TREBALL DE FI DE GRAU

Grau en Estadística

Títol: Models per a sèries temporals no estacionàries: L'exemple de la Covid-19.

Autor: Marc Ríos Masferrer

Director: David Moriña Soler

Departament: Econometria, Estadística i Economia Aplicada

Convocatòria: Juny 2021

UNIVERSITAT DE BARCELONA

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Resum

En aquest treball s'utilitzen els enfocaments clàssics/estocàstics i bayesians per modelitzar i analitzar el comportament de les sèries temporals del nombre d'infectats per la Covid-19 per cada 100.000 habitants a Nova Zelanda, Mèxic i Xina. També s'analitzen les previsions resultants de cada metodologia i se'n comparen els EPAM de cada ajust. Finalment, a partir de la inferència causal utilitzada en els anàlisis contrafactuals es quantifica l'efecte que han tingut les diverses mesures de confinament aplicades en cada país de l'estudi.

Paraules clau:

COVID-19; Sèries temporals; No estacionarietat; Metodologies d'anàlisi de sèries temporals estocàstiques; Metodologies d'anàlisi de sèries temporals bayesianes; Anàlisi causal

Abstract

This work uses classical stochastical and Bayesian approaches to model and analyze the behavior of the time series representing the number of Covid-19 cases for 100.000 habitants in New Zealand, Mexico and China. The forecasts produced by each methodology are also analyzed and their performance is evaluated. Finally, the causal inference used in the counterfactual analysis allow us to quantify the effect that the various containment measures applied in each considered country.

Keywords:

COVID-19; Time series; Non-stationarity; Stochastic time series analysis methodologies; Bayesian time series analysis methodologies; Causal analysis

Classificació AMS

Seguint la classificació publicada per l'American Mathematical Society (http://www.ams.org/mathscinet/msc/pdfs/classifications2010.pdf) el treball correspondria a les següents temàtiques:

- 37M10 Time sèries analysis
- 62F15 Bayesian inference
- 62M10 Time sèries, auto-correlation, regression, etc.
- 62M20 Prediction
- 62P10 Applications to biology and medical sciences

Índex

1. Introducció 3
    1.1. Presentació del treball 3
    1.2. Motivacions 3
    1.3. Hipòtesis i Objectius 3
    1.4. Metodologia 4
    1.5. Parts de la memòria 4
    1.6. Agraïments 5
2. Memòria 6
  I. INTRODUCCIÓ 6
    1. Dades 6
    2. Introducció a l'anàlisi de sèries temporals 7
    3. Introducció a l'anàlisi clàssica de sèries temporals 8
      3.1. Teoria models ARIMA 12
    4. Introducció a l'anàlisi bayesiana 13
      4.1. Teoria dels models BSTS 14
  II. PART PRÀCTICA: ESTIMACIÓ DELS MODELS CLÀSSICS VS BAYESIANS 18
    1. Sèrie I: Nombre d'infectats per la covid-19 a Nova Zelanda durant l'any 2020 18
      1.2. Identificació del model 19
      1.3. Anàlisi clàssica 21
        1.3.1. Estimació i validació 21
        1.3.2. Predicció 22
      1.4. Anàlisi bayesiana 23
        1.4.1. Estimació i validació 23
        1.4.2. Predicció 25
      1.5. Comparació dels diferents models obtinguts 26
        1.5.1. Període mostral 26
        1.5.2. Període extramostral 27
        1.5.3. Resum 28
    2. Sèrie II: Nombre d'infectats per la covid-19 a Mèxic durant l'any 2020 28
      2.2. Identificació del model 30
      2.3. Anàlisi clàssica 34
        2.3.1. Estimació i validació 34
        2.3.2. Predicció 35
      2.4. Anàlisi bayesiana 36

2.4.1. Estimació i validació 36
2.4.2. Predicció 39
2.5. Comparació dels diferents models obtinguts 40
2.5.1. Període mostral 40
2.5.2. Període extramostral 41
2.5.3. Resum 42
3. Sèrie III: Nombre d'infectats per la Covid-19 a Xina durant l'any 2020 42
3.2. Identificació del model 44
3.3. Anàlisi clàssica 46
3.3.1. Estimació i validació 46
3.3.2. Predicció 46
3.4. Anàlisi bayesiana 47
3.4.1. Estimació i validació 48
3.4.2. Predicció 50
3.5. Comparació dels diferents models obtinguts 51
3.5.1. Període mostral 51
3.5.2. Període extramostral 52
3.5.3. Resum 53
III. COMPARACIÓ DELS RESULTATS ENTRE SÈRIES 54
1. Metodologia clàssica 54
2. Metodologia bayesiana 56
3. Conclusions 59
IV. AMPLIACIÓ DELS MODELS BSTS 60
2. Mesures de confinament 62
3. Resultats 65
V. CONCLUSIONS 72
3. Conclusions 74
4. Bibliografia 76
5. Annexos 78

1. Introducció

El següent apartat consta d'un breu resum del treball realitzat. S'hi inclouen les motivacions, els objectius, la metodologia emparada i les parts del treball:

1.1. Presentació del treball

L'anàlisi de sèries temporals es la branques de l'estadística encarregada d'identificar patrons i tendències en una sèrie de valors ordenats cronològicament per tal de preveure el seu esdevenir.

En el següent treball s'ha realitzat un estudi pràctic de l'aplicació i modelització de sèries temporals tenint en compte diverses metodologies d'anàlisi diferents. Concretament, s'ha fet èmfasi en els punts de vista freqüentista estocàstic i bayesià, resultants de l'ajust de models ARIMA i BSTS. Per aquest exercici, s'han escollit unes dades referents al numero d'infectats per la Covid-19 en diversos països que han reaccionat de maneres diferents davant l'epidèmia.

Addicionalment, s'ha analitzat un dels possibles causants de l'obtenció de sèries tant distants a partir d'un mateix fenomen, com son les mesures preventives implementades a cada país estudiat. Per fer això s'ha realitzar un estudi contrafactual dels escenaris de confinament més destacables, per veure si han tingut el mateix impacte a cada país.

1.2. Motivacions

La motivació principal per escollir aquest tipus de treball recau en el desig d'ampliar i aprofundir en els conceptes d'anàlisi de sèries temporals observats a classe durant el primer quadrimestre de quart d'estadística. El futur o esdevenir, la incertesa que aquest provoca, i tots els possibles escenaris que es poden arribar a desenvolupar, és un dels conceptes que em generen més interès i curiositat. Un cop impartida l'assignatura, i fascinat pel tipus d'inferència i estimacions que aquesta permet fer en base a les dades, per preveure els possibles esdeveniments d'aquestes, va despertar en mi l'interès de seguir treballant aquesta disciplina. A més, donada la situació excepcional que ens ha imposat a viure diàriament la pandèmia de la malaltia coneguda com a Covid-19, on els casos detectats la setmana anterior podien provocar noves mesures restrictives que; o bé ens privessin de realitzar activitats o trobades, o bé ens retornessin algunes de les activitats que formaven part de la nostra vida diària. Em donava l'oportunitat perfecta per utilitzar dades d'actualitat per assolir el meu objectiu.

1.3. Hipòtesis i Objectius

Es considera que l'evolució de la pandèmia mundial provocada pel virus SARS-CoV-2 ha distat significativament entre els diferents països, degut a les aplicacions personalitzades de les mesures de control i prevenció efectuades per cada govern en qüestió. Un cop finalitzat el treball, es pretén comparar els diferents protocols utilitzats

en els països estudiats i, donada la naturalesa de sèries temporals de les dades, es vol ampliar els coneixements sobre la seva anàlisi. Més concretament, l'anàlisi de sèries temporals no estacionàries des d'un punt de vista estocàstic i complementar-ho des d'una aproximació bayesiana. L'enfocament bayesià resulta molt interessant en aquest estudi, donat que permet una estimació més realista i que no subestima tant la incertesa i variabilitat que presenta aquesta malaltia infecciosa.

Per assolir aquests objectius s'ha utilitzat una base de dades sobre el nº d'infectats de la Covid-19 en diferents països i s'han definit els següents objectius:

- Analitzar les sèries temporals no estacionàries a partir de metodologies clàssiques i bayesianes per cada país de l'estudi.

- Comparar els diferents enfocaments estadístics i valorar-ne el ajustos i rendiments de les diverses aproximacions.

- Avaluar l'impacte de les mesures implementades per fer front a l'epidèmia de Covid-19 a cada país a partir d'un anàlisi contrafactual.

1.4. Metodologia

Per al desenvolupament d'aquest treball s'ha pres la base de dades "COVID-19-geographic-disbtribution-worldwide-2020-12-14.xlsx" obtinguda en la web oficial del ECDC (European Centre for Disease Prevention and Control) (https://www.ecdc.europa.eu) i s'han seleccionat tres sèries de països diferents segons les puntuacions atorgades a cada país en el rànquing confeccionat per l'empresa de serveis financers Bloombergs, en l'article "Coronavirus: los mejores y peores países donde pasar la pandemia" (BBC News Mundo n.d.), publicat el 26 de novembre del 2020.

Posterior a l'elecció de les sèries, la qual s'explica detalladament a la introducció de la memòria, s'han estandarditzat les dades per facilitar-ne la comparació entre elles, treballant amb el número relatiu d'infectats per 100.000 habitants en tots els casos.

Després s'ha procedit a la modelització de sèries temporals amb R. Aquí s'han identificat, estimat i validat pas per pas models estocàstics ARIMA i models de sèries temporals estructurals bayesianes (BSTS) per cada país. Amb els models finals, escollits per cada metodologia, s'han realitzat previsions de les diverses sèries.

Finalment, per completar l'anàlisi i aprofundir en les aplicacions dels models BSTS, s'ha utilitzat la inferència causal que proporciona el paquet d'R "CausalImpact" per fer una anàlisi causal. Aquest permetrà analitzar l'impacte, de les mesures de prevenció de la pandèmia aplicades a cada país, en el nombre d'infectats observats a les respectives sèries.

1.5. Parts de la memòria

Per tal d'assolir els objectius plantejats s'ha definit la següent estructura:

- Identificació del tipus de dades a tractar
- Introducció teòrica als models clàssics i bayesians per l'anàlisi de sèries temporals
- Anàlisi estocàstica: ajust de models, validació i prediccions.
- Anàlisi bayesiana: ajust de models, validació i prediccions.
- Comparació del rendiment de les dues metodologies (ARIMA vs BSTS).
- Avaluació de l'impacte de les mesures implementades per fer front a l'epidèmia a cada país.

1.6. Agraïments

Abans de procedir amb el treball m'agradaria agrair a totes aquelles persones que m'han ajudat, tant en la realització del treball com en el transcurs de la carrera en general.

Pel que fa a la realització d'aquesta memòria, la dedicació, ajuda i temps dedicat pel meu tutor David Moriña Soler.

També, i imprescindible, el suport dels meus companys, amb els quals he afrontat la carrera i tots els desafiaments que ens ha portat el nostre pas per la universitat de Barcelona (UB) i posteriorment per la facultat d'estadística i matemàtiques (FEM).

Finalment, i no menys important, tota l'ajuda i suport rebut de part de la meva família. Sense ells no hauria estat possible res d'això.

2. Memòria

I. INTRODUCCIÓ

1. Dades

Per a la realització d'aquest treball s'ha escollit la base de dades "COVID-19-geographic-disbtribution-worldwide-2020-12-14.xlsx" obtinguda en la web oficial europea del ECDC (European Centre for Disease Prevention and Control). (https://www.ecdc.europa.eu)

En la base original es recullen, diàriament, les dades del nombre de morts i infectats per la pandèmia de la covid-19 al llarg de l'any 2020, per un total de 214 països diferents, per fer un seguiment territorial de l'impacte de la pandèmia. A partir d'aquestes dades, i enfocant-se en l'objectiu d'aquest treball, s'ha realitzat una selecció dels diversos països més interessants o rellevants segons l'enfocament i les mesures preventives i de contenció enfront la malaltia, per poder comparar-los posteriorment. S'ha buscat seleccionar metodologies de treball diferents i països amb característiques distintives per estudiar els diferents escenaris que han portat o haurien pogut portar a terme (Apartat: "IV.Ampliació dels models BSTS") a les dades en qüestió.

Per tal de fer aquesta selecció, s'ha basat la tria en les puntuacions atorgades a cada país segons el ranking proporcionat per l'empresa de serveis financers Bloomberg (Figura I.1), en l'article "Coronavirus: los mejores y peores países donde pasar la pandemia", publicat el 26 de novembre del 2020 (BBC News Mundo n.d.).

Clasificación de resiliencia al Covid-19

País	Puntuación promedio de Bloomberg
Nueva Zelanda	85,4
Japón	85
Taiwán	82,9
Corea del Sur	82,3
Finlandia	82

● ● ●

País	Puntuación promedio de Bloomberg
México	37,6
Argentina	41,1
Perú	41,6
Bélgica	45,6
Chequia	46,8

Fuente: Bloomberg BBC

Figura I.1: Inici i final del ranking proporcionat per l'empresa de serveis financers Bloomberg

Aquesta llista de països està ponderada segons: la situació i l'accés a la sanitat; el nombre de casos i morts per cada 100.000 habitant al mes; el total de morts; el percentatge d'infectats; l'accés a les vacunes; la qualitat de vida; les mesures restrictives; el creixement econòmic i l'índex de desenvolupament humà. De manera que intenta establir un mètode de valoració el més objectiu possible, considerant tots els factors claus i decisius econòmicament i sanitària, respecte la pandèmia en qüestió.

Basant-se en aquestes premisses, s'ha escollit treballar amb els següents països com a representants de diferents casos i situacions enfront la covid-19:

- El primer país de l'estudi és **Nova Zelanda**. S'escull en representació d'un dels països que millor ha reaccionat en front de la pandèmia. Nova Zelanda està format per un conjunt d'illes amb una població estimada de 4.942.500 habitants ($18{,}44\ hab/Km^2$). També és un clar exemple d'un programa de prevenció ràpid i contundent per controlar la pandèmia.

- Per contraposició, també s'inclourà l'anàlisi dels casos a **Mèxic**, considerat com l'últim país de la classificació donada la gran taxa de positius i morts per covid-19. Mèxic és una república federal comunicada territorialment amb diverses regions d'Estats Units amb una població de 124.777.324 habitants ($63{,}26\ hab/Km^2$). A diferencia de Nova Zelanda, la densitat de població és molt major i encara no s'ha aconseguit controlar significativament la propagació del virus.

- Finalment també s'ha decidit analitzar la **Xina**. Aquí és on va aparèixer el primer cas registrat i també el primer país que es va veure obligat a desenvolupar mesures preventives sense gaire informació prèvia de les característiques i naturalesa del virus. La Xina és una república popular amb una població de 1.409.517.397 habitants ($146{,}87\ hab/Km^2$). Curiosament, aquest és l'únic país de l'estudi amb un creixement positiu en el seu PIB, durant l'any d'estudi.

Un cop seleccionats els diversos països, s'ha decidit estandarditzar les sèries corresponents per facilitar-ne les comparacions. Per fer això s'ha passat la sèrie del nombre cru de casos de Covid-19 a la taxa d'infectats per 100.000 habitants, fent un ajust segons el cens de població durant el 2020 de cadascun dels països considerats.

En conclusió, per aquest treball s'ha escollit treballar amb les sèries temporals, corresponents a l'any 2020, del número d'infectats per 100.000 habitants per la Covid-19 a Xina, Mèxic i Nova Zelanda.

2. Introducció a l'anàlisi de sèries temporals

Una sèrie temporal és una seqüencia de N observacions, ordenades cronològicament, sobre una característica (sèrie escalar) o diverses característiques (sèrie vectorial) (Peña Sánchez de Rivera 2005).

Les sèries escalars que es treballaran durant el treball, es representen matemàticament com:

$$y_1, y_2, \dots, y_N \; ; (y_t)_{t=1}^{N} \; ; (y_t : t = 1, \dots, N)$$

$on\ y_t$ és la observació nº t $(1 \leq t \leq N)$ de la sèrie i N és el nº total d'observacions que la compon.

Les N observacions es poden recollir en un vector columna -> $y \equiv [y_1, y_2, \dots, y_N]'$ d'ordre Nx1.

Totes les sèries es componen per una combinació o el conjunt dels següents components:

$$X_t = T_t + S_t + I_t$$

$Valor\ observat\ (X_t) = \text{Tendència}\ (T_t) + Estacionalitat\ (S_t) + Component\ irregular (I_t)$

Tendència: comportament o moviment de la sèrie al llarg del temps.

Estacionalitat: moviments d'oscil·lació dins del període d'un any.

Component irregular: variacions aleatòries dels components anteriors.

L'objectiu de l'anàlisi economètrica d'una sèrie temporal consisteix en elaborar un model estadístic que reculli les oscil·lacions d'aquesta permetent-ne preveure el resultat a futur més probable. Per tal de fer això, es poden prendre diversos enfocaments: Anàlisi determinista, estocàstica, bayesiana,... Concretament l'estudi profunditzarà en l'anàlisi estocàstica i l'anàlisi bayesiana.

3. Introducció a l'anàlisi clàssica de sèries temporals

Un procés estocàstic és una seqüencia de variables aleatòries, ordenades i equidistants cronològicament, referides a una (procés escalar) o varies (procés vectorial) característiques d'una unitat observable.

A l'hora de treballar amb un procés estocàstic cal identificar si aquest és estacionari o no. Un procés estocàstic és estacionari quan les propietats estadístiques de qualsevol seqüencia finita d'ell mateix (sèrie temporal) són similars per qualsevol segmentació. Això implica que totes les variables aleatòries que el componen estan idènticament distribuïdes, independentment del moment del temps en el qual han estat generades. En aquest cas es considera que les propietats són constants al llarg del temps i això en facilita l'obtenció de prediccions, permetent fer ús dels valors constants de la mitjana per obtenir observacions futures i generar intervals de predicció.

Per altra banda, si no es compleix aquesta condició el procés estocàstic és no estacionari. Les propietats estadístiques en un procés no estacionari són més complexes, però es pot intentar modelitzar a partir d'alguna transformació senzilla per tal de definir-ne la seva

estructura probabilística completa a partir d'una única realització finita del mateix procés.

Els processos escalars que és treballaran durant el treball, és representen matemàticament com:

$$\dots, Y_{-1}, Y_0, Y_1, \dots ; (Y_t : t = 0, \pm 1, \pm 2, \dots); (Y_t)$$

$on\ Y_t$ és una variable aleatòria escalar referida a la unitat observable considerada en el moment t.

Segons el teorema de Wold, qualsevol procés estocàstic (y_t) es pot representar per la suma d'un procés de soroll blanc (ϵ_t) i un de purament determinista (z_t).

$$y_t = z_t + \sum_{j=1}^{\infty} \psi_j * \epsilon_{t-j}$$

La part no determinista es pot escriure com el resultat d'una transformació lineal del procés de soroll blanc (procés estocàstica constant amb esperança zero i sense correlació estadística). Basant-nos en aquesta definició, podem trobar 3 tipus diferents, processos autoregressius (AR), mitjana mòbil (MA) i mixtes (ARMA):

Els **processos autoregressius** o AR

$$y_t = \sum_{i=1}^{p} \phi_i * y_{t-i} + \epsilon_t \sim AR(p)$$

suposen que el present de la sèrie depèn única i directament dels p-valors immediatament anteriors a aquesta (Rodríguez and Vírseda 2019). El problema d'aquesta casuística recau en que no poden representar bé sèries de memòria molt curta, on l'últim valor està correlacionat amb molts pocs valors històrics. Aquests models compleixen les següents propietats:

- Són invertibles
- És estacionari si i només si es invertible i estable (no te tendència determinista).
- $E[y_t] = \mu$
- $Var[y_t] = E[y_t - \mu]^2 = \gamma_0$
- Autocovariància: $\gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)] = \phi^k * \gamma_0\ ;\ on\ k = 1, 2, ..$
- Funció d'autocorrelació simple (FAS): $\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k$
- Funció d'autocorrelació parcial (FAP): $\alpha_k = \frac{\rho_k - \rho_{k-1}^2}{1 - \rho_{k-1}^2}$

Els **processos mitjana mòbil** o MA

$$y_t = \epsilon_t + \sum_{j=1}^{q} -\theta_j * \epsilon_{t-j} \sim MA(q)$$

generen cada valor de la sèrie com una mitjana ponderada de pertorbacions aleatòries amb un retràs de q períodes. Aquests models compleixen les següents propietats:

- Són estacionaris
- L'operador de retard és invertible si les seves arrels estan fora del cercle unitari.
- És estacionari si i només si es invertible i estable.
- $E[y_t] = \mu$
- $Var[y_t] = E[y_t - \mu]^2 = \gamma_0$
- Autocovariància: $\gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)] = \phi^k * \gamma_0 \; ; \; on \; k = \; 1, 2, ..$
- Funció d'autocorrelació simple (FAS): $\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k$
- Funció d'autocorrelació parcial (FAP): $\alpha_k = \frac{\rho_k - \rho_{k-1}^2}{1 - \rho_{k-1}^2}$

Els **processos mixtes** o autoregressiu mitjana mòbil (ARMA)

$$y_t = \sum_{i=1}^{p} \phi_i * y_{t-i} + \epsilon_t - \sum_{j=1}^{q} \theta_j * \epsilon_{t-j}$$

són més complexos que els dos anteriors, donat que en representen la unió dels dos. Aquí es te tant en compte que els nous valors depenen directament dels valors passats com que depenen d'una successió finita d'aquests. Aquests models compleixen les següents propietats:

- $E[y_t] = \mu$
- $Var[y_t] = E[y_t - \mu]^2 = \gamma_0$
- Autocovariància: $\gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)] = \phi^k * \gamma_0 \; ; \; on \; k = \; 1, 2, ..$
- Funció d'autocorrelació simple (FAS): $\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k$
- Funció d'autocorrelació parcial (FAP): $\alpha_k = \frac{\rho_k - \rho_{k-1}^2}{1 - \rho_{k-1}^2}$

Per poder modelitzar els diferents processos vistos, s'ha de complir la propietat de l'estacionarietat, encara que a la pràctica, la majoria de sèries temporals no es poden considerar generades per processos estocàstics estacionaris. Això passa perquè

acostumen a presentar certes tendències al llarg de la seva evolució temporal, no presenten una dispersió constant (són estacionals) o una combinació de les dues característiques anteriors. També es pot donar el cas que presentin efectes estacionals o patrons al llarg del temps. No obstant això, moltes d'aquestes sèries es poden transformar per adoptar una forma d'aparença estacionària, la qual cosa en permet modelitzar aquestes variacions a partir de models estadístics similars als models ARMA utilitzats en les sèries estacionals. Els models ARIMA (la sèrie presenta no estacionarietat), SARIMA (la sèrie presenta no estacionarietat i estacionalitat), ...

Donat que aquest treball es centra en l'anàlisi de processos estocàstics escalars no estacionaris caldrà tenir en compte aquestes transformacions en la sèrie original.

Les transformacions més utilitzades són les que busquen l'estacionarietat de la variància (Box-Cox) i la estacionarietat de la mitjana (Diferències regulars):

- **Transformacions de Box-Cox:**

Sigui (Y_t) un procés estocàstic no estacionari tal que

$$\mu_t \equiv E[Y_t] \; i \; \sigma_t^2 \equiv Var[Y_t]$$

existeixen i depenen de t (no constants). Si

$$\sigma_t^2 = \sigma^2 * h(\mu_t)^2,$$

on σ^2>0 és una constant i h(·) és una funció real tal que h(·)≠ 0 $\forall \mu_t$, llavors una transformació estabilitzadora de la variància de (Y_t) es qualsevol funció real g(·) que compleixi:

$$g'(\cdot) = \frac{1}{h(\cdot)} \; \forall \mu_t$$

D'altre manera, $Y_t' \equiv g(Y_t) = \frac{(Y_t+m)^\lambda - 1}{\lambda}$ és denomina com una transformació de Box-Cox de paràmetre λ i m on $|\lambda| \leq 2$ i el -1 del numerador s'utilitza perquè $\lim_{\lambda \to 0}\{\frac{(Y_t+m)^\lambda - 1}{\lambda}\} = \ln{(Y_t + m)}$.

- **Diferències regulars:**

L'operador de diferències regulars d'ordre d ($d \geq 1 \; enter$) es defineix com $\nabla^d \equiv (1-B)^d$ on B es un operador de retard, de manera que $\nabla^1 Y_t \equiv Y_t - Y_{t-1}, \nabla^2 Y_t \equiv Y_t - 2Y_{t-1} + Y_{t-2}$, …

Un procés estocàstic (Y_t) es integrat d'ordre d si i només si el procés $\nabla^d Y_t$ segueix un model ARMA(p,q) estacionari i invertible.

Es poden trobar mes exemples i característiques sobre els models utilitzats en l'anàlisi clàssica de sèries temporals en els llibres exposats a la bibliografia (Mauricio 2007).

3.1. Teoria models ARIMA

Donat que es treballa amb unes dades temporals no estacionaries i sense estacionalitat, cal fer ús dels models ARIMA amb les transformacions vistes anteriorment per ajustar-ne l'estacionarietat. Es fa ús del model ARIMA per contra del ARMA donat que aquest permet incloure les variacions causades per la no estacionarietat (BOX et al. 2016; Nishi 2012).

Un procés estocàstic (Y_t) es integrat d'ordre d si i només si (Y_t) segueix un model *AutoRegressive-Integrated-Moving Averge* d'ordre (p, d, q) o ARIMA(p, d, q):

$$\phi_p(B)\Phi_{\mathrm{P}}(B^S)[\nabla^d\nabla_S^D Y_t' - \mu_W] = \theta_q(B)\Theta_Q(B_S)A_t$$

Aquesta seria la representació més general del model ARIMA, tenint en compte totes les característiques dels models ARMA i els processos no estacionaris. Per tal d'identificar totes les parts d'aquesta formula s'ha de seguir el següent procés:

Partint d'una sèrie temporal $(y_t)_{t=1}^N$ i un període estacional S:

- S'identifiquen λ i m segons les transformacions de Box-Cox utilitzades per l'estacionarietat de la variància per obtenir: $(y_t') = \begin{cases} \ln(y_t + m) & si\ \lambda = 0 \\ (y_t + m)^\lambda & altrament \end{cases}$
- S'identifica d i D (nº de diferències efectuades en la sèrie original abans de poder ajustar un model ARMA) segons el gràfic temporal de tendències o la ACF mostral de (y_t') per obtenir: $w_t \equiv \nabla^d\nabla_S^D y_t'$
- S'obté μ_W a partir del contrast de significació de $\mu_w \equiv E[W_t] \equiv E[\nabla^d\nabla_S^D Y_t']$ fent ús de la mitjana mostral de la sèrie (w_t)
- Finalment s'identifica p, P, q i Q a partir de les ACF i PACF mostrals de la sèrie (w_t), comparades amb les ACF i PACF teòriques de models $ARMA(p,q)xARMA(P,Q)_S$

Un cop identificat el model ARIMA de la sèrie temporal estudiada, cal estimar-ne els seus components i procedir a la seva validació, abans de fer-ne ús per obtenir prediccions de la sèrie.

Per tal d'estimar els paràmetres del model identificat, es parteix de la sèrie, suposada estacionaria, $(w_t)_{t=1}^n \equiv (\nabla^d\nabla_S^D y_t')_{t=1}^n$ del procés $(W_t) \equiv (\nabla^d\nabla_S^D Y_t')$, s'escull un mètode d'estimació com l'estimador màxim versemblant (MV):

- $\mathrm{MV} \Leftrightarrow \mathrm{Min}(\widetilde{\mathrm{w}}'\sum^{-1}\widetilde{\mathrm{w}})|\sum|^{1/n}$ on $\sum = \sigma_A^{-2}Var[W]$
- Donat el problema de l'estimació no lineal de la forma quadràtica de $\widetilde{\mathrm{w}}'\sum^{-1}\widetilde{\mathrm{w}}$ i el determinant de $\sum$ es proposa l'alternativa: $\mathrm{MC} \Leftrightarrow \mathrm{Min}(\widetilde{\mathrm{w}}'\sum^{-1}\widetilde{\mathrm{w}})$

Després de l'etapa d'estimació s'obté un model estimat ($\hat{\mu}_W$; $\hat{\phi} \equiv \left[\hat{\phi}_1, \dots, \hat{\phi}_p\right]'$; $\hat{\Phi} \equiv \left[\hat{\Phi}_1, \dots, \hat{\Phi}_p\right]'$; $\hat{\theta} \equiv \left[\hat{\theta}_1, \dots, \hat{\theta}_p\right]'$; $\hat{\Theta} \equiv \left[\hat{\Theta}_1, \dots, \hat{\Theta}_p\right]'$; $\hat{\sigma}_A^2$), una matriu de covariàncies estimades (V) i una sèrie de residus $(\hat{a}_t)_{t=1}^n$.

Abans de poder fer ús d'aquestes estimacions per a la predicció de valors futurs de la sèrie, cal validar-ne les següents característiques:

- Significació individual i conjunta dels paràmetres del model (es poden eliminar els no significatiu).
- Correlacions menyspreables en la matriu de correlacions (sinó, hi ha sobre parametrització).
- Estacionarietat de la part AR i invertibilitat de la part MA.
- Normalitat dels residus (pot presentar errors en les transformacions de Box-Cox).

Sinó es compleixen algunes de les condicions anteriors, es pot plantejar un model diferent i tornar a repetir els processos d'estimació i validació. En cas contrari es pot procedir a fer prediccions de la sèrie.

4. Introducció a l'anàlisi bayesiana

A vegades utilitzar els mètodes d'anàlisi de sèries temporals clàssics no es la millor opció, ja sigui perquè requereixen de que es compleixin moltes propietats estadístiques, o directament perquè els canvis que presenten les sèries no sempre son ajustables per aquests models. Això ha creat la necessitat de buscar altres tipus de models que no parteixin dels mateixos supòsits (VALENCIA CARDENAS, CORREA MORALES, and SERNA 2015).

Els models vistos anteriorment, com el model ARIMA, incorporen característiques del passat de la mateixa sèrie, segons la seva autocorrelació. Inclouen p variables autoregressives (model AR) i q termes dels errors del passats (model MA) per tal d'ajustar les dades estudiades, donant un enfocament totalment objectiu, sense donar joc a les suposicions subjectives.

Per altre banda, l'anàlisi bayesiana neix de l'ampliació d'aquestes inferències clàssiques, sumant-li una interpretació més subjectiva de la probabilitat.

Aquí es consideren paràmetres de distribucions de probabilitat com a variables aleatòries θ, sobre les quals tenim una informació a priori, quantificada en una distribució de probabilitat:

$$Cas\ discret\colon \{\Pr(\theta_1), \Pr(\theta_2), \dots\}\ on \sum_j \Pr(\theta_j|H) = 1$$

$$Cas\ continu\text{:}\ \{\mathrm{p}(\theta|H)\,, \theta \in \Theta\}\ on\ \int_{\Theta} \mathrm{p}(\theta|H)\ d\theta = 1$$

$$on\ H\ són\ les\ condicions\ anteriors\ i\ \theta\ la\ variable\ aleatòria$$

amb una funció de densitat, suposadament, coneguda i una informació mostral $(y_1, \dots, y_n)$ amb la seva funció de versemblança corresponent $(L(y_1, \dots, y_n|\theta))$. A partir d'aquesta informació a priori, i fent ús del teorema de Bayes, se'n busca la distribució a posteriori. Donat que és molt complicat associar la distribució a posteriori amb una distribució coneguda, es fa ús de processos de mostreig per trobar valors de la variable aleatòria θ. Alguns dels mètodes bayesians amb els que es treballa són els models lineals dinàmics bayesians (MLDB), els models de regressió lineal bayesiana (RLB), els models de sèries temporals estructurals bayesianes (BSTS), ... En aquest treball es farà èmfasi en els models BSTS.

Per tal de constituir un model bayesià, es molt important conèixer el teorema de Bayes. Aquest teorema apareix amb la necessitat d'atribuir una nova definició al terme probabilitat, sobre el grau de convicció sobre la certesa d'una hipòtesi, més que no pas el fet de ser la freqüència relativa d'un succés.

El teorema de Bayes calcula les probabilitats de que succeeixi un esdeveniment A condicionada un altre esdeveniment (B). Si els dos esdeveniments són decrets s'expressa com:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

En casos de dues variables o esdeveniments continus es representa com:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

Si es vol ampliar la informació sobre el context històric que va portar a la necessitat de fer ús de l'estadística bayesiana, o es vol entrar més en detall sobre el teorema esmentat o les interpretacions d'aquest es pot complementar amb les següents lectures: (Bernardo, 2002; Solhjell, 2009; Patrícia Oliveres Luna, 2017).

4.1. Teoria dels models BSTS

El model de sèries temporals estructurals bayesianes o BSTS es una tècnica estadística utilitzada per l'anàlisi de sèries temporals entre d'altres. Aquest model amplia la mateixa estructura vista pels models ARIMA, afegint una metodologia diferent: les distribucions a priori i a posteriori.

La implementació d'aquest model varia segons si la sèrie estudiada presenta o no una component regressiva:

En el cas de **no tenir una component regressiva**, com podria ser un model que només depengués de valors passats:

$$y_t = \mu_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$
$$\mu_{t+1} = \mu_t + \xi_t \quad \xi_t \sim N(0, \sigma_\xi^2)$$

Els paràmetres del model serien les variàncies dels termes d'errors, per tant, necessitaríem especificar una distribució a priori per definir-los. Després, i basant-nos en les dades numèriques, l'ajustaríem a una distribució a posteriori fent ús del mètode de mostreig de la cadena de Markov de Monte Carlo (MCMC). Aquest mètode consisteix es un algorisme de mostreig a partir de distribucions de probabilitat.

Per altre banda, si la sèrie temporal presenta components regressius, seria necessari aplicar el mètode de l'espiga i llosa. En aquest cas, es considera com si aquests coeficients del model fossin fixats en el temps, per evitar que aquests es poguessin afegir com a variables d'estat addicionals.

- **Mètode de l'espiga i llosa:**

Com s'observa en l'exemple de "Predicting the present with Bayesian structural time sèries" (Scott and Varian 2014) i "Bayesian Variable Selection for Nowcasting Economic Time Sèries" (Scott and Varian 2013), el mètode de l'espiga i llosa és una tècnica de selecció de variables bayesianes.

L'espiga (π) representa la probabilitat de que un coeficient del model sigui zero, mentre que la llosa seria la distribució a priori dels valors del coeficient de regressió.

Per tal d'identificar l'espiga, suposant un vector γ on $\begin{matrix}\gamma_k = 1 \; si \; \beta_k \neq 0 \\ \gamma_k = 0 \; si \; \beta_k = 0\end{matrix}$ on β és el vector dels coeficients de regressió del model. S'utilitza β_γ per seleccionar el subconjunt del vector β on $\gamma = 1$, i per conseqüent, $\beta \neq 0$.

