

# PREDICCIÓ DE LA TAXA D'ATUR ESPANYOLA: UN ANÀLISI REGIONAL

Anton Figuerola Wischke

Tutoria: Dr. Oscar Claveria González  
Ensenyament: Administració i Direcció d'Empreses



UNIVERSITAT DE  
BARCELONA



## RESUM I PARAULES CLAU – ABSTRACT AND KEY WORDS

Aquest projecte d'investigació té per objectiu principal la realització i estudi previ de les prediccions de l'atur espanyol, posant èmfasi a l'anàlisi regional. Per un costat, es fa ús dels mètodes deterministes per la taxa de desocupació nacional. Mitjançant els contrastos de Daniel i Kruskal-Wallis es detecta que la sèrie temporal presenta tendència però no component estacional. És per això que els mètodes emprats en la primera fase del exercici són: la TL (Tendència Lineal), les DMM (Dobles Mitjanes Mòbils) i l'AEH (Allisat Exponencial de Holt). Per l'altre costat, es generen prediccions a través de models ARIMA (Model Autoregressiu de Mitjana Mòbil) a nivell nacional i per a cada una de les disset CCAA (Comunitats Autònomes). En darrer lloc, s'avalua la capacitat predictiva dels mètodes i models utilitzats. La conclusió principal d'aquest estudi subratlla la importància de generar prediccions independents per a cada regió.

### Forecasting Unemployment in Spain: A Regional Analysis

*The goal of this research project is to conduct a forecasting experiment for the Spanish unemployment rate at a regional level. We use a sampling size that starts at the first quarter of the year 2002 and ends at the last quarter of the year 2017. First, we apply a deterministic approach at the national level. We run Daniel and Kruskal-Wallis tests and find that the unemployment growth rate presents a significant trend component although the seasonal component is not significant. As a result, we use the Linear Trend, Double Moving Average and Holt Exponential Smoothing methods. Second, we generate forecasts by means of ARIMA models at the national and regional level. At the national exercise we use the Box-Jenkins Methodology with Gretl and compare it to the output generated with the auto-ARIMA function with R designed by Hyndman and Khandakar (2008). We find that the auto-ARIMA function provides better outcomes, for this reason we use it for model selection at the regional level. We obtain different optimal ARIMA models in each region. This finding suggests that unemployment forecasting should be implemented by means of region-specific models. Finally, we compute the out-of-sample forecast accuracy for the four quarters of the year 2017. We obtain the best results for Galicia and Valencian Community, as opposed to Aragon.*

Taxa d'atur | *Unemployment Rate*  
Mètodes de previsió | *Forecasting Methods*  
Economia aplicada | *Applied Economics*  
Sèries temporals | *Time Series*  
Mètodes deterministes | *Deterministic Methods*  
Models estocàstics | *Stochastic Models*  
Model Autoregressiu de Mitjana Mòbil | *Autoregressive Integrated Moving Average*  
Metodologia Box-Jenkins | *Box-Jenkins Methodology*  
Funció Auto-ARIMA | *Auto-ARIMA Function*

# ÍNDEX

<b>Abreviacions i notacions</b>	2
<b>I. Introducció</b>	3
<b>II. Anàlisi nacional</b>	4
1. Anàlisi determinista	4
2. Anàlisi estocàstica	8
2.1. Metodologia Box-Jenkins	8
2.2. Auto-ARIMA	18
3. Avaluació de la capacitat predictiva	21
<b>III. Anàlisi regional</b>	22
<b>IV. Conclusions</b>	35
<b>Bibliografia</b>	36
<b>Annexos</b>	37

## ABREVIACIONS I NOTACIONS

AEH	Allisat Exponencial de Holt o Doble
AIC	<i>Akaike's Information Criteria</i> Criteri d'Informació Akaike
AICc	<i>Akaike's Information Criteria Corrected</i> Criteri d'Informació Akaike Corregit
AR	<i>Autoregressive Model</i> Model Autoregressiu
ARIMA	<i>Autoregressive Integrated Moving Average</i> Model Autoregressiu Integrat de Mitjana Mòbil
BJ	Box-Jenkins
CCAA	Comunitats Autònomes
DMM	Dobles Mitjanes Mòbils
EAM	Error Absolut Mitjà
EPAM	Error Percentual Absolut Mitjà
EQM	Error Quadràtic Mitjà
FAS	Funció d'Autocorrelació Simple (o FAC)
FAP	Funció d'Autocorrelació Parcial (o FACP)
H	Període extra-mostrat
H <sub>A</sub>	Hipòtesi alternativa
H <sub>0</sub>	Hipòtesi nul·la
HQC	<i>Hannan-Quinn's Information Criterion</i> Criteri d'Informació Hannan-Quinn
M	Mètode
MA	<i>Moving Average</i> Mitjana Mòbil
R	Entorn i llenguatge de programació estadístic
SBIC	<i>Schwarz's Information Criteria</i> Criteri d'Informació Schwarz
TL	Tendència Lineal

# I. INTRODUCCIÓ

L'atur és una de les macro-variables amb més consideració pels agents econòmics. La forma més freqüent de mesurar-lo és a través de la taxa d'atur.

A què fa referència la taxa d'atur? La taxa d'atur fa referència a la part de la població activa que no es troba ocupada. Cal recordar que s'entén per població activa aquella que estan en edat, condicions i disposició legal de treballar, compta amb un lloc de treball o aspira a poder accedir-ne en algun.

La taxa d'atur és un indicador molt important, sobretot en termes polítics, ja que està estretament relacionada amb el creixement econòmic d'un país. Així, la disminució de la taxa d'atur a un país s'acaba reflectint en un increment dels ingressos de les famílies i consegüentment en un augment de la qualitat de vida. I no només es beneficien les famílies, sinó que també ho fan les empreses a través d'un increment de les vendes. El contrari succeeix amb elevades taxes d'atur. Per tant, un dels principals objectius a nivell de política econòmica cau en la reducció de la taxa d'atur.

Així doncs, és fonamental predir amb precisió l'evolució de l'atur, ja que ens permet anticipar el futur o afrontar-lo de la manera més convenient. Igual que quan anem de viatge consultem prèviament quin temps farà al lloc de destí per tal d'escollir la roba més adient; de la mateixa manera un govern prendrà les mesures necessàries, com per exemple una millora educativa o ajustos fiscals, si les prediccions de la taxa d'atur són negatives.

Però no només és necessari tenir en compte la taxa d'atur a nivell nacional, sinó que també ho és a nivell regional, ja que existeix heterogeneïtat entre les diferents CCAA (Comunitats Autònomes) de l'Estat espanyol. Presenten diversitat respecte l'estructura productiva, els cicles econòmics, etc. donant lloc a diferents evolucions de l'ocupació a nivell regional. Per exemple, Andalusia va registrar una taxa d'atur del 24% el darrer trimestre de l'any 2017, mentre que a Catalunya va ser del 13% (Institut Nacional d'Estadística, 2018).

L'estudi es centra en l'anàlisi predictiu de l'atur, tant a nivell nacional com a regional. D'una banda s'utilitzen els mètodes deterministes adients per la taxa nacional. D'altra banda es cerca el model ARIMA òptim a nivell agregat i també per les diferents CCAA. Finalment s'avalua la capacitat predictiva. Totes les prediccions comprenen els quatre trimestres de la l'any 2017.

Amb l'objectiu de facilitar la selecció dels models estocàstics s'ha emprat la funció auto-ARIMA dissenyada per Hyndman i Khandakar (2008).

A tall de cloenda, volia expressar el meu agraïment més sincer al Dr. Oscar Claveria per la seva valuosa contribució en l'elaboració d'aquest projecte.

## II. ANÀLISI NACIONAL

L'anàlisi nacional es divideix en tres parts. La primera compren l'anàlisi determinista, la segona l'anàlisi estocàstica i, l'ultima, compara la capacitat predictiva d'ambdós enfocaments.

### 1. Anàlisi determinista

Un mètode determinista és un model matemàtic on les mateixes entrades produiran invariablement les mateixes sortides, sense contemplar l'existència de l'atzar ni de la incertesa.

