIC, si no que l'estimació de la mitjana mòbil d'ordre 1 no resulta ser significativa.

Partint que la sèrie que millor representa els preus d'electricitat per a la indústria és un ARIMA(1,1,0), es valora si cal parametritzar regressors o mitjanes mòbils estacionals. Basant-nos en la poca presència de retards que podrien venir donats per estacionalitat, és possible que parametritzar la part estacional de la sèrie no aporti més capacitat predictiva al model. Es procedeix a fer l'anàlisi dels residus per a ARIMA(1,1,0).

2.2. Anàlisi dels residus

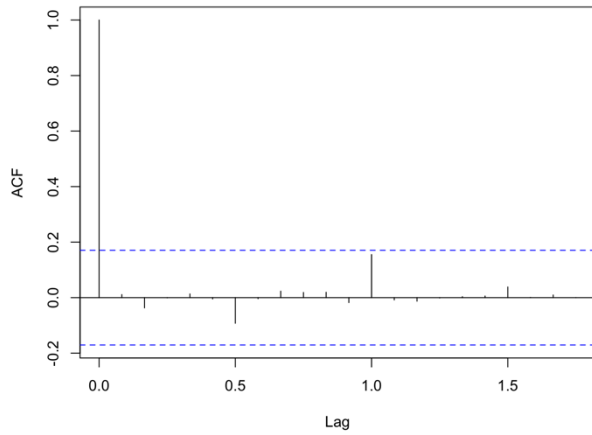


Figura III-19 Autocorrelació dels residus dels preus per a la indústria d'electricitat amb el Model1

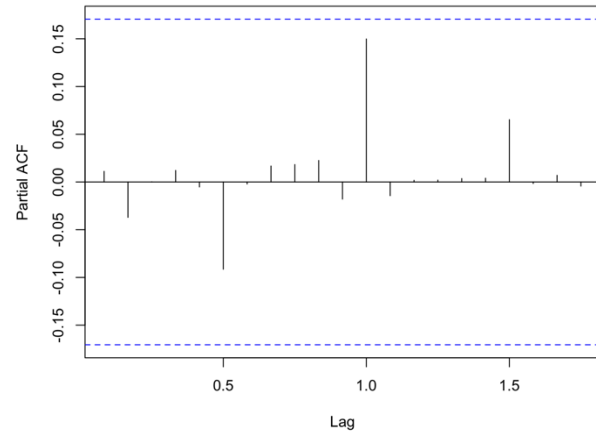


Figura III-18 Autocorrelació parcial dels residus dels preus per a la indústria d'electricitat amb el Model1

S'observa a les gràfiques d'autocorrelació i autocorrelació parcial dels residus, com a aquests es mostren com soroll blanc. També es realitza una prova de *Dickey-Fuller* on s'assumeix que la sèrie de residus no conté arrels unitàries i que, per tant, es pot considerar com a estacionària. A continuació es comprova gràficament si els valors de la sèrie residual es pot distribuir de forma normal.

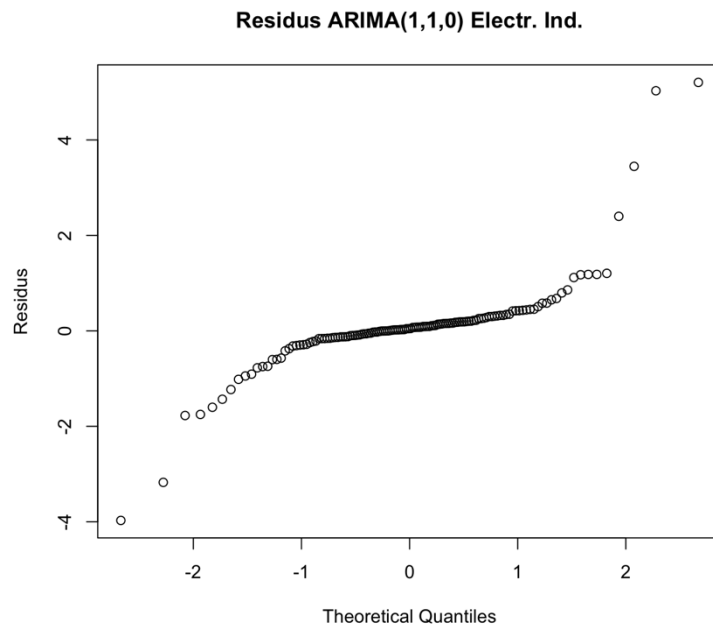


Figura III-20 QQPLOT dels residus dels preus per a la indústria d'electricitat amb el Model1

De la mateixa forma que passava amb la sèrie de preus d'electricitat dirigits al consumidor domèstic, els que van dirigits a la indústria també presenta uns residus amb cues que escapen de la normalitat.

2.3. Predicció de valors futurs

Havent escollit el model univariant adient d'acord amb l'anàlisi exploratori dels preus d'electricitat per a la indústria i havent fet l'anàlisi pertinent dels residus, es confirma que es farà servir un model ARIMA(1,1,0).

A continuació es fa una representació dels valors del *forecasting*.

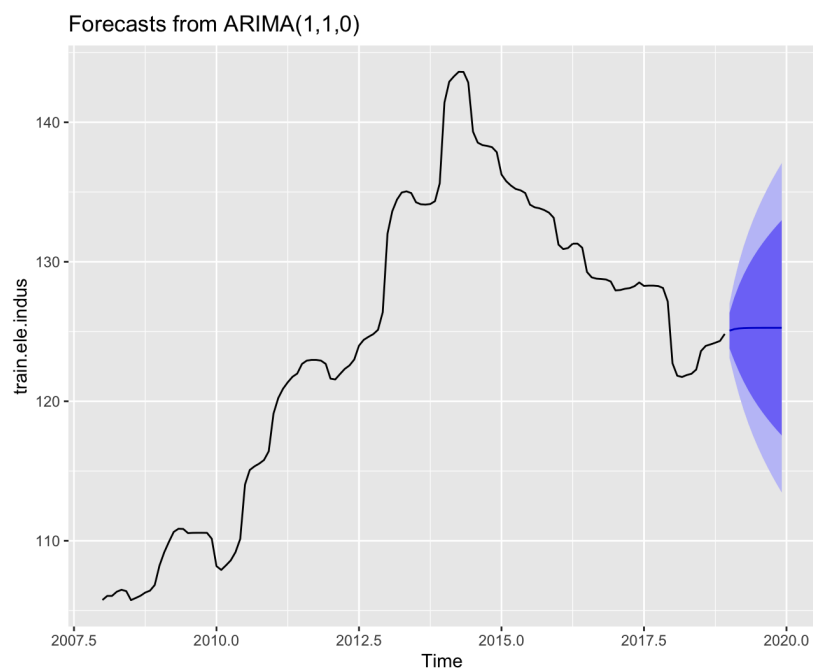


Figura III-21 Forecasts del model ARIMA(1,1,0) dels preus d'electricitat de la indústria amb intervals de confiança (de fins al 95%) en blau

S'observa com els valors predits són molt similars, debut a la poca memòria autoregressiva que té el model amb els paràmetres especificats.

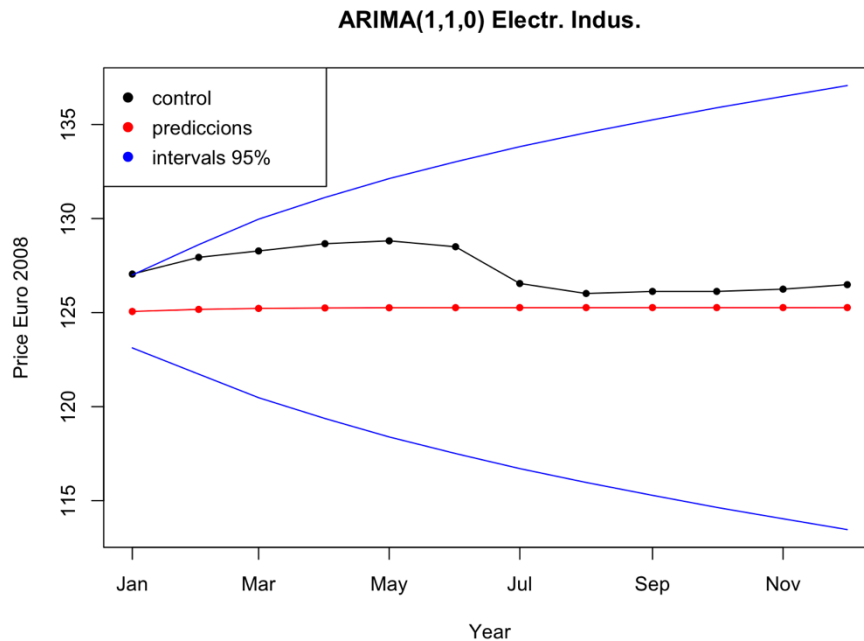


Figura III-22 Comparació amb test de les prediccions dels preus d'electricitat de la indústria

Valors predits tan estacionaris i amb tan poca variabilitat fa que s'allunyin fàcilment de les dades reals si aquestes estan sotmeses a canvis bruscos. Com és el cas del primer semestre del 2019, on les dades reals estan molt per sobre de les prediccions. És a partir del segon semestre, que els preus de l'electricitat baixen i s'apropen a les dades previstes pel model.

TEST	PREDICT	Lower95%	Upper95%
127.0493	125.0620	123.1239	127.0002
127.9381	125.1714	121.7299	128.6129
128.2781	125.2225	120.4772	129.9678
128.6638	125.2463	119.3707	131.1220
128.8165	125.2575	118.3889	132.1260
128.5022	125.2627	117.5077	133.0177
126.5492	125.2651	116.7066	133.8237

126.0163	125.2663	115.9697	134.5629
126.1263	125.2668	115.2850	135.2486
126.1263	125.2671	114.6435	135.8906
126.2463	125.2672	114.0381	136.4962
126.4863	125.2672	113.4636	137.0709

Taula III-7 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat per a la indústria

A la taula on es recullen els valors reals, els valors predits i els intervals de confiança del 95%, es veu com en gairebé tots els mesos, els intervals contempnen la dada real del preu.

6. Preus del mercat majorista

Pel que fa a l'últim indicador de preus de l'electricitat a analitzar, la sèrie dels preus dirigits al mercat majorista té un comportament que es diferencia dels ja analitzats. El preu majorista és més volàtil al llarg del temps i, exceptuant un comportament irregular al principi de la sèrie (del 2008 al 2009 aproximadament) es podria arribar a considerar estacionari.

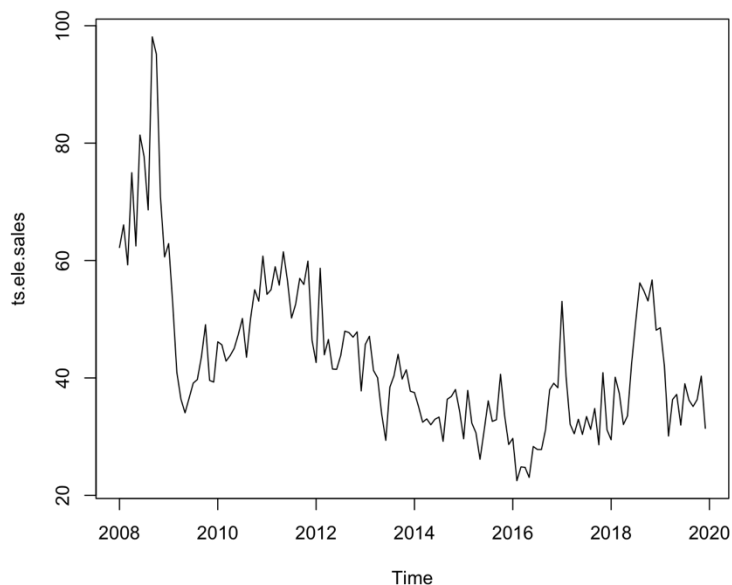


Figura III-23 Sèrie mensual dels preus d'electricitat del mercat majorista

Mínim	1r quartil	Mediana	Mitjana	3r quartil	Màxim
22.49€	33.41€	40.16€	43.20€	49.69€	98.10€

Taula III-8 Estadístics de la sèrie temporal dels preus d'electricitat del mercat majorista

Amb l'objectiu de poder mitigar sobretot el valor màxim (98.10€) en aquesta sèrie també s'aplicarà el mètode de suavitzat de *Kalman*.

Suavitzat de Electricitat Majorista

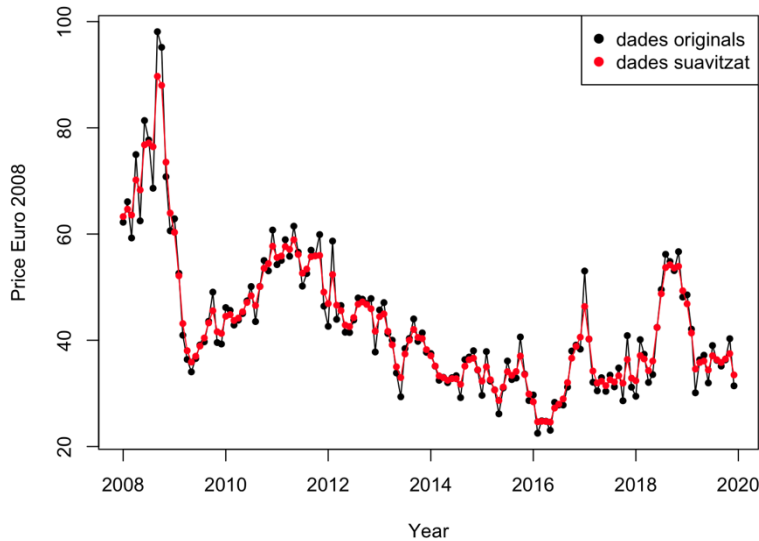


Figura III-24 Dades suavitzades de la sèrie temporal dels preus d'electricitat del mercat majorista

Amb el suavitzat s'aconsegueix mitigar punts atípics. Realitzant una prova de *Dickey-Fuller* es rebutja hipòtesi nul·la d'absència d'arrels unitàries, per tant, s'assumeix que la sèrie estacionària.

Suavitzat de diferencials Electr. Maj.

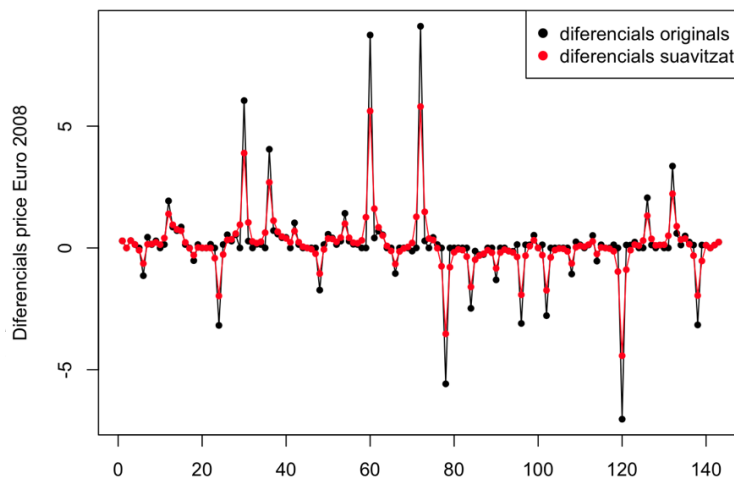


Figura III-27 Diferencials de les dades suavitzades de la sèrie temporal dels preus d'electricitat del mercat majorista

Aquest fenomen podria ser rebutjable tal com s'ha vist visualment, per tant, es farà un diagnòstic del model univariat de sèries temporals òptim sense cap ordre integrador, si aquest no compleix els requisits es plantejarà un model que contempli les dades diferenciades, tal com s'observen a la **figura III- 27**.

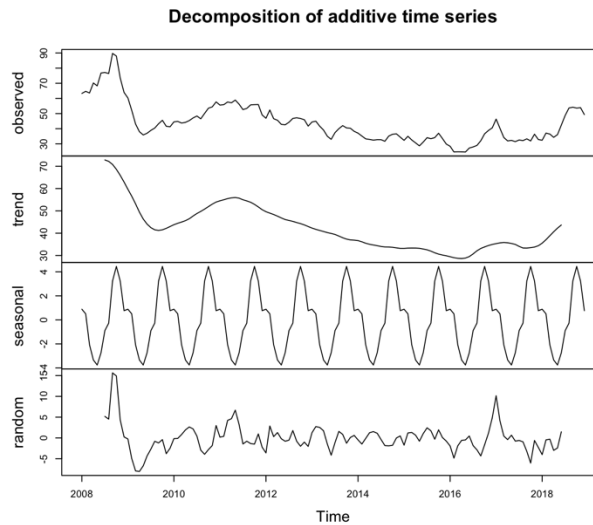


Figura III-26 Descomposició additiva de la sèrie de preus d'electricitat del mercat majorista

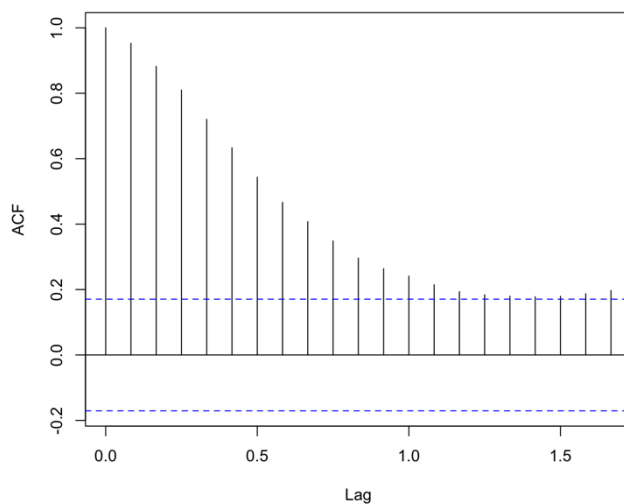


Figura III-28 Autocorrelació dels preus d'electricitat del mercat majorista

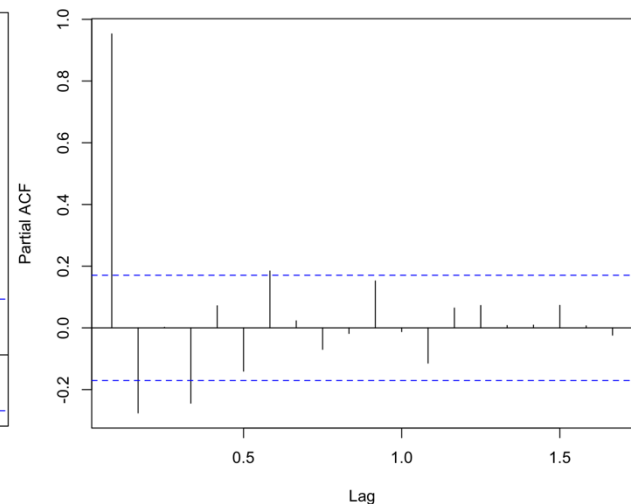


Figura III-27 Autocorrelació parcial dels preus d'electricitat del mercat majorista

En el gràfic de la descomposició de la sèrie es pot veure, que un cop només ens quedem el component de tendència de les dades, aquesta sembla. El tram de preus alts situats a l'inici de la mostra sembla ser un inconvenient a l'hora de prendre la sèrie com estacionària.

Per altra banda, el gràfic d'autocorrelació mostra uns retards significatius en gairebé tots els nombres que van disminuint. Suggereix que la sèrie hauria de tenir un regressor, però també suggereix que la sèrie no es estacionària i que s'haurà de tractar amb diferències.

Per la part d'autocorrelació parcial, els patrons de retards parcials significatius, alguns positius i d'altres negatius dona a pensar que pot haver-hi un component estacional que serà necessari parametritzar.

A continuació, malgrat que amb la prova de *Dickey-Fuller* s'hagués arribat a la conclusió contrària, es realitza l'anàlisi d'autocorrelació i autocorrelació parcial amb la sèrie transformada mitjançant diferències, és a dir, amb un ordre d'integració.

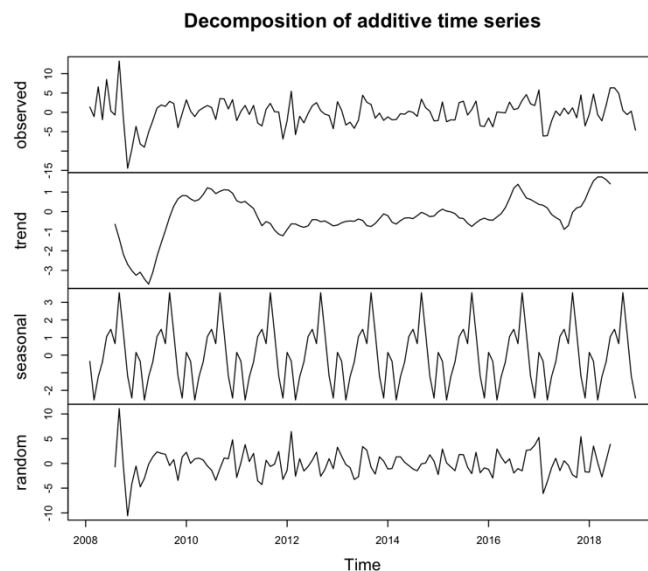


Figura III-29 Descomposició additiva de la sèrie de diferències de preus d'electricitat del mercat majorista

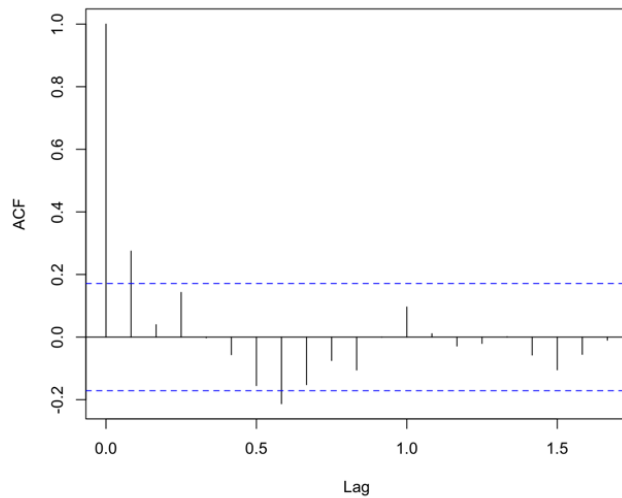


Figura III-31 Autocorrelació de diferències dels preus d'electricitat del mercat majorista

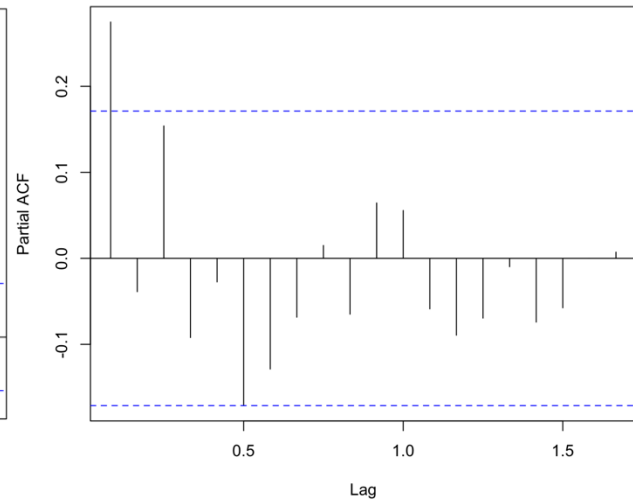


Figura III-30 Autocorrelació parcial de diferències dels preus d'electricitat del mercat majorista

Observant la representació gràfica de la sèrie temporal descomposta, es veu com el component de tendència es manté més estable que quan es tractava de la sèrie original de dades dels preus d'electricitat del mercat majorista. Pel que fa al component d'estacionalitat té un patró diferent del qual s'anava veient de les altres sèries de preus d'electricitat, on a part d'observar un pic que es repeteix a finals del primer semestre de cada any, també s'observa un pic menor a principis de cada any.

