

Doble Grau en Estadística i Economia

Títol: Predicció de la volatilitat implícita a partir d'indicadors d'incertesa.

Autor: Jaquelina Cera Piquet

Director: Salvador Torra Porras

Departament: Economia, Estadística i Economia Aplicada.

Convocatòria: Setembre 2021



Resum

Aquest treball està compost de dues parts degut a que és un projecte de final de grau de la doble titulació d'economia i estadística que pretén relacionar els conceptes d'incertesa política-econòmica i risc geopolític amb la volatilitat implícita de l'economia d'Estats Units.

La primera part, vinculada amb la part del grau d'economia, es basa en dur a terme una presentació dels conceptes d'incertesa política-econòmica, l'indicador de risc geopolític i la volatilitat (VIX), de manera que s'especifica com es construeixen aquests indicadors i quins són els seus principals efectes en l'economia.

La segona part del treball, consta d'analitzar si es possible predir la volatilitat (VIX) a partir del indicador d'incertesa política-econòmica (EPU) i el indicador de risc geopolític (GPR). Donada l'estructura de les dades, la metodologia emprada consta d'aplicar diferents models per a fer la predicció, com és el cas del model ARIMA o bé tècniques d'aprenentatge automàtic, com són les xarxes neuronals. En tots els casos es proposa realitzar una modelització univariant del VIX en comparació amb una modelització multivariant, on l'objectiu és predir el VIX a partir dels indicadors EPU i GPR.

Paraules clau: Incertesa, risc geopolític, EPU, GPR, VIX, volatilitat, xarxes neuronals, NNAR, ARIMA.

Abstract

This work is composed of two parts because it is a final degree project of the double degree in economics and statistics that aims to relate the concepts of political-economic uncertainty and geopolitical risk with the implicit volatility of the economy of United States.

The first part, linked to the part of the economics degree, is based on a presentation of the concepts of political-economic uncertainty, the geopolitical risk indicator and volatility (VIX), so that it is specified how these indicators are constructed and what their main effects are on the economy.

The second part of the work consists of analyzing whether it is possible to predict volatility (VIX) from the indicator of political-economic uncertainty (EPU) and the indicator of geopolitical risk (GPR). Given the structure of the data, the methodology used consists of applying different models to make the prediction, such as the ARIMA model or machine learning techniques, such as neural networks. In all cases, it is proposed to perform a univariate modeling of the VIX compared to a multivariate modeling, where the goal is to predict the VIX from the EPU and GPR indicators.

Keywords: Uncertainty, geopolitical risk, EPU, GPR, VIX, volatility, neural networks, NNAR, ARIMA.

Classificació AMS (American Mathematical Society)

37M10 - Time series analysis.

62M20 – Prediction.

62M10 – Time series, auto-correlation, regression.

68T07 – Artificial neural networks and deep learning.

Index

1. Introducció	5
2. Metodologia	6
3. Primeres definicions d'incertesa	7
4. L'efecte de la incertesa en l'economia.....	7
4.1 Consum.....	7
4.2 Empreses	7
4.3 Expectatives.....	8
5. Indicadors d'incertesa econòmica.....	8
5.1. Indicadors basats en articles de premsa	8
5.1.1 Economic Policy Uncertainty Index (EPU)	8
5.1.1.1 Índex EPU Global	9
5.1.1.2 Índex EPU a Estats Units.....	10
5.1.1.3 Índex EPU a Espanya	11
5.1.2 Geopolitical risk (GPR).....	14
5.1.2.1 Altres indicadors de risc geopolític: BlackRock Geopolitical Risk Indicator (BGRI)	18
5.1.3 Monetary Policy Uncertainty (MPU)	20
6. Comparativa entre indicadors d'incertesa	22
7. Efectes dels indicadors d'incertesa en el mercat.	23
7.1 Efectes del EPU.....	23
7.1.1 La incertesa i els seus efectes macroeconòmics - cas espanyol.....	24
7.2 Efectes del GPR.....	25
8. Índex de volatilitat.....	27
8.1 VIX.....	27
9. Anàlisi Previ.....	29
9.1 Preparació de les dades.....	29
9.2 Anàlisi estadístic dels indicadors	33
9.3 Definició de les particions	42
10. Models lineals estocàstics (ARIMA).....	43
10.1 ARIMA Univariant del VIX.....	44
10.2 ARIMA Multivariant del VIX.....	49
11. Models no lineals (Xarxes neuronals).....	53
11.1 Models neuronals (feed-forward)	53

11.2 Models neuronals autoregressius (NNAR)	57
11.2.1 Model Neuronal Autoregressiu Univariant.	59
11.2.2 Model Neuronal Autoregressiu Multivariant.	60
12. Conclusions.....	63
Índex de figures	65
Índex de taules	67
Bibliografia.....	68
Annexes	70
Codi RStudio	70

1. Introducció

En els últims anys el concepte de globalització ha ocasionat múltiples conseqüències positives en àmbits socioculturals, demogràfics, polítics i econòmics entre d'altres. Així doncs, es lògic pensar que tot i que aquests esdeveniments tinguin bones implicacions en l'àmbit econòmic, la globalització també ha suposat esdeveniments que han generat un gran dalt a baix dins del sector financer, com poden ser conflictes bèl·lics, conflictes entre països, decisions polítiques que prenen les gran potencies mundials...

Fa exactament vint anys, el mon va quedar consternat en veure l'atac terrorista que va tenir lloc el 11 de setembre a les Torres Bessones de Nova York. Tenint com a referencia aquest esdeveniment i el gran impacte que va ocasionar a nivell mundial, sorgeix una inquietud personal per estudiar com aquests fets tenen un impacte en l'economia.

D'aquesta manera, sorgeix el dubte de si es possible que indicadors relacionats amb la incertesa política – econòmica o bé indicadors de risc geopolític son capaços de predir el índex de volatilitat en l'economia d'Estats Units. L'objectiu principal d'aquest treball és identificar si aquests indicadors poden predir la volatilitat, fent servir les dades proporcionades per les pagines webs oficials dels indicadors, agafant com a referencia els últims trenta anys.

Degut a que una de les majors qüestions que estan sorgint avui en dia és la preocupació que pugui haver pel futur, han aparegut una sèrie de models matemàtics que a partir de l'anàlisi de dades massives intenten aproximar el més acuradament possible els valors de la realitat, com poden ser models de regressió lineal simple o bé tècniques de *Machine Learning*, de manera que al llarg d'aquest treball s'utilitzaran les dos tipologies de models esmentats anteriorment. El primer, més clàssic, és el model ARIMA (model autoregressiu integrat de mitjanes mòbils) i el segon és un model de xarxes neuronals, ambdós realitzats tant des de un punt de vista univariant com multivariant.

2. Metodologia

Aquest treball està dividit en dos parts diferenciades entre si. La primera correspon als aspectes més teòrics dels conceptes i dels indicadors que s'utilitzaran al llarg del treball, mentre que la segona es tracta d'un anàlisi empíric que es basa en l'aplicació de tècniques de regressió lineal i models de xarxes neuronals per a realitzar la predicció de la volatilitat implícita.

La primera part, com s'ha mencionat anteriorment, es basa en definir els conceptes bàsics vinculats amb el treball així com en explicar la construcció dels indicadors d'incertesa, dels quals destaca el indicador d'incertesa política-econòmica; el indicador de risc geopolític i el índex de volatilitat tot indicant quins son els seus efectes en el mercat. Per a poder realitzar aquest anàlisi, es fan ús de diferents documents metodològics així com articles acadèmics, com per exemple els documents metodològics que es poden trobar a les pàgines oficials dels indicadors corresponents.

La segona part, que consta d'un anàlisi empíric utilitzant el *software* RStudio, tracta d'aplicar la modelització ARIMA, així com models de xarxes neuronals per a poder realitzar la predicció de la volatilitat implícita a partir de la seva pròpia informació o bé a partir dels indicadors d'incertesa política-econòmica (EPU) i el indicador de risc geopolític (GPR). Per a poder realitzar aquest estudi, s'ha creat una base de dades diària amb el mateix marc històric, sent aquest des del gener del 1990 fins març del 2020.

3. Primeres definicions d'incertesa

Frank Knight (1921), va ser un dels primers economistes que va diferenciar el concepte d'incertesa i el risc. En el cas del risc, a priori es coneix la distribució dels resultats d'un conjunt d'esdeveniments. En canvi, pel que fa a la incertesa, es caracteritza per la incapacitat d'avaluar la probabilitat d'ocurrència en els determinats casos.

4. L'efecte de la incertesa en l'economia

En el context de la literatura econòmica, es consideren tres àmbits en els quals la incertesa pot afectar les decisions que prenen els agents en l'economia: el consum, les decisions empresarials i les pròpies expectatives.

4.1 Consum

Segons Bloom (2014) en períodes on la incertesa es molt elevada, la elasticitat de demanda dels bens duradors disminueix, ja que els agents esperen el moment en que esdevingui una millor oportunitat per adquirir aquesta tipologia de bens. En conseqüència a aquest fet, l'estalvi dels agents econòmics augmenta. Seguint amb l'argument, es podria concloure que aquest fet ocasionaria un augment en la inversió en el moment en que disminueix la taxa d'interès i el nivell de preus. Ara bé l'efecte advers que podria ocasionar-se es que hi hagués un estancament dels preus i que per tant no disminueixi ni la taxa d'interès ni el nivell de preus fins al punt necessari per a poder incentivar la inversió, de manera que el resultat final seria que la producció total disminueix.

4.2 Empreses

Pel que fa a les empreses, davant la incertesa es veuen afectats el nivell d'ocupació i les decisions d'inversió. Les expectatives dels agents d'inversió suposen que davant d'un escenari amb alta incertesa tindran lloc els pitjors resultats, de manera que es les decisions d'inversió no es duran a terme per tal d'evitar resultats no satisfactoris Bloom (2014).

4.3 Expectatives

La incertesa afecta a les creences dels inversors sobre les polítiques que el govern pugui dur a terme en un futur, augmentant així la prima de risc exigida (Pastor i Varonesi, 2013), que ocasiona que el cost de finançar-se augmenti. Davant aquest fet, es podria ocasionar una fuga de capitals, degut a que els inversionistes estrangers vinguin el capital que tenen en el país afectat per la incertesa politico-econòmica i invertint aquests capitals, en un mercat on presenti una millor estabilitat (Hillier i Loncan, 2019).

Bali et al.,(2017) argumenten que un augment de la incertesa redueix el consum dels inversors, de manera que aquests canvien el seu consum i inversions per tal de protegir-se davant del empitjorament de la activitat econòmica. On aquests, busquen invertir en accions que estiguin correlacionades positivament amb la incertesa econòmica.

5. Indicadors d'incertesa econòmica.

Els economistes utilitzen metodologies empíriques amb l'objectiu de mesurar la incertesa. Per exemple, Bachmann et al., (2013) proposen utilitzar com a indicador el desacord entre les expectatives dels agents. Bloom (2009) proposa utilitzar la volatilitat observada en el mercat de valors. Baker et al.,(2015) proposen utilitzar la freqüència de referències en articles de premsa de conceptes relacionats amb la incertesa de la política econòmica. Donadelli (2014) proposa que un indicador per a mesurar la incertesa es pot reflectir en el volum de recerques a Google de temes relacionats amb la política.

A continuació s'expliquen amb detall alguns dels indicadors esmentats anteriorment, i que són de principal interès al llarg d'aquest document.

5.1. Indicadors basats en articles de premsa

A continuació es mostraran aquells indicadors on la seva metodologia es basa en el recompte d'articles de premsa publicats.

5.1.1 Economic Policy Uncertainty Index (EPU)

El índex d'incertesa política-econòmica, o EPU (*Economic Policy Uncertainty Index*), construït per Baker et al., (2015), es basa en calcular la freqüència de certes paraules en diferents articles de premsa. On el seu principal objectiu és captar la incertesa de qui pren les decisions polítiques-econòmiques, quines mesures es duren a terme i en quin moment i finalment tenir en compte els efectes de les polítiques econòmiques realitzades o bé la seva falta d'acció.

Per tal de veure amb més detall la construcció del indicador d'incertesa política-econòmica, s'estudiarà els següents casos: Mundial, Estats Units i Espanya.

5.1.1.1 Índex EPU Global

Davis(2016), construeix el indicador d'incertesa política-econòmica a nivell mundial (GEPU), basant-se amb la metodologia duta a terme per Baker et al.,(2015).

La metodologia que segueix l'autor per a construir el indicador GEPU es basa en, com a primer pas, normalitzar cada indicador EPU a nivell nacional (16 països) per tal de que s'obtingui un valor de 100 en el període compres entre 1997 fins al 2015. Degut a que no tots els països tenen la informació corresponent des del 1997, s'imputen els valors que falten basant-se en mètodes de regressió. Per exemple en el cas d'Espanya, s'obtenen observacions des del 2001, de manera que el autor realitza una regressió del indicador EPU per aquesta economia des del 2001 en endavant amb els valors de tots els altres països que no tenen valors mancants. Els valors predits s'utilitzaran per imputar els valors d'Espanya entre el 1997 i el 2001.

Al següent gràfic es pot observar la evolució temporal del índex d'incertesa politico-econòmica a nivell mundial en el període de temps compres entre el 1997 i el 2020. Tenint en compte aquest eix temporal, es pot observar que hi ha diferents pics on hi hagut un augment de la incertesa, com son: l'atemptat a Nova York el 2011, la invasió d'Iraq el 2003, la crisi financera del 2007, la crisi europea al 2011, el Brexit de Regne Unit, les eleccions en les quals Donald Trump era candidat a la presidència d'Estats Units i finalment la pandèmia deguda al Covid – 19 al 2020.

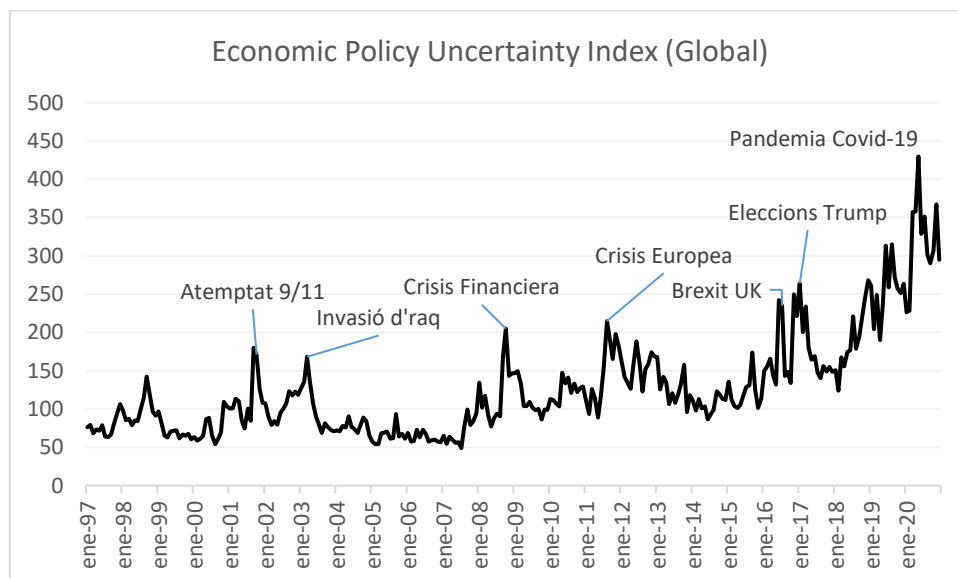


Figura 1: Evolució de l'Índex d'incertesa politico-econòmica Global
Font: Pagina web: www.policyuncertainty.com. Elaboració pròpia

5.1.1.2 Índex EPU a Estats Units

Per a construir el indicador d'incertesa política-econòmica per a l'economia d'Estats Units, es construeix el índex a partir de tres components principals, Baker et al., (2015).

El primer component quantifica la incertesa econòmica relacionada amb les polítiques dins de l'àmbit periodístic. Aquest indicador s'obté a partir dels resultats de la recerca d'informació de política econòmica a deu diaris seleccionats: *USA Today*, *Miami Herald*, *Chicago Tribune*, *Washington Post*, *Los Angeles Times*, *Boston Globe*, *San Francisco Chronicle*, *Dallas Morning News*, *New York Times* i *Wall Street Journal*. Les paraules que es tenen en consideració per a la construcció d'aquest índex aquells articles que continguin: "Incertesa", "economia" i un dels següents termes de política: "congres", "dèficit", "Reserva Federal", "Casa Blanca", etc.. És a dir, els seus criteris son que un article de diari ha de contenir paraules relacionades amb les categories: política, economia i incertesa.

La construcció d'aquest índex a dur a terme amb la següent manera:

Primer de tot es calcula per a cada diari (i) i cada mes (t) la proporció d'articles relacionats amb la incertesa política-econòmica:

$$n(i, t) = \frac{\# \text{ Articles relacionats amb la incertesa politica - econòmica}}{\# \text{ Articles totals publicats}}$$

Posteriorment, per a cada sèrie a nivell de diari, s'estandarditza el valor obtingut, de manera que s'obtingui una desviació estàndard unitària per el període compres entre el gener de 1985 fins al desembre de 2009 i es realitza la mitjana dels deus diaris per a cada mes. Un cop s'obté la sèrie dels 10 diaris, es realitza una normalització de manera que aquesta sèrie tingui una mitjana de 100 en període compres entre el 1985 a 2009.

El segon component es basa en informes de la Oficina de Pressupost del Congres (CBO)¹, ja que es considera que les mesures tributàries generen incertesa per els agents econòmics.

El tercer component utilitza el desacord entre les expectatives econòmiques com a indicador d'incertesa, a partir de les enquestes dutes a terme per professionals del Banc de la Reserva Federal de Filadèlfia. On mesuren la dispersió a nivell individual per les següents variables (prèviament seleccionades ja que estan influenciades per la política monetària i les accions de la política fiscal): el IPC, compra de bens i serveis realitzats pels governs estatals/locals i la compra de bens i serveis realitzats pels governs federals. Baker et al., (2015).

¹ La Oficina de Pressupost del Congres (*Congressional Budget Office*) es una agencia federal dels Estats Units que proporciona informació econòmica i pressupostària al congres.

Per a poder construir el indicador general d'incertesa política- econòmica per a Estats Units requereix que es normalitzin cada component. L'equació final d'aquest indicador per a aquesta economia és la següent:

$$EPU = \frac{1}{2} * primer\ component + \frac{1}{6} * segon\ component + \frac{1}{3} * tercer\ component$$

Al següent gràfic es pot observar la evolució temporal del índex d'incertesa politico-econòmica en el període de temps compres entre el 1985 i el 2020 per a Estats Units. Tenint en compte aquest eix temporal, es pot observar que hi ha diferents pics, on hi hagut un augment de la incertesa, com son: l'atemptat a Nova York el 2011, la invasió d'Iraq el 2003, la crisi financera del 2007, la crisi de deute d'Estats Units el 2011, les eleccions en les quals Donald Trump era candidat a la presidència d'Estats Units i finalment la pandèmia deguda al Covid – 19 al 2020.

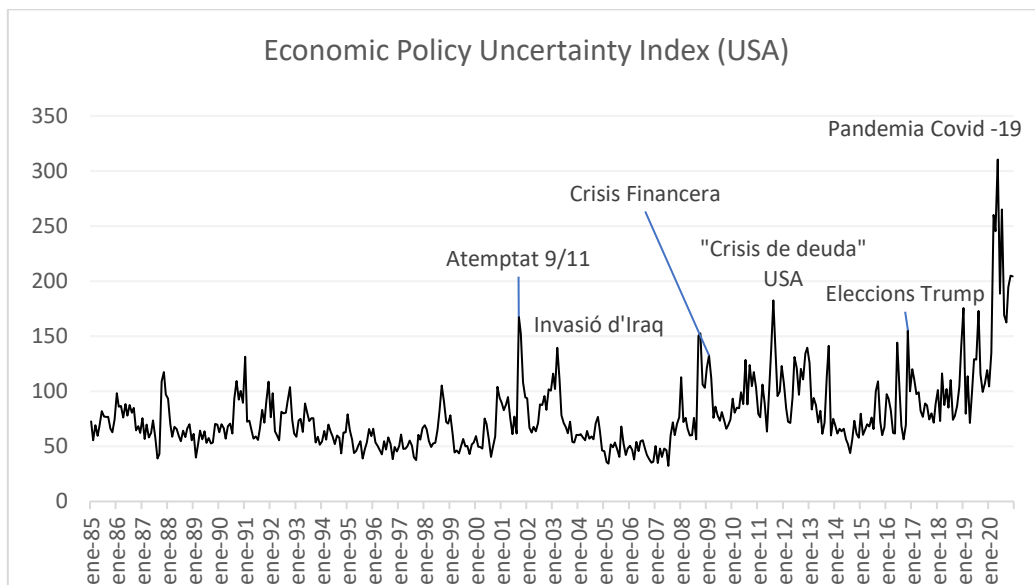


Figura 2: Evolució de l'Índex d'incertesa politico-econòmica a USA.
Font: Pagina web: www.policyuncertainty.com. Elaboració pròpia

5.1.1.3 Índex EPU a Espanya

Dins de l'àmbit de calcular l'índex d'incertesa de la política econòmica dins de l'economia d'Espanya, s'han dut a terme a partir de dos equips d'investigació independents, amb la similitud que tots dos han utilitzat la metodologia de tenir en compte els articles publicats en diaris. A continuació s'exposarà la construcció d'aquests indicadors:

Ghirelli et al., (2019) construeixen el índex d'incertesa de la política econòmica basada en la freqüència de les publicacions impreses en set diaris espanyols: *El País*, *El Mundo*, *La Vanguardia*, *ABC*, *Expansión*, *Cinco días* i *El Economista*, des de 1997 fins a la actualitat, seguint així la metodologia proposada per Baker et al. (2016), comptant cada article que mencioni alguna paraula vinculada amb "incertesa", "economia" o bé "política". Cal destacar que totes les recerques es duen a terme utilitzant el servei *Factiva Dow Jones*², restringint així tots aquells articles que tenen vinculació amb Espanya.