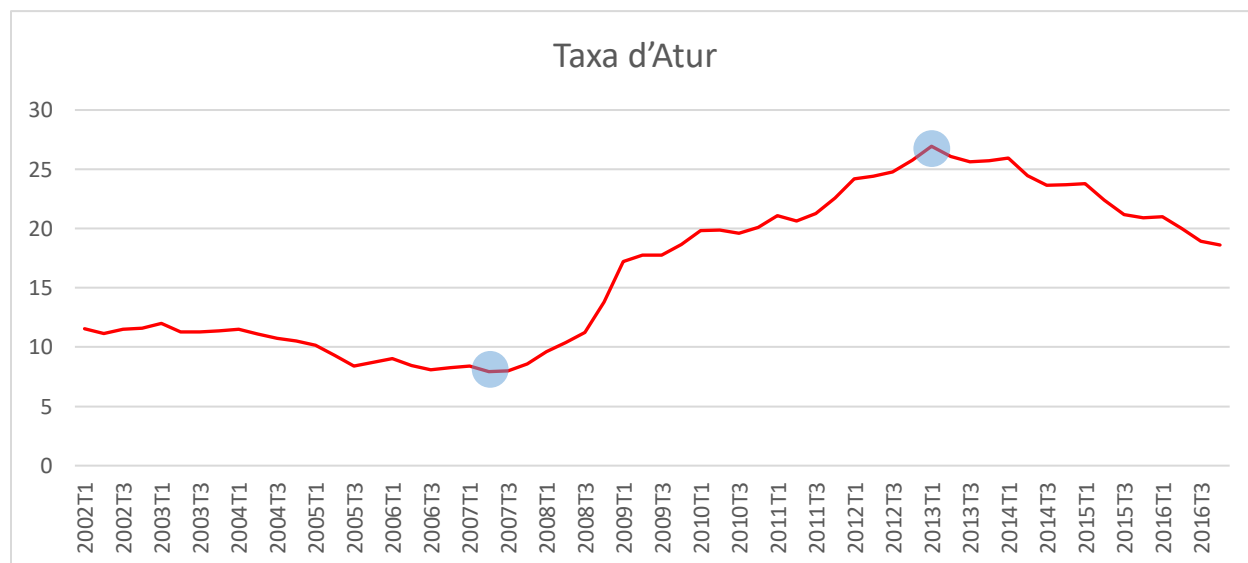
Per tal d'implementar aquest tipus d'anàlisi, en primer lloc cal determinar si la sèrie temporal analitzada presenta tendència i component estacional. Amb aquest objectiu s'implementen dos tests estadístics: el contrast de Daniel (contrastava si la sèrie temporal manifesta tendència) i el contrast de Kruskal-Wallis (contrastava si la sèrie temporal presenta component estacional). Un cop aplicats, es pot categoritzar la sèrie temporal en quatre classes: series tipus 1 (sense tendència i sense component estacional), series tipus 2 (sense tendència i amb component estacional), series tipus 3 (amb tendència i sense component estacional) i series tipus 4 (amb tendència i amb component estacional). Aquesta taxonomia permet aplicar els mètodes més idonis en cada cas.

---

#### Identificació del tipus de sèrie

---

L'estudi analitza la taxa d'atur des del primer trimestre de l'any 2002 fins el quart trimestre de l'any 2017 (grandària de la mostra), la qual cosa permet avaluar l'efecte de la crisi sobre l'ocupació.



**Gràfic 1.** Representació de la sèrie "Taxa d'Atur".

Al Gràfic 1 es pot observar com el procés de destrucció de llocs de treball que va acompanyar la crisi financera de l'any 2008 sembla revertir-se a partir del primer trimestre de l'any 2013.

A continuació s'analitza si la sèrie temporal presenta component tendencial així com estacional per tal de categoritzar-la posteriorment. En primer lloc s'utilitza la interpretació gràfica i, en segon lloc, es porten a terme dos tests estadístics.

D'una banda, basant-nos en el *Gràfic 1* podem intuir que la sèrie presenta tendència. D'entrada, sembla que la funció lineal traçada tendeix a decreixer lleugerament fins assolir el primer punt de gir, situat al segon trimestre de l'any 2007. Després, comença a créixer de manera bastant accentuada fins a l'inici de l'any 2013, moment on es produeix el segon punt de gir. Arribats al segon punt de gir, la sèrie tendeix novament a decreixer fins al final del període mostral.

D'altra banda, basant-nos en el *Gràfic 1*, és difícil apreciar si hi ha o no component estacional. Sembla que es manifesta un patró de lleuger descens del percentatge de persones a l'atur durant els tercers trimestres. Tot i això, no es podria considerar com a significatiu.

Per tal de contrastar estadísticament l'existència de tendència i estacionalitat s'apliquen dos contrastos:

### 1. Contrast de Daniel

Partint del següent contrast d'hipòtesis:

$$\begin{cases} H_0 = \text{sèrie sense tendència} \\ H_A = \text{sèrie amb tendència} \end{cases}$$

$$\begin{cases} \text{Si } |Z| > Z_{\alpha/2} \approx 1.96 \rightarrow \text{Rebutgem } H_0 \\ \text{Si } |Z| \leq Z_{\alpha/2} \approx 1.96 \rightarrow \text{No rebutgem } H_0 \end{cases}$$

Donat que l'estadístic de prova  $Z$  de la sèrie temporal estudiada és igual a 5.67 per a un nivell de significació del 5%, rebutgem la hipòtesis nul·la. Per la qual cosa podem afirmar que la sèrie presenta un component tendencial significatiu. (*Annex 1*)

### 2. Contrast de Kruskal-Wallis

En aquest cas les dues hipòtesis possibles són les següents:

$$\begin{cases} H_0 = \text{sèrie sense component estacional} \\ H_A = \text{sèrie amb component estacional} \end{cases}$$

Amb una periodicitat trimestral ( $s=4$ ), el valor crític és de 7.81, així doncs:

$$\begin{cases} \text{Si } H > X^2_{\alpha, s-1} \approx 7.81 \rightarrow \text{Rebutgem } H_0 \\ \text{Si } H \leq X^2_{\alpha, s-1} \approx 7.81 \rightarrow \text{No rebutgem } H_0 \end{cases}$$

Pel contrast Kruskal-Wallis, l'estadístic de prova,  $H$ , dona un valor de 0.35. Com que  $H$  és inferior al valor crític per un nivell de significació del 5%, no es rebutja la hipòtesis nul·la, és a dir, la sèrie per a la variable "taxa d'atur" no té component estacional. (*Annex 1*)



En conseqüència, podem concloure que ens trobem davant d'una sèrie tipus 3, ja que la sèrie temporal estudiada presenta tendència però no mostra un component estacional significatiu al llarg del període mostral. Conseqüentment els mètodes més adients per executar les estimacions pertinents són: la TL, les DMM i l'AEH.

---

### Prediccions

---

El primer mètode que apliquem és el de la TL. Concretament aquesta tècnica d'estructura fixa regressa la variable dependent (taxa d'atur) sobre un rang temporal a través d'una equació lineal. Els paràmetres obtinguts s'utilitzen per generar les prediccions pertanyents. (*Annex 3*)

El segon mètode és el de les DMM el qual consisteix en calcular un conjunt de mitjanes mòbils de la sèrie original i tot seguit calcular un segon conjunt com a mitjana mòbil del primer. Aquest, a diferència del mètode de la TL, es caracteritza per ser d'estructura variable i no pas d'estructura fixa. (*Annex 3*)

El darrer mètode utilitzat és el de l'AEH. És una tècnica d'estructura variable la qual minora directament la tendència al obtenir la diferència entre els valors successius (de la minoració exponencial). Incorpora dues constants d'allistament ( $\alpha$  i  $\gamma$ ) les quals prenen valors compresos entre 0 i 1. Com més grans siguin els valors de les constants, major és la importància atorgada a les dades recents. (*Annex 3*)

Arrel de la informació fins el quart trimestre de l'any 2016 es generen les prediccions extra-mostrals pels quatre trimestres de l'any 2017.

DATA	Yt (Valor Real)	M.TL	M. DMM	M. AEH <sup>1</sup>
2017T1	18.75%	25.68%	18.69%	18.69%
2017T2	17.22%	25.98%	17.83%	17.75%
2017T3	16.38%	26.28%	16.48%	16.43%
2017T4	16.55%	26.58%	15.96%	15.98%

**Taula 1.** *Prediccions dels mètodes deterministes.*

Si donem un ràpid cop d'ull a la *Taula 1*, crida a l'atenció les elevades prediccions obtingudes mitjançant el mètode de la TL, les quals divergeixen notablement amb les dades reals (Yt). Pel que fa l'output dels dos mètodes restants emprats, sembla tenir un elevat grau d'assertivitat.