Pel que fa a la representació de l'autocorrelació de la sèrie diferenciada dels preus de l'electricitat del mercat majorista, s'observa un patró amb només un ordre de retards significatiu, sense cap patró que suggereixi que la sèrie realment no és estacionària. Malgrat que només hi hagi un element significatiu s'observa un cert comportament que podria ser induït per algun ordre d'estacionalitat, patró que també s'observa a la representació de l'autocorrelació parcial de la sèrie.

En termes d'autocorrelació parcial, destaca un primer ordre de retard parcial. S'observa com al principi de la representació gràfica es veu com l'autocorrelació parcial s'intercala amb valors positius i negatius fins que a partir del sisè retard parcial, les significances s'estabilitzen i agafen una mena de comportament estacional. Suggereix que serà necessari aplicar una mitjana mòbil i que serà necessari tractar l'estacionalitat de la sèrie.

3.1. Modelització univariant

A continuació es presentarà dues propostes de models ARIMA per parametritzar la part regular de la sèrie, és a dir, sense tenir en compte la part estacional.

Model	AIC	Estimació dels coeficients	$P(> z)$
Model1: ARIMA(0,1,1)	694.8817	ma1 = 0.32705	ar1 = 0.0006092 **
Model2: ARIMA(1,1,1)	694.8849	ar1 = -0.43501 ma1 = 0.74089	ar1 = 0.02886 * ma1 = 2.229e-06 **

Taula III-9 Resultats de models de la part regular de la sèrie temporal de preus d'electricitat del mercat majorista

El valor AIC dels dos models proposats són similars, sent l'ARIMA(0,1,1) lleugerament menor. Prenent com a criteri de selecció això i que a més el p-valor que acompanya al coeficient d'autoregressió de primer ordre no és tant vaig com el de la mitjana mòbil, s'escull com a parametrització de la part regular de la sèrie l'ARIMA(0,1,1).

A continuació es comprova si aplicant un primer ordre d'autoregressió i autoregressió parcial per la part estacional disminueix el valor AIC.

Model	AIC	Estimació dels coeficients	$P(> z)$
Model 1.1: ARIMA(0,1,1)(0,0,1)[12]	694.4688	ma1 = 0.32705 sma1 = 0.187445	ma1 = 0.0005436*** sma1 = 0.1242855
Model 1.2: ARIMA(0,1,1)(1,0,0)[12]	694.7575	ar1 = 0.322365 ma1 = 0.159976	ar1 = 0.0006003 *** ma1 = 0.1415186

Taula III-10 Resultats de models de la part estacional de la sèrie temporal de preus d'electricitat del mercat majorista

Aplicant un ordre d'autocorrelació o bé un ordre d'autocorrelació parcial per la part de la parametrització parcial es disminueix molt lleugerament el valor AIC, però cap d'aquests

coeficients nous resulten ser significatius. En conseqüència s'acaba optant per un model de sèrie temporal pels preus d'electricitat del mercat majorista més senzill, que no contempla memòria de valors passats estacionals de la mostra, el model ARIMA(0,1,1).

3.2. Anàlisi dels residus

La representació dels residus del model finalment triat és la següent:

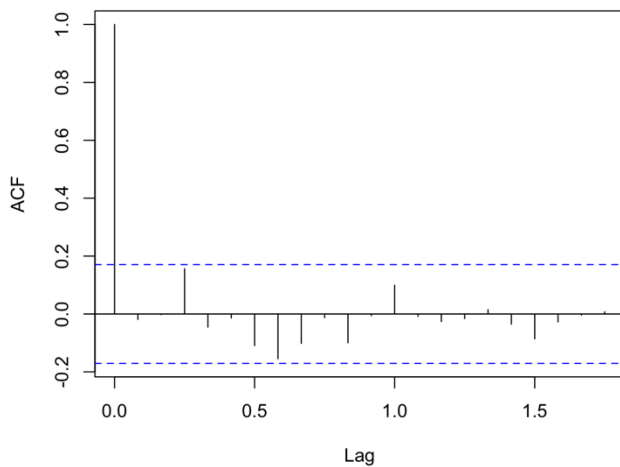


Figura III-33 Autocorrelació dels residus dels preus d'electricitat del mercat majorista amb el Model1

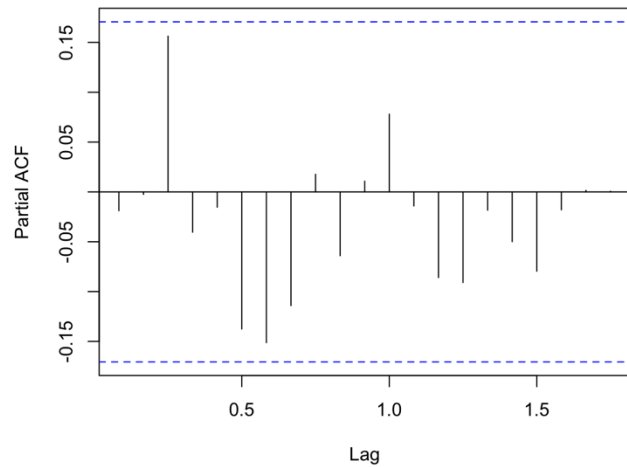


Figura III-32 Autocorrelació parcial dels residus dels preus d'electricitat del mercat majorista amb el Model1

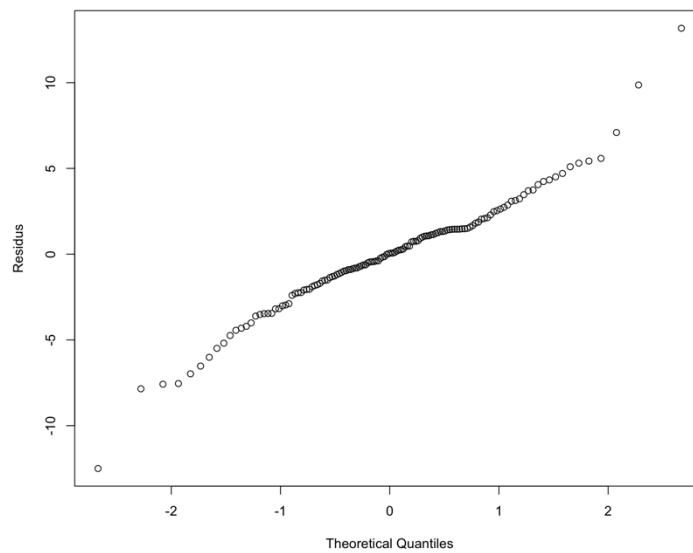


Figura III-34 QQPLOT dels residus dels preus d'electricitat del mercat majorista amb el Model1

Els residus del model de sèrie temporal ARIMA(0,1,1) pels del mercat majorista resulten no presentar cap retard en termes d'autocorrelació ni d'autocorrelació. Destaca com la distribució dels residus representats en un QQPLOT, a diferència del que es porta observant en les altres modalitats de preus de l'electricitat, no té tanta presència de cues descentralitzades. La volatilitat dels preus de cara a l'engròs afecta que els seus valors es consideri com una distribució més independent, si comparem amb els altres preus, que al final és més fàcil de parametritzar mitjançant un model de sèrie temporal univariat.

3.3. Predicció dels valors futurs

A continuació es fa el *forecasting* per a les dades de l'any 2019 amb el model escollit, ARIMA(0,1,1). Es farà també la comparació dels preus d'electricitat del mercat majorista reals.

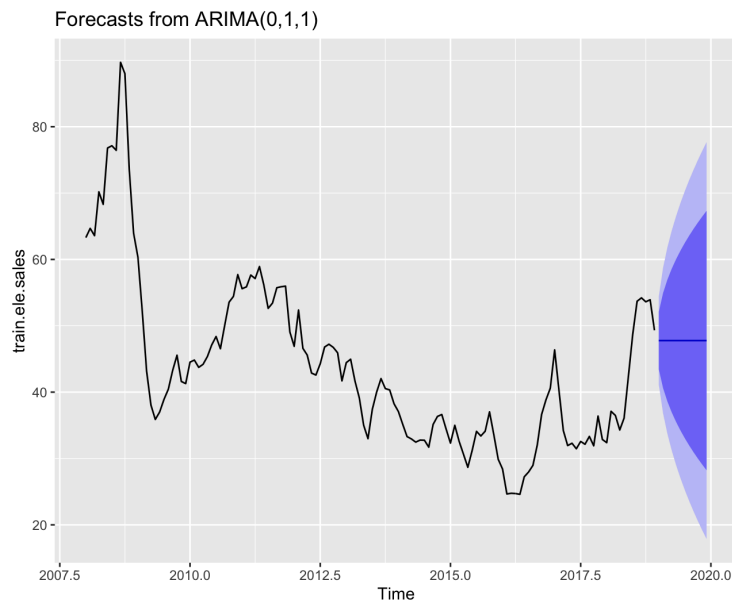


Figura III-35 Forecasts del model ARIMA(0,1,1) dels preus d'electricitat del mercat majorista amb intervals de confiança (de fins el 95%) en blau

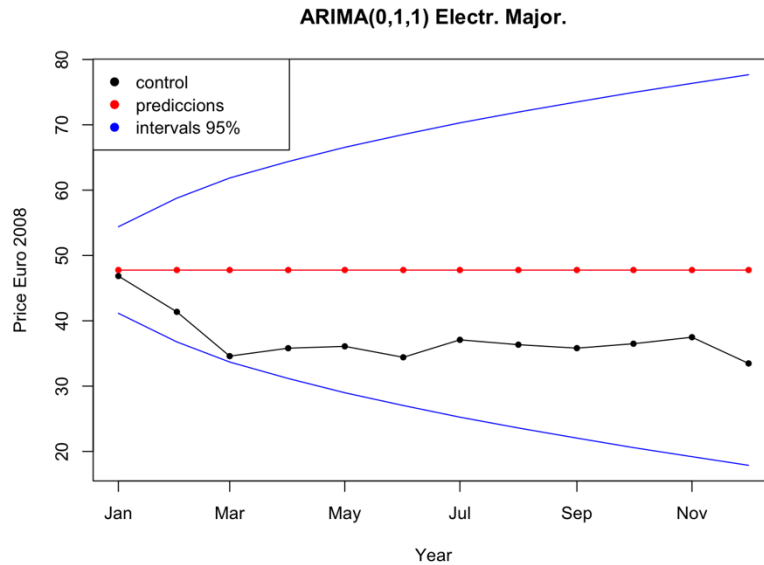


Figura III-36 Comparació amb test de les prediccions dels preus d'electricitat del mercat majorista

Malgrat els intervals de confiança que proporciona el model aconsegueix capturar totes les observacions reals, les prediccions dels preus de l'electricitat del mercat majorista es mantenen igual en tot el període, com era d'esperar, ja que no s'estigui tenint en compte un ordre de mitjanes mòbils.

TEST	PREDICT	Lower95%	Upper95%
46.85315	47.77153	41.14871	54.39435
41.36540	47.77153	36.76675	58.77630
34.59104	47.77153	33.68791	61.85514
35.80513	47.77153	31.17058	64.37248
36.08792	47.77153	28.98763	66.55543
34.41088	47.77153	27.03320	68.50985
37.08698	47.77153	25.24773	70.29532
36.34057	47.77153	23.59375	71.94930
35.81298	47.77153	22.04590	73.49716
36.48850	47.77153	20.58603	74.95702
37.48610	47.77153	19.20066	76.34239
33.47380	47.77153	17.87942	77.66363

Taula III-11 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat pel mercat majorista

Capítol 4

IV. ANÀLISI DE SÈRIES TEMPORALS MULTIVARIANT

1. Anàlisi descriptiu de la resta de variables

A continuació, es farà un anàlisi exploratori de les sèries temporals dels preus del gas i de combustibles fòssils, considerant els diferents consumidors a qui van dirigits. L'objectiu d'aquest anàlisi és tenir una percepció dels principals estadístics que representen aquestes variables, perquè aquestes es tindran en compte en el model multivariant de sèries temporals. S'espera que el conjunt de variables de preus més la informació de temperatures ajudin a fer un model multivariant més robust amb més capacitat predictiva i en conseqüència, tenir unes prediccions dels preus d'electricitat més ajustades.

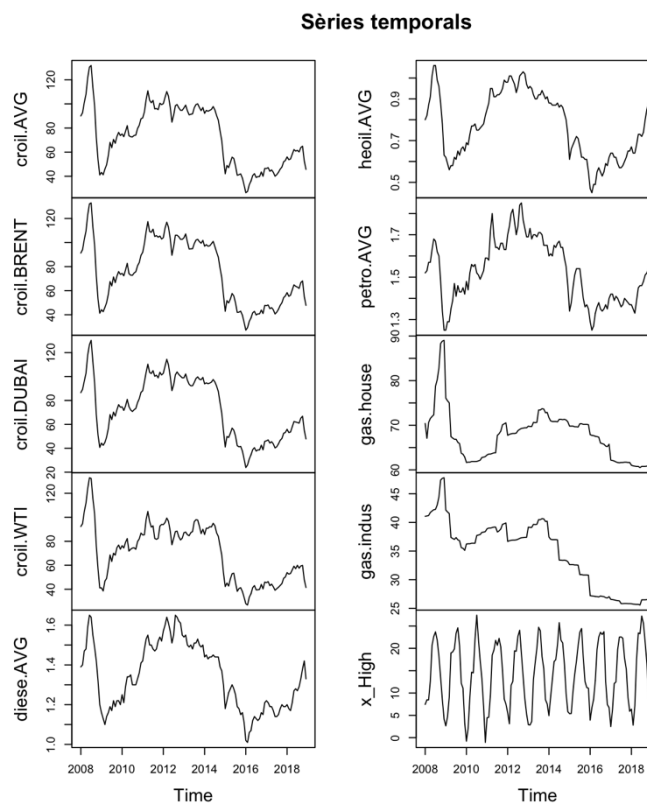


Figura IV-1 Sèries temporals dels preus del gas, combustibles fòssils i temperatura

Les sèries temporals de preus de combustibles fòssils són les *croil.avg*, *croil.BRENT*, *croil.DUBAI*, *croil.WTI*, *diese.AVG*, *heoil.AVG* i *petro.AVG*; totes elles es comporten de forma similar. Entre els anys 2008 i 2009 presenten un pic màxim molt pronunciat, segurament per algun tipus de regulació dels preus o per factors externs. A partir de l'any 2009 fins a aproximadament el 2013 la tendència ha sigut alcista. En els últims anys de la mostra el canvi de tendència canvia i sembla que els preus s'estabilitzen.

Pel que fa al gas, tant el domèstic com l'industrial (les variables *gas.house* i *gas.indus*) presenta pujades intermitents. Aquest comportament també s'observava a les sèries d'electricitat, evidenciant que els preus d'aquests recursos energètics realment són intervinguts per figures externes a les dades.

Pel que fa als màxims mensuals de la temperatura, el comportament que es reflecteix al gràfic és l'esperable. Molt estacional i estable.

	Sèrie temporal	Mínim	1r quartil	Mediana	Mitjana	3r quartil	Màxim
<i>croil.AVG</i>	Petroli cru	26.31€	47.51€	73.31€	72.81€	95.29€	131.85€
<i>croil.BRENT</i>	Brent	27.21€	49.88€	73.22€	75.31€	100.53€	132.89€
<i>croil.DUBAI</i>	Dubai	23.85€	48.83€	72.22€	73.08€	97.06€	130.26€
<i>croil.WTI</i>	WTI	26.85€	45.84€	74.38€	70.04€	88.19€	132.94€
<i>diese.AVG</i>	Diesel	1.01€	1.19€	1.33€	1.34€	1.49€	1.65€
<i>heoil.AVG</i>	Combustible líquid	0.45€	0.63€	0.77€	0.77€	0.91€	1.06€
<i>petro.AVG</i>	Petroli	1.25€	1.39€	1.52€	1.53€	1.65€	1.85€
<i>gas.house</i>	Gas domèstic	60.51€	62.63€	67.92€	67.73€	70.59€	89.03€
<i>gas.indus</i>	Gas industrial	25.58€	27.14€	36.84€	34.62€	39.03€	47.79€
<i>x_High</i>	Temperatures màximes	-1.06°C	7.56 °C	13.92 °C	14.16 °C	20.68 °C	27.43 °C

Taula IV-1 Estadístics de sèries temporals del gas, combustibles fòssils i temperatura

A la taula descriptiva de cada sèrie temporal s'observa com tots els indicadors de preu del petroli cru (el mitjà, el *Brent* i el *Dubai*) s'observa que presenten uns estadístics molt similars, serà pertinent prescindir d'alguna d'aquestes mesures però per fer-ho caldrà estudiar-les en un correlograma.

2. Selecció de la base mostral definitiva

En el següent anàlisi es construirà un model de sèries temporals multivariant amb l'objectiu d'obtenir unes prediccions dels preus de l'electricitat a futur que s'ajustin el màxim possible a la realitat de les dades. Les prediccions dels preus de l'electricitat ara també es construiran de la mateixa forma que s'ha fet a l'anàlisi univariant, es reserva una part de la mostra per testejar la capacitat predictiva del model i amb aquesta es confirmarà si el model proposat és adequat o si cal replantejar els paràmetres.

El següent anàlisi exploratori permetrà tenir una percepció de com es comporten conjuntament les sèries temporals, mitjançant un correlograma.

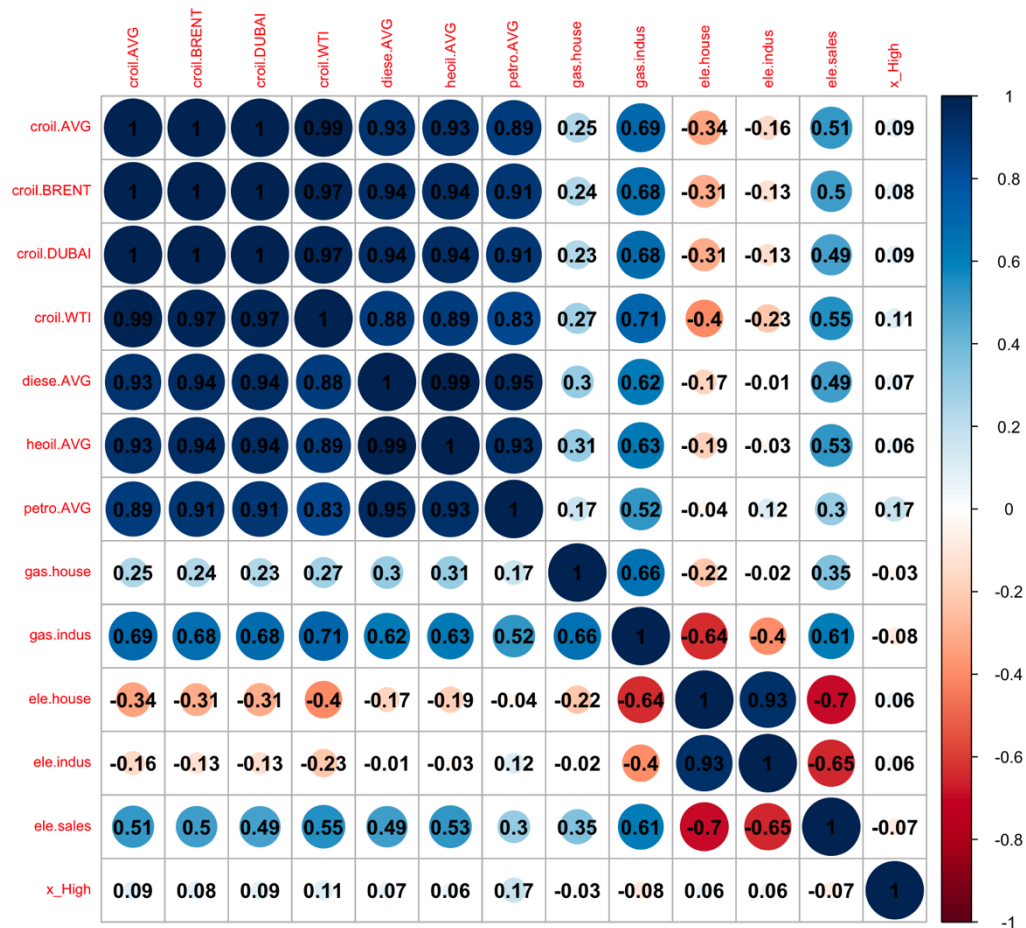


Figura IV-2 Correlograma del conjunt de sèries temporals

En avaluar els resultats del correlograma, es pren la decisió de descartar algunes variables, els motius són:

- 1) Evitar multicol·linealitat del model amb variables altament correlacionades amb altres, ja que aquestes acabaran aportant informació molt similar entre elles i pot provocar *overfitting*.
- 2) Optimitzar el nombre de variables a afegir al model. Eliminant variables redundants se simplifica el model i la seva interpretació.