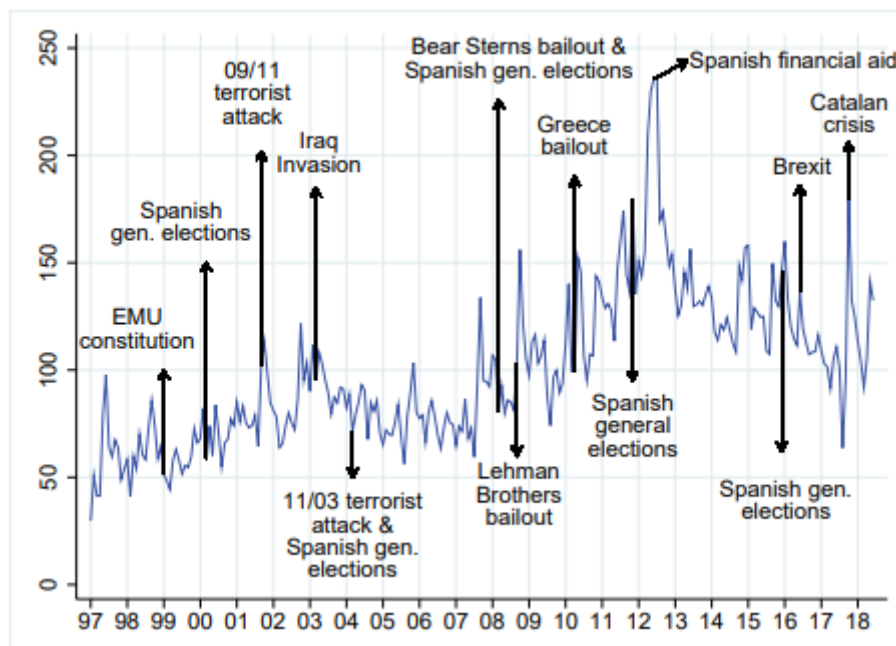


Figura 3: Evolució de l'Índex d'incertesa política-econòmica a Espanya.
 Font: Banco de España. Documento de Trabajo nº 1906

La figura 3 mostra el indicador d'incertesa de la política econòmica en l'economia d'Espanya. Aquest es pot observar que augmenta en quan va tenir lloc: l'atac terrorista el 9/11 del 2001 a Estats Units, la crisi de *Lehman Brothers* al 2006. Particularment, es pot observar que el pic més alt és al 2012, quan Espanya va demanar ajuda financera. Posteriorment s'observa que aquest presenta una tendència de disminuir, fins que torna a augmentar durant la crisi catalana el 2017.

² Factiva és una base de dades comercial d'informació de premsa, corporativa i comercial que ofereix Dow Jones. Aquesta inclou unes 35.000 fonts de notícies procedents de 200 països en 26 idiomes diferents.

Metodològicament les dos propostes divergeixen en que el cas de Baker et al., (2016) proposava dos diaris en el cas d'Espanya (El Mundo i El País) en comparació amb els 7 diaris que proposen els altres autors. Per altra banda també cal destacar que aquests últims, consideren altres paraules a analitzar, fet que provoca que la cobertura de paraules a estudiar sigui més enriquidora, adaptant-la així, al cas espanyol.

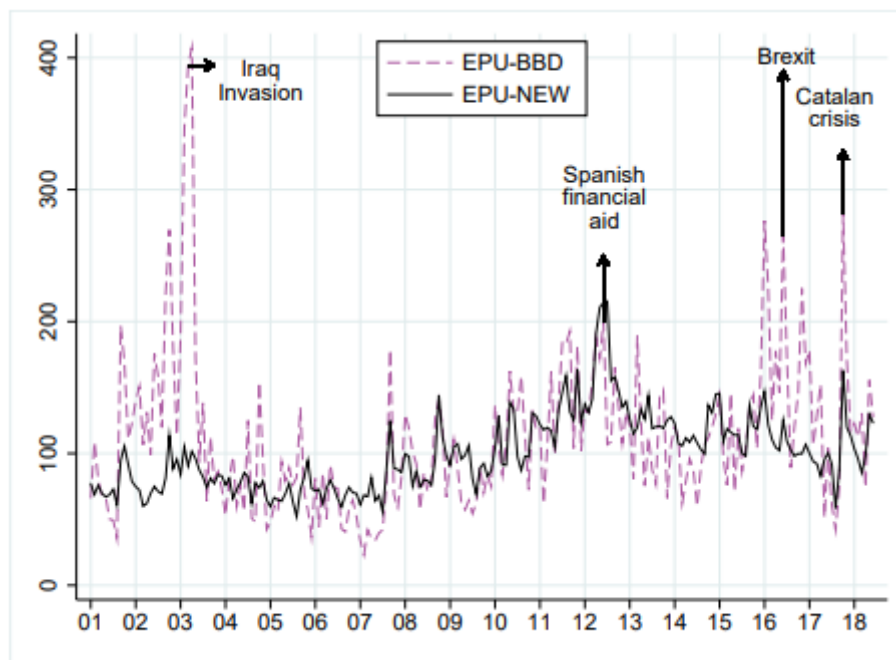


Figura 4: Comparativa dels dos indicadors d'incertesa espanyola.
Font: Banco de España. Documento de Trabajo nº 1906

La figura 4 mostra la comparativa dels dos indicadors que s'han realitzat per a estudiar la incertesa politicoeconòmica en l'economia espanyola. EPU-BBD representa el EPU construït per Baker et al.,(2016) i el EPU-NEW es el que esta construït per Ghirelli et al., (2019). Tal i com es pot observar, el EPU-BBD te una major oscil·lació, i per tant, es més volàtil que el EPU-NEW. Mostrant així, que els dos indicadors no coincideixen en tots el punts auges, com es el cas de la invasió d'Iraq, on el EPU-BBD te el seu punt màxim en el període estudiat, i en canvi el EPU-NEW, no s'observa aquest fet.

A partir del 2004, es pot observar que ambdós indicadors segueixen una tendència semblant, tot i mostrant certes diferències en pics d'incertesa. Cal destacar que el EPU-NEW es troba per sobre del EPU-BBD en el moment del rescat financer que va tenir el país.

5.1.2 Geopolitical risk (GPR)

Caldara i Iacoviello (2018) construeixen el índex de risc geopolític basant-se en la metodologia que van dur a terme *Baker, Bloom i Davis. (2015)* Saiz i Simonsohn (2013), on construeixen un algoritme que compta la freqüència d'articles relacionats amb el risc geopolític a diaris internacionals publicats a Estats Units, Regne Unit i Canadà. A continuació s'esmenta les assumpcions que aquests autors han pres i es veurà amb més detall la metodologia que han dut a terme.

La **geopolítica** es un mètode d'estudi que pretén veure els efectes de la política exterior per entendre, explicar i predir el comportament polític internacional a través de variables geogràfiques.

Aquests autors defineixen el **risc geopolític** com aquell que està associat amb guerres, atacs terroristes o bé tensions entre dos estats que afecten la normalitat de les relacions internacionals.

Com s'ha comentat anteriorment, la metodologia que han dut a terme per a construir aquest índex s'ha basat en comptar el nombre de vegades que sortia en onze diaris qualsevol paraula que tingués relació amb les següents categories:

- *Geopolitical threats* (amenaces geopolítiques): aquest primer grup inclou paraules que mencionen explícitament el risc geopolític o bé, paraules que descriuen tensions entre Estats Units amb altres països del món. Per exemple: "*Geopolitical, Risk, concern, tension, uncertain...*".
- *Nuclear threats* (amenaces nuclears): aquest grup conté paraules relacionades amb les amenaces nuclears. Per exemple: "*Nuclear war, atomic war, nuclear missile...*".
- *War threats* (amenaces de guerra): aquest grup conté paraules relacionades amb amenaces de guerra. Per exemple: "*War risk, risk of war, fear of war...*"
- *Terrorist threats* (amenaces de guerra): aquest grup conté paraules relacionades amb amenaces terroristes. Per exemple: "*Terrorist risk, menace of terrorism, terror risk...*"
- *War acts* (actes de guerra): aquest grup conté la cobertura de premsa d'actes reals relacionats amb la guerra des del punt de vista negatiu. Per exemple: "*Beginning of war, outbreak of the war, start of the war...*"
- *Terrorist acts* (atacs terroristes): aquest grup conté la cobertura de premsa d'actes reals relacionats amb atacs terroristes des del punt de vista negatiu. Per exemple: "*Terrorist act, terrorist acts...*"

Considerant els sis grups esmentats anteriorment, es pot diferenciar entre aquells que parlen de amenaces, es a dir, possibles esdeveniments que poder succeir i actes, de manera que aquests ja estan tenint lloc en el mon internacional. Es per aquest motiu que Caldara i Iacoviello (2018) decideixen diferenciar dos índexs.

- *Geopolitical threats index (GPT)* on es tindran en compte només aquells articles que tinguin a veure amb possibles amenaces.
- *Geopolitical acts index (GPA)* on es tindran en compte aquells articles que parlin de fets.

La construcció d'aquest índex a dur a terme amb la següent manera:

$$GPR = \frac{\# \text{ Articles que contenen les paraules d' estudi}}{\# \text{ Articles totals publicats}}$$

Posteriorment, aquest índex es normalitza per tenir un valor de 100 en el període del 2000-2009.

El GPR índex està disponible per els següents països: Aràbia Saudí, Argentina, Brasil, Colòmbia, Corea, Hong Kong, Índia, Indonèsia, Israel, Malàisia, Mèxic, Filipines, Rússia, Sud-Àfrica, Tailàndia, Turquia, Ucraïna, Veneçuela i Xina.

A continuació, es mostra les gràfiques corresponents al índex de risc geopolític i els seus dos subíndexs, esmentats anteriorment:

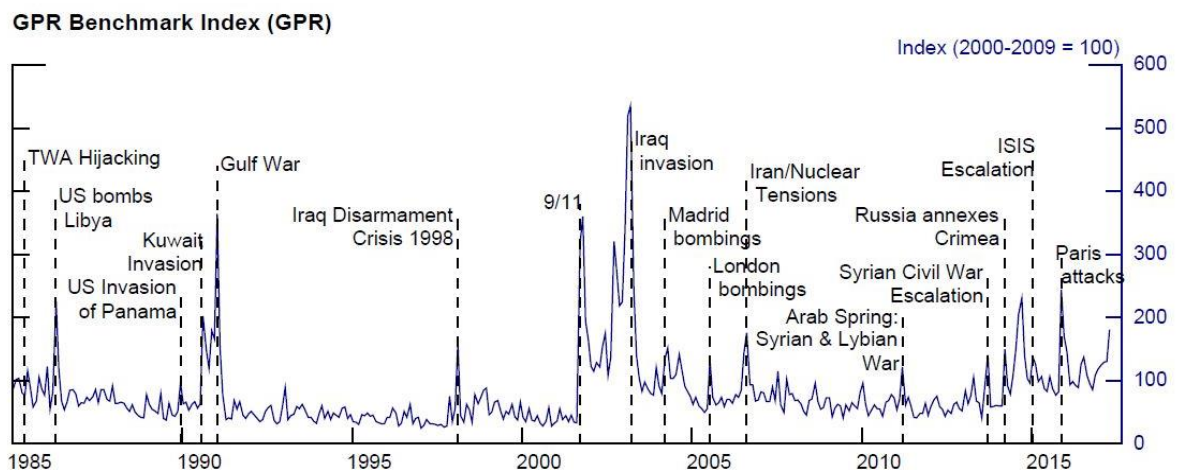


Figura 5: Geopolitical Risk Index.
Font: www.policyuncertainty.com/gpr.html

GPR THREATS AND ACTS

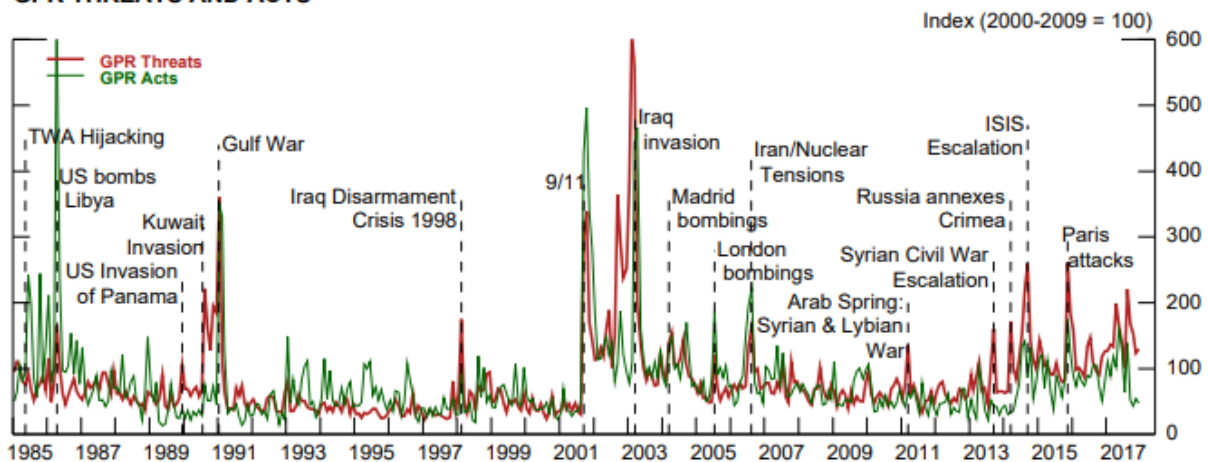


Figura 6: Geopolitical Risk index: els dos subíndex.
Font: Measuring Geopolitical Risk, Caldara and Matteo.

Tenint en compte la figura 5 es pot observar que hi ha tres grans esdeveniments que van provocar que augmentes el indicador del risc geopolític.

El primer escenari es el cas de la guerra del Golf (1991). Després aquest índex es redueix fins que te lloc l'atemptat de les torres bessones a Nova York el 2001, que provoca així un gran augment en el índex. Fruit d'aquest fet els valors del índex es mantenen relativament alts en comparació amb els valors abans de l'atemptat, assolint així el màxim valor durant aquest període en quan te lloc la invasió d'Iraq el 2003.

En el cas del gràfic dels dos subíndexs s'observa que aquests segueixen la mateixa tendència que l'índex de risc geopolític, és a dir, aquests indicadors son relativament alts en el cas de la guerra del Golf, atemptat 9/11 i la invasió de Iraq.

Ara bé, cal destacar que si comparem el índex de basat en els fets (*GPR Acts*) respecte el índex de risc geopolític (*GPR*), el primer te el seu valor màxim el 1986, quan Estats Units va bombardejar Líbia, mentre que el segon aquest te lloc el 2003 amb la invasió d'Iraq.

Cal esmentar que, tal i com indiquen Caldara i Iacoviello (2018) i tenint en compte la metodologia de construcció del índex, aquest es veu influenciat per la cobertura que els diaris atorguin els esdeveniments geopolítics. De manera que pot provocar fluctuacions en el índex GPR ja sigui bé per falta de cobertura o bé per la constància d'aquests fets al llarg de diferents publicacions. Com podria ser el cas del atemptat del 11 de setembre del 2001, on els alts nivells del podrien estar influenciats més aviat per la por que aquest atemptat va ocasionar mundialment que no pas pel risc real.

Un altre fet que pot afectar en la cobertura dels temes geopolítics, i mencionat anteriorment, es la falta de cobertura a la premsa produïda per altres esdeveniments d'interès periodístics, com per exemple, fets esportius o bé desastres naturals.

A continuació es mostra una gràfica on recopila la informació sobre com les publicacions a la premsa poden afectar la construcció del índex. En els dos gràfics es mostra el índex GPR i un índex alternatiu que s'ha construït tenint en compte els articles que han sortit als diaris vinculats amb esdeveniments naturals i esportius, respectivament.

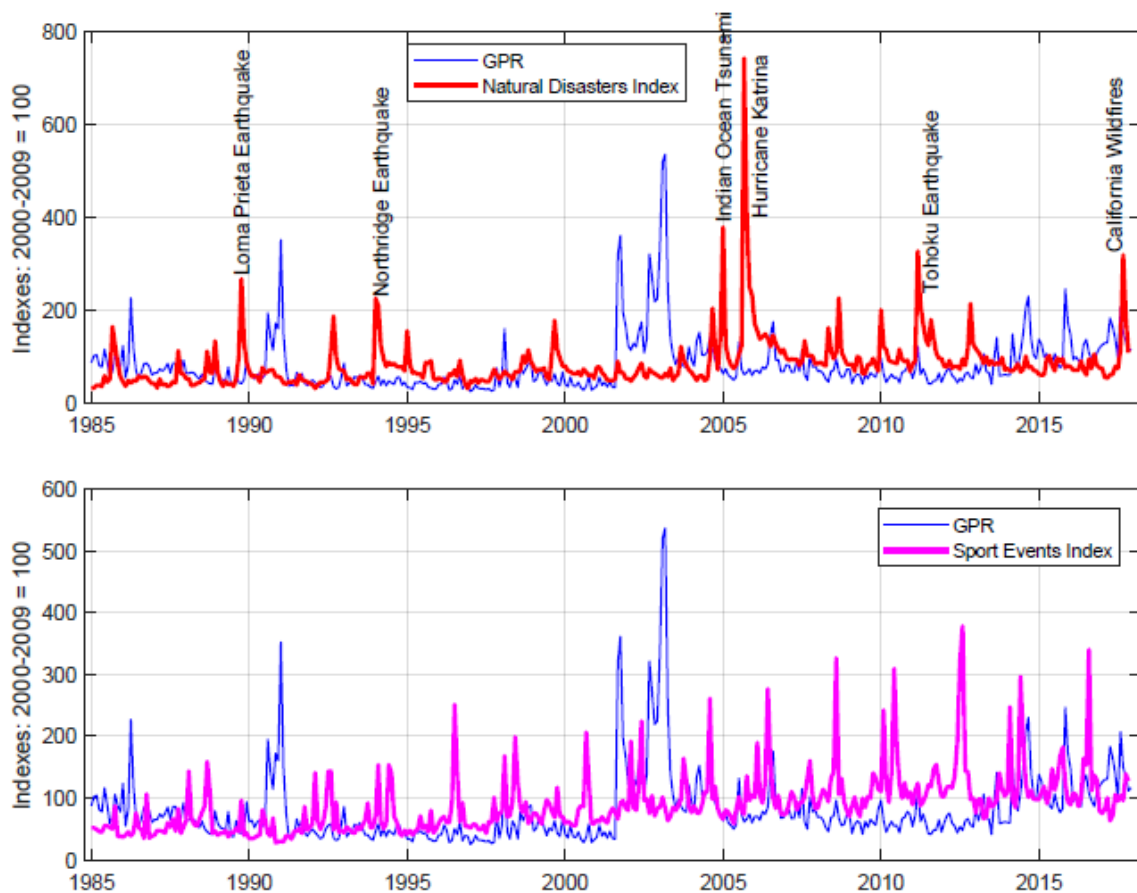


Figura 7: GPR i la influència de la premsa.
Font: *Measuring Geopolitical Risk*, Caldara and Matteo

Tal i com s'observa en el gràfic es pot observar que la cobertura d'articles que parlen de temes polítics es veu influenciada per si en aquell estan tenint lloc altres esdeveniments d'interès periodístic. Tot i que aquells punts on el GPR representa un índex més elevat, entre 2000 – 2005, les mencions a la premsa d'altres esdeveniments es mostren per sota dels articles que tenen relació amb els temes vinculats amb la construcció del índex de risc geopolític.

5.1.2.1 Altres indicadors de risc geopolític: BlackRock Geopolitical Risk Indicator (BGRI)

Una de les iniciatives que ha pres *BlackRock Investment Institute*,³ és crear el baròmetre que mesura els riscos geopolítics. Aquest baròmetre es coneix com *BlackRock Geopolitical Risk Indicator* (BGRI). Aquest indicador analitza una sèrie de dades històriques que es revisen mensualment amb l'objectiu d'identificar un top 10 de riscos geopolítics.

Els autors defineixen els *shocks* geopolítics com els esdeveniments que tenen a veure amb atacs terroristes, guerres o bé altres actes que provoquen que augmentin les tensions entre estats i que afecten la normalitat de la política interna del propi país i les relacions internacionals.

Tal i com indica la pròpia empresa, *Geopolitical risk dashboard (2021)*, els components qualitius del seu indicador de risc geopolític incorporen dos mesures de risc diferents: la primera es basa en l'atenció del mercat als possibles esdeveniments de risc i la segona correspon al moviment del mercat relacionat amb aquests esdeveniments propis.

Per al primer component, és a dir, l'atenció al del mercat es calcular la freqüència de les paraules relacionades amb el risc geopolític (com per exemple: conflicte) en diferents informes d'analistes, notícies financeres i *tweets* de comptes verificades, ajustant a través de sentiments positius com negatius a les quals se li assigna una puntuació. Per tal de classificar si una paraula està escrita amb sentiment positiu o negatiu, tenen com a referència un diccionari d'aproximadament 150 paraules respectivament com a mostra, on per exemple, es considera sentiment positiu la paraula "millora" i, com a sentiment negatiu la paraula "caiguda".

La puntuació, a la qual aquella vinculada amb el sentiment se li assigna un pes del 20%, reflexa el nivell d'atenció que se li dona a cada risc en comparació amb un historial de cinc anys. Seguidament apliquen tècniques d'aprenentatge automàtic que a través d'analitzar informes i notícies financeres realitzen els següents passos:

- Classificar la rellevància de cada frase per el risc geopolític individual per a poder generar una puntuació.
- Classificar el sentiment de cada frase per a generar una puntuació de sentiment.