---

### Capacitat predictiva

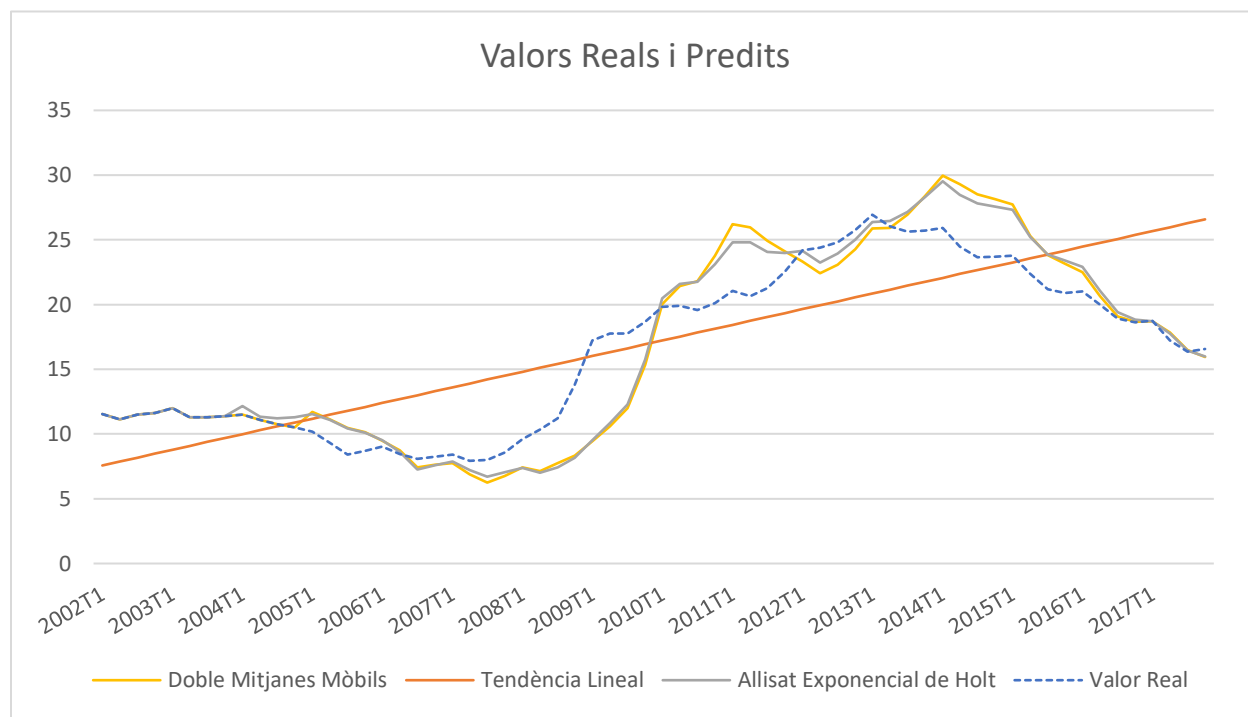
---

Per tal d'estudiar la qualitat de les prediccions de cada mètode utilitzarem d'una banda l'anàlisi gràfic i d'altra banda un conjunt d'estadístics no paramètrics: EAM (Error Absolut Mitjà), EQM (Error Quadràtic Mitjà) i EPAM (Error Percentual Absolut Mitjà) (*Annex 2*). En tots tres estadístics, com més petit sigui el quocient més bona és la capacitat predictiva. Així mateix, dels tres

---

<sup>1</sup> Per al mètode de l'AEH la constant d'allistament de la tendència o alpha ( $\alpha$ ) utilitzat és de 0.8 punts i per la pendent o gamma ( $\gamma$ ) és 0.7.

estadístics l'EPAM és l'únic adimensional (no depèn de les unitats de mesura de la variable analitzada), és per això que se li dóna més importància. Aquesta avaluació de la capacitat predictiva es du a terme pel període extra-mostrat comprès pels quatre trimestres de l'any 2017.



**Gràfic 2.** Representació dels valors reals i predits mitjançant mètodes deterministes.

Al traçar un gràfic lineal amb les prediccions de cada respecte mètode determinista per a sèries tipus 3, juntament amb els valors reals (*Gràfic 2*), podem observar que el mètode més asserit és el de l'AEH, seguit de molt a prop del de les DMM. En canvi, el mètode de la TL mostra tenir les desviacions més elevades de totes al llarg del temps.

Mètode	EAM(H)	EQM(H)	EPAM(H)
TL	8.91	80.86	52.22%
DMM	0.34	0.18	2.00%
AEH	0.30	0.15	1.79%

**Taula 2.** Capacitat predictiva dels mètodes deterministes.

Tenint en compte els indicadors EAM, EQM i EPAM facilitats a la *Taula 2*, podem afirmar definitivament que l'AEH és el mètode que genera les prediccions més acurades.

Un EPAM per sota del 3% denota molt bona capacitat predictiva i, en aquest cas tant el mètode de les DMM com el de l'AEH ho aconsegueixen.

Finalment, podem observar l'elevat percentatge del EPAM obtingut en el mètode de la TL, indicant altre cop una baixa capacitat predictiva. A més a més, el EQM, el qual assigna una ponderació major als errors grans que no pas als petits, també és considerablement elevat en

comparació al dels altres mètodes emprats. Aquest resultat es podria explicar pel fet que el mètode de la TL obté les prediccions suposant que la pendent és constant en tot el període mostral. Això perjudica significativament les prediccions, ja que tal i com mostra el *Gràfic 1*, a inicis de l'any 2013 es produeix un punt de gir important (la sèrie temporal té una forta tendència creixent des de l'any 2008 fins al 2013 i, després passa a ser decreixent).

## 2. Anàlisi estocàstica

Un procés estocàstic (conegut també com a model probabilístic) es pot definir formalment com “una família de variables aleatòries  $\{Y_t\}$ , on  $t$  denota el temps, de tal manera que per a qualsevol elecció finita de valors de  $t, t_1, t_2, \dots, t_n$  existeix la distribució de probabilitat conjunta corresponent a les variables aleatòries  $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ ” (Pérez-Ramírez, 2006).

Per la qual cosa podem interpretar un procés estocàstic com a una successió de variables aleatòries on les característiques de les quals poden variar al llarg del temps.

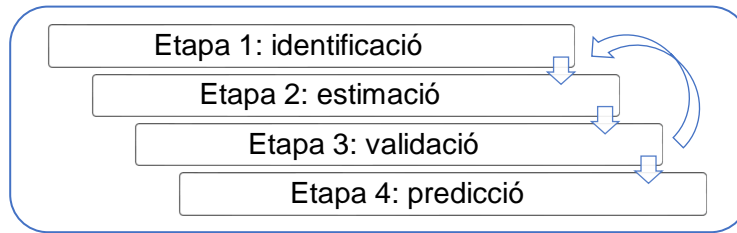
Així doncs, a diferència dels mètodes deterministes, fonamentat en la no aleatorietat, els processos estocàstics incorporen l'efecte del atzar per a obtenir les relatives estimacions.

Gran part de l'anàlisi estocàstica dut a terme en aquest apartat gira al voltant del plantejament aportat pels matemàtics Box i Jenkins (1964), els quals van proposar la modelització temporal mitjançant els processos ARIMA, el qual es pot entendre com un model matemàtic dinàmic dissenyat amb el propòsit d'obtenir prediccions a partir d'una sèrie temporal originalment no estacionària. Aquest està format per una part regular (p,d,q) i en alguns casos també per una d'estacional (P,D,Q).

Amb l'objectiu d'identificar el model ARIMA òptim s'aplica d'una banda la metodologia Box-Jenkins utilitzant *Gretl* i d'altra banda la implantació d'un programa de selecció automàtica, auto-ARIMA, mitjançant R.

### 2.1. Metodologia Box-Jenkins

A la metodologia Box-Jenkins es du a terme una seqüència de 4 etapes esquematitzades al *Gràfic 3*. A la primera etapa, identificació, primer es determina si la sèrie temporal és estacionària (en cas de no ser-ho caldrà aplicar les diferències i/o logaritmes oportuns per a la seva transformació) i, a continuació, trobar els valors  $p$  i  $q$  (que indiquen l'ordre dels potencials esquemes AR (autoregressiu) i MA (mitjana mòbil) respectivament) del model ARIMA. Seguidament, a l'etapa d'estimació, s'estima el model prèviament identificat als correlogrames. A la tercera etapa, validació, s'avalua la constant i els paràmetres estimats. Si aquests són significatius i, a més a més, compleixen amb les condicions d'invertibilitat i estacionarietat, es porta a terme un anàlisi dels residus. En cas de que els residus mostrin un comportament propi d'un procés de soroll blanc es procedeix a generar les prediccions (darrera etapa), en cas contrari es tornaria al punt de partida.