Basant-se en això, la distribució a priori de la γ seguiria una Bernoulli independent on per cada component de γ_k de γ, π_k es la probabilitat de que aquest sigui igual a 1:

$$\gamma \sim \prod_{k=1}^{K} \pi_k^{\gamma_k} (1 - \pi_k)^{1-\gamma_k}$$

Per poder especificar el valor de π_k, és necessari definir també la llosa. Suposant una matriu simètrica Ω^{-1} on Ω_γ^{-1} representa la submatriu, les files i columnes de la qual corresponen als índex K on $\gamma_k = 1$. La llosa a priori seria:

$$\beta_\gamma | \sigma_\epsilon^2, \gamma \sim N(b_\gamma, \sigma_\epsilon^2 \Omega_\gamma)$$
$$\frac{1}{\sigma_\epsilon^2} | \gamma \sim Gamma \left(\frac{v}{2}, \frac{ss}{2}\right)$$

Si es multipliquen les tres distribucions a priori anteriors se'n obté la distribució a priori completa:

$$p(\beta, \gamma, \sigma_\epsilon^2) = p(\beta_\gamma|\gamma, \sigma_\epsilon^2)p(\sigma_\epsilon^2|\gamma)p(\gamma)$$

A continuació es procedeix a identificar els paràmetres "ss" i "v" utilitzats en les distribucions anteriors. El "ss" és la suma a priori dels quadrats i "v" és la mida mostral prèvia a la implementació del paquet d'R "bsts" (Steven Scott and Steven Scott 2020). La manera d'identificar aquests valors consisteix en definir un R^2 esperat a partir de la regressió i una mida del model esperada. D'aquesta manera, i seguint la igualtat:

$$\frac{ss}{v} = (a - R^2)s_y^2$$

on s_y^2 la desviació estàndard marginal de la variable resposta. Es troben els valors dels paràmetres de la distribució a priori. Altrament, els valors per defecte són $R^2 = 0{,}5$, v=0,01 i $\pi = 0{,}5$.

Per especificar el π_k cal basar-se en la informació i coneixement previs que es coneix de la seria estudiada. Cal formular-se la següent pregons: Com de segurs estem de que β_k sigui diferent a zero? Si el grau de certesa és elevat, i per conseqüent es creu que β ha de pertànyer al model, π_1 pot pendre valor 0,9. Per contra, sinó es tenen coneixements previs sobre quins components formen part del model, π_k pot prendre valor 0,5 $\forall\ k \in Model$, fent ús del model per defecte, formulat pel paquet d'R "bsts".

Per triar la matriu de covariància inversa (Ω^{-1}) cal basar-se en el model de la matriu X. Donat que el paquet "bsts" d'R estableix que $\Omega^{-1} = \frac{k}{n}X^TX$, per un valor específic de k, això reflecteix el coneixement previ sobre la proximitat de la β amb el vector mitjà anterior. El valor per defecte del paquet d'R és k=1.

Un cop definits aquests paràmetres es prossegueix a l'ajust de la distribució a posteriori de la llosa, seguint les igualtats i premisses següents:

- Z_t^* és la matriu d'observacions Z_t amb $\beta^T x_t = 0$
- $y_t^* = y_t - Z_t^{*T}\alpha_t$
- $y^* = (y_1^*, \dots, y_n^*)^T$

Es poden identificar les distribucions a posteriori:

$$\beta_\gamma \left| \sigma_\epsilon, \gamma, y^* \sim N\left(\widetilde{\beta_\gamma}, \sigma_\epsilon^2 V_\gamma\right) \quad i \quad \frac{1}{\sigma_\epsilon^2} \right| \gamma^* \sim Gamma\ \left(\frac{N}{2}, \frac{SS_\gamma}{2}\right)$$

on:

$$V_\gamma^{-1} = (X^TX)_\gamma + \Omega_\gamma^{-1}$$

$$\widetilde{\beta_\gamma} = V_\gamma(X_\gamma^T y^* + \Omega_\gamma^{-1} b\_\gamma)$$

$$N = v + n$$

$$SS_\gamma = ss + {y^*}^T y^* + b_\gamma^T \Omega_\gamma^{-1} b_\gamma - \widetilde{\beta_\gamma}^T V_\gamma^{-1} \widetilde{\beta_\gamma}$$

I també s'identifica la distribució a posteriori de y:

$$\gamma | y^* \sim C(y^*) \frac{\left|\Omega_\gamma^{-1}\right|^{\frac{1}{2}} \quad p(\gamma)}{\left|V_\gamma^{-1}\right|^{1/2} \quad SS_\gamma^{\frac{N}{2}-1}}$$

Finalment, per obtenir una mostra de la distribució a posteriori, es fa ús del mètode de mostreig de la cadena de Markov de Monte Carlo (MCMC).

II. PART PRÀCTICA: ESTIMACIÓ DELS MODELS CLÀSSICS VS BAYESIANS

En aquest apartat s'identificaran i estimaran els models estadístics, segons les diverses metodologies introduïdes, per a les diferents sèries estudiades. Per cada sèrie se'n ajustarà, validarà i és realitzaran les prediccions adients per cada tipus de model, amb la intenció de, posteriorment, comparar-ne els resultats segons els enfocaments estadístics utilitzats.

1. Sèrie I: Nombre d'infectats per la covid-19 a Nova Zelanda durant l'any 2020

La primera sèrie de l'estudi consisteix en el nombre d'infectats per la Covid-19 a Nova Zelanda durant l'any 2020 (Figura II.1). Aquesta recull diàriament el nombre d'infectats pel virus, reportats pel govern i els centres mèdics, iniciant el dia 1 de gener del 2020 i finalitzant el 14 de desembre del mateix any. Per facilitar la validació i l'ajust dels models proposats a continuació, s'ha decidit prendre els 11 primers mesos com a període mostral, deixant les ultimes dues setmanes com a període extramostral.

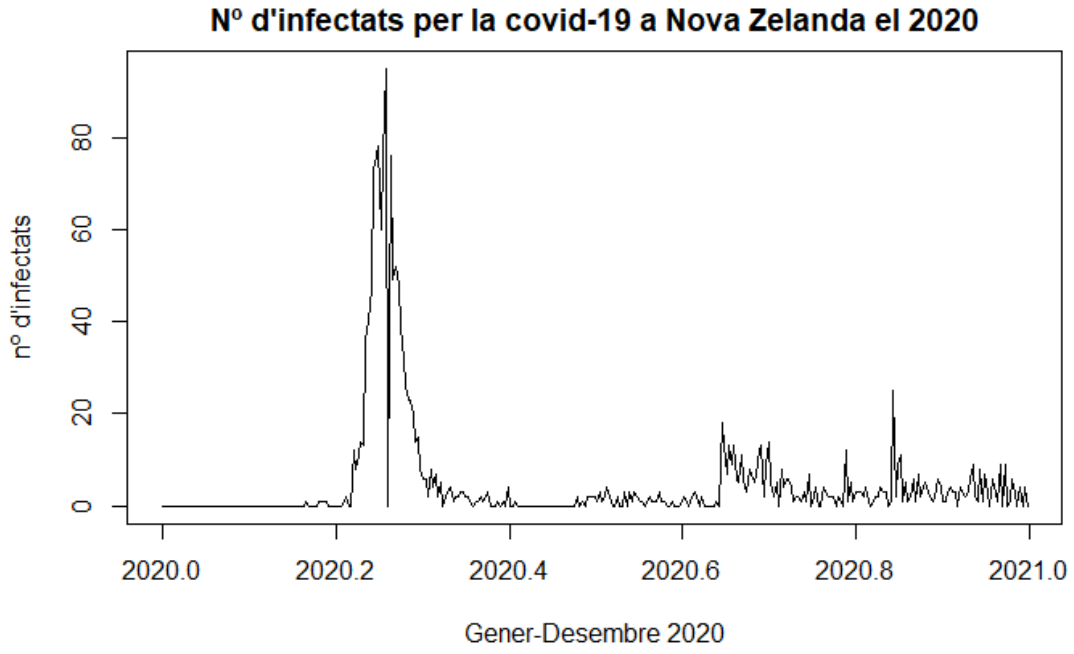


Figura II.1: Gràfic de la sèrie I:Nº d'infectats a Nova Zelanda

Abans de procedir a la modelització de la sèrie, i com s'ha esmentat prèviament, s'utilitza el cens de la població de Nova Zelanda durant l'any de les dades per estandarditzar les unitats i passar del nombre d'infectats al nombre d'infectats per 100.000 habitants seguin la formula que es veu a continuació:

$$Y = S * \frac{100.000\ habitants}{4.822.232\ habitants\ a\ Nova\ Zelanda}$$

on S és la sèrie original i Y la resultant de la transformació.

Aquest procés es necessari per tenir les tres sèries temporals amb les mateixes unitats de mesura, facilitant-ne així les comparacions i els comportaments d'aquestes.

Finalment, la sèrie a analitzar és la que es representa en la Figura II.2:

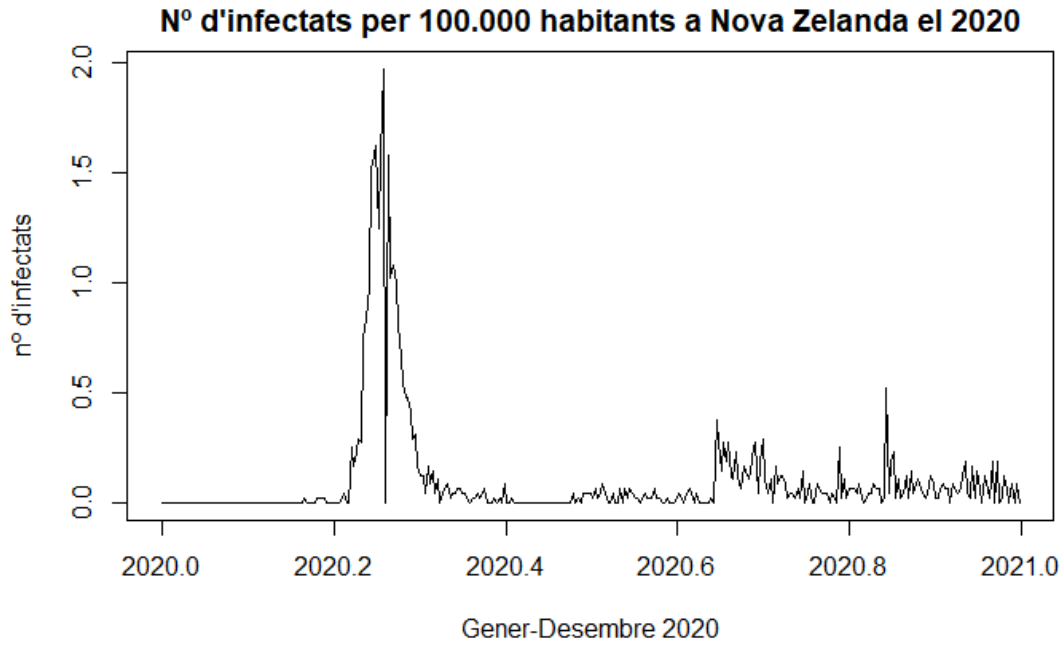


Figura II.2: Gràfic de la sèrie I ajustada

1.2. Identificació del model

Donada la sèrie temporal I, es pretén identificar els diversos components que la formen: Tendència, estacionalitat, ... També cal recordar que per poder modelitzar-la, aquesta ha de complir el supòsit d'estacionarietat mencionat anteriorment. Donat que les sèries utilitzades en l'anàlisi són no estacionaries caldrà aplicar diverses transformacions, segons els components que les conformin, per poder assolir aquesta propietat.

Primerament s'estudia si la sèrie presenta algun tipus de tendència. Per fer això és crea un model lineal amb les dades com a variable dependent i un vector seqüencial del 1 fins al nombre total de dades recollides per la sèrie com a variable explicativa. Utilitzant aquest model es busca quantificar l'efecte que té aquest nou vector, donat que si fos estadísticament significatiu implicaria que ajuda a explicar els valors de la variable dependent, i com a conseqüència, que existeix una tendència.

Aplicant aquests conceptes s'obté un p-valor de 0,0834, superior al 5% de significació, trobant-se al límit d'acceptar o no que la sèrie presenta una tendència. En aquest cas considerem que hi ha tendència i és necessari corregir-la. Conseqüentment, s'apliquen les diferències regulars per intentar estabilitzar la mitjana i se'n grafiquen els resultats (Figura II.3).

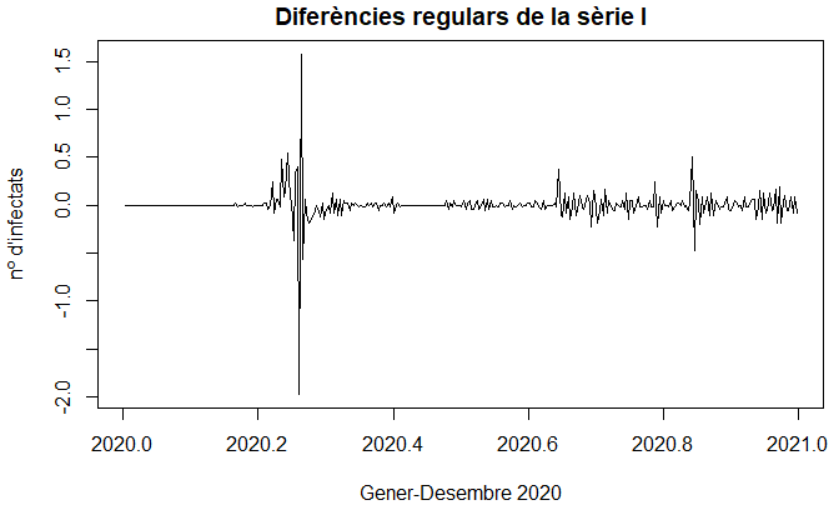


Figura II.3: Gràfic de les diferències regulars de la sèrie I

S'observa com la mitjana s'ha estabilitzat al voltant del zero i sembla ser constant al llarg de tota la sèrie. Si es repeteix la comprovació numèrica per la nova sèrie transformada, s'observa com el coeficient de la pendent a deixat de ser significativament diferent de zero (p-valor = 0,908).

Per altra banda, a excepció d'unes poques observacions, possiblement atípics, la variància també sembla mantenir-se constant durant tot el període i el gràfic no presenta patrons ni cicles fàcilment identificables. Es pot assumir doncs, una variància constant i l'absència de la component estacional.

Per acabar amb el procés d'identificació, es comproven els gràfics de les funcions d'autocorrelació simple (Figura II.4) i de la funció d'autocorrelació parcial (Figura II.5) per definir els components mitjana mòbil i/o autoregressius:

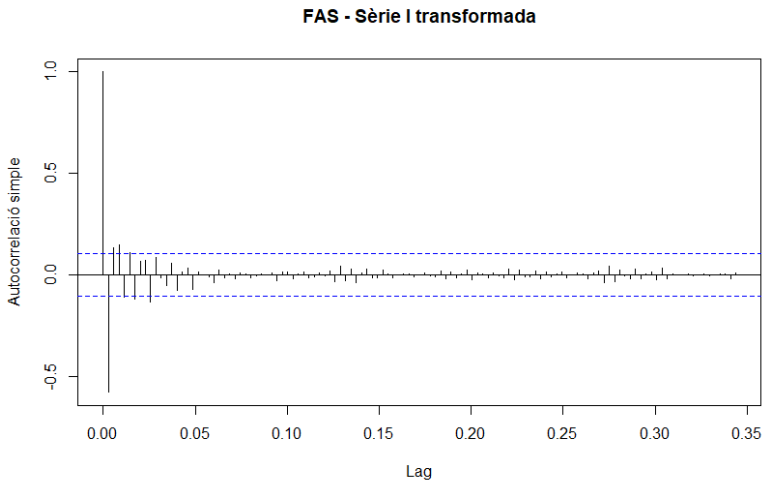


Figura II.4: Gràfic de les autocorrelacions simples de la sèrie I transformada

Donats dos *lags* significatius, previs a una sobtada caiguda cap a zero, la sèrie sembla presentar un component mitjana mòbil d'ordre 2→ $MA(q = 2)$.

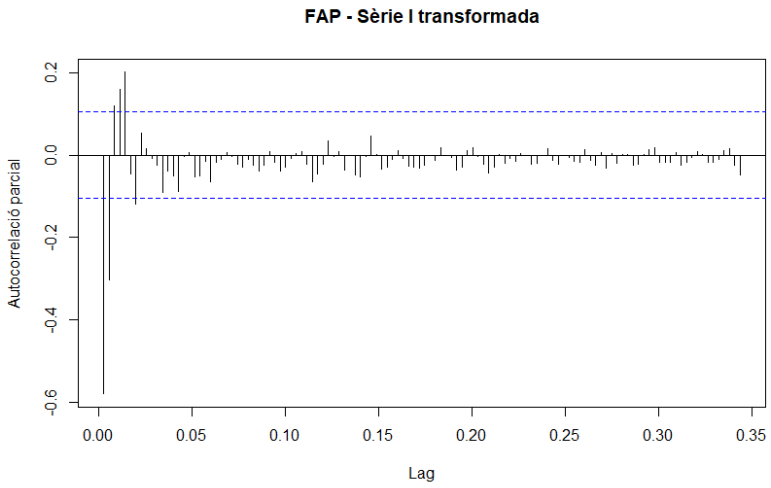


Figura II.5: Gràfic de les autocorrelacions parcials de la sèrie I transformada

Vista la tendència progressiva cap a zero, la sèrie sembla presentar un component autoregressiu d'ordre 0→ $AR(p = 0)$.

En resum, la sèrie I sobre el numero d'infectats per 100.000 habitants a Nova Zelanda durant el 2020 presenta una tendència corregida amb diferències regulars, una component $MA(q = 2)$ i $AR(p = 0)$.

1.3. Anàlisi clàssica

En aquest apartat s'intentarà modelitzar la sèrie estudiada a partir de les metodologies clàssiques. Donada una sèrie amb tendència i sense estacionalitat, es procedirà a l'estimació d'un model tipus ARIMA.

1.3.1. Estimació i validació

Com s'ha trobat en la identificació de la sèrie, s'aplicarà un model ARIMA amb unes diferències regulars i una $MA(q = 2) \rightarrow ARIMA(p = 0, d = 1, q = 2)$:

	Coeficients	Error estàndard	p-valor
ma1	-0,8006	0,0447	3,955626e-68
ma2	0,4956	0,0501	8,522465e-22

Figura II.6: Taula dels coeficients i p-valors del model clàssic per la sèrie I

Per poder utilitzar el model proposat per preveure els pròxims valors de la sèrie I, interessa que els paràmetres estimats siguin significativament diferents de zero i que els residus reemplacin a un procés de soroll blanc (test de Ljung-Box). En cas de que no es complissin aquestes condicions seria un indicador de la necessitat de buscar un model diferent.

Donats uns p-valors significatius en tots els paràmetres del model i l'acceptació d'H_0: $Els\ residus\ segueixen\ un\ soroll\ blanc$, en el test de Ljung-Box, amb un p-valor de 0,6987, es dona per validat el model.

1.3.2. Predicció

La funció de previsió puntual d'un procés estocàstic Y_t donat un origen N i un horitzó l, es representa com el valor esperat de Y_{N+l} condicionat per tota la informació històrica fins al punt d'origen k. Donat que segueix un model ARIMA del tipus $\phi'(B)Y_t = \mu + \theta(B)A_t$, quedaria la següent funció:

$$Y_N(l) \equiv E_N[Y_{N+l}] = \mu + \sum_{i=1}^{p+d} \phi' \; E_N[Y_{N+l-i}] + E_N[A_{N+l}] - \sum_{i=1}^{q} \theta_i E_N[A_{N+l-i}]$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen les prediccions pel període extramostral de la sèrie.

L'error de previsió de la funció de previsió es una variable aleatòria representada com la diferencia entre el valor observat en un moment N+l i la previsió donada per aquell mateix valor calculat amb l'origen N i l'horitzó l. Donat que segueix un model ARIMA del tipus $\phi'(B)Y_t = \mu + \theta(B)A_t$, quedaria la següent funció:

$$E_N \equiv Y_{N+l} - Y_N(l) = \psi^*(B)A_{N+l} - E[\psi^*(B)A_{N+l}]$$

$$on \qquad \psi^*(B)A_{N+l} = \sum_{i=0}^{\infty} \psi_i^* A_{N+l-i}$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen els errors de predicció pel període extramostral de la sèrie.

Finalment, si es tenen en compte aquestes previsions a continuació del període mostral de la sèrie I, s'esperaria l'evolució observada en la Figura II.7.

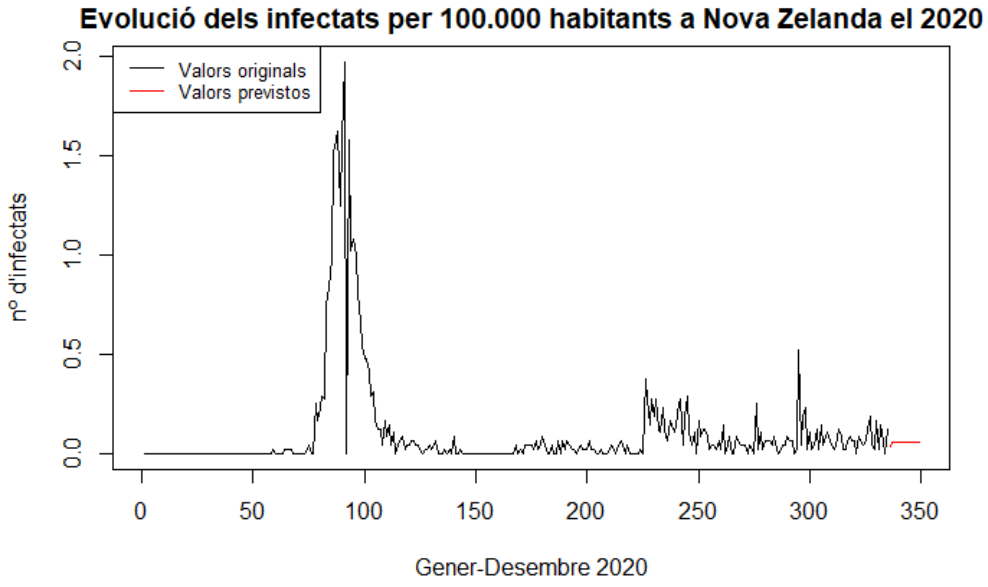


Figura II.7: Gràfic de la previsió d'infectats per 100.000 habitants a Nova Zelanda el desembre

1.4. Anàlisi bayesiana

En aquest apartat s'intentarà modelitzar la sèrie estudiada a partir de les metodologies bayesianes. Donada una sèrie amb tendència i sense estacionalitat, es procedirà a l'estimació d'un model tipus BSTS.

1.4.1. Estimació i validació

Per estimar un model bayesià cal centrar-se en la informació a priori que es coneix sobre la sèrie temporal. Prèviament s'ha identificat que la sèrie no té component regressiva, però si es necessita una manera de modelitzar-ne la tendència.

La manera més comuna de modelitzar la tendència es fent ús de la tendència lineal local. Aquesta component que permet adaptar ràpidament les variacions locals de la sèrie, assumint que la tendència segueix un camí aleatori. Fent-ne ús, el model *bsts* s'escriuria com:

$$y_t = \mu_t + \epsilon_t$$
$$\mu_{t+1} = \mu_t + \delta_t + \eta_{\mu,t}$$
$$\delta_{t+1} = \delta_t + \eta_{\delta,t}$$
$$\eta_{\mu,t} \sim N\left(0, \sigma^2_{\mu,t}\right) \quad , \quad \eta_{\delta,t} \sim N(0, \sigma^2_{\delta,t})$$

on μ_t recull el valor de la tendència en l'observació t i δ_t és l'increment esperat de μ entre t i t+1. Pel que fa $\eta_{\mu,t}$ i $\eta_{\delta,t}$, representen els termes d'error.

Donada la tipologia de les dades, i contràriament a les especificacions del model ARIMA, semblaria que la sèrie hauria de tenir una forta autocorrelació amb els p valors passats directes. En el model bayesià expressaria aquesta component amb la funció "AddAutoAR", però comparant l'error absolut acumulat entre els dos models (Figura II.8) es prefereix la modelització de la tendència prèviament esmentada.

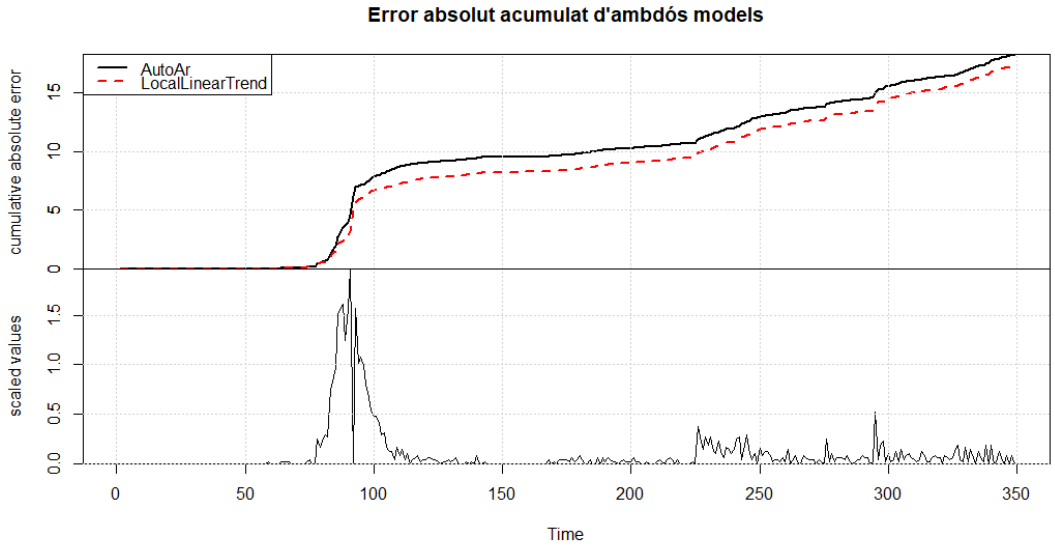


Figura II.8: Gràfic de l'error absolut acumulat dels models bayesians per la sèrie I

Fent ús del component "LocalLinearTrend" es modelitza el model bsts i s'identifica la distribució a posteriori de la sèrie (Figura II.9).

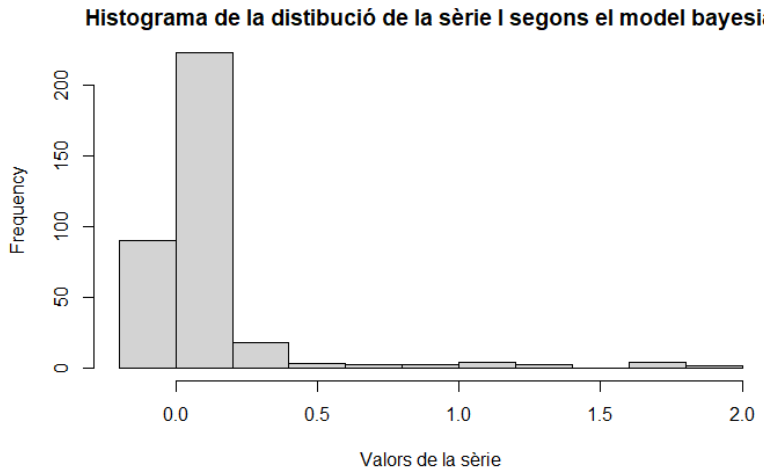


Figura II.9: Histograma de la distribució a posteriori del model bayesià per la sèrie I

Segons aquest model, la sèrie té una mediana sobre els 0,0287 infectats per cada 100.000 habitants, amb una desviació típica de 0,2686 casos amunt i avall.

Per poder utilitzar el model proposat per preveure els pròxims valors de la sèrie I, interessa els residus segueixin una distribució normal i siguin estacionaris. En cas de que no es complissin aquestes premisses, seria un indicador de la necessitat de buscar un model diferent.

Donat el següent gràfic sobre la distribució dels residus del model proposat (Figura II.10), s'observa com, a excepció del extrems, els valors tendeixen als quantils de la distribució normal estàndard. Es suposa normalitat als residus.

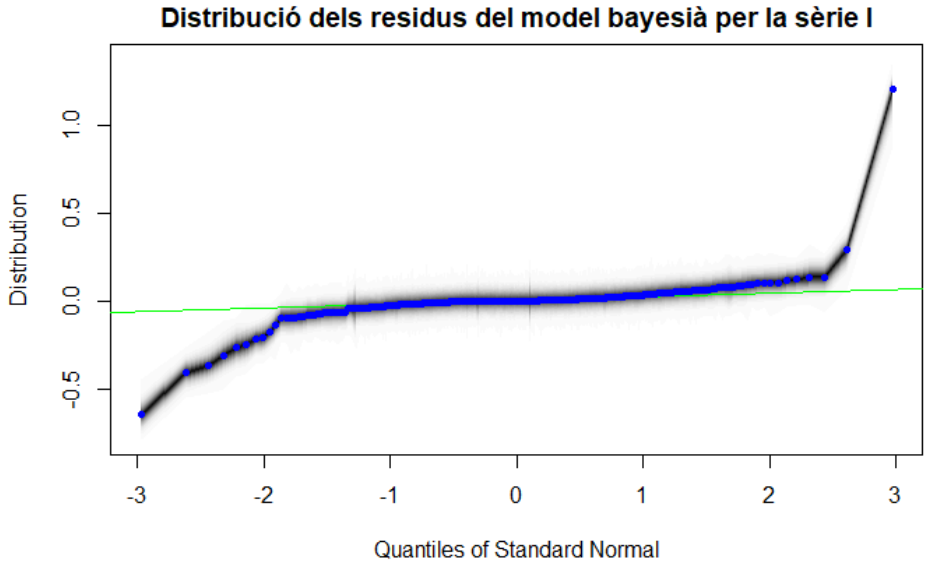


Figura II.10: Gràfic de la distribució dels residus del model bayesià per la sèrie I

Per comprovar l'estacionarietat dels residus es pretén que aquests tendeixin a no tenir autocorrelació (autocorrelació zero) (Figura II.11).

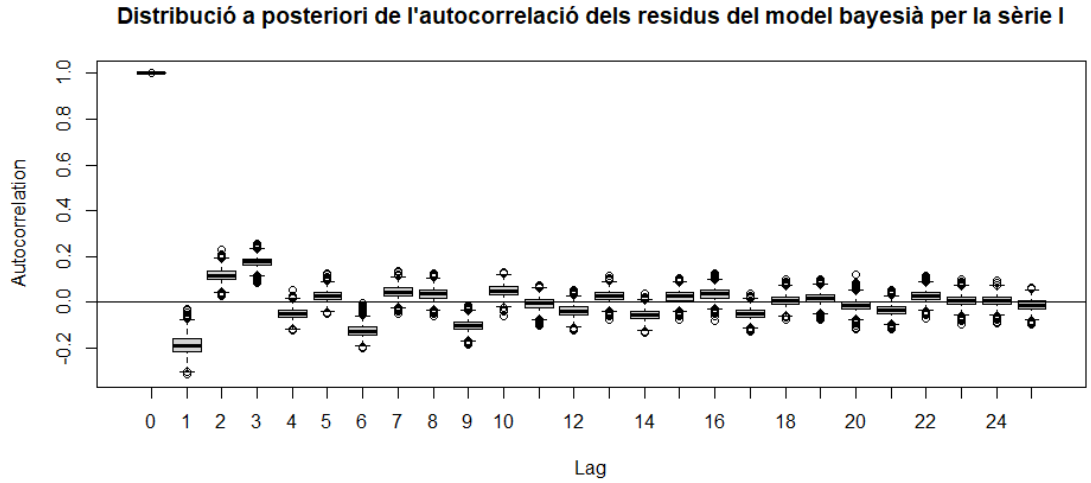


Figura II.11: Gràfic de la distribució a posteriori de l'autocorrelació dels residus del model bayesià per la sèrie I

Observant el gràfic, a partir del *lag* 10 els residus del model bayesià tendeixen a zero. Havent-se complert les dues validacions es pot procedir a realitzar previsions per la sèrie I.

1.4.2. Predicció

La funció de previsió puntual d'un procés estocàstic Y_t donat un origen N i un horitzó l, es representa com el valor esperat de Y_{N+l} condicionat per tota la informació històrica fins al punt d'origen k:

$$Y_N(l) \equiv E_N[Y_{N+l}] \ \ per \ \ l = 1,2, \dots$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen les prediccions pel període extramostral de la sèrie.

L'error de previsió de la funció de previsió es una variable aleatòria representada com la diferencia entre el valor observat en un moment N+l i la previsió donada per aquell mateix valor calculat amb l'origen N i l'horitzó l:

$$E_N \equiv Y_{N+l} - Y_N(l) \ \ per \ \ l = 1,2, \dots$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen els errors de predicció pel període extramostral de la sèrie.

Finalment, si es tenen en compte aquestes previsions a continuació del període mostral de la sèrie I, s'esperaria l'evolució observada en la Figura II.12.

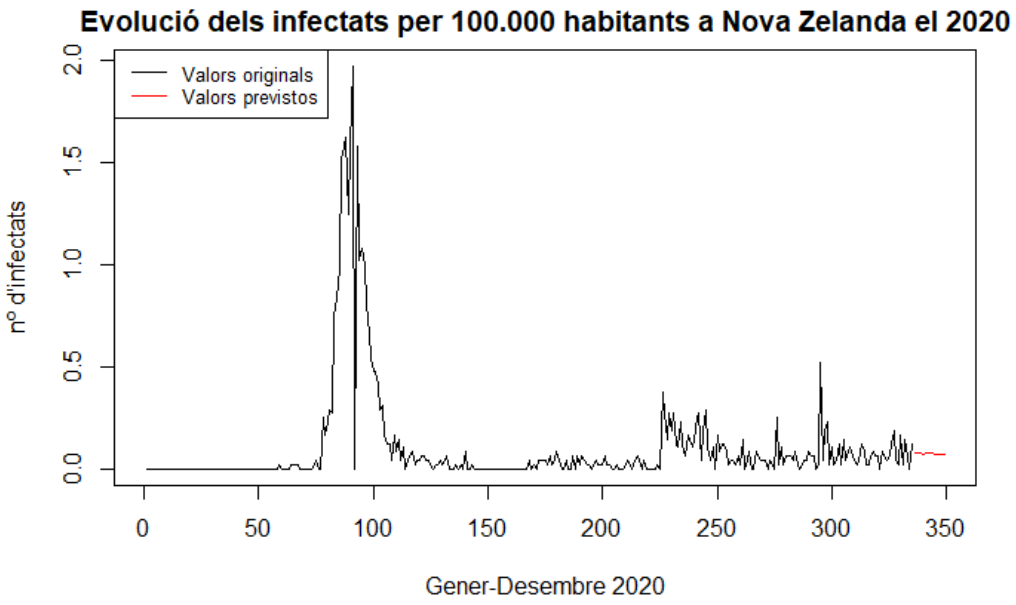


Figura II.12: Gràfic de la previsió d'infectats per 100.000 habitants a Nova Zelanda el desembre

1.5. Comparació dels diferents models obtinguts

Un cop escollits els diferents models, segons la metodologia clàssica i la bayesiana, per ajustar la sèrie I, es procedeix a estudiar quin dels dos enfocaments permet ajustar unes millors previsions.