Aquestes dos puntuacions, la d'atenció i sentiment s'agreguen per tal d'obtenir una puntuació de risc geopolític compost.

³ BlackRock Investment Institute genera investigació pròpia per proporcionar informació sobre la economia global, els mercats, la geopolítica i la assignació d'actius a llarg termini.

Respecte el segon component, moviments del mercat, utilitzen diferents escenaris impulsats per el mercat (*Market-Driven Scenario, MDS*)⁴ que estan associats amb cada esdeveniment del risc geopolític. Un cop obtenen els resultats d'aquests escenaris, a través de models econòmètrics troben l'equivalència d'aquests resultats en diferents *shocks* d'indicadors globals del mercat i factors de risc. A continuació compilen un índex de moviment del mercat per a cada risc, on aquest esta compost de dues parts, la semblança i la magnitud.

On seguint amb la línia amb el que els autors indiquen, la semblança recull com de semblant es el entorn actual del mercat en comparació a com aquest es veuria afectat si s'apliqués un *MDS*. Per altre banda, la part de la magnitud recull l'impacte dels rendiments d'un mes enrere dels actius del *MDS* rellevants, *Geopolitical risk dashboard (2021)*.

El *BlackRock Investment Institute* consta amb un anàlisi històric de 68 esdeveniments claus que han tingut un gran impacte dins del àmbit de risc geopolític des del 1962. Aquest llistat es distribueix en tres grans grups de risc geopolític: dates concretes de possibles eleccions – referèndums, riscos exògens, és a dir, aquells que no es podrien preveure, com son atacs terroristes i per últim, els riscos temàtics, que son aquells que tenen constància durant un temps prolongat, com pot ser la tensió entre dos països. En el llistat esmentat, es poden detectar els següents esdeveniments: Referèndum del Brexit (2016), atac terrorista Nova York (2001), accident nuclear de Fukushima (2011), guerra de Iran – Iraq (1980), etc...

A partir del estudi realitzat per aquests investigadors, *BlackRock (2019)*, es que per aquells esdeveniments inesperats, la resposta que implica en el mercat, ha estat relativament reduït i de poca duració, posant com exemple el cas del atemptat del 11 de setembre del 2001, on el índex de S&P 500 va caure casi un 12% a la primera setmana posterior al atac, però que el mercat de valors als vint-i-cinc dies hàbils posteriors al esdeveniment, ja havia recuperat les totes les pèrdues.

El objectiu clau es com avaluar el panorama geopolític actual i com els possibles riscos futurs poden influir en els mercats. Els riscos geopolítics tenen quatre pilars clau: identificar els riscos geopolítics, on un cop detectats, es procedeix a analitzar-los. La finalitat es mesurar quina es la seva probabilitat relativa i el seu impacte potencial. Finalment, el pas final es aplicar el coneixement obtingut a les carteres.

⁴ Per veure més detall sobre com es construeixen aquests escenaris, veure el document: *Market Driven Scenarios: An Approach for Plausible Scenario Construction*, (2018), *BlackRock*.

A continuació es mostra un gràfic del indicador creat per BlackRock:

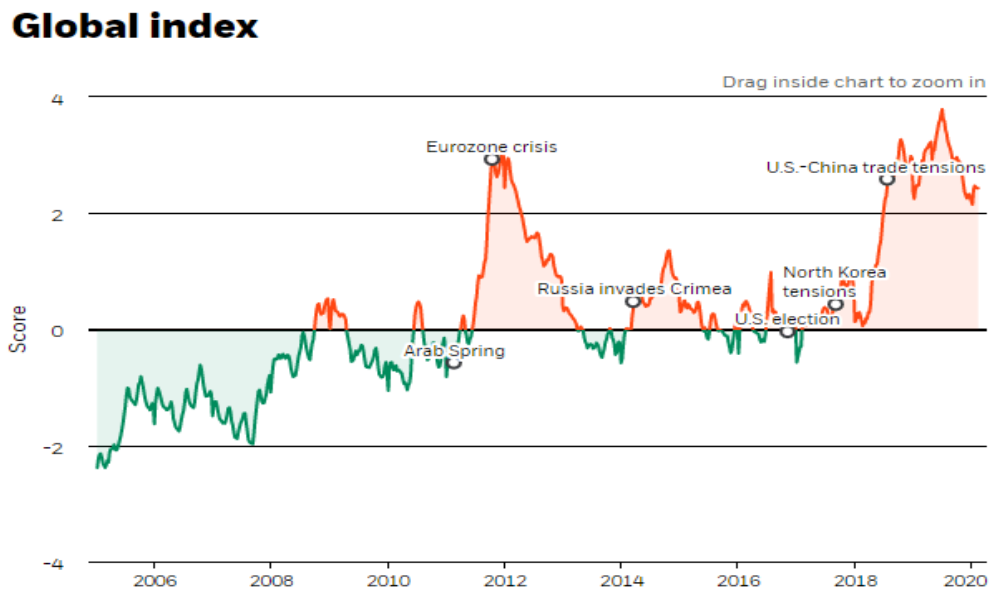


Figura 8: BlackRock Geopolitical Risk Indicator, 2005-2020.
Font: BlackRock Investment Institute

Tenint en compte el gràfic anterior, la puntuació reflecteix el nivell de l'atenció en el mercat de cada risc versus cinc anys d'històric. El gràfic corresponent també mostra si les paraules identificades mostren un sentiment positiu o bé negatiu, assignant així una puntuació corresponent. En aquest cas, una puntuació d'un punt significa que el nivell del indicador BGRI es una desviació estàndard per sobre de la mitja. Per tant, com es pot observar, la crisi financera europea va representar sentiments negatius, implicant així una gran diferencia respecte la mitjana.

5.1.3 Monetary Policy Uncertainty (MPU)

Husted et al., (2017) construeixen el índex de la incertesa de la política monetària (MPU) basant-se en la freqüència dels articles publicats a través de ProQuest⁵. Construeixen aquest índex buscant paraules clau publicades pels diaris: *Washington Post*, *Wall Street Journal*, *New York Times* relacionades amb la incertesa de la política monetària, buscant així, articles que contenen les següents paraules: "incertesa", "política monetària", "taxa d'interès", "Reserva Federal", etc... La periodicitat d'aquest indicador es mensual i es realitza des del 1985.

Com indiquen els autors, per tal de precisar en el recull d'articles, posen com a màxim una distancia entre les paraules que tenen en consideració, és a dir, restringeixen que la paraula "incertesa" estigui a 5,10 o 20 paraules de "Reserva Federal", per exemple.

⁵ ProQuest és una companyia editorial que publica recursos en formats electrònics.

La metodologia que segueixen tenint en compte aquests criteris de selecció d'articles és basa, com a primer pas, fer un recompte de quants articles publicats contenen els criteris de selecció esmentats anteriorment i un recompte que per a tots aquells articles que mencionen la Reserva Federal, on es realitza l'equació d'ambdós indicadors per a cada diari en un període de temps determinat:

$$n(i, t) = \frac{\# \text{ articles relacionats amb el MPU } (i, t)}{\# \text{ articles que mencionen la Reserva Federal } (i, t)}$$

La proporció d'articles es normalitza posteriorment, amb l'objectiu de tenir una desviació estàndard unitària per a cada diari durant el temps de la mostra, Husted et al., (2017). Es realitza de la següent manera:

$$nn(i, t) = \frac{n(i, t)}{stdev(n(i, 1985: 2015))}$$

Finalment, el indicador de incertesa de la política monetària (MPU) es construeix com es mostra a continuació:

$$MPU(t) = \left[\frac{\sum_i nn(t)}{avg(\sum_i nn(1985: 2015))} \right] * 100$$

Com es pot observar al següent gràfic, es mostra l'evolució del indicador de política econòmica. Si comparem l'evolució d'aquest amb els altres indicadors de risc econòmic es pot observar com els tres indicadors esmentats, coincideixen en els punts auge que van tenir lloc en quan va succeir l'atemptat del 11/9 a Nova York, la invasió d'Iraq, Brexit. Cal esmentar però, que aquest últim indicador divergeix respecte els altres, amb les guerres comercials entre Estats Units i Xina, aquest fet podria estar explicar tenint en compte les paraules que formen part de la metodologia del propi indicador.

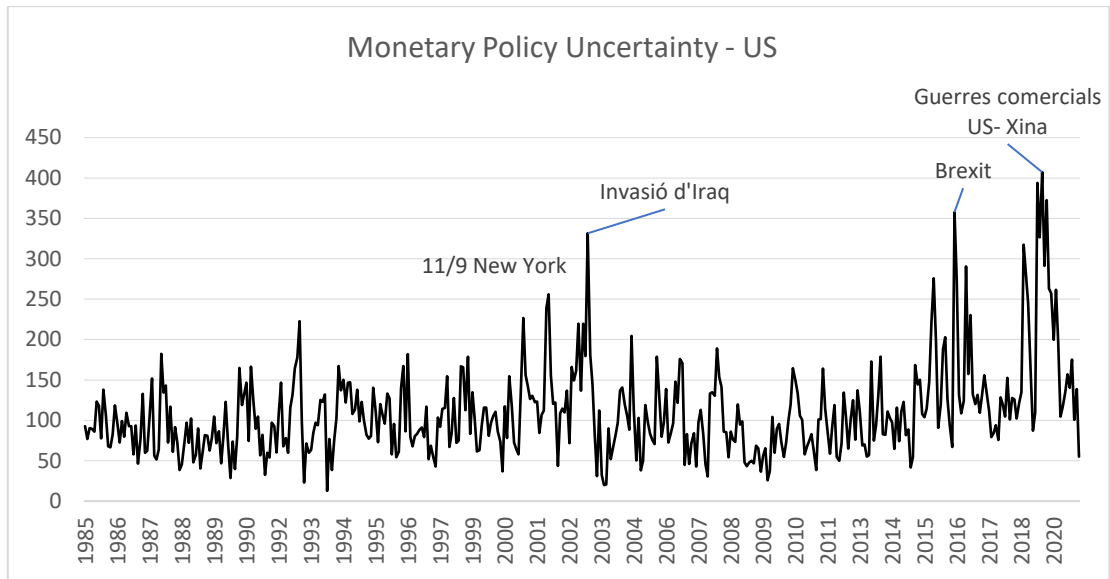


Figura 9: Evolució de l'índex d'incertesa política monetària.

Font: Pagina web: <https://www.policyuncertainty.com/monetary.html>. Elaboració pròpia.

6. Comparativa entre indicadors d'incertesa

Tal i com realitzen Caldara i Iacoviello (2018) es mostra una comparativa del índex geopolític respecte els indicadors d'incertesa: el índex d'incertesa politicoeconòmica (EPU) i l'índex de volatilitat (VIX), ja esmentats anteriorment.

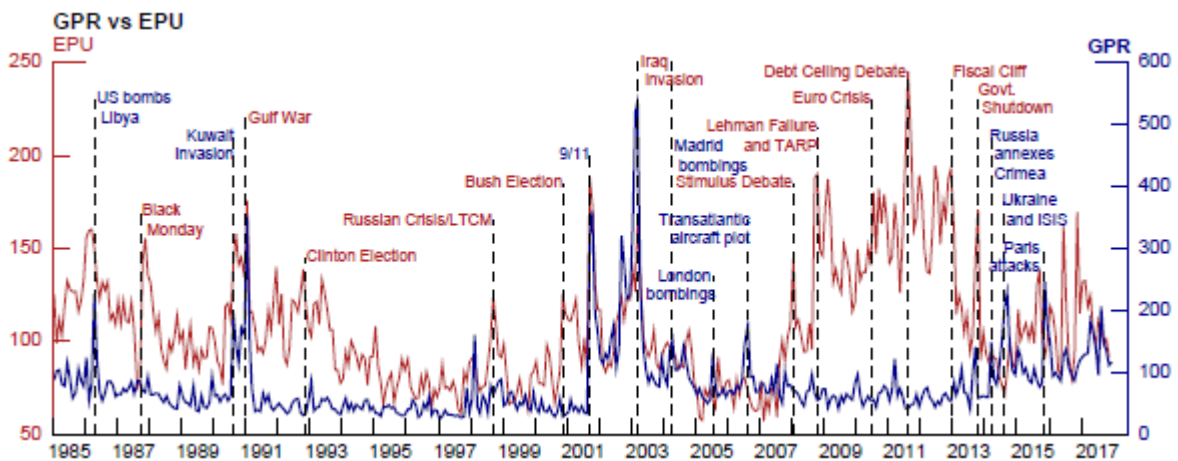


Figura 10: Comparativa GPR – EPU.

Font: Measuring Geopolitical Risk, Caldara and Matteo

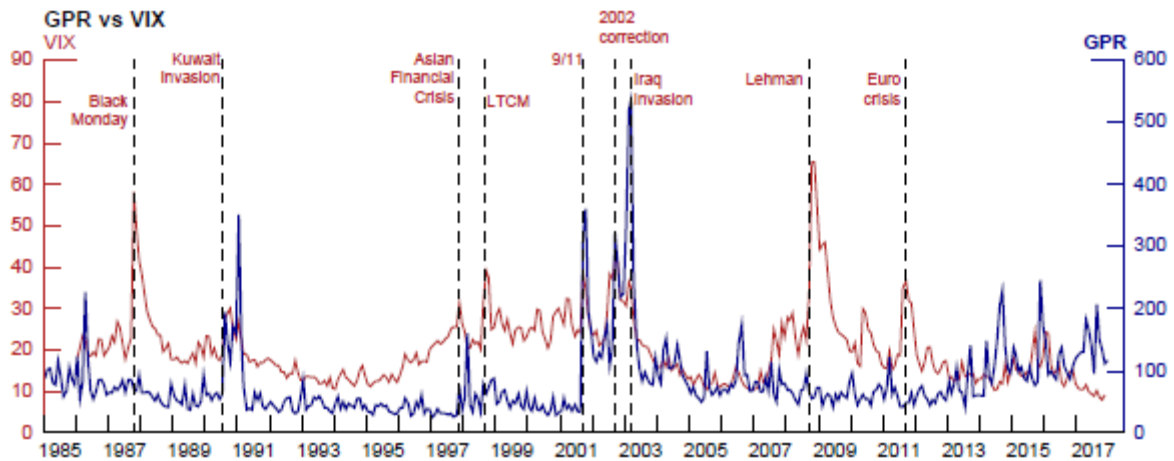


Figura 11: Comparativa GPR - VIX.
 Font: *Measuring Geopolitical Risk*, Caldara and Matteo

Durant el període establert entre el 1985 – 2017, es pot observar que els tres índex comparats tenen una concordança en diferents punts d'auge. Aquests tenen lloc a l'atemptat del 9/11 i a la invasió d'Iraq el 2003. Tot i així es pot observar que no sempre el GPR augmenta en quan el mercat presenta elevats valors d'incertesa politicoeconòmiques o bé dificultats econòmiques i financeres, ja que en la major part del període analitzat el GPR queda per sota del altres indicadors, com es el cas del període 2007 – 2013, moment on va tenir lloc la crisi financera global.

7. Efectes dels indicadors d'incertesa en el mercat.

A continuació es mostra quins son els efectes dels indicadors esmentats anteriorment en el mercat.

7.1 Efectes del EPU

S'han realitzat diferents estudis per estudiar quina es la relació que hi ha entre el índex d'incertesa politicoeconòmica i el mercat. Baker et al., (2016) troben que el EPU condueix a una disminució de les inversions i l'ocupació, impacte que ja s'ha mencionat anteriorment en quan s'han esmentat quins eren els efectes de la presència d'incertesa en un mercat. Aquests autors ressalten que el efecte es mes pronunciat depenent del sector de l'economia a tractar, com pot ser, el sector polític, defensa, salut, financer...

Altres estudis demostren que la incertesa politicoeconòmica pot afectar els preus dels actius, de diferents formes. Una d'ella es que el EPU es relaciona negativament amb les inversions de capital. Kang et al., (2013) estudien els *shocks* de la política econòmica te un efecte advers en les inversions a nivell empresa . Per altre banda Liu i Zhang (2015) investiguen la predicció del

EPU respecte la volatilitat del mercat de valors, on s'observa que un valor elevat del EPU provoca una major volatilitat. Aquests autors defensen que incloure el EPU com a paràmetre de volatilitat podria arribar a millorar significativament les capacitats de predicció de la volatilitat del mercat.

Brogaard i Detzel (2015) estudien la relació que hi ha entre el EPU i els rendiments de la borsa, arribant a la conclusió que el EPU redueix significativament la rendibilitat de les accions i que aquesta relació es persistent en períodes d'alta volatilitat i ressalten en la importància que té el EPU com a factor de risc per la renda variable.

7.1.1 La incertesa i els seus efectes macroeconòmics - cas espanyol

Ghirelli et al.,(2019) van realitzar un anàlisi sobre el impacte d'una major incertesa macroeconòmica sobre la activitat econòmica espanyola mesurada a partir del PIB, el consum privat i la inversió. On a partir d'un model estructural de valors autoregressius (VAR model) on inclouen com a variables endògenes tres indicadors sintètics d'incertesa que ells mateixos construeixen a partir de la tècnica de d'anàlisi de components principals (PCA):

- Indicador sintètic d'incertesa financera.
- Indicador sintètic d'incertesa basat en el desacord econòmic.
- Indicador de incertesa de la política econòmica.

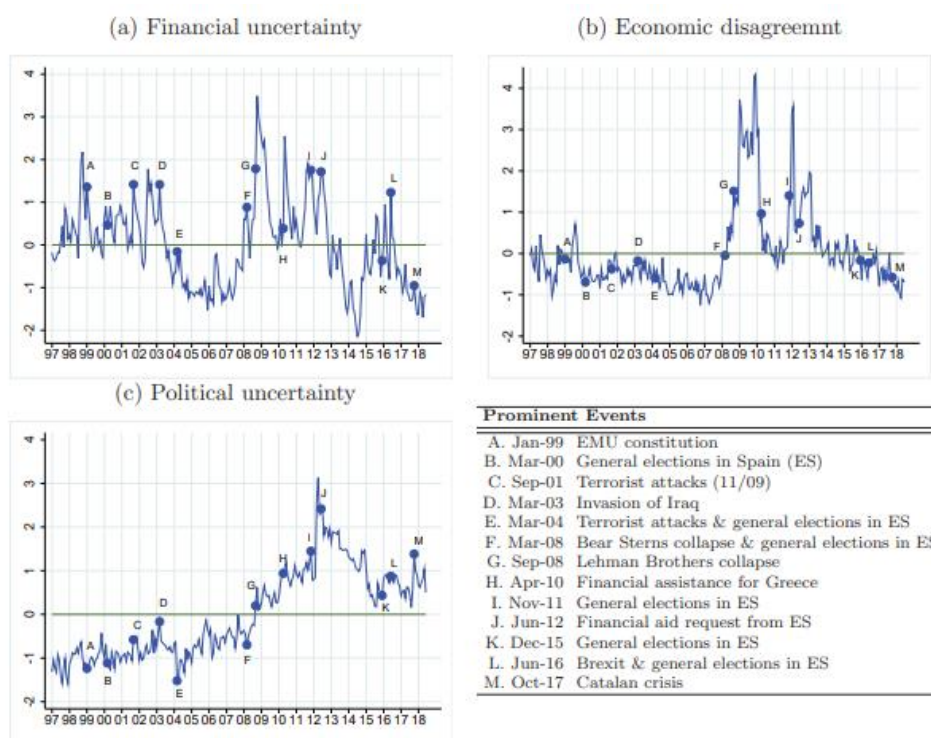


Figura 12: Indicadors sintètics d'incertesa.
Font: Banco de España. Documento de Trabajo nº 1905.

Adicionalment, afegixen el índex EPU calculat per Baker et al.,(2016), la volatilitat del Eurostoxx-50 com a variables exògenes del model.

La figura 5 mostra com responen els indicadors sintètics que els autors han construït davant de diferents esdeveniments importants per la economia espanyola. Aquests indicadors no responen de la mateixa manera als esdeveniments. Per exemple l'indicador sintètic d'incertesa financera respon davant de fenòmens com pot la crisi bancaria de Lehman Brothers però no reacciona davant esdeveniments polítics, com es la crisi a Catalunya el 2017.

A partir de diferents proves que s'apliquen al model, arriben a concloure que els *shocks* d'incertesa afecten significativament la dinàmica de la economia espanyola, del consum privat i la inversió de bens de capital. De manera que un augment inesperat de la incertesa provoca una disminució del consum privat, la inversió en bens de capital i una disminució del PIB.

7.2 Efectes del GPR

Un tema recurrent d'estudi és la relació que hi ha entre la política i variables econòmiques, es per aquest motiu que al llarg dels últims anys, aquests anàlisis han anat augmentant ja que avui en dia varis economistes valoren la importància que presenten els riscos geopolítics per avaluar amb precisió les perspectives econòmiques.

Diferents estudis s'han centrat en analitzar els efectes de la existència de risc geopolític en els diferents cicles econòmics que presenta una economia, en els preus del habitatge, en el mercat del petroli, en el preu de l'or, la dinàmica del mercat de valors, entre d'altres.

Per exemple, Gupta et al., (2019) després d'aplicar varies tècniques d'estimació, conclouen que el índex de risc geopolític (GPR) esta associat negativament amb el **volum comercial**⁶ en 164 països.

Antonakakis et al., (2017) utilitzen el índex de risc geopolític (GPR) per investigar els efectes que provoca en el **mercat del petroli i el mercat de valors** a través del model VAR-BEKK-GARCH. Per estudiar el petroli, utilitzen el indicador WTI⁷ i pel mercat de valors el índex borsari S&P 500. On arriben a concloure que el mercat del petroli es veu més afectat per tensions polítiques que no pas el mercat de valors. Tot i així en ambdós casos, hi ha una correlació negativa entre el risc geopolític i la influencia en els mercats estudiats.