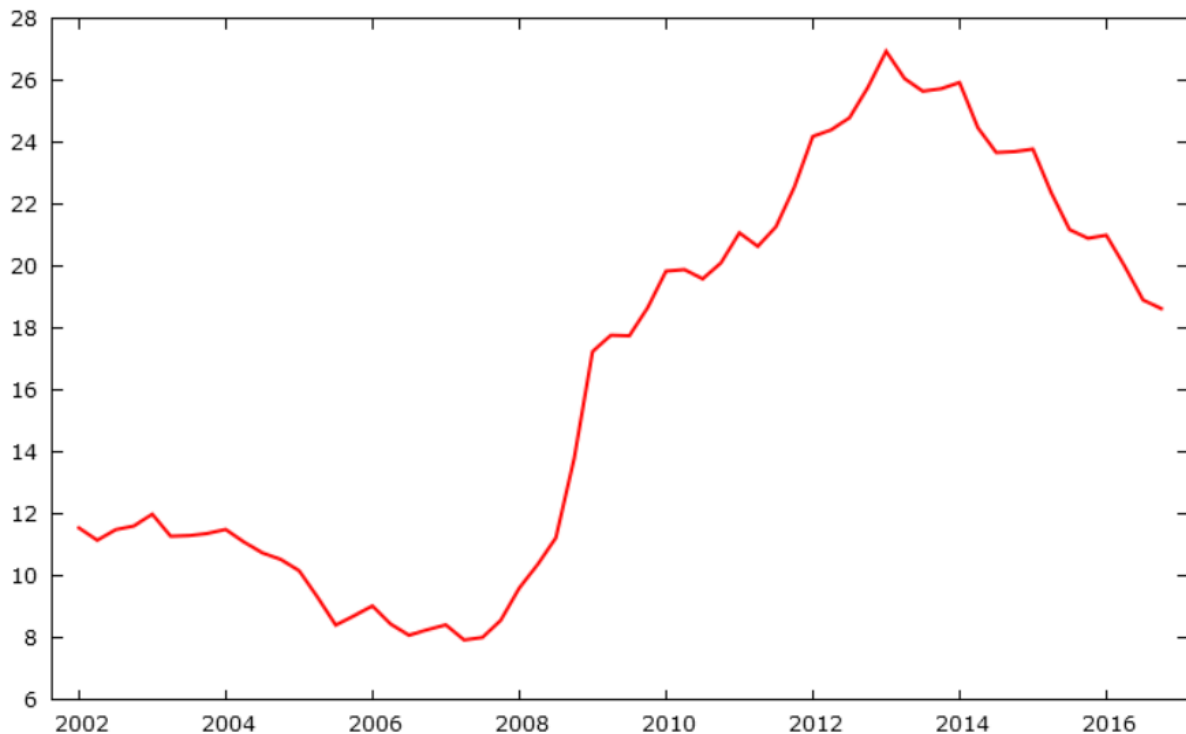
**Gràfic 3.** *Etapes metodologia Box-Jenkins.*

### Primer etapa: identificació

A la fase d'identificació es determina en primer lloc si la sèrie temporal és estacionària en mitjana i variància, tant a la part regular com a la part estacional, ja que en cas de no ser-ho caldrà transformar-la aplicant les diferències necessàries (d,D).

D'entrada, al *Gràfic 4* es pot observar com l'evolució de l'atur a Espanya assoleix un mínim (vall) al segon trimestre del 2007 (8%). Aquest punt fa referència a l'inici de la crisi financera de l'any 2008. El gràfic lineal també mostra un cim (part més alta), el qual marca una taxa d'atur propera al 27% (rècord de 6.202.700 persones parades), superant el 30% en les CCAA d'Andalusia, Extremadura, les Illes Canàries, Castella La Manxa i la Regió de Múrcia (Manjón, 2013)).

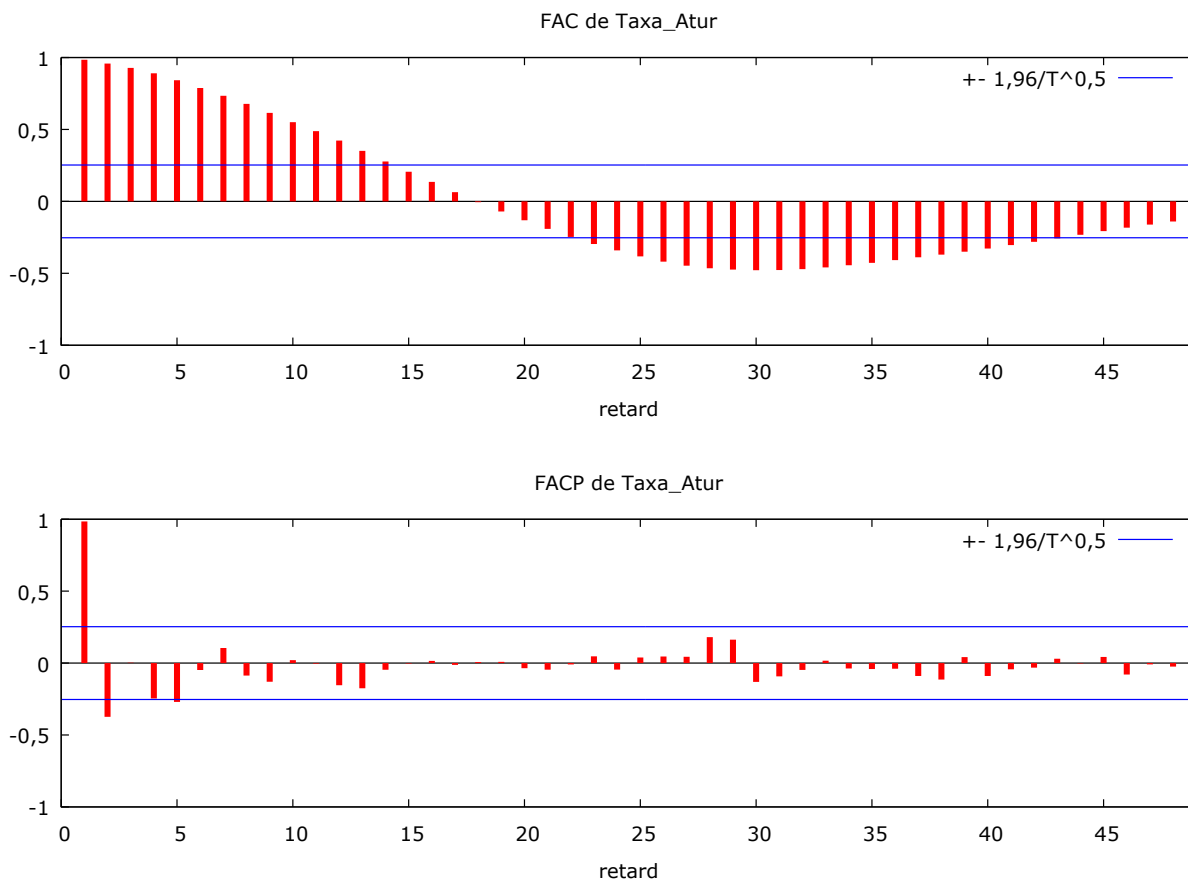
Podem observar una tendència creixent (durant el període compres entre el segon trimestre de l'any 2007 fins al primer trimestre de l'any 2013) i també decreixent (dos períodes: el primer, des de l'inici de l'any 2002 fins a mitats de l'any 2007 i, el segon, des de l'inici de l'any 2013 fins al final de l'any 2016), on la mitjana no és constant al llarg del temps i, conseqüentment la variància també canvia.



**Gràfic 4.** *Representació de la sèrie original. Taxa\_Atur.*

Per la qual cosa, basant-nos en l'anàlisi visual del *Gràfic 1*, podem afirmar en una primera instància que la sèrie temporal no presenta estacionarietat.

A continuació es generen els correlogrames de la sèrie original (*Gràfic 5*) per tal de completar l'anàlisi gràfic i identificar el procés generador de dades. Al correlograma superior es gràfica la FAS (Funció d'Autocorrelació Simple) i al correlograma inferior la FAP (Funció d'Autocorrelació Parcial), ambdós per quaranta-vuit retards ( $k$ ). FAS: en el retard  $k$ , és l'autocorrelació entre els valors de les series que es troben a  $k$  intervals de distància. FAP: en el retard  $k$ , és l'autocorrelació entre els valors de la sèrie que es troben a  $k$  intervals de distància, tenint en compte els valors dels intervals intermedis. Per la qual cosa la FAP, a diferència de la FAS, isola els efectes de les possibles variables restants. Un cop s'ha donat a entendre el significat de les FAS i FAP podem procedir a l'anàlisi dels correlogrames subseqüents.