1.5.1. Període mostral

Donada l'anterior sèrie, s'ha pres com a període mostral els primers 11 mesos, resultant en un total de 335 observacions històriques. A partir d'aquestes, i fent ús dels dos models, s'han obtingut els següents ajustos (Figura II.13 i Figura II.14):

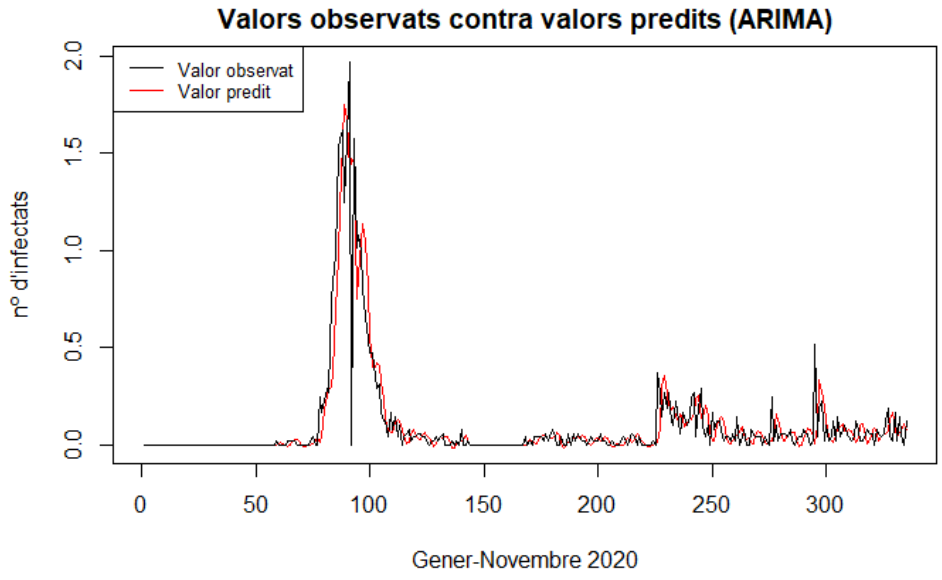


Figura II.13: Gràfic dels valors predits pel model ARIMA en el període mostral

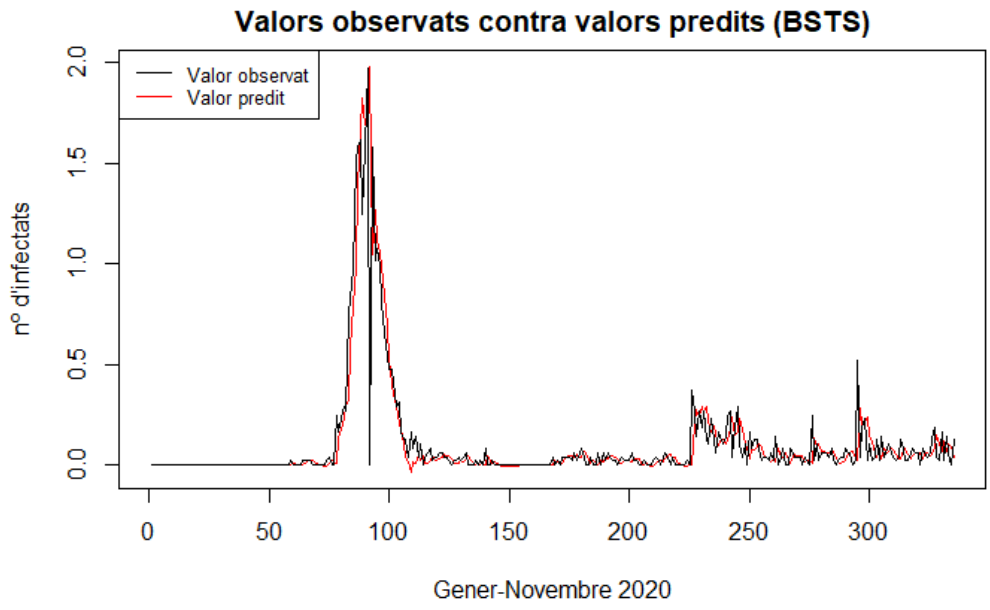


Figura II.14: Gràfic dels valors predits pel model BSTS en el període mostral

Gràficament sembla que ambdós models escollits ajusten considerablement bé les dades. En el primer pic (al voltant de l'observació 80), el model ARIMA subestima els valors de la sèrie durant la tendència creixent d'aquesta, mentre que sobreestima la caiguda del nombre d'infectats. Per altra banda, el model bayesià segueix el mateix patró però s'ajusta millor als valors observats, encara que després de la caiguda subestima considerablement les dades observades. Malgrat aquestes petites diferències no es pot considerar que l'ajust dels dos models distin molt un de l'altre, i es procedirà a calcular-ne l'EPAM de cada un, per quantificar-ne l'ajust.

$$EPAM = \frac{\frac{100 * \sum_{i=1}^{n} |y_i - \hat{y}_i|}{y_i}}{n}$$

Aplicant la formula s'obté un EPAM del 79,19% segons el model ARIMA i un EPAM del 76,77% segons el model BSTS. Cal remarcar que encara que són uns percentatges d'error considerables, són raonables tenint en compte el tipus de dades que componen la sèrie, i l'elevat grau d'incertesa que aquestes presenten.

Es reafirma que els dos models ajusten de manera similar el període mostral, encara que sembla que el model bayesià permet fer unes estimacions lleugerament més properes als valors registrats.

1.5.2. Període extramostral

Donada l'anterior sèrie, s'ha pres com a període extramostral l'últim mes, resultant en un total de 14 previsions per cada model (Figura II.15).

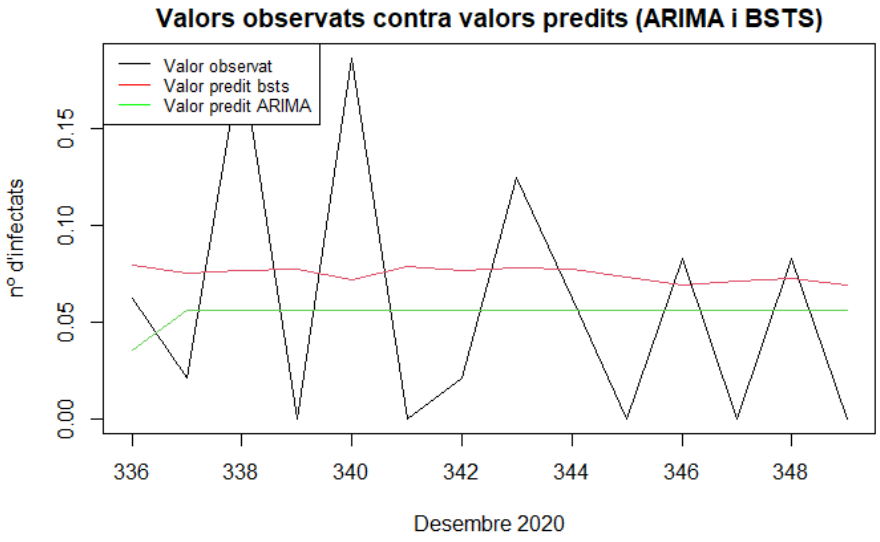


Figura II.15: Gràfic dels valors predits pel model ARIMA i BSTS en el període extramostral

Observant el gràfic, el model bayesià presenta uns valors previsions superiors als del model ARIMA, encara que és l'únic dels dos que preveu la tendència decreixent que te la sèrie original. Pel que fa a les previsions del model clàssic, es mantenen constants a partir de la segona observació, mantenint-se entre les oscil·lacions dels valors reals.

Aplicant un altre cop la formula del EPAM per les noves dades, s'obtenen uns errors del 72,75% i del 85,64% pels models ARIMA i BSTS respectivament.

Encara que sembla que les previsions del model ARIMA s'ajustin millor i presentin un EPAM inferior, a llarg termini no aconsegueixen modelitzar correctament la tendència de la sèrie. Això també s'observa en la segona predicció, donat que tant la sèrie original com el model bayesià han presentat una disminució dels casos d'infectats, mentre que el model ARIMA en preveu un augment.

1.5.3. Resum

Amb els resultats anteriors no sembla haver diferències rellevants entre l'ajust dels dos models durant el període mostral. Donats uns EPAMs similars, costaria decantar-se per una de les dues metodologies. On si s'han donat diferències de més interès és en el període extramostral. Aquí el model ARIMA semblava imposar-se amb unes previsions més ajustades, però és el model bayesià el que realment ha modelitzat de manera més adequada les tendències de la sèrie, predint, de forma més ajustada, el seu comportament.

2. Sèrie II: Nombre d'infectats per la covid-19 a Mèxic durant l'any 2020

La segona sèrie de l'estudi consisteix en el nombre d'infectats per la covid-19 a Mèxic durant l'any 2020 (Figura II.16). Aquesta recull diàriament el nombre d'infectats pel virus, reportats pel govern i els centres mèdics, iniciant el dia 1 de gener del 2020 i finalitzant el 14 de desembre del mateix any. Per facilitar-ne la validació i l'ajust dels

models proposats a continuació, s'ha decidit prendre els 11 primers mesos com a període mostral, deixant les últimes dues setmanes com a període extramostral.

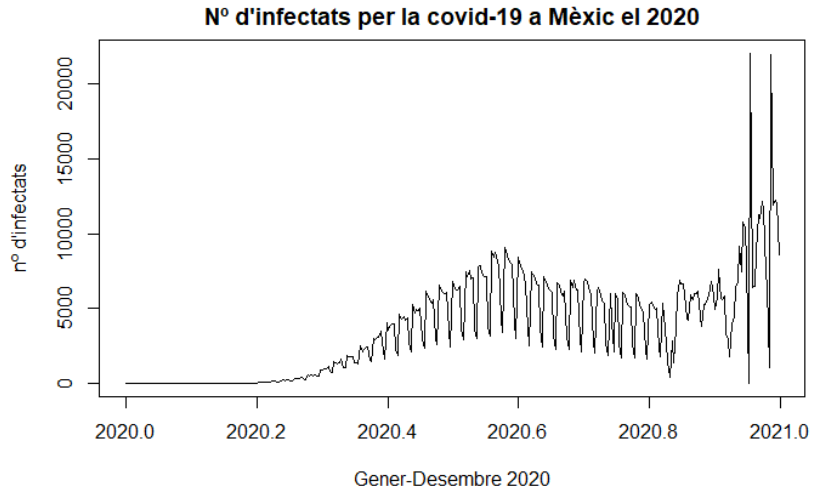


Figura II.16: Gràfic de la sèrie II:Nº d'infectats a Mèxic

Abans de procedir a la modelització de la sèrie, i com s'ha esmentat prèviament, s'utilitza el cens de la població de Mèxic durant l'any de les dades per estandarditzar les unitats i passar del nombre d'infectats al nombre d'infectats per 100.000 habitants seguin la formula que es veu a continuació:

$$Y = S * \frac{100.000\ habitants}{128.932.753\ habitants\ a\ Mèxic}$$

on S és la sèrie original i Y la resultant de la transformació.

Aquest procés es necessari per tenir les tres sèries temporals amb les mateixes unitats de mesura, facilitant-ne així les comparacions i els comportaments d'aquestes.

Finalment, la sèrie a analitzar és la que es representa en la Figura II.17:

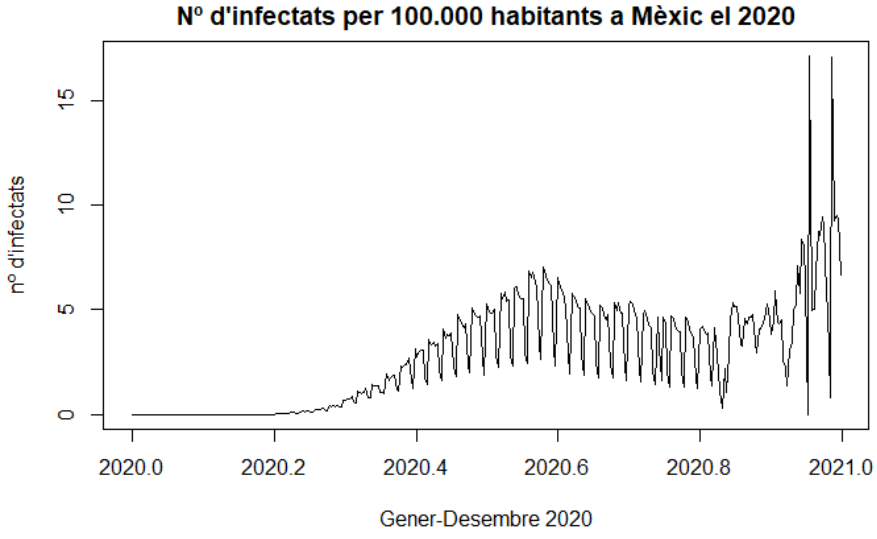


Figura II.17: Gràfic de la sèrie II ajustada

2.2. Identificació del model

Donada la sèrie temporal II, es pretén identificar els diversos components que la formen: Tendència, estacionalitat, ... També cal recordar que per poder modelitzar-la, aquesta ha de complir el supòsit d'estacionarietat mencionat anteriorment. Donat que les sèries utilitzades en l'anàlisi són no estacionaries caldrà aplicar diverses transformacions, segons els components que les conformin, per poder assolir aquesta propietat.

Primerament s'estudia si la sèrie presenta algun tipus de tendència. Per fer això és crea un model lineal amb les dades com a variable dependent i un vector seqüencial del 1 fins al nombre total de dades recollides per la sèrie com a variable explicativa. Utilitzant aquest model es busca quantificar l'efecte que té aquest nou vector, donat que si fos estadísticament significatiu implicaria que ajuda a explicar els valors de la variable dependent, i com a conseqüència, que existeix una tendència.

Aplicant aquests conceptes s'obté un p-valor menor a 2e-16, inferior al 5% de significació, obligant a acceptar la hipòtesi de que hi ha tendència i és necessari corregir-la. Com a conseqüent, s'apliquen les diferències regulars per intentar estabilitzar la mitjana i se'n grafiquen els resultats (Figura II.18).

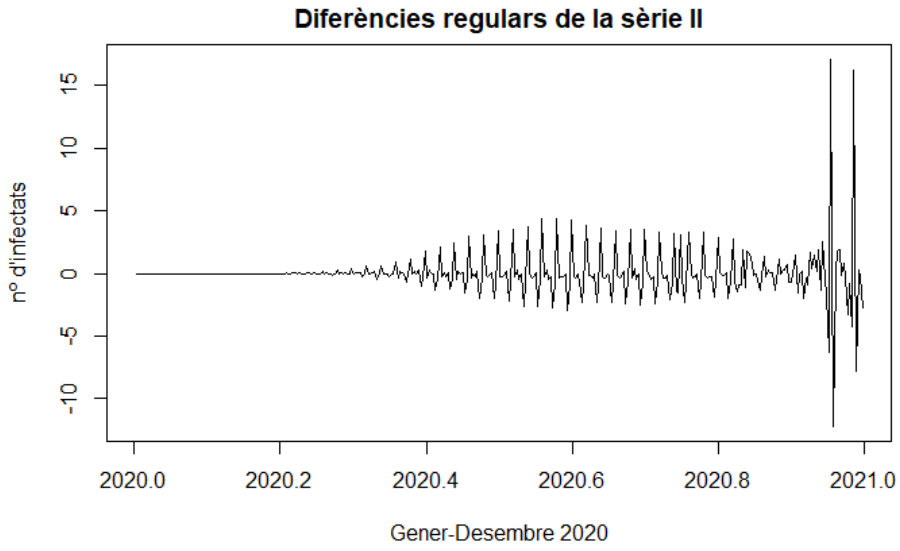


Figura II.18: Gràfic de les diferències regulars de la sèrie II

S'observa com la mitjana s'ha estabilitzat al voltant del zero i sembla ser constant al llarg de tota la sèrie. Si es repeteix la comprovació numèrica per la nova sèrie transformada, s'observa com el coeficient de la pendent a deixat de ser significativament diferent de zero (p-valor=0,957).

Per altra banda, sembla que la variància no es manté constant, a la vegada que la sèrie presenta una successió de patrons al llarg de tota la seva trajectòria. Detectats aquests problemes caldrà actuar en conseqüent, aplicant transformacions a la sèrie i diferències estacionals.

Primerament es busca, una transformació de la sèrie que estabilitzi la variància, transformant-la en constant durant tot el període estudiat. S'ha trobat com la nova sèrie transformada que satisfà aquesta condició seria l'arrel sisena de la sèrie original (Figura II.19).

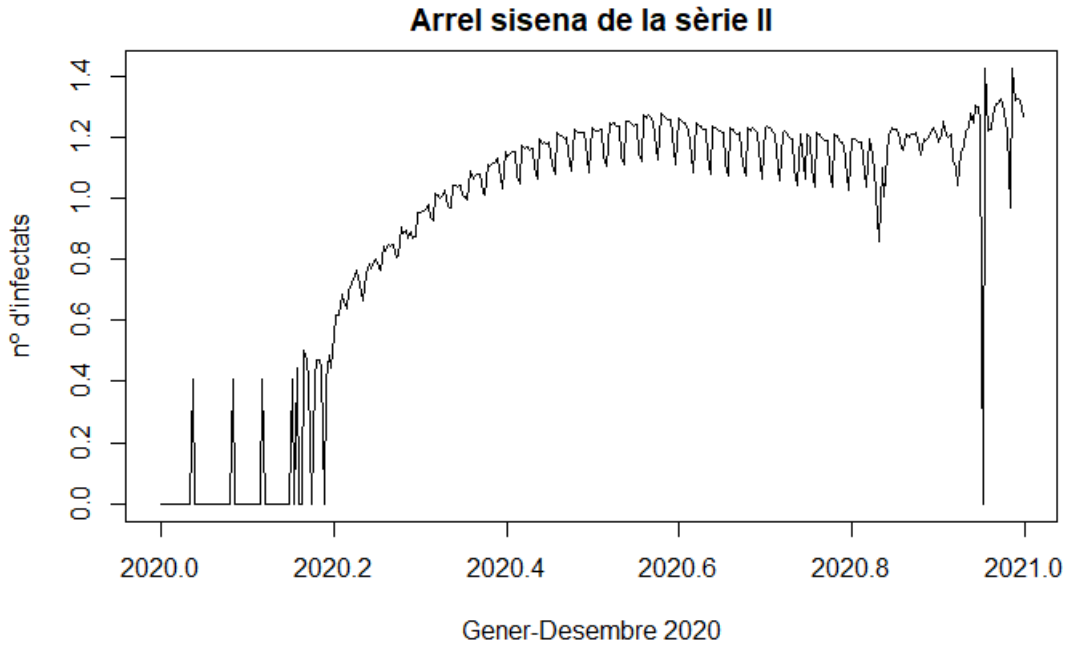


Figura II.19: Gràfic de la nova sèrie II transformada

Donada ara si una variància que pot ser considerada constant, es tornen a aplicar les diferències regulars (Figura II.20).

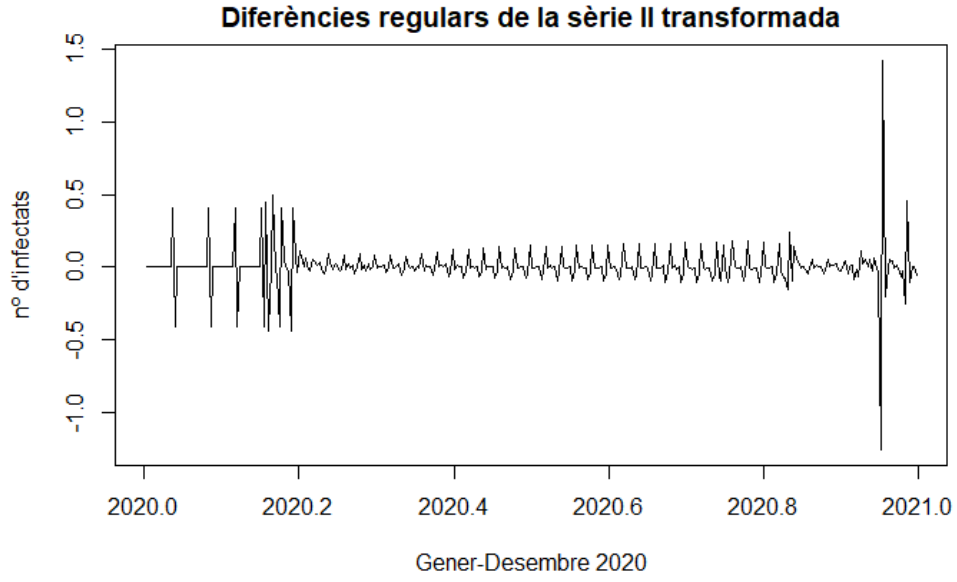


Figura II.20: Gràfic de les diferències regulars de la sèrie II transformada

S'observa com tant la mitjana com la variància es mantenen constants, però encara presenta un patró de nous casos d'infectats, amb oscil·lacions de set dies. Cal aplicar unes diferències estacionals per mirar d'eliminar-ho (Figura II.21).

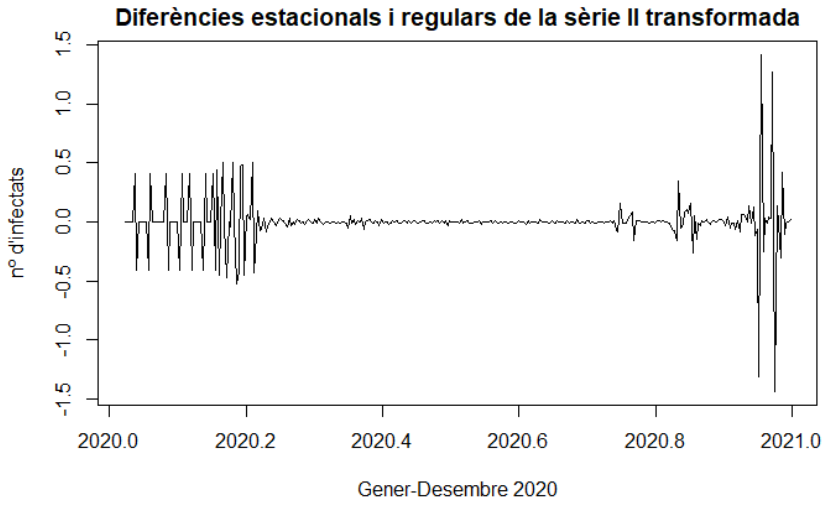


Figura II.21: Gràfic de les diferències regulars i estacionals de la sèrie II transformada

Finalment ens trobem amb una sèrie sense estacionalitat. Per acabar amb el procés d'identificació, es comproven els gràfics de les autocorrelacions simples (Figura II.22 i Figura II.24) i parcials (Figura II.23 i Figura II.25) per definir els components mitjana mòbil i/o autoregressius:

- Part regular:

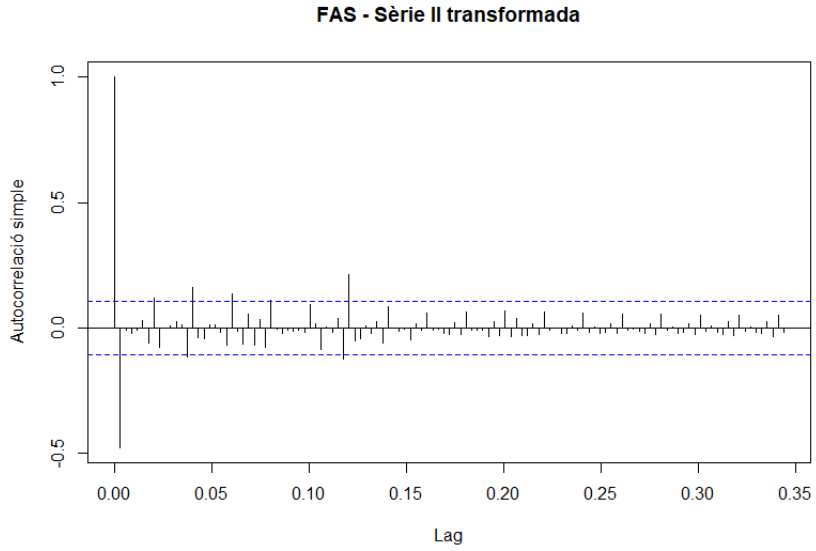


Figura II.22: Gràfic de les autocorrelacions simples de la sèrie II transformada sense diferències

Donats dos *lags* significatius, previs a una sobtada caiguda cap a zero, la sèrie sembla presentar una component mitjana mòbil d'ordre 2→ $MA(q = 2)$.

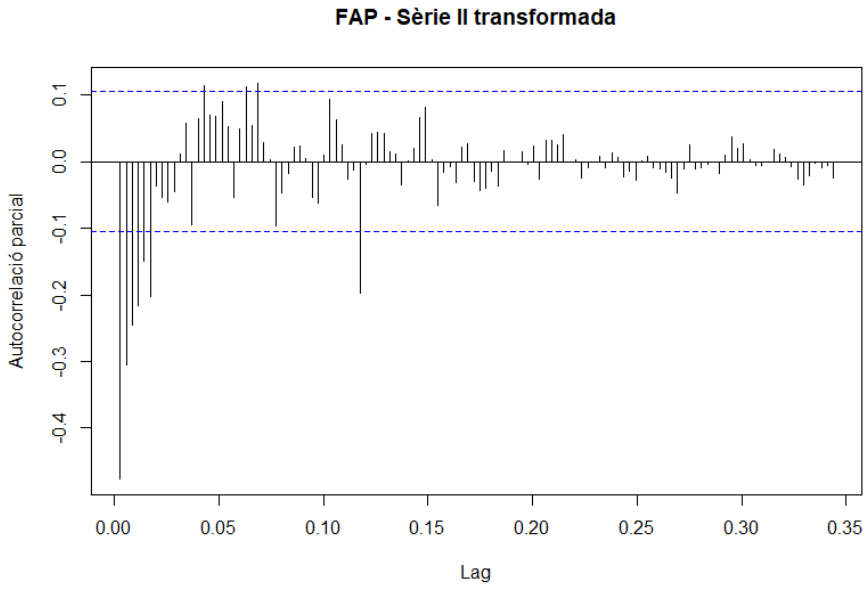


Figura II.23: Gràfic de les autocorrelacions parcials de la sèrie II transformada sense diferències

Vista la tendència progressiva cap a zero, la sèrie sembla presentar una component autoregressiu d'ordre 0→ $AR(p = 0)$.

- Part estacional:

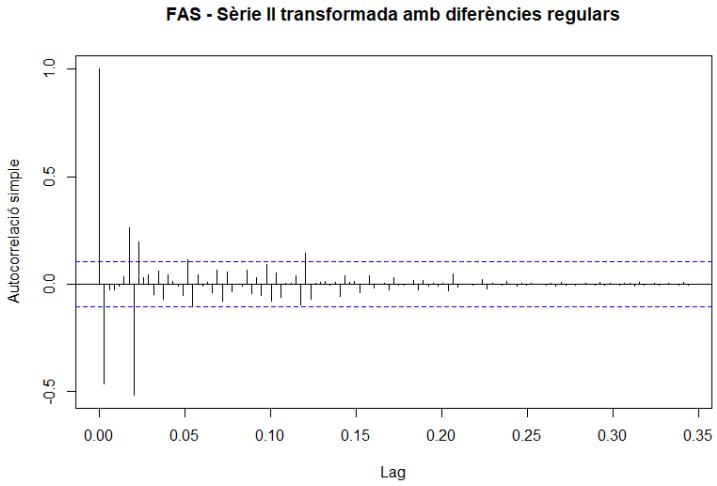


Figura II.24: Gràfic de les autocorrelacions simples de la sèrie II transformada amb diferències regulars

Donats dos *lags* significatius, previs a una sobtada caiguda cap a zero, la sèrie sembla presentar una component mitjana mòbil d'ordre 2→ $MA(q = 2)$.

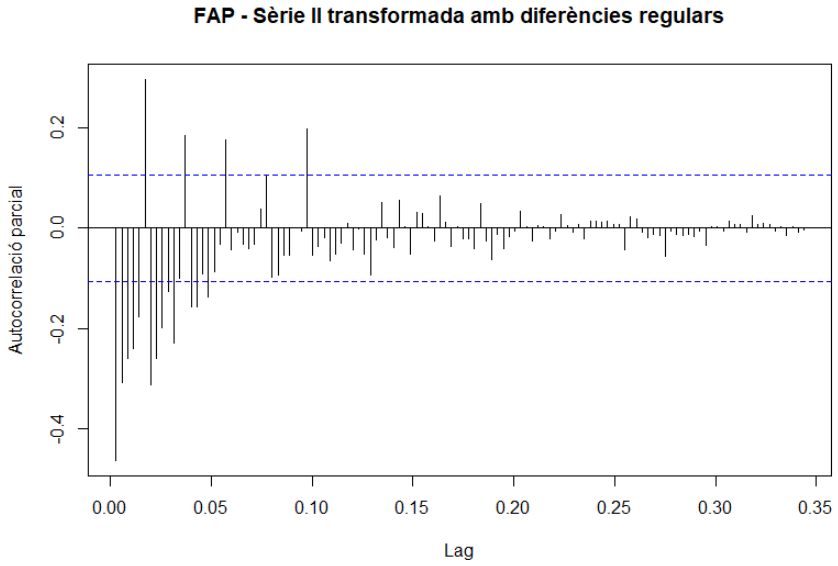


Figura II.25: Gràfic de les autocorrelacions parcials de la sèrie II transformada amb diferències regulars

Vista la tendència progressiva cap a zero, la sèrie sembla presentar una component autoregressiu d'ordre 0→ $AR(p=0)$.

En resum, la sèrie II sobre el numero d'infectats per 100.000 habitants a Mèxic durant el 2020 presenta una combinació de una component $MA(q=2)$ i $AR(p=0)$ en el període regular i una combinació de una component $MA(q=2)$ i $AR(p=0)$ en el període estacionari. Additivament, també s'ha recorregut a una arrel sexta de la sèrie per estabilitzar-ne la variància.

2.3. Anàlisi clàssica

En aquest apartat s'intentarà modelitzar la sèrie estudiada a partir de les metodologies clàssiques. Donada una sèrie amb tendència i amb estacionalitat, es procedirà a l'estimació d'un model tipus SARIMA, el qual és una variació del model ARIMA, afegint una part estacional.

2.3.1. Estimació i validació

Com s'ha trobat en la identificació de la sèrie, s'aplicarà un model SARIMA. Tant la part regular com l'estacional tindran una component $MA(q=2)$ i tindra diferències regulars i estacionals amb un període de 7 dies $\rightarrow SARIMA(p=0,d=1,q=2)(P=0,D=1,Q=2)^{s=7}$:

	Coeficients	Error estàndard	p-valor
ma1	-0,8298	0,0575	3,015526e-47
ma2	0,0364	0,0554	5,102576e-01
sma1	-0,9524	0,0602	2,369380e-56
sma2	0,1095	0,0660	9,745048e-02

Figura II.26: Taula dels coeficients i p-valors del model clàssic per la sèrie II

Per poder utilitzar el model proposat per preveure els pròxims valors de la sèrie II, interessa els paràmetres d'aquest siguin significatius i que els residus reemplacin a un procés de soroll blanc (test de Ljung-Box). En cas de que no es complissin seria un indicador de la necessitat de buscar un model diferent.

Donat un p-valor no significatiu en el segon paràmetre del model, en la part regular, es creu que aquest pot estar sobreestimat i s'ajusta un $SARIMA(p=0,d=1,q=1)(P=0,D=1,Q=2)^{s=7}$:

	Coeficients	Error estàndard	p-valor
ma1	-0,7975	0,0288	1,930686e-168
sma1	-0,9447	0,0589	6,475920e-58
sma2	0,1022	0,0649	1,149377e-01

Figura II.27: Taula dels coeficients i p-valors del model clàssic final per la sèrie II

Novament es mira la significació dels coeficients i que els residus es comportin com un procés de soroll blanc.

Donats un p-valor significatiu en l'únic paràmetre del model i l'acceptació d'H_0: Els residus segueixen un soroll blanc , en el test de Ljung-Box, amb un p-valor de 0,5571, es dona per validat el nou model.

2.3.2. Predicció

La funció de previsió puntual d'un procés estocàstic Y_t donat un origen N i un horitzó l, es representa com el valor esperat de Y_{N+l} condicionat per tota la informació històrica fins al punt d'origen k. Donat que segueix un model ARIMA del tipus $\phi'(B)Y_t = \mu + \theta(B)A_t$, quedaria la següent funció:

$$Y_N(l) \equiv E_N[Y_{N+l}] = \mu + \sum_{i=1}^{p+d} \phi' \; E_N[Y_{N+l-i}] + E_N[A_{N+l}] - \sum_{i=1}^{q} \theta_i E_N[A_{N+l-i}]$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen les prediccions pel període extramostral de la sèrie.

L'error de previsió de la funció de previsió es una variable aleatòria representada com la diferencia entre el valor observat en un moment N+l i la previsió donada per aquell mateix valor calculat amb l'origen N i l'horitzó l. Donat que segueix un model ARIMA del tipus $\phi'(B)Y_t = \mu + \theta(B)A_t$, quedaria la següent funció:

$$E_N \equiv Y_{N+l} - Y_N(l) = \psi^*(B)A_{N+l} - E[\psi^*(B)A_{N+l}]$$

$$on \qquad \psi^*(B)A_{N+l} = \sum_{i=0}^{\infty} \psi_i^* A_{N+l-i}$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen els errors de predicció pel període extramostral de la sèrie.

Finalment, si es tenen en compte aquestes previsions a continuació del període mostral de la sèrie II, s'esperaria l'evolució observada en la Figura II.28.

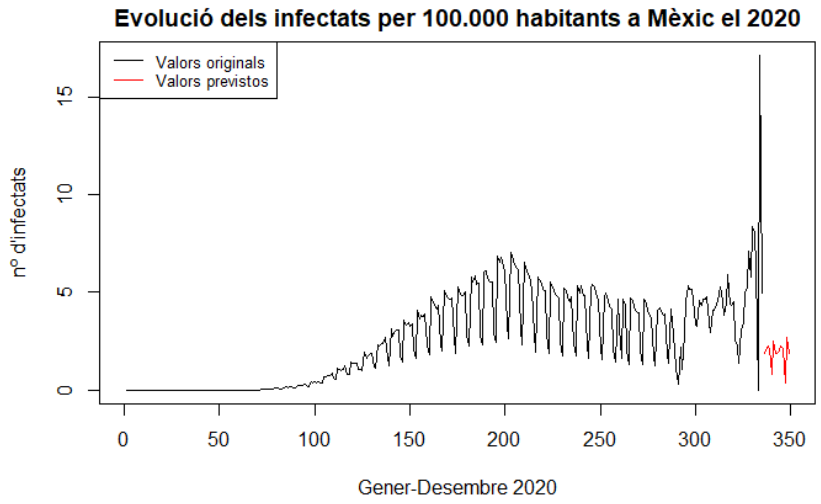


Figura II.28: Gràfic de la previsió d'infectats per 100.000 habitants a Mèxic el desembre

2.4. Anàlisi bayesiana

En aquest apartat s'intentarà modelitzar la sèrie estudiada a partir de les metodologies bayesianes. Donada una sèrie amb tendència i sense estacionalitat, es procedirà a l'estimació d'un model tipus BSTS.