⁶ S'entén per volum comercial la quantitat de transaccions comercials que es produeixen en un determinat període de temps entre dues unitats econòmiques, per exemple, entre països.

⁷ West Texas Intermediate (WTI) es un indicador que s'utilitza com a referencia en el mercat del petroli a Estats Units.

CaixaBank Research (2018) va analitzar la relació entre el indicador de risc geopolític (GPR) i el **creixement del PIB mundial**, on un augment de la incertesa normalment està associat a una disminució del creixement del PIB, on el seu efecte es manté al llarg dels sis – dotze mesos posteriors.

Per altre banda diferents autors han volgut analitzar si un augment de la incertesa geopolítica global, té el mateix impacte tant en països desenvolupats com països emergents, entenent com països desenvolupats aquells països que gaudeixen d'un alt nivell de vida i com països emergents aquells que la seva economia es troba en ple desenvolupament econòmic. CaixaBank Research (2018) afirma que el impacte d'un *shock* geopolític té un impacte de 1,45 vegades en termes de reducció del PIB en una economia emergent que no pas a una economia avançada. A més a més també demostra que els efectes que aquest fet provoca perduren més en l'economia emergent. Per altre banda, Balcilar et al., (2016) examinen el efecte de la incertesa geopolítica sobre la **dinàmica de la renda i la volatilitat** en els mercats borsaris BRICS⁸ on afirma que tot i que aquests països siguin tots considerats emergents, el efecte del risc geopolític no afecta de manera uniforme en els països que formen part del grup estudiat. D'aquests, els països que es veuen més afectats són Rússia i Xina, en comparativa a Índia que seria el país que resisteix més a la incertesa geopolítica.

Seguint amb la línia d'estudiar les economies emergents, Cheng i Chiu (2018) utilitzant un model autoregressiu vectorial estructural (SVAR) demostren que els *shocks* del GPR poden explicar entre el 13% i el 22% les variacions de la **producció del mercat** de 38 mercats emergents.

Per altra banda Lu et al., (2020) es basa en estudiar 18 economies emergents de diferents continents amb l'objectiu d'analitzar com afecta en índex de risc geopolític (GPR) **al crèdit intern del sector financer**, arribant a la conclusió, que un valor alt del índex s'associa negativament amb els préstecs interns de les economies emergents.

Demir et al., (2019) pretenen estudiar si existeixen efectes negatius dels riscos geopolítics (GPR) sobre el **turisme**, amb una mostra de 18 països. Arribant a la conclusió que aquells països on el turisme és important per a la economia del propi país (per exemple: Tailàndia i Filipines) els riscos geopolítics tenen efectes més durs en el la producció del PIB. Demostrant així que hi ha una relació negativa entre el risc geopolític del país i el turisme.

⁸ BRICS és el acrònim d'un grup de països emergents que inclou Brasil, Rússia, Índia, Xina i Sud-Àfrica.

8. Índex de volatilitat

Tot i que existeixen altres indicadors per mesurar la volatilitat com son el VXN per el *Nasdaq 100*, el RVX per el *Russell 2000*, el VXD per el *Dow Jones Industrial Average*, el EURO STOXX 50 que fa referència a les 50 empreses europees més grans que cotitzen en borsa, el VIX en base al *S&P 500*, sent aquest més famós i utilitzat. De manera que ens els últims anys els analistes es basen en ell prenent-lo com a referència per les seves inversions i com a indicador del comportament de la borsa americana a través del sentiment de por que transmet el propi VIX al incrementar-se, sent aquest últim considerat com un referent per tot el mercat de valors a nivell mundial.

8.1 VIX

El índex de volatilitat va ser creat per el *Chicago Board Options Exchange (CBOE)* el 1993 amb l'objectiu de mesurar les probabilitats de canvi i la inestabilitat del mercat de valors. Aquest es basava en analitzar els preus del *S&P 500*⁹ durant els 30 dies següents, calculant-se així a temps real.

Deu anys mes tard, en el 2003 el CBOE juntament amb *Goldman Sachs*¹⁰, van actualitzar la metodologia del índex per reflectir una nova manera de mesurar la volatilitat esperada a partir de la agregació dels preus ponderats de les opcions *puts*¹¹ i *calls*¹² del *S&P 500* sobre una amplia gama de preus de *strike*¹³.

Tal i com indiquen a CBOE (2019) el índex de volatilitat compost de opcions, on el preu de cada opció reflecteix la expectació de la futura volatilitat. El VIX, igual que altres índexs convencionals, aplica regles per la selecció dels seus components i una formula per a calcular el valor del índex. La formula general per el càlcul del VIX és la següent:

$$\sigma^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} e^{RT} Q(K_t) - \frac{1}{T} \left[\frac{F}{K_0} - 1 \right]^2$$

On:

- T es el temps per l'expiració
- F es el índex de nivell Forward derivat del preu de les opcions.

⁹ El S&P 500 es un índex borsari que recopila les 500 empreses més grans d'Estats Units. La seva elaboració es dur a terme per la empresa Standard & Poor's, d'on provenen les sigles "S&P". El valor 500 corresponent a les empreses que hi formen part.

¹⁰ Goldman Sachs és un grup de banca i inversió situat a la ciutat de Nova York.

¹¹ Un put és un contracte d'opcions que dona al comprador el dret de vendre l'actiu subjacent en qualsevol moment fins a la data de caducitat.

¹² Un call és un contracte d'opcions que dona al comprador el dret de comprar l'actiu subjacent en qualsevol moment fins a la data de caducitat.

¹³ Strike es el preu al que es compra o es ven un actiu financer i que ve definit per l'emissor de la opció.

- K_0 es el primer strike per sota del nivell d'índex F .
- R tipo de interès lliure de risc fins la expiració.
- K_i és el preu de strike de la i -essima opció "out-of-the-money"¹⁴, una *call* si $K_i > K_0$ i una *put* si $K_i < K_0$ i tots dos casos en cas que siguin iguals.
- $Q(K_i)$ és el punt mig per el diferencial de la oferta i de la demanda per cada opció amb strike.

El índex VIX mesura la volatilitat esperada a trenta dies del índex de S&P 500. Els components del índex som opcions *call* i *put* amb més de 23 i menys de 37 dies per la seva expiració.

A continuació es mostra un gràfic de l'evolució del VIX en el període compres entre 1997 i 2020.

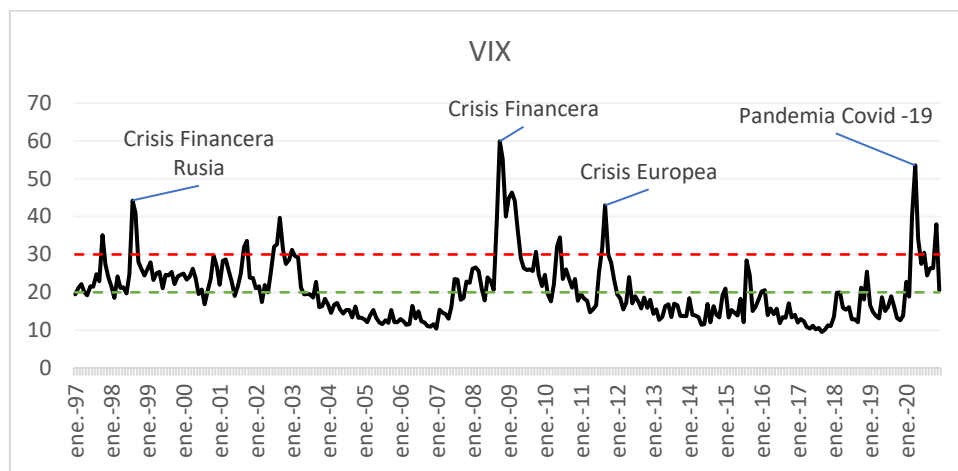


Figura 13: Evolució del VIX.
Font: Yahoo Finance. Elaboració pròpia

Cal destacar que es considera que si el valor del VIX està per sobre del 30, es considera que hi ha una gran incertesa en el mercat.

Com es pot observar, que hi ha tres períodes diferenciats on va tenir lloc una volatilitat elevada. El primer d'ells va tenir lloc als anys 1998 – 1999 on va tenir lloc la Crisis Financera Russa. El següent pic te lloc a la crisis financera del 2008, arribant així al seu punt màxim, fet que indica que hi havia una gran incertesa en el mercat, ja que aquest casi arriba a assolir un valor de 60. Consecutivament ens trobem davant la crisis financera Europea al 2012, on es torna a observar un altre augment de l'índex. Tot i que sembla que aquest augment disminueix durant el 2013 – 2019, presentant així uns nivells de volatilitat al voltant del 20, aquest fet es veu altament influenciat per la pandèmia del Covid – 19, que ha tingut lloc aquest darrer 2020, provocant així que torni a augmentar la volatilitat en el mercat i per tant la incertesa.

¹⁴ Out-of-money és una expressió que fa referència a que el preu de strike de una opció, està per sobre del seu instrument subjacent.

9. Anàlisi Previ

Un cop explicats com es construeixen aquests indicadors i quin és el seu impacte en l'economia, es procedeix a realitzar la segona part del treball, és a dir l'àmbit empíric, que consta, com s'ha dit anteriorment, d'aplicar la modelització ARIMA, així com models de xarxes neuronals per a poder realitzar la predicció de la volatilitat implícita a partir de la seva pròpia informació o bé a partir dels indicadors d'incertesa política-econòmica (EPU) i el indicador de risc geopolític (GPR).

Prèviament es realitzarà un breu anàlisi descriptiu per de conèixer la base de dades, considerant que aquesta només recull informació sobre Estats Units¹⁵, que s'utilitzarà al llarg d'aquest bloc.

9.1 Preparació de les dades

Com s'ha esmentat anteriorment, la base de dades a estudiar es una combinació dels següents indicadors:

- **Indicador d'incertesa política-econòmica (EPU):** Aquestes dades s'obtenen de la pàgina principal del propi indicador especificant el país que es vol estudiar sent en aquest cas, es Estats Units. (*Economic Policy Uncertainty Index* - https://www.policyuncertainty.com/us_monthly.html). Aquesta base de dades conté les següents variables, sent la primera observació a 01/01/1985:
 - *Day*: correspon al dia de la mostra.
 - *Month*: correspon al mes de la mostra.
 - *Year*: correspon al any de la mostra.
 - *Daily_policy_index*: variable que correspon al valor del propi indicador d'incertesa política-econòmica.

Sent ***Daily_policy_index*** la variable que formaria part de la base de dades a estudiar.

Destacar que la descarrega de la base de dades indicada en el *link* esmentat s'obté una mostra que va des de el 01/01/1985 fins a un dia abans del moment de la descarrega.

¹⁵ S'ha escollit l'economia d'Estats Units donat que és una de les economies més influents a nivell mundial. Addicionalment també es considera oportú estudiar Estats Units, ja que el índex de volatilitat (VIX) te com a referencia el índex de S&P 500, que està compost per industries del mercat nord-americà.

- **Indicador de risc geopolític (GPR):** Aquestes dades s'obtenen de la pàgina principal *Matteo Iacoviello: Geopolitical Risk Index*. <https://www.matteoiacoviello.com/gpr.htm#data>. Aquesta base de dades conté les següents variables, sent la primera observació a 01/01/1985:
 - *Date*: que correspon a la data de la observació, sent aquesta un format ["yyyymmdd"]
 - *Day_numbers*: on es fracciona un any natural, de manera que els 365 dies que te un any, seria prendria un valor 100. Per exemple: a 31/08/2016, aquesta variable té un valor de 2016,66 i a 31/12/2016, pren per valor 2016,99.
 - *N11D*: correspon al **nombre d'articles** de diaris.
 - *GRPD_THREAT_RAW*: nombre d'articles que mencionen paraules relacionades amb possibles **amenaces**.
 - *GRPD_ACT_RAW*: nombre d'articles que menciones paraules relacionades amb **fets**.
 - *GRPD_RAW*: nombre **d'articles totals** relacionats amb el **risc geopolític**. Correspon a realitzar la suma entre *GRPD_THREAT_RAW* i *GRPD_ACT_RAW*.
 - *GRPD_SHARE*: es el quocient entre *GRPD_RAW* i *N11D*.
 - *GRPD_THREAT_SHARE*: es el quocient entre *GRPD_THREAT_RAW* i *N11D*.
 - *GRPD_ACT_SHARE*: es el quocient entre *GRPD_ACT_RAW* i *N11D*.
 - *GRPD*: correspon al **indicador** diari de **risc geopolític**. Aquest es calcula a partir de *GRPD_SHARE*, i com s'ha dit anteriorment, s'estandarditza de manera que prengui valor 100 en el període compres entre 2000-2009.
 - *GRPD_THREAT*: es el component del indicador que fa referencia als articles de diaris que mencionen possibles **amenaces**. Aquest es calcula a partir de *GRPD_THREAT_SHARE*, estandarditzant el valor de manera que tingui valor 100 en el període 2000-2009.
 - *GRPD_ACT*: : es el component del indicador que fa referencia als articles de diaris que mencionen **esdeveniments**. Aquest es calcula a partir de *GRPD_ACT_SHARE*, estandarditzant el valor de manera que tingui valor 100 en el període 2000-2009.
 - *EVENT*: on mencionen esdeveniments que han suposat un impacte en el indicador. Ja sigui o bé el propi índex de risc geopolític o bé qualsevol dels dos components que hi formen part, és a dir el indicador d'amenaces com el de esdeveniments.

A continuació es mostren alguns exemples de la base de dades corresponent:

Date	Event	GPRD	GPRD_THREAT	GPRD_ACT
19910115	Gulf War - Operation Desert Storm	710	809	235
20010912	9/11 Terrorist Attacks	628	487	1348
20030318	U.S Prepares for War with Iraq	1067	1244	269
20151117	Paris Terrorist Attacks	653	678	534
20170810	North Korea Tensions	560	655	62

*Taula 1: Indicador de risc geopolític.
Font: Elaboració pròpia.*

Tenint en compte les variables mencionades anteriorment, es selecciona la **GPRD** com a indicador de risc geopolític a analitzar.

- **Índex de volatilitat del S&P 500 (VIX):** Les dades per aquest índex es poden obtenir a la pàgina web *Yahoo Finance*. <https://finance.yahoo.com/quote/%5EVIX/history?p=%5EVIX>, sent el valor històric màxim el 01/01/1990. La informació que s'obté d'aquest índex és la següent:
 - *Date:* Variable indicadora que mostra la data amb el següent format: ["dd/mm/yyyy"].
 - *Open:* variable numèrica que mesura el valor d'obertura al inici del dia del índex.
 - *High:* variable numèrica que mesura el valor màxim al que arriba el índex al llarg del dia.
 - *Low:* variable numèrica que mesura el valor mínim al que arriba el índex al llarg del dia.
 - *Close:* variable numèrica que mesura el valor de tancament al finalitzar el dia.
 - *Adj Close:* variable numèrica que mesura el valor de tancament al finalitzar el dia ajustant el índex per factors que poden afectar el preu després del tancament del mercat¹⁶.
 - *Volumne:* variable numèrica que mesura el nombre d'interseccions realitzades durant un dia. Tot i que en el cas del VIX, al ser un índex artificial que depèn dels valors de *S&P500* i no de un mercat propi, el valor de volum pren valors zero ja que no existeixen transaccions.

¹⁶ El preu d'una acció normalment es veu afectat per la oferta i la demanda dels participants del mercat. Tot i així, alguns fets, com per exemple, la divisió d'accions, els dividendes i les ofertes de drets, que afecten el preu de les accions.

Tenint en compte les variables mencionades anteriorment, es selecciona la **Adj Close** com a indicador de la volatilitat.

Observant que en els tres indicadors seleccionats no coincideix el marc històric i amb l'objectiu d'homogeneïtzar aquest es selecciona el següent històric: gener 1990 – març 2020.

Cal destacar que tant el indicador d'incertesa com el indicador de risc geopolític s'observen valors per tots els dies de l'any, en canvi el VIX només pren per valors els dies entre setmana i no festius.

Es per aquest motiu que per tal de facilitar el anàlisi del comportament estadístic, per aquells valors mancants del VIX, és a dir, caps de setmana es pren per valor el que correspon al VIX de divendres, i en el cas dels dies festius, s'agafa per valor el que correspon al dia anterior.

A continuació es pot veure un exemple de la casuística comentada anteriorment,

Data	Dia setmana	EPU	GPR	VIX
02/03/2020	Dilluns	153,53	56,31	33,419998
03/03/2020	Dimarts	132,55	178,91	36,82
04/03/2020	Dimecres	197,2	51,84	31,99
05/03/2020	Dijous	118,02	44,68	39,619999
06/03/2020	Divendres	132,5	42,43	41,939999
07/03/2020	Dissabte	122,57	56,19	41,939999
08/03/2020	Diumenge	250,63	75,18	41,939999

*Taula 2: Mostra indicadors setmana 2-8 març del 2020 .
Font: Elaboració pròpia.*

Per altre banda, tenint en compte que el VIX es tracta d'un indicador que mesura la volatilitat implícita del S&P 500 en els pròxims trenta dies, es considera convenient que el indicador EPU i GPR també estiguin expressats en la mateixa mesura, és a dir, en termes de volatilitat (en aquest cas, històrica).

Es per aquest motiu que es creen dues noves variables a partir dels indicadors corresponents per tal d'obtenir les seves volatilitats, tenint en compte una finestra mòbil de trenta dies, per tal de tenir la mateixa base que el indicador de volatilitat. Per al càlcul d'aquestes variables, s'aplica la fórmula de la desviació típica, que és la següent:

$$Vol_{EPU} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}}$$

$$Vol_{GPR} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}}$$

9.2 Anàlisi estadístic dels indicadors

En aquest primer apartat es treballarà amb la base de dades que conté dades diàries dels indicadors d'incertesa política-econòmica, indicador de risc geopolític i el VIX per a l'economia d'Estats Units tenint en compte tots els dies de l'any.

Es realitza un anàlisi preliminar de les respectives sèries temporals amb la finalitat de detectar les característiques més importants de la pròpia sèrie, tal com, la seva tendència (creixent o decreixent), la existència de cicles, etc.

Cal destacar que es realitzarà la anàlisi preliminar sobre els propis indicadors i a continuació es realitza el mateix anàlisi sobre les noves variables, és a dir, la volatilitat del indicador.

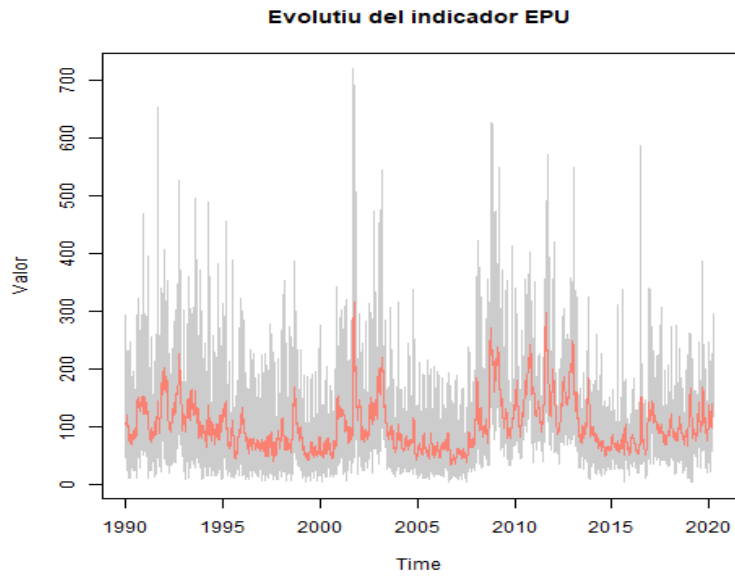
En primer lloc, es mostren els estadístics bàsics per a cada un dels indicadors corresponents:

	EPU	GPR	VIX
Min	3.32	0.00	9.14
1st Qu	54.69	37.16	13.50
Median	84.89	62.86	17.11
Mean	101.31	87.51	19.09
3rd Qu	130.50	108.15	22.41
Max	719.07	1168.88	80.86

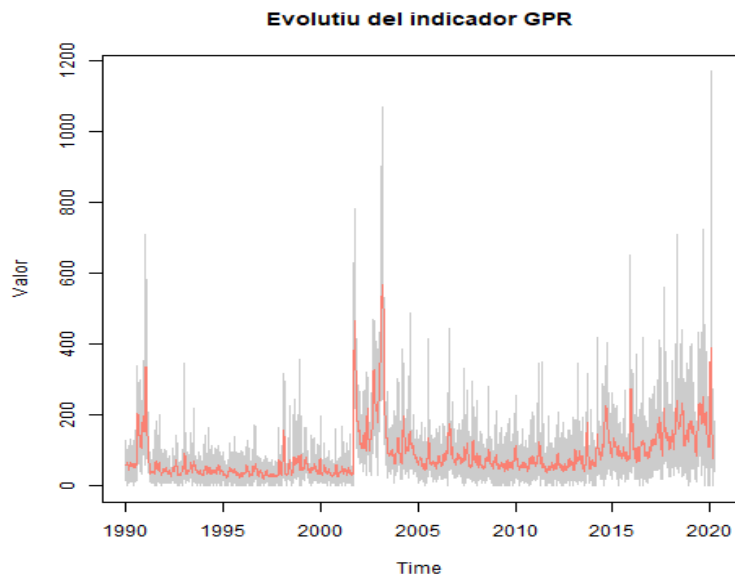
*Taula 3: Estadístics descriptius de les variables.
Font: Elaboració pròpia*

Els valors màxims de cada sèrie temporal corresponent el 18 de setembre del 2001 en el cas del EPU, que coincideix amb uns dies posterior al atac terrorista de Nova York. En el cas del GPR el màxim s'assoleix en 7 de gener del 2020, es podria arribar a pensar que està vinculat amb els inicis de la pandèmia Covid. Pel que fa al VIX el màxim va tenir lloc el 20 de novembre del 2008 que tenint en compte el context històric aquest te lloc en quan va esdevenir la crisi financera.