**Gràfic 5.** Correlogrames FAS i FAP de la sèrie original.

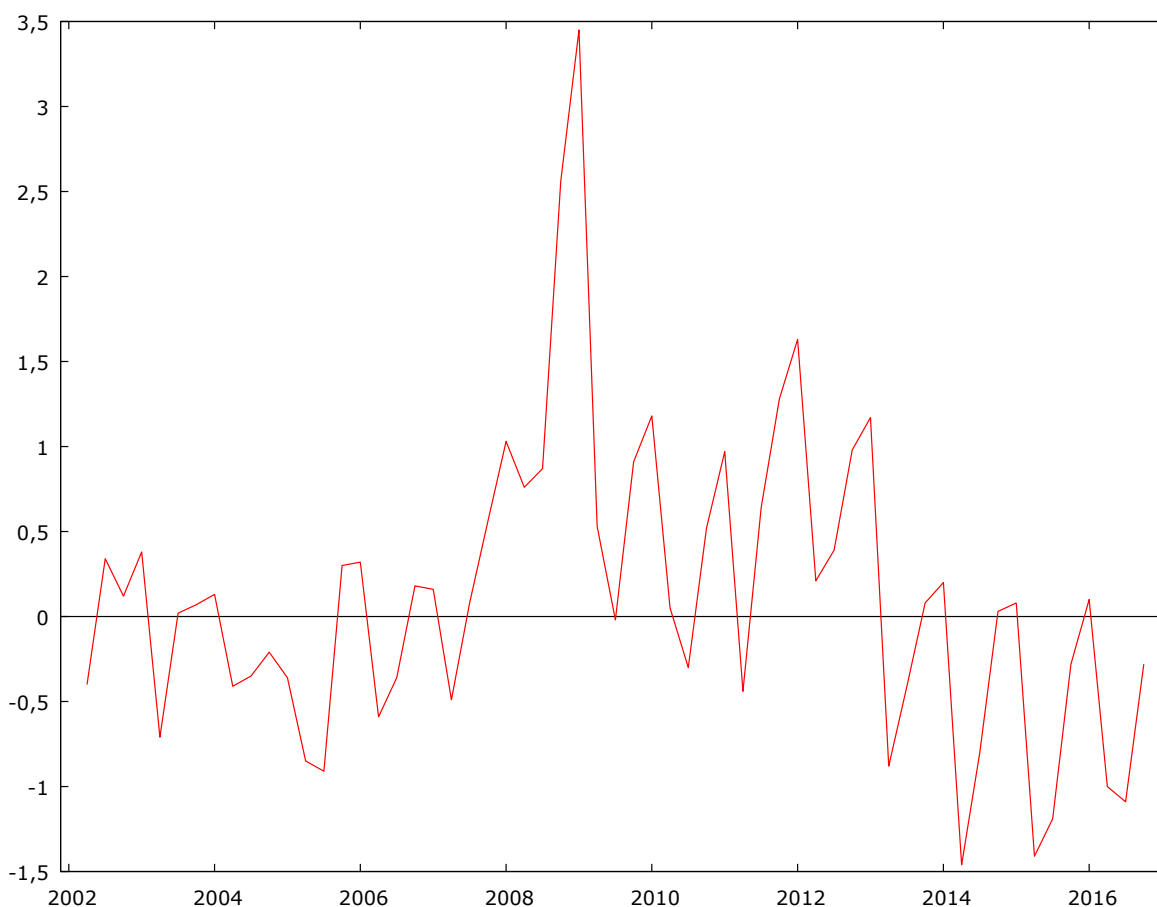
Les característiques de la FAS (*Gràfic 5*) coincideixen amb les d'una sèrie temporal generada per un procés de camí aleatori (també conegut en anglès com a *random walk*). D'una banda, els primers coeficients d'autocorrelació de la FAS són propers a la unitat i, d'altra banda, presenta un decreixement suau al llarg del gràfic.

Pel que fa la FAP, també mostra un comportament característic d'un procés de camí aleatori donat que el primer retard és notablement significatiu i proper a la unitat i, la resta pròxims a zero.

Per la qual cosa, caldrà aplicar una diferència a la part regular per tal d'aconseguir que la sèrie sigui estacionària en mitjana. Conseqüentment l'ordre d'integració  $d$  del model ARIMA serà més gran o igual 1.

Observant el *Gràfic 6* no s'aprecia una clara tendència creixent o decreixent en comparació amb la sèrie original. Tot i això no sembla tenir una estabilitat prou forta al llarg del temps com per a finalitzar la primera etapa de la metodologia Box-Jenkins. Per la qual cosa, basant-nos en els fonaments d'un anàlisi subjectiu, serà necessari transformar novament la sèrie aplicant una segona diferència.

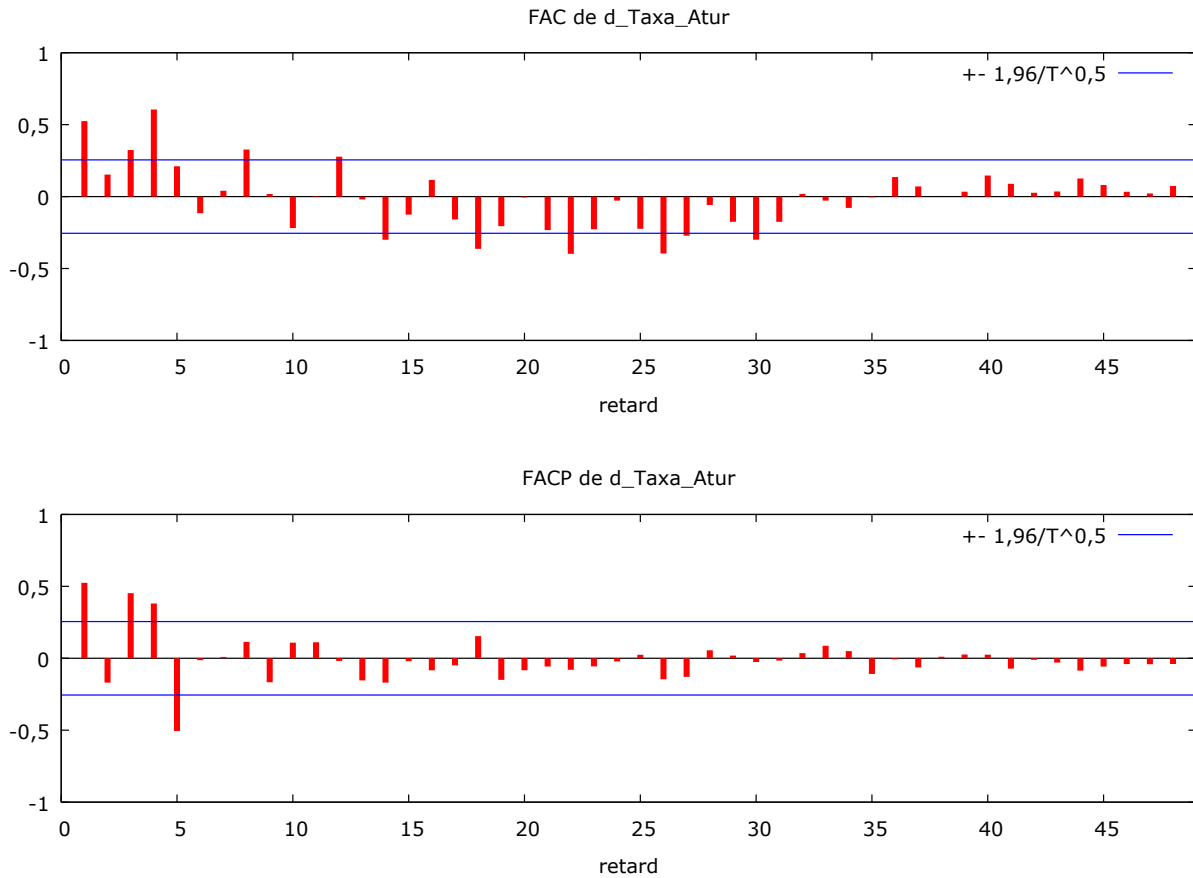
També apareix un *outlier* (observació atípica) comprès entre l'any 2008 i 2010, cosa que podria afectar negativament a la qualitat de les prediccions (Trivez, 1994). Tot i això, en aquest cas no té rellevància ja que la sèrie original no presenta cap *outlier*. Així doncs, no serà necessària la implantació d'un anàlisi d'intervenció.