2.4.1. Estimació i validació

Per estimar un model bayesià cal centrar-se en la informació a priori que es coneix sobre la sèrie temporal. Prèviament s'ha identificat que la sèrie no té component regressiva, però si es necessita una manera de modelitzar-ne la tendència. Additivament, en aquest cas també es té una component estacional, la qual s'haurà d'introduir en el model.

La manera més comuna de modelitzar la tendència es fent ús de la tendència lineal local. Aquesta component que permet adaptar ràpidament les variacions locals de la sèrie, assumint que la tendència segueix un camí aleatori. Fent-ne ús, el model *bsts* s'escriuria com:

$$y_t = \mu_t + \epsilon_t$$
$$\mu_{t+1} = \mu_t + \delta_t + \eta_{\mu,t}$$
$$\delta_{t+1} = \delta_t + \eta_{\delta,t}$$
$$\eta_{\mu,t} \sim N(0, \sigma^2_{\mu,t}) \quad , \quad \eta_{\delta,t} \sim N(0, \sigma^2_{\delta,t})$$

on μ_t recull el valor de la tendència en l'observació t i δ_t és l'increment esperat de μ entre t i t+1. Pel que fa $\eta_{\mu,t}$ i $\eta_{\delta,t}$, representen els termes d'error.

Donada la tipologia de les dades, i contràriament a les especificacions del model ARIMA, semblaria que la sèrie hauria de tenir una forta autocorrelació amb els p valors passats directes. En el model bayesià expressaria aquesta component amb la funció "AddAutoAR", però comparant l'error absolut acumulat entre els dos models (Figura II.29) es prefereix la modelització de la tendència prèviament esmentada.

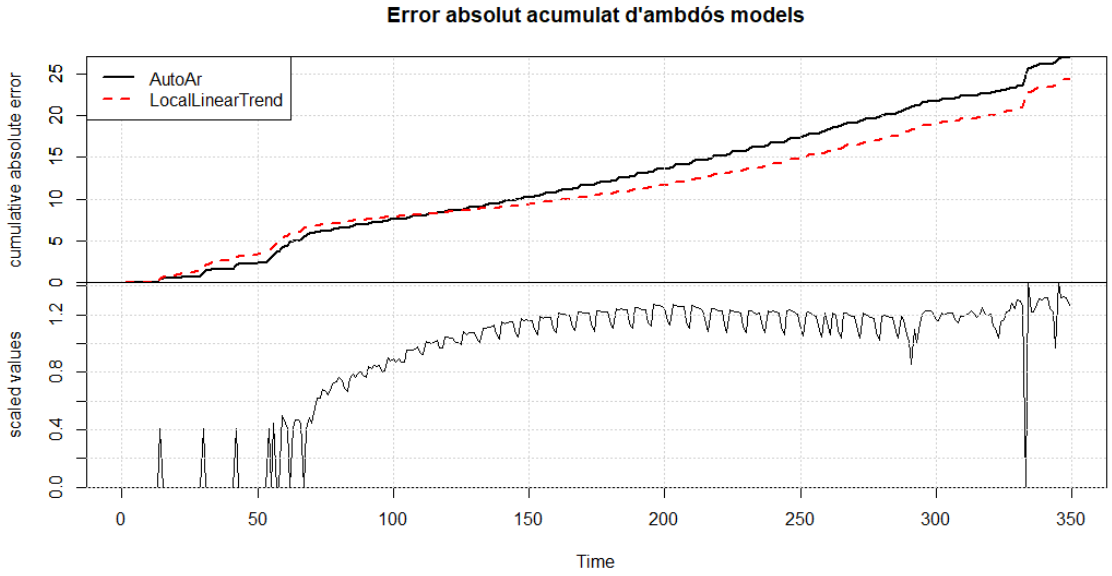


Figura II.29: Gràfic de l'error absolut acumulat dels models bayesians per la sèrie I

Es prefereix fer ús del component "LocalLinearTrend", donat que presenta un menor error absolut acumulat.

Pel que refereix a l'estacionalitat, el component més freqüentment usat per captar aquestes periodicitats en el model es la "Regression with Seasonal Dummy Variables":

$$y_t = \gamma_t + \epsilon_t$$

$$\gamma_{t+d} = -\sum_{i=0}^{s-2} \gamma_{t-i*d} + \eta_{\gamma,t}$$

on s és el nº de períodes o repeticions i d la seva durada. El model es pot considerar com una regressió on s variables fictícies representen els s períodes i γ_t en recull la seva contribució conjunta a la resposta observada.

Fent ús d'aquests components es pot modelitzar el model *bsts* i s'identifica la distribució a posteriori de la sèrie (Figura II.30).

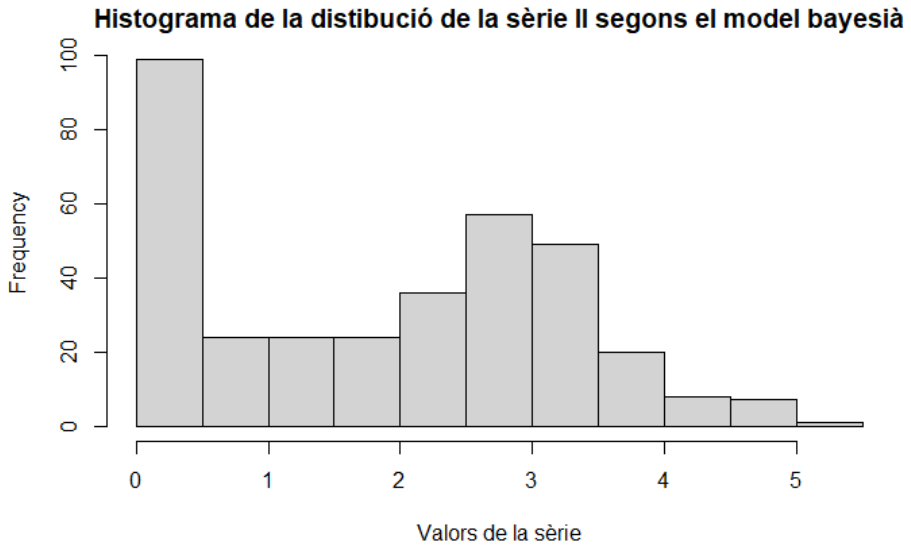


Figura II.30: Histograma de la distribució a posteriori del model bayesià per la sèrie II

Segons aquest model, la sèrie te una mediana sobre els 2,06 infectats per cada 100.000 habitants, amb una desviació típica de 1,40 casos amunt i avall.

Per poder utilitzar el model proposat per preveure els pròxims valors de la sèrie II, interessa els residus segueixin una distribució normal i siguin estacionaris. En cas de que no es complissin aquestes premisses, seria un indicador de la necessitat de buscar un model diferent.

Donat el següent gràfic sobre la distribució dels residus del model proposat (Figura II.31), s'observa com, a excepció del extrems, els valors tendeixen als quantils de la distribució normal estàndard. Es suposa normalitat als residus.

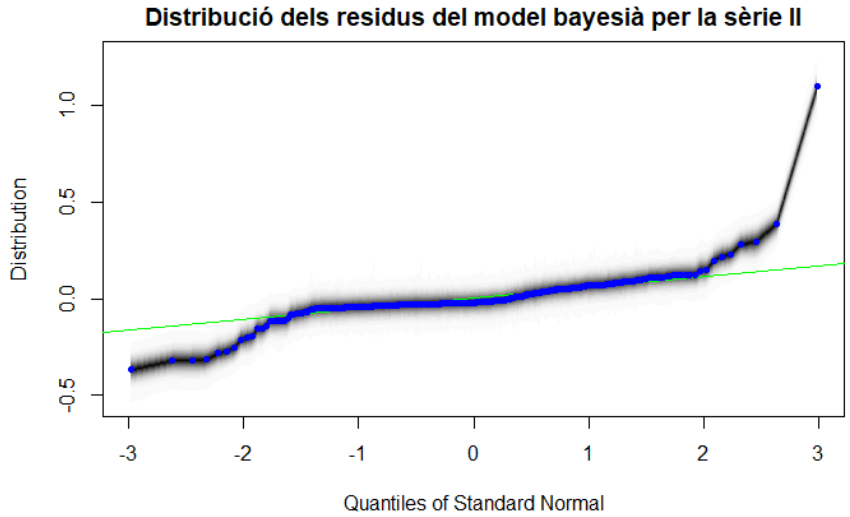


Figura II.31: Gràfic de la distribució dels residus del model bayesià per la sèrie II

Per comprovar l'estacionarietat dels residus es pretén que aquests tendeixin a no tenir autocorrelació (autocorrelació zero) (Figura II.32).

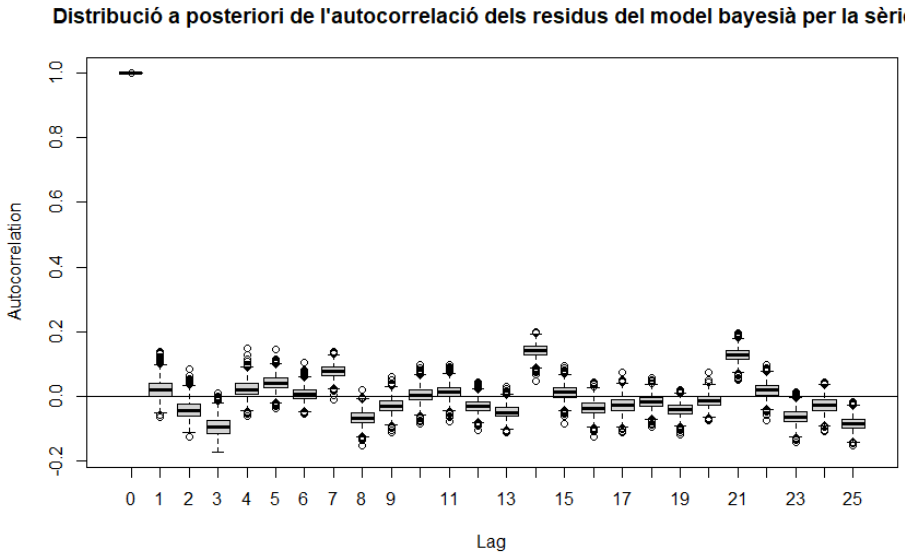


Figura II.32: Gràfic de la distribució a posteriori de l'autocorrelació dels residus del model bayesià per la sèrie II

Observant el gràfic, sembla com l'autocorrelació dels residus és contant al voltant del zero. A excepció d'alguns *lags* on aquesta augmenta lleugerament és pren com a que no hi ha autocorrelació en els residus del model. Havent-se complert les dues validacions es pot procedir a realitzar previsions per la sèrie II.

2.4.2. Predicció

La funció de previsió puntual d'un procés estocàstic Y_t donat un origen N i un horitzó l, es representa com el valor esperat de Y_{N+l} condicionat per tota la informació històrica fins al punt d'origen k:

$$Y_N(l) \equiv E_N[Y_{N+l}] \ \ per \ \ l = 1,2, \dots$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen les prediccions pel període extramostral de la sèrie.

L'error de previsió de la funció de previsió es una variable aleatòria representada com la diferencia entre el valor observat en un moment N+l i la previsió donada per aquell mateix valor calculat amb l'origen N i l'horitzó l:

$$E_N \equiv Y_{N+l} - Y_N(l) \ \ per \ \ l = 1,2, \dots$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen els errors de predicció pel període extramostral de la sèrie.

Finalment, si es tenen en compte aquestes previsions a continuació del període mostral de la sèrie II, s'esperaria l'evolució observada en la Figura II.33.

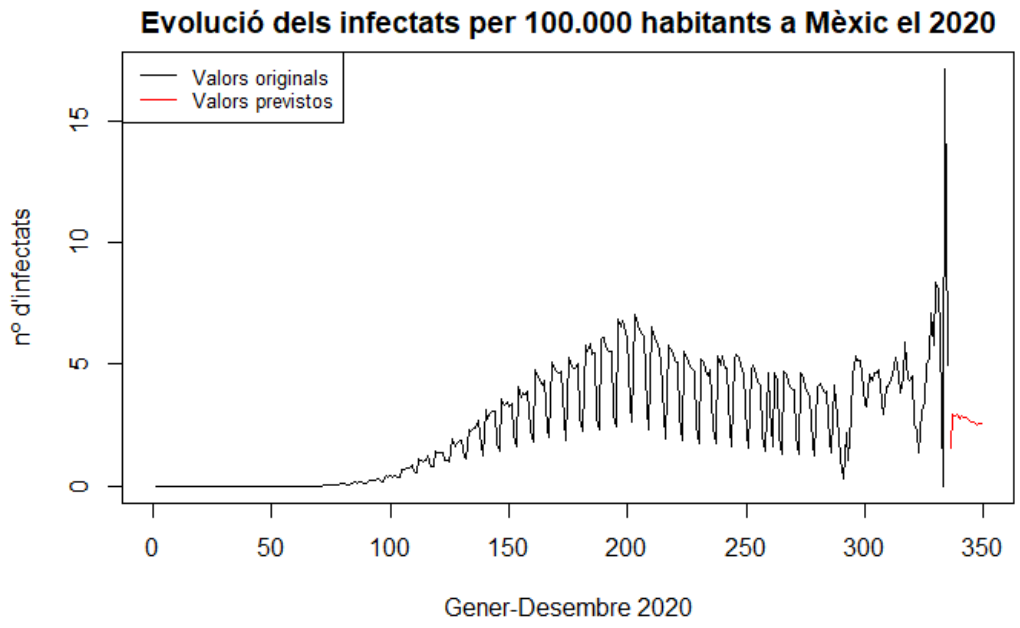


Figura II.33: Gràfic de la previsió d'infectats per 100.000 habitants a Mèxic el desembre

2.5. Comparació dels diferents models obtinguts

Un cop escollits els diferents models, segons la metodologia clàssica i la bayesiana, per ajustar la sèrie II, es procedeix a estudiar quin dels dos enfocaments permet ajustar unes millors previsions.

2.5.1. Període mostral

Donada l'anterior sèrie, s'ha pres com a període mostral els primers 11 mesos, resultant en un total de 335 observacions històriques. A partir d'aquestes, i fent ús dels dos models, s'han obtingut els següents ajustos (Figura II.34 i Figura II.35):

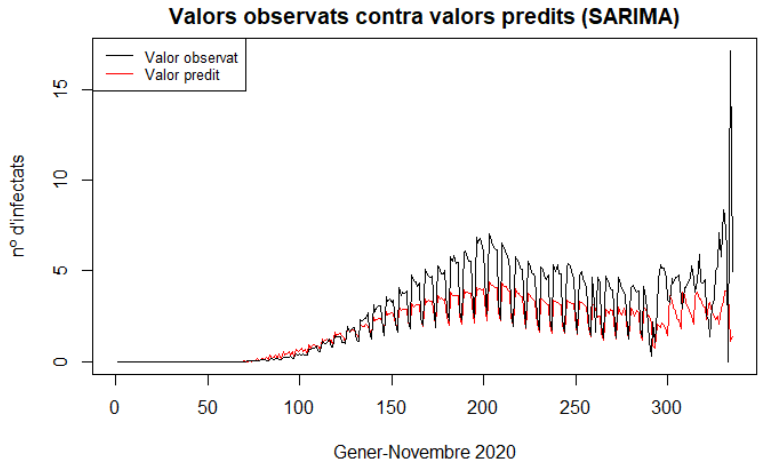


Figura II.34: Gràfic dels valors predits pel model SARIMA en el període mostral

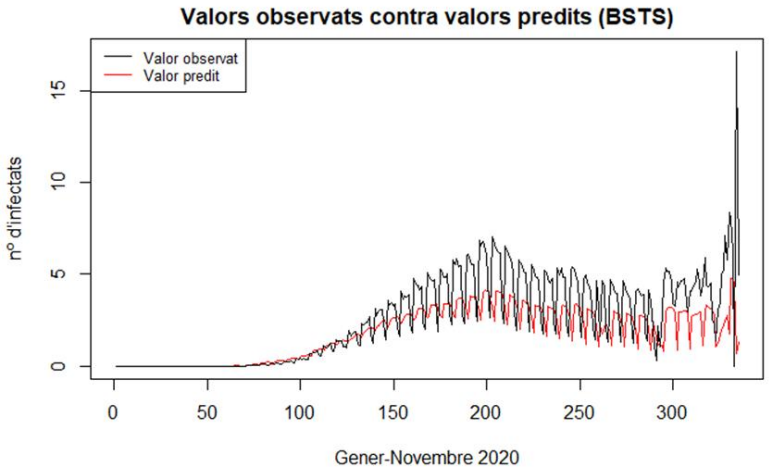


Figura II.35: Gràfic dels valors predits pel model BSTS en el període mostral

Gràficament sembla que ambdós models escollits ajusten considerablement bé les dades. Ambdós models presenten dificultats per estimar correctament les pujades de casos reportats als inicis de cada període, infravalorant d'aquesta manera les dades originals. També s'observa com a partir de les 300 observacions, en el moment del canvi de tendència de la sèrie original, el model SARIMA canvia la forma del patró que s'observa en els últims períodes. Malgrat aquestes petites diferències no es pot considerar que l'ajust dels dos models distin molt un de l'altre, i es procedirà a calcular-ne l'EPAM de cada un, per quantificar-ne l'ajust.

$$EPAM = \frac{\frac{100 * \sum_{i=1}^{n} |y_i - \hat{y}_i|}{y_i}}{n}$$

Aplicant la formula s'obté un EPAM del 45,20% segons el model SARIMA i un EPAM del 55,61% segons el model BSTS. Cal remarcar que encara que són uns percentatges d'error considerables, són raonables tenint en compte el tipus de dades que componen la sèrie, i l'elevat grau d'incertesa que aquestes presenten.

Es reafirma que els dos models ajusten de manera similar el període mostral, encara que sembla que el model clàssic s'imposa una mica.

2.5.2. Període extramostral

Donada l'anterior sèrie, s'ha pres com a període extramostral l'últim mes, resultant en un total de 14 previsions per cada model (Figura II.36).

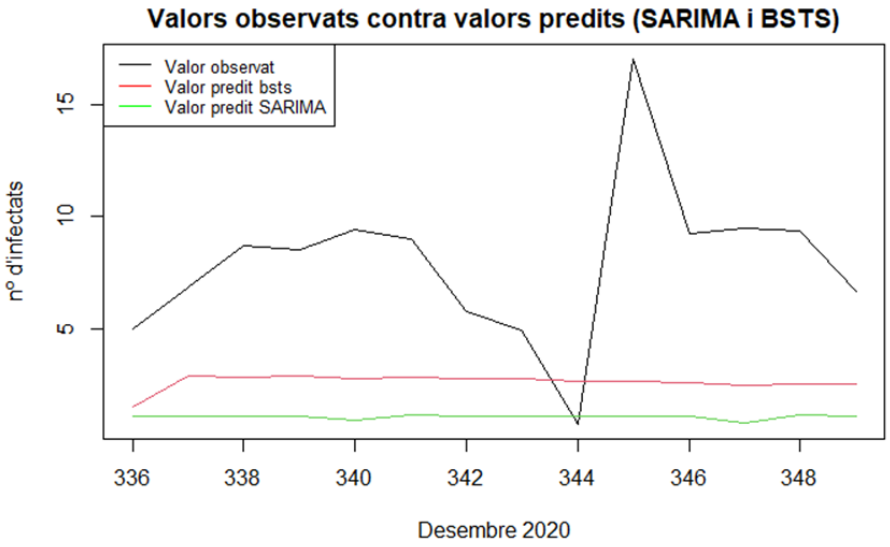


Figura II.36: Gràfic dels valors predits pel model SARIMA i BSTS en el període extramostral

Observant el gràfic, cap dels dos models ha estat capaç d'anticipar l'augment exponencial en el nombre d'infectat. El model bayesià ha recollit de forma més precisa el primer augment en el nombre de casos, però els següents valors atípics han impossibilitat les següents prediccions. Encara que cap dels dos models recull significativament el període extramostral de la sèrie, les oscil·lacions pròpies dels model *bsts* són més afins al comportament volàtil de les dades, per contraposició de la tendència contant del model SARIMA.

Aplicant un altre cop la formula del EPAM per les noves dades, s'obtenen uns errors del 81,37% i del 78,47% pels models SARIMA i *bsts* respectivament. Es reafirma com en el període extramostral el model *bsts* ajusta millor les dades, encara que aquesta diferència només s'aprecia gràcies al primer valor.

2.5.3. Resum

Amb els resultats anteriors sembla que ambdós models permeten ajustar significativament el període mostral. És en el període extramostral on només el model bayesià és capaç de seguir, al llarg dels primers valors, la tendència aleatòria de la sèrie II.

3. Sèrie III: Nombre d'infectats per la Covid-19 a Xina durant l'any 2020

La tercera sèrie de l'estudi consisteix en el nombre d'infectats per la Covid-19 a Xina durant l'any 2020 (Figura II.37). Aquesta recull diàriament el nombre d'infectats pel virus, reportats pel govern i els centres mèdics, iniciant el dia 1 de gener del 2020 i finalitzant el 14 de desembre del mateix any. Per facilitar la validació i l'ajust dels models proposats a continuació, s'ha decidit prendre els 11 primers mesos com a període mostral, deixant les ultimes dues setmanes com a període extramostral.

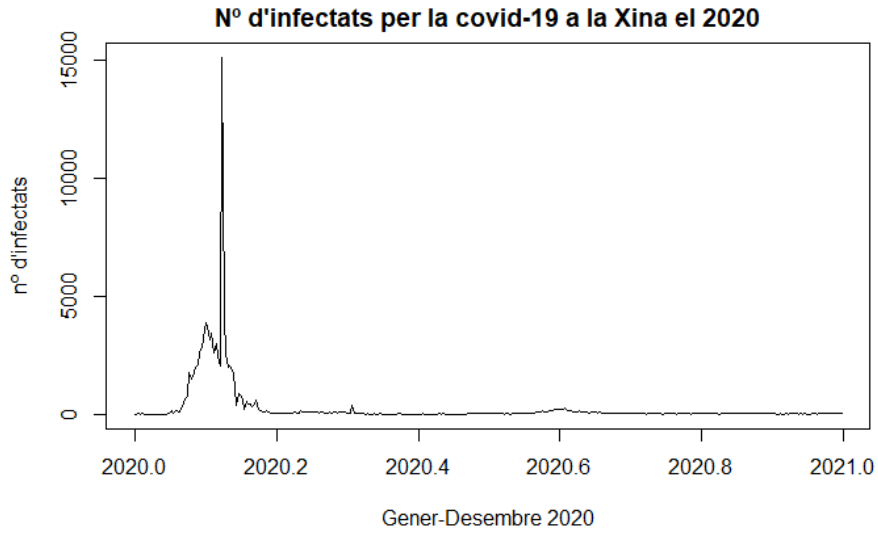


Figura II.37: Gràfic de la sèrie III:Nº d'infectats a Xina

Abans de procedir a la modelització de la sèrie, i com s'ha esmentat prèviament, s'utilitza el cens de la població de Xina durant l'any de les dades per estandarditzar les unitats i passar del nombre d'infectats al nombre d'infectats per 100.000 habitants seguin la formula que es veu a continuació:

$$Y = S * \frac{100.000\ habitants}{1.439.323.774\ habitants\ a\ Xina}$$

on S és la sèrie original i Y la resultant de la transformació.

Aquest procés es necessari per tenir les tres sèries temporals amb les mateixes unitats de mesura, facilitant-ne així les comparacions sobre els comportaments d'aquestes.

Finalment, la sèrie a analitzar és la que es representa en la Figura II.38:

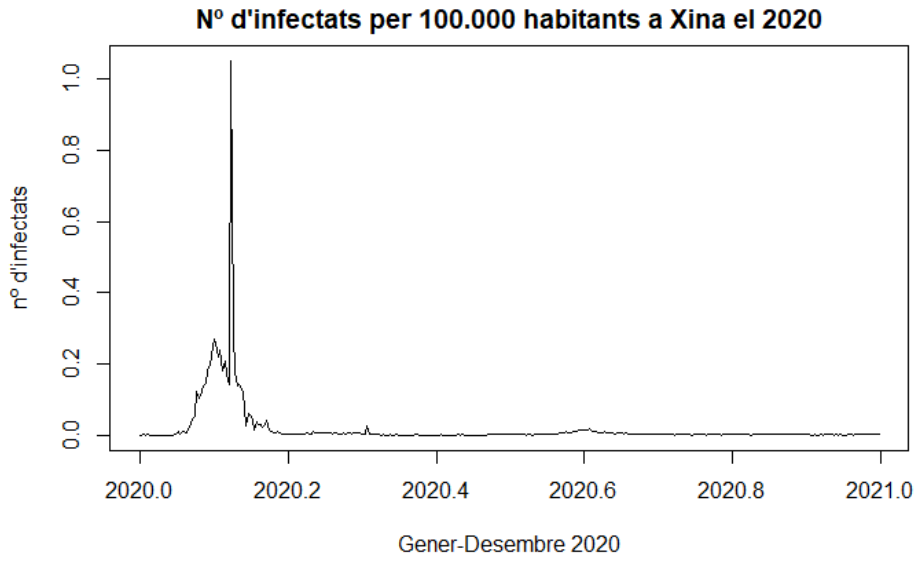


Figura II.38: Gràfic de la sèrie III ajustada

3.2. Identificació del model

Donada la sèrie temporal III, es pretén identificar els diversos components que la formen: Tendència, estacionalitat, ... També cal recordar que per poder modelitzar-la, aquesta ha de complir el supòsit d'estacionarietat mencionat anteriorment. Donat que les sèries utilitzades en l'anàlisi són no estacionaries caldrà aplicar diverses transformacions, segons els components que les conformin, per poder assolir aquesta propietat.

Primerament s'estudia si la sèrie presenta algun tipus de tendència. Per fer això és crea un model lineal amb les dades com a variable dependent i un vector seqüencial del 1 fins al nombre total de dades recollides per la sèrie com a variable explicativa. Utilitzant aquest model es busca quantificar l'efecte que té aquest nou vector, donat que si fos estadísticament significatiu implicaria que ajuda a explicar els valors de la variable dependent, i com a conseqüència, que existeix una tendència.

Aplicant aquests conceptes s'obté un p-valor de 1,90e-08, inferior al 5% de significació, obligant a acceptar la hipòtesi de que hi ha tendència i és necessari corregir-la. Com a conseqüent, s'apliquen les diferències regulars per intentar estabilitzar la mitjana i se'n grafiquen els resultats (Figura II.39).

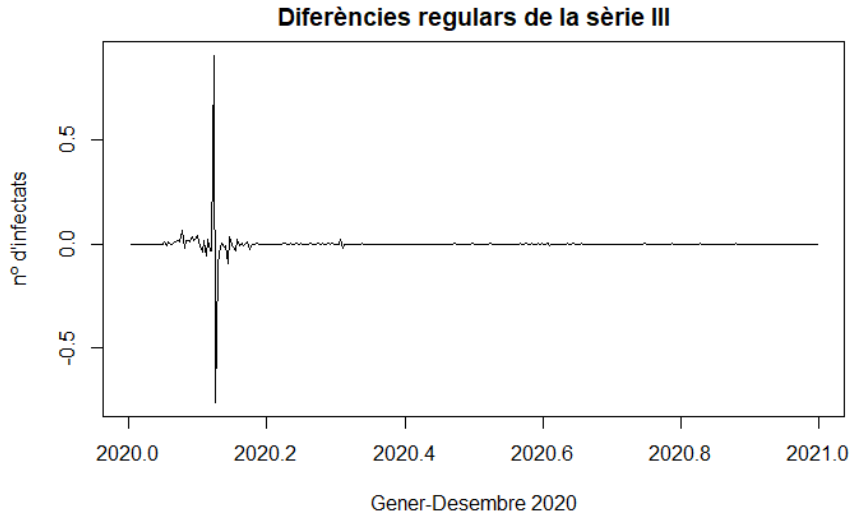


Figura II.39: Gràfic de les diferències regulars de la sèrie III

S'observa com la mitjana s'ha estabilitzat al voltant del zero i sembla ser constant al llarg de tota la sèrie. Si es repeteix la comprovació numèrica per la nova sèrie transformada, s'observa com el coeficient de la pendent a deixat de ser significativament diferent de zero (p-valor=0,959).

Per altra banda, a excepció d'uns valors, possiblement atípics, la variància també sembla mantenir-se constant durant tot el període i el gràfic no presenta patrons ni cicles

fàcilment identificables. Es pot assumir doncs, una variància constant i l'absència de la component estacional.

Per acabar amb el procés d'identificació, es comproven els gràfics de les autocorrelacions simples (Figura II.40) i parcials (Figura II.41) per definir els components mitjana mòbil i/o autoregressius:

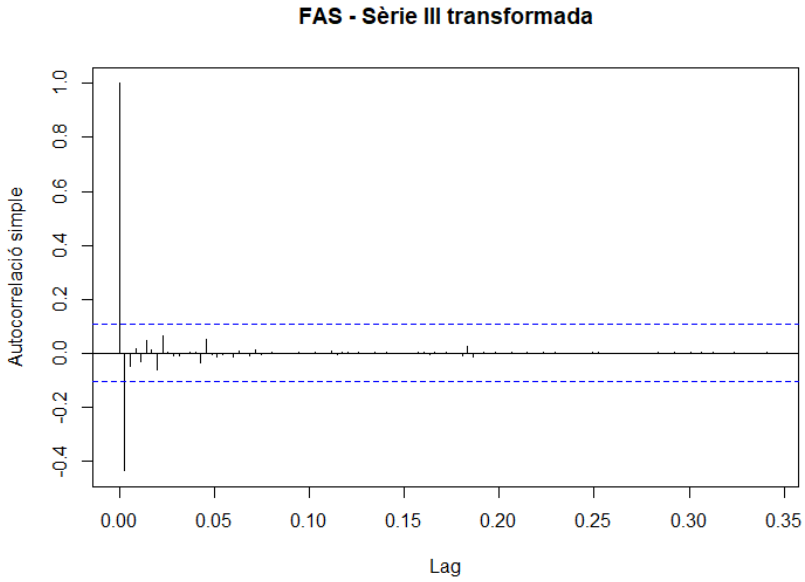


Figura II.40: Gràfic de les autocorrelacions simples de la sèrie III transformada

Donats dos *lags* significatius, previs a una sobtada caiguda cap a zero, la sèrie sembla presentar un component mitjana mòbil d'ordre 2→ $MA(q = 2)$.

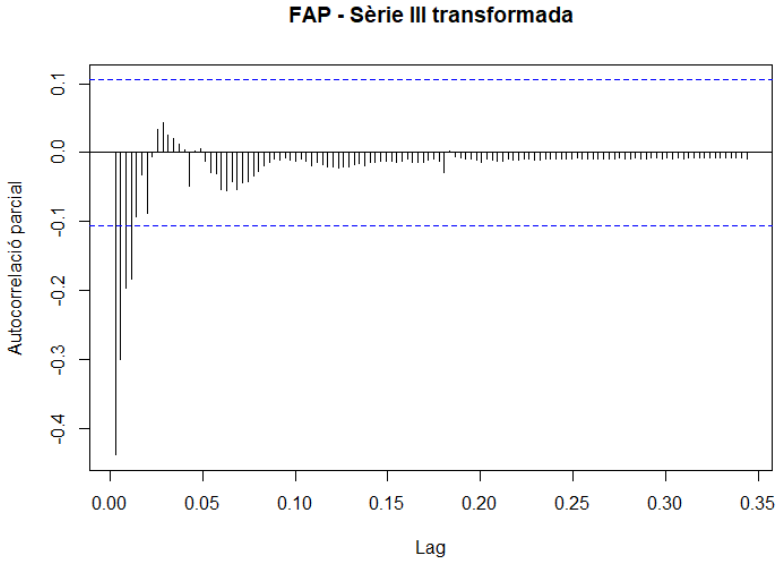


Figura II.41: Gràfic de les autocorrelacions parcials de la sèrie III transformada

Vista la tendència progressiva cap a zero, la sèrie sembla presentar un component autoregressiu d'ordre 0→ $AR(p = 0)$.

En resum, la sèrie III sobre el numero d'infectats per 100.000 habitants a Xina durant el 2020 presenta una tendència corregida amb diferències regulars, una component $MA(q = 2)$ i $AR(p = 0)$.

3.3. Anàlisi clàssica

En aquest apartat s'intentarà modelitzar la sèrie estudiada a partir de les metodologies clàssiques. Donada una sèrie amb tendència i sense estacionalitat, es procedirà a l'estimació d'un model tipus ARIMA.

3.3.1. Estimació i validació

Com s'ha trobat en la identificació de la sèrie, s'aplicarà un model ARIMA amb unes diferències regulars i una $MA(q = 2) \rightarrow ARIMA(p = 0, d = 1, q = 2)$:

	Coeficients	Error estàndard	p-valor
ma1	-0,6806	0,0555	1,556305e-34
ma2	-0,0252	0,0539	6,400952e-01

Figura II.42: Taula dels coeficients i p-valors del model clàssic per la sèrie III

Per poder utilitzar el model proposat per preveure els pròxims valors de la sèrie III, interessa els paràmetres d'aquest siguin significatius i que els residus reemplacin a un procés de soroll blanc (test de Ljung-Box). En cas de que no es complissin seria un indicador de la necessitat de buscar un model diferent.

Donats un p-valor no significatiu en el segon paràmetre del model es creu que aquest pot estar sobreestimat i s'ajusta un $ARIMA(p = 0, d = 1, q = 1)$:

	Coeficients	Error estàndard	p-valor
ma1	-0,7005	0,0367	2,200376e-81

Figura II.43: Taula dels coeficients i p-valors del model clàssic final per la sèrie III

Novament es mira la significació dels coeficients i que els residus reemplacin a un procés de soroll blanc.

Donats un p-valor significatiu en l'únic paràmetre del model i l'acceptació d'H_0: Els residus segueixen un soroll blanc , en el test de Ljung-Box, amb un p-valor<2,2e-16, es dona per validat el nou model.