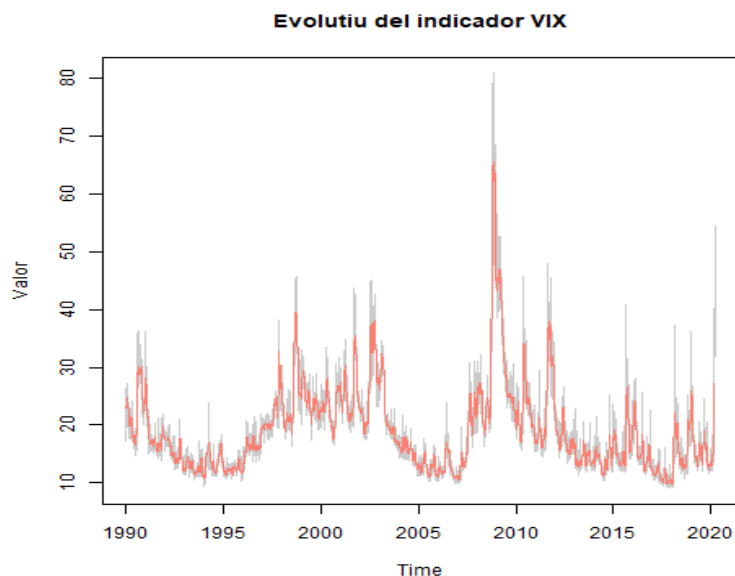
Un cop realitzat el anàlisi descriptiu inicial, es procedeix a realitzar un anàlisi gràfic, on en tots els gràfics de sèries temporals, s'ha calculat una mitjana mòbil de trenta dies amb l'objectiu de suavitzar la sèrie temporal, representada en color vermell tal i com es mostra a continuació:



*Figura 14: Evolució del indicador EPU, període 1990-2020.
Font: Elaboració pròpia*



*Figura 15: Evolució del indicador GPR, període 1990-2020.
Font: Elaboració pròpia.*



*Figura 16: Evolució del indicador VIX, període 1990-2020.
Font: Elaboració pròpia.*

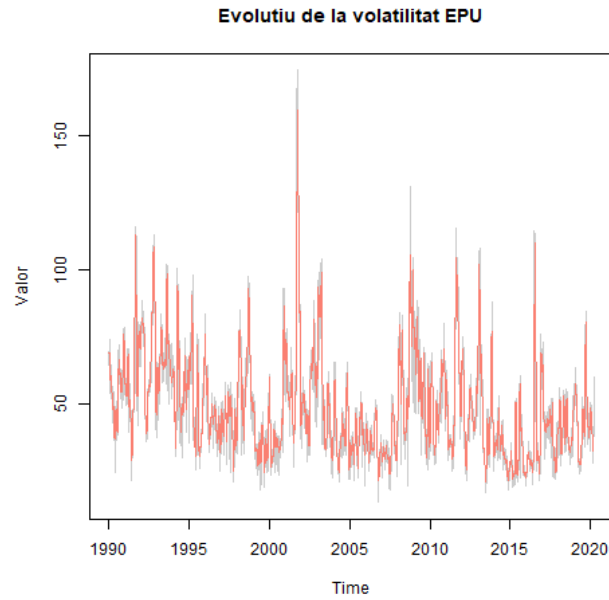
En els gràfics anteriors es pot observar com el període comprès entre el 2000-2005 va suposar un augment elevat del indicador corresponent. Com s'ha mencionat anteriorment, al llarg d'aquests anys van esdevenir molts conflictes polítics, sobretot relacionats amb guerra o atacs terroristes, es per aquest motiu que s'observen aquests pics. Per altre banda, el EPU al llarg del 2010-2015 també mostra un comportament elevat juntament amb el VIX aquests augments vindrien explicats per les crisis financeres i polítiques que es van tenir lloc durant aquell període.

Tal i com s'ha mencionat anteriorment, a continuació es realitza el mateix anàlisi preliminar sobre les variables creades, és a dir, la volatilitat dels indicadors. A continuació es mostren els estadístics bàsics per a cada una d'elles:

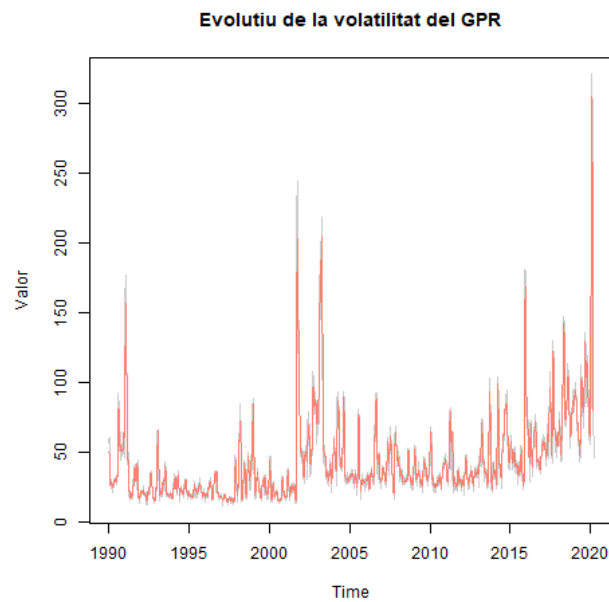
	vol.epu	vol.gpr	vol.vix
Min	13,77	11,75	9,14
1st Qu.	34,39	24,73	13,5
Median.	45,08	34,36	17,11
Mean.	49	44,1	19,09
3rd Qu.	58,92	51,59	22,41
Max.	173,97	321,19	80,86

*Taula 4: Estadístics descriptius de les variables.
Font: Elaboració pròpia*

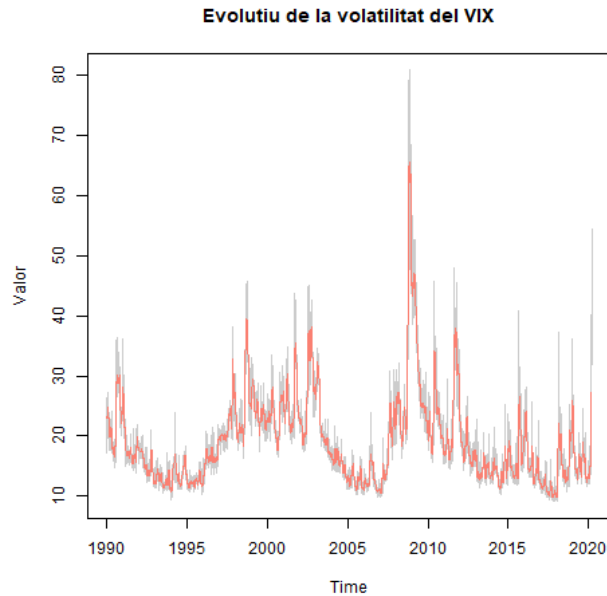
Tot seguit, es mostra el gràfic evolutiu corresponent on s'ha calculat una mitjana mòbil de trenta dies amb l'objectiu de suavitzar la sèrie temporal, representada en color vermell tal i com es mostra a continuació:



*Figura 17: Evolució de la volatilitat del EPU, període 1990-2020.
Font: Elaboració pròpia.*



*Figura 18: Evolució de la volatilitat del GPR, període 1990-2020.
Font: Elaboració pròpia.*



*Figura 19: Evolució del VIX, període 1990-2020.
Font: Elaboració pròpia.*

Com es pot veure en els gràfics que tenen en compte la volatilitat dels indicadors, s'observa la mateixa tendència mencionada anteriorment, tots tres indicadors tenen un augment considerat entre els anys 2000-2005, explicat pels conflictes bèl·lics, i posteriorment tant el EPU com el VIX mostren un augment al voltant del 2010, que podria estar explicat per les crisis financeres que van tenir lloc.

En el cas del GPR es pot observar que sembla que a partir del 2015 hi ha una tendència creixent de la volatilitat, tot i així no es pot veure amb claredat si existeixen cicles de volatilitat en els tres casos corresponents.

Adicionalment, s'estudia la correlació que hi ha entre variables que corresponent a la volatilitat dels indicadors, considerant els tres tipus de coeficients de correlació¹⁷. Aquests són els següents: *Pearson, Spearman i Kendall*.

¹⁷ El coeficient de correlació pren valors entre el -1 i el 1. Un valor igual a 0 indica que no hi ha associació entre les dues variables. Un valor positiu implica que hi ha una associació positiva, és a dir, que a mesura que augmenta el valor d'una variable, també augmenta el valor de l'altre, per el contrari, un valor negatiu indica que la associació és negativa, és a dir, que a mesura que augmenta el valor d'una variable, l'altre variable disminueix.

- *Coefficient de correlació de Pearson* es defineix com:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

PEARSON			
	vol.epu	vol.gpr	vol.vix
vol.epu	1	0,0686791	0,3380782
vol.gpr	0,0686791	1	0,028575
vol.vix	0,3380782	0,028575	1

*Taula 5: Coeficient correlació Pearson.
Font: Elaboració pròpia.*

- *Coefficient de correlació de Spearman* es defineix com:

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

On D es la diferència entre els corresponents valors i N és el nombre de parelles.

SPEARMAN			
	vol.epu	vol.gpr	vol.vix
vol.epu	1	-0,1231974	0,2322685
vol.gpr	-0,1231974	1	-0,0289206
vol.vix	0,2322685	-0,0289206	1

*Taula 6: Coeficient correlació Spearman.
Font: Elaboració pròpia.*

- *Coefficient de correlació de Kendall* es defineix com:

$$\tau = \frac{2(N_c - N_d)}{N(N - 1)}$$

On N_c correspon a una parella amb concordança i N_d correspon a aquelles parelles sense concordança.

KENDALL			
	vol.epu	vol.gpr	vol.vix
vol.epu	1	-0,0833926	0,1591358
vol.gpr	-0,0833926	1	-0,019461
vol.vix	0,1591358	-0,019461	1

*Taula 7: Coeficient correlació Kendall.
Font: Elaboració pròpia.*

A les tres taules anteriors que corresponent als coeficients de correlació corresponents s'observa que en tots els casos hi ha una correlació positiva, i la més alta en tots els casos, entre el indicador d'incertesa política-econòmica i el VIX, això podria estar explicat sota la suposició de la metodologia que es dur a terme per a construir aquest indicador, ja que aquesta inclou aspectes econòmics a diferència de la metodologia que es dur a terme per la construcció del indicador de risc geopolític. Tenint en compte el valor de les correlacions, al tenir uns valors baixos, es pot podria concloure inicialment que es possible que no hi hagi multicol·linealitat ¹⁸entre les variables.

Seguidament i tenint en compte, que s'estan analitzant series temporals, en procedeix a analitzar la sèrie temporal des de el punt de vista de les seves components estructurals:

$$\text{Serie Observada} = \text{Tendencia} + \text{Efecte estacional} + \text{Errors aleatoris}$$

- Tendència: es pot definir com el comportament o moviment a llarg termini.
- Efecte estacional: moltes series temporals presenten certa periodicitat, és a dir, mostren una variació en un cert període determinat (anual, mensual). Un exemple clar sobre aquest fet és l'atur, ja que aquest augmenta normalment al hivern i en canvi al estiu disminueix.
- Component aleatòria: es tracta de variacions aleatòries al voltant dels components anteriors.

Cal ressaltar que aquest anàlisi es realitza sobre la volatilitat de les variables. S'utilitza la funció *decompose*¹⁹ () a R on el resultat és el següent:

¹⁸ Multicol·linealitat es produeix quan existeix una correlació entre les variables independents, de manera que la regressió entre la dependent i les independents, en trobar-se aquestes correlacionades entre elles, queda emmascarada.

¹⁹ La seva sortida en R conte la següent informació: *Observed*: la sèrie de temps observada, *Seasonal*: sèrie temporal del component estacional estimat, *Trend*: sèrie temporal de la tendència estimada, *Random*: sèrie temporal dels errors aleatoris.

Decomposition of additive time series

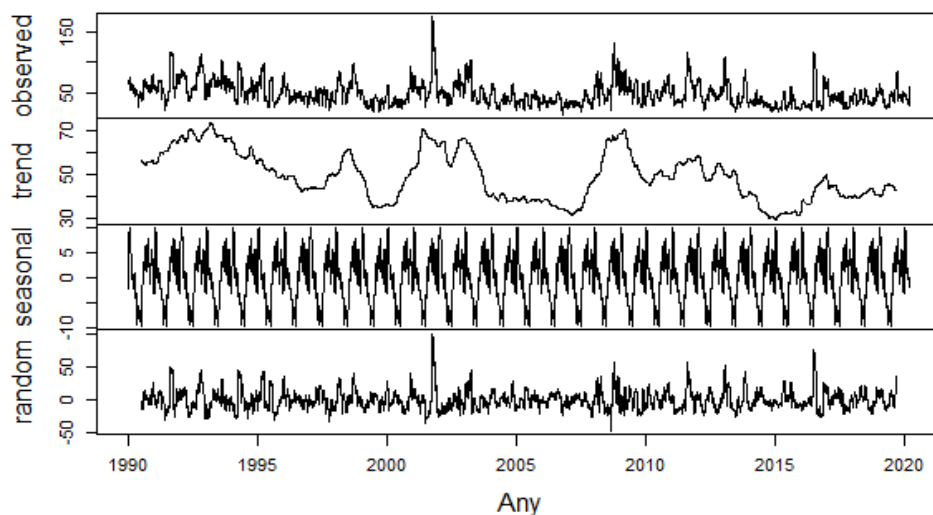


Figura 20: Sèrie temporal volatilitat EPU i els seus components.
Font: Elaboració pròpia.

Decomposition of additive time series

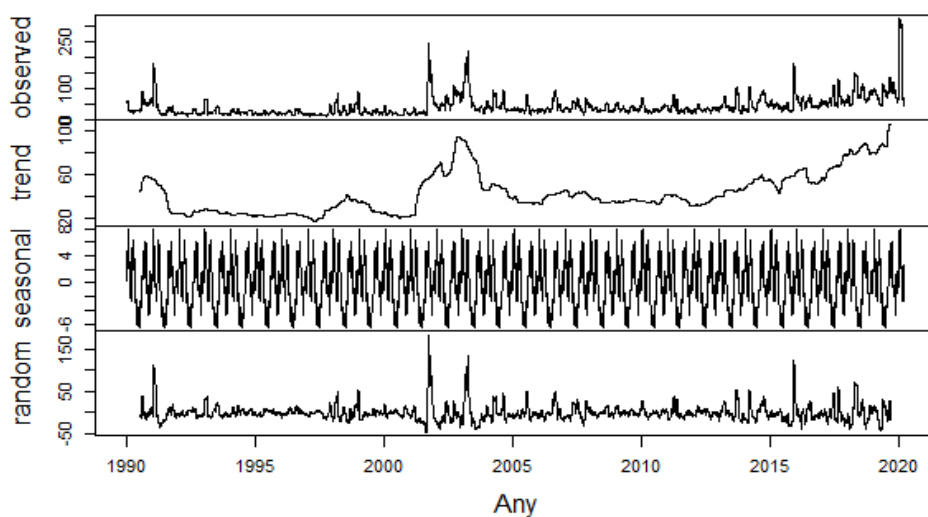
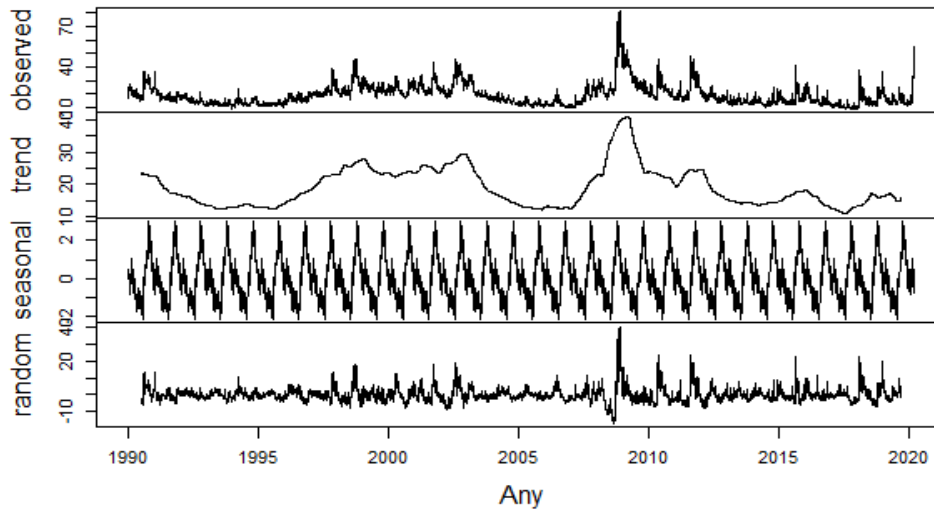


Figura 21: Sèrie temporal volatilitat GPR i els seus components.
Font: Elaboració pròpia.

Decomposition of additive time series



*Figura 22: Sèrie temporal VIX i els seus components.
Font: Elaboració pròpia.*

Tenint en compte els gràfics anteriors on es mostren els components de la sèrie temporal es pot concloure que tant per la volatilitat del EPU com el VIX no s'observa que tinguin una tendència clara, però sí que entre elles sembla que hi ha certa relació. En canvi en el cas de la volatilitat del GPR, sí que es pot observar com a partir del 2005, hi ha una tendència creixent.

Si tenim en conte el context històric en el que estem vivint avui en dia, observar que el indicador de risc geopolític te una tendència creixent en aquests últims anys es degut a que han augmentat les possibles amenaces o bé els conflictes bèl·lics estan d'acord amb el que succeeix en el mon últimament, ja que tenint en compte les notícies que s'escolten avui en dia relacionades amb guerres, conflictes, etc.. cada vegada va en augment.

9.3 Definició de les particions

Per a poder realitzar un model que aprengui a predir uns valors a partir de les dades i que posteriorment permeti generalitzar aquests resultats en un nou conjunt de dades, es necessari dividir la mostra de dades en tres subconjunts anomenats: entrenament, validació i test.

En el **conjunt d'entrenament** es troba el major percentatge de les dades on l'objectiu principal és que el model ha d'aprendre a detectar les característiques essencials de les dades. Posteriorment haurà de generalitzar el procés d'aprenentatge amb la finalitat de ser capaç de reconèixer i classificar les característiques en un conjunt de dades totalment nou. De manera que com major sigui aquest conjunt de dades, major serà la varietat i característiques diferents que podrà aprendre durant la etapa de entrenament. Entenen el concepte d'aprenentatge el procés en el qual el model ajusta els paràmetres automàtics.

El **conjunt de validació** està format per un percentatge més petit del total de dades disponibles. El conjunt de validació és útil perquè ajuda a reconèixer si el model presenta *overfitting* o *underfitting*.

Finalment, el **conjunt de test** te el mateix percentatge que el conjunt de validació. Aquest conjunt serveix per fer una avaluació final del model, per tant en aquesta fase no s'ajusta cap paràmetre. Simplement s'utilitzen els paràmetres i hiperparametres calculats als dos conjunts anteriors, amb la finalitat de veure si el model es capaç de reproduir aproximadament els mateixos resultats amb un conjunt de dades que fins el moment, no ha sigut observable.

En aquest treball però, per tal de realitzar les prediccions corresponents, s'ha dividit el conjunt de dades en les següents parts:

- Entrenament: que conte les dades des del el 01-01-1990 fins el 31-12-2015 (85% de les dades)
- Test: que conte les dades des del 01-01-2016 fins el 10-03-2020 (15% de les dades).

Només s'utilitzen dues fases ja que Hyndman (2018) calculen els pronòstics diferenciant únicament entre els subconjunt de entrenament i el conjunt de test, i tenint en compte que al llarg de la part empírica del treball, s'aplicaran funcions del R automàtiques que han estat programades per aquests autors, es considera oportú, seguir aplicant la mateixa diferenciació mencionada pels propis autors.

10. Models lineals estocàstics (ARIMA)

El model ARIMA (*autoregressive integrated moving average*) és un model estadístic desenvolupat per *Box i Jenkins (1970)* per a series temporals que tenen en compte la dependència existent entre les dades, és a dir, la observació en el moment donat es modelitza a partir de dades anteriors.

Aquesta metodologia permet descriure el valor com una funció lineal de les dades anteriors i errors aleatoris. Els components del que està format i com a conseqüència donen nom al model són els següents:

- AR: correspon a un model autoregressiu.
- I: es tracta de la diferenciació de la variable.
- MA: correspon a les mitjanes mòbils.

Els autors van definir una metodologia a seguir per tal de realitzar les tècniques de predicció, aquesta es coneix com **metodologia Box-Jenkins**, que es posarà en practica al llarg d'aquest treball:

- Identificació del model: consta de fer un anàlisi de la sèrie temporal i decidir quines són les transformacions que s'han d'aplicar amb l'objectiu que la sèrie sigui estacionaria. Seguidament es determina quin és el model ARMA a aplicar i per tant identificar quin són les ordres de p i q de la seva estructura autoregressiva i de mitjana mòbil.
- Estimació dels paràmetres: s'estimen els paràmetres AR i MA a partir de màxima versemblança on s'obtenen els errors estàndards i els residus del model.
- Validació: s'ha de verificar que es compleixen els supòsits bàsics del model. Aquests són: significació dels paràmetres, condicions d'estacionarietat i que el terme de pertorbació segueixi una distribució normal.
- Predicció: finalment un cop es dona per vàlid el model, es realitzen les prediccions amb el mateix.