**Gràfic 6.** Representació de la sèrie transformada amb una primera diferència a la part regular.

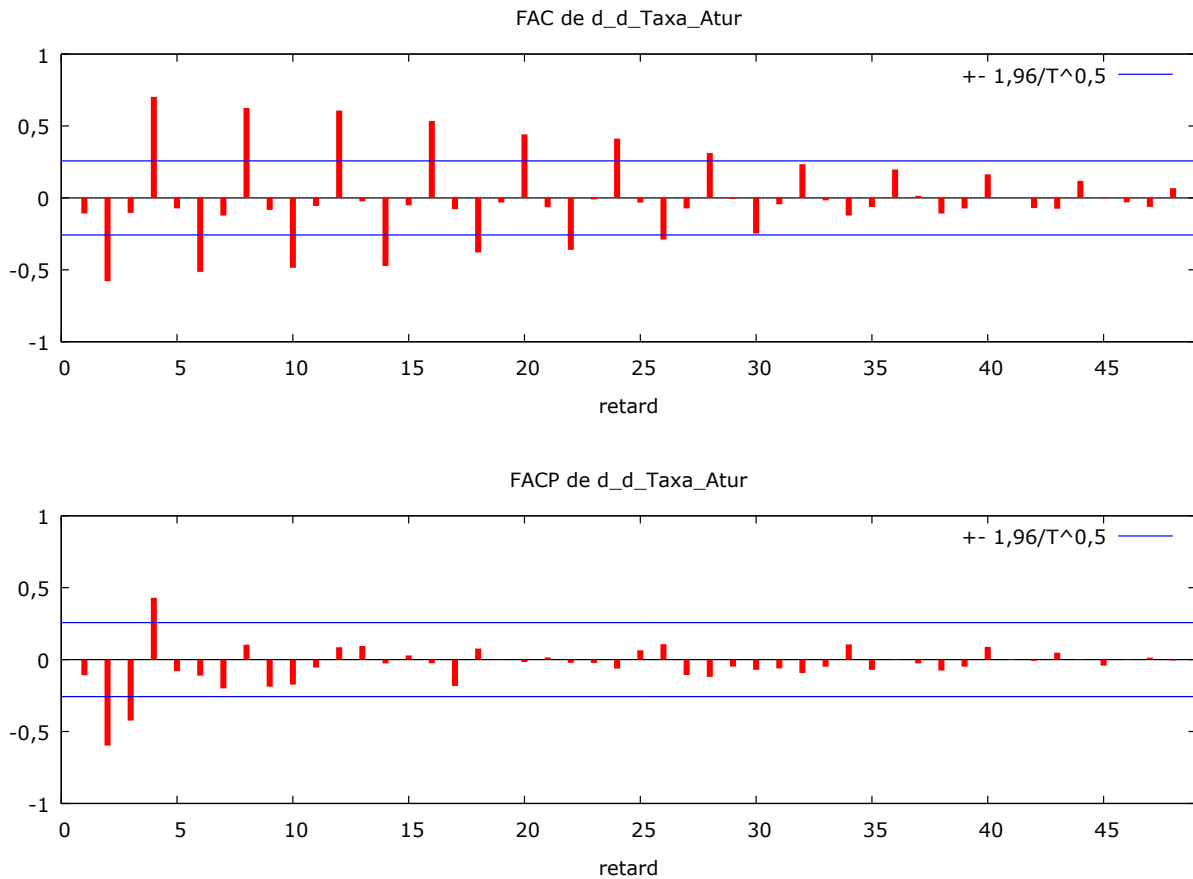
Si observem els correlogrames al *Gràfic 7* podem concloure que la part regular (primers retards) es caracteritza per tenir un decreixement lent tant en la FAS com en la FAP. El contrari succeeix a la part estacional (retards múltiples de 4), on es dona un decreixement ràpid.

En conclusió, recolzant-nos en l'anàlisi de la sortida obtinguda de la sèrie transformada, la part regular segueix sent no estacionària en mitjana. Així doncs, caldrà aplicar una segona diferència, afirmant la hipòtesis anteriorment plantejada.



**Gràfic 7.** Correlogrames FAS i FAP de la sèrie transformada amb una primera diferència a la part regular.

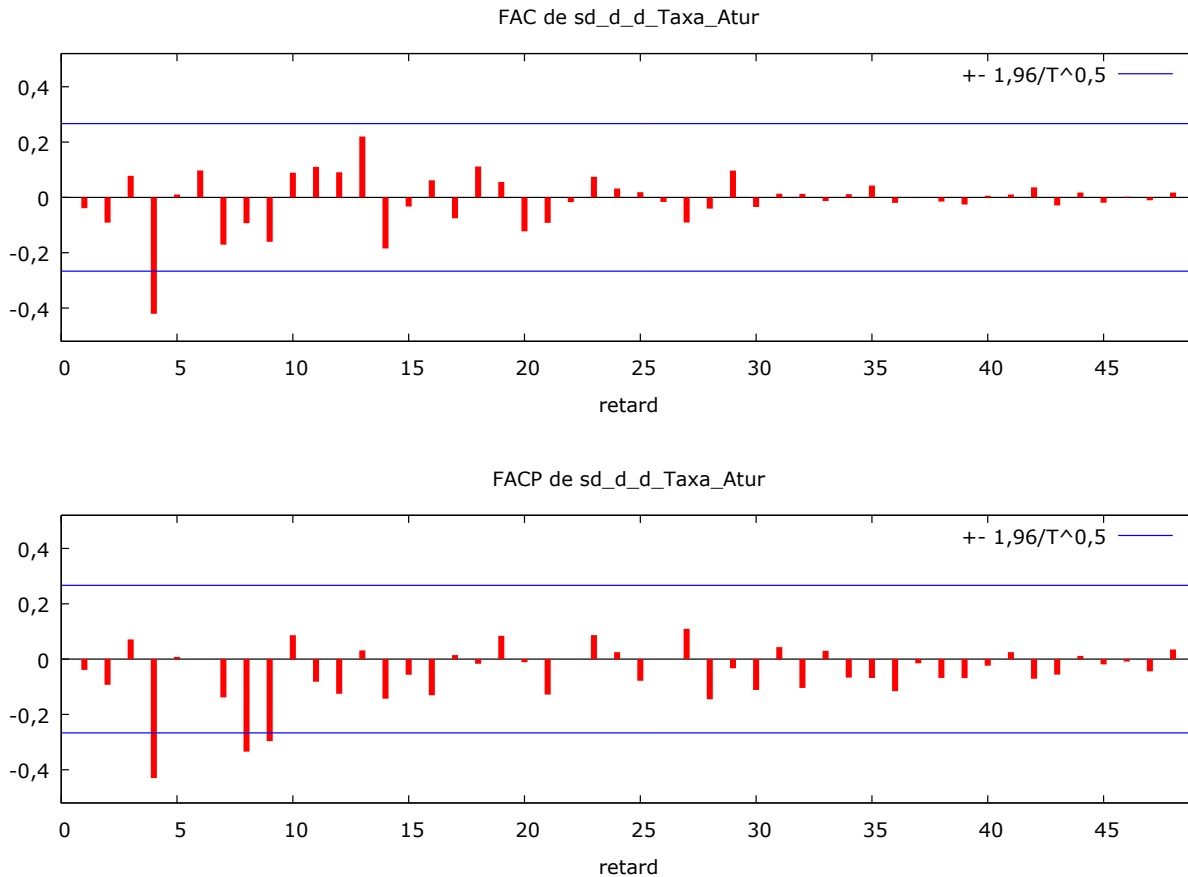
D'una banda, tal i com s'observa al correlograma de la FAS i també de la FAP al *Gràfic 8*, la part regular sembla guanyar estacionarietat donat que els primers retards de la sèrie modificada mostren un decreixement més ràpid. D'altra banda, la part estacional (retards: 4,8,12,...,48) presenta un decreixement lent, indicant la conveniència de prendre diferències estacionals. De manera que caldrà modificar novament la sèrie, però aquesta vegada amb una diferència a la part estacional.



**Gràfic 8.** Correlogrames FAS i FAP de la sèrie transformada amb una segona diferència a la part regular.



Al Gràfic 9 es pot observar com arrel de les successives transformacions de la sèrie original, s'ha aconseguit que aquesta compleixi les condicions d'estacionarietat. Conseqüentment l'orde de  $d$  serà igual a 2 i el de  $D$  serà igual a 1.



**Gràfic 9.** Correlogrames FAS i FAP de la sèrie transformada amb una segona diferència a la part regular i una primera diferència a la part estacional.

Basant-nos en els patrons típics de les funcions d'autocorrelació simple i parcial per a cada model, juntament amb el principi de parsimònia (és preferible un model amb el menor nombre de paràmetres possibles), s'identifica inicialment un model ARIMA  $(0,2,0) \times (0,1,1)_4$ , el qual es correspon amb la següent forma funcional:

$$(1-L)^2 (1-L^4) Y_t = (1-\theta_1 L^4) u_t$$

Per tant podem donar per acabada la primera etapa de la metodologia Box-Jenkins i conseqüentment iniciar la fase d'estimació juntament amb la de validació.

## Segona i tercera etapa: estimació i validació

Per començar s'estimen els paràmetres del model. Després es du a terme el procés de validació, el qual es divideix en tres parts. La primera part tracta de validar la significació dels coeficients estimats així com de la constant. La segona part comprova que es compleixi tant el principi d'estacionarietat com el d'invertibilitat. Finalment, l'última part verifica que els residus estiguin incorrelacionats.

D'entrada, donat que la constant no és significativa per un nivell de significació del 5%, no s'inclourà al model estimat.