3.3.2. Predicció

La funció de previsió puntual d'un procés estocàstic Y_t donat un origen N i un horitzó l, es representa com el valor esperat de Y_{N+l} condicionat per tota la informació històrica fins al punt d'origen k. Donat que segueix un model ARIMA del tipus $\phi'(B)Y_t = \mu + \theta(B)A_t$, quedaria la següent funció:

$$Y_N(l) \equiv E_N[Y_{N+l}] = \mu + \sum_{i=1}^{p+d} \phi' \; E_N[Y_{N+l-i}] + E_N[A_{N+l}] - \sum_{i=1}^{q} \theta_i E_N[A_{N+l-i}]$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen les prediccions pel període extramostral de la sèrie.

L'error de previsió de la funció de previsió es una variable aleatòria representada com la diferencia entre el valor observat en un moment N+l i la previsió donada per aquell mateix valor calculat amb l'origen N i l'horitzó l. Donat que segueix un model ARIMA del tipus $\phi'(B)Y_t = \mu + \theta(B)A_t$, quedaria la següent funció:

$$E_N \equiv Y_{N+l} - Y_N(l) = \psi^*(B)A_{N+l} - E[\psi^*(B)A_{N+l}]$$

$$on \qquad \psi^*(B)A_{N+l} = \sum_{i=0}^{\infty} \psi_i^* A_{N+l-i}$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen els errors de predicció pel període extramostral de la sèrie.

Finalment, si es tenen en compte aquestes previsions a continuació del període mostral de la sèrie III, s'esperaria l'evolució observada en la Figura II.44.

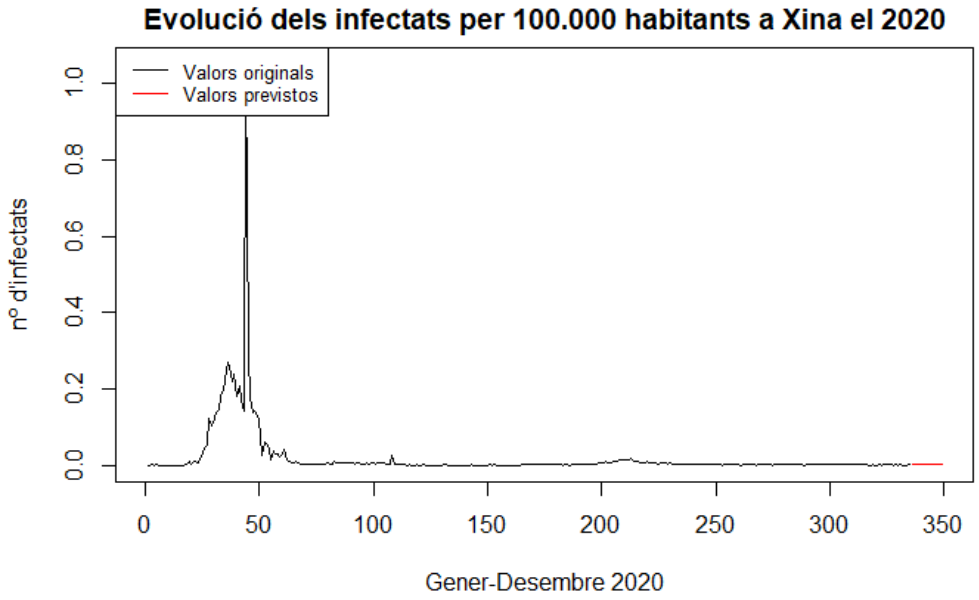


Figura II.44: Gràfic de la previsió d'infectats per 100.000 habitants a Xina el desembre

3.4. Anàlisi bayesiana

En aquest apartat s'intentarà modelitzar la sèrie estudiada a partir de les metodologies bayesianes. Donada una sèrie amb tendència i sense estacionalitat, es procedirà a l'estimació d'un model tipus *bsts*.

3.4.1. Estimació i validació

Per estimar un model bayesià cal centrar-se en la informació a priori que es coneix sobre la sèrie temporal. Prèviament s'ha identificat que la sèrie no té component regressiva, però si que serà necessària una manera de modelitzar-ne la tendència.

La manera més comuna de modelitzar la tendència es fent ús de la tendència lineal local, però encara que no s'ha detectat una component autoregressiva, per la tipologia de les dades sembla que els valors de la sèrie temporal depenen linealment dels seus propis valors anterior. Com a conseqüència és modelitzarà un component autoregressiu d'ordre p, el qual recull aquesta relació més un terme estocàstic per tal de representar la tendència de la sèrie:

$$y_t = \mu_t + \epsilon_t$$
$$\mu_{t+1} = \sum_{i=0}^{p-1} \phi_i \mu_{t-i} + \eta_t$$

Curiosament, al modelitzar el model ARIMA s'havia pres un component mitjana mòbil en comptes d'un autoregressiu, però observant l'error absolut acumulat per les dues maneres de representar la tendència (Figura II.45), el nou ajust és més proper als valors de la sèrie.

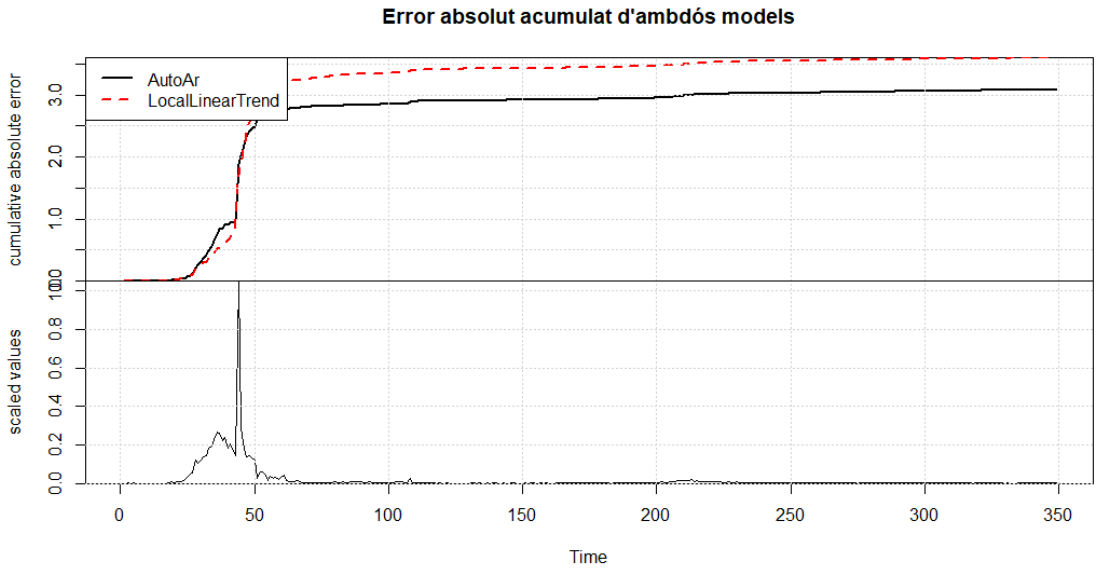


Figura II.45: Gràfic de l'error absolut acumulat dels models bayesians per la sèrie III

Fent ús d'aquests components es pot ajustar el model *bsts* i s'identifica la distribució a posteriori de la sèrie (Figura II.46).

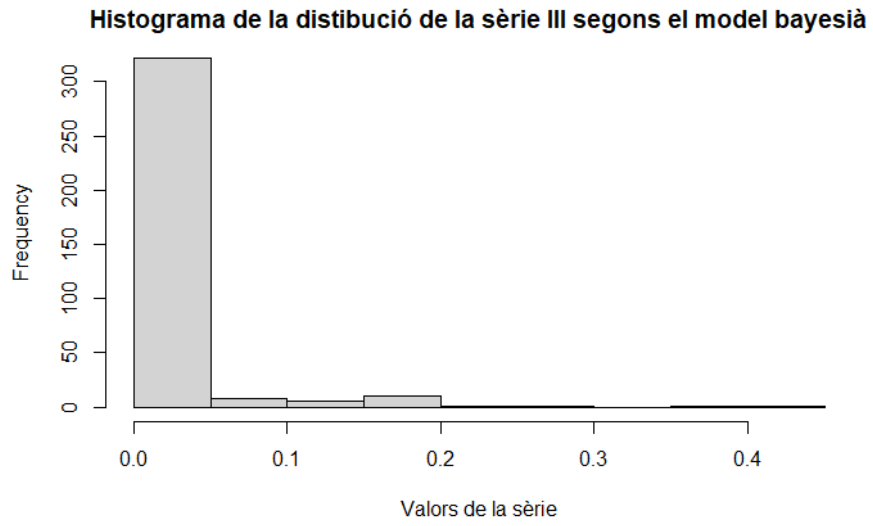


Figura II.46: Histograma de la distribució a posteriori del model bayesià per la sèrie III

Segons aquest model, la sèrie te una mediana sobre els 0,0016 infectats per cada 100.000 habitants, amb una desviació típica de 0,0484 casos amunt i avall.

Per poder utilitzar el model proposat per preveure els pròxims valors de la sèrie III, interessa que els residus segueixin una distribució normal i siguin estacionaris. En cas de que no es complissin aquestes premisses, seria un indicador de la necessitat de buscar un model diferent.

Donat el següent gràfic sobre la distribució dels residus del model proposat (Figura II.47), s'observa com, a excepció del extrems, els valors tendeixen als quantils de la distribució normal estàndard. Es suposa normalitat als residus.

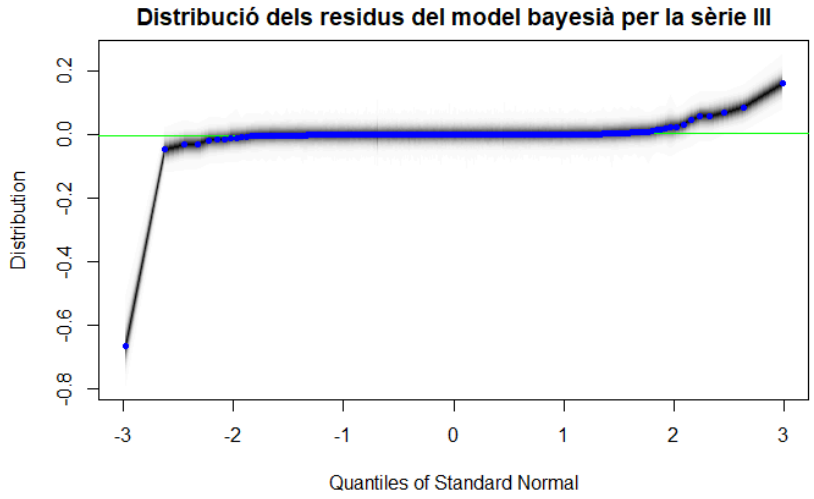


Figura II.47: Gràfic de la distribució dels residus del model bayesià per la sèrie III

Per comprovar l'estacionarietat dels residus es pretén que aquests tendeixin a no tenir autocorrelació (autocorrelació zero) (Figura II.48).

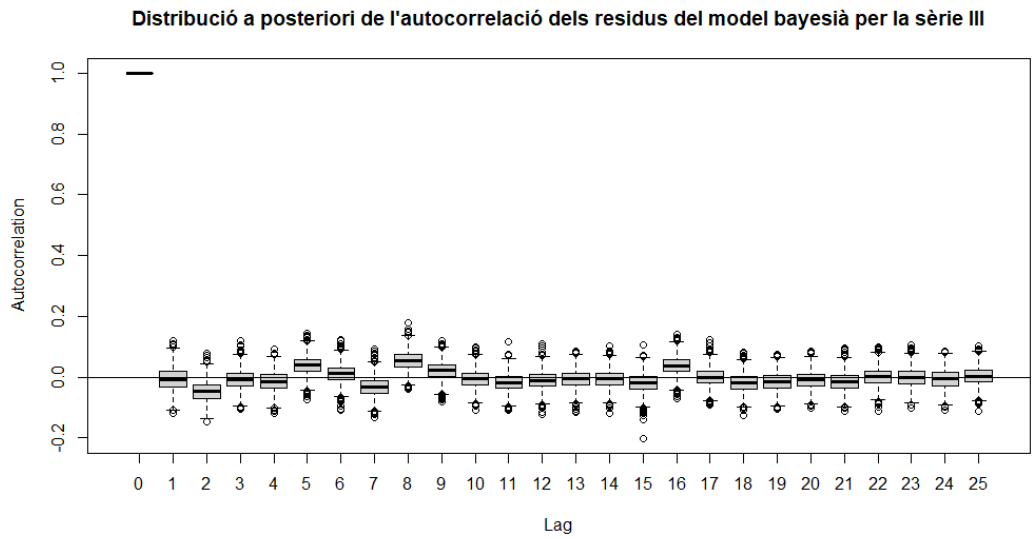


Figura II.48: Gràfic de la distribució a posteriori de l'autocorrelació dels residus del model bayesià per la sèrie III

Observant el gràfic, a partir del *lag* 10 els residus del model bayesià tendeixen a zero. Havent-se complert les dues validacions es pot procedir a realitzar previsions per la sèrie III.

3.4.2. Predicció

La funció de previsió puntual d'un procés estocàstic Y_t donat un origen N i un horitzó l, es representa com el valor esperat de Y_{N+l} condicionat per tota la informació històrica fins al punt d'origen k:

$$Y_N(l) \equiv E_N[Y_{N+l}] \ \ per \ \ l = 1,2, \dots$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen les prediccions pel període extramostral de la sèrie.

L'error de previsió de la funció de previsió es una variable aleatòria representada com la diferencia entre el valor observat en un moment N+l i la previsió donada per aquell mateix valor calculat amb l'origen N i l'horitzó l:

$$E_N \equiv Y_{N+l} - Y_N(l) \ \ per \ \ l = 1,2, \dots$$

Aplicant aquest principi, reemplaçant els paràmetres de la formula pels seus valors estimats segons el model, s'obtenen els errors de predicció pel període extramostral de la sèrie.

Finalment, si es tenen en compte aquestes previsions a continuació del període mostral de la sèrie III, s'esperaria l'evolució observada en la Figura II.49.

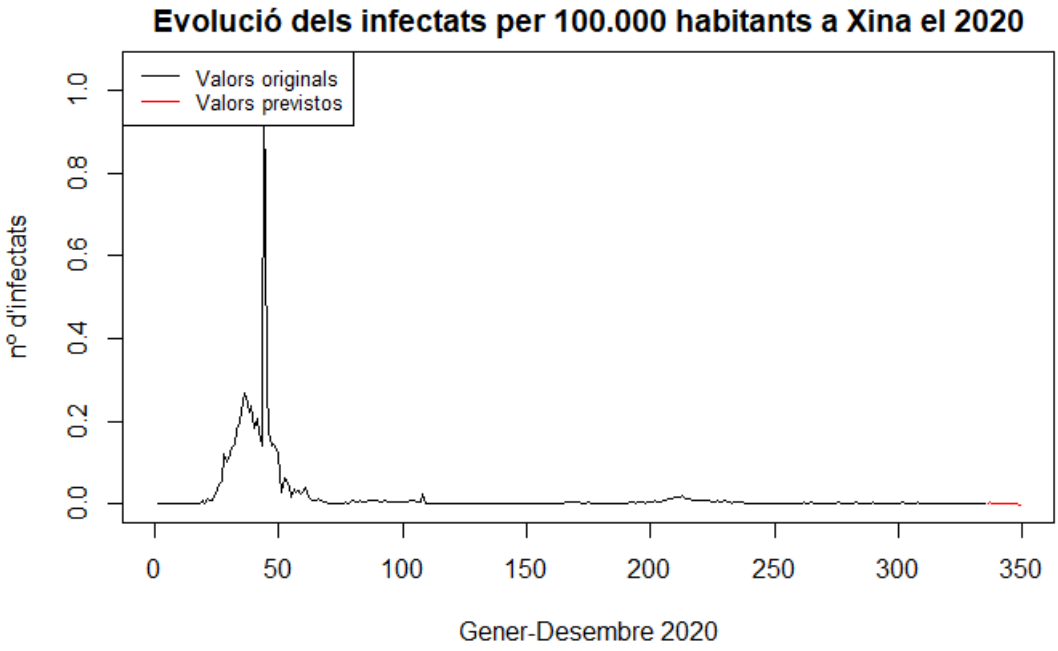


Figura II.49: Gràfic de la previsió d'infectats per 100.000 habitants a Xina el desembre

3.5. Comparació dels diferents models obtinguts

Un cop escollits els diferents models, segons la metodologia clàssica i la bayesiana, per ajustar la sèrie III, es procedeix a estudiar quin dels dos enfocaments permet ajustar unes millors previsions.

3.5.1. Període mostral

Donada l'anterior sèrie, s'ha pres com a període mostral els primers 11 mesos, resultant en un total de 335 observacions històriques. A partir d'aquestes, i fent ús dels dos models, s'han obtingut els següents ajustos (Figura II.50 i Figura II.51):

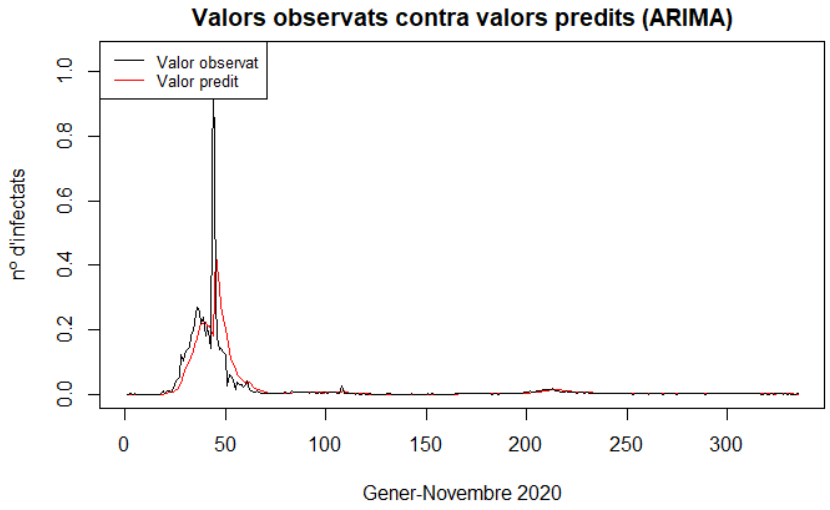


Figura II.50: Gràfic dels valors predits pel model ARIMA en el període mostral

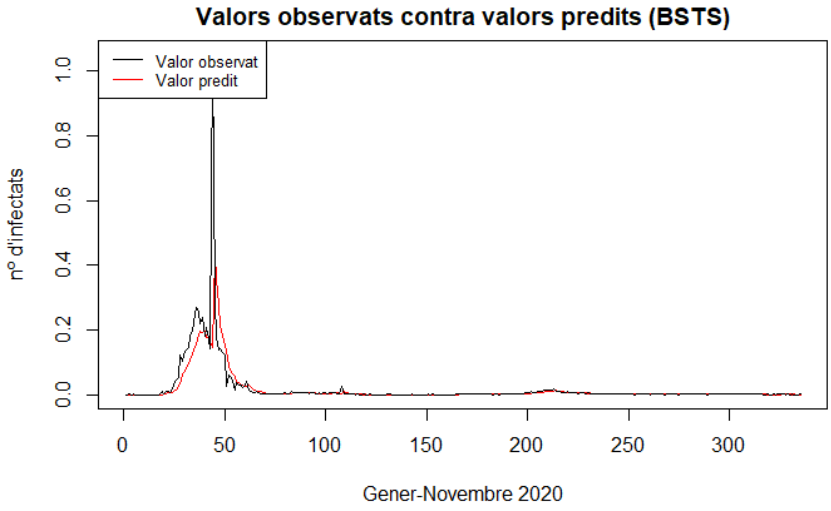


Figura II.51: Gràfic dels valors predits pel model BSTS en el període mostral

Gràficament sembla que ambdós models escollits ajusten considerablement les dades. Ambdós presenten dificultats per predir el pic observat al voltant de les 45 observacions, però sembla com el model bayesià tarda menys en recuperar-se de la influencia d'aquest valor atípic, retornant ràpidament a valors similars als reals. Malgrat aquestes petites diferències no es pot considerar que l'ajust dels dos models distin molt un de l'altre, i es procedirà a calcular-ne l'EPAM de cada un, per quantificar-ne l'ajust.

$$EPAM = \frac{\frac{100 * \sum_{i=1}^{n} |y_i - \hat{y}_i|}{y_i}}{n}$$

Aplicant la formula s'obté un EPAM del 74,95% segons el model ARIMA i un EPAM del 57,71% segons el model *bsts*. Cal remarcar que encara que són uns percentatges d'error considerables, són raonables tenint en compte el tipus de dades que componen la sèrie, i l'elevat grau d'incertesa que aquestes presenten.

Es reafirma que els dos models ajusten de manera similar el període mostral, encara que sembla que el model bayesià s'imposa una mica.

3.5.2. Període extramostral

Donada l'anterior sèrie, s'ha pres com a període extramostral l'últim mes, resultant en un total de 14 previsions per cada model (Figura II.52).

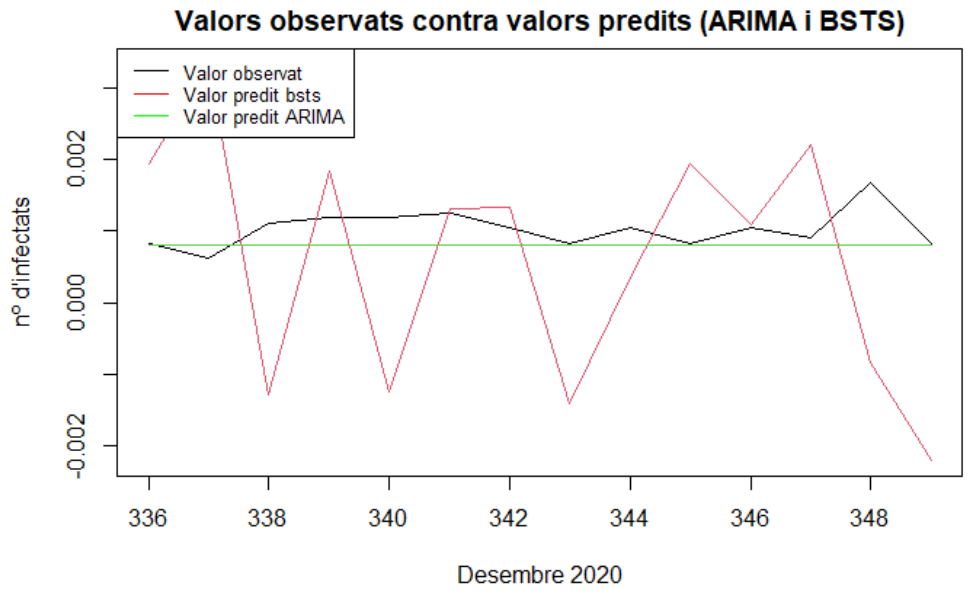


Figura II.52: Gràfic dels valors predits pel model ARIMA i BSTS en el període extramostral

Observant el gràfic, el model bayesià presenta unes oscil·lacions majors sobre el valor real. Sembla que li costa més ajustar aquests valors, donat que són pràcticament constants. A la vegada, també presenta valors negatius, els quals no es poden donar a la pràctica. Aquí, el model ARIMA sembla que és capaç de produir un millor ajust, amb una previsió constant molt propera al valor real. Sembla que el model clàssic subestima una mica els valors observats durant el període extramostral.

Aplicant un altre cop la formula del EPAM per les noves dades, s'obté un ajuts del 20,78% i del 157,31% pels models ARIMA i *bsts* respectivament. Es reafirma com en el període extramostral el model ARIMA ajusta millor que el bayesià.

3.5.3. Resum

Amb els resultats anteriors sembla que el model *bsts* ajusta millor les dades mostrals, recuperant-se, abans que el model clàssic, de les influencies de valors atípics. És a partir de l'estabilització dels valors originals, on l'error acumulat per les variacions fa que el model ARIMA sigui preferible.

III. COMPARACIÓ DELS RESULTATS ENTRE SÈRIES

Un cop observades les diferències en l'ajust de les metodologies clàssiques i bayesianes, per la mateixa sèrie, es pretén focalitzar en la capacitat d'ajust d'aquestes metodologies en diferents situacions (sèries I, II i III). En el següent apartat s'analitzaran les capacitats d'ajust de cada metodologia al trobar-se amb sèries composades per components distints. Es pretén analitzar les capacitats d'ajust i predicció de les dues metodologies en diversos escenaris.

1. Metodologia clàssica

Segons l'estimació realitzada a l'apartat "II. PART PRÀCTICA: ESTIMACIÓ DELS MODELS CLÀSSICS VS BAYESIANS" els models clàssics per les sèries estudiades són els següents:

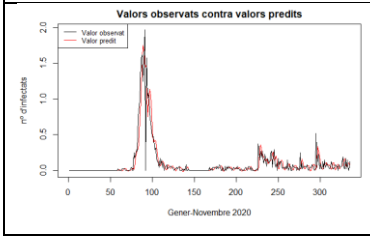
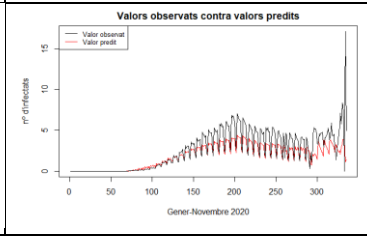
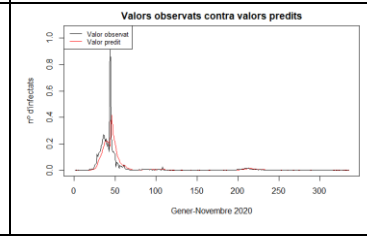
Sèrie I: Nova Zelanda	Sèrie II: Mèxic	Sèrie III: Xina
$ARIMA(0,1,2)$	$SARIMA(0,1,1)(0,1,2)^7$	$ARIMA(0,1,1)$
		

Figura III.1: Taula dels models clàssics per cada sèrie i la seva representació gràfica respecte l'original

Per tal de comparar l'ajust dels models clàssics en aquest tipus de sèries, se'n han calculat alguns estadístics bàsics per els dos períodes estudiats, i s'han comparat amb els valors originals:

La taula següent mostra els resultats de diversos estadístics calculats durant el període mostral de les tres sèries i els seus respectius models clàssics associats:

Estadístics	Sèrie I: Nova Zelanda		Sèrie II: Mèxic		Sèrie III: Xina	
	Model	Sèrie	Model	Sèrie	Model	Sèrie
Mitjana	0,1050	0,1052	1,8348	2,5631	0,0190	0,0190
Mediana	0,0277	0,0207	1,9914	2,2516	0,002	0,0019
Desviació típica	0,2591	0,2676	1,3954	2,3443	0,0567	0,0728
Valor màx.	1,7477	1,9700	4,3667	17,1322	0,4424	1,052
Valor min.	-0,0119	0,0000	0,0000	0,0000	0,0000	0,0000
RIQ	0,0768	0,0829	2,8967	4,4985	0,0048	0,005

Figura III.2: Taula amb els estadístics de les sèries i els seus models durant el període mostral

En la primera sèrie s'ha ajustat un model amb component mitjana mòbil, al qual se li ha aplicat les diferències regulars. La mitjana del nombre d'infectats calculats pel model és pràcticament igual a la dels valors originals. El mateix passa amb la mediana i és molt similar amb la desviació típica. Encara que el model permet l'aparició de valors negatius

(valor mínim de -0,0119), un escenari inversemblant en dades reals, l'amplitud del rang de valors observats pel RIQ torna a reafirmar la fiabilitat del model.

Pel que fa a la segona sèrie, s'ha modelitzat una component mitjana mòbil amb diferències regulars i estacionalitat, prèviament aplicada una transformació a les dades originals. Aquí és on els estadístics del model disten més dels valors reals. Per començar el nombre d'infectats per 100.000 habitants a Mèxic es major al de les altres sèrie, causant un rang de valors més ampli i una desviació típica major a les altres sèries. Encara que no es tenen en compte els valors extrems, el RIQ segueix presentant una diferència de 0,8 ($\frac{RIQ_{sèrie}-RIQ_{model}}{2}$) casos d'infectats més i menys en el primer i tercer quartil de les dades extretes del model. Aquestes diferències que també s'observen en la mitjana i la mediana, s'acaben disparant en el valor màxim de la sèrie original, on el model ha subestimat significativament el pic màxim d'infectats reportats el mes de novembre.

Finalment, la tercera sèrie és molt similar a la primera, afegint una transformació a les dades. Aquí s'observa com tant la mitjana com la mediana obtingudes del model són pràcticament iguals a les de les dades originals. La diferència recau en l'augment sobtat de casos detectats el març, donat que el model infravalora el pic de les dades, resultant en una diferencia significativa en els valors màxims i la desviació típica. No obstant això, el RIQ, que ignora els valors més extrems, torna a donar molt similar.

La taula següent mostra els resultats de diversos estadístics calculats durant el període extramostral de les tres sèries i els seus respectius models clàssics associats:

Estadístics	Sèrie I: Nova Zelanda		Sèrie II: Mèxic		Sèrie III: Xina	
	Model	Sèrie	Model	Sèrie	Model	Sèrie
Mitjana	0,0549	0,0592	1,8941	7,9207	0,0008	0,001
Mediana	0,0564	0,0415	1,9979	8,6406	0,0008	0,001
Desviació típica	0,0056	0,0670	0,6149	3,6095	0	0,0003
Valor màx.	0,0564	0,1866	2,6743	17,043	0,0008	0,0017
Valor min.	0,0353	0	0,3661	0,7803	0,0008	0,0006
RIQ	0	0,0829	0,3658	3,3147	0	0,0003

Figura III.3: Taula amb els estadístics de les sèries i els seus models durant el període extramostral

Pel que respecta a les prediccions extretes dels diversos models, comparades amb el període extramostral de les sèries, s'obtenen ajustos similars als del període mostral.

La sèrie de Nova Zelanda manté una mitjana i una mediana considerablement similars. Les diferències apareixen en el rang de valors. El model té un interval de valors de [0,0353 , 0,0564] amb una desviació de 0,0056, mentre que els valors originals oscil·len

entre [0 , 0,1866] amb una desviació de 0,076. Cal destacar que el RIQ del model es 0 degut a que, a excepció de la primera previsió, la resta de valors són constants.

El període extramostral de la sèrie de Mèxic torna a presentar un pic sobtat de casos d'infectats per la Covid-19 el 10 de desembre i al tenir únicament una mostra de 14 dades queda molt reflectit en els resultats numèrics. La diferència entre les desviacions típiques és de 3 infectats per cada 100.000 habitant i encara que no es tingues en compte aquesta dada atípica, el model segueix subestimat considerablement les dades històriques. En aquest model si que s'han obtingut observacions menys constants però no s'ha ajustat satisfactòriament la volatilitat de les dades.

Finalment, la sèrie de Xina torna a ser similar al cas de Nova Zelanda. Les previsions produïdes pel model són constants, però donada la poca variació dels resultats reals, aquestes s'ajusten de forma satisfactòria.

En general s'observa com les sèries amb menys variabilitat i uns valors més constants són les que queden millor ajustades pels models clàssics. Com a exemple s'observen els valors obtinguts per les sèries de Nova Zelanda i Xina contra la de Mèxic. El mateix problema es troba amb el període extramostral, on les previsions més fiables són les de la sèrie de Xina, que recull pràcticament el mateix nombre d'infectats detectats al llarg de tot el mes de desembre. Si es té en compte Nova Zelanda, que ja presenta un RIQ una mica superior, aquí l'ajust segueix essent bo però no es recullen bé els màxims i els mínims. Finalment Mèxic, per culpa de l'estacionalitat i algun valor atípic, presenta un RIQ considerablement superior a les altres sèries i aquí els valors previstos disten molt dels reals.

Curiosament, s'observa com totes les sèries analitzades sobre el nombre d'infectats per la Covid-19 presenten components mitjana mòbil, mentre que no tenen part autoregressiva. Prèviament a l'anàlisi semblaria que les dades haurien de tenir una forta correlació amb valors passats, donada com a font d'obtenció de les dades una malaltia infecciosa que es transmet exponencialment tenint en compte el nombre d'infectats previs, però no es així. Sembla ser que la component autoregressiva s'amaga darrera la variabilitat i la no estacionarietat de les dades. Per altre banda, també s'explica la component mitjana mòbil tenint en compte l'alta volatilitat i grau d'incertesa de les dades, donat que la propagació del virus és molt exponencial i difícil d'analitzar únicament amb una regressió lineal directe d'un cert nombre de valors anteriors.

Independentment del país d'obtenció de les dades tots els models són relativament similars, a excepció de Mèxic que presenta un component estacional afegit.

2. Metodologia bayesiana

Segons l'estimació realitzada a l'apartat "II. PART PRÀCTICA: ESTIMACIÓ DELS MODELS CLÀSSICS VS BAYESIANS" els models bayesians per les sèries estudiades són els següents:

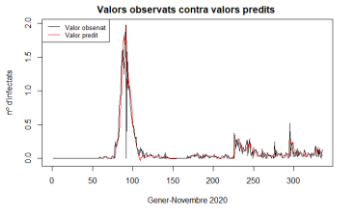
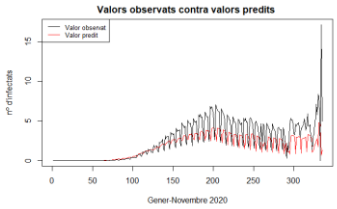
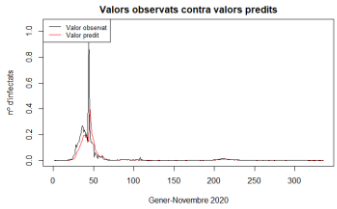
Sèrie I: Nova Zelanda	Sèrie II: Mèxic	Sèrie III: Xina
bsts::AddLocalLinearTrend	bsts::AddLocalLinearTrend bsts::AddSeasonal	bsts::AddAutoAr
		

Figura III.4: Taula dels models bayesians per cada sèrie i la seva representació gràfica respecte l'original

Per tal de comparar l'ajust dels models bayesians en aquest tipus de sèries, se'n han calculat alguns estadístics bàsics per els dos períodes estudiats, i s'han comparat amb els valors originals:

La taula següent mostra els resultats de diversos estadístics calculats durant el període mostral de les tres sèries i els seus respectius models clàssica associats:

Estadístics	Sèrie I: Nova Zelanda		Sèrie II: Mèxic		Sèrie III: Xina	
	Model	Sèrie	Model	Sèrie	Model	Sèrie
Mitjana	0,1052	0,1052	1,7796	2,5631	0,016	0,019
Mediana	0,0269	0,0207	2,0319	2,2516	0,0016	0,0019
Desviació típica	0,2787	0,2676	1,3662	2,3443	0,0494	0,0728
Valor màx.	1,9813	1,97	4,7816	17,1322	0,4333	1,052
Valor min.	-0,0327	0	0	0	0	0
RIQ	0,0655	0,0829	2,7443	4,4985	0,0042	0,005

Figura III.5: Taula amb els estadístics de les sèries i els seus models durant el període mostral

En la primera sèrie s'ha ajustat un model amb component "AddLocalLinearTrend". La mitjana del nombre d'infectats estimats pel model és pràcticament igual a la dels valors originals. El mateix passa amb la mediana i és molt similar amb la desviació típica. Encara que el model permet l'aparició de valors negatius (valor mínim de -0,0327), anàlogament al model clàssic, l'amplitud del rang de valors observats pel RIQ torna a reafirmar la fiabilitat del model.