A continuació es posa en pràctica la metodologia definida anteriorment.

10.1 ARIMA Univariant del VIX

Es proposa realitzar un model ARIMA on només es tingui en compte la informació del propi indicador VIX.

Per a poder fer la predicció mitjançant aquest model, un dels requisits que s'ha de complir, és que les dades siguin **estacionaries**. Es comprova si la sèrie és estacionària mitjançant gràfics de la FAS i la FAP i per tal de contrastar la informació sobre si és estacionari o no, també es realitzarà el test de *Dickey-Fuller Ampliat* i el test de KPSS (*Kwiatkowski – Phillips – Schimidt – Shin*)

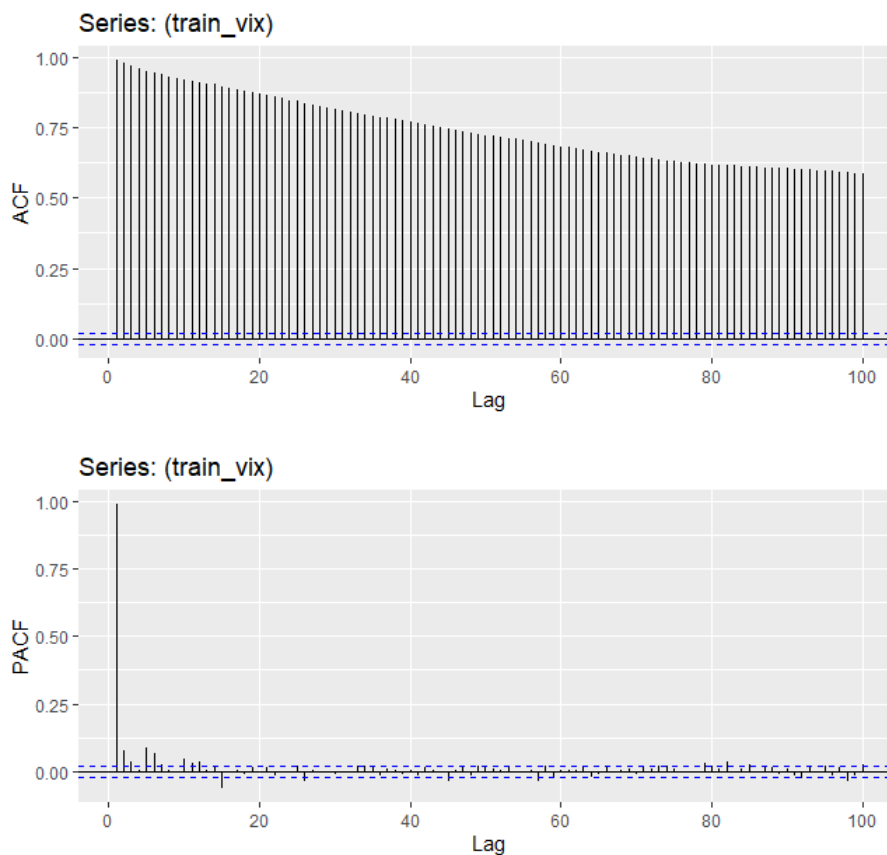


Figura 23: Correlograma del VIX.
Font: Elaboració pròpia.

Tenint en compte el gràfic del component regular (FAS) no es poden extreure conclusions clares respecte si es tracta d'una sèrie temporal estacionària o no, ja que tot i que aquests disminueixen a mesura que augmenten els retards, aquests no disminueixen ràpidament, de manera que es realitza el test comentat anteriorment a partir de la funció *adf.test* i *kps.test* que ens facilita el R.

El test de Dickey-Fuller Ampliat té la següent hipòtesis:

$$\begin{cases} H_0: \text{La sèrie no és estacionaria} \equiv \text{Te arrel unitaria} \\ H_1: \text{La sèrie és estacionaria} \equiv \text{No té arrel unitaria} \end{cases}$$

Augmented Dickey-Fuller Test

```
data: log(train_vix)
Dickey-Fuller = -4.832, Lag order = 21, p-value = 0.01
alternative hypothesis: stationary
```

*Figura 24: Test de Dickey-Fuller Ampliat.
Font: Elaboració pròpia.*

Tenint en compte p-valor del test realitzat, aquest és inferior al 5%, per tant es rebutja la hipòtesis nul·la i com a conseqüència s'arriba a la conclusió que es tracta d'una sèrie estacionaria.

Adicionalment es comprova a partir de la funció *ndiffs* que automatitza la cerca de la quantitat de diferències necessàries per tal de poder reafirmar que la sèrie es estacionaria. Aquest retorna un valor de 0, per tant no cal diferenciar la variable, estant en línia amb les afirmacions realitzades anteriorment.

Per altre banda, el test de KPSS té la següent hipòtesi:

$$\begin{cases} H_0: \text{La sèrie és estacionaria} \\ H_1: \text{La sèrie no és estacionaria} \end{cases}$$

KPSS Test for Trend Stationarity

```
data: log(train_vix)
KPSS Trend = 2.781, Truncation lag parameter = 12, p-value = 0.01
```

*Figura 25: Test de KPSS.
Font: Elaboració pròpia.*

Tenint en compte p-valor del test realitzat, aquest és inferior al 5%, per tant es rebutja la hipòtesis nul·la i com a conseqüència s'arriba a la conclusió que no es tracta d'una sèrie estacionaria. Es comprova a partir de la funció *ndiffs* que automatitza la cerca de la quantitat de diferències necessàries. Aquest retorna un valor de 1, per tant cal diferenciar la variable.

De manera que si es té en compte els resultats del test de KPSS, si que es necessari diferenciar tal i com ens mostra la sortida del ARIMA, en comparació amb el test de Dickey – Fuller que tenint en compte el p-valor resultant, aquest no era necessari diferenciar. Al observar que la sèrie no presenta estacionarietat en variància, d'ara en endavant, es treballarà amb les variables transformades en escala logarítmica.

Seguint amb les fases comentades anteriorment, es procedeix a identificar el model ARIMA a aplicar. Per fer aquest apartat, es farà ús de *auto.arima()* que retorna el millor model ARIMA segons el valor de *Akaike* (AIC), Criteri de informació corregit de Akaike (AICc) i Criteri de Informació de Bayes(BIC).

Els resultats obtinguts amb la implementació del *auto.arima* son les següents:

```
Series: log(train_vix)
ARIMA(2,1,1)

Coefficients:
      ar1      ar2      ma1
    0.8548  0.0056 -0.9234
s.e.  0.0229  0.0120  0.0205

sigma^2 estimated as 0.002681:  log likelihood=14631.48
AIC=-29254.96  AICc=-29254.96  BIC=-29226.33
```

Figura 26: Sortida model univariant *auto.arima*.
Font: Elaboració pròpia.

De manera que tenint en compte aquest resultat, el millor model automatitzat trobat es un ARIMA (2,1,1). Cal destacar que tenint en compte aquests resultats, si que seria necessària una diferenciació de la variable, fet que prèviament tenint en compte el test de Dickey – Fuller, s’havia dit que no era necessari però en canvi amb el test de KPSS, aquesta diferenciació si que era necessària per a complir la condició d’estacionarietat.

El següent pas, un cop obtingut el model, és validar-lo. Per a validar un model s’ha de comprovar que els paràmetres estimats són estadísticament significatius, així com comprovar que els residus es distribueixen com una distribució normal, aplicant el test de *Shapiro-Wilk*.

Tenint en conte el model obtingut anteriorment, ARIMA (2,1,1) es comprova si els paràmetres son significatius a partir de la funció *coefstest()*. El resultat és el següent:

```
z test of coefficients:
      Estimate Std. Error  z value Pr(>|z|)
ar1  0.8547893  0.0228723  37.3723  <2e-16 ***
ar2  0.0056203  0.0120270   0.4673  0.6403
ma1 -0.9233962  0.0204615 -45.1285  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figura 27: Sortida coeficients model ARIMA (2,1,1) univariant.
Font: Elaboració pròpia.

Com es pot observar, no tots els coeficients són estadísticament significatius. És per aquest motiu, que es comprova el model ARIMA(1,1,1) conjuntament amb si els coeficients d'aquest son estadísticament significatius.

```
Call:
arima(x = log(train_vix), order = c(1, 1, 1))

Coefficients:
      ar1      ma1
  0.8536  -0.9186
s.e.  0.0230  0.0179

sigma^2 estimated as 0.002681:  log likelihood = 14631.34,  aic = -29256.67
```

Figura 28: Sortida model ARIMA(1,1,1) univariant.
Font: Elaboració pròpia.

```
z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.853613   0.023006  37.104 < 2.2e-16 ***
ma1 -0.918585   0.017877 -51.385 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 29: Sortida coeficients ARIMA (1,1,1) univariant
Font: Elaboració pròpia.

Obtenint tots els coeficients significatius i tenint en compte que els valors de Akaike dels dos models estimats, sent -29254,96 (ARIMA(2,1,1)) respecte -29256,67. (ARIMA(1,1,1)) no difereixen gaire es procedeix a realitzar el test de *Shapiro-Wilk* per tal de comprovar que els residus del model ARIMA (1,1,1) es distribueixen com una distribució normal.

La hipòtesis del test de normalitat *Shapiro-Wilk* és la següent:

$$\begin{cases} H_0: X \sim N(\mu, \sigma^2) \\ H_1: X \not\sim N(\mu, \sigma^2) \end{cases}$$

El resultat obtingut és el següent:

```
shapiro-wilk normality test

data:  arima.uni2$residuals
W = 0.99468, p-value = 0.08057
```

Figura 30: Test de normalitat model ARIMA (1,1,1) univariant Shapiro-Wilk.
Font: Elaboració pròpia.

Tenint en compte p-valor del test realitzat, aquest és superior al 5%, per tant no es tenen evidències suficients per a rebutjar la hipòtesis nul·la i com a conseqüència s'arriba a la conclusió que els residus es distribueixen com una distribució normal.

Adicionalment també es comprova si el model ARIMA (1,1,1) és invertible i estacionari estudiant les arrels del model a partir de la funció *autoplot()* que permet analitzar les arrels invertides. Aquestes, si es troben dins del cercle unitari permet identificar si el model és estacionari i invertible, Hyndman (2018).

El resultat obtingut és el següent:

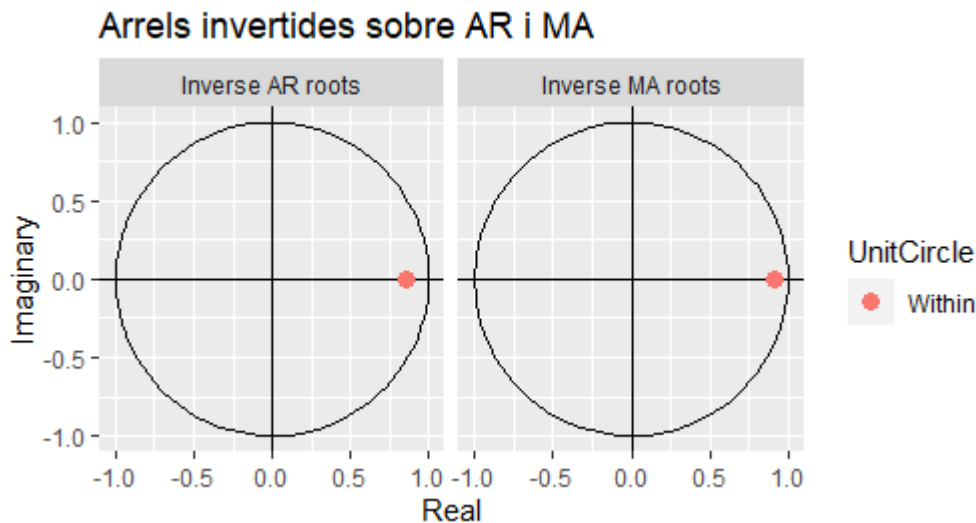


Figura 31: Arrels estacionaries inverses pel model ARIMA(1,1,1) univariant.
Font: Elaboració pròpia.

Tenint en compte el gràfic anterior, es pot verificar que el model es estacionari i estable, tot i que els punts estiguin al límit de la circumferència, estan dins del límit establert.

Seguint amb aquestes conclusions, el següent pas, és fer la predicció del model, per comprovar si les prediccions del model realitzat, s'ajusten correctament o si pel contrari, el model proporciona unes males prediccions. Per poder concloure si les prediccions son correctes, es farà servir la funció *accuracy()* de R, ja que la sortida d'aquest conte: el error mig (EM), el error quadràtic mitja de l'arrel (RMSE), error absolut mitja (MAE), error percentual mitja (MPE), error de percentatge absolut mitja (MAPE), etc. Els resultats obtinguts son els següents:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	-9,5982	0,0518	0,0327	-0,0222	1,1127	0,1187	-0,0035	NA
Test set	-1,5637	0,3130	0,2653	-6,8947	10,2928	0,9628	0,9613	4,9226

Taula 8: Resultats de la predicció model ARIMA (1,1,1) univariant.
Font: Elaboració pròpia.

Tenint en compte els resultats obtinguts, no es pot afirmar que el model realitzi una bona predicció dels valors del VIX, ja que el valor del MAPE es superior al 5%, donat que el valor resultant de l'EPAM té la següent interpretació:

$$\left\{ \begin{array}{l} MAPE < 1\%: \text{Molt bona predicció} \\ MAPE \in (1\%, 3\%]: \text{Capacitat predictiva bona} \\ MAPE \in (3\%, 5\%]: \text{Capacitat predictiva baixa} \\ MAPE > 5\%: \text{Capacitat predictiva pobre} \end{array} \right.$$

Per tant, predir el VIX de forma univariant amb models lineals estocàstics sembla ser que no es factible sense tenir errors elevats.

10.2 ARIMA Multivariant del VIX

Es proposa realitzar un model ARIMA per tal de fer una predicció de la volatilitat implícita a partir de les volatilitats dels indicadors d'incertesa.

Ja que la variable endògena es la mateixa, el VIX, la fase en la que s'estudia si es tracta d'una sèrie estacionaria es poden concloure les mateixes afirmacions realitzades a l'apartat anterior. De manera que s'arribarien a les mateixes conclusions si es tornés a realitzar el test de Dickey-Fuller o bé el test de KPSS. Per tant, al no tractar-se d'una sèrie estacionaria, s'aplica una transformació logarítmica previ a la sèrie. De manera que totes les variables expressades en els següents models, estan en escala logarítmica.

Es procedeix a identificar el model ARIMA a aplicar. Per fer aquest apartat, es farà ús de *auto.arima ()* amb l'argument *xreg* que permet ajusta a una regressió lineal multivariant, Hyndman, (2018) . El model que s'ajusta quan s'utilitza aquest argument és el següent:

$$\begin{aligned} X_t &= \alpha + \phi_1 C_{t,1} + \phi_2 C_{t,2} + \dots + z_t \\ z_t &= \beta_1 z_{t-1} + \dots + \beta_p z_{t-p} + e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \\ e_t &\sim N(0, \sigma) \end{aligned}$$

On *xreg* es la matriu que conte els valors C_t de cada indicador corresponent, sent aquests el logaritme de les volatilitats dels indicadors.

Els resultats obtinguts amb la implementació del *auto.arima* son les següents:

```
Series: log(train_vix)
Regression with ARIMA(2,1,1) errors

Coefficients:
      ar1      ar2      ma1      xreg1      xreg2
0.8523  0.0049 -0.9200 -0.0192 -0.0038
s.e.  0.0235  0.0120  0.0212  0.0080  0.0085

sigma^2 estimated as 0.00268:  log likelihood=14634.5
AIC=-29257  AICc=-29256.99  BIC=-29214.05
```

Figura 32: Sortida model multivariant *auto.arima*.
Font: Elaboració pròpia.

De manera que tenint en compte aquest resultat²⁰, el millor model automatitzat trobat es un ARIMA (2,1,1).

Es realitzaran els mateixos passos realitzats anteriorment, mirar si els coeficients son estadísticament significatius, que els residus del model es distribueixen com una distribució normal i comprovar si el model és estacionari i invertible tenint en compte les arrels invertides del model.

A continuació es mostra la sortida de R aplicant la funció *coefstest()*:

```
z test of coefficients:

      Estimate Std. Error  z value Pr(>|z|)
ar1    0.8523218  0.0234793  36.3009 < 2e-16 ***
ar2    0.0048539  0.0120395   0.4032  0.68683
ma1   -0.9200086  0.0211630 -43.4724 < 2e-16 ***
xreg1  -0.0192191  0.0079992  -2.4026  0.01628 *
xreg2  -0.0038259  0.0084849  -0.4509  0.65206
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figura 33: Sortida coeficients ARIMA (2,1,1) multivariant.
Font: Elaboració pròpia

Com es pot observar, no tots els coeficients són estadísticament significatius. És per aquest motiu, que es comprova el model ARIMA(1,1,1) on solament es tingui en compte la volatilitat del EPU com a variable exògena (xreg1), ja que aquest si que és significatiu.

²⁰ Xreg1 i xreg2 son les dues variables independents generades com a volatilitats dels indicadors d'incertesa política-econòmica i de risc geopolític, respectivament.

```

Call:
arima(x = log(train_vix), order = c(1, 1, 1), xreg = matriu_epu)

Coefficients:
      ar1      ma1  matriu_epu
  0.8534 -0.9176   -0.0195
s.e.  0.0234  0.0182    0.0080

sigma^2 estimated as 0.002679:  log likelihood = 14634.28,  aic = -29260.57

```

Figura 34: Sortida model ARIMA(1,1,1) multivariant.
Font: Elaboració pròpia

z test of coefficients:

```

              Estimate Std. Error  z value Pr(>|z|)
ar1           0.8534314  0.0233505  36.5488 < 2e-16 ***
ma1           -0.9175692  0.0181841 -50.4601 < 2e-16 ***
matriu_epu   -0.0195403  0.0079948  -2.4441  0.01452 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 35: Sortida coeficients ARIMA (1,1,1) multivariant
Font: Elaboració pròpia.

Obtenint tots els coeficients significatius i tenint en compte que els valors de Akaike dels dos models estimats, no difereixen gaire es procedeix a realitzar el test de *Shapiro-Wilk* per tal de comprovar que els residus del model ARIMA (1,1,1) multivariant amb només la volatilitat del EPU com a variable exògena es distribueixen com una distribució normal.

La hipòtesis del test de normalitat *Shapiro-Wilk* és la següent:

$$\begin{cases} H_0: X \sim N(\mu, \sigma^2) \\ H_1: X \not\sim N(\mu, \sigma^2) \end{cases}$$

El resultat obtingut és:

```

shapiro-wilk normality test

data:  arima.epu$residuals
w = 0.99933, p-value = 0.061

```

Figura 36: Test de normalitat model ARIMA (1,1,1) multivariant Shapiro-Wilk.
Font: Elaboració pròpia.

Tenint en compte p-valor del test realitzat, aquest és superior al 5%, per tant no es tenen evidències suficients per a rebutjar la hipòtesis nul·la i com a conseqüència s'arriba a la conclusió que els residus es distribueixen com una distribució normal.

El següent pas és comprovar si el model ARIMA (1,1,1) multivariant amb la variable de la volatilitat del EPU com a variable exògena és invertible i estacionari estudiant les arrels invertides.

El resultat obtingut és el següent:

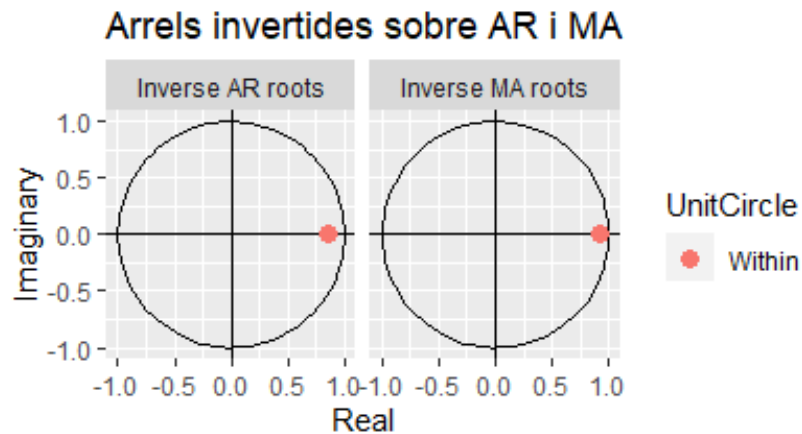


Figura 47: Arrels estacionàries inverses pel model ARIMA(1,1,1) multivariant.
Font: Elaboració pròpia.

Tenint en compte el gràfic anterior, es pot verificar que el model es estacionari i estable, tot i que els punts estiguin al límit de la circumferència, estan dins del límit establert.

Un cop validat el model, el següent pas és fer la predicció del model, per comprovar si les prediccions del model realitzat, s'ajusten correctament o si pel contrari, el model proporciona unes males prediccions. Per poder concloure si les prediccions son correctes, es farà servir la funció *accuracy ()* de R, ja utilitzada en l'apartat anterior. Els resultats obtinguts son els següents:

	ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's U
Test set	-0,0884	0,2851	0,2338	-4,3197	8,9320	0,9612	4,3435

Taula 9: Resultats de la predicció model ARIMA (1,1,1) multivariant.
Font: Elaboració pròpia.

Tenint en compte els resultats obtinguts, no es pot afirmar que el model realitzi una bona predicció dels valors del VIX, ja que si el valor del MAPE es superior al 5% es considera que la capacitat predictiva del model es pobre. Per tant, predir el VIX de forma multivariant amb models lineals estocàstics sembla ser que no es factible sense tenir errors elevats.