El criteri que es segueix per descartar variables serà principalment eliminar les que tenen una correlació a prop de l'1, com es pot observar que passa amb les variables situades a la zona superior esquerra del correlograma. Totes elles variables de combustibles fòssils. Naturalment aquestes són molt similars. Dins d'aquestes es seleccionaran les que podrien explicar més el comportament dels preus d'electricitat i de gas. Així doncs, s'acaben seleccionant les següents variables:

- Preus de combustibles fòssils:
 - Petroli (*petro.AVG*)
 - Petroli cru WTI (*croil.WTI*)

- Preus del gas:
 - Preu del gas dirigit al consumidor domèstic (*gas.house*)
 - Preu del gas dirigit a la indústria (*gas.indus*)

- Mesures de temperatura:
 - Temperatura màxima mensual (*x_High*)

- **Preus de l'electricitat i objectes d'interès per analitzar les prediccions:**
 - Preu de l'electricitat dirigit al consumidor domèstic (*ele.house*)
 - Preu de l'electricitat dirigit a la indústria (*ele.indus*)
 - Preu de l'electricitat dirigit al mercat majorista (*ele.sales*)

D'aquesta manera es passa d'un conjunt inicial de 13 sèries temporals a un base de 8 sèries temporals que conformaran el model VARMA.

Un cop determinat això, es fa una visualització de les sèries i es compara amb el mateix conjunt però havent filtrat les observacions amb un suavitzat de *Kalman*.

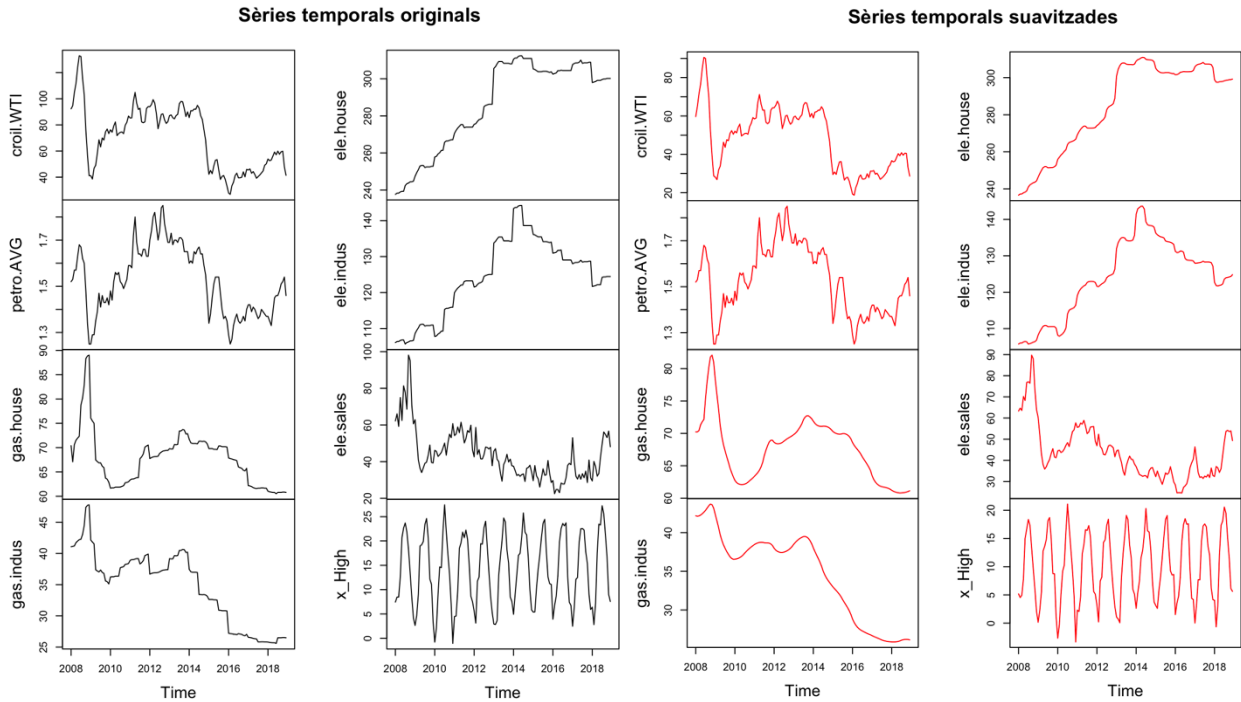


Figura IV-4 Conjunt de sèries temporals

Figura IV-3 Conjunt de sèries temporals suavitzades

En els gràfics anteriors s’observa com, a les dades no suavitzades, en els preus afegits de gas i combustibles fòssils passa quelcom similar al que es portava observant amb els indicadors de preu de l’electricitat. Es poden veure pics i comportaments similars a altiplans, on en algun punt del temps sembla que es realitza alguna regulació o factor extern a les dades, que fa canviar el preu radicalment, i aquest es manté durant un temps. Les mateixes dades havent aplicat el suavitzat de *Kalman*, malgrat que encara presenten certs pics molt elevats o inferiors a la gran part de les seves pròpies observacions, mitiga certs punts de la mostra que un altre cas es prendrien com valors atípics. Es descarta aplicar el suavitzat a les variables de petroli i temperatures, ja que tots els valors registrats es mostren més adequats sobre el conjunt de tota la seva història.

Tenint en compte que a l’anàlisi univariant que s’ha fet anteriorment per les sèries temporals dels preus de l’electricitat s’ha determinat que era necessari utilitzar un ordre d’integració de les

dades, es procedeix a aplicar també una transformació mitjançant diferències també a la resta de variables afegides.

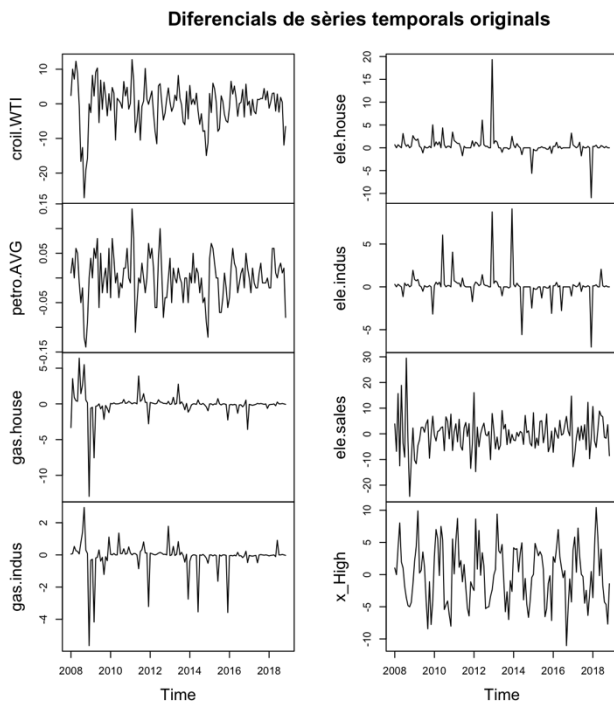


Figura IV-5 Conjunt de diferencials de les sèries temporals

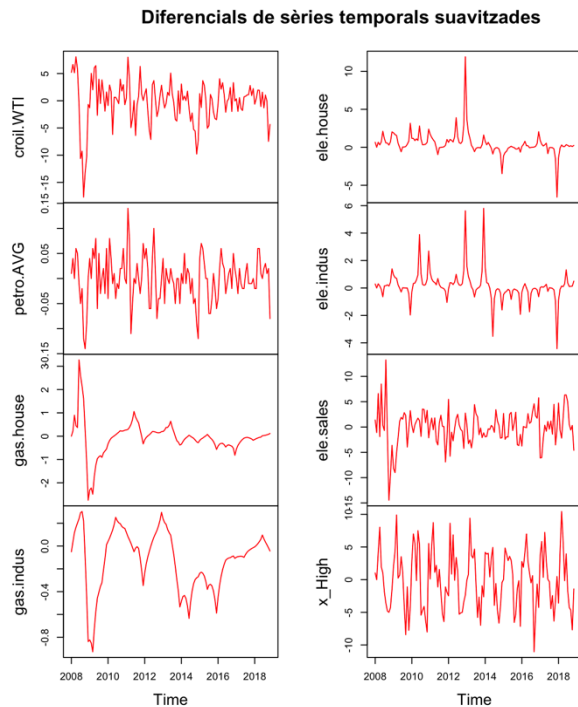


Figura IV-6 Conjunt de diferencials de les sèries temporals suavitzades

Aplicant una prova de *Dickey-Fuller* per a cada una de les sèries temporals diferenciades que conformen el total de la mostra d'estudi, es conclou que totes elles són estacionàries i que ja és pertinent poder-hi aplicar un model multivariant.

3. Modelització multivariant

En aquest apartat es pretén buscar el model de sèries temporals multivariant que millor s'adapti als preus de tots els recursos energètics. Per aconseguir-ho, l'enfoc inicial es r un model VARMA, és a dir, que tingui un ordre autoregressiu p i de mitjanes mòbils q .

Com trobar una parametrització adequada del model $VARMA(p, q)$ pot ser complex, pel fet que pot haver-hi una ampla gamma de combinacions entre ordres autoregressius i ordres de mitjanes mòbils sobre una configuració multivariada on les sèries temporals s'interrelacionen; es farà un procés iteratiu per trobar els ordres p i q òptims. El procés iteratiu només recorrerà

certs models $VARMA(p, q)$ que puguin tenir sentit basant-nos en els models $ARIMA(p, d, q)$ de cada una de les sèries temporals de la mostra.

Els models de sèries temporals univariants dels preus de l'electricitat ja han sigut plantejats i provats. Falta tenir una percepció de quins serien els models univariants dels preus del gas, dels combustibles fòssils i de la temperatura. En aquest cas es farà mitjançant la funció *auto.arima* de la llibreria *forecast* d'R, que dona com a resultat els valors AIC de diferents ARIMA plantejats a la sèrie en qüestió.

A continuació es recullen els models $ARIMA(p, d, q)$ idonis de totes les sèries temporals:

Sèrie temporal	Model univariant	AIC
<i>Electricitat domestic</i>	ARIMA(1,1,0)(0,1,1)[12]	435.8136
<i>Electricitat industrial</i>	ARIMA(1,1,0)	373.0786
<i>Electricitat majorista</i>	ARIMA(0,1,1)	694.8817
<i>Petrolí cru</i>	ARIMA(1,1,0)(0,0,2)[12]	811.63
<i>Petrolí cru</i>	ARIMA(1,1,0)(0,0,1)[12]	-455.3
<i>Gas domestic</i>	ARIMA(1,1,1)	511.5926
<i>Gas industrial</i>	ARIMA(1,1,1)(1,0,0)[12]	363.3263

Taula IV-2 Models univariants de les sèries temporals de tots els recursos energètics

Es pot observar com tots els models ARIMA de totes les sèries, menys en el cas de l'electricitat majorista, contenen un ordre d'autoregressió mentre que pel que fa a l'ordre de mitjanes mòbils a tenir en compte hi ha més discrepància.

Així doncs, el procés iteratiu que executarà els models $VARMA(p, q)$, recorrerà totes les opcions de parametrització de p i q que s'observen a la **taula IV-2** i també es provaran ordres de fins a 2 retards i 2 mitjanes mòbils⁵. A continuació es farà una visualització de tots els valors AIC extrets i es busca identificar quina és la combinació de p i q òptima.

⁵ El nombre de variables així com la mida mostral fa que implementar un model VARMA sigui costós en termes de temps de màquina i recursos. Es busca fer un procés iteratiu que no sigui molt extens, executant només aquelles parametritzacions del model que es creu que aportaran conclusions.

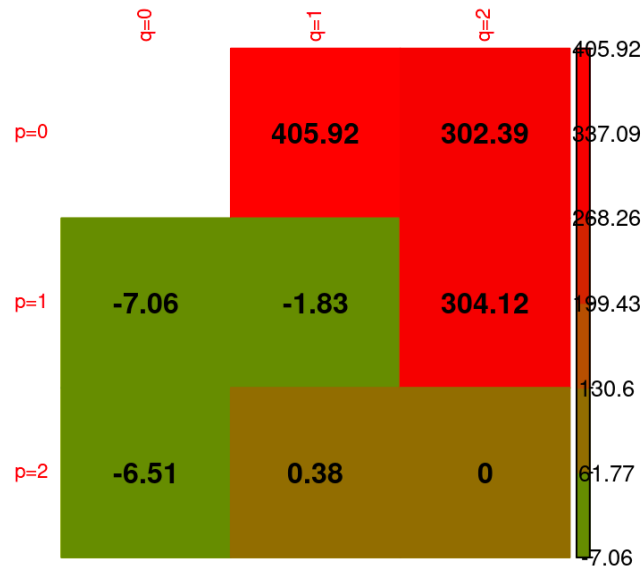


Figura IV-7 Valors AIC dels models $VARMA(p,q)$

A la representació gràfica anterior es pot observar els valors AIC resultants d'aplicar un model VARMA amb els valors p i q (especificats en l'eix vertical i horitzontal, respectivament). Es veu com l'AIC augmenta quan s'apliquen ordres de mitjanes mòbils.

Es decideix descartar l'opció d'aplicar un model $VARMA(p, q)$ i procedir a determinar el nombre òptim de retards que seran utilitzats en el model $VAR(p)$. La parametrització d'un model on només s'ha d'escollir l'ordre autoregressor és més senzill, per tant també es configura un procés iteratiu on es provarà fins a 12 retards, que és el nombre d'observacions del component estacional de totes les sèries temporals.

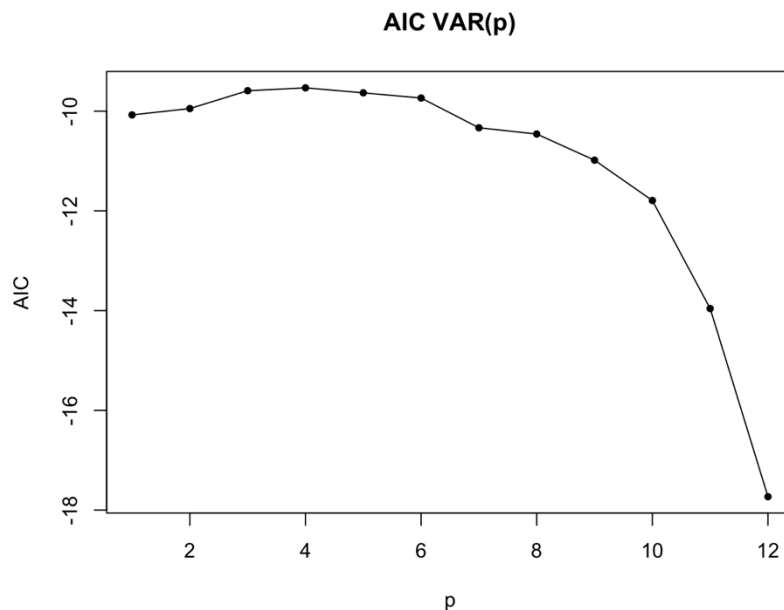


Figura IV-8 Sèrie de valors AIC en funció del número de retards al model VAR

En aquest cas el valor p que minimitza l'AIC és 12. Pel que s'utilitzarà un model VAR amb 12 retards, sent aquest el que proporciona més informació al model.

4. Causalitat de Granger

Un cop modelitzat el conjunt de sèries temporals es fa una prova de Causalitat de Granger. Amb aquesta prova es pot determinar, mitjançant test d'hipòtesis, si una sèrie temporal es pot utilitzar per explicar les altres i viceversa, si les altres sèries temporals poden explicar a la sèrie en concret. Sent la hipòtesis nul·la la suposició la sèrie temporal en concret no es causada per les demés i l'alternativa que si que hi ha causalitat.

En aquest cas d'estudi, totes les sèries temporals rebutgen la hipòtesi nul·la de no causalitat, és a dir que es pot assumir que hi ha causalitat.

Amb la realització d'aquest test d'hipòtesis es pot concloure que:

- Els preus d'electricitat són causats entre ells mateixos. És a dir, el preu d'electricitat dirigit a l'ús domèstic variarà segons els preus que vagin per les indústries i al determinat en el

mercat majorista. És dona el mateix en totes les combinacions de causalitat de preus d'electricitat.

- El preu d'electricitat, dirigit a qualsevol consumidor, és causat pels preus que es donen del gas (dirigit a la indústria i a l'ús domèstic), de la mateixa manera que es causat pels preus dels combustibles fòssils (preu del petroli i l'indicador del preu *WTI*).
- Els preus dels recursos energètics també són causats per les temperatures màximes que enregistrades mensualment.

5. Anàlisi dels residus

Per poder avaluar la qualitat del model utilitzat, es grafiquen les sèries dels residus. Es busca que aquests siguin estacionaris i que no presentin cap patró destacable d'autocorrelació ni autocorrelació parcial. Altrament dit, que es comportin com soroll blanc.

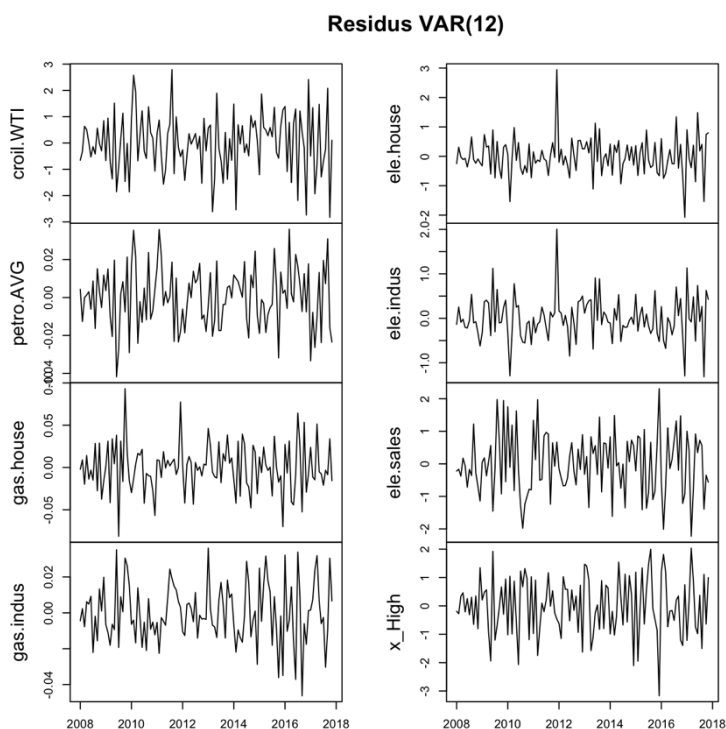


Figura IV-9 Sèrie dels residus del model VAR(12)

Jutjant els gràfics de cada una de les distribucions dels residus, es conclou que els residus són estacionaris. També s'aplica una prova de *Dickey-Fuller* en totes les sèries de residus i es conclou que no tenen arrels unitàries.

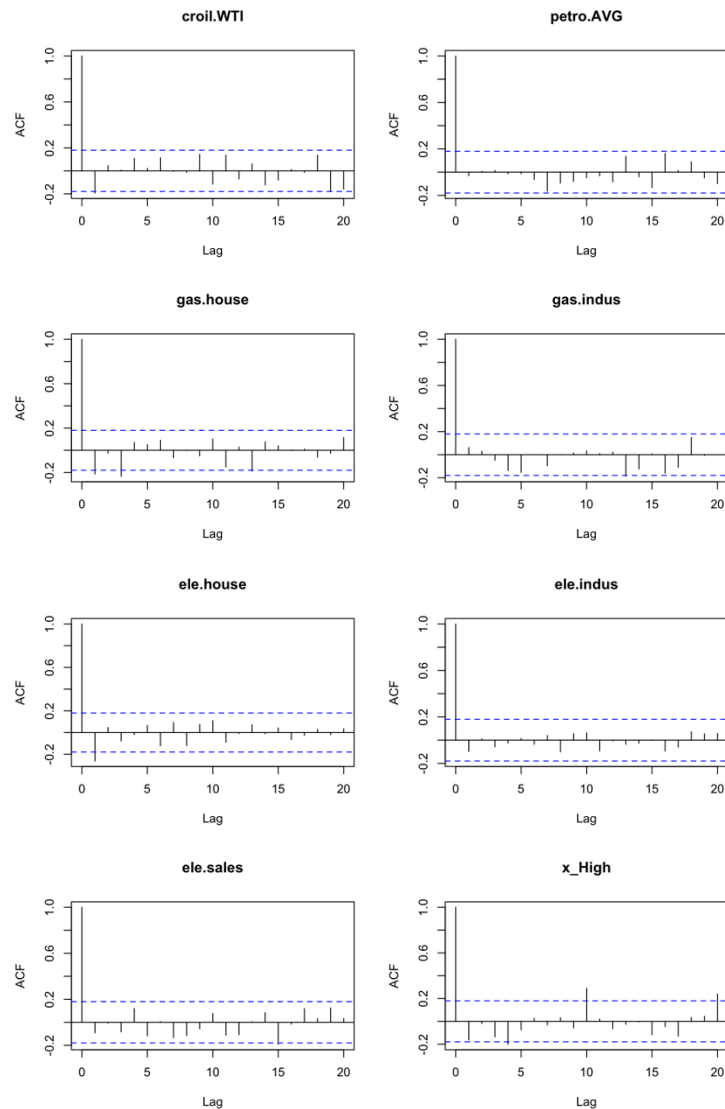


Figura IV-10 Autocorrelacions dels residus del model VAR(12)

A la **figura IV- 10** estan representades les autocorrelacions de totes les sèries de residus resultants del model *VAR(12)*. No es percep que hi hagi algun comportament molt destacable de significació. Tret de les sèries del gas domèstic i electricitat domèstica (variables *gas.house* i *ele.house*) on s'observa un primer nombre d'autocorrelació que sobresurt lleugerament del llindar de significació.

Tenint en compte que els preus dirigits al consumidor domèstics són els més intervinguts per l'estat i figures externes, És comprensible que un model multivariant presenti dificultats per ajustar-se al màxim a aquest tipus de situacions.

De totes maneres, el comportament en termes d'autocorrelació del conjunt de les sèries de residus es correcte i podria considerar-se similar al del soroll blanc. Caldrà comprovar que l'autocorrelació parcial dels residus tampoc presenti número de retards significatius.