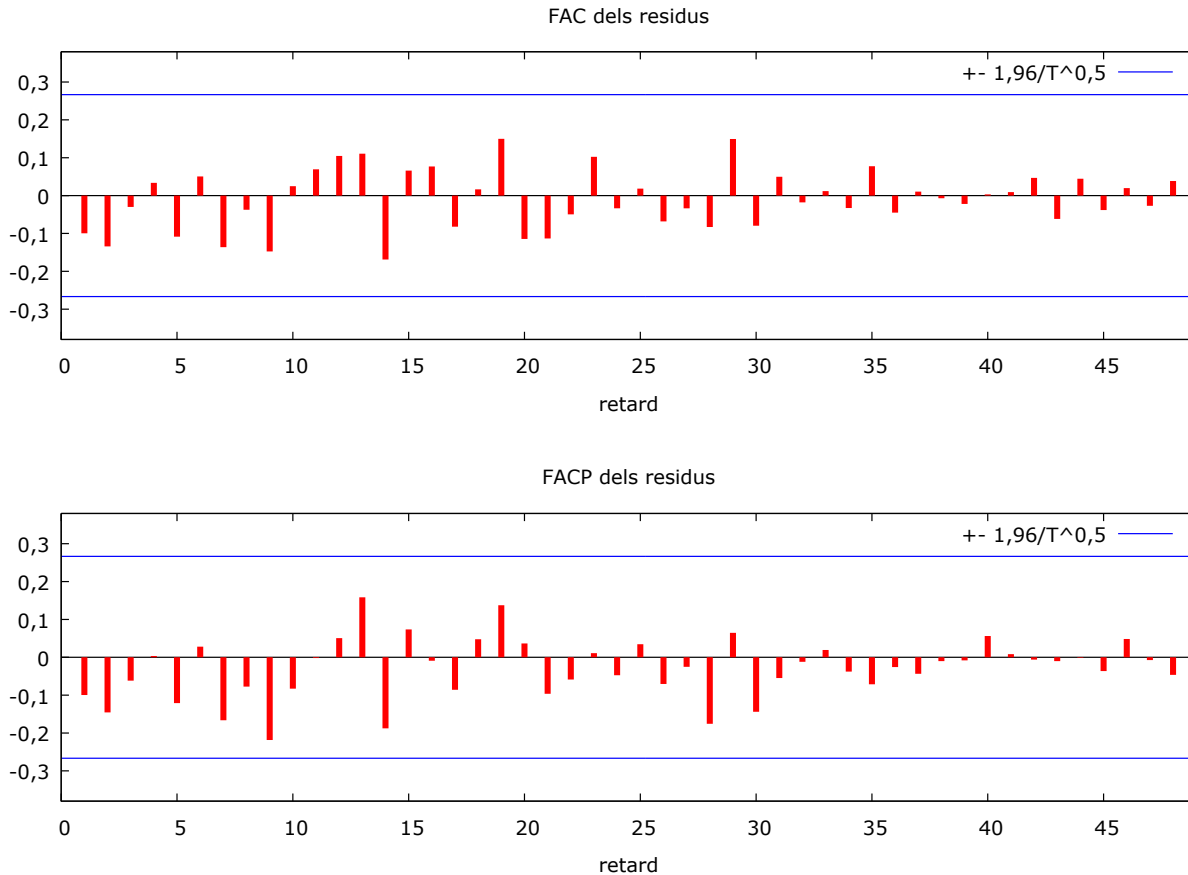
Un cop estimat el model sense constant (definit a la *Taula 3*), podem observar com el paràmetre  $\theta_1$  de la MA, denominat Theta\_1, és eloqüentment diferent de 0 per un nivell de significació del 5%. Per la qual cosa, donaríem per bona la primera part del procés de validació.

<b>Model 2: ARIMA, emprant les observacions 2003:3-2016:4 (T = 54)</b>					
Variable dependent: $(1-L)^2(1-L_s)$ Taxa_Atur					
Desviacions típiques basades en el Hessià					
	<i>Coefficient</i>	<i>Desv. Típica</i>	<i>z</i>	<i>Valor p</i>	
Theta_1	-0,720583	0,135999	-5,2984	<0,00001	***
Mitj. de la vble. dep.	0,005741	D.T. de la vble. dep.		0,634735	
Mitjana de les innovacions	0,000895	D.T. innovacions		0,515023	
Log-versemblança	-42,25591	Criteri d'Akaike		88,51182	
Criteri de Schwarz	92,48979	Crit. de Hannan-Quinn		90,04596	
	<i>Real</i>	<i>Imaginària</i>	<i>Mòdul</i>	<i>Freqüència</i>	
MA (estacional)					
Arrel 1	1,3878	0,0000	1,3878	0,0000	

**Taula 3.** Estimació del model ARIMA  $(0,2,0) \times (0,1,1)_4$  excloent la constant a la regressió.

A continuació, caldrà comprovar que els coeficients estimats compleixin les condicions d'invertibilitat i d'estacionarietat. En aquest cas particular, el model estimat segueix un esquema MA(1), per la qual cosa sempre serà estacionari per definició. Pel que fa la condició d'invertibilitat, el paràmetre  $\theta_1$  (Theta\_1), ha de ser menor que la unitat en valor absolut. Com que  $\theta_1 = |-0.720583| < 1$ , podem afirmar que el model estimat també és invertible.

En darrer terme, per tal de que el model es consideri definitivament vàlid, els errors han de comportar-se segons un procés de soroll blanc (mitjana igual a zero i variància constant). Aquesta tercera fase de l'etapa de validació es du a terme mitjançant l'estudi dels correlogrames de la FAC i FAP del residus.



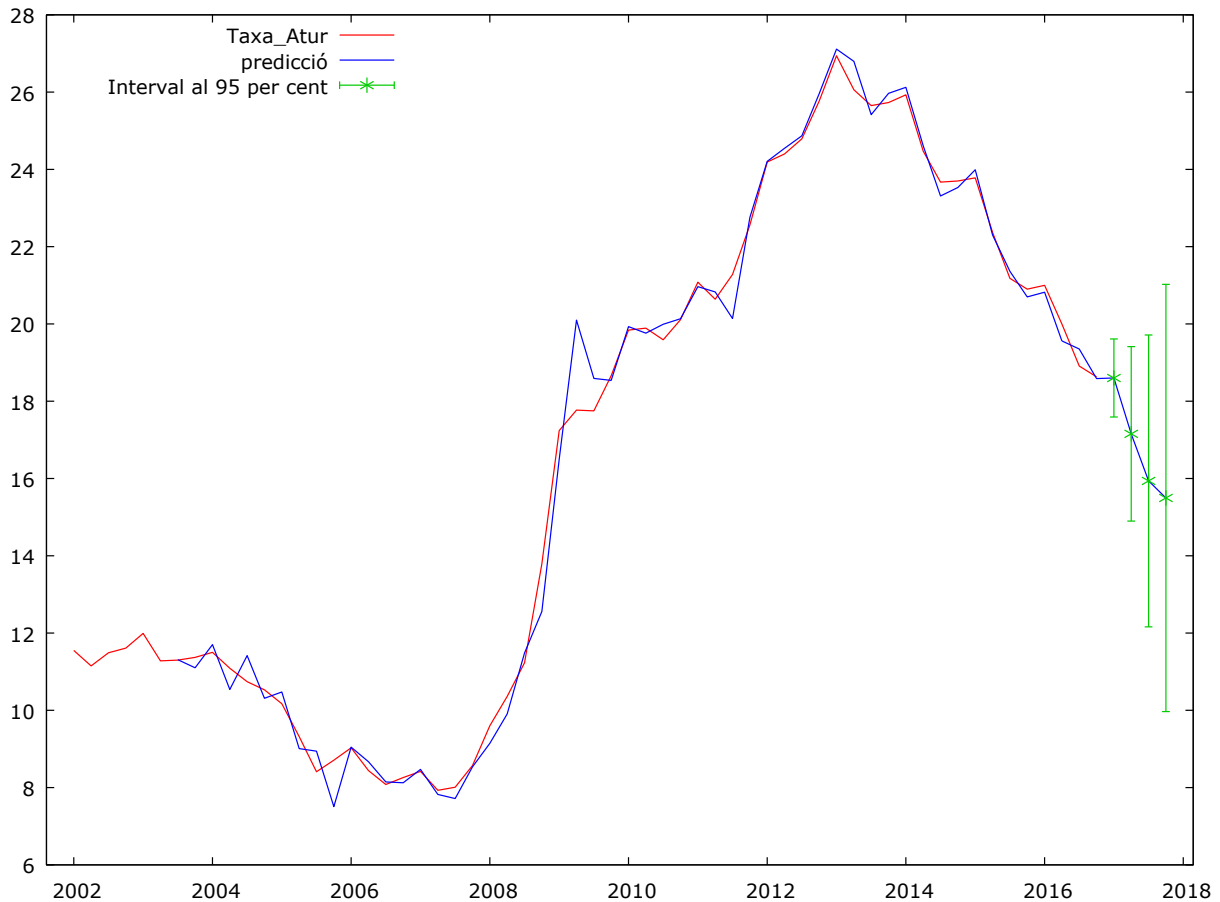
**Gràfic 10.** Correlogrames FAS i FAP dels residus del model estimat.