Tot i que aplicant la metodologia ARIMA en els dos models calculats (tant univariant com multivariant) s'ha arribat a la conclusió que no realitzen una bona predicció del VIX. Si es té en compte el valor del MAPE, el que millor ajustaria els resultats, és el model multivariant, ja que és el que té un error de percentatge absolut mitjà menor, tot i així els valors són molt semblants. Cosa que confirma, com s'ha dit anteriorment, que els models ARIMA no realitzen unes bones prediccions. Aquest fet podria estar explicat degut a que les relacions entre les variables són més aviat no lineals que lineals.

11. Models no lineals (Xarxes neuronals)

Abans de definir en què consisteix una xarxa neuronal artificial, es realitza una petita introducció al concepte d'aprenentatge automàtic (*Machine Learning*), ja que aquest està vinculat amb el funcionament de la xarxa. Es defineix aprenentatge automàtic com el desenvolupament de sistemes amb la capacitat de canviar el seu comportament de forma autònoma a base d'experiència de manera que s'aconsegueix una capacitat d'aprenentatge a partir d'un seguit de valors donats. Entenent, capacitat d'aprenentatge com desenvolupar una tasca de forma correcta i que aquesta al mateix temps, minimitza el error durant la seva execució, Nilsson (1998).

El aprenentatge automàtic es classifica principalment en dos tipus:

- Aprenentatge supervisat (*Supervised Learning*): es reben dades d'entrada i unes etiquetes associades a cada una d'aquestes dades, de manera que s'ha de trobar el algoritme correcte que sigui capaç d'associar cada valor amb la seva etiqueta corresponent, de manera que a les dades futures, es pugi assignar correctament l'etiqueta assignada.
- Aprenentatge no supervisat (*Unsupervised Learning*): es reben dades d'entrada amb diferents característiques de les quals l'objectiu es aprendre aquestes propietats i trobar patrons que permeten classificar les dades o bé estructurar-les.

11.1 Models neuronals (feed-forward)

Són tècniques de aprenentatge automàtic, concretament d'aprenentatge supervisat, que es basen en models matemàtics que pretenen simular el funcionament de la pròpia xarxa neuronal biològica. Aquesta metodologia permet que hi hagi relacions no lineals entre la variable resposta i els seus predictors, és a dir, les variables d'entrada.

Les xarxes neuronals es poden imaginar com bé indica el nom, com una xarxa de neurones que estan organitzades en diferents capes on cada una d'elles té un funcionament diferenciat. La primera capa, anomenada capa inferior està formada per els perceptrons (entrades) mentre que la capa superior està composta per els pronòstics (sortida). Entre la capa inferior i la capa superior poden haver capes entremig, aquestes reben el nom de capes ocultes. De manera que resumidament, les capes que componen les xarxes neuronals son:

- Capa d'entrada: es per on entren les dades inicials.
- Capa de sortida: es la encarregada de tornar les dades finals.
- Capes ocultes: es localitzen entre la capa d'entrada i la de sortida i son les encarregades de processar les dades per aconseguir els resultats desitjats.

Juntament amb cada perceptró si adjunten uns coeficients que es denominen ponderacions, de manera que el pronòstic s'obté a partir d'una combinació lineal de les entrades corresponents. Els pesos de cada ponderació es seleccionen en el marc de la xarxa neuronal utilitzant algoritmes d'aprenentatge que minimitza la funció de cost, Hyndman (2018). És a dir, que el procés d'aprenentatge de la xarxa neuronal consisteix en ajustar els diferents pesos en tots els perceptrons per aconseguir obtenir la sortida final.

Les xarxes més simples son aquelles que no contenen capes ocultes i serien equivalents a les regressions lineals, de manera que en el moment que s'afegeix una capa oculta, la xarxa neuronal passa a ser no lineal. A continuació es mostra una xarxa neuronal d'una regressió lineal amb quatre entrades i una altra figura que fa referència a la mateixa xarxa neuronal amb afegint una capa oculta:

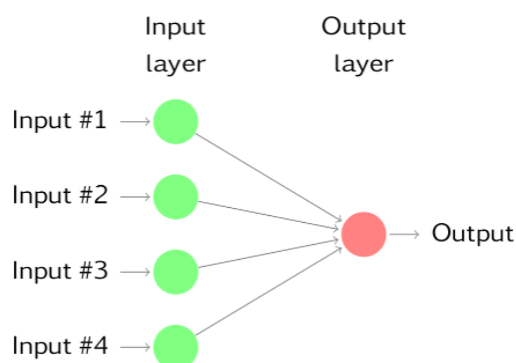


Figura 38: Xarxa neuronal simple.

Font: Hyndman, R.J. Forecasting: principles and practice.

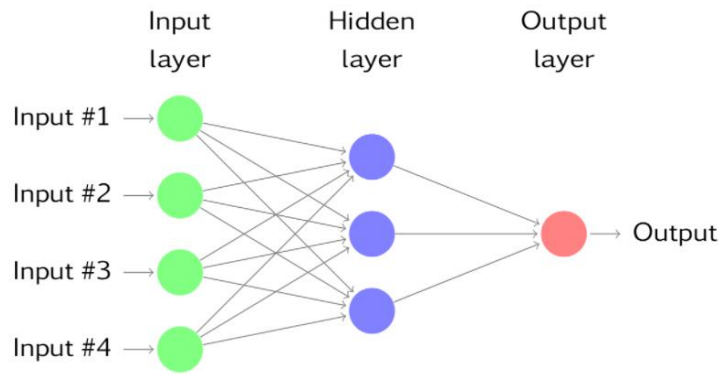


Figura 39: Xarxa neuronal amb quatre entrades i una capa oculta.
 Font: Hyndman, R.J. Forecasting: principles and practice.

La figura 39 es coneix com una *xarxa d'alimentació directe multicapa*, on cada capa de nodes rep entrades de les capes anteriors, i on les sortides d'aquest node son entrades per la següent capa, i així, successivament.

Per a obtenir la sortida de la xarxa neuronal, aquesta utilitza el que s'anomena **funció d'activació**, que es la que s'encarrega de calcular cada sortida en funció de totes les entrades i pesos. Existeixen diferents tipologies de funcions d'activació i aquestes es seleccionen segons les necessitats de la xarxa. Les més utilitzades son les següents:

- Lineal: funció lineal que retorna el valor del vector d'entrada al qual se li aplica.

$$G(x_i) = x_i \text{ on } x_i \in \mathbb{R}^2$$

- Sigmoide: es tracta d'una funció no lineal. Retorna un valor entre 0 i 1, depenen del valor d'entrada del vector que se li aplica. Aquesta es sol aplicar en problemes de classificació binaria donades les limitacions que presenta. Es fàcilment calculable amb la següent formulació matemàtica:

$$G(x_i) = \frac{1}{1 + e^{-x_i}} \text{ on } x_i \in \mathbb{R}^2$$

- Rectifier Liner Unit (ReLU): es tracta d'una funció no lineal que permet obtenir poques representacions i impedeix que els pesos d'una capa es quedin estancats durant l'entrenament. Normalment s'utilitza per problemes de tipologies profundes. Consisteix en escollir el màxim valor per a cada mostra que entra i 0:

$$G(x_i) = \max(0, x_i) \text{ on } x_i \in \mathbb{R}^2$$

- Softmax: es una funció no lineal similar al funcionament de la sigmoide. Retorna el que es podria considerar la probabilitat a posteriori. El seu càlcul es el següent:

$$G(x_i) = \frac{e^{x_i}}{\sum \forall i e^{x_i}} \text{ on } x_i \in \mathbb{R}^2$$

Un cop s'ha dissenyat tota l'arquitectura de la xarxa neuronal el que s'ha de fer es **entrenar-la** i es per aquest motiu que es necessita el que es denomina *set* d'entrenament. Un cop es disposa d'aquest *set* s'introdueix a la xarxa neuronal que s'encarregarà d'actualitzar tots els pesos a partir d'algoritmes iteratius.

Com s'ha dit anteriorment, l'objectiu es **minimitzar l'error** durant el procés d'aprenentatge. Per a calcular aquest error es poden utilitzar les següents formules:

- Error quadràtic mig:

$$ECM = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2$$

- Error mitja absolut:

$$EAM = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|^2$$

Un altre concepte clau relacionat amb les xarxes neuronals és la **optimització**, ja que en problemes complexos no sol ser possible aconseguir els pesos que obtingut un error nul de forma analítica, es per aquest motiu que s'apliquen algoritmes iteratius d'optimització per poder apropar-se a la solució que proporciona un error baix.

Les opcions en aquest cas son molt variades, però casi totes tenen en comú el us de la variació del error respecte la variació dels pesos aplicant algoritme del gradient descendent²¹ o bé aplicant el mètode de Newton, Geoffrey (2012).

²¹ L'algoritme del gradient descendent és un mètode iteratiu d'optimització de primer ordre que busca el mínim d'una funció.

En les següents equacions, w representa els pesos, γ el paràmetre d'aprenentatge i E el error corresponent al resultat de càlcul.

- Algorisme del gradient descendent:

$$\Delta w[t] = -\gamma \frac{\partial E_n}{\partial w[t]}$$

- Mètode de Newton:

$$\Delta w[t] = -\gamma \left(\frac{\partial^2 E_n}{\partial w[t]^2} \right)^{-1} \frac{\partial E_n}{\partial w[t]}$$

En tots els mètodes d'optimització hi ha el paràmetre d'aprenentatge, que indica el moment de salt en cada iteració de l'algorisme. Aquest paràmetre es molt important per assolir el error mínim, ja que un valor alt pot ocasionar que l'algorisme no convergeixi, fent que no sigui possible un entrenament. En canvi un valor baix si que pot assegurar que la sèrie convergeixi però no necessàriament sigui el valor òptim, ara bé pot causar que el nombre d'iteracions sigui molt elevat i per tant que el cost computacional sigui elevat, Geoffrey (2012).

11.2 Models neuronals autoregressius (NNAR)

Una primera aproximació del us de les xarxes neuronals per al problema de pronòstic és utilitzar les conegudes com xarxes neuronals autoregressives.

La clau per al us d'aquesta tipologia de xarxes es considerar les dades ordenades com a sèrie temporal com entrades per a una xarxa neuronal de la mateixa manera que s'utilitzen els valors antics en un model d'autoregressió lineal conjuntament amb un perceptró multicapa [Figura 39]. Aquest models se'ls anomena xarxes neuronals d'autoregressió o model NNAR (*neural network autoregression*).



Figura 40: Arquitectura NNAR bàsica.
Font: Métodos basados en deep learning.html

Quan es tracta de pronòstics que utilitzen aquest algoritme, la xarxa s'aplica de forma iterativa de manera que si es vol pronosticar un valor futur, es tenen en compte els valors històrics determinats, en canvi si es vol pronosticar dos valors futurs, llavors es té en compte el pronòstic anterior i el històric de la mateixa sèrie temporal.

Aquestes xarxes neuronals presenten avantatges respecte aplicar un model autoregressiu clàssic AR com venen a ser que les NNAR seleccionen i estimen automàticament els coeficients importants d'un procés AR dispers, eliminant la necessitat de conèixer l'ordre real del procés AR.

Considerant la següent sèrie temporal expressada com un procés AR:

$$y_t = c + \sum_{i=1}^p w_i y_{t-i} + e_t$$

Per predir el següent pas y_t es multiplica per un pes w_i que aprèn la xarxa neuronal, anomenat coeficient AR, p són els retards i e és el soroll blanc. De manera que la xarxa neuronal s'entrena i a partir del mètode de descens del gradient s'aprenen aquests coeficients AR.

El disseny d'aquesta xarxa neuronal correspon a que aquells paràmetres que es troben a la primera capa són equivalents als coeficients AR. Aquest model es pot ampliar opcionalment amb capes ocultes per aconseguir una millor precisió de predicció.

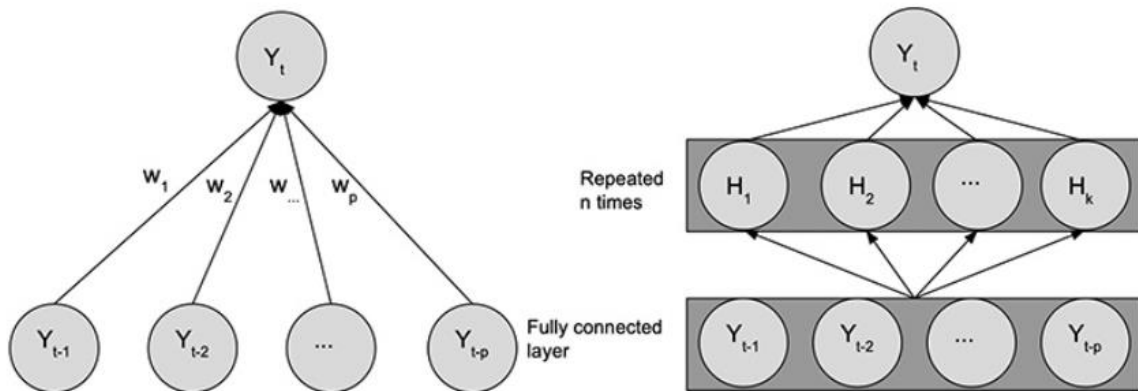


Figura 41: AR equivalent a una xarxa neuronal sense cap capa oculta (esquerra). I una xarxa neuronal autoregressiva amb n capes ocultes. (dreta).

Font: AR-Net: A simple auto-regressive neural network for time – series.

Hyndman (2018) han implantat aquesta tipologia de xarxes neuronals des de zero. Aquesta funció es *nnetar* del paquet *forecast*. Aquesta només te en compte una capa oculta i retorna una notació NNAR (pag, k) on el pag indica el valors antics a tenir en compte i k fa referència als nodes que hi ha a la capa oculta. Un NNAR ($pag, 0$) es equivalent a un model ARIMA ($pag, 0, 0$), però sense tenir en compte els paràmetres que assegurin la estacionalitat.

Amb dades estacionals, és útil agregar els últims valors observables de la mateixa temporada com entrada. Per exemple un model $NNAR(3, 1, 2)_{12}$ te en compte els tres últims valors i el dotze, és a dir te les següents entrades: $y_{t-1}, y_{t-2}, y_{t-3}, y_{t-12}$, tenint dues neurones en una capa oculta.

A continuació, es procedeix a realitzar un model NNAR univariant i un model NNAR multivariant per tal de realitzar els pronòstics del VIX, utilitzant com s'ha dit anteriorment, la funció *nnetar* que proporciona el programa R.

11.2.1 Model Neuronal Autoregressiu Univariant.

En aquest apartat, es proposa realitzar una predicció del VIX, mitjançant tècniques de predicció de xarxes neuronals autoregressives, on només es tingui en compte la seva pròpia informació.

El valor de les dades resultants d'aplicar la funció *nnetar* en R són les següents:

```
Series: log(train_vix)
Model: NNAR(15,1,8) [365]
Call: nnetar(y = log(train_vix), lambda = "auto")

Average of 20 networks, each of which is
a 16-8-1 network with 145 weights
options were - linear output units

sigma^2 estimated as 4.377e-05
```

Figura 42: Sortida Xarxa Neuronal Autoregressiva (NNAR) univariant.
Font: Elaboració pròpia.

Tenint en compte el valor resultant, es tracta d'un model NNAR (15,1,8)[365], que es tracta d'un model on es te en compte els quinze últims valors i el valor de fa un any: $y_{t-1}, y_{t-2} \dots y_{t-15}, y_{t-365}$ tenint vuit neurones en la capa oculta.

A continuació es mostra el gràfic de pronòstic del model obtingut, representat per una línia blava i els valors reals de la sèrie temporal, representants per una línia vermella.

Forecasts from NNAR(15,1,8)[365]

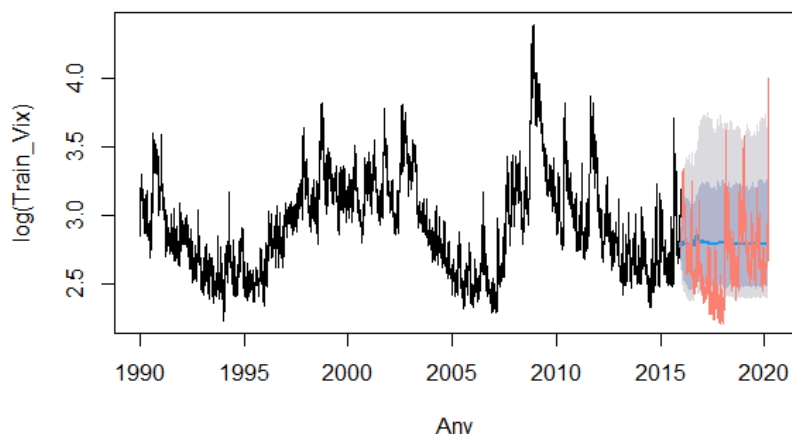


Figura 43: Pronòstic Model NNAR univariant.
Font: Elaboració pròpia.

Tenint en compte el gràfic anterior, es pot observar com les prediccions realitzades amb el model són dèbils. Per contrastar aquesta informació s'utilitza la funció *accuracy ()* de R, ja utilitzada en l'apartat anterior.

Els resultats obtinguts són els següents:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0,0009	0,0501	0,0321	0,0013	1,0957	0,1165	0,0047	NA
Test set	-0,1324	0,3019	0,2532	-5,9864	9,7789	0,9191	0,9614	4,7019

Taula 10: Resultats de la predicció model NNAR univariant.
Font: Elaboració pròpia.

Tenint en compte els resultats obtinguts, no es pot afirmar que el model realitzi una bona predicció dels valors del VIX, ja que si el valor del MAPE és superior al 5% es considera que la capacitat predictiva del model és baixa, fet que es s'havia observat anteriorment, tenint en compte el gràfic anterior.

11.2.2 Model Neuronal Autoregressiu Multivariant.

En aquest apartat, es proposa realitzar una predicció del VIX, mitjançant tècniques de predicció de xarxes neuronals autoregressives, on es tingui en compte la seva pròpia informació i la volatilitat dels indicadors de incertesa política-econòmica i de risc geopolític, respectivament. Per a poder realitzar un model multivariant, s'utilitza la funció *nnetar* de la llibreria *forecast*; però afegint el paràmetre *xreg*. Ja que com bé indica en el document oficial d'aquesta llibreria: *xreg* és un paràmetre opcional que potser un vector o una matriu de regressors externs, que ha de tenir el mateix nombre de files que la variable endògena de

manera que aquests valors han de ser numèrics. En quan aquest paràmetre està definit, les seves columnes també s'utilitzen com a entrades.

En aquest cas *xreg* és la matriu que conte els valors de cada indicador corresponent, sent aquests el logaritme de les volatilitats dels indicadors.

El valor de les dades resultants d'aplicar la funció en R són les següents:

```
Series: log(train_vix)
Model: NNAR(15,1,10)[365]
Call: nnetar(y = log(train_vix), xreg = matriu, lambda = "auto")

Average of 20 networks, each of which is
a 18-10-1 network with 201 weights
options were - linear output units

sigma^2 estimated as 4.337e-05
```

Figura 44: Sortida Xarxa Neuronal Autoregressiva (NNAR) multivariant.
Font: Elaboració pròpia.

Tenint en compte els resultats, es tracta d'un model NNAR (15,1,10)[365], que es tracta d'un model on es té en compte els quinze últims valors i el valor de fa un any: $y_{t-1}, y_{t-2} \dots y_{t-15}, y_{t-365}$ tenint deu neurones en la capa oculta. Aquesta proposta de model multivariant no selecciona cap de les variables independents, és a dir, la volatilitat dels indicadors. Per lo tant, el model multivariant com el model univariant son molt semblants entre si.

Com s'ha realitzat en el apartat anterior, a continuació, es mostra el gràfic de pronòstic del model obtingut (línia blava) en comparació amb els valors reals de la sèrie (línia vermella).

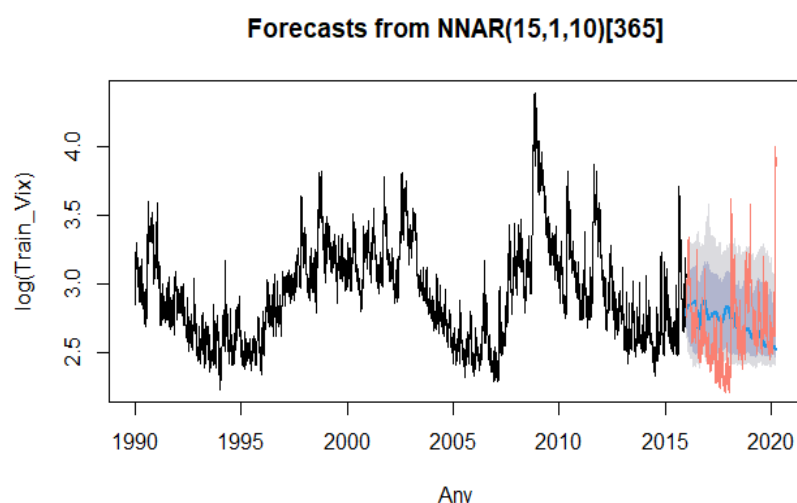


Figura 45 Pronòstic Model NNAR multivariant
Font: Elaboració pròpia.

Tenint en compte el gràfic anterior, es pot observar com les prediccions realitzades amb el model son dèbils.

Per contrastar aquesta informació s'utilitza la funció *accuracy ()* de R, ja utilitzada en l'apartat anterior.

Els resultats obtinguts son els següents:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	0,0008	0,0498	0,0321	-0,0013	1,0961	0,1166	0,0031	NA
Test set	-0,0483	0,3069	0,2456	-2,8531	9,2903	0,8913	0,9674	4,6053

*Taula 11: Resultats de la predicció model NNAR multivariant.
Font: Elaboració pròpia.*

Considerant el valor del MAPE, sent aquest major que el 5%, es considera que la capacitat predictiva del model es pobre.