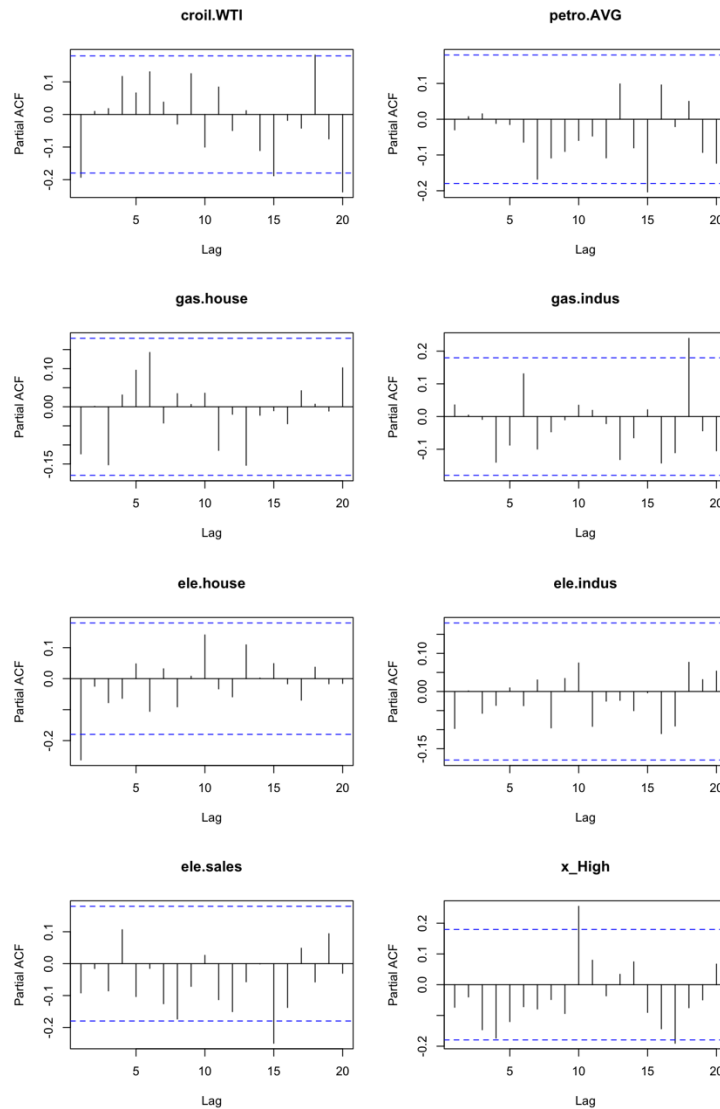


Figura IV-11 Autocorrelacions parcials dels residus del model VAR(12)

No destaquen retards destacables en termes d'autocorrelació parcial de les sèries de residus. Destaca un primer nombre de retards parcials lleugerament per sobre del llindar de significació de la variable de preus domèstics de l'electricitat, a causa del que és una sèrie de preus més intervinguts per factors externs.

Tenint en compte una visió general de les autocorrelacions i autocorrelacions de totes les sèries de residus, es determina que són molt similars a soroll blanc.

6. Predicció de valors futurs

Després de corroborar, mitjançant l'anàlisi dels residus, que l'aplicació del model $VAR(12)$ és adient pel conjunt de sèries temporals es procedeix a realitzar les prediccions de preus futurs per a cada una de les sèries temporals. L'objectiu és aconseguir unes prediccions de les sèries, especialment les dels preus de l'electricitat, que convergeixin i que en conseqüència, tinguin un millor pronòstic.

Es farà una predicció de 12 observacions que comprendrien l'any 2019. Això permetrà fer una comparació de les prediccions amb la sèrie real de 2019, que ha sigut reservada per fer aquestes comprovacions com a grup control.

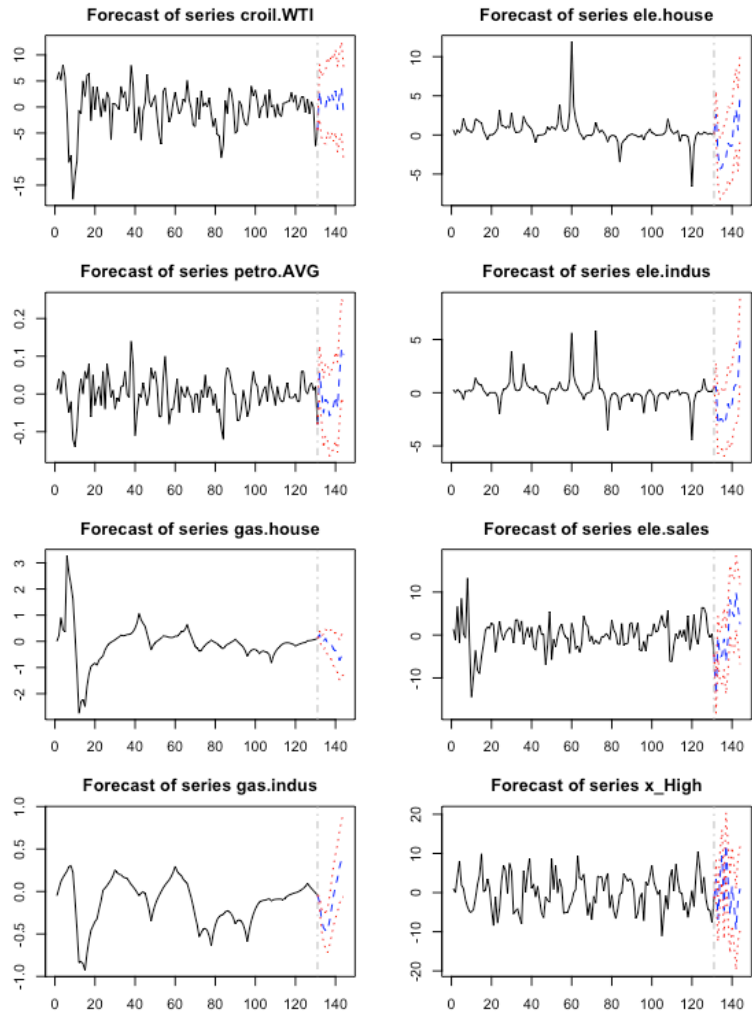


Figura IV-12 Representació de les prediccions en blau amb intervals de confiança (del 95%) en vermell

A les gràfiques representades a la **figura IV-12** s'observen els *forecasts* de totes les sèries diferenciades amb uns intervals de confiança del 95%.

Aquests són els punts observables a destacar d'aquestes representacions dels *forecasts*:

- Diferencials dels preus de combustibles fòssils (*croil.WTI*, *petro.AVG*):
 - o Pel que fa les prediccions del preu WTI s'observa una continuació de la sèrie estable al voltant del 0. En el cas del preu mitjà de petroli també es mantenen al voltant del 0 excepte a finals de 2019 on preveu un augment pronunciat.
- Diferencials dels preus del gas domèstic i industrial (*gas.house*, *gas.indus*):

- Mentre que els diferencials predits pels preus del gas domèstic disminueixen, els preus del gas dirigit a les indústries preveu uns augments doncs es tracta d'una sèrie amb més variabilitat.
- Diferencials dels màxims mensuals de les temperatures (*x_High*):
 - Les prediccions dels diferencials de la temperatura presenta un comportament esperable, amb forts augments a la mitat de l'any, amb disminucions per a finals d'any.
- **Diferencials dels preus de l'electricitat domèstic, industrial i majorista** (*ele.house, ele.indus, ele.sales*):
 - Pel que fa el preu domèstic, els valors predits indiquen una forta disminució dels preus durant a principis d'any, mentre que els següents valors predits continuen disminuint cada vegada amb menys intensitat.
 - Els preus de la indústria presenten un comportament similar al domèstic però amb una menor variació. S'espera una estabilització dels preus més evident que en el cas domèstic.
 - La predicció dels diferencials dels preus del mercat resulten ser més volàtils amb presència de 2 tendències, una primera de disminució de preus durant el primer semestre de 2019 i una recuperació en el segon semestre.

A continuació es farà la comparació dels valors predits de l'electricitat amb els valors reals del 2019 que s'han reservat com a grup control i no han participat dins la mostra d'entrenament del model.

7. Comparació amb grup control

A continuació es farà la transformació dels valors en forma de diferències entre observacions de les sèries per poder fer la comparació amb valors de preus del grup control.

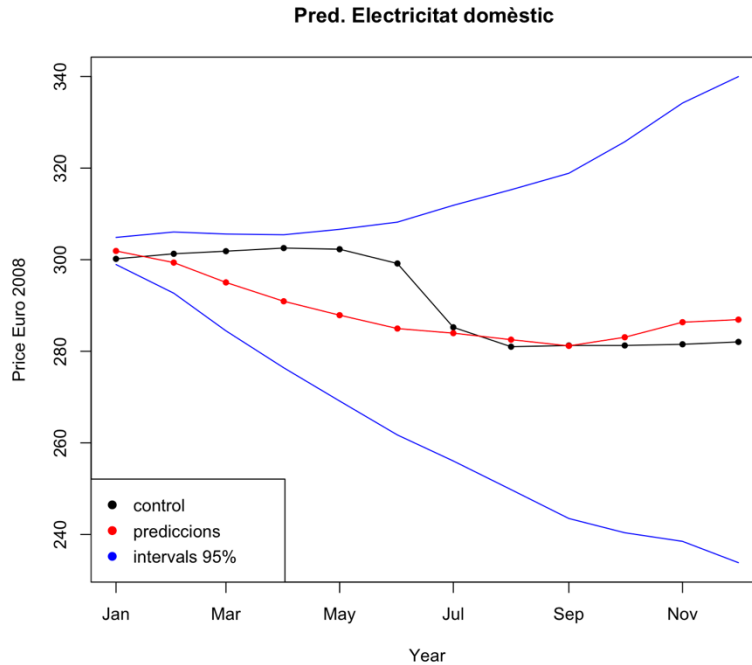


Figura IV-13 Representació de les prediccions contra el grup control de preus d'electricitat pel consumidor domèstic

TEST	PREDICT	Lower95%	Upper95%
300.1782	301.9034	298.9450	304.8618
301.2868	299.3637	292.6793	306.0481
301.8557	295.0310	284.4637	305.5984
302.5402	290.9197	276.3983	305.4411
302.2856	287.8827	269.1387	306.6268
299.1675	284.9644	261.7323	308.1965
285.2353	283.9577	256.0480	311.8674
280.9941	282.5338	249.7948	315.2728
281.2641	281.1776	243.4892	318.8661
281.2641	283.0603	240.3765	325.7442
281.5241	286.3481	238.4966	334.1996
282.0441	286.9086	233.8381	339.9792

Taula IV-3 Comparació de test, prediccions i intervals de confiança dels preus domèstics d'electricitat amb model multivariant

En comparar les prediccions dels preus d'electricitat dirigits al consumidor domèstic amb el grup de control, s'observa com els valors predits indicaven una disminució al llarg del període de l'any 2019. Encara que aquesta disminució dels preus, a la sèrie del grup control es presenti com un canvi més abrupte i a les prediccions es presenti amb una tendència més suau al llarg del temps, el model fa unes prediccions raonables.

En aquest cas el model té una cobertura total de les observacions reals dels preus de l'electricitat domèstics.

A continuació es fa la visualització dels preus de l'electricitat dirigits a les indústries:

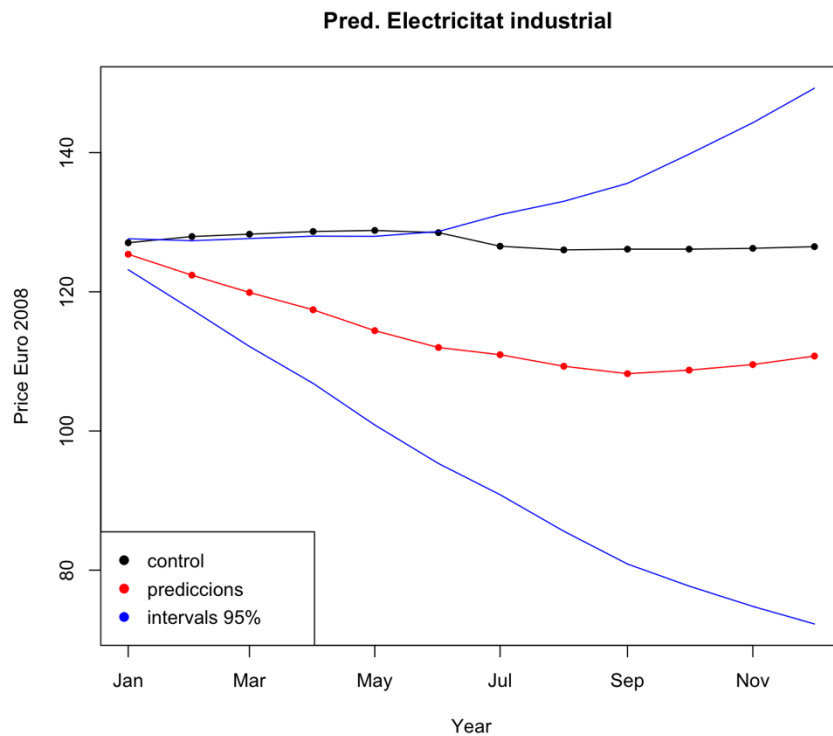


Figura IV-14 Representació de les prediccions contra el grup control de preus d'electricitat per a la indústria

TEST	PREDICT	Lower95%	Upper95%
127.0493	125.3973	123.17035	127.6242
127.9381	122.3854	117.43568	127.3351
128.2781	119.8992	112.14771	127.6508
128.6638	117.4147	106.83664	127.9928
128.8165	114.4148	100.85977	127.9697
128.5022	111.9941	95.32619	128.6620
126.5492	110.9598	90.84001	131.0797
126.0163	109.2992	85.59666	133.0017
126.1263	108.2346	80.88522	135.5840
126.1263	108.7429	77.71049	139.7753
126.2463	109.5402	74.79385	144.2866
126.4863	110.7620	72.27089	149.2532

Taula IV-4 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat per a la indústria amb model multivariant

En el cas de les prediccions dels preus d'electricitat dirigits a les indústries, s'observa com els primers punts predits presenten una discrepància important amb les dades reals del grup control. Això provoca que, encara que les prediccions dels diferencials de finals de 2019 siguin més similars a les dades reals, el *gap* que es genera a les primeres observacions persisteixi en tot el període del 2019.

Per últim, es fa la representació d'observacions de *train/test* pels preus de l'electricitat del mercat majorista:

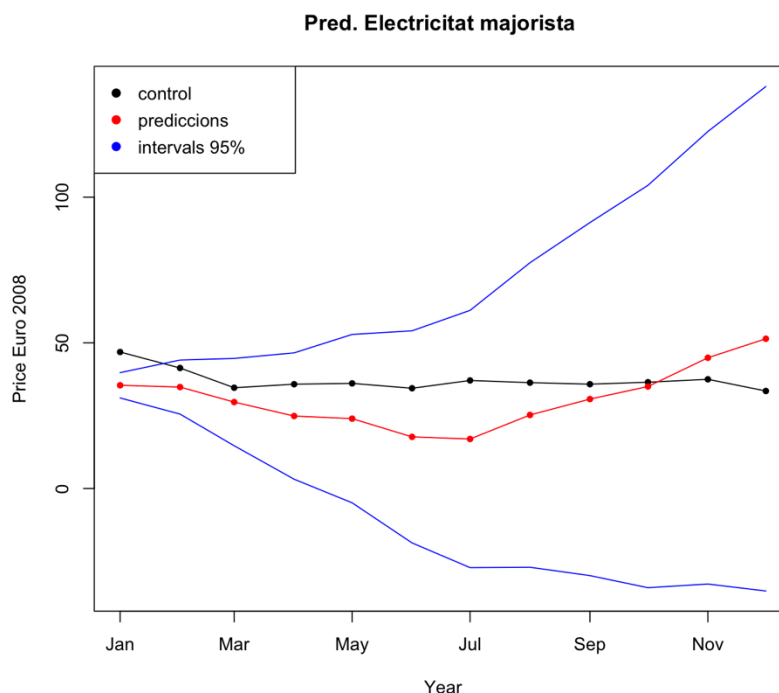


Figura IV-15 Representació de les prediccions contra el grup control de preus d'electricitat del mercat majorista

TEST	PREDICT	Lower95%	Upper95%
46.85315	35.43088	31.091208	39.77056
41.36540	34.82301	25.545218	44.10079
34.59104	29.66813	14.679163	44.65711
35.80513	24.87720	3.193786	46.56061
36.08792	23.99187	-4.903848	52.88760
34.41088	17.72844	-18.671209	54.12809
37.08698	16.99767	-27.153151	61.14849
36.34057	25.23906	-27.027467	77.50558
35.81298	30.69716	-29.886804	91.28112
36.48850	35.00638	-34.048020	104.06079
37.48610	44.87579	-32.785262	122.53684
33.47380	51.40703	-35.168319	137.98238

Taula IV-5 Comparació de test, prediccions i intervals de confiança dels preus de l'electricitat majorista amb model multivariant

A les prediccions dels preus de l'electricitat del mercat majorista, s'observen uns preus predits amb una tendència més baixa durant el primer semestre de 2019 respecte el grup control. Malgrat que el model ha sobreestimat la disminució de la sèrie en aquest tram temporal, les prediccions del segon semestre augmenten trobant-se de nou amb observacions del grup control.

Els intervals de confiança associats a cada predicció donen cobertura a totes les observacions del grup control.

Capítol 5

V. DISCUSSIÓ DEL RENDIMENT DELS MODELS DELS PREUS D'ELECTRICITAT

A la taula proporcionada de continuació es presenten indicadors de mesura de la capacitat predictiva dels models de sèries temporals univariants i multivariants dels preus dels recursos energètics d'interès, els de l'electricitat. En concret consta de l'indicador d'arrel de l'error quadràtic mig (*RMSE*) i la taxa de cobertura dels intervals de confiança del 95% que té assignada cada predicció.

Sèrie temporal	Model utilitzat	RMSE	Taxa de cobertura (95% de confiança)
Electricitat - Domèstic	SARIMA(1,1,0)(0,1,1)[12]	11.17725	50%
	VAR(12)	7.369013	100%
Electricitat - Industrial	ARIMA(1,1,0)	2.27145	92%
	VAR(12)	14.08712	67%
Electricitat - Majorista	ARIMA(0,1,1)	11.17725	100%
	VAR(12)	11.80729	100%

Taula 5.1 Resultats de models de sèries temporals univariants i multivariants dels preus de l'electricitat

Examinant els resultats pel que la primera sèrie, la de preus de l'electricitat del consumidor domèstic, denota un rendiment molt bo al model *VAR(12)*. La cobertura de les observacions del grup control es total i el valor *RMSE* és notablement menor que el del model univariant *SARIMA(1,1,0)(0,1,1)[12]*. Aquesta sèrie temporal es caracteritza per un comportament de les observacions inusual amb canvis bruscos dels preus en períodes curts de temps, on després el preu es mantenia estable durant uns mesos fins que tornava a variar bruscament el preu, evidenciant que la sèrie de preus del consumidor domèstic té intervencions per part de l'estat o d'una figura externa reguladora. Sota aquestes característiques, jutjant per les mesures del rendiment que ha tingut el model univariant, és evident que aquest no és capaç d'ajustar-se correctament i captar amb precisió la realitat de les dades. L'aplicació del model multivariant, en aquest cas, millora significativament les prediccions. La convergència dels preus dels altres

recursos energètics (en totes les seves modalitats) més la informació de les temperatures, fa que el model $VAR(12)$ s'adapti millor a la complexitat de la sèrie i que generi prediccions més precises.

Pel que fa a la sèrie temporal dels preus de l'electricitat dirigits a la indústria, els resultats de rendiment demostren un millor comportament en el model univariant $ARIMA(1,1,0)$, amb un RMSE menor i una taxa de cobertura molt elevada amb un 92% (mentre que el model multivariant té un 67% de cobertura). Cal destacar que el model multivariant, com es pot veure a la representació gràfica de les seves prediccions, capta els moviments els increments i decreixements dels preus del grup control mentre que les prediccions del model univariant es mantenen estàtiques en el temps.

En quant a la sèrie de preus de l'electricitat del mercat majorista, tant el model univariant $ARIMA(0,1,1)$ com el multivariant $VAR(12)$, ambdós tenen una taxa de cobertura del 100% de les observacions del grup control. Pel que fa el valor RMSE són molt similars, sent el del model univariant lleugerament menor (11,18 contra 11,81). Cal destacar que, malgrat que ambdós tenen uns rendiments adequats, les prediccions del model multivariant és més dinàmic i és capaç de captar més el moviment dels preus entre prediccions. El model univariant es manté estable en el temps donat que només té un ordre de mitjanes mòbils.

Es conclou que per la primera sèrie és beneficiós utilitzar un model multivariant a cause de les seves característiques particulars. Per les altres dues sèries, la decisió d'utilitzar models univariants o multivariants és més ambigua i l'elecció entre les metodologies a utilitzar pot dependre de consideracions addicionals, com recursos disponibles i necessitats específiques per les qual es vol predir aquesta tipologia de preus.

Capítol 6

VI. CONCLUSIONS

En conclusió, l'anàlisi de sèries temporals ha demostrat ser una eina molt interessant pel que fa a la predicció d'un conjunt de dades amb una informació limitada. Com ha sigut el cas d'aquest projecte on, a partir de les dades proporcionades per la *web* de la Unió Europea, només es podia obtenir informació molt agregada dels preus dels recursos energètics. S'ha de tenir en compte, però que per aquesta mateixa raó, les prediccions extretes de models de sèries temporals també tenen les seves limitacions.

El tractament previ al modelatge de les dades ha sigut un pas crucial. L'estudi exhaustiu de cada una de les sèries i posteriorment, el suavitzat de les observacions ha sigut un element molt necessari. Quelcom que tenien en comú totes les sèries temporals de la mostra seleccionada era la presència de canvis inesperats i abruptes dels preus. S'ha de tenir en compte que per la mateixa naturalesa de les dades estudiades és comprensible que passi, ja que els preus dels recursos energètics estan sotmesos a un entorn macroeconòmic i, amb més o menys mesura, entitats estatals i econòmiques cíclicament regularan aquests preus. El *pre-processing* i el suavitzat emprat en aquest projecte ha fet que aquestes característiques es poguessin estimar dins l'anàlisi de sèries temporals.