Pel que fa a la segona sèrie, s'ha modelitzat amb la combinació d'una component "AddLocalLinearTrend" per la tendència i una "AddSeasonal" per l'estacionalitat de les dades. Aquí, igual al model SARIMA, tornem a trobar discrepàncies entre els valors dels estadístics del model i els de la sèrie original. Ara el RIQ ha augmentat la diferència d'infectats més i menys en el primer i tercer quartil de les dades extretes del model de 0,8 a 1,3 ($\frac{RIQ_{sèrie}-RIQ_{model}}{2}$). Aquestes diferències que també s'observen en la mitjana i

la mediana, s'acaben disparant en el valor màxim de la sèrie original, on el model ha subestimat significativament el pic màxim d'infectats reportats el mes de novembre.

Finalment, la tercera sèrie hauria de ser molt similar a la primera, encara que ara la tendència 'ha modelitzat amb un component "AddAutoAr". Aquí s'observa com tant la mitjana com la mediana obtingudes del model són pràcticament iguals a les de les dades originals. La diferència recau en l'augment sobtat de casos detectats el març, donat que el model infravalora el pic de les dades, resultant en una diferencia significativa en els valors màxims i la desviació típica. No obstant això, el RIQ, que ignora els valors més extrems, torna a donar molt similar.

La taula següent mostra els resultats de diversos estadístics calculats durant el període extramostral de les tres sèries i els seus respectius models clàssica associats:

Estadístics	Sèrie I: Nova Zelanda		Sèrie II: Mèxic		Sèrie III: Xina	
	Model	Sèrie	Model	Sèrie	Model	Sèrie
Mitjana	0,0747	0,0592	2,648	7,9207	0,0006	0,001
Mediana	0,0758	0,0415	2,7221	8,6406	0,0012	0,001
Desviació típica	0,0035	0,067	0,3437	3,6095	0,0017	0,0003
Valor màx.	0,0797	0,1866	2,9282	17,043	0,0033	0,0017
Valor min.	0,069	0	1,5625	0,7803	-0,0022	0,0006
RIQ	0,0053	0,0829	0,2457	3,3147	0,003	0,0003

Figura III.6: Taula amb els estadístics de les sèries i els seus models durant el període extramostral

Pel que respecta a les prediccions extretes dels diversos models, comparades amb el període extramostral de les sèries, s'obtenen ajustos pitjors als del període mostral.

La sèrie de Nova Zelanda presenta una mitjana i una mediana superiors a les dades reals, encara que el rang de valors es inferior. Les previsions del model es troben més concentrades, reduint la diferència entre aquestes i subestimat la variabilitat de les dades.

El període extramostral de la sèrie de Mèxic torna a presentar un pic sobtat de casos d'infectats per la Covid-19 el 10 de desembre i al tenir únicament una mostra de 14 dades queda molt reflectit en els resultats numèrics. La diferència entre les desviacions típiques és de més de 3 infectats per cada 100.000 habitant i encara que no es tingues en compte aquesta dada atípica, el model segueix subestimat considerablement les dades històriques. Es torna a observar un rang d'observacions més concentrat i dins del rang real de les dades.

Finalment, el model per la sèrie de Xina sembla ser el que millor ha ajustat les dades encara que és la que presentava un pitjor EPAM de totes, degut a l'ús de valors molt

més reduïts. Aquí el RIQ és major al de la sèrie original, però tant les mitjanes com les s'assemblen considerablement. A diferència de les sèries anteriors, la previsió del nombre d'infectats a Xina presenta major variabilitat que les dades reals.

En general sembla que els models bayesians presenten una major variabilitat, resultant en diferències més significatives en els estadístics utilitzats. No obstant això, considerant que les dades també presenten aquestes variacions, permeten tenir una idea més realista de l'evolució de la sèrie. Durant el període mostral la sèrie de Nova Zelanda i la de Xina han ajustat significativament les dades, encara que en l'esdevenir del període extramostral, les fortes desviacions típiques han resultat en estadístics més diferents.

Curiosament, la sèrie de la Xina ha sigut la única en reflectir el component autoregressiu que s'esperaria d'aquest tipus de dades. Encara que l'EPAM d'aquesta semblava molt dolent, degut a que al treballar amb dades tant petites ressalta molt les diferències , numèricament sembla que és la que s'ajusta més als valors reals.

3. Conclusions

Vistos els resultats anteriors, s'observa com ambdós metodologies han presentat resultats similars.

Per una banda, els models clàssics presenten resultats molt més constants i infravaloren les variacions pròpies de les dades. També s'ha vist com insten a una modelització contradictòria al que s'esperaria de les dades (no apareix cap component autoregressiu), encara que en sèries amb valors molt més constants, preveuen resultats molt similars als reals.

Per altra banda, els models bayesians destaquen per sobremodelitzar la variació de les dades originals, permeten preveure més fidelment el possible esdevenir de les dades o les conseqüències de la gran volatilitat d'aquestes. Encara que es veritat que semblen ajustar millor les tendències que els models clàssics, el gran rang de valors ens causa resultats dels estadístics més distants als originals.

Malgrat els problemes de les dades, ambdues metodologies presenten unes prediccions a curt termini similars als valors reals. Certament, si es pretén observar uns resultats més a futur, els models bayesians permeten ajustar més el tipus de comportament i tendència de les dades.

IV. AMPLIACIÓ DELS MODELS BSTS

Un cop finalitzat l'ajust pràctic per les dades de la Covid-19 segons les dues metodologies presentades s'ha intentat ampliar una mica els usos pràctics que ofereix la metodologia bayesiana descrita. Un d'aquests usos consisteix en mesurar l'efecte potencial que ha tingut un esdeveniment passat en la sèrie estudiada. Per quantificar aquest efecte s'utilitza el paquet d'R "CausalImpact".

"CausalImpact" d'R és un paquet utilitzat per a realitzar inferència causal basada en models *bsts*. Aquest paquet implementa una aproximació per estimar l'efecte causal d'una intervenció "dissenyada" en una sèrie temporal. Donada la sèrie d'estudi, crea un model *bsts* per intentar predir el contrafactual, és a dir, com hauria evolucionat la sèrie estudiada després d'una intervenció o esdeveniment si aquest no hagués arribat a succeir. La mesura d'aquest efecte es realitza a partir de l'anàlisi de les diferències entre el comportament esperat i el comportament observat.

Per tal de realitzar aquesta inferència causal, el model necessita realitzar algun supòsits:

- Existeix una sèrie temporal controlada que no es va veure afectada per la intervenció de l'esdeveniment.
- La relació entre la covariable i la sèrie estudiada es manté estable durant tot el període posterior a la intervenció.

❖ Exercici simulat per veure el funcionament del paquet "*CausalImpact*"

Donada un conjunt de dades format per una variable resposta 'y' i un predictor 'x1', ambdós amb 100 observacions, arbitràriament s'imposa un efecte causat artificialment, augmentant en 10 unitats els valors de la variable resposta després de l'observació 71.

```
> set.seed(1)
> x1 <- 100 + arima.sim(model = list(ar = 0.999), n = 100)
> y <- 1.2 * x1 + rnorm(100)
> y[71:100] <- y[71:100] + 10
> data <- cbind(y, x1)
```

Sigui 'data' una sèrie temporal amb dues columnes: la variable estudiada, calculada a partir del predictor 'x1', i el seu predictor:

```
> head(data)
Time Series:
Start = 1
End = 6
Frequency = 1
          y       x1
1 105.2950 88.21513
2 105.8943 88.48415
3 106.6209 87.87684
4 106.1572 86.77954
5 101.2812 84.62243
6 101.4484 84.60650
```

Es busca detectar l'escenari de l'augment de les 10 unitats en la variable resposta, que prèviament s'ha implementat. Per fer això es defineix un període pre-intervenció, per entrenar el model, i un període post-intervenció, per calcular la predicció contrafactual, i s'introdueix a la funció "CausalImpact" per inferències:

```
> pre.period <- c(1, 70)
> post.period <- c(71, 100)
> impact <- CausalImpact(data, pre.period, post.period)
```

D'aquesta manera, el paquet construeix un model estructural de sèries temporals amb el qual realitzar inferències posteriors i calcular estimacions de l'efecte causal a partir de l'observació 71.

Un cop obtingut el model es poden visualitzar els resultats de forma gràfica (Figura IV.1) o numèrica (Figura IV.2):

```
> plot(impact)
```

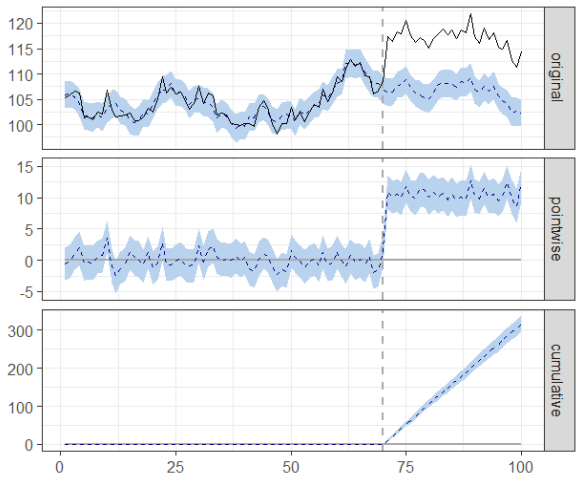


Figura IV.1: Gràfics de l'efecte causal per a la simulació d'exemple

El primer gràfic torna la sèrie temporal de la variable resposta, en negre, i una predicció contrafactual a partir de l'observació 71, en blau.

El segon gràfic mostra les diferències observades entre les dades reals i la predicció. Aquest seria l'efecte causal puntual segons el model. Aquí es veu clarament un canvi en el punt de la intervenció simulada.

Finalment a l'últim s'observa l'efecte causal acumulat. Es trona a veure com a partir de la intervenció, aquestes diferències augmenten significativament.

```
> summary(impact)
```

```
Posterior inference {CausalImpact}

                          Average        Cumulative
Actual                    117            3511
Prediction (s.d.)         107 (0.37)     3196 (11.16)
95% CI                    [106, 107]     [3173, 3217]

Absolute effect (s.d.)    11 (0.37)      316 (11.16)
95% CI                    [9.8, 11]      [294.7, 338]

Relative effect (s.d.)    9.9% (0.35%)   9.9% (0.35%)
95% CI                    [9.2%, 11%]    [9.2%, 11%]

Posterior tail-area probability p:   0.001
Posterior prob. of a causal effect:  99.8996%
```

Figura IV.2: Taula de l'efecte causal per a la simulació d'exemple

Si es reporten els mateixos resultats de forma numèrica, s'observa que d'una mitjana de 117 unitats observades en el període post-intervenció, si aquesta no s'hagués donat es tindrien nomes 107 unitats. L'efecte absolut es d'un augment de 11 unitats, el qual representa un augment del 9,9%.

Encara que les dades representen l'observació desitjada, es important mirar la probabilitat de que aquest efecte s'hagi donat únicament per casualitat. Amb una p=0,001 de que l'efecte es doni per casualitat, la probabilitat de que el canvi en les dades sigui degut a una intervenció en l'observació 71 es del 99,8996%.

Cal recordar que totes les inferències anteriors depenen críticament de que les suposicions esmentades anteriorment es compleixin.

Un exemple pràctic de d'utilització d'aquest paquet seria el mencionat en l'article *Estimating Causal Effects on Financial Time-Sèries with Causal Impact BSTS* (Price 2021), sobre el càlcul de l'efecte causal del col·lapse de la presa Vale a Brasil, sobre el preu del mineral de ferro.

En el cas de les dades estudiades, es buscarà quantificar l'efecte de diferents mesures implementades pels governs, en el nombre d'infectats per el virus de la Covid-19 en els respectius països considerats en aquest treball. Per poder realitzar aquest anàlisi, s'han de definir prèviament els diferents protocols de contenció que es van aplicar a cada país per veure'n l'efectivitat que van tenir.

Un cop identificats els diferents escenaris, s'utilitza el paquet d'R per calcular-ne l'efecte causal que han tingut sobre el nombre d'infectats. Cal tenir en compte que es segueix fent ús de les sèries originals estandarditzades i que, donada la naturalesa del virus, s'ha estudiat l'efecte contrafactual 14 dies després de que fossin efectives les mesures. Aquest retard és degut al període d'espera entre que una persona es contagia i és identificada com a infectada pel virus, donat que aquest impàs no és instantani. Additivament, el període post-intervenció comprèn únicament les 4 setmanes posteriors a la implementació de la mesura (28 dies), per intentar minimitzar els efectes d'escenaris aliens a la pròpia mesura estudiada en cada cas.

2. Mesures de confinament

Donada una pandèmia que va aparèixer de forma casi simultània a tot al mon, es desconeixia totalment el seu efecte, les conseqüències que podia portar i la manera de fer front a una situació tant excepcional. Degut a la gran incertesa que presentava la situació, cada país va adaptar diverses mesures de contenció davant l'epidèmia, però van ser igual d'eficients totes elles? A continuació s'exposaran les mesures més destacables que van realitzar els diversos països tractats durant l'estudi per definir els diversos esdeveniments claus del confinament i analitzar-ne l'efecte causal davant el nombre d'infectats registrats.

Nova Zelanda:

El primer cas de covir-19 reportat al país va ser el 28 de febrer, però fins el 14 de març no va iniciar el protocol de contenció amb una quarantena obligatòria per tota la gent que entrés al país. Aquesta mesura de prevenció davant la infecció pel virus es va intensificar el 19 del mateix mes, prohibint l'entrada a tots els no residents.

El següent gran escenari que es va implementar va ser el tancament d'escoles i fronteres, sumat a un confinament estricte de tota la població (23 de març) que va concloure amb el bloqueig del país a nivell nacional el dia 25 de maig. Aquest es va mantenir fins finals d'abril, on es va permetre el retorn dels residents si mantenien un aïllament de dues setmanes als seus domicilis.

La des-escalada es va mantenir progressivament al llarg de maig fins que el dia 8 de juny es va donar per erradicat el virus dins el país(Ara (Diari online) 2020).

Puntualment, el 12 d'agost es va tornar a reportar l'aparició de nous casos a la ciutat d'Auckland, intensificant novament les restriccions preventives.

Finalment, al setembre es va retornar novament a un període de "normalitat".

Vistes les mesures de confinament prèviament esmentades, es creu interessant estudiar l'efecte que s'hauria produït en el nombre d'infectats, en cas de no haver implementat els següents escenaris:

- 19 de març – Prohibició de l'entrada al país dels no residents.
- 25 de maig – Bloqueig del país a nivell nacional
- 12 d'agost – Confinament de la ciutat d'Auckland

Segons l'estudi realitzat per la revista The Lancet (The Lancet 2020), les fortes mesures preventives instaurades al país han contribuït a una disminució de la mortalitat.

Mèxic:

Les etapes de la pandèmia a Mèxic es componen segons tres fases, la ultima de les quals segueix vigent a l'actualitat, donat el gran nombre de contagis que encara presenta el país.

La primera fase, del 28 de febrer al 23 de març, es la més laxa de totes. Aquí les mesures restrictives són inexistents, donat que es creia que tots els casos d'infectats provenien de persones de fora del país i no hi havia contagis interns. Del 14 al 18 d'abril es va implementar la distància social en espais públics i l'extensió dels períodes de vacances de Setmana Santa, però no va ser fins el 22 que es van tancar esglésies, teatres, museu, ... amb la promesa d'una reobertura no molt llunyana.

La segona fase inicia al finalitzar la primera, quan es registren els primers contagis dins el país, i s'estén fins el 20 d'abril. Donada una prèvia recomanació de confinament

domiciliari, es van començar a suspendre les activitats no essencials el 26 de març, finalitzant amb la declaració de l'estat d'emergència sanitària el dia 30. El 8 d'abril, es va habilitar el centre mèdic naval per casos greus de Covid-19 a Ciudad de México.

La tercera fase, que segueix vigent actualment (17/4/2021), va iniciar el 21 d'abril amb la suspensió definitiva de qualsevol activitat no essencial i d'implementació de protocols de confinaments més estrictes.

Malgrat totes les mesures implementades, el govern del país no ha aconseguit fer front de manera efectiva al virus i ha estat molt qüestionat per la seva falta d'accions i la poca quantitat de proves de detecció realitzades, en comparació a altres estats.

Vistes les mesures de confinament prèviament mencionades, es creu interessant estudiar l'efecte que s'hauria produït en el nombre d'infectats, en cas de no haver implementat els següents escenaris:

- 22 de març – Tancament d'espais públics interiors (Teatres, museu, ...)
- 30 de març – Declaració de l'estat d'emergència sanitària
- 8 d'abril – Obertura del centre mèdic naval per casos greus de Covid-19
- 21 d'abril – Suspensió de les activitats no essencials i mesures estrictes de confinament

Xina:

Xina va ser el país d'origen del contagi. El focus de la infecció es va detectar en el mercat de majoristes de Wuhan a finals del 2019 i va ser el primer país en trobar-se obligat a implementar mesures preventives sense tenir coneixements de a que s'enfrontava.

Amb el previ aïllament del focus de la infecció, el 23 de gener ja va declarar el confinament estricte de la població, juntament amb el tancament d'escoles i la posada en quarantena de la província de Hubei, on s'havien detectat els primers casos d'infectats.

Posteriorment es van seguir confinant més ciutats, i per fer front al gran nombre d'infectats, el 3 de febrer es va inaugurar l'hospital d'emergència de Wuhan.

Les mesures preventives es van mantenir a tot el país, i es va instaurar un confinament domiciliari voluntari de dues setmanes a tots els residents que tornaven al país el 12 de febrer, després de les vacances d'any nou.

Finalment, el 23 de març es va tornar a obrir Wuhan i la des escalada va durar dins el 8 d'abril, on es va declarar el final del confinament.

Malgrat ser el focus de la infecció, Xina destaca en la seva velocitat i efectivitat en prendre mesures per fer front a la pandèmia, sent un país referent en mesures preventives davant la Covid-19.

Vistes les mesures de confinament prèviament mencionades, es creu interessant estudiar l'efecte que s'hauria produït en el nombre d'infectats, de no implementar els següents escenaris:

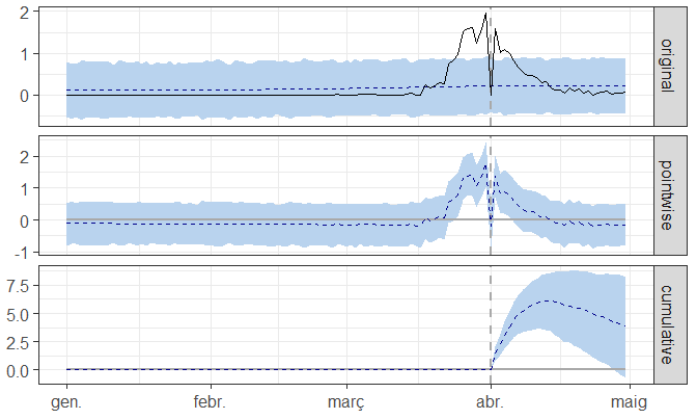
- 23 de gener – Confinament de la població, tancament d'escoles i quarantena de Hubei
- 3 de febrer – Inauguració de l'hospital d'emergència de Wuhan

3. Resultats

- **<u>Efectes del confinament a Nova Zelanda</u>**

Escenari 1: Prohibició de l'entrada al país dels no residents

Aplicant l'anàlisi contrafactual del que hauria passat si no s'hagués aplicat la restricció d'entrada al país pels no residents de Nova Zelanda (Figura IV.3), s'observa el següent:



Posterior inference {CausalImpact}

	Average	Cumulative
Actual	0.34	10.00
Prediction (s.d.)	0.21 (0.08)	6.11 (2.31)
95% CI	[0.06, 0.37]	[1.74, 10.74]
Absolute effect (s.d.)	0.13 (0.08)	3.89 (2.31)
95% CI	[-0.026, 0.28]	[-0.745, 8.25]
Relative effect (s.d.)	64% (38%)	64% (38%)
95% CI	[-12%, 135%]	[-12%, 135%]

Posterior tail-area probability p: 0.04728
Posterior prob. of a causal effect: 95.272%

Figura IV.3: Resultats de l'anàlisi contrafactual de l'escenari 1 a Nova Zelanda

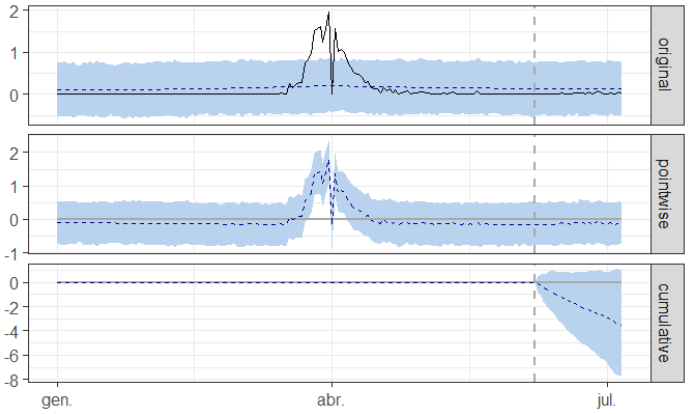
Segons l'anàlisi, sinó s'hagués donat la intervenció, el nombre mitjà d'infectats per 100.000 habitants el següent mes seria de 0,21, de manera que s'ha observat un augment del 64% en el nombre d'infectats. D'aquesta manera semblaria que la intervenció ha causat un augment en el nombre d'infectats per la Covid-19, donant a entendre que la mesura preventiva no hauria estat prou efectiva per frenar l'augment exponencial de casos detectats. Igualment, donat que l'efecte contrafactual ha resultat en una disminució dels casos, es podria pensar que l'escenari que ha detectat el model sigui algun dels brots de contagi del virus.

Donada una probabilitat de 0,04727 de que l'efecte observat sigui donat per casualitat, es creuria que els resultats obtinguts són estadísticament significatius amb un nivell de significació del 0,1.

Cal remarcar que hi ha una observació estranya (el pic inferior) immediatament prèvia a l'anàlisi, la qual podria haver influït significativament en els resultats. Si no es té en compte, la intervenció passa a ser no significativa estadísticament.

Escenari 2: Bloqueig del país a nivell nacional

Aplicant l'anàlisi contrafactual del que hauria passat si no s'hagués aplicat el bloqueig a nivell Nacional a Nova Zelanda (Figura IV.4), s'observa el següent:




```
Posterior inference {CausalImpact}

                         Average          Cumulative
Actual                   0.021            0.601
Prediction (s.d.)        0.14 (0.076)     4.16 (2.218)
95% CI                   [-0.012, 0.29]   [-0.356, 8.35]

Absolute effect (s.d.)   -0.12 (0.076)    -3.56 (2.218)
95% CI                   [-0.27, 0.033]   [-7.75, 0.958]

Relative effect (s.d.)   -86% (53%)       -86% (53%)
95% CI                   [-186%, 23%]     [-186%, 23%]

Posterior tail-area probability p:   0.06963
Posterior prob. of a causal effect:  93%
```

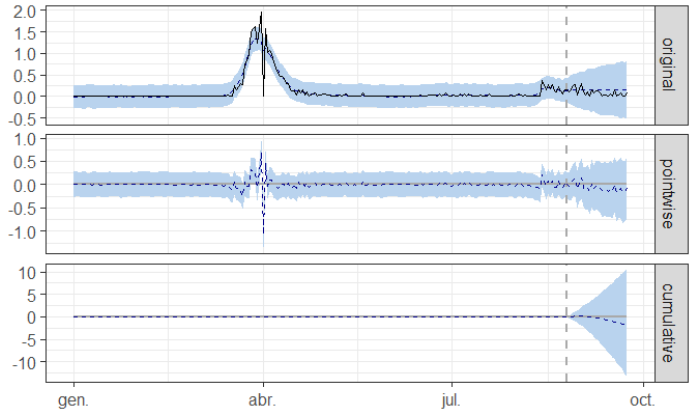
Figura IV.4: Resultats de l'anàlisi contrafactual de l'escenari 2 a Nova Zelanda

Segons l'anàlisi, si no s'hagués donat la intervenció, el nombre mitjà d'infectats per 100.000 habitants el següent mes seria de 0,14, de manera que s'ha observat una disminució del 86% en el nombre d'infectats. D'aquesta manera semblaria que la intervenció ha causat una disminució considerable en el nombre d'infectats per la Covid-19, donant a entendre que la mesura preventiva hauria estat efectiva. Amb aquesta intervenció s'haurien evitat un total de 3,56 infectats per cada 100.000 habitants el següent mes.

Donada una probabilitat de 0,06963 de que l'efecte observat sigui donat per casualitat, es podria considerar que els resultats obtinguts són estadísticament significatius amb un nivell de significació del 0,1.

Escenari 3: Confinament de la ciutat d'Auckland

Aplicant l'anàlisi contrafactual del que hauria passat si no s'hagués aplicat el confinament a la ciutat d'Aucklad a Nova Zelanda (Figura IV.5), s'observa el següent:




```
Posterior inference {CausalImpact}

                         Average          Cumulative
Actual                   0.092            2.675
Prediction (s.d.)        0.15 (0.2)       4.47 (5.9)
95% CI                   [-0.28, 0.55]    [-8.00, 15.92]

Absolute effect (s.d.)   -0.062 (0.2)     -1.797 (5.9)
95% CI                   [-0.46, 0.37]    [-13.25, 10.68]

Relative effect (s.d.)   -40% (133%)      -40% (133%)
95% CI                   [-296%, 239%]    [-296%, 239%]

Posterior tail-area probability p:   0.37963
Posterior prob. of a causal effect:  62%
```

Figura IV.5: Resultats de l'anàlisi contrafactual de l'escenari 3 a Nova Zelanda

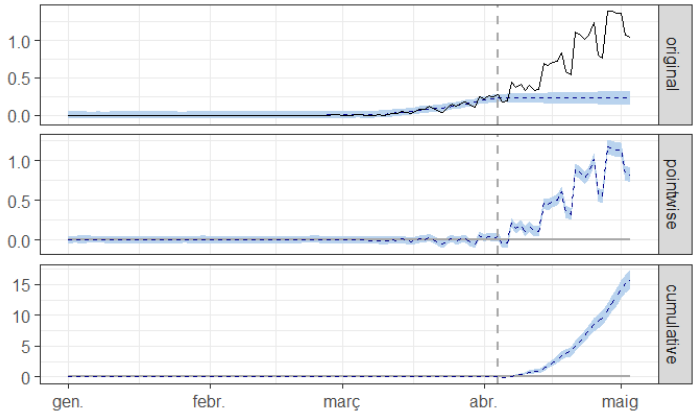
Segons l'anàlisi, si no s'hagués donat la intervenció, el nombre mitjà d'infectats per 100.000 habitants el següent mes seria de 0,15, de manera que s'ha observat una disminució del 40% en el nombre d'infectats. D'aquesta manera semblaria que la intervenció ha causat una disminució en el nombre d'infectats per la Covid-19, donant a entendre que la mesura preventiva hauria estat efectiva. Amb aquesta intervenció s'haurien evitat un total de 1,8 infectats per cada 100.000 habitants el següent mes.

Malgrat que l'efecte observat seria el desitjat, donada una probabilitat de 0,37963 de que l'efecte observat sigui donat per casualitat, no es podria considerar que els resultats obtinguts són estadísticament significatius amb un nivell de significació del 0,1.

- **Efectes del confinament a Mèxic**

Escenari 1: Tancament d'espais públics interiors (Teatres, museu, ...)

Aplicant l'anàlisi contrafactual del que hauria passat si no s'hagués aplicat el tancament d'espais públics interiors a Mèxic (Figura IV.6), s'observa el següent:



Posterior inference {CausalImpact}

	Average	Cumulative
Actual	0.78	22.48
Prediction (s.d.)	0.23 (0.028)	6.59 (0.798)
95% CI	[0.18, 0.28]	[5.08, 8.20]
Absolute effect (s.d.)	0.55 (0.028)	15.89 (0.798)
95% CI	[0.49, 0.6]	[14.28, 17.4]
Relative effect (s.d.)	241% (12%)	241% (12%)
95% CI	[217%, 264%]	[217%, 264%]

Posterior tail-area probability p: 0.00116
Posterior prob. of a causal effect: 99.88426%

Figura IV.6: Resultats de l'anàlisi contrafactual de l'escenari 1 a Mèxic

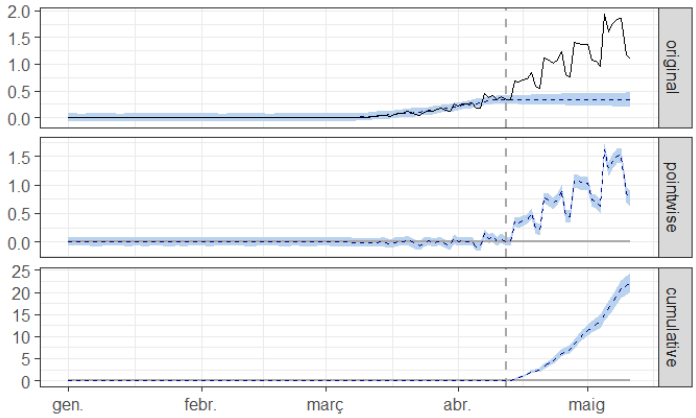
Segons l'anàlisi, si no s'hagués donat la intervenció, el nombre mitjà d'infectats per 100.000 habitants el següent mes seria de 0,23, de manera que s'ha observat un augment del 241% en el nombre d'infectats. D'aquesta manera semblaria que la intervenció ha causat un augment considerable en el nombre d'infectats per la Covid-19, donant a entendre que la mesura preventiva no hauria estat efectiva.

Aquest efecte és podria atribuir a la mala praxis en el sistema de prevenció i contenció de la pandèmia efectuat a Mèxic, de manera que aniria acord amb els fets de que les mesures realitzades no han aconseguit controlar l'expansió del virus en el període estudiat.

Donada una probabilitat de 0,00116 de que l'efecte observat sigui donat per casualitat, es podria considerar que els resultats obtinguts són estadísticament significatius amb un nivell de significació del 0,1.

Escenari 2: Declaració de l'estat d'emergència sanitària

Aplicant l'anàlisi contrafactual del que hauria passat si no s'hagués declarat l'estat d'emergència sanitària a Mèxic (Figura IV.7), s'observa el següent:




```
Posterior inference {CausalImpact}

                               Average        Cumulative
Actual                         1.1            32.1
Prediction (s.d.)              0.33 (0.038)   9.70 (1.112)
95% CI                         [0.26, 0.41]   [7.59, 11.84]

Absolute effect (s.d.)         0.77 (0.038)   22.37 (1.112)
95% CI                         [0.7, 0.84]    [20.2, 24.48]

Relative effect (s.d.)         231% (11%)     231% (11%)
95% CI                         [208%, 252%]   [208%, 252%]

Posterior tail-area probability p:   0.00109
Posterior prob. of a causal effect:  99.89142%
```

Figura IV.7: Resultats de l'anàlisi contrafactual de l'escenari 2 a Mèxic

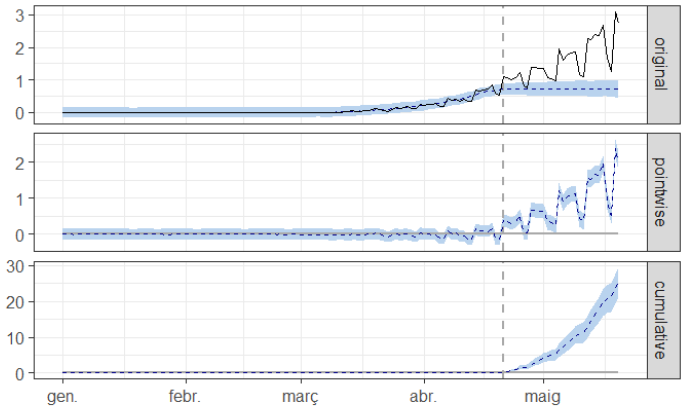
Segons l'anàlisi, si no s'hagués donat la intervenció, el nombre mitjà d'infectats per 100.000 habitants el següent mes seria de 0,33, de manera que s'ha observat un augment del 231% en el nombre d'infectats. D'aquesta manera semblaria que la intervenció ha causat un augment considerable en el nombre d'infectats per la Covid-19, donant a entendre que la mesura preventiva no hauria estat efectiva.

Reafirmant el que s'ha vist en l'escenari 1, aquest efecte és podria atribuir a la mala praxis en el sistema de prevenció i contenció de la pandèmia efectuat a Mèxic, de manera que aniria acord amb els fets de que les mesures realitzades no han aconseguit controlar l'expansió del virus en el període estudiat.

Donada una probabilitat de 0,00109 de que l'efecte observat sigui donat per casualitat, es podria considerar que els resultats obtinguts són estadísticament significatius amb un nivell de significació del 0,1.

Escenari 3: Obertura del centre mèdic naval per casos greus de Covid-19

Aplicant l'anàlisi contrafactual del que hauria passat si no s'hagués habilitat el centre mèdic naval per casos greus de Covid-19 a Mèxic (Figura IV.8), s'observa el següent:




```
Posterior inference {CausalImpact}

                               Average          Cumulative
Actual                         1.6              46.6
Prediction (s.d.)              0.72 (0.077)     20.98 (2.222)
95% CI                         [0.58, 0.88]     [16.85, 25.49]

Absolute effect (s.d.)         0.88 (0.077)     25.63 (2.222)
95% CI                         [0.73, 1]        [21.12, 30]

Relative effect (s.d.)         122% (11%)       122% (11%)
95% CI                         [101%, 142%]     [101%, 142%]

Posterior tail-area probability p:   0.00105
Posterior prob. of a causal effect:  99.89474%
```

Figura IV.8: Resultats de l'anàlisi contrafactual de l'escenari 3 a Mèxic

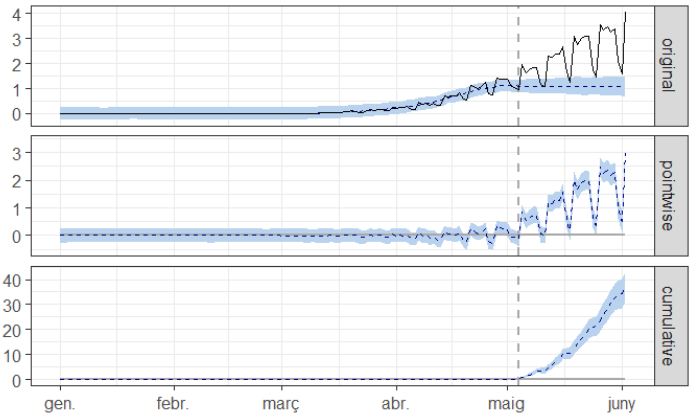
Segons l'anàlisi, si no s'hagués donat la intervenció, el nombre mitjà d'infectats per 100.000 habitants el següent mes seria de 0,72, de manera que s'ha observat un augment del 122% en el nombre d'infectats. D'aquesta manera semblaria que la intervenció ha causat un augment considerable en el nombre d'infectats per la Covid-19, donant a entendre que la mesura preventiva no hauria estat efectiva.