Els resultats obtinguts d'aplicar la metodologia de xarxes neuronals autoregressives en els dos models proposats (tant univariant com multivariant) son que no realitzen una bona predicció del VIX. Però si es te en compte el resultat del MAPE, aquell que millor ajustaria el model és el model NNAR multivariant, tot i que la predicció sigui pobre.

A aquesta mateixa conclusió s'ha arribat si es comparava el model ARIMA univariant i el model ARIMA multivariant. Que tot i que la capacitat predictiva dels dos models era baixa, el que millor ajustava els valors, era el model multivariant.

12. Conclusions

En aquest treball s'han analitzat diferents tipus d'indicadors, com és el cas del indicador d'incertesa política-econòmica, el indicador de risc geopolític i el indicador de volatilitat implícita, des d'un punt de vista teòric com des d'un punt de vista pràctic, que es basa en predir la volatilitat a partir dels indicadors mencionats o bé predir la volatilitat amb la seva pròpia informació.

Pel que fa a la part teòrica, la metodologia que es duu a terme per a la construcció dels indicadors d'incertesa i risc geopolític es veu influenciada per la cobertura que es doni a aquest tipus de notícies en el diari. Ja que com s'ha vist al llarg del treball, si un esdeveniment bèl·lic coincideix amb un esdeveniment esportiu, com pot ser el cas de la *SuperBowl*, no reflecteix correctament el grau d'incertesa/ risc geopolític que pugi haver en aquell moment. De manera que altres esdeveniments d'interès periodístic limiten la fiabilitat dels propis indicadors.

Vinculat amb el risc geopolític, ressaltar que s'ha observat que aquest en els últims anys, està tenint una tendència creixent des de aproximadament el 2005, fet que concorda en que fruit de l'atac de les Torres Bessones, Estats Units s'ha involucrat més en les guerres del Orient Mitjà, com es és cas d'Afganistan i que últimament els atacs terroristes arreu del món han tingut mes pes, com és el cas per exemple dels atemptats a Paris.

Per altre banda, destacar que la incertesa en una economia te un gran impacte pel que fa a les decisions d'inversió, ocupació, preu dels actius tenint així un impacte negatiu. I que un augment del risc geopolític també causa impactes negatius com per exemple en el PIB d'un país, el turisme, sector financer, mercat del petroli, etc. Però un augment d'aquest indicador no requereix el mateix impacte si es tracta d'una economia desenvolupada o en vies de desenvolupament, ja que en aquest últim cas, l'impacte es major.

Respecte la capacitat predictiva dels models plantejats, ja sigui la metodologia univariant com multivariant dels models ARIMA com de les xarxes neuronals autoregressives, no realitzen una bona predicció dels valors del VIX. Tot i així, dins de cada metodologia aplicada, si tenim en compte el model univariant en comparació amb el multivariant, el que explicaria millor els resultats és el model multivariant, és a dir, el que conte la pròpia informació del VIX juntament amb el indicador d'incertesa política-econòmica (EPU) i el indicador de risc geopolític (GPR), ja que en aquesta tipologia de models es te en compte més informació provinent dels propis indicadors. De manera que es conclou, que les dues metodologies proposades (ARIMA com Models Neuronals Autoregressius), no son vàlides per a realitzar una predicció correcte de la volatilitat implícita en l'economia d'Estats Units.

Una de les raons per les quals es creu que no ha funcionat el pronòstic de la volatilitat implícita ha estat treballar amb els indicadors d'incertesa expressats en termes de volatilitat, tenint aquests una finestra mòbil de trenta dies, tal i com està expressat el propi VIX.

Seguint amb aquest argument i tenint en compte la metodologia empleada per a calcular aquestes noves variables, aquestes es van crear a partir de les últimes trenta observacions corresponents, ara bé, el VIX expressa la volatilitat esperada en els pròxims trenta dies, fet que podria condicionar la proposta de treballar en termes de volatilitat tenint en compte el valor del últim més.

De manera que una possible línia de recerca seria realitzar la predicció del VIX tenint en compte els indicadors d'incertesa política-econòmica i el indicador de risc geopolític, però sense que aquests dos estiguin expressats en termes de volatilitat. Per altra banda, també es podria comprovar el pronòstic aplicant altres tipologies de xarxes neuronals, com per exemple el cas de les LSTM (*Long Short – Term Memory*).

Índex de figures

Figura 1: Evolució de l'Índex d'incertesa politico-econòmica Global Font: Pagina web: www.policyuncertainty.com . Elaboració pròpia	9
Figura 2: Evolució de l'Índex d'incertesa politico-econòmica a USA. Font: Pagina web: www.policyuncertainty.com . Elaboració pròpia	11
Figura 3: Evolució de l'Índex d'incertesa política-econòmica a Espanya. Font: Banco de España. Documento de Trabajo nº 1906	12
Figura 4: Comparativa dels dos indicadors d'incertesa espanyola. Font: Banco de España. Documento de Trabajo nº 1906	13
Figura 5: Geopolitical Risk Index. Font: www.policyuncertainty.com/gpr.html	15
Figura 6: Geopolitical Risk index: els dos subíndex. Font: Measuring Geopolitical Risk, Caldara and Matteo	16
Figura 7: GPR i la influencia de la premsa. Font: Measuring Geopolitical Risk, Caldara and Matteo ..	17
Figura 8: BlackRock Geopolitical Risk Indicator, 2005-2020. Font: BlackRock Investment Institute ..	20
Figura 9: Evolució de l'Índex d'incertesa politica monetària. Font: Pagina web: https://www.policyuncertainty.com/monetary.html . Elaboració pròpia.	22
Figura 10: Comparativa GPR – EPU. Font: Measuring Geopolitical Risk, Caldara and Matteo	22
Figura 11: Comparativa GPR - VIX. Font: Measuring Geopolitical Risk, Caldara and Matteo	23
Figura 12: Indicadors sintètics d'incertesa. Font: Banco de España. Documento de Trabajo nº 1905.24	24
Figura 13: Evolució del VIX. Font: Yahoo Finance. Elaboració pròpia	28
Figura 14: Evolució del indicador EPU, període 1990-2020. Font: Elaboració pròpia	34
Figura 15: Evolució del indicador GPR, període 1990-2020. Font: Elaboració pròpia.	34
Figura 16: Evolució del indicador VIX, període 1990-2020. Font: Elaboració pròpia	35
Figura 17: Evolució de la volatilitat del EPU, període 1990-2020. Font: Elaboració pròpia	36
Figura 18: Evolució de la volatilitat del GPR, període 1990-2020. Font: Elaboració pròpia.	36
Figura 19: Evolució del VIX, període 1990-2020. Font: Elaboració pròpia	37
Figura 20: Sèrie temporal volatilitat EPU i els seus components. Font: Elaboració pròpia.	40
Figura 21: Sèrie temporal volatilitat GPR i els seus components. Font: Elaboració pròpia.	40
Figura 22: Sèrie temporal VIX i els seus components. Font: Elaboració pròpia	41
Figura 23: Correlograma del VIX. Font: Elaboració pròpia	44
Figura 24: Test de Dickey-Fuller Ampliat. Font: Elaboració pròpia.	45
Figura 25: Test de KPSS. Font: Elaboració pròpia	45
Figura 26: Sortida model univariant auto.arima. Font: Elaboració pròpia.	46
Figura 27: Sortida coeficients model ARIMA (2,1,1) univariant. Font: Elaboració pròpia.	46
Figura 28: Sortida model ARIMA(1,1,1) univariant. Font: Elaboració pròpia	47
Figura 29: Sortida coeficients ARIMA (1,1,1) univariant Font: Elaboració pròpia	47

Figura 30: Test de normalitat model ARIMA (1,1,1) univariant Shapiro-Wilk. Font: Elaboració pròpia.	47
Figura 31: Arrels estacionaries inverses pel model ARIMA(1,1,1) univariant. Font: Elaboració pròpia.	48
Figura 32: Sortida model multivariant auto.arima. Font: Elaboració pròpia.	50
Figura 33: Sortida coeficients ARIMA (2,1,1) multivariant. Font: Elaboració pròpia	50
Figura 35: Sortida coeficients ARIMA (1,1,1) multivariant Font: Elaboració pròpia.	51
Figura 36: Test de normalitat model ARIMA (1,1,1) multivariant Shapiro-Wilk. Font: Elaboració pròpia.	51
Figura 37: Arrels estacionaries inverses pel model ARIMA(1,1,1) multivariant. Font: Elaboració pròpia.	52
Figura 38: Xarxa neuronal simple. Font: Hyndman, R.J. Forecasting: principles and practice.....	54
Figura 39: Xarxa neuronal amb quatre entrades i una capa oculta. Font: Hyndman, R.J. Forecasting: principles and practice.	55
Figura 40: Arquitectura NNAR bàsica. Font: Métodos basados en deep learning.html	57
Figura 41: AR equivalent a una xarxa neuronal sense cap capa oculta (esquerra). I una xarxa neuronal autoregressiva amb n capes ocultes. (dreta). Font: AR-Net: A simple auto-regressive neural network for time – series.....	58
Figura 42: Sortida Xarxa Neuronal Autoregressiva (NNAR) univariant. Font: Elaboració pròpia.	59
Figura 43: Pronòstic Model NNAR univariant. Font: Elaboració pròpia.	60
Figura 44: Sortida Xarxa Neuronal Autoregressiva (NNAR) multivariant. Font: Elaboració pròpia.	61
Figura 45 Pronòstic Model NNAR multivariant Font: Elaboració pròpia.....	61

Índex de taules

Taula 1: Indicador de risc geopolític. Font: Elaboració pròpia.	31
Taula 2: Mostra indicadors setmana 2-8 març del 2020 . Font: Elaboració pròpia.	32
Taula 3: Estadístics descriptius de les variables. Font: Elaboració pròpia.....	33
Taula 4: Estadístics descriptius de les variables. Font: Elaboració pròpia.....	35
Taula 5: Coeficient correlació Pearson. Font: Elaboració pròpia.	38
Taula 6: Coeficient correlació Spearman. Font: Elaboració pròpia.	38
Taula 7: Coeficient correlació Kendall. Font: Elaboració pròpia.	38
Taula 8: Resultats de la predicció model ARIMA (1,1,1) univariant. Font: Elaboració pròpia.	48
Taula 9: Resultats de la predicció model ARIMA (1,1,1) multivariant. Font: Elaboració pròpia.	52
Taula 10: Resultats de la predicció model NNAR univariant. Font: Elaboració pròpia.	60
Taula 11: Resultats de la predicció model NNAR multivariant. Font: Elaboració pròpia.	62

Bibliografia

- Antonakakis, N., Gupta, R., Kollias, C. and Papadamou, S. (2017), "Geopolitical risks and the oil-stock nexus over 1899–2016", *Finance Research Letters*, Vol. 23, pp. 165-173.
- Bachmann, R., S Elstner y E. Sims (2013). «*Uncertainty and Economic Activity: Evidence from Business Survey Data*», *American Economic Journal: Macroeconomics*, 5, pp. 217-24
- Baker, S.R., Bloom, N., y Davis, S.J. (2015). *Measuring economic policy uncertainty (NBER Working Paper nº. 21633)*
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). *Measuring Economic Policy Uncertainty. The Quarterly Journal of Economics*, 131(4), 1593–1636
- Balcilar, M., Bonato, M., Demirer, R., Gupta, R. (2016) *Geopolitical Risk and Stock Market Dynamics of the BRICS*. Economic systems.
- Bali, T. G., Brown, S. J., & Tang, Y. (2017). *Is economic uncertainty priced in the cross-section of stock returns? Journal of Financial Economics*, 126(3), 471–489.
- BlackRock (2019) *Gauging Geopolitics: A framework to assess and price geopolitical risks*.
- BlackRock (2021) *Geopolitical risk dashboard*.
- BlackRock (2018) *Market Driven Scenarios: An Approach for Plausible Scenario Construction*.
- Bloom, N. (2009). *The Impact of Uncertainty Shocks. Econometrica* 77 (3), 623–685.
- Brogaard, J., & Detzel, A. (2015). *The Asset-Pricing Implications of Government Economic Policy Uncertainty. Management Science*, 61(1), 3–18
- Box, G., Jenkins, G. (1970) *Time Series analysis: forecasting and control*.
- CaixaBank Research (2018). *Geopolitical uncertainty and economics. Deep impact?*. International Economy.
- Caldara, Dario and Matteo Iacoviello (2018). *Measuring Geopolitical Risk. International Finance Discussion Papers 1222*.
- CBOE VIX (2019). *White Paper CBOE Volatility Index*.
- Cheng, C.H.J., & Chiu, C.W.J. (2018). *How Important Are Global Geopolitical Risks to Emerging Countries? International Economics*, 156 (1), 305–325.
- Davis, S. J. (2016). *An Index of global economic policy uncertainty*. National Bureau of Economic Research.
- Demir, E., Gozgor, G. and Paramati, S.R. (2019), "Do geopolitical risks matter for inbound tourism?", *Eurasian Business Review*, Vol. 9 No. 2, pp. 183-191.
- Donadelli, M (2014). *Google search – based metrics, policy – related uncertainty and macroeconomic conditions*. *Applied Economics Letters*, 801 – 807.
- Geoffrey Hinton (2012). *Neural Networks for Machine Learning*.

Ghirelli, C., Gil, M., Pérez, J.J. et al. *Measuring economic and economic policy uncertainty and their macroeconomic effects: the case of Spain*. *Empir Econ* **60**, 869–892 (2021).

Ghirelli, C, Pérez, J.J, Urtasun.A. *A new economic policy uncertainty index for Spain*. *Economics Letters*. Volume 182, Pages 64-67,

Gupta, R., Gozgor, G., Kaya, H., & Demir, E. (2019). *Effects of Geopolitical Risks on Trade Flows: Evidence from the Gravity Model*. *Eurasian Economic Review*, 9 (4), 515–530.

Hillier, D., y Loncan, T. (2019). *Political uncertainty and Stock returns: Evidence from Brazilian Political Crisis*. *Pacific-Basin Finance Journal*, 54, 1-12.

Husted, Lucas, John Rogers, and Bo Sun (2017). *Monetary Policy Uncertainty*. *International Finance Discussion Papers* 1215.

Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.

Hyndman, RJ (2018). *Dynamic Regression*. Monash University.

Kang, Wensheng and Lee, Kiseok and Ratti, Ronald A. (2013): *Economic Policy Uncertainty and Firm-Level Investment*. Forthcoming in: *Journal of Macroeconomics*

Knight, F.H. (1921) *Risk, Uncertainty, and Profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company

Liu, L., & Zhang, T. (2015). *Economic policy uncertainty and stock market volatility*. *Finance Research Letters*, 15, 99–105.

Lu, Z., Gozgor, G., Huang, M., & Lau, C. K. M. (2020). *The Impact of Geopolitical Risks on Financial Development: Evidence from Emerging Markets*. *Journal of Competitiveness*, 12(1), 93–107.

N.J.Nilsson (1998). *Introduction to Machine Learning*. Stanford University, Department of computer Sciene.

Pástor, L., & Veronesi, P. (2012). *Uncertainty about Government Policy and Stock Prices*. *The Journal of Finance*, 67(4), 1219–1264.

Pástor, L., y Veronesi, P. (2013). *Political uncertainty and risk premia*. *Journal of Financial Economics*, 110 (3), 520-545.

Saiz, A., and U. Simonsohn (2013): *Proxying for unobservable variables with internet document-frequency*. *Journal of the European Economic Association*, 11(1), 137-165.

Annexes

Codi RStudio

```
library(readxl)
library(quantmod)
library(writexl)
library(tseries)
library(forecast)
library(rlang)
library(zoo)
library(lmtest)
library(ggplot2)
library(neuralnet)
set.seed(1234)

# Importació de la base de dades dies complets

df= read.csv2("USA_Dades_final.csv")
names(df)[1] = "Fecha"
df$Fecha = as.Date(df$Fecha, format="%d/%m/%Y")
df$any = substr(df$Fecha, 1, 4)

# Construcció de les series temporals dels propis indicadors i de la seva
volatilitat.

# propis indicadors

evol.epu=ts(df$EPU, start = c(1990, 1), frequency = 365)
evol.gpr= ts(df$GPR, start = c(1990, 1), frequency = 365)

# volatilitat dels indicadors

vol.epu = ts(df$Vol_epu, start = c(1990, 1), frequency = 365)
vol.gpr = ts(df$Vol_gpr, start = c(1990, 1), frequency = 365)
vol.vix = ts(df$Vol_vix, start = c(1990, 1), frequency = 365)
```

```

data_int = ts.intersect(vol.epu,vol.gpr,vol.vix)
dim(data_int)
summary(data_int)

#Gràfics evolutius dels indicadors i de les seves volatilitats:

png("evol.epu.png")
plot(evol.epu,type="l",main="Evolutiu del indicador EPU",ylab="Valor",
col="grey80")
lines(rollmean(evol.epu,30),col='salmon')
dev.off()

png("evol.gpr.png")
plot(evol.gpr,type="l",main="Evolutiu del indicador GPR",ylab="Valor",
col="grey80")
lines(rollmean(evol.gpr,30),col='salmon')
dev.off()

png("evol.vol.vix.png")
plot(vol.vix,type="l",main="Evolutiu del indicador VIX",ylab="Valor",
col="grey80")
lines(rollmean(vol.vix,30),col='salmon')
dev.off()

png("evol.vol.epu.png")
plot(vol.epu,type="l",main="Evolutiu de la volatilitat EPU",ylab="Valor",
col="grey80")
lines(rollmean(vol.epu,30),col='salmon')
dev.off()

png("evol.vol.gpr.png")
plot(vol.gpr,type="l",main="Evolutiu de la volatilitat del
GPR",ylab="Valor", col="grey80")
lines(rollmean(vol.gpr,30),col='salmon')
dev.off()

```



```

png("evol.vol.vix.png")

plot(vol.vix,type="l",main="Evolutiu de la volatilitat del
VIX",ylab="Valor", col="grey80")

lines(rollmean(vol.vix,30),col='salmon')

dev.off()

# Descomposició de la sèrie temporal
#EPU
boxplot(vol.epu ~ cycle(vol.epu))
vol.epu.desc = decompose(vol.epu)
plot(vol.epu.desc,xlab="Any")

#GPR
boxplot(vol.gpr ~ cycle(vol.gpr))
vol.gpr.desc = decompose(vol.gpr)
plot(vol.gpr.desc,xlab="Any")

#VIX
boxplot(vol.vix ~ cycle(vol.vix))
vol.vix.desc = decompose(vol.vix)
plot(vol.vix.desc,xlab="Any")

# Correlacions
dat = data.frame(vol.epu,vol.gpr,vol.vix)
cor_pearson = cor(dat,method = c("pearson"))
cor_kendall = cor(dat,method = c("kendall"))
cor_spearman = cor(dat,method = c("spearman"))

# MODEL ARIMA

train_vix = ts(vol.vix,frequency = 365,start = c(1990,1),end=c(2015,365))
test_vix = ts(vol.vix[9491:11027],frequency=365, start = c(2016,1))
plot(train_vix)

p1 = ggAcf((train_vix),lag.max = 100)
p2 = ggPacf((train_vix),lag.max = 100)

```

```

adf.test(log(train_vix))
ndiffs(log(train_vix),test="adf")

kpss.test (log(train_vix),null="Trend")
ndiffs(log(train_vix),test="kpss")

# UNIVARIANT

arima.uni = auto.arima(log(train_vix))
arima.uni
coefstest(arima.uni)
arima.uni2 = arima(log(train_vix),order=c(1,1,1))
coefstest(arima.uni2)
r1.uni = ggAcf(arima.uni2$residuals)
r2.uni = ggPacf(arima.uni2$residuals)
shapiro.test(arima.uni2$residuals)
checkresiduals(arima.uni2)
autoplot(arima.uni2, main= "Arrels invertides sobre AR i MA")
forecast.uni = forecast(arima.uni2,h=1537)
accuracy(forecast.uni,log(test_vix))

# MULTIVARIANT
# diferenciació entre entrenament i test dels indicadors

matriu = cbind(log(df$Vol_epu[1:9490]),log(df$Vol_gpr[1:9490]))
matriu2 = cbind(log(df$Vol_epu[9491:11027]),log(df$Vol_gpr[9491:11027]))
model = auto.arima(log(train_vix),xreg=matriu)
coefstest(model)
matriu_epu=cbind(log(df$Vol_epu[1:9490]))
matriu_epu_test= cbind(log(df$Vol_epu[9491:11027]))
arima.epu = arima(log(train_vix), xreg=matriu_epu,order=c(1,1,1))
arima.epu
coefstest(arima.epu)

```

```

shapiro.test(arima.epu$residuals)
autoplot(arima.epu, main= "Arrels invertides sobre AR i MA")
forecast2 = predict(arima.epu,newxreg = matriu_epu_test)
plot(forecast2$pred)
lines(log(test_vix),col="salmon")
accuracy(forecast2$pred,log(test_vix))

# XARXA NEURONAL AUTOREGRESSIVA (NNAR)
# UNIVARIANT

fit.uni = nnetar(log(train_vix),lambda="auto")
fcast.uni = forecast(fit.uni,PI=TRUE,h=1537)
accuracy(fcast.uni,log(test_vix))

# MULTIVARIANT

fit.multi = nnetar(log(train_vix), xreg=matriu,lambda = "auto")
fcast.multi = forecast(fit.multi,xreg=matriu2,PI=TRUE)
accuracy(fcast.multi,log(test_vix))

# GRÀFIC

plot(fcast.uni,xlab = "Any",ylab="log(Train_Vix)")
lines(log(test_vix),col="salmon")

plot(fcast.multi,xlab="Any",ylab="log(Train_Vix)")
lines(log(test_vix),col="salmon")

```