Amb l'aplicació de models univariants de sèries temporals, s'ha pogut arribar a un dels objectius d'aquest projecte: poder tenir una percepció del comportament dels preus de l'electricitat a futur per poder estimar l'impacte que pot tenir a l'economia de l'usuari. Si bé les mitjanes de les prediccions mensuals poden no ser molt dinàmiques en tot el període predit mitjançant models univariants, els intervals de confiança han resultat tenir una taxa de cobertura elevada. Sobretot en aquelles tipologies de preus més volàtils com són els preus de les indústries i els del mercat majorista.

Les prediccions obtingudes del model multivariant, però permet al consumidor tenir la percepció de com l'electricitat pot impactar a la seva economia amb més exactitud. Les mitjanes mensuals predites capturen millor els augments i els decreixements dels preus entre observacions. A més l'exploració de les sèries d'electricitat amb un enfocament multivariant també ha permès

examinar la relació entre els preus de l'electricitat i els de combustibles fòssils, així com les temperatures ambientals. Els resultats assolits confirmen una de les hipòtesis inicials que s'havien plantejat a l'inici del projecte: existeix una connexió significativa entre aquestes sèries i les fluctuacions dels preus de l'electricitat.

Es conclou que no hi ha una metodologia (univariant o multivariant) clarament millor que l'altra, si més no serà convenient utilitzar mètodes multivariants per aquelles sèries que es sospiti que són més susceptibles a factors externs o que siguin intervingudes amb més mesura; com és el cas de la sèrie de preus de l'electricitat de la llar.

Amb aquest projecte es dona evidències de què és plausible poder fer prediccions dels preus de l'electricitat amb sèries temporals, que siguin d'utilitat per a l'usuari. Aquest pot tenir una idea de com aquests preus poden impactar a la seva economia en un futur pròxim i poder planificar i prendre decisions informades en relació amb el seu consum energètic. Tanmateix, es planteja de què a futures investigacions s'incorpori informació sobre el context regulador dels preus estudiats, el que implicaria una recopilació addicional d'informació de polítiques, impostos o canvis en les tarifes que afectin directament als preus. Això permetria tenir unes prediccions més precises, ja que es tindria una comprensió més completa del mercat energètic.

VII. BIBLIOGRAFIA

- 1) Àrea d'estadística del Departament de Matemàtiques de la ULPGC. *Objetos en R: Series temporales* (2016). URL: <https://estadistica-dma.ulpgc.es/cursoR4ULPGC/14-seriesTemporales.html> (visitat el 20/5/2023)
- 2) Athanasopoulos, George, and Farshif Vahid. "VARMA versus VAR for Macroeconomic Forecasting." *Journal of Business & Economic Statistics*, 26 (2), 2008, pp. 237-52
- 3) Bundesnetzagentur. *Entstehung*. URL: <https://www.bundesnetzagentur.de/DE/Allgemeines/DieBundesnetzagentur/Ueberuns/sta.html> (visitat el 30/5/2023)
- 4) Dictetorate-General for Energy, European Comission. *Energy prices and costs in Europe. 2020*. URL: https://energy.ec.europa.eu/data-and-analysis/energy-prices-and-costs-europe_en (visitat el 10/1/2023)
- 5) Dictetorate-General for Energy, European Comission. "Study on energy prices, costs and their impact on industry and households." ISBN: 978-92-76-18954-1. DOI: 10.2833/49063
- 6) George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung. "Time Series Analysis: Forecasting and Control." ISBN: 978-1-118-67502-1
- 7) Georgia State University Library, Institute of Insight. *Web Scraping (Research Guides)*. (2023). URL: <https://research.library.gsu.edu/c.php?g=1050939&p=7632068>
- 8) Horsley, S. "Soaring electricity bills are the latest inflation flashpoint." NPR. URL: <https://www.npr.org/2022/09/13/1122371879/electricity-utilities-gasoline-gas-prices-inflation-august-cpi-consumer-prices> (visitat el 3/4/2023)
- 9) Johannes Stroebel, Joseph Vavra, NYU Stern, CEPR Chicago Booth, NBER. "House Prices, Local Demand, and Retail Prices." *Journal of Political Economy*. 127(3). 2019. DOI: 10.1086/701422
- 10) John Elder and Peter E. Kennedy. "Testing for Unit Roots: What Should Students Be Taught?" *The Journal of Economic Education*. 32 (2). 2001, pp. 137-146. URL: <http://www.jstor.org/stable/1183489>
- 11) Krotov, Vlad & Silva, Leiser. "Legality and Ethics of Web Scraping." *Twenty-fourth America's Conference on Information Systems, New Orleans*. (2018)
- 12) Pfaff, B. "VAR, SVAR and SVEC Models: Implementation Within R Package vars." *Journal of Statistical Software*. 27 (4), pp. 1-32. DOI: 10.18637. URL: <https://doi.org/10.18637/jss.v027.i04>

- 13) RDocumentation, Bernhard Pfaff. *VAR: Estimation of a VAR(p)*. URL: <https://www.rdocumentation.org/packages/vars/versions/1.5-9/topics/VAR> (visitat el 23/5/2023)
- 14) R.H. Shumway, D.S. Stoffer. "An approach to time series smoothing and forecasting using the EM algorithm." *Journal of Time Series Analysis*, 3 (4). 1982. DOI: 0143-9782/04 0253-12
- 15) The MathWorks. *Modelo de espacio de estados*. 2006. URL: <https://la.mathworks.com/help/control/ref/ss.html#d124e28102> (visitat el 15/6/2023)
- 16) V.A. Profillidis, G.N. Botzoris. *Modeling of transport Demand* (2019). URL: <https://www.sciencedirect.com/topics/social-sciences/granger-causality-test> (visitat el 9/5/2023)
- 17) George Müller, Rogier Floors. *Wetterzentrale*. URL: https://www.wetterzentrale.de/averages_de.php?station=691&jaar=2008&maand=1&dag=1&jaar2=2008&maand2=12&dag2=31
- 18) Xinghua Chang, Meng Gao, Yan Wang, Xiyong Hou. "Seasonal Autoregressive Integrated Moving Average Model for Precipitation Time Series." *Journal of Mathematics and Statistics*. 8 (4): 500-505. 2012. ISSN 1549-3644. DOI: 10.3844/jmssp.2012.500.505

VIII. ANNEX A

Durant l'elaboració d'aquest treball, se'm va brindar l'oportunitat de participar a un projecte de divulgació estadística a alumnes d'ESO i Batxillerat com a activitat complementària. El projecte s'anomena *Stat Wars* i, en un format amè i didàctic, pretén apropar l'estadística als estudiants del territori espanyol.

<https://www.proyectostatwars.es/>

Es presenta en format vídeo i els temes a tractar poden ser diversos, en el meu cas vaig presentar l'aportació estadística de la infermera, matemàtica i estadística Florence Nightingale.

A continuació es deixa el material que va ser presentat:

Florence Nightingale nació en Florencia el 12 de mayo de 1820, en el seno de una familia inglesa acomodada. Recibió una educación muy privilegiada y muy completa. Des de pequeña se sentía interesada por las matemáticas.

En el 1854 empezó como enfermera en hospitales de campaña en la Guerra de Crimea. Durante este tiempo recopiló información y datos sobre la mortalidad de los soldados, pues a menudo las causas de defunción eran por las malas circunstancias de higiene y atención médica que recibían.

Abogó por la educación en la salud y la prevención de enfermedades, utilizando la estadística descriptiva para mostrar a la sociedad que camino se debía tomar para mejorar la sanidad.

Murió el 13 de agosto de 1910, considerada Fundadora de la enfermería moderna y pionera en la estadística de la salud.

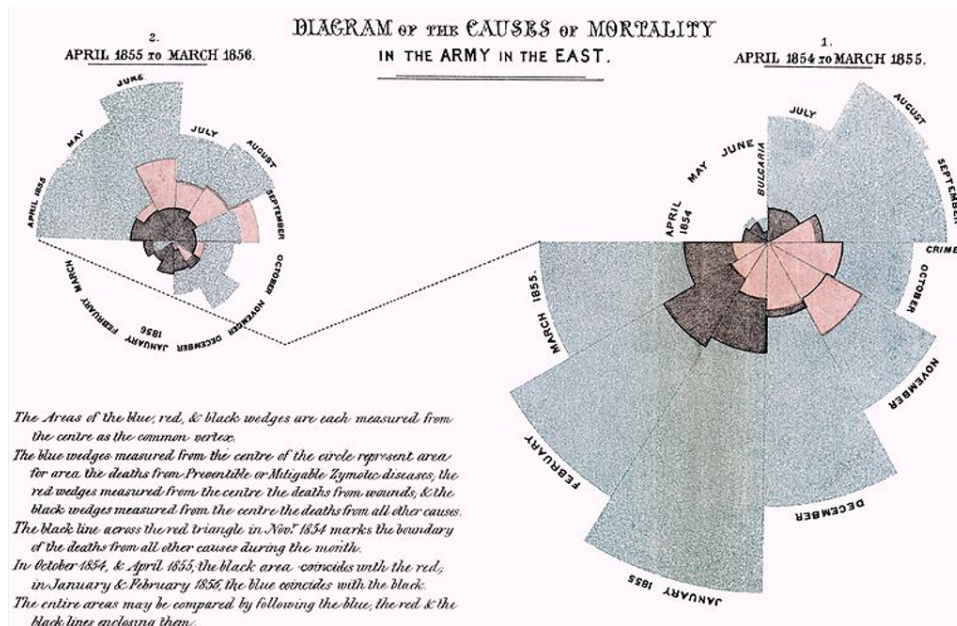


Figura Anexo A-1 Diagrama de área polar de Florence Nightingale

Una de sus aportaciones estadísticas más importantes fue el diagrama de área polar, conocido como diagrama de rosas o diagrama de Nightingale.

Es un gráfico que permite visualizar las causas de la mortalidad durante la Guerra de Crimea.

El gráfico se divide en doce secciones, cada una correspondiente a un mes del año. La anchura de cada sección representa la duración del tiempo y el área total del gráfico representa el número total de muertes.

El gráfico tiene además dos partes: una parte coloreada en rojo (representa las muertes causadas por heridas de guerra) y otra en azul (representa las muertes causadas por enfermedades). De este modo el gráfico permite ver cómo las enfermedades eran la principal causa de muerte entre los soldados, en vez de las heridas de guerra.

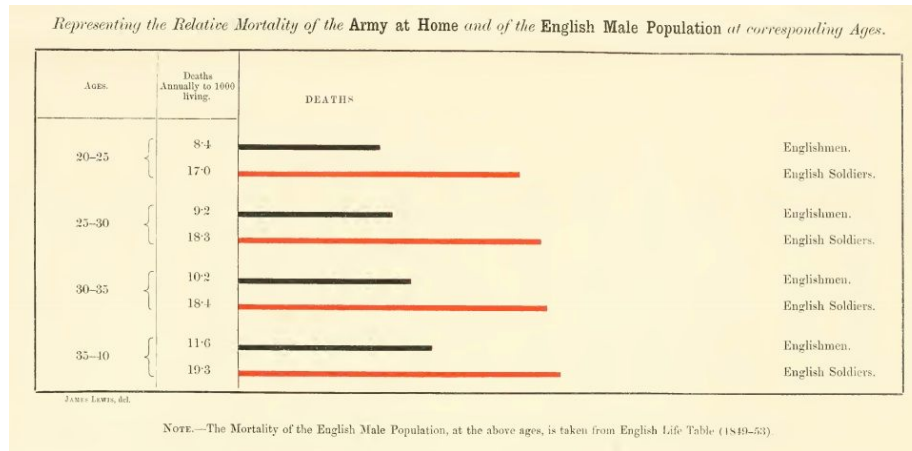


Figura Anexo A-2 Representación de la mortalidad relativa entre soldados

También hizo otras aportaciones dirigidas al entendimiento de la salud de los soldados cuando estos volvían a sus hogares (por edad). En las barras negras se ve la mortalidad de la población inglesa del siglo XIX mientras que la roja es la de los soldados.

El gráfico permite ver la gran diferencia entre la mortalidad de los soldados, y además, qué efecto tiene en ella la edad.

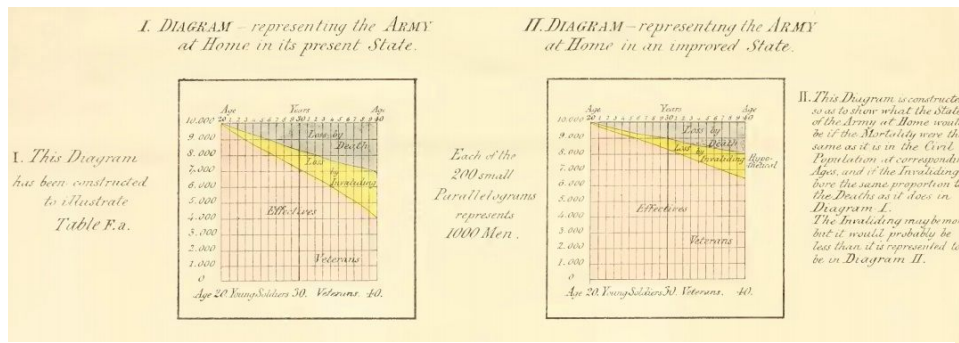


Figura Anexo A-3 Representación de la efectividad de las tropas

Mediante diagramas de área, Florence Nightingale, también pudo medir la pérdida de la efectividad de las tropas inglesas a causa de las malas condiciones sanitarias. En el primer diagrama se mide el ratio de pérdidas mortales y por invalidez, según el sistema sanitario del momento. En el segundo diagrama se hace una representación de lo que serían las pérdidas en un sistema de sanidad más efectivo, que claramente resulta ser mejor.

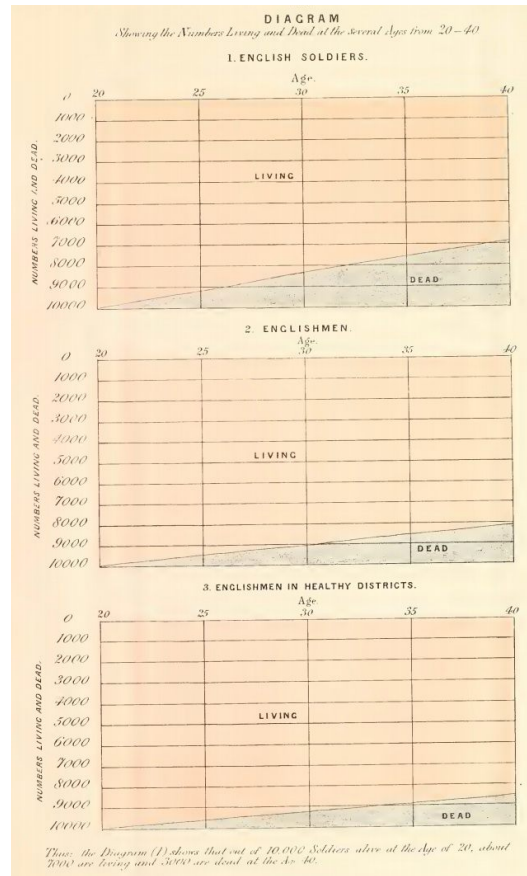


Figura Anexo A-4 Diagrama de la mortalidad de los soldados en función de su servicio

Y por último, otra forma de demostrar la mortalidad de los soldados una vez que han terminado su servicio es peor que a la del resto de la población; hizo 3 diagramas donde se trazan las tasas de mortalidad de 3 objetos de estudio.

- La de los soldados.
- La de la población inglesa general.
- Y la de la población inglesa de distritos sanos.

El trabajo de Florence Nightingale en la Guerra de Crimea conllevó una mejora significativa en la atención médica, estableciendo estándares que aún en la actualidad se utilizan en la enfermería moderna.

Sus análisis y visualizaciones de datos sobre la mortalidad en la guerra, han sido reconocidos como hitos en la visualización de datos y en la aplicación de la estadística a la medicina y la salud pública.

IX. ANNEX B

```
library(corrplot)
library(vars)
library(MTS)
library(astsa)
library(ggplot2)
library(lubridate)
library(rjson)
library(stringr)
library(dplyr)
library(tseries)
library(forecast)
library(MASS)
library(zoo)
```

Scraping (Python)

```
import requests
import re
import json
import time
```

```
data = {}
```

```
for year in range(2008, 2020):
    response = requests.get(
```

```
#f"http://www.wetterzentrale.de/averages_de.php?jaar={year}&maand=1&dag=1&jaar2={year}
&maand2=12&dag2=31&station=691&country=1") #bremen
```

```
f"http://www.wetterzentrale.de/averages_de.php?jaar={year}&maand=1&dag=1&jaar2={year}
&maand2=12&dag2=31&station=257&country=3") #wijk aan zee
```

```
text = response.text
```

```
parsed_json = re.search("var jdata = (. *?);", text).group(1)
```

```
year_data = json.loads(parsed_json)
data[year] = year_data
print(f"Saved data for year {year}. len: {len(year_data)}")
time.sleep(5)
```

```
with open("data_ne.json","w") as output:
    output.write(json.dumps(data))
```

Tractament de les dades (R)

```
ele.house <- read.csv(file = 'electricity-household-re.csv', header = T, sep = ";")
### DATASET
#####
## Electricity
# Household
ele.house$Germany <- as.numeric(gsub(",", ". ", gsub("\\.", "", ele.house$Germany)))
as.Date(ele.house$Category, format='%Y-%m')
ele.house$Category[nchar(ele.house$Category) == 6] <-
paste0(substr(ele.house$Category[nchar(ele.house$Category) ==
6],1,5),0,substr(ele.house$Category[nchar(ele.house$Category) == 6],6,6))
ele.house$Category <- paste0(ele.house$Category,'-01')
ele.house$Category <- as.Date(ele.house$Category, format='%Y-%m-%d')

# Industry
ele.indus <- read.csv(file = 'electricity-industry-ret.csv', header = T, sep = ";")
ele.indus$Germany <- as.numeric(gsub(",", ". ", gsub("\\.", "", ele.indus$Germany)))
ele.indus$Category[nchar(ele.indus$Category) == 6] <-
paste0(substr(ele.indus$Category[nchar(ele.indus$Category) ==
6],1,5),0,substr(ele.indus$Category[nchar(ele.indus$Category) == 6],6,6))
ele.indus$Category <- paste0(ele.indus$Category,'-01')
ele.indus$Category <- as.Date(ele.indus$Category, format='%Y-%m-%d')

# Wholesale
ele.sales <- read.csv(file = 'electricity-wholesale-pr.csv', header = T, sep = ";")
ele.sales$Germany <- as.numeric(gsub(",", ". ", gsub("\\.", "", ele.sales$Germany)))
ele.sales$Category[nchar(ele.sales$Category) == 6] <-
paste0(substr(ele.sales$Category[nchar(ele.sales$Category) ==
6],1,5),0,substr(ele.sales$Category[nchar(ele.sales$Category) == 6],6,6))
ele.sales$Category <- paste0(ele.sales$Category,'-01')
```

```
ele.sales$Category <- as.Date(ele.sales$Category, format='%Y-%m-%d')
```

```
## Natural Gas
```

```
# Household
```

```
gas.house <- read.csv(file = 'natural-gas-household-re.csv', header = T, sep = ";")
```

```
names(gas.house) <- c("Category", "Germany")
```

```
gas.house$Germany <- as.numeric(gsub(",", ".", gsub("\\.", "", gas.house$Germany)))
```

```
gas.house$Category[nchar(gas.house$Category) == 6] <-
```

```
paste0(substr(gas.house$Category[nchar(gas.house$Category) ==
```

```
6],1,5),0,substr(gas.house$Category[nchar(gas.house$Category) == 6],6,6))
```

```
gas.house$Category <- paste0(gas.house$Category, '-01')
```

```
gas.house$Category <- as.Date(gas.house$Category, format='%Y-%m-%d')
```

```
# Industry
```

```
gas.indus <- read.csv(file = 'natural-gas-industry-ret.csv', header = T, sep = ";")
```

```
gas.indus$Germany <- as.numeric(gsub(",", ".", gsub("\\.", "", gas.indus$Germany)))
```

```
gas.indus$Category[nchar(gas.indus$Category) == 6] <-
```

```
paste0(substr(gas.indus$Category[nchar(gas.indus$Category) ==
```

```
6],1,5),0,substr(gas.indus$Category[nchar(gas.indus$Category) == 6],6,6))
```

```
gas.indus$Category <- paste0(gas.indus$Category, '-01')
```

```
gas.indus$Category <- as.Date(gas.indus$Category, format='%Y-%m-%d')
```

```
# Wholesale
```

```
gas.sales <- read.csv(file = 'natural-gas-wholesale-pr.csv', header = T, sep = ";")
```

```
gas.sales$Germany <- as.numeric(gsub(",", ".", gsub("\\.", "", gas.sales$Germany)))
```

```
gas.sales$Category[nchar(gas.sales$Category) == 6] <-
```

```
paste0(substr(gas.sales$Category[nchar(gas.sales$Category) ==
```

```
6],1,5),0,substr(gas.sales$Category[nchar(gas.sales$Category) == 6],6,6))
```

```
gas.sales$Category <- paste0(gas.sales$Category, '-01')
```

```
gas.sales$Category <- as.Date(gas.sales$Category, format='%Y-%m-%d')
```

```
## Petroleum products
```

```
# Crude Oil
```

```
pet.croil <- read.csv(file = 'crude-oil-prices-eur2018.csv', header = T, sep = ";")
```

```
names(pet.croil) <- c("Category", "Average", "Brent", "Dubai", "WTI", "World")
```

```

pet.croil$Average <- as.numeric(gsub(",", ".", gsub("\\.", "", pet.croil$Average)))
pet.croil$Brent <- as.numeric(gsub(",", ".", gsub("\\.", "", pet.croil$Brent)))
pet.croil$Dubai <- as.numeric(gsub(",", ".", gsub("\\.", "", pet.croil$Dubai)))
pet.croil$WTI <- as.numeric(gsub(",", ".", gsub("\\.", "", pet.croil$WTI)))
pet.croil$World <- as.numeric(gsub(",", ".", gsub("\\.", "", pet.croil$World)))
pet.croil$Category[nchar(pet.croil$Category) == 6] <-
paste0(substr(pet.croil$Category[nchar(pet.croil$Category) ==
6],1,5),0,substr(pet.croil$Category[nchar(pet.croil$Category) == 6],6,6))
pet.croil$Category <- paste0(pet.croil$Category,'-01')
pet.croil$Category <- as.Date(pet.croil$Category, format='%Y-%m-%d')

# Diesel
pet.diese <- read.csv(file = 'diesel-retail-prices-wit.csv', header = T, sep = ";")
pet.diese$Germany <- as.numeric(gsub(",", ".", gsub("\\.", "", pet.diese$Germany)))
pet.diese$Category[nchar(pet.diese$Category) == 6] <-
paste0(substr(pet.diese$Category[nchar(pet.diese$Category) ==
6],1,5),0,substr(pet.diese$Category[nchar(pet.diese$Category) == 6],6,6))
pet.diese$Category <- paste0(pet.diese$Category,'-01')
pet.diese$Category <- as.Date(pet.diese$Category, format='%Y-%m-%d')

# Heating Oil
pet.heoil<- read.csv(file = 'heating-oil-retail-with.csv', header = T, sep = ";")
pet.heoil$Germany <- as.numeric(gsub(",", ".", gsub("\\.", "", pet.heoil$Germany)))
pet.heoil$Category[nchar(pet.heoil$Category) == 6] <-
paste0(substr(pet.heoil$Category[nchar(pet.heoil$Category) ==
6],1,5),0,substr(pet.heoil$Category[nchar(pet.heoil$Category) == 6],6,6))
pet.heoil$Category <- paste0(pet.heoil$Category,'-01')
pet.heoil$Category <- as.Date(pet.heoil$Category, format='%Y-%m-%d')

# Petrol
pet.petro<- read.csv(file = 'petrol-retail-prices-wit.csv', header = T, sep = ";")
pet.petro$Germany <- as.numeric(gsub(",", ".", gsub("\\.", "", pet.petro$Germany)))
pet.petro$Category[nchar(pet.petro$Category) == 6] <-
paste0(substr(pet.petro$Category[nchar(pet.petro$Category) ==
6],1,5),0,substr(pet.petro$Category[nchar(pet.petro$Category) == 6],6,6))