Els resultats segueixen el mateix comportament que els dos escenaris anteriors.

Donada una probabilitat de 0,00105 de que l'efecte observat sigui donat per casualitat, es podria considerar que els resultats obtinguts són estadísticament significatius amb un nivell de significació del 0,1.

Escenari 4: Suspensió de les activitats no essencials i mesures estrictes de confinament

Aplicant l'anàlisi contrafactual del que hauria passat si no s'haguessin suspès les activitats no essencials a Mèxic (Figura IV.9), s'observa el següent:




```
Posterior inference {CausalImpact}

                               Average          Cumulative
Actual                         2.4              69.2
Prediction (s.d.)              1.1 (0.11)       31.9 (3.07)
95% CI                         [0.89, 1.3]      [25.92, 37.8]

Absolute effect (s.d.)         1.3 (0.11)       37.2 (3.07)
95% CI                         [1.1, 1.5]       [31.4, 43.3]

Relative effect (s.d.)         117% (9.6%)      117% (9.6%)
95% CI                         [98%, 135%]      [98%, 135%]

Posterior tail-area probability p:   0.00108
Posterior prob. of a causal effect:  99.89247%
```

Figura IV.9: Resultats de l'anàlisi contrafactual de l'escenari 4 a Mèxic

Segons l'anàlisi, si no s'hagués donat la intervenció, el nombre mitjà d'infectats per 100.000 habitants el següent mes seria de 1,1, de manera que s'ha observat un augment del 117% en el nombre d'infectats. D'aquesta manera semblaria que la intervenció ha

causat un augment considerable en el nombre d'infectats per la Covid-19, donant a entendre que la mesura preventiva no hauria estat efectiva.

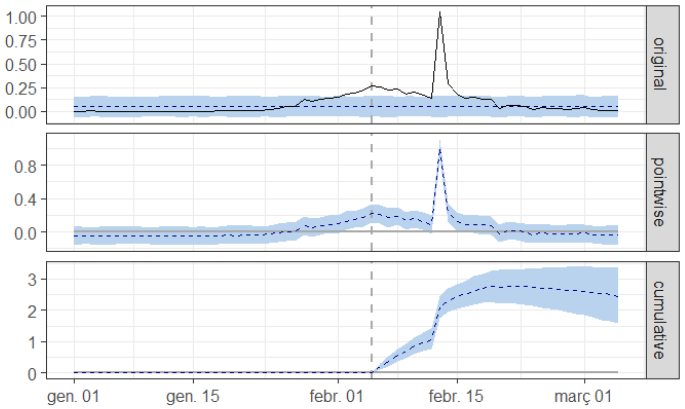
Els resultats segueixen el mateix comportament que els tres escenaris anteriors.

Donada una probabilitat de 0,00108 de que l'efecte observat sigui donat per casualitat, es podria considerar que els resultats obtinguts són estadísticament significatius amb un nivell de significació del 0,1.

- **Efectes del confinament a Xina**

Escenari 1: Confinament de la població, tancament d'escoles i quarantena de la província de Hubei

Aplicant l'anàlisi contrafactual del que hauria passat si no s'hagués confinat la població a Xina (Figura IV.10), s'observa el següent:




```
Posterior inference {CausalImpact}

                               Average          Cumulative
Actual                         0.13             3.90
Prediction (s.d.)              0.05 (0.015)     1.45 (0.436)
95% CI                         [0.019, 0.08]    [0.543, 2.33]

Absolute effect (s.d.)         0.085 (0.015)    2.453 (0.436)
95% CI                         [0.054, 0.12]    [1.575, 3.36]

Relative effect (s.d.)         169% (30%)       169% (30%)
95% CI                         [109%, 232%]     [109%, 232%]

Posterior tail-area probability p:   0.00101
Posterior prob. of a causal effect:  99.8994%
```

Figura IV.10: Resultats de l'anàlisi contrafactual de l'escenari 1 a Xina

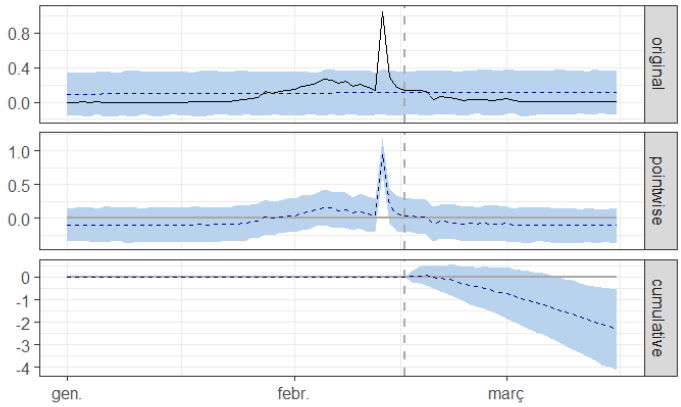
Segons l'anàlisi, si no s'hagués donat la intervenció, el nombre mitjà d'infectats per 100.000 habitants el següent mes seria de 0,05, de manera que s'ha observat un augment del 169% en el nombre d'infectats. D'aquesta manera semblaria que la intervenció ha causat un augment en el nombre d'infectats.

En aquest cas, donada la ràpida intervenció del govern xines hi ha un històric molt petit de les dades pre-intervenció, de manera que l'anàlisi contrafactual realitzat podria estar detectant encara l'aparició del virus, donada la falta de dades en temps de pandèmia, prèvies al confinament efectuat.

Donada una probabilitat de 0,00101 de que l'efecte observat sigui donat per casualitat, es creuria que els resultats obtinguts són estadísticament significatius amb un nivell de significació del 0,1. Si es pren com a escenari l'aparició del virus, l'efecte quedaria explicat.

Escenari 2: Inauguració de l'hospital d'emergència de Wuhan

Aplicant l'anàlisi contrafactual del que hauria passat si no s'hagués inaugurat l'Hospital d'emergència de Wuhan a Xina (Figura IV.11), s'observa el següent:




```
Posterior inference {CausalImpact}

                               Average          Cumulative
Actual                         0.03             0.87
Prediction (s.d.)              0.11 (0.032)     3.16 (0.924)
95% CI                         [0.049, 0.17]    [1.432, 5.00]

Absolute effect (s.d.)         -0.079 (0.032)   -2.296 (0.924)
95% CI                         [-0.14, -0.02]   [-4.14, -0.57]

Relative effect (s.d.)         -73% (29%)       -73% (29%)
95% CI                         [-131%, -18%]    [-131%, -18%]

Posterior tail-area probability p:   0.00702
Posterior prob. of a causal effect:  99.29789%
```

Figura IV.11: Resultats de l'anàlisi contrafactual de l'escenari 2 a Xina

Segons l'anàlisi, si no s'hagués donat la intervenció, el nombre mitjà d'infectats per 100.000 habitants el següent mes seria de 0,11, de manera que s'ha observat una disminució del 73% en el nombre d'infectats. D'aquesta manera semblaria que la intervenció ha causat una disminució considerable en el nombre d'infectats per la Covid-19, donant a entendre que la mesura preventiva hauria estat efectiva. Amb aquesta intervenció s'haurien evitat un total de 2,3 infectats per cada 100.000 habitants el següent mes.

Donada una probabilitat de 0,00702 de que l'efecte observat sigui donat per casualitat, es podria considerar que els resultats obtinguts són estadísticament significatius amb un nivell de significació del 0,1. Aquesta seria la intervenció estudiada amb un efecte més clar i en línia amb els resultats esperats.

V. CONCLUSIONS

El cos del treball es movia al voltant de la hipòtesi sobre que *l'evolució de la pandèmia mundial provocada pel virus de la COVID-19 ha distat significativament entre els diferents països, degut a les aplicacions personalitzades de les mesures de control i prevenció efectuades per cada govern en qüestió*. Vista la tipologia de les dades, les quals presenten un grau d'incertesa i volatilitat molt elevat, s'han presentat diverses metodologies per modelitzar-ne la seva distribució i intentar preveure'n valors futurs. Això permetia comparar les metodologies clàssiques utilitzades per referencia amb la metodologia bayesiana, la qual, amb les dades estudiades, permetia estimar d'una forma més realista les tendències sense subestimar la incertesa i volatilitat de les dades.

Es important recordar que l'estudi s'ha fet a partir de les dades obtingudes d'una infecció vírica molt contagiosa i impossible de detectar o identificar a simple vista, abans de la manifestació dels símptomes posteriorment descoberts. Additivament, la naturalesa d'aquesta malaltia dificulta la identificació del nombre d'infectats, donat la existència de les persones asimptomàtiques. Aquests individus no apareixen registrats en els recomptes de casos però tenen un efecte igual d'important en l'evolució de la pandèmia. Altrament també s'ha de tenir en compte el lapsus de temps entre el contagi i la manifestació de la malaltia.

Des del punt de vista estocàstica s'ha observat que encara que els models ARIMA permeten modelar correctament les sèries del nombre d'infectats per la Covid-19, no ens donen informació sobre la història d'aquesta, ja que les estimacions resultants són valors concrets. Aquests models resulten en estimacions puntuals, que encara que també tenen associades un component d'incertesa, no es tant ajustada com les distribucions de probabilitat resultats dels models *bsts*.

Amb els resultats obtinguts es pot afirmar que aqueta metodologia permet ajustar correctament les dades mitjançant una combinació de transformacions i mitjanes mòbils. Curiosament, la naturalesa de la malaltia infecciosa faria pensar sobre la necessitat de components autoregressives que expliquin el nombre d'infectats com una evolució del mateix nombre el dia anterior o els k dies anteriors, però aquesta component no ha aparegut en cap de les sèries temporals tractades. Es creu que la volatilitat i estacionalitat de les dades han amagat aquesta component causant que els models no la detectessin.

Per contraposició també s'ha analitzat un enfocament bayesià, a partir dels models *bsts*. El model bayesià no es basa en diferències, retard i mitjanes mòbils, sinó en gestionar i quantificar la incertesa segons la interpretació visual, permetent entendre realment el funcionament del model i el comportament de la sèrie. Els models *bsts* utilitzen un enfocament probabilístic per modelar un problema de sèries temporals, és a dir, retornen una distribució predictiva posterior sobre la qual podem mostrejar per proporcionar no només una previsió, sinó també un mitjà per quantificar la incertesa del

model directament a partir de les variables aleatòries que retorna. Addicionalment, aquesta anàlisi també permet afegir altres coneixements a priori sobre les dades, finalitzant en un conjunt de variables aleatòries distribuïdes de forma més precisa o exacta sobre la sèrie real.

Encara que no hi ha cap model que s'imposi clarament sobre l'altre en la modelització de les sèries considerades en aquest treball, si que s'ha pogut detectar com segons les components de la sèrie s'obtenien ajustos diferents, tant en el període mostral com pel que fa a les previsions. Ambdós metodologies s'intenten adequar als valors reals, amb petits punts distintius. Les previsions dels models ARIMA són més constants i tenen menys variabilitat, mentre que les dels models *bsts* intenten reproduir les tendències reals, proporcionant estimacions mes realistes de la incertesa, a la vegada que més volàtils.

Finalment s'ha afegit un anàlisi causal per contrastar les diferències observades amb les sèries dels tres països analitzats i les mesures preventives i de confinament que s'havien aplicat en cadascun d'ells. Aquí, com era d'esperar, s'ha tornat a veure com la ràpida intervenció dels governs Xinesos i Neozelandesos han tingut un paper clau en l'evolució i la desescalada de les sèries d'infectats per 100.000 habitants. Contràriament, com era d'esperar, en les intervencions de Mèxic no s'ha vist el mateix efecte. Aquí les mesures restrictives han estat més laxes i menys agressives, fet que ha causat que l'efecte preventiu no hagi resultat suficient per reduir el creixement de la infecció. A Mèxic no s'ha aconseguit controlar la malaltia en el període estudiat.

3. Conclusions

Al inici del treball s'havia exposat la intenció de treballar i aprofundir en l'anàlisi de sèries temporals. Per assolir aquest propòsit s'havia escollit un tema d'actualitat com és la Covid-19, la qual presenta unes dades difícils de treballar i que poden donar resultats controvertibles i s'havia definit com a hipòtesis la creença de que l'evolució d'aquesta sèrie varia segons cada país i les diverses mesures preventives que cada un ha aplicat. D'aquesta manera també s'havien definit una objectius acordes, els quals s'han anat assolint amb èxit.

A partir de l'ajust dels models clàssics i bayesians per a les sèries del nombre d'infectats per la Covid-19 per 100.000 habitants s'han distingit ràpidament dos patrons diferents, acords amb la prèvia classificació que s'havia fet sobre els millors i pitjors països on passar la pandèmia:

En els casos de Nova Zelanda i Xina, països pioners en la lluita contra la pandèmia, s'han obtingut resultats molt similars. La rapidesa i serietat en la resposta a l'imminent infecció ha resultat en una contenció exitosa del viris. Ambdues sèries, petites diferències apart, es composaven dels mateixos component, tendència i falta d'estacionalitat. En l'anàlisi contrafactual, el qual mesurava l'efecte de les mesures del confinament de cada país, també s'ha reafirmat que en ambdós casos aquestes havien sigut efectives i estadísticament significatives, minvant així l'augment de casos i permetent assolit una desescalada exitosa.

El cas contrari seria Mèxic. Aquí ja des del principi la sèrie tenia una forma molt estranya, amb una forta estacionalitat setmanal que no s'observa en cap dels altres casos. El govern mexicà ha despertat fortes polèmiques sobre possibles falsejaments en els recomptes d'infectats i morts detectats i la seva actuació en la prevenció i rastreig de nous brots d'infecció. Aquestes discussions, entre d'altres, portarien a explicar aquest comportament de la sèrie temporal tant contrari als altres dos estudiats. Encara que en els anàlisis de les previsions han donat uns percentatges d'ajustos molt similars a les altres sèries, en l'estudi contrafactual s'ha reafirmat com les mesures preventives que es van prendre no van tenir aquest efecte minvant en l'augment del nombre d'infectats per la pandèmia.

Addicionalment, també s'ha pogut treballar les diverses maneres proposades d'estimar i estudiar les sèries temporals, observant-ne així els seus funcionaments i les aportacions en les futures previsions de les sèries temporals. Per una banda, les sèries analitzades des del punt de vista estocàstica presentaven molts bons ajustos com més constant fos la sèrie original, encara que, contraries al comportament esperat de les dades, no presentaven cap component autoregressiva. Aquest tipus de models tampoc aportaven informació sobre l'historia de la sèrie o la seva distribució. Per l'altra banda, l'enfocament bayesià sí que permet entendre el comportament de la sèrie i atribuir una

distribució a les dades, encara que presenta unes previsions molts més volàtils. En la sèrie de Xina, la qual presentava un pic inicial però la resta d'observacions eren constants al voltant del zero, aquestes fortes variacions de les previsions no fen justícia al comportament real de la sèrie.

En concussió, s'ha observat com l'evolució de la mateixa sèrie temporal ha distat significativament segons el país i s'ha vist que un dels motius d'aquestes diferències ha estat la velocitat i el rigor de l'aplicació de mesures preventives i de confinament. A més a més s'ha pogut treballar diverses metodologies d'anàlisi de sèries temporals i entendre les diferències darrera els seus enfocaments i els resultats que proporcionen.

4. Bibliografia

Ara (Diari online). 2020. “Nova Zelanda, Primer País Que Dona per Eliminat El Covid-19.” https://www.ara.cat/internacional/nova-zelanda-primer-pais-Covid-19-coronavirus-eliminat_1_1129063.html (April 17, 2021).

BBC News Mundo. “Coronavirus: Los Mejores y Peores Países Donde Pasar La Pandemia - BBC News Mundo.” https://www.bbc.com/mundo/noticias-55048757 (March 1, 2021).

Bernardo, José M. 2002. 17 Societat Catalana de Matemàtiques *Una Introducció a l’estadística Bayesiana **.

BOX, GEORGE E. P., GWILYM M. JENKINS, GREGORY C. REINSEL, and GRETA M. LJUNG. 2016. *TIME SERIES ANALYSIS Forecasting and Control*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Mauricio, José Alberto. 2007. *Análisis de Series Temporales* _____ *SERIES TEMPORALES PÁGINA II OBSERVACIONES PRELIMINARES*.

Nishi, Norio. 2012. 43 Journal of the Yamashina Institute for Ornithology *The First Breeding Record of Blue Rockthrush in Yamanashi Prefecture*.

Patrícia Oliveres Luna. 2017. *UNA INTRODUCCIÓ A L’ESTADÍSTICA BAYESIANA*.

Peña Sánchez de Rivera, Daniel. 2005. Alianza Ed *Análisis de Series Temporales*. Alianza Ed. ed. Alianza Editorial. http://295045.fromeriverfriends.org.uk/descargar/295045/Analisis%2Bde%2Bseries%2Btemporales.pdf (February 25, 2021).

Price, Chris. 2021. “Estimating Causal Effects on Financial Time-Series with Causal Impact BSTS | by Chris Price | Towards Data Science.” *4 juny*. https://towardsdatascience.com/estimating-causal-effects-on-financial-time-series-with-causal-impact-bsts-85d3c403f4a0 (April 17, 2021).

Rodríguez, Francisco Javier Parra, and Juan Antonio Vicente Vírseda. 2019. *Análisis de Series Temporales*. https://bookdown.org/franciscoparrod/analisis_series/Analisis_Series.html (February 24, 2021).

Scott, Steven L., and Hal Varian. 2013. “Bayesian Variable Selection for Nowcasting Economic Time Series.” *Economics of Digitization.* (July 2012): 1–22.

Scott, Steven L., and Hal R. Varian. 2014. “Predicting the Present with Bayesian Structural Time Series.” *International Journal of Mathematical Modelling and Numerical Optimisation* 5(1–2): 4–23.

Solhjell, Ida Kjersem. 2009. *BAYESIAN FORECASTING AND DYNAMIC MODELS APPLIED TO STRAIN DATA FROM THE GÖTA RIVER BRIDGE*.

Steven Scott, Author L, and Maintainer L Steven Scott. 2020. “Bayesian Structural Time Series R.”

The Lancet. 2020. “Nova Zelanda Guanya El Pols Al Coronavirus i, Fins i Tot, Redueix La Mortalitat En 2020 - À Punt.” *2020*. https://www.apuntmedia.es/Covid-19/nova-zelanda-guanya-pols-coronavirus-i-tot-redueix-mortalitat-2020_1_1342933.html (April 17, 2021).

VALENCIA CARDENAS, MARISOL, JUAN CARLOS CORREA MORALES, and FRANCISCO DIAZ SERNA. 2015. “CLASSICAL AND BAYESIAN STATISTICAL METHODS FOR DEMAND

FORECASTING. A COMPARATIVE ANALYSIS."
https://www.researchgate.net/profile/Marisol-Valencia-Cardenas/publication/303894667_Metodos_estadisticos_clasicos_y_bayesianos_para_el_pronostico_de_demanda_Un_analisis_Comparativo/links/575ae8ad08aed884620d9586/Metodos-estadisticos-clasicos-y-bayesianos (February 25, 2021).

5. Annexos

Codi R:

```
---
title: "Models per a sèries temporals no estacionàries. L'exemple de la Covid-19"
author: "Marc Ríos Masferrer"
date: "15/2/2021"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```


```{r}
library(tseries)
library(seastests)
library(bsts)
library(CausalImpact)
```


Part pràctica

Dades:

```{r warning=FALSE}
library (openxlsx)
library(readxl)
setwd("C:/Users/Marc Ríos Masferrer/Desktop/TFG/BBDD")
d<-read_excel("C:/Users/Marc Ríos Masferrer/Desktop/TFG/BBDD/COVID 19.xlsx")

x<-d[d$countriesAndTerritories=="New_Zealand",]
x<-t(x[-350,5])
NZ<-ts(rev(x),start=c(2020,1,1), frequency=c(349))

x<-d[d$countriesAndTerritories=="Mexico",]
x<-t(x[-350,5])
Mex<-ts(rev(x),start=c(2020,1,1), frequency=c(349))

x<-d[d$countriesAndTerritories=="China",]
x<-t(x[-350,5])
Xin<-ts(rev(x),start=c(2020,1,1), frequency=c(349))
```


Gràfics:
```{r}
plot.ts(NZ,main="N° d'infectats per la covid-19 a Nova Zelanda el 2020",ylab="n° d'infectats",xlab="Gener-Desembre 2020")
plot.ts(Mex,main="N° d'infectats per la covid-19 a Mèxic el 2020",ylab="n° d'infectats",xlab="Gener-Desembre 2020")
plot.ts(Xin,main="N° d'infectats per la covid-19 a la Xina el 2020",ylab="n° d'infectats",xlab="Gener-Desembre 2020")
```


Nova Zelanda:

Dades:
```{r}
poblacio_NZ<- 4822232
NZ<-NZ*100000/poblacio_NZ
```
```

```
head(NZ)

p_mostral_NZ<-ts(NZ[1:335])
p_extramostral_NZ<-ts(NZ[336:349])
```

Gràfic infectats cada 100.000 habitants:

```{r}
plot.ts(NZ,main="N° d'infectats per 100.000 habitants a Nova Zelanda el 2020",ylab="n° d'infectats",xlab="Gener-Desembre 2020")
```

Identificació:

- Tendència?

```{r}
summary(lm(NZ~seq(1:length(NZ))))
```

- Component estacional?

```{r}
isSeasonal(NZ, freq=30)
isSeasonal(NZ, freq=12)
isSeasonal(NZ, freq=7)
```

- Diferències regulars:

```{r}
d_NZ<-diff(NZ)
par(mfrow=c(1,1))
plot.ts(d_NZ,main="Diferències regulars de la sèrie I",ylab="n° d'infectats",xlab="Gener-Desembre 2020")
```

Comprovació:

```{r}
#Té tendència?
summary(lm(d_NZ~seq(1:length(d_NZ))))
```

- Identificació dels components de la sèrie:

```{r}
par(mfrow=c(2,1))
acf(d_NZ,main="FAS - Sèrie I transformada",ylab="Autocorrelació simple",xlab="Lag",lag=120)
pacf(d_NZ,main="FAP - Sèrie I transformada",ylab="Autocorrelació parcial",xlab="Lag",lag=120)
```

Models ARIMA

Estimació:

```{r}
model_c_NZ<-arima(NZ,order = c(0,1,2), seasonal = list (order = c(0,0,0)))

model_c_NZ
```

Validació:

- Significació dels coeficients:

```{r}
2*pnorm(c(abs(model_c_NZ$coef)/sqrt(diag(model_c_NZ$var.coef))), mean=0, sd=1, lower.tail=FALSE)
```

```

- Normalitat dels residus
```{r}
library(tseries)
jarque.bera.test(model_c_NZ$residuals)
```

Comprovació a partir de l'histograma:
```{r}
hist(model_c_NZ$residuals)
```

- Residus soroll blanc:
```{r}
Box.test(model_c_NZ$residuals,type = c( "Ljung-Box"))
```

Prediccions:

```{r}
model_c_NZ<-arima(p_mostral_NZ,order = c(0,1,2), seasonal = list (order = c(0,0,0)))
```

- Valors predits:
```{r}
predic_p_extramostral_NZ<-predict(model_c_NZ,n.ahead=14)
predic_p_extramostral_NZ
```

- Gràfic període mostral:
```{r}
p_mostral_arima_NZ<-(as.numeric(-model_c_NZ[["residuals"]]+p_mostral_NZ))

ts.plot(p_mostral_arima_NZ,p_mostral_NZ,col=c("red","black"),ylab="n° d'infectats",xlab="Gener-Novembre 2020")
title("Valors observats contra valors predits (ARIMA)")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Gràfic període extramostral:
```{r}
par(mfrow=c(1,1))
ts.plot(p_extramostral_NZ,predic_p_extramostral_NZ$pred[1:14],col=1:2)
title("Valors observats vs predits")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Avaluació de la capacitat predictiva
```{r}
errors<-p_extramostral_NZ-predic_p_extramostral_NZ$pred[1:14]
ts.plot(errors,col=2)

epam_mostral_NZ_1<-sum(100*abs(p_mostral_NZ[-c(which(p_mostral_NZ==0))]-p_mostral_arima_NZ[-c(which(p_mostral_NZ==0))])/abs(p_mostral_NZ[-c(which(p_mostral_NZ==0))]))/length(p_mostral_NZ[-c(which(p_mostral_NZ==0))])
epam_mostral_NZ_1

epam_extramostral_NZ_1<-sum(100*abs(errors[-c(4,6,10,12,14)])/abs(p_extramostral_NZ[-c(4,6,10,12,14)]))/9  #El 9 és el tamany de la mostra
epam_extramostral_NZ_1
```

- Predicció a futur:
```

```{r}
ts.plot(p_mostral_NZ,predic_p_extramostral_NZ$pred,
col=c("black","red"),ylab="n° d'infectats",xlab="Gener-Desembre 2020")
title("Evolució dels infectats per 100.000 habitants a Nova Zelanda el 2020")
legend("topleft",legend = c("Valors originals", "Valors previstos"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Intervals de prediccions:
```{r}
inf = predic_p_extramostral_NZ$pred - 2*predic_p_extramostral_NZ$se
sup = predic_p_extramostral_NZ$pred + 2*predic_p_extramostral_NZ$se

ts.plot(p_mostral_NZ,predic_p_extramostral_NZ$pred, col=c(4,2))
lines(inf, col="blue", lty="dashed")
lines(sup, col="blue", lty="dashed")
title("Intervals de prediccions")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("blue","red"), lty=1, cex=0.8)
```

Models bayesians

```{r}
plot.ts(NZ,main="N° d'infectats per la covid-19 a Nova Zelanda el 2020",ylab="n°
d'infectats",xlab="Gener-Desembre 2020")
```

Estimació:

```{r}
model_components <- list()
model_components <- bsts::AddLocalLinearTrend(model_components, y = NZ)

model_b_NZ <- bsts(NZ, state.specification = model_components, niter = 2000)
```

```{r}
components <- list() #list buit
components <- bsts::AddAutoAr(components, y = NZ) # Afegir tendència lineal

model_b_lineal_NZ <- bsts(NZ, state.specification = components, niter = 2000)
```

```{r}
CompareBstsModels(list("AutoAr" = model_b_lineal_NZ,
                       "LocalLinearTrend" = model_b_NZ),
                  colors = c("black", "red"), main = "Error absolut acumulat
d'ambdós models")
```

```{r}
hist(as.numeric(-colMeans(model_b_NZ$one.step.prediction.errors)+NZ),xlab    =
"Valors de la sèrie",main = "Histograma de la distibució de la sèrie I segons
el model bayesià")

median(as.numeric(-colMeans(model_b_NZ$one.step.prediction.errors)+NZ))
sd(as.numeric(-colMeans(model_b_NZ$one.step.prediction.errors)+NZ))
```

Validació:

- Normalitat dels residus per test
```{r}
res_NZ<-residuals(model_b_NZ, mean.only = TRUE)

jarque.bera.test(res_NZ)
```

Comprovació a partir de l'histograma:
```{r}
hist(res_NZ)
```

- Normalitat dels residus per gràfica:
```{r}
res_NZ<-residuals(model_b_NZ)

qqdist(res_NZ)
title("Distribució dels residus del model bayesià per la sèrie I")
```

- Residus estacionaris?
```{r}
AcfDist(res_NZ)
title("Distribució a posteriori de l'autocorrelació dels residus del model bayesià per la sèrie I")
```

Predicció:

```{r}
model_b_NZ <- bsts(p_mostral_NZ, state.specification = model_components, niter = 2000)
```

- Valors predits:
```{r}
pred_b_NZ <- predict(model_b_NZ, horizon = 14)
predic_b_p_extramostral_NZ<-pred_b_NZ$mean
predic_b_p_extramostral_NZ
```

- Grafic periode mostral
```{r}
p_mostral_b_NZ<-as.numeric(-colMeans(model_b_NZ$one.step.prediction.errors)+p_mostral_NZ)
ts.plot(p_mostral_b_NZ,p_mostral_NZ,col=c("red","black"),ylab="n° d'infectats",xlab="Gener-Novembre 2020")
title("Valors observats contra valors predits (BSTS)")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Gràfic del període extramostral:
```{r}
par(mfrow=c(1,1))
ts.plot(p_extramostral_NZ,predic_b_p_extramostral_NZ,col=1:2)
title("Valors observats vs predits")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Avaluació de la capacitat predictiva
```{r}
errors<-p_extramostral_NZ-predic_b_p_extramostral_NZ

epam_mostral_NZ_2<-sum(100*abs(p_mostral_NZ[-c(which(p_mostral_NZ==0))]-p_mostral_b_NZ[-c(which(p_mostral_NZ==0))])/abs(p_mostral_NZ[-c(which(p_mostral_NZ==0))]))/length(p_mostral_NZ[-c(which(p_mostral_NZ==0))])
epam_mostral_NZ_2

epam_extramostral_NZ_2<-sum(100*abs(errors[-c(4,6,10,12,14)])/abs(p_extramostral_NZ[-c(4,6,10,12,14)]))/9
epam_extramostral_NZ_2
```

- Predicció a futur:
```{r}
predic_b_p_extramostral_NZ<-ts(predic_b_p_extramostral_NZ,start=336)

ts.plot(p_mostral_NZ,predic_b_p_extramostral_NZ,  col=c("black","red"),ylab="n°
d'infectats",xlab="Gener-Desembre 2020")
title("Evolució dels infectats per 100.000 habitants a Nova Zelanda el 2020")
legend("topleft",legend = c("Valors originals", "Valors previstos"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Gràfic de les prediccions:
```{r}
pred_b_NZ <- predict(model_b_NZ, horizon = 14)
plot(pred_b_NZ, plot.original = 335)
```

Comparació:

- Gràfic del període extramostral:
```{r}
par(mfrow=c(1,1))
ts.plot(as.numeric(p_extramostral_NZ),predic_b_p_extramostral_NZ,predic_p_extr
amostral_NZ$pred,col=1:3,ylab="n° d'infectats",xlab="Desembre 2020")
title("Valors observats contra valors predits (ARIMA i BSTS)")
legend("topleft",legend = c("Valor observat", "Valor predit bsts", "Valor predit
ARIMA"),
       col=c("black","red","green"), lty=1, cex=0.8)
```

Mèxic:

Dades:
```{r}
poblacio_Mex<- 128932753
Mex<-Mex*100000/poblacio_Mex

head(Mex)

p_mostral_Mex<-ts(Mex[1:335])
p_extramostral_Mex<-ts(Mex[336:349])
```

Gràfic infectats cada 100.000 habitants:
```{r}
plot.ts(Mex,main="N° d'infectats per 100.000 habitants a Mèxic el 2020",ylab="n°
d'infectats",xlab="Gener-Desembre 2020")
```

Identificació:

- Tendència?
```{r}
summary(lm(Mex~seq(1:length(Mex))))
```

- Component estacional?
```{r}
isSeasonal(Mex, freq=7)
```

- Diferències regulars:
```{r}
d_Mex<-diff(Mex)
par(mfrow=c(1,1))
plot.ts(d_Mex,main="Diferències regulars de la sèrie II",ylab="n°
d'infectats",xlab="Gener-Desembre 2020")
```

```

- Tendència?
```{r}
summary(lm(d_Mex~seq(1:length(d_Mex))))
```

- Aplicar arrel sexta a la sèrie per estabilitzar la variància:
```{r}
l_Mex<-sqrt(sqrt(sqrt(Mex)))

par(mfrow=c(1,1))
plot.ts(l_Mex,main="Arrel sexta de la sèrie II",ylab="n° d'infectats",xlab="Gener-Desembre 2020")

#library(MVT)
#homogeneity.test(studentFit(l_Mex))  #Comprovar que estadisticament és constant
```

- Afegim diferències regulars:
```{r}
dl_Mex<-diff(l_Mex)
par(mfrow=c(1,1))
plot.ts(dl_Mex,main="Diferències regulars de la sèrie II transformada",ylab="n° d'infectats",xlab="Gener-Desembre 2020")
```

- Afegim diferències estacionals:
```{r}
dd7l_Mex<-diff(dl_Mex,lag=7)
par(mfrow=c(1,1))
plot.ts(dd7l_Mex,main="Diferències estacionals i regulars de la sèrie II transformada",ylab="n° d'infectats",xlab="Gener-Desembre 2020")
```

- Identificació dels components de la sèrie:

- Part regular:
```{r}
par(mfrow=c(2,1))
acf(dl_Mex,main="FAS - Sèrie II transformada",ylab="Autocorrelació simple",xlab="Lag",lag=120)
pacf(dl_Mex,main="FAP - Sèrie II transformada",ylab="Autocorrelació parcial",xlab="Lag",lag=120)
```

- Part estacional:
```{r}
par(mfrow=c(2,1))
acf(dd7l_Mex,main="FAS - Sèrie II transformada amb diferències regulars",ylab="Autocorrelació simple",xlab="Lag",lag=120)
pacf(dd7l_Mex,main="FAP - Sèrie II transformada amb diferències regulars",ylab="Autocorrelació parcial",xlab="Lag",lag=120)
```

Models ARIMA

Estimació:

```{r}
model_c_Mex<-arima(l_Mex,order = c(0,1,1), seasonal = list (order = c(0,1,2),period=7))

model_c_Mex
```

Validació:
```

- Significació dels coeficients:
```{r}
2*pnorm(c(abs(model_c_Mex$coef)/sqrt(diag(model_c_Mex$var.coef))),      mean=0,
sd=1, lower.tail=FALSE)
```

- Normalitat dels residus
```{r}
library(tseries)
jarque.bera.test(model_c_Mex$residuals)
```

Comprovació a partir de l'histograma:
```{r}
hist(model_c_Mex$residuals)
```

- Residus soroll blanc:
```{r}
Box.test(model_c_Mex$residuals,type = c( "Ljung-Box"))
```

Prediccions:

```{r}
p_mostral_l_Mex<-ts(l_Mex[1:335])
p_extramostral_l_Mex<-ts(l_Mex[336:349])
```

```{r}
model_c_Mex<-arima(p_mostral_l_Mex,order = c(0,1,1), seasonal = list (order =
c(0,1,2),period=7))
```

- Valors predits:
```{r}
predic_p_extramostral_Mex<-predict(model_c_Mex,n.ahead=14)
predic_p_extramostral_Mex$pred<-predic_p_extramostral_Mex$pred^6

predic_p_extramostral_Mex$pred
```