```

```

pet.petro$Category <- paste0(pet.petro$Category, '-01')
pet.petro$Category <- as.Date(pet.petro$Category, format='%Y-%m-%d')

## Temperatures
temp<- fromJSON(file = 'data.json')

plot(temp$"2008" $"Lowest Min. Temp.")

matplot(temp$"2008" $"m", temp$"2008" $"Max. Temp", type='o', pch=20, lty=1,
        ylab="Price Euro 2008", xlab="Year"#, main='Valors no nulls de Crude Oil World'
        )
matplot(temp$"2008" $"m", temp$"2008" $"Min. Temp", type='o', pch=20, lty=1,
        ylab="Price Euro 2008", xlab="Year"#, main='Valors no nulls de Crude Oil World'
        )
matplot(temp$"2008" $"m", temp$"2008" $"Gem. Temp", type='o', pch=20, lty=1,
        ylab="Price Euro 2008", xlab="Year"#, main='Valors no nulls de Crude Oil World'
        )

df_temp <- NULL

for(i in 1:length(names(temp))){
  Normal = temp[[i]]$"Gem. Temp"
  High = temp[[i]]$"Max. Temp"
  Low = temp[[i]]$"Min. Temp"
  m = temp[[i]]$m+1

  names(Normal) = NULL
  names(High) = NULL
  names(Low) = NULL
  names(m) = NULL

  Normal = unlist(Normal)
  High = unlist(High)
  Low = unlist(Low)
  #d = unlist(d)
}

```

```

m = unlist(m)

x = data.frame(date=paste0(rep(as.numeric(names(temp[i])), length(Normal)), "-
",str_pad(m,2,side="left", pad=0),"-01"), Normal=Normal, High=High, Low=Low)
x$date = as.Date(x$date, format='%Y-%m-%d')

df_temp = rbind(df_temp, x)

}

matplot(df_temp$date, df_temp[,c("Normal","High","Low")], type='o', pch=20, lty = 1
, ylab="Temperatura C", xlab="Year", main='Temperatures Alemanya')
legend("topright", legend=c("Normal","High","Low"), col=c(1:10),pch=15:19,
bg= ("white"), horiz=F)
df_temp_m = df_temp %>% group_by(date) %>%
summarise(x_Normal = mean(Normal),
x_High = mean(High),
x_Low = mean(Low),
mx_High = max(High),
mm_High = min(Low),
.groups = 'drop') %>%
as.data.frame()

matplot(df_temp_m$date, df_temp_m[,c("x_Normal","x_High","x_Low")], type='o', pch=20, lty
= 1
, ylab="Temperatura C", xlab="Year", main='Temperatures Alemanya')
legend("bottomright", legend=c("x_Normal","x_High","x_Low"), col=c(1:10),pch=15:19,
bg= ("white"), horiz=F)

#### PREPROCESSING
#####
## Electricity

```

```
summary(ele.house)
summary(ele.indus)
summary(ele.sales)
```

```
nrow(ele.house);nrow(ele.indus);nrow(ele.sales)
```

```
ele.house$Category[ele.house$Category != ele.indus$Category]
ele.house$Category[ele.house$Category != ele.sales$Category] #ok
```

```
ELE <- data.frame(date = ele.house$Category
, house.AVG = ele.house$Germany
, indus.AVG = ele.indus$Germany
, sales.AVG = ele.sales$Germany)
```

```
summary(ELE[,c(1,4)])
```

```
names(ELE)
```

```
par(mfrow = c(1, 1))
```

```
matplot(ELE$date, ELE[,substr(names(ELE),1,5)=="house"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Electricity - Household')
```

```
matplot(ELE$date, ELE[,substr(names(ELE),1,5)=="indus"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Electricity - Industry')
```

```
matplot(ELE$date, ELE[,substr(names(ELE),1,5)=="sales"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Electricity - Sales')
```

```
## Natural Gas
```

```
summary(gas.house)
summary(gas.indus)
summary(gas.sales)
```

```
nrow(gas.house);nrow(gas.indus);nrow(gas.sales)
```

```
gas.house$Category[gas.house$Category != gas.indus$Category]
gas.house$Category[gas.house$Category != gas.sales$Category] #ok
```


gas.indus\$Category[is.na(gas.indus\$Germany)] #hi ha un NA, grafiquem i veiem com el tractem

```
GAS <- data.frame(date = gas.house$Category
, house.AVG = gas.house$Germany
, indus.AVG = gas.indus$Germany
, sales.AVG = gas.sales$Germany)
```

```
par(mfrow = c(1, 1))
```

```
matplot(GAS$date, GAS[,substr(names(GAS),1,5)=="house"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Natural Gas - Household')
```

```
matplot(GAS$date, GAS[,substr(names(GAS),1,5)=="indus"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Natural Gas - Industry')
```

```
matplot(GAS$date, GAS[,substr(names(GAS),1,5)=="sales"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Natural Gas - Sales')
```

```
## Petroleum
```

```
summary(pet.croil)
summary(pet.diese)
summary(pet.heoil)
summary(pet.petro)
```

```
nrow(pet.croil);nrow(pet.diese);nrow(pet.heoil);nrow(pet.petro)
```

```
pet.croil$Category[pet.croil$Category != pet.diese$Category]
pet.croil$Category[pet.croil$Category != pet.heoil$Category]
pet.croil$Category[pet.croil$Category != pet.petro$Category] #ok
```

```
PET <- data.frame(date = pet.croil$Category
, croil.AVG = pet.croil$Average, croil.BRENT = pet.croil$Brent, croil.DUBAI =
pet.croil$Dubai, croil.WTI = pet.croil$WTI, croil.World = pet.croil$World
, diese.AVG = pet.diese$Germany
, heoil.AVG = pet.heoil$Germany
, petro.AVG = pet.petro$Germany)
names(PET)
```

```

df_PET <-
PET[,c("croil.AVG", "croil.BRENT", "croil.DUBAI", "croil.WTI", "croil.World", "diese.AVG", "heoil.AVG", "petro.AVG")]

par(mfrow = c(1, 1))
matplot(PET$date, PET[,substr(names(PET),1,5)=="croil"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Petroleum - Crude Oil')
legend("topright", legend=names(PET[,substr(names(PET),1,5)=="croil"]),
col=c(1:10),pch=15:19,
bg= ("white"), horiz=F)
matplot(PET$date, PET[,substr(names(PET),1,5)=="diese"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Natural Gas - Diesel')
matplot(PET$date, PET[,substr(names(PET),1,5)=="heoil"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Natural Gas - Heating Oil') # Heating Oil tiene
Outliers 2016 y 2017
matplot(PET$date, PET[,substr(names(PET),1,5)=="petro"], type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Natural Gas - Petrol')

summary(PET)

PET_World <- subset(PET, !is.na(croil.World))
par(mfrow = c(1, 1))
matplot(PET_World$date, PET_World$croil.World, type='o', pch=20, lty=1,
ylab="Price Euro 2008", xlab="Year", main='Valors no nulls de Crude Oil World')

## SELECCIONEM VARIABLES #####
names(ELE)
df_ELE <- ELE[,c("house.AVG", "indus.AVG", "sales.AVG")]
names(df_ELE) <- c("house", "indus", "sales")

names(GAS)
df_GAS <- GAS[,c("house.AVG", "indus.AVG", "sales.AVG")]
names(df_GAS) <- c("house", "indus", "sales")

names(PET)

```

```
df_PET <-
PET[,c("croil.AVG","croil.BRENT","croil.DUBAI","croil.WTI","croil.World","diese.AVG","heoil.AV
G","petro.AVG")]
names(df_PET) <- c("croil","brent","dubai","wti","world","diese","heoil","petro")
```

Suavitzat de Kalman (R)

```
y=matrix(ts.ele.house,nc=1)
# iniciem valors
A<-1
mu0<-0
Sigma0<-10000
Phi0<-0.1
cQ0<-0.1
cR0<-0.1

Linn<-function(p){
  Phi<-p[1]
  cQ<-p[2]
  cR<-p[3]
  kf<-Kfilter(y, A, mu0, Sigma0, Phi, cQ, cR)
  kf$like
}

##### Electricitat domestic #####
par(mfrow=c(1,1))
y=matrix(ts.ele.house,nc=1)
init_p<-(c(0.1,0.1,0.1))

mod_opt<-optim(init_p,Linn,NULL,method="CG",hessian=TRUE)
mod_opt$p
#method := ("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN","Brent")

Phi<-mod_opt$p[1]; cQ<-mod_opt$p[2]; cR<-mod_opt$p[3]
```

```

kf<-Ksmooth(y, A, mu0, Sigma0, Phi, cQ, cR)
kf$Xs
ts.ele.house.suav<-ts(c(unlist(kf$Xs)),start = c(2008,1), frequency = 12)

ts.ele.house.suav[1:4] <- ts.ele.house[1:4] - (ts.ele.house[5] - ts.ele.house.suav[5])
ts.ele.house.suav[140:144] <- ts.ele.house[140:144] - (ts.ele.house[5] - ts.ele.house.suav[5])
adf.test(ts.ele.house.suav)
ts.ele.house.suav.dif = diff(ts.ele.house.suav)

matplot(ELE$date, data.frame(as.vector(ts.ele.house), as.vector(ts.ele.house.suav)) , type='o',
pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Suavitzat de Electricitat Domèstic')
legend("topleft", legend=c("dades originals", "dades suavitzat"), col=c(1:10),pch=19:19,
bg= ("white"), horiz=F, cex = 0.8)

matplot(y = data.frame(dif = as.vector(ts.ele.house.dif), dif.suav =
as.vector(ts.ele.house.suav.dif) )
, type='o', pch=20, lty = 1
, ylab="Diferencials price Euro 2008", xlab="Year", main='Suavitzat de diferencials Electr.
Dom.')
```

```

legend("bottomleft", legend=c("diferencials originals", "diferencials suavitzat"),
col=c(1:10),pch=19:19,
bg= ("white"), horiz=F, cex = 0.8)

##### Electricitat industrial #####
y=matrix(ts.ele.indus,nc=1)
init_p<-(c(0.1,0.1,0.1))

mod_opt<-optim(init_p,Linn,NULL,method="CG",hessian=TRUE)
mod_opt$p
#method := ("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN","Brent")

Phi<-mod_opt$p[1]; cQ<-mod_opt$p[2]; cR<-mod_opt$p[3]

```

```

kf<-Ksmooth(y, A, mu0, Sigma0, Phi, cQ, cR)
kf$Xs
ts.ele.indus.suav<-ts(c(unlist(kf$Xs)),start = c(2008,1), frequency = 12)

ts.ele.indus.suav[1:4] <- ts.ele.indus[1:4] - (ts.ele.indus[5] - ts.ele.indus.suav[5])
ts.ele.indus.suav[140:144] <- ts.ele.indus[140:144] - (ts.ele.indus[5] - ts.ele.indus.suav[5])
adf.test(ts.ele.indus.suav)
ts.ele.indus.suav.dif = diff(ts.ele.indus.suav)

matplot(ELE$date, data.frame(as.vector(ts.ele.indus), as.vector(ts.ele.indus.suav)) , type='o',
pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Suavitzat de Electricitat Industrial')
legend("topleft", legend=c("dades originals", "dades suavitzat"), col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)

matplot(y = data.frame(dif = as.vector(ts.ele.indus.dif), dif.suav =
as.vector(ts.ele.indus.suav.dif) )
, type='o', pch=20, lty = 1
, ylab="Diferencials price Euro 2008", xlab="Year", main='Suavitzat de diferencials Electr.
Ind.')
```

```

legend("topright", legend=c("diferencials originals", "diferencials suavitzat"),
col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)

##### Electricitat majorista #####
y=matrix(as.vector(ts.ele.sales),nc=1)
init_p<-(c(0.1,0.1,0.1))

mod_opt<-optim(init_p,Linn,NULL,method="BFGS",hessian=TRUE)
mod_opt$p
#method := ("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN","Brent")

Phi<-mod_opt$p[1]; cQ<-mod_opt$p[2]; cR<-mod_opt$p[3]

```

```

kf<-Ksmooth(y, A, mu0, Sigma0, Phi, cQ, cR)
kf$Xs
ts.ele.sales.suav<-ts(c(unlist(kf$Xs)),start = c(2008,1), frequency = 12)
adf.test(ts.ele.sales.suav)
ts.ele.sales.suav.dif = diff(ts.ele.sales.suav)

matplot(ELE$date, data.frame(as.vector(ts.ele.sales), as.vector(ts.ele.sales.suav)) , type='o',
pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Suavitzat de Electricitat Majorista')
legend("topright", legend=c("dades originals", "dades suavitzat"), col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)
matplot(y = data.frame(dif = as.vector(ts.ele.indus.dif), dif.suav =
as.vector(ts.ele.indus.suav.dif) )
, type='o', pch=20, lty = 1
, ylab="Diferencials price Euro 2008", xlab="Year", main='Suavitzat de diferencials Electr.
Maj.')
```

```

legend("topright", legend=c("diferencials originals", "diferencials suavitzat"),
col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)

##### Gas domestic #####
y=matrix(as.vector(ts.gas.house),nc=1)
init_p<-(c(0.1,0.1,0.1))

mod_opt<-optim(init_p,Linn,NULL,method="BFGS",hessian=TRUE)
mod_opt$p
#method := ("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN", "Brent")

Phi<-mod_opt$p[1]; cQ<-mod_opt$p[2]; cR<-mod_opt$p[3]

kf<-Ksmooth(y, A, mu0, Sigma0, Phi, cQ, cR)
kf$Xs
```

```

ts.gas.house.suav<-ts(c(unlist(kf$Xs)),start = c(2008,1), frequency = 12)
adf.test(ts.gas.house.suav)
ts.gas.house.suav[1:5] <- ts.gas.house[1:5] - (ts.gas.house[6] - ts.gas.house.suav[6])

ts.gas.house.suav.dif = diff(ts.gas.house.suav)

matplot(ELE$date, data.frame(as.vector(ts.gas.house), as.vector(ts.gas.house.suav)) , type='o',
pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Gas domestic')
legend("topright", legend=c("dades originals", "dades suavitzat"), col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)

matplot(y = data.frame(dif = as.vector(ts.gas.house.dif), dif.suav =
as.vector(ts.gas.house.suav.dif) )
, type='o', pch=20, lty = 1
, ylab="Diferencials price Euro 2008", xlab="Year", main='Diferencials electricity Industrial')
legend("topright", legend=c("diferencials originals", "diferencials suavitzat"),
col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)

##### Gas industrial #####
y=matrix(as.vector(ts.gas.indus),nc=1)
init_p<-(c(0.1,0.1,0.1))

mod_opt<-optim(init_p,Linn,NULL,method="Nelder-Mead",hessian=TRUE)
mod_opt$p
#method := ("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN","Brent")

Phi<-mod_opt$p[1]; cQ<-mod_opt$p[2]; cR<-mod_opt$p[3]

kf<-Ksmooth(y, A, mu0, Sigma0, Phi, cQ, cR)
kf$Xs
ts.gas.indus.suav<-ts(c(unlist(kf$Xs)),start = c(2008,1), frequency = 12)
adf.test(ts.gas.indus.suav)

```

```

ts.gas.indus.suav.dif = diff(ts.gas.indus.suav)

matplot(ELE$date, data.frame(as.vector(ts.gas.indus), as.vector(ts.gas.indus.suav)) , type='o',
pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Gas industrial')
legend("topright", legend=c("dades originals", "dades suavitzat"), col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)

matplot(y = data.frame(dif = as.vector(ts.gas.house.dif), dif.suav =
as.vector(ts.gas.house.suav.dif) )
, type='o', pch=20, lty = 1
, ylab="Diferencials price Euro 2008", xlab="Year", main='Diferencials electricity Industrial')
legend("topright", legend=c("diferencials originals", "diferencials suavitzat"),
col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)

##### Combustibles fòssils Crude Oil #####
y=matrix(ts.pet.wti,nc=1)
init_p<-(c(0.1,0.1,0.1))

mod_opt<-optim(init_p,Linn,NULL,method="CG",hessian=TRUE)
mod_opt$p
#method := ("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN","Brent")

Phi<-mod_opt$p[1]; cQ<-mod_opt$p[2]; cR<-mod_opt$p[3]

kf<-Ksmooth(y, A, mu0, Sigma0, Phi, cQ, cR)
kf$Xs
ts.pet.wti.suav<-ts(c(unlist(kf$Xs)),start = c(2008,1), frequency = 12)

adf.test(ts.pet.wti.suav)
ts.pet.wti.suav.dif = diff(ts.pet.wti.suav)

```



```
matplot(ELE$date, data.frame(as.vector(ts.pet.wti), as.vector(ts.pet.wti.suav)) , type='o',
pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Combustible fòssil Cru')
legend("topright", legend=c("dades originals", "dades suavitzat"), col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)
```

```
matplot(y = data.frame(dif = as.vector(ts.pet.wti.dif), dif.suav = as.vector(ts.pet.wti.suav.dif) )
, type='o', pch=20, lty = 1
, ylab="Diferencials price Euro 2008", xlab="Year", main='Diferencials Combustible fòssil
Cru')
legend("bottomright", legend=c("diferencials originals", "diferencials suavitzat"),
col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)
```

Sèries temporals univariades (R)

```
(ts.ele.house = ts(data = df_ELE$house, start = c(2008,1), frequency = 12))
(ts.ele.indus = ts(data = df_ELE$indus, start = c(2008,1), frequency = 12))
(ts.ele.sales = ts(data = df_ELE$sales, start = c(2008,1), frequency = 12))
```

```
(ts.gas.house = ts(data = df_GAS$house, start = c(2008,1), frequency = 12))
(ts.gas.indus = ts(data = df_GAS$indus, start = c(2008,1), frequency = 12))
(ts.gas.sales = ts(data = df_GAS$sales, start = c(2008,1), frequency = 12))
```

```
(ts.pet.croil = ts(data = df_PET$croil, start = c(2008,1), frequency = 12))
(ts.pet.brent = ts(data = df_PET$brent, start = c(2008,1), frequency = 12))
(ts.pet.dubai = ts(data = df_PET$dubai, start = c(2008,1), frequency = 12))
(ts.pet.wti = ts(data = df_PET$wti, start = c(2008,1), frequency = 12))
(ts.pet.world = ts(data = df_PET$world, start = c(2008,1), frequency = 12))
(ts.pet.diese = ts(data = df_PET$diese, start = c(2008,1), frequency = 12))
(ts.pet.heoil = ts(data = df_PET$heoil, start = c(2008,1), frequency = 12))
(ts.pet.petro = ts(data = df_PET$petro, start = c(2008,1), frequency = 12))
```

```
#ELECTRICITAT
```

```
plot(ts.ele.house)
```

```
summary(ts.ele.house)
```

```
plot(ts.ele.indus)
```

```
summary(ts.ele.indus)
```

```
plot(ts.ele.sales)
```

```
summary(ts.ele.sales)
```

```
#GAS
```

```
plot(ts.gas.house)
```

```
summary(ts.gas.house)
```

```
plot(ts.gas.indus)
```

```
summary(ts.gas.indus)
```

```
plot(ts.gas.sales)
```

```
summary(ts.gas.sales)
```

```
#COMBUSTIBLES FÒSSILS
```

```
plot(ts.pet.croil)
```

```
summary(ts.pet.croil)
```

```
#### AMB VARIABLES SUAVITZADES
```

```
#####
```

```
colnames(multi_ts.suav) #treballat a script de series MULTIVARIADES
```

```
#data.suav <- data.frame(croil.WTI = ts.pet.wti.suav
```

```
#           , petro.AVG = PET_multi[,"petro.AVG"]
```

```
#           , gas.house = ts.gas.house.suav
```

```
#           , gas.indus = ts.gas.indus.suav
```

```
#           , ele.house = ts.ele.house.suav
```

```
#           , ele.indus = ts.ele.indus.suav
```

```

#           , ele.sales = ts.ele.sales.suav
#           , x_High = df_temp_m["x_Normal"]
#)

##### ELECTRICITAT #####
n <- length(ELE$date[ELE$date<'2019-01-01'])
N <- nrow(ELE)
summary(ts.ele.house)

train.ele.house = ts(data = ts.ele.house.suav[1:n], start = c(2008,1), frequency = 12)
test.ele.house = ts(data = ts.ele.house.suav[(n+1):N], start = c(2008,1), frequency = 12)

class(train.ele.house)
## Electricitat domèstic
train.ele.house
plot(train.ele.house)
hist(train.ele.house, breaks = "Sturges", main="")
adf.test(train.ele.house) # No es rebutja hipòtesis nula de no estacionarietat = no és
estacionària

train.ele.house.dif <- diff(train.ele.house)
train.ele.house.dif

adf.test(train.ele.house.dif)

plot(train.ele.house.dif)

# descomposem la sèrie:
plot(decompose(train.ele.house.dif))
# Autocorrelacions:
acf(train.ele.house.dif, lag.max = 20, plot = T, main= "")
pacf(train.ele.house.dif, lag.max =20, plot = T, main= "")
tsdisplay(train.ele.house.dif)

```

```
# ARIMA(1,1,0)
ARIMA110.ele.house = arima(train.ele.house, order = c(1,1,0))
AIC(ARIMA110.ele.house) # 446.4922
```

```
summary(ARIMA110.ele.house)
coefstest(ARIMA110.ele.house)
confint(ARIMA110.ele.house)
```

```
# ARIMA(0,1,1)
ARIMA011.ele.house = arima(train.ele.house, order = c(0,1,1))
AIC(ARIMA011.ele.house) # 458.2939 # augmenta
```

```
summary(ARIMA011.ele.house)
coefstest(ARIMA011.ele.house)
confint(ARIMA111.ele.house)
```

```
# ARIMA(1,1,1)
ARIMA111.ele.house = arima(train.ele.house, order = c(1,1,1))
AIC(ARIMA111.ele.house) # 448.4717 # augmenta
```

```
summary(ARIMA111.ele.house)
coefstest(ARIMA111.ele.house) # MA(1) no significatiu AIC aumgenta, es descarta
confint(ARIMA111.ele.house)
```

#Estacionalitat?

```
# ARIMA(1,1,0)(1,1,0)[12]
ARIMA110_110.ele.house = arima(train.ele.house, order = c(1,1,0), seasonal = list(order =
c(1,1,0), period = 12))
AIC(ARIMA110_110.ele.house) # 458.3291 # AIC major
```

```
summary(ARIMA110_110.ele.house)
coefstest(ARIMA110_110.ele.house)
confint(ARIMA110_110.ele.house)
```

```

# ARIMA(1,1,0)(0,1,1)[12] ##### s'escull aquesta
ARIMA110_011.ele.house = arima(train.ele.house, order = c(1,1,0), seasonal = list(order =
c(0,1,1), period = 12))
AIC(ARIMA110_011.ele.house) # 435.8136

ARIMA110_011.ele.house$intercept

summary(ARIMA110_011.ele.house)
coeftest(ARIMA110_011.ele.house)
confint(ARIMA110_011.ele.house)

coef(ARIMA110_011.ele.house)

# Anàlisi dels residus
ARIMA110_011.ele.house
res.ele.house_110_011 <- residuals(ARIMA110_011.ele.house)

res.ele.house_110_011
adf.test(res.ele.house_110_011)# estacionari
acf(res.ele.house_110_011)
pacf(res.ele.house_110_011)
tsdisplay(res.ele.house_110_011)

qqnorm(res.ele.house_110_011, main = "Residus ARIMA(1,1,0)(0,1,1)[12] Electr. Dom.", ylab =
"Residus")
res.ele.house_110_011

(n_ahead = length(test.ele.house))

pred.ele.house <- forecast::forecast(ARIMA110_011.ele.house,h = n_ahead)
pred.ele.house$mean
Apred.ele.house <- as.data.frame(pred.ele.house)

```

```

autoplot(pred.ele.house)
pred.ele.house$upper[,2]

Aval.ele.house <- data.frame(test=as.vector(test.ele.house), pred=pred.ele.house$mean,
Lo95=pred.ele.house$lower[,2], Up95=pred.ele.house$upper[,2])

val.ele.house <- data.frame(test=as.vector(test.ele.house), pred=pred.ele.house$mean)

matplot(ELE$date[(n+1):N], Aval.ele.house, type=c('o','o','l','l'), pch=20, lty = 1,
col=c('black','red', 'blue','blue')
, ylab="Price Euro 2008", xlab="Year", main='ARIMA(1,1,0)(0,1,1)[12] Electr. Dom.')
legend("topleft", legend=c("control", "prediccions", "intervals 95%"), col=c('black','red',
'blue','blue'),pch=19:19,
bg= ("white"), horiz=F)

```

```
## Electricitat industrial
```

```

n <- length(ELE$date[ELE$date<'2019-01-01'])
N <- nrow(ELE)
plot(ts.ele.indus)
summary(ts.ele.indus)

```

```

train.ele.indus = ts(data = ts.ele.indus.suav[1:n], start = c(2008,1), frequency = 12)
test.ele.indus = ts(data = ts.ele.indus.suav[(n+1):N], start = c(2008,1), frequency = 12)

```

```

class(train.ele.indus)
## Electricitat domèstic
train.ele.indus

```

```

plot(train.ele.indus)
hist(train.ele.indus, breaks = "Sturges", main="")

```

```
adf.test(train.ele.indus) # No es rebutja hipòtesis nula de no estacionarietat = no és estacionària
```

```
train.ele.indus.dif <- diff(train.ele.indus)  
train.ele.indus.dif
```

```
adf.test(train.ele.indus.dif)
```

```
plot(train.ele.indus.dif)
```

```
# descomposem la sèrie:
```

```
plot(decompose(train.ele.indus.dif))
```

```
# Autocorrelacions:
```

```
acf(train.ele.indus.dif, lag.max = 20, plot = T, main= "")
```

```
pacf(train.ele.indus.dif, lag.max =20, plot = T, main= "")
```

```
tsdisplay(train.ele.indus.dif)
```

```
# ARIMA(1,1,0)
```

```
ARIMA110.ele.indus = arima(train.ele.indus, order = c(1,1,0))
```

```
AIC(ARIMA110.ele.indus) # 373.0786
```

```
summary(ARIMA110.ele.indus)
```

```
coeftest(ARIMA110.ele.indus)
```

```
confint(ARIMA110.ele.indus)
```

```
# ARIMA(2,1,0)
```

```
ARIMA210.ele.indus = arima(train.ele.indus, order = c(2,1,0))
```

```
AIC(ARIMA210.ele.indus) # 373.0786
```

```
summary(ARIMA210.ele.indus)
```

```
coeftest(ARIMA210.ele.indus)
```

```
confint(ARIMA210.ele.indus)
```

```

# ARIMA(0,1,1)
ARIMA011.ele.indus = arima(train.ele.indus, order = c(0,1,1))
AIC(ARIMA011.ele.indus) # 377.3428 # augmenta

summary(ARIMA011.ele.indus)
coefstest(ARIMA011.ele.indus)
confint(ARIMA011.ele.indus)

# ARIMA(0,1,2)
ARIMA012.ele.indus = arima(train.ele.indus, order = c(0,1,2))
AIC(ARIMA012.ele.indus) # 375.7991 # augmenta

summary(ARIMA012.ele.indus)
coefstest(ARIMA012.ele.indus)# no significatiu
confint(ARIMA012.ele.indus)

# ARIMA(1,1,1)
ARIMA111.ele.indus = arima(train.ele.indus, order = c(1,1,1))
AIC(ARIMA111.ele.indus) # 374.9571 # major que AR(1)

summary(ARIMA111.ele.indus)
coefstest(ARIMA111.ele.indus) # MA(1) no significatiu AIC aumgenta, es descarta
confint(ARIMA111.ele.indus)

#Abans de mirar paràmetres d'estacionalitat fem un primer anàlisi dels residus:
res.ele.indus_110 <- residuals(ARIMA110.ele.indus)

adf.test(res.ele.indus_110)# estacionari
acf(res.ele.indus_110)
pacf(res.ele.indus_110)
tsdisplay(res.ele.indus_110)

qqnorm(res.ele.indus_110, main = "Residus ARIMA(1,1,0) Electr. Ind.", ylab = "Residus");
qqline(res.ele.indus_110)

```



```

pred.ele.indus <- forecast::forecast(ARIMA110.ele.indus,h = n_ahead)
pred.ele.indus$mean

autoplot(pred.ele.indus)

Aval.ele.indus <- data.frame(test=as.vector(test.ele.indus), pred=pred.ele.indus$mean,
Lo95=pred.ele.indus$lower[,2], Up95=pred.ele.indus$upper[,2])
val.ele.indus <- data.frame(test=as.vector(test.ele.indus), pred=pred.ele.indus$mean)

matplot(ELE$date[(n+1):N], Aval.ele.indus, type=c('o','o','l','l'), pch=20, lty = 1,
col=c('black','red', 'blue','blue')
      , ylab="Price Euro 2008", xlab="Year", main='ARIMA(1,1,0) Electr. Indus.')
legend("topleft", legend=c("control", "prediccions", "intervals 95%"), col=c('black','red',
'blue','blue'),pch=19:19,
      bg= ("white"), horiz=F)

## Electricitat majorista

n <- length(ELE$date[ELE$date<'2019-01-01'])
N <- nrow(ELE)

train.ele.sales = ts(data = ts.ele.sales.suav[1:n], start = c(2008,1), frequency = 12)
test.ele.sales = ts(data = ts.ele.sales.suav[(n+1):N], start = c(2019,1), frequency = 12)

class(train.ele.sales)
## Electricitat domèstic
plot(ts.ele.sales)
summary(ts.ele.sales)
49.69 + 1.5*(49.69 - 33.41)

train.ele.sales
plot(train.ele.sales)

```

```

hist(train.ele.sales, breaks = "Sturges", main="")
adf.test(train.ele.sales) # Es rebutja hipòtesis nula de no estacionarietat = és estacionària

train.ele.sales.dif <- diff(train.ele.sales)
train.ele.sales.dif

adf.test(train.ele.sales)

plot(train.ele.sales.dif)

# descomposem la sèrie:
plot(decompose(train.ele.sales))
plot(decompose(train.ele.sales.dif))
# Autocorrelacions:
acf(train.ele.sales, lag.max = 20, plot = T, main= "")
pacf(train.ele.sales, lag.max =20, plot = T, main= "")
tsdisplay(train.ele.sales)

acf(train.ele.sales.dif, lag.max = 20, plot = T, main= "")
pacf(train.ele.sales.dif, lag.max =20, plot = T, main= "")
tsdisplay(train.ele.sales.dif)

# ARIMA(0,0,1)
ARIMA001.ele.sales = arima(train.ele.sales, order = c(0,0,1))
AIC(ARIMA001.ele.sales) # 897.9117

summary(ARIMA001.ele.sales)
coefest(ARIMA001.ele.sales)
confint(ARIMA001.ele.sales)

# ARIMA(1,0,1)
ARIMA101.ele.sales = arima(train.ele.sales, order = c(1,0,1))
AIC(ARIMA101.ele.sales) # 702.6984

summary(ARIMA101.ele.sales)

```

```

coefstest(ARIMA101.ele.sales)
confint(ARIMA101.ele.sales)

# S'ha de tractar l'estacionalitat

# ARIMA(1,0,1)(0,0,1)
ARIMA101_001.ele.sales = arima(train.ele.sales, order = c(1,0,1), seasonal =
list(order=c(0,0,1),period=12))
AIC(ARIMA101_001.ele.sales) # 702.6984

summary(ARIMA101_001.ele.sales)
coefstest(ARIMA101_001.ele.sales)
confint(ARIMA101_001.ele.sales)

# ARIMA(1,1,1)
ARIMA111.ele.sales = arima(train.ele.sales, order = c(1,1,1))
AIC(ARIMA111.ele.sales) # 694.8849

summary(ARIMA111.ele.sales)
coefstest(ARIMA111.ele.sales)
confint(ARIMA111.ele.sales)

tsdisplay(train.ele.sales.dif) #suggereix una mitjana mobil

# ARIMA(0,1,1)
ARIMA011.ele.sales = arima(train.ele.sales, order = c(0,1,1))
AIC(ARIMA011.ele.sales) # 694.8817

summary(ARIMA011.ele.sales)
coefstest(ARIMA011.ele.sales) ## s'escull aquest
confint(ARIMA011.ele.sales)

# ARIMA(1,1,0)
ARIMA110.ele.sales = arima(train.ele.sales, order = c(1,1,0))

```

```
AIC(ARIMA110.ele.sales) # 696.0776
```

```
summary(ARIMA110.ele.sales)  
coefstest(ARIMA110.ele.sales) ## s'escull aquest  
confint(ARIMA110.ele.sales)
```

```
# ARIMA(0,1,1)(0,0,1)  
ARIMA011_001.ele.sales = arima(train.ele.sales, order = c(0,1,1), seasonal=list(order=c(0,0,1),  
period=12))  
AIC(ARIMA011_001.ele.sales) # 694.4688
```

```
summary(ARIMA011_001.ele.sales)  
coefstest(ARIMA011_001.ele.sales)  
confint(ARIMA011_001.ele.sales)
```

```
# ARIMA(0,1,1)(1,0,0)  
ARIMA011_100.ele.sales = arima(train.ele.sales, order = c(0,1,1), seasonal=list(order=c(1,0,0),  
period=12))  
AIC(ARIMA011_100.ele.sales) # 694.7575
```

```
summary(ARIMA011_100.ele.sales)  
coefstest(ARIMA011_100.ele.sales)  
confint(ARIMA011_100.ele.sales)
```

```
# Es comprova distribució de residus  
res.ele.sales_011 <- residuals(ARIMA011.ele.sales)
```

```
adf.test(res.ele.sales_011)# estacionari  
acf(res.ele.sales_011)  
pacf(res.ele.sales_011)  
tsdisplay(res.ele.sales_011)
```

```
qqnorm(res.ele.sales_011, main = "Residus ARIMA(0,1,1) Electr. Maj.", ylab = "Residus")#;  
qqline(res.ele.indus_121_111)
```

```
qqnorm(res.ele.indus_110, main = "Residus ARIMA(1,1,0) Electr. Ind.", ylab = "Residus")#;  
qqline(res.ele.indus_110) # ens quedem amb aquest
```

```
res.ele.indus_121_100 <- residuals(ARIMA121_100.ele.indus)  
qqnorm(res.ele.indus_121_100, main = "Residus ARIMA(1,2,1)(100) Electr. Indus.", ylab =  
"Residus"); qqline(res.ele.indus_121_100)
```

```
pred.ele.sales <- forecast::forecast(ARIMA011.ele.sales,h = n_ahead)  
pred.ele.sales$mean
```

```
autoplot(pred.ele.sales)
```

```
Aval.ele.sales <- data.frame(test=as.vector(test.ele.sales), pred=pred.ele.sales$mean,  
Lo95=pred.ele.sales$lower[,2], Up95=pred.ele.sales$upper[,2])
```

```
val.ele.sales <- data.frame(test=as.vector(test.ele.sales), pred=pred.ele.sales$mean)
```

```
matplot(ELE$date[(n+1):N], Aval.ele.sales, type=c('o','o','l','l'), pch=20, lty = 1, col=c('black','red',  
'blue','blue')  
  , ylab="Price Euro 2008", xlab="Year", main='ARIMA(0,1,1) Electr. Major.')
```

```
legend("topleft", legend=c("control", "prediccions", "intervals 95%"), col=c('black','red',  
'blue','blue'),pch=19:19,  
  bg= ("white"), horiz=F)
```

Sèries temporals multivariades (R)

```
PET_multi <- PET  
GAS_multi <- GAS  
ELE_multi <- ELE  
TEMP_multi <- df_temp_m[,c("date", "x_Normal", "x_High", "x_Low")]
```

```

names(PET_multi)
names(GAS_multi) <- c("date","gas.house","gas.indus","gas.sales")
names(ELE_multi) <- c("date","ele.house","ele.indus","ele.sales")
names(TEMP_multi)

GAS_multi$gas.house[is.na(GAS_multi$gas.house)] <-
mean(GAS_multi$gas.house[!is.na(GAS_multi$gas.house)])
GAS_multi$gas.indus[is.na(GAS_multi$gas.indus)] <-
mean(GAS_multi$gas.indus[!is.na(GAS_multi$gas.indus)])
GAS_multi$gas.sales[is.na(GAS_multi$gas.sales)] <-
mean(GAS_multi$gas.sales[!is.na(GAS_multi$gas.sales)])

#### DESCRIPTIU GAS, COMBUSTIBLES, TEMPERATURA #####
altres <-
cbind(PET_multi[1:n,c("croil.AVG","croil.BRENT","croil.DUBAI","croil.WTI","diese.AVG","heoil.A
VG","petro.AVG")]
      , GAS_multi[1:n,c("gas.house","gas.indus")]
      , TEMP_multi[1:n,c("x_High")])
names(altres) <-
c("croil.AVG","croil.BRENT","croil.DUBAI","croil.WTI","diese.AVG","heoil.AVG","petro.AVG","ga
s.house","gas.indus","x_High")

multi_ts_ALTRES = ts(altres, start = c(2008,1), frequency = 12)
plot(multi_ts_ALTRES, main='Sèries temporals')
summary(multi_ts_ALTRES)

#### SELECCIÓ VARIABLES I TRAIN/TEST #####
n <- length(ELE_multi$date[ELE_multi$date<'2019-01-01'])
(N <- nrow(PET_multi))

train <-
cbind(PET_multi[1:n,c("croil.AVG","croil.BRENT","croil.DUBAI","croil.WTI","diese.AVG","heoil.A
VG","petro.AVG")]

```

```

      , GAS_multi[1:n,c("gas.house","gas.indus")]
      , ELE_multi[1:n,c("ele.house","ele.indus","ele.sales")]
      , TEMP_multi[1:n,c("x_High")])
names(train) <-
c("croil.AVG","croil.BRENT","croil.DUBAI","croil.WTI","diese.AVG","heoil.AVG","petro.AVG","gas.house","gas.indus","ele.house","ele.indus","ele.sales","x_High")
head(train)
summary(train)

test <-
cbind(PET_multi[(n):N,c("croil.AVG","croil.BRENT","croil.DUBAI","croil.WTI","diese.AVG","heoil.AVG","petro.AVG")]
      , GAS_multi[(n):N,c("gas.house","gas.indus")]
      , ELE_multi[(n):N,c("ele.house","ele.indus","ele.sales")]
      , TEMP_multi[(n):N,c("x_High")])
names(test) <-
c("croil.AVG","croil.BRENT","croil.DUBAI","croil.WTI","diese.AVG","heoil.AVG","petro.AVG","gas.house","gas.indus","ele.house","ele.indus","ele.sales","x_High")
head(test)
summary(test)

# Correlograma
M_cor <- cor(train)

corrplot(M_cor, method = "circle", tl.cex = 0.7, cl.lim = 0.2, addCoef.col = TRUE)

# Descartem variables que estan molt correlacionades
train <- cbind(PET_multi[1:n,c("croil.WTI","diese.AVG","petro.AVG")]
              , GAS_multi[1:n,c("gas.house","gas.indus")]
              , ELE_multi[1:n,c("ele.house","ele.indus","ele.sales")]
              , TEMP_multi[1:n,c("x_High")])
names(train) = c("croil.WTI", "diese.AVG", "petro.AVG", "gas.house", "gas.indus", "ele.house", "ele.indus", "ele.sales", "x_High")
test <- cbind(PET_multi[(n):N,c("croil.WTI","diese.AVG","petro.AVG")]
              , GAS_multi[(n):N,c("gas.house","gas.indus")]

```

```

      , ELE_multi[(n):N,c("ele.house","ele.indus","ele.sales")]
      , TEMP_multi[(n):N,c("x_High")]
names(test) = c("croil.WTI", "diese.AVG", "petro.AVG", "gas.house", "gas.indus", "ele.house",
"ele.indus", "ele.sales", "x_High")

# Diferenciació necessària de les dades:
train_diff = data.frame(lapply(train, diff))
lapply(train_diff,adf.test)# totes son estacionaries

# Correlacions de base diferenciada
M_cor_diff <- cor(train_diff)
corrplot(M_cor_diff, method = "circle", tl.cex = 0.7, cl.lim = 0.2, addCoef.col = TRUE)
## S'elimina diese.AVG per estar molt correlacionada amb el preu del petroli i perque mostra 0
correlacio amb el gas domestic

train <- subset(train, select = -diese.AVG)
test <- subset(test, select = -diese.AVG)

train_diff = data.frame(lapply(train, diff))
test_diff = data.frame(lapply(test, diff))

##### MODELITZACIÓ
#####
# Creació del data set en format ts multivariant
multi_ts_ORIGINAL = ts(train, start = c(2008,1), frequency = 12) # La creem per veure plots
multi_ts = ts(train_diff, start = c(2008,1), frequency = 12)

test.multi_ts_ORIGINAL = ts(test, start = c(2019,1), frequency = 12)
test.multi_ts = ts(test_diff, start = c(2019,1), frequency = 12)

plot(multi_ts_ORIGINAL, main='Sèries temporals originals')
plot(multi_ts, main='Diferencials de sèries temporals originals')

plot(test.multi_ts_ORIGINAL, main='Sèries temporals originals')
plot(test.multi_ts, main='Diferencials de sèries temporals originals')