- Gràfic període mostral:
```{r}
p_mostral_arima_Mex<-(as.numeric(-
model_c_Mex[["residuals"]]+p_mostral_l_Mex))^6

ts.plot(p_mostral_arima_Mex,p_mostral_Mex,col=c("red","black"),ylab="n°
d'infectats",xlab="Gener-Novembre 2020")
title("Valors observats contra valors predits (SARIMA)")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Gràfic període extramostral:
```{r}
par(mfrow=c(1,1))
ts.plot(as.numeric(p_extramostral_Mex),predic_p_extramostral_Mex$pred,col=1:2,
xlab="Període extramostral")
title("Valors observats vs predits")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Avaluació de la capacitat predictiva
```{r}
errors<-as.numeric(p_extramostral_Mex)-predic_p_extramostral_Mex$pred
ts.plot(errors,col=2)
```

```
epam_mostral_Mex_1<-sum(100*abs(p_mostral_Mex[-c(which(p_mostral_Mex==0))]-
p_mostral_arima_Mex[-c(which(p_mostral_Mex==0))])/abs(p_mostral_Mex[-
c(which(p_mostral_Mex==0))]))/length(p_mostral_Mex[-
c(which(p_mostral_Mex==0))])
epam_mostral_Mex_1

epam_extramostral_Mex_1<-
sum(100*abs(errors)/abs(as.numeric(p_extramostral_Mex)))/14
epam_extramostral_Mex_1
```

- Predicció a futur:
```{r}
ts.plot(p_mostral_Mex,predic_p_extramostral_Mex$pred,
col=c("black","red"),ylab="n° d'infectats",xlab="Gener-Desembre 2020")
title("Evolució dels infectats per 100.000 habitants a Mèxic el 2020")
legend("topleft",legend = c("Valors originals", "Valors previstos"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Intervals de prediccions:
```{r}
predic_p_extramostral_Mex<-predict(model_c_Mex,n.ahead=14)

inf = (predic_p_extramostral_Mex$pred - 2*predic_p_extramostral_Mex$se)^6
sup = (predic_p_extramostral_Mex$pred + 2*predic_p_extramostral_Mex$se)^6

ts.plot(p_mostral_Mex,(predic_p_extramostral_Mex$pred)^6, col=c(4,2))
lines(inf, col="blue", lty="dashed")
lines(sup, col="blue", lty="dashed")
title("Intervals de prediccions")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("blue","red"), lty=1, cex=0.8)
```

Models bayesians

```{r}
plot.ts(l_Mex,main="Arrel sexta de la sèrie II",ylab="n°
d'infectats",xlab="Gener-Desembre 2020")
```

Estimació:

```{r}
model_components_Mex <- list()
model_components_Mex <- bsts::AddLocalLinearTrend(model_components_Mex, y =
l_Mex)
model_components_Mex <- bsts::AddSeasonal(model_components_Mex, y = l_Mex,
nseasons = 50, season.duration = 7)

model_b_Mex <- bsts(l_Mex, state.specification = model_components_Mex, niter =
2000)
```

```{r}
components <- list() #list buit
components <- bsts::AddAutoAr(components, y = l_Mex) # Afegir tendència lineal
components <- bsts::AddSeasonal(components, y = l_Mex, nseasons = 50,
season.duration = 7) #La sèrie te 50 setmanes i l'estacionalitat es setmenal

model_b_AutoAr_Mex <- bsts(l_Mex, state.specification = components, niter =
2000)

CompareBstsModels(list("AutoAr" = model_b_AutoAr_Mex,
                       "LocalLinearTrend" = model_b_Mex),
```

```
                  colors = c("black", "red"), main = "Error absolut acumulat
d'ambdós models")
```

```{r}
hist((as.numeric(-
colMeans(model_b_Mex$one.step.prediction.errors)+l_Mex))^6,xlab = "Valors de la
sèrie",main = "Histograma de la distibució de la sèrie II segons el model
bayesià")

median((as.numeric(-
colMeans(model_b_Mex$one.step.prediction.errors)+l_Mex))^6)
sd((as.numeric(-colMeans(model_b_Mex$one.step.prediction.errors)+l_Mex))^6)
```

Validació:

- Normalitat dels residus per test
```{r}
res_Mex<-residuals(model_b_Mex, mean.only = TRUE)

jarque.bera.test(res_Mex)
```

Comprovació a partir de l'histograma:
```{r}
hist(res_Mex)
```

- Normalitat dels residus per gràfica:
```{r}
res_Mex<-residuals(model_b_Mex)

qqdist(res_Mex)
title("Distribució dels residus del model bayesià per la sèrie II")
```

- Residus estacionaris?
```{r}
AcfDist(res_Mex)
title("Distribució a posteriori de l'autocorrelació dels residus del model
bayesià per la sèrie II")
```

Predicció:

```{r}
p_mostral_l_Mex<-ts(l_Mex[1:335])
p_extramostral_l_Mex<-ts(l_Mex[336:349])
```

```{r}
model_components_Mex <- list()
model_components_Mex  <-  bsts::AddLocalLinearTrend(model_components_Mex,  y  =
p_mostral_l_Mex)
model_components_Mex    <-    bsts::AddSeasonal(model_components_Mex,    y    =
p_mostral_l_Mex, nseasons  = 48, season.duration = 7)

model_b_Mex      <-      bsts(p_mostral_l_Mex,      state.specification      =
model_components_Mex, niter = 2000)
```

- Valors predits:
```{r}
predic_b_p_extramostral_Mex <- predict(model_b_Mex, horizon = 14)
predic_b_p_extramostral_Mex$mean<-(predic_b_p_extramostral_Mex$mean)^6
predic_b_p_extramostral_Mex$mean
```

- Gràfic període mostral
```{r}
p_mostral_b_Mex<-(as.numeric(-
colMeans(model_b_Mex$one.step.prediction.errors)+p_mostral_l_Mex))^6
ts.plot(p_mostral_b_Mex,p_mostral_Mex,col=c("red","black"),ylab="nº
d'infectats",xlab="Gener-Novembre 2020")
title("Valors observats contra valors predits (BSTS)")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Gràfic del període extramostral:
```{r}
par(mfrow=c(1,1))
ts.plot(ts(p_extramostral_Mex),predic_b_p_extramostral_Mex$mean,col=1:2)
title("Valors observats vs predits")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Avaluació de la capacitat predictiva
```{r}
errors<-p_extramostral_Mex-(predic_b_p_extramostral_Mex$mean)

epam_mostral_Mex_2<-sum(100*abs(p_mostral_Mex[-c(which(p_mostral_Mex==0))]-
p_mostral_b_Mex[-c(which(p_mostral_Mex==0))])/abs(p_mostral_Mex[-
c(which(p_mostral_Mex==0))]))/length(p_mostral_Mex[-
c(which(p_mostral_Mex==0))])
epam_mostral_Mex_2

epam_extramostral_Mex_2<-sum(100*abs(errors)/abs(p_extramostral_Mex))/14
epam_extramostral_Mex_2
```

- Predicció a futur:
```{r}
predic_b_p_extramostral_Mex<-ts(predic_b_p_extramostral_Mex$mean,start=336)

ts.plot(p_mostral_Mex,predic_b_p_extramostral_Mex,
col=c("black","red"),ylab="nº d'infectats",xlab="Gener-Desembre 2020")
title("Evolució dels infectats per 100.000 habitants a Mèxic el 2020")
legend("topleft",legend = c("Valors originals", "Valors previstos"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Gràfic de les prediccions:
```{r}
predic_b_p_extramostral_Mex <- predict(model_b_Mex, horizon = 14)

inf = ts((predic_b_p_extramostral_Mex$interval[1,]),start=336)^6
sup = ts((predic_b_p_extramostral_Mex$interval[2,]),start=336)^6

predic_b_p_extramostral_Mex<-ts(predic_b_p_extramostral_Mex$mean,start = 336)^6

ts.plot(p_mostral_Mex,predic_b_p_extramostral_Mex, col=c(4,2))
lines(inf, col="blue", lty="dashed")
lines(sup, col="blue", lty="dashed")
title("Intervals de prediccions")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("blue","red"), lty=1, cex=0.8)
```

Comparació:

- Gràfic del període extramostral:
```{r}
par(mfrow=c(1,1))
```

```
ts.plot(as.numeric(p_extramostral_Mex),predic_b_p_extramostral_Mex,predic_p_ex
tramostral_Mex$pred,col=1:3,ylab="n°     d'infectats",xlab="Desembre     2020")
#Graficat les prediccions
title("Valors observats contra valors predits (SARIMA i BSTS)")
legend("topleft",legend = c("Valor observat", "Valor predit bsts", "Valor predit
SARIMA"),
       col=c("black","red","green"), lty=1, cex=0.8)
```

Xina:

Dades:
```{r}
poblacio_Xin<- 1439323774
Xin<-Xin*100000/poblacio_Xin

head(Xin)

p_mostral_Xin<-ts(Xin[1:335])
p_extramostral_Xin<-ts(Xin[336:349])
```

Gràfic infectats cada 100.000 habitants:
```{r}
plot.ts(Xin,main="N° d'infectats per 100.000 habitants a Xina el 2020",ylab="n°
d'infectats",xlab="Gener-Desembre 2020")
```

Identificació:

- Tendència?
```{r}
summary(lm(Xin~seq(1:length(Xin))))
```

- Component estacional?
```{r}
isSeasonal(Xin, freq=30)
isSeasonal(Xin, freq=12)
isSeasonal(Xin, freq=7)
```

- Diferències regulars:
```{r}
d_Xin<-diff(Xin)
par(mfrow=c(1,1))
plot.ts(d_Xin,main="Diferències   regulars   de   la   sèrie   III",ylab="n°
d'infectats",xlab="Gener-Desembre 2020")
```

Comprovació:
```{r}
summary(lm(d_Xin~seq(1:length(d_Xin))))
```

- Identificació dels components de la sèrie:
```{r}
par(mfrow=c(2,1))
acf(d_Xin,main="FAS    -    Sèrie    III    transformada",ylab="Autocorrelació
simple",xlab="Lag",lag=120)
pacf(d_Xin,main="FAP    -    Sèrie    III    transformada",ylab="Autocorrelació
parcial",xlab="Lag",lag=120)
```

Models ARIMA

Estimació:

```{r}
model_c_Xin<-arima(Xin,order = c(0,1,1), seasonal = list (order = c(0,0,0)))

model_c_Xin
```

Validació:

- Significació dels coeficients:
```{r}
2*pnorm(c(abs(model_c_Xin$coef)/sqrt(diag(model_c_Xin$var.coef))),      mean=0,
sd=1, lower.tail=FALSE)
```

- Normalitat dels residus
```{r}
library(tseries)
jarque.bera.test(model_c_Xin$residuals)
```

Comprovació a partir de l'histograma:
```{r}
hist(model_c_Xin$residuals)
```

- Residus soroll blanc:
```{r}
Box.test(model_c_Xin$residuals,type = c( "Ljung-Box"))
```

Prediccions:

```{r}
model_c_Xin<-arima(p_mostral_Xin,order = c(0,1,1), seasonal = list (order =
c(0,0,0)))
```

- Valors predits:
```{r}
predic_p_extramostral_Xin<-predict(model_c_Xin,n.ahead=14)
predic_p_extramostral_Xin
```

- Gràfic període mostral:
```{r}
p_mostral_arima_Xin<-(as.numeric(-model_c_Xin[["residuals"]]+p_mostral_Xin))

ts.plot(p_mostral_arima_Xin,p_mostral_Xin,col=c("red","black"),ylab="n°
d'infectats",xlab="Gener-Novembre 2020")
title("Valors observats contra valors predits (ARIMA)")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Gràfic període extramostral:
```{r}
par(mfrow=c(1,1))
ts.plot(p_extramostral_Xin,predic_p_extramostral_Xin$pred[1:14],col=1:2)
title("Valors observats vs predits")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Avaluació de la capacitat predictiva
```{r}
errors<-p_extramostral_Xin-predic_p_extramostral_Xin$pred[1:14]
ts.plot(errors,col=2)
```

```
epam_mostral_Xin_1<-sum(100*abs(p_mostral_Xin[-c(which(p_mostral_Xin==0))]-
p_mostral_arima_Xin[-c(which(p_mostral_Xin==0))])/abs(p_mostral_Xin[-
c(which(p_mostral_Xin==0))]))/length(p_mostral_Xin[-
c(which(p_mostral_Xin==0))])
epam_mostral_Xin_1

epam_extramostral_Xin_1<-sum(100*abs(errors)/abs(p_extramostral_Xin))/14
epam_extramostral_Xin_1
```

- Predicció a futur:
```{r}
ts.plot(p_mostral_Xin,predic_p_extramostral_Xin$pred,
col=c("black","red"),ylab="n° d'infectats",xlab="Gener-Desembre 2020")
title("Evolució dels infectats per 100.000 habitants a Xina el 2020")
legend("topleft",legend = c("Valors originals", "Valors previstos"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Intervals de prediccions:
```{r}
inf = predic_p_extramostral_Xin$pred - 2*predic_p_extramostral_Xin$se
sup = predic_p_extramostral_Xin$pred + 2*predic_p_extramostral_Xin$se

ts.plot(p_mostral_Xin,predic_p_extramostral_Xin$pred, col=c(4,2))
lines(inf, col="blue", lty="dashed")
lines(sup, col="blue", lty="dashed")
title("Intervals de prediccions")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("blue","red"), lty=1, cex=0.8)
```

Models bayesians

```{r}
plot.ts(Xin,main="N° d'infectats per la covid-19 a Xina el 2020",ylab="n° de
persones",xlab="Gener-Desembre 2020")
```

Estimació:

```{r}
model_components <- list()
model_components <- bsts::AddAutoAr(model_components, y = Xin)

model_b_Xin <- bsts(Xin, state.specification = model_components, niter = 2000)
```

```{r}
components <- list()
components <- bsts::AddLocalLinearTrend(components, y = Xin)

model_b_lineal_Xin <- bsts(Xin, state.specification = components, niter = 2000)
```

```{r}
CompareBstsModels(list("AutoAr" = model_b_Xin,
                       "LocalLinearTrend" = model_b_lineal_Xin),
                  colors = c("black", "red"), main = "Error absolut acumulat
d'ambdós models")
```

```{r}
hist(as.numeric(-colMeans(model_b_Xin$one.step.prediction.errors)+Xin),xlab   =
"Valors de la sèrie",main = "Histograma de la distibució de la sèrie III segons
el model bayesià")
```

```
median(as.numeric(-colMeans(model_b_Xin$one.step.prediction.errors)+Xin))
sd(as.numeric(-colMeans(model_b_Xin$one.step.prediction.errors)+Xin))
```

Validació:

- Normalitat dels residus per test
```{r}
res_Xin<-residuals(model_b_Xin, mean.only = TRUE)

jarque.bera.test(res_Xin)
```

Comprovació a partir de l'histograma:
```{r}
hist(res_Xin)
```

- Normalitat dels residus per gràfica:
```{r}
res_Xin<-residuals(model_b_Xin)

qqdist(res_Xin)
title("Distribució dels residus del model bayesià per la sèrie III")
```

- Residus estacionaris?
```{r}
AcfDist(res_Xin)
title("Distribució a posteriori de l'autocorrelació dels residus del model
bayesià per la sèrie III")
```

Predicció:

```{r}
model_components <- list()
model_components <- bsts::AddAutoAr(model_components, y = Xin)

model_b_Xin <- bsts(p_mostral_Xin, state.specification = model_components, niter
= 2000)
```

- Valors predits:
```{r}
pred_b_Xin <- predict(model_b_Xin, horizon = 14)
predic_b_p_extramostral_Xin<-pred_b_Xin$mean
predic_b_p_extramostral_Xin
```

- Gràfic període mostral
```{r}
p_mostral_b_Xin<-as.numeric(-
colMeans(model_b_Xin$one.step.prediction.errors)+p_mostral_Xin)
ts.plot(p_mostral_b_Xin,p_mostral_Xin,col=c("red","black"),ylab="n°
d'infectats",xlab="Gener-Novembre 2020")
title("Valors observats contra valors predits (BSTS)")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Gràfic del període extramostral:
```{r}
par(mfrow=c(1,1))
ts.plot(p_extramostral_Xin,predic_b_p_extramostral_Xin,col=1:2)
title("Valors observats vs predits")
legend("topleft",legend = c("Valor observat", "Valor predit"),


```

```
       col=c("black","red"), lty=1, cex=0.8)
```

- Avaluació de la capacitat predictiva
```{r}
errors<-p_extramostral_Xin-predic_b_p_extramostral_Xin

epam_mostral_Xin_2<-sum(100*abs(p_mostral_Xin[-c(which(p_mostral_Xin==0))]-
p_mostral_b_Xin[-c(which(p_mostral_Xin==0))])/abs(p_mostral_Xin[-
c(which(p_mostral_Xin==0))]))/length(p_mostral_Xin[-
c(which(p_mostral_Xin==0))])
epam_mostral_Xin_2

epam_extramostral_Xin_2<-sum(100*abs(errors)/abs(p_extramostral_Xin))/14 #El 14
és el tamany de la mostra
epam_extramostral_Xin_2
```

- Predicció a futur:
```{r}
predic_b_p_extramostral_Xin<-ts(predic_b_p_extramostral_Xin,start=336)

ts.plot(p_mostral_Xin,predic_b_p_extramostral_Xin,
col=c("black","red"),ylab="nº d'infectats",xlab="Gener-Desembre 2020")
title("Evolució dels infectats per 100.000 habitants a Xina el 2020")
legend("topleft",legend = c("Valors originals", "Valors previstos"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Gràfic de les prediccions:
```{r}
pred_b_Xin <- predict(model_b_Xin, horizon = 14)
plot(pred_b_Xin, plot.original = 335)
```

Comparació:

- Gràfic del període extramostral:
```{r}
par(mfrow=c(1,1))
ts.plot(as.numeric(p_extramostral_Xin),predic_b_p_extramostral_Xin,predic_p_ex
tramostral_Xin$pred,col=1:3,ylab="nº d'infectats",xlab="Desembre 2020")
title("Valors observats contra valors predits (ARIMA i BSTS)")
legend("topleft",legend = c("Valor observat", "Valor predit bsts", "Valor predit
ARIMA"),
       col=c("black","red","green"), lty=1, cex=0.8)
```

Resum EPAM:
```{r}
EPAM_NZ<-
c(epam_mostral_NZ_1,epam_extramostral_NZ_1,epam_mostral_NZ_2,epam_extramostral
_NZ_2)
EPAM_Mex<-
c(epam_mostral_Mex_1,epam_extramostral_Mex_1,epam_mostral_Mex_2,epam_extramost
ral_Mex_2)
EPAM_Xin<-
c(epam_mostral_Xin_1,epam_extramostral_Xin_1,epam_mostral_Xin_2,epam_extramost
ral_Xin_2)

resultats<-matrix(c(EPAM_NZ,EPAM_Mex,EPAM_Xin),byrow=T,ncol=4)

colnames(resultats)<-c("EPAM mostral ARIMA","EPAM extramostral ARIMA","EPAM
mostral BSTS","EPAM extramostral BSTS")
row.names(resultats)<-c("Nova Zelanda","Mèxic","Xina")
resultats
```

Anàlisi entre sèries:

```{r}
model_c_NZ
model_c_Mex
model_c_Xin
```

```{r}
p_mostral_arima_NZ<-(as.numeric(-model_c_NZ[["residuals"]]+p_mostral_NZ))
ts.plot(p_mostral_arima_NZ,p_mostral_NZ,col=c("red","black"),ylab="n°
d'infectats",xlab="Gener-Novembre 2020")
title("Valors observats contra valors predits")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)

p_mostral_arima_Mex<-(as.numeric(-
model_c_Mex[["residuals"]]+p_mostral_l_Mex))^6
ts.plot(p_mostral_arima_Mex,p_mostral_Mex,col=c("red","black"),ylab="n°
d'infectats",xlab="Gener-Novembre 2020")
title("Valors observats contra valors predits")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)

p_mostral_arima_Xin<-(as.numeric(-model_c_Xin[["residuals"]]+p_mostral_Xin))
ts.plot(p_mostral_arima_Xin,p_mostral_Xin,col=c("red","black"),ylab="n°
d'infectats",xlab="Gener-Novembre 2020")
title("Valors observats contra valors predits")
legend("topleft",legend = c("Valor observat", "Valor predit"),
       col=c("black","red"), lty=1, cex=0.8)
```

- Estadístics:

Models clàsics

```{r}
#NZ
mean_NZ<-
round(c(mean(p_mostral_arima_NZ),mean(p_mostral_NZ),mean(predic_p_extramostral
_NZ$pred),mean(p_extramostral_NZ)),4)
median_NZ<-
round(c(median(p_mostral_arima_NZ),median(p_mostral_NZ),median(predic_p_extram
ostral_NZ$pred),median(p_extramostral_NZ)),4)
sd_NZ<-
round(c(sd(p_mostral_arima_NZ),sd(p_mostral_NZ),sd(predic_p_extramostral_NZ$pr
ed),sd(p_extramostral_NZ)),4)
max_NZ<-
round(c(max(p_mostral_arima_NZ),max(p_mostral_NZ),max(predic_p_extramostral_NZ
$pred),max(p_extramostral_NZ)),4)
min_NZ<-
round(c(min(p_mostral_arima_NZ),min(p_mostral_NZ),min(predic_p_extramostral_NZ
$pred),min(p_extramostral_NZ)),4)
IQR_NZ<-
round(c(IQR(p_mostral_arima_NZ),IQR(p_mostral_NZ),IQR(predic_p_extramostral_NZ
$pred),IQR(p_extramostral_NZ)),4)

estadistics_NZ_1<-
matrix(c(mean_NZ,median_NZ,sd_NZ,max_NZ,min_NZ,IQR_NZ),byrow=T,ncol=4)
colnames(estadistics_NZ_1)<-c("Model ARIMA (pmost.)","Sèrie (pmost.)","Model
ARIMA (pextramost.)","Sèrie (pextramost.)")
row.names(estadistics_NZ_1)<-c("Mitjana","Mediana","Desviació
típica","Màx.","Min.","RIQ")
estadistics_NZ_1

#Mex
predic_p_extramostral_Mex<-predict(model_c_Mex,n.ahead=14)$pred^6
```

```
mean_Mex<-
round(c(mean(p_mostral_arima_Mex),mean(p_mostral_Mex),mean(predic_p_extramostr
al_Mex),mean(p_extramostral_Mex)),4)
median_Mex<-
round(c(median(p_mostral_arima_Mex),median(p_mostral_Mex),median(predic_p_extr
amostral_Mex),median(p_extramostral_Mex)),4)
sd_Mex<-
round(c(sd(p_mostral_arima_Mex),sd(p_mostral_Mex),sd(predic_p_extramostral_Mex
),sd(p_extramostral_Mex)),4)
max_Mex<-
round(c(max(p_mostral_arima_Mex),max(p_mostral_Mex),max(predic_p_extramostral_
Mex),max(p_extramostral_Mex)),4)
min_Mex<-
round(c(min(p_mostral_arima_Mex),min(p_mostral_Mex),min(predic_p_extramostral_
Mex),min(p_extramostral_Mex)),4)
IQR_Mex<-
round(c(IQR(p_mostral_arima_Mex),IQR(p_mostral_Mex),IQR(predic_p_extramostral_
Mex),IQR(p_extramostral_Mex)),4)

estadistics_Mex_1<-
matrix(c(mean_Mex,median_Mex,sd_Mex,max_Mex,min_Mex,IQR_Mex),byrow=T,ncol=4)
colnames(estadistics_Mex_1)<-c("Model ARIMA (pmost.)","Sèrie (pmost.)","Model
ARIMA (pextramost.)","Sèrie (pextramost.)")
row.names(estadistics_Mex_1)<-c("Mitjana","Mediana","Desviació
típica","Màx.","Min.","RIQ")
estadistics_Mex_1

#Xin
mean_Xin<-
round(c(mean(p_mostral_arima_Xin),mean(p_mostral_Xin),mean(predic_p_extramostr
al_Xin$pred),mean(p_extramostral_Xin)),4)
median_Xin<-
round(c(median(p_mostral_arima_Xin),median(p_mostral_Xin),median(predic_p_extr
amostral_Xin$pred),median(p_extramostral_Xin)),4)
sd_Xin<-
round(c(sd(p_mostral_arima_Xin),sd(p_mostral_Xin),sd(predic_p_extramostral_Xin
$pred),sd(p_extramostral_Xin)),4)
max_Xin<-
round(c(max(p_mostral_arima_Xin),max(p_mostral_Xin),max(predic_p_extramostral_
Xin$pred),max(p_extramostral_Xin)),4)
min_Xin<-
round(c(min(p_mostral_arima_Xin),min(p_mostral_Xin),min(predic_p_extramostral_
Xin$pred),min(p_extramostral_Xin)),4)
IQR_Xin<-
round(c(IQR(p_mostral_arima_Xin),IQR(p_mostral_Xin),IQR(predic_p_extramostral_
Xin$pred),IQR(p_extramostral_Xin)),4)

estadistics_Xin_1<-
matrix(c(mean_Xin,median_Xin,sd_Xin,max_Xin,min_Xin,IQR_Xin),byrow=T,ncol=4)
colnames(estadistics_Xin_1)<-c("Model ARIMA (pmost.)","Sèrie (pmost.)","Model
ARIMA (pextramost.)","Sèrie (pextramost.)")
row.names(estadistics_Xin_1)<-c("Mitjana","Mediana","Desviació
típica","Màx.","Min.","RIQ")
estadistics_Xin_1
```

Models bayesians

```{r}
#NZ
mean_NZ_2<-
round(c(mean(p_mostral_b_NZ),mean(p_mostral_NZ),mean(predic_b_p_extramostral_N
Z),mean(p_extramostral_NZ)),4)
median_NZ_2<-
round(c(median(p_mostral_b_NZ),median(p_mostral_NZ),median(predic_b_p_extramos
tral_NZ),median(p_extramostral_NZ)),4)
```

```
sd_NZ_2<-
round(c(sd(p_mostral_b_NZ),sd(p_mostral_NZ),sd(predic_b_p_extramostral_NZ),sd(
p_extramostral_NZ)),4)
max_NZ_2<-
round(c(max(p_mostral_b_NZ),max(p_mostral_NZ),max(predic_b_p_extramostral_NZ),
max(p_extramostral_NZ)),4)
min_NZ_2<-
round(c(min(p_mostral_b_NZ),min(p_mostral_NZ),min(predic_b_p_extramostral_NZ),
min(p_extramostral_NZ)),4)
IQR_NZ_2<-
round(c(IQR(p_mostral_b_NZ),IQR(p_mostral_NZ),IQR(predic_b_p_extramostral_NZ),
IQR(p_extramostral_NZ)),4)

estadistics_NZ_2<-
matrix(c(mean_NZ_2,median_NZ_2,sd_NZ_2,max_NZ_2,min_NZ_2,IQR_NZ_2),byrow=T,nco
l=4)
colnames(estadistics_NZ_2)<-c("Model BSTS (pmost.)","Sèrie (pmost.)","Model
BSTS (pextramost.)","Sèrie (pextramost.)")
row.names(estadistics_NZ_2)<-c("Mitjana","Mediana","Desviació
típica","Màx.","Min.","RIQ")
estadistics_NZ_2

#Mex
mean_Mex_2<-
round(c(mean(p_mostral_b_Mex),mean(p_mostral_Mex),mean(predic_b_p_extramostral
_Mex),mean(p_extramostral_Mex)),4)
median_Mex_2<-
round(c(median(p_mostral_b_Mex),median(p_mostral_Mex),median(predic_b_p_extram
ostral_Mex),median(p_extramostral_Mex)),4)
sd_Mex_2<-
round(c(sd(p_mostral_b_Mex),sd(p_mostral_Mex),sd(predic_b_p_extramostral_Mex),
sd(p_extramostral_Mex)),4)
max_Mex_2<-
round(c(max(p_mostral_b_Mex),max(p_mostral_Mex),max(predic_b_p_extramostral_Me
x),max(p_extramostral_Mex)),4)
min_Mex_2<-
round(c(min(p_mostral_b_Mex),min(p_mostral_Mex),min(predic_b_p_extramostral_Me
x),min(p_extramostral_Mex)),4)
IQR_Mex_2<-
round(c(IQR(p_mostral_b_Mex),IQR(p_mostral_Mex),IQR(predic_b_p_extramostral_Me
x),IQR(p_extramostral_Mex)),4)

estadistics_Mex_2<-
matrix(c(mean_Mex_2,median_Mex_2,sd_Mex_2,max_Mex_2,min_Mex_2,IQR_Mex_2),byrow
=T,ncol=4)
colnames(estadistics_Mex_2)<-c("Model BSTS (pmost.)","Sèrie (pmost.)","Model
BSTS (pextramost.)","Sèrie (pextramost.)")
row.names(estadistics_Mex_2)<-c("Mitjana","Mediana","Desviació
típica","Màx.","Min.","RIQ")
estadistics_Mex_2

#Xin
mean_Xin_2<-
round(c(mean(p_mostral_b_Xin),mean(p_mostral_Xin),mean(pred_b_Xin$mean),mean(p
_extramostral_Xin)),4)
median_Xin_2<-
round(c(median(p_mostral_b_Xin),median(p_mostral_Xin),median(pred_b_Xin$mean),
median(p_extramostral_Xin)),4)
sd_Xin_2<-
round(c(sd(p_mostral_b_Xin),sd(p_mostral_Xin),sd(pred_b_Xin$mean),sd(p_extramo
stral_Xin)),4)
max_Xin_2<-
round(c(max(p_mostral_b_Xin),max(p_mostral_Xin),max(pred_b_Xin$mean),max(p_ext
ramostral_Xin)),4)
min_Xin_2<-
round(c(min(p_mostral_b_Xin),min(p_mostral_Xin),min(pred_b_Xin$mean),min(p_ext
ramostral_Xin)),4)
```

```
IQR_Xin_2<-
round(c(IQR(p_mostral_b_Xin),IQR(p_mostral_Xin),IQR(pred_b_Xin$mean),IQR(p_ext
ramostral_Xin)),4)

estadistics_Xin_2<-
matrix(c(mean_Xin_2,median_Xin_2,sd_Xin_2,max_Xin_2,min_Xin_2,IQR_Xin_2),byrow
=T,ncol=4)
colnames(estadistics_Xin_2)<-c("Model BSTS (pmost.)","Sèrie (pmost.)","Model
BSTS (pextramost.)","Sèrie (pextramost.)")
row.names(estadistics_Xin_2)<-c("Mitjana","Mediana","Desviació
típica","Màx.","Min.","RIQ")
estadistics_Xin_2
```

Causal Impact BSTS:

```{r}
times <- seq.Date(as.Date("2020-01-01"), by = 1, length.out = 349)

data_NZ<- zoo(cbind(c(NZ),rep(0,n=349)), times)
data_Mex<- zoo(cbind(c(Mex),rep(0,n=349)), times)
data_Xin<- zoo(cbind(c(Xin),rep(0,n=349)), times)
```

Intervencions Nova Zelanda:

- 19 de març (pos.79) - Prohibició de l'entrada al país dels no residents

```{r}
a<-79+14
pre.period <- c(1, a-1)
post.period <- c(a, a+28)
data<-data_NZ[c(1:(a+28)),]

int1_NZ <- CausalImpact(data,times[pre.period], times[post.period])

summary(int1_NZ)
summary(int1_NZ, "report")

plot(int1_NZ)
plot(int1_NZ, "original")
```

- 25 de maig (pos.146) - Bloqueig del país a nivell Nacional

```{r}
a<-146+14
pre.period <- c(1, a-1)
post.period <- c(a, a+28)
data<-data_NZ[c(1:(a+28)),]

int2_NZ <- CausalImpact(data,times[pre.period], times[post.period])

summary(int2_NZ)
summary(int2_NZ, "report")

plot(int2_NZ)
plot(int2_NZ, "original")
```

- 12 d'agost (pos.225) - Confinament de la ciutat d'Auckland

```{r}
a<-225+14
pre.period <- c(1, a-1)
post.period <- c(a, a+28)
data<-data_NZ[c(1:(a+28)),]

int3_NZ <- CausalImpact(data,times[pre.period], times[post.period])

summary(int3_NZ)
```

```
summary(int3_NZ, "report")

plot(int3_NZ)
plot(int3_NZ, "original")
```

Intervencions Mèxic:

- 22 de març (pos.82) - Tancament d'espais públics interiors (Teatres, museu, ...)
```{r}
a<-82+14
pre.period <- c(1, a-1)
post.period <- c(a, a+28)
data<-data_Mex[c(1:(a+28)),]

int1_Mex <- CausalImpact(data,times[pre.period], times[post.period])

summary(int1_Mex)
summary(int1_Mex, "report")

plot(int1_Mex)
plot(int1_Mex, "original")
```

- 30 de març (pos.90) - Declaració de l'estat d'emergència sanitària
```{r}
a<-90+14
pre.period <- c(1, a-1)
post.period <- c(a, a+28)
data<-data_Mex[c(1:(a+28)),]

int2_Mex <- CausalImpact(data,times[pre.period], times[post.period])

summary(int2_Mex)
summary(int2_Mex, "report")

plot(int2_Mex)
plot(int2_Mex, "original")
```

- 8 d'abril (pos.99) - Obertura del centre mèdic naval per casos greus de covid-19
```{r}
a<-99+14
pre.period <- c(1, a-1)
post.period <- c(a, a+28)
data<-data_Mex[c(1:(a+28)),]

int3_Mex <- CausalImpact(data,times[pre.period], times[post.period])

summary(int3_Mex)
summary(int3_Mex, "report")

plot(int3_Mex)
plot(int3_Mex, "original")
```

- 21 d'abril (pos.112) - Suspensió de les activitats no essencials i mesures estrictes de confinament
```{r}
a<-112+14
pre.period <- c(1, a-1)
post.period <- c(a, a+28)
data<-data_Mex[c(1:(a+28)),]

int4_Mex <- CausalImpact(data,times[pre.period], times[post.period])
```

```
summary(int4_Mex)
summary(int4_Mex, "report")

plot(int4_Mex)
plot(int4_Mex, "original")
```

Intervencions Xina:

- 23 de gener (pos.23) - Confinament de la població, tancament d'escoles i quarantena de Hubei

```{r}
a<-23+14
pre.period <- c(1, a-1)
post.period <- c(a, a+28)
data<-data_Xin[c(1:(a+28)),]

int1_Xin <- CausalImpact(data,times[pre.period], times[post.period])

summary(int1_Xin)
summary(int1_Xin, "report")

plot(int1_Xin)
plot(int1_Xin, "original")
```

- 3 de febrer (pos.34) - Inauguració de l'hospital d'emergència de Wuhan

```{r}
a<-34+14
pre.period <- c(1, a-1)
post.period <- c(a, a+28)
data<-data_Xin[c(1:(a+28)),]

int2_Xin <- CausalImpact(data,times[pre.period], times[post.period])

summary(int2_Xin)
summary(int2_Xin, "report")

plot(int2_Xin)
plot(int2_Xin, "original")
```