```



```
# Amb variables suavitzades
```

```
#Es fa un suavitzat de les següents variables:
```

```
# - croil.WTI
```

```
# - gas.house
```

```
# - gas.indus
```

```
# - ele.house
```

```
# - ele.sales
```

```
PET_multi[1:n,c("croil.AVG", "croil.BRENT", "croil.DUBAI", "croil.WTI", "diese.AVG", "heoil.AVG", "petro.AVG")]
```

```
names(train_diff)
```

```
data.suav <- data.frame(croil.WTI = ts.pet.wti.suav
```

```
, petro.AVG = PET_multi[, "petro.AVG"]
```

```
, gas.house = ts.gas.house.suav
```

```
, gas.indus = ts.gas.indus.suav
```

```
, ele.house = ts.ele.house.suav
```

```
, ele.indus = ts.ele.indus.suav
```

```
, ele.sales = ts.ele.sales.suav
```

```
, x_High = df_temp_m[, "x_High"]
```

```
)
```

```
names(data.suav)
```

```
summary(data.suav)
```

```
train.suav <- data.suav[(1:n),]
```

```
test.suav <- data.suav[((n):N),]
```

```
# Creació del data set en format ts multivariant
```

```
multi_ts.suavORIGINAL = ts(train.suav, start = c(2008,1), frequency = 12)
```

```
multi_ts.suav = ts(train_diff.suav, start = c(2008,1), frequency = 12)
```

```
plot(multi_ts.suav, main='Sèries temporals suavitzades', col='red')
```

```
plot(multi_ts.suav, main='Diferencials de sèries temporals suavitzades', col='red')
```

```
##### VARMA
```

```
#####
```

```
# Exploració d'ARIMA de sèries restants
```

```
ts.pet.croil.suav
```

```
train.pet.croi = ts(data = ts.pet.croil.suav[1:n], start = c(2008,1), frequency = 12)
```

```
test.pet.croi = ts(data = ts.pet.croil.suav[(n+1):N], start = c(2008,1), frequency = 12)
```

```
auto.arima(train.pet.croi, trace = T) #ARIMA(1,1,0)(0,0,2)[12]
```

```
#811.63
```

```
ts.pet.petro
```

```
train.pet.petro = ts(data = ts.pet.petro[1:n], start = c(2008,1), frequency = 12)
```

```
test.pet.petro = ts(data = ts.pet.petro[(n+1):N], start = c(2008,1), frequency = 12)
```

```
auto.arima(train.pet.petro, trace = T) #ARIMA(1,1,0)(0,0,1)[12]
```

```
#-455.3
```

```
ts.gas.house
```

```
train.gas.house = ts(data = ts.gas.house[1:n], start = c(2008,1), frequency = 12)
```

```
test.gas.house = ts(data = ts.gas.house[(n+1):N], start = c(2008,1), frequency = 12)
```

```
auto.arima(train.gas.house, trace = T) #ARIMA(0,1,0)
```

```
#509.62
```

```
#ARIMA(1,1,1)
```

```
#511.5926
```

```
ts.gas.indus
```

```
train.gas.indus = ts(data = ts.gas.indus[1:n], start = c(2008,1), frequency = 12)
```

```
test.gas.indus = ts(data = ts.gas.indus[(n+1):N], start = c(2008,1), frequency = 12)
```

```
auto.arima(train.gas.indus, trace = T) #ARIMA(0,1,0)(1,0,0)[12] with drift
```

```
#359.72
```

```
#ARIMA(1,1,1)(1,0,0)[12] with drift
```

```
#363.3263
```

```
## PRIMERA DIFERENCIACIO
```

```

# Modelització
AIC_m <- matrix(rep(0,9),nrow=3, ncol=3)

for(i in 0:2){
  for (j in 0:2){
    VARMA.suav = VARMA(multi_ts.suav, p=i, q=j)
    AIC_m[i+1,j+1] = VARMA.suav$aic
  }
}
AIC_m[1,1] <- NULL
AIC_m[3,3] <- 0
rownames(AIC_m) <- c("p=0","p=1","p=2")
colnames(AIC_m) <- c("q=0","q=1","q=2")
colors <- colorRampPalette(c("green", "red"))(10)
corrplot(AIC_m,is.corr = FALSE, method = 'color',col = colors, tl.cex = 0.7, cl.lim = 0.2,
addCoef.col = TRUE)

##### VAR
#####
## PRIMERA DIFERENCIACIO
# Modelització
summary(multi_ts.suav)
varselect1.suav = VARselect(multi_ts.suav, type = "none", lag.max = 12)
varselect2.suav = VARselect(multi_ts.suav, type = "const", lag.max = 12)
varselect3.suav = VARselect(multi_ts.suav, type = "both", lag.max = 12)

AICs.suav <- data.frame(none=varselect1.suav$criteria[1,1:12],
const=varselect2.suav$criteria[1,1:12], both=varselect3.suav$criteria[1,1:12])

matplot(AICs.suav , type='o', pch=20, lty = 1
, ylab="AIC", xlab="AIC", main='AIC de models proposats')
legend("topright", legend=c("none", "const", "both"), col=c(1:10),pch=19:19,
bg= ("white"), horiz=F)

```

```
plot(varselect3.suav$criteria[1,1:12], type='o', pch=20, lty=1, ylab="AIC", xlab="p", main='AIC  
VAR(p)')
```

```
# Obtenim valors optims
```

```
(P <- varselect$selection[[1]])
```

```
#Diferents propostes de models amb varietat de tipus de regressors
```

```
VAR2.suav <- vars::VAR(multi_ts.suav, lag.max = 12, ic = "AIC",type = "const")
```

```
summary(VAR2.suav)
```

```
VAR3.suav <- vars::VAR(multi_ts.suav, lag.max = 12, ic = "AIC",type = "both")
```

```
summary(VAR3.suav)
```

```
res2.suav <- residuals(VAR2.suav)
```

```
res3.suav <- residuals(VAR3.suav)
```

```
#Causalitat de Granger
```

```
causality(VAR2.suav, cause = c("croil.WTI"))
```

```
causality(VAR2.suav, cause = c("petro.AVG"))
```

```
causality(VAR2.suav, cause = c("gas.house"))
```

```
causality(VAR2.suav, cause = c("gas.indus"))
```

```
causality(VAR2.suav, cause = c("ele.house"))
```

```
causality(VAR2.suav, cause = c("ele.indus"))
```

```
causality(VAR2.suav, cause = c("ele.sales"))
```

```
causality(VAR2.suav, cause = c("x_High"))
```

```
causality(VAR3.suav, cause = c("croil.WTI"))
```

```
causality(VAR3.suav, cause = c("petro.AVG"))
```

```
causality(VAR3.suav, cause = c("gas.house"))
```

```
causality(VAR3.suav, cause = c("gas.indus"))
```

```
causality(VAR3.suav, cause = c("ele.house"))
```

```
causality(VAR3.suav, cause = c("ele.indus"))
```

```
causality(VAR3.suav, cause = c("ele.sales"))
```

```
causality(VAR3.suav, cause = c("x_High"))
```

```

# Revisió dels errors
plot(as.data.frame(res3.suav))
RES3 = ts(res3.suav, start = c(2008,1), frequency = 12)
plot(x=RES3#, y=1:ncol(RES3)
     , main="Residus VAR(12)", xlab="")

par(mfrow = c(3, 1))
plot(res2.suav[,1], type = "l", main = paste("Residus Model 2 -", colnames(res2.suav)[1]), ylab =
"Residus")
plot(res3.suav[,1], type = "l", main = paste("Residus Model 3 -", colnames(res3.suav)[1]), ylab =
"Residus")

plot(res2.suav[,2], type = "l", main = paste("Residus Model 2 -", colnames(res2.suav)[2]), ylab =
"Residus")
plot(res3.suav[,2], type = "l", main = paste("Residus Model 3 -", colnames(res3.suav)[2]), ylab =
"Residus")

plot(res2.suav[,3], type = "l", main = paste("Residus Model 2 -", colnames(res2.suav)[3]), ylab =
"Residus")
plot(res3.suav[,3], type = "l", main = paste("Residus Model 3 -", colnames(res3.suav)[3]), ylab =
"Residus")

plot(res2.suav[,4], type = "l", main = paste("Residus Model 2 -", colnames(res2.suav)[4]), ylab =
"Residus")
plot(res3.suav[,4], type = "l", main = paste("Residus Model 3 -", colnames(res3.suav)[4]), ylab =
"Residus")

plot(res2.suav[,5], type = "l", main = paste("Residus Model 2 -", colnames(res2.suav)[5]), ylab =
"Residus")
plot(res3.suav[,5], type = "l", main = paste("Residus Model 3 -", colnames(res3.suav)[5]), ylab =
"Residus")

```

```
plot(res2.suav[,6], type = "l", main = paste("Residus Model 2 -", colnames(res2.suav)[6]), ylab = "Residus")
```

```
plot(res3.suav[,6], type = "l", main = paste("Residus Model 3 -", colnames(res3.suav)[6]), ylab = "Residus")
```

```
plot(res2.suav[,7], type = "l", main = paste("Residus Model 2 -", colnames(res2.suav)[7]), ylab = "Residus")
```

```
plot(res3.suav[,7], type = "l", main = paste("Residus Model 3 -", colnames(res3.suav)[7]), ylab = "Residus")
```

```
plot(res2.suav[,8], type = "l", main = paste("Residus Model 2 -", colnames(res2.suav)[8]), ylab = "Residus")
```

```
plot(res3.suav[,8], type = "l", main = paste("Residus Model 3 -", colnames(res3.suav)[8]), ylab = "Residus")
```

```
# Estacionarietat dels residus
```

```
paste('m2:', adf.test(res2.suav[,1])$p.value, 'm3:', adf.test(res3.suav[,1])$p.value)
```

```
paste('m2:', adf.test(res2.suav[,2])$p.value, 'm3:', adf.test(res3.suav[,2])$p.value)
```

```
paste('m2:', adf.test(res2.suav[,3])$p.value, 'm3:', adf.test(res3.suav[,3])$p.value)
```

```
paste('m2:', adf.test(res2.suav[,4])$p.value, 'm3:', adf.test(res3.suav[,4])$p.value)
```

```
paste('m2:', adf.test(res2.suav[,5])$p.value, 'm3:', adf.test(res3.suav[,5])$p.value)
```

```
paste('m2:', adf.test(res2.suav[,6])$p.value, 'm3:', adf.test(res3.suav[,6])$p.value)
```

```
paste('m2:', adf.test(res2.suav[,7])$p.value, 'm3:', adf.test(res3.suav[,7])$p.value)
```

```
# Retars dels residus
```

```
tsdisplay(res2.suav[,1], main = colnames(res2.suav)[1])
```

```
tsdisplay(res2.suav[,2], main = colnames(res2.suav)[2])
```

```
tsdisplay(res2.suav[,3], main = colnames(res2.suav)[3])
```

```
tsdisplay(res2.suav[,4], main = colnames(res2.suav)[4])
```

```
tsdisplay(res2.suav[,5], main = colnames(res2.suav)[5])
```

```
tsdisplay(res2.suav[,6], main = colnames(res2.suav)[6])
```

```
tsdisplay(res2.suav[,7], main = colnames(res2.suav)[7])
```

```
tsdisplay(res2.suav[,8], main = colnames(res2.suav)[8])
```

```
tsdisplay(res3.suav[,1], main = colnames(res3.suav)[1])
tsdisplay(res3.suav[,2], main = colnames(res3.suav)[2])
tsdisplay(res3.suav[,3], main = colnames(res3.suav)[3])
tsdisplay(res3.suav[,4], main = colnames(res3.suav)[4])
tsdisplay(res3.suav[,5], main = colnames(res3.suav)[5])
tsdisplay(res3.suav[,6], main = colnames(res3.suav)[6])
tsdisplay(res3.suav[,7], main = colnames(res3.suav)[7])
tsdisplay(res3.suav[,8], main = colnames(res3.suav)[8])
```

```
par(mfrow=c(2,2))
```

```
acf(res3.suav[,1], lag.max = 20, plot = T, main= colnames(res3.suav)[1])
acf(res3.suav[,2], lag.max = 20, plot = T, main= colnames(res3.suav)[2])
acf(res3.suav[,3], lag.max = 20, plot = T, main= colnames(res3.suav)[3])
acf(res3.suav[,4], lag.max = 20, plot = T, main= colnames(res3.suav)[4])
acf(res3.suav[,5], lag.max = 20, plot = T, main= colnames(res3.suav)[5])
acf(res3.suav[,6], lag.max = 20, plot = T, main= colnames(res3.suav)[6])
acf(res3.suav[,7], lag.max = 20, plot = T, main= colnames(res3.suav)[7])
acf(res3.suav[,8], lag.max = 20, plot = T, main= colnames(res3.suav)[8])
```

```
pacf(res3.suav[,1], lag.max = 20, plot = T, main= colnames(res3.suav)[1])
pacf(res3.suav[,2], lag.max = 20, plot = T, main= colnames(res3.suav)[2])
pacf(res3.suav[,3], lag.max = 20, plot = T, main= colnames(res3.suav)[3])
pacf(res3.suav[,4], lag.max = 20, plot = T, main= colnames(res3.suav)[4])
pacf(res3.suav[,5], lag.max = 20, plot = T, main= colnames(res3.suav)[5])
pacf(res3.suav[,6], lag.max = 20, plot = T, main= colnames(res3.suav)[6])
pacf(res3.suav[,7], lag.max = 20, plot = T, main= colnames(res3.suav)[7])
pacf(res3.suav[,8], lag.max = 20, plot = T, main= colnames(res3.suav)[8])
```

```
#### VALORACIÓ DELS MODELS PROPOSATS
```

```
#####
```

```
AIC(VAR2.suav)#750.6197
```

```
AIC(VAR3.suav)#591.578
```

```

#### FORECASTING
#####
n_pred <- nrow(test_diff.suav)
par(mar = c(2.5,2.5,2.5,2.5))

pred2 = predict(VAR2.suav, n.ahead = n_pred)
plot(pred2)

pred3 = predict(VAR3.suav, n.ahead = n_pred)
plot(pred3)

# Extracció dels valors
## Model VAR

PRED_ele.house = data.frame(m1.diff = pred2$fcst["ele.house"]
, m2.diff = pred3$fcst["ele.house"]
, test = test_diff$ele.house)

PRED_ele.indus = data.frame(m1 = pred2$fcst["ele.indus"]
, m2 = pred3$fcst["ele.indus"]
, test = test_diff.suav$ele.indus)

PRED_ele.sales = data.frame(m1 = pred2$fcst["ele.sales"]
, m2 = pred3$fcst["ele.sales"]
, test = test_diff.suav$ele.sales)

PRED_gas.house = data.frame(m1 = pred2$fcst["gas.house"]
, m2 = pred3$fcst["gas.house"]
, test = test_diff.suav$gas.house)

PRED_gas.indus = data.frame(m1 = pred2$fcst["gas.indus"]
, m2 = pred3$fcst["gas.indus"]
, test = test_diff.suav$gas.indus)

```



```
PRED_croil.WTI = data.frame(m1 = pred2$fcst["croil.WTI"]
, m2 = pred3$fcst["croil.WTI"]
, test = test_diff.suav$croil.WTI)
```

```
PRED_petro.AVG = data.frame(m1 = pred2$fcst["petro.AVG"]
, m2 = pred3$fcst["petro.AVG"]
, test = test_diff.suav$petro.AVG)
```

```
PRED_x_High = data.frame(m1 = pred2$fcst["x_High"]
, m2 = pred3$fcst["x_High"]
, test = test_diff.suav$x_High)
```

```
#### COMPARACIÓ AMB GRUP CONTROL
```

```
#####
```

```
par(mfrow = c(2,1))
```

```
n <- length(ELE_multi$date[ELE_multi$date<'2019-01-01'])
```

```
N <- nrow(PET_multi)
```

```
#electricitat domèstica
```

```
val2.ele.house <- data.frame(test=test.suav$ele.house, pred=cumsum(c(test.suav$ele.house[1],
PRED_ele.house[["m2.diff.ele.house.fcst"]]))
```

```
, lower=cumsum(c(test.suav$ele.house[1,
```

```
PRED_ele.house[["m2.diff.ele.house.lower"]])) )
```

```
, upper=cumsum(c(test.suav$ele.house[1,
```

```
PRED_ele.house[["m2.diff.ele.house.upper"]])) )
```

```
)
```

```
matplot(ELE$date[(n+1):N],
```

```
PRED_ele.house[,c("test", "m2.diff.ele.house.fcst", "m2.diff.ele.house.lower",
```

```
"m2.diff.ele.house.upper" )], type=c('o','o','l','l'), pch=20, lty = 1
```

```
, col=c('black','red','blue','blue'), ylab="Price Euro 2008", xlab="Year", main='Pred.
```

```
Diferencials electricitat domèstica ')
```

```
legend("bottomright", legend=c("control", "prediccions", "interval 95%"),
```

```
col=c('black','red','blue','blue'), pch=19:19,
```

```
bg= ("white"), horiz=F)
```

```

matplot(ELE$date[(n+1):N], val2.ele.house[2:13,], type=c('o','o','l','l'), pch=20, lty = 1
        ,col=c("black","red","blue","blue"), ylab="Price Euro 2008", xlab="Year", main='Pred.
Electricitat domèstic ')
legend("bottomleft", legend=c("control", "prediccions",'intervals 95%'),
col=c("black","red","blue","blue"),pch=19:19,
      bg= ("white"), horiz=F)

```

```
#electricitat industrial
```

```

val2.ele.indus <- data.frame(test=test.suav$ele.indus, pred=cumsum(c(test.suav$ele.indus[1],
PRED_ele.indus[["m2.ele.indus.fcst"]]))
        , lower=cumsum(c(test.suav$ele.indus[1],
PRED_ele.indus[["m2.ele.indus.lower"]])) )
        , upper=cumsum(c(test.suav$ele.indus[1],
PRED_ele.indus[["m2.ele.indus.upper"]])) )
)
matplot(ELE$date[(n+1):N], PRED_ele.indus[,c("test","m2.ele.indus.fcst")] , type='o', pch=20, lty
= 1
        , ylab="Price Euro 2008", xlab="Year", main='Pred. Diferencials electricitat industrial ')
legend("bottomright", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

```

```

matplot(ELE$date[(n):N], data.frame(test.suav$ele.indus , cumsum(c(test.suav$ele.indus[1],
PRED_ele.indus[["m2.ele.indus.fcst"]])) ), type='o', pch=20, lty = 1
        , ylab="Price Euro 2008", xlab="Year", main='Pred. Electricitat industrial ')
legend("bottomleft", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

```

```

matplot(ELE$date[(n+1):N], val2.ele.indus[2:13,], type=c('o','o','l','l'), pch=20, lty = 1
        ,col=c("black","red","blue","blue"), ylab="Price Euro 2008", xlab="Year", main='Pred.
Electricitat industrial ')

```

```

legend("bottomleft", legend=c("control", "prediccions","intervals 95%"),
col=c("black","red","blue","blue"),pch=19:19,
      bg= ("white"), horiz=F)

#Electricitat majorista
val2.ele.sales <- data.frame(test=test.suav$ele.sales, pred=cumsum(c(test.suav$ele.sales[1],
PRED_ele.sales[["m2.ele.sales.fcst"]]))
      , lower=cumsum(c(test.suav$ele.sales[1],
PRED_ele.sales[["m2.ele.sales.lower"]]) )
      , upper=cumsum(c(test.suav$ele.sales[1],
PRED_ele.sales[["m2.ele.sales.upper"]]) )
)
matplot(ELE$date[(n+1):N], PRED_ele.sales[,c("test", "m2.ele.sales.fcst")] , type='o', pch=20, lty
= 1
      , ylab="Price Euro 2008", xlab="Year", main='Pred. Diferencials electricitat majorista ')
legend("bottomright", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

matplot(ELE$date[(n):N], data.frame(test$ele.sales , cumsum(c(test.suav$ele.sales[1],
PRED_ele.sales[["m2.ele.sales.fcst"]]) )), type='o', pch=20, lty = 1
      , ylab="Price Euro 2008", xlab="Year", main='Pred. Electricitat majorista ')
legend("bottomleft", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

matplot(ELE$date[(n+1):N], val2.ele.sales[2:13,], type=c('o','o','l','l'), pch=20, lty = 1
      ,col=c("black","red","blue","blue"), ylab="Price Euro 2008", xlab="Year", main='Pred.
Electricitat majorista ')
legend("topleft", legend=c("control", "prediccions","intervals 95%"),
col=c("black","red","blue","blue"),pch=19:19,
      bg= ("white"), horiz=F)

gcontrol <- ELE_multi$ele.sales[(length(ELE_multi$ele.sales)-11):length(ELE_multi$ele.sales)]
gpred <- cumsum(c(ELE_multi$ele.sales[length(ELE_multi$ele.sales)-12],
PRED_ele.sales[["m2.ele.sales.fcst"]]) )

```

```

gpred <- gpred[-1]
all_hist <- ELE_multi$ele.sales
all_gpred <- c(rep(NA,length(ELE_multi$ele.house)-12), gpred)

matplot( data.frame(all_hist, all_gpred), type='o', pch=20, lty = 1
        , ylab="Price Euro 2008", xlab="Year", main='Pred. Electricitat majorista ')
matplot(ELE$date[(n+1):N], data.frame(gcontrol , gpred ), type='o', pch=20, lty = 1
        , ylab="Price Euro 2008", xlab="Year", main='Pred. Electricitat majorista ')
legend("bottomright", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

```

#Altres

```

matplot(ELE$date[(n+1):N], PRED_gas.house[,c("m2.gas.house.fcst","test")] , type='o', pch=20,
lty = 1
        , ylab="Price Euro 2008", xlab="Year", main='Gas domèstic m2')
legend("topright", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

```

```

matplot(ELE$date[(n+1):N], PRED_gas.indus[,c("m2.gas.indus.fcst","test")] , type='o', pch=20,
lty = 1
        , ylab="Price Euro 2008", xlab="Year", main='Gas industrial m2')
legend("topright", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

```

```

matplot(ELE$date[(n+1):N], PRED_croil.WTI[,c("m2.croil.WTI.fcst","test")] , type='o', pch=20, lty
= 1
        , ylab="Price Euro 2008", xlab="Year", main='Combustible fòssil Cru WTI')
legend("topright", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

```

```

matplot(ELE$date[(n+1):N], PRED_petro.AVG[,c("m2.petro.AVG.fcst","test")] , type='o', pch=20,
lty = 1

```

```

, ylab="Price Euro 2008", xlab="Year", main='Combustible fòssil Petroli')
legend("topright", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

matplot(ELE$date[(n+1):N], PRED_x_High[,c("m2.x_High.fcst", "test")] , type='o', pch=20, lty = 1
, ylab="Price Euro 2008", xlab="Year", main='Temperatures màximes')
legend("topright", legend=c("control", "prediccions"), col=c(1:10),pch=19:19,
      bg= ("white"), horiz=F)

```

```
#### Càlcul dels RMSE
```

```
#####
```

```
# Del model multivariant
```

```

RMSE <- function(pred,test){
  mse <- mean((pred - test)^2)
  print(sqrt(mse))
}

```

```
RMSE.ele.house.multi = RMSE(val2.ele.house$pred[2:13],val2.ele.house$test[2:13])
```

```
RMSE.ele.indus.multi = RMSE(val2.ele.indus$pred[2:13],val2.ele.indus$test[2:13])
```

```
RMSE.ele.sales.multi = RMSE(val2.ele.sales$pred[2:13],val2.ele.sales$test[2:13])
```

```
#Dels models univariants
```

```
RMSE.ele.house.uni = RMSE(val.ele.sales$pred,val.ele.sales$test)
```

```
RMSE.ele.indus.uni = RMSE(val.ele.indus$pred,val.ele.indus$test)
```

```
RMSE.ele.sales.uni = RMSE(val.ele.sales$pred,val.ele.sales$test)
```