Els correlogrames de la FAS i de la FAP dels residus (*Gràfic 10*) mostren un comportament propi d'un procés de soroll blanc, tant en la part regular (primeres autocorrelacions) com en la part estacional (autocorrelacions múltiples de 4), donat que cap retard és significatiu. És a dir, no existeix una correlació entre els residus.

Així doncs, donat que el model estimat compleix les tres condicions imprescindibles per tal de poder ser validat, es procedirà a realitzar les respectives prediccions.

## Quarta etapa: predicció

Al Gràfic 11 es presenta l'evolució de la sèrie "taxa d'atur espanyola" conjuntament amb l'estimació i predicció pel període extra-mostrat (els quatre trimestres de l'any 2017).



**Gràfic 11.** Representació de les prediccions del model estimat.

Per a intervals de confiança 95%,  $z(0,025) = 1,96$

Observacions	Taxa_Atur	predicció	Desv. Típica	Interval de 95%
2017:1	indefinit	18,6012	0,515023	(17,5918, 19,6106)
2017:2	indefinit	17,1551	1,15163	(14,8979, 19,4122)
2017:3	indefinit	15,9369	1,92704	(12,1600, 19,7138)
2017:4	indefinit	15,4962	2,82090	(9,96736, 21,0251)

**Taula 4.** Prediccions puntuals, desviacions estàndard i intervals de confiança del model estimat.

A la *Taula 5* es mostren els resultats dels diferents estadístics d'avaluació de la capacitat predictiva obtinguts amb el model estimat anteriorment i pel període extra-mostrat.

<b>EAM(H)</b>	<b>EQM(H)</b>	<b>EPAM(H)</b>
0.43	0.33	2.56%

**Taula 5.** Capacitat predictiva del model estimat.

La metodologia Box Jenkins està basada en part en un anàlisi subjectiu, per la qual cosa els models ARIMA obtinguts poden variar en funció de la interpretació donada per l'analista pertinent. A la *Taula 6* es presenten altres models alternatius avaluats juntament amb els diferents criteris d'informació associats a cadascun d'ells.

<b>Model ARIMA</b>	<b>AIC</b>	<b>HQC</b>	<b>SBIC</b>
$(1,2,1) \times (1,0,1)_4$	94.33960	98.35252	104.6418
$(2,1,1) \times (1,1,0)_4$	94.17533	98.05659	104.2120
$(0,1,2) \times (1,1,0)_4$	98.22812	101.3331	106.2575
$(1,1,0) \times (1,0,0)_4$	100.1307	102.5637	106.3633
$(0,2,0) \times (0,1,1)_4$	88.51182	90.04596	92.48979
$(0,2,0) \times (2,1,0)_4$	92.36809	94.66931	98.33505

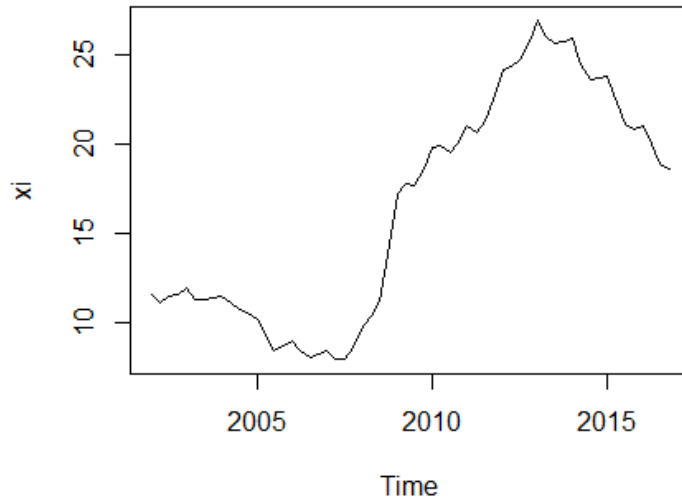
**Taula 6.** Criteris d'informació dels models ARIMA designats.

Tenint en compte que els criteris de selecció AIC (Criteri d'Informació Akaike), HQC (Criteri d'Informació Hannan-Quinn) i SBIC (Criteri d'Informació Schwarz) estan basats en la penalització de paràmetres, un menor valor indica un millor ajust de la regressió. És per això que entre les diferents possibilitats, el model escollit és l'ARIMA  $(0,2,0) \times (0,1,1)_4$ , el qual és el mateix que s'ha generat anteriorment de forma detallada.

## 2.2. Auto-ARIMA

Una altra opció més innovadora per a l'anàlisi i predicció de sèries temporals és mitjançant l'ús d'un procés de selecció automàtica de models implementat en R, conegut com a funció auto-ARIMA (Hyndman i Khandakar, 2008). Concretament aquesta escull entre tots els models possibles aquell que presenta el millor resultat en termes del AICc (Criteri d'Informació Akaike Corregit) i màxima versemblança.

A l'estudi apliquem els passos recomanats per Figuerola (2018). És per això que en aquest subapartat utilitzarem el entorn de desenvolupament integrat *RStudio*.



**Gràfic 12.** Representació de la sèrie original de la taxa d'atur a nivell nacional.

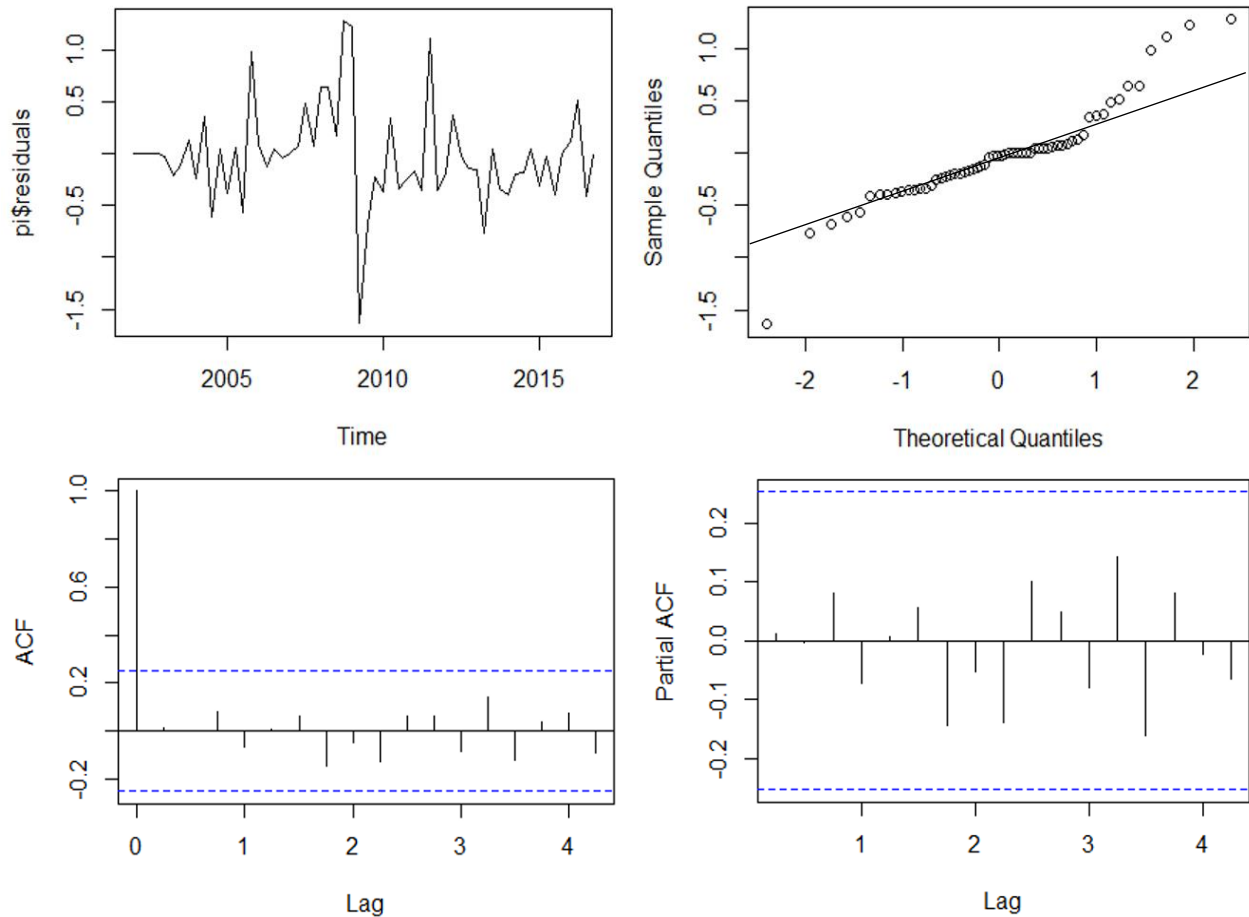
<b>Series: Espanya</b>				
<b>ARIMA(1,1,0)(2,1,0)[4]</b>				
Coefficients:				
	ar1	sar1	sar2	
	0.7418	-0.5232	-0.3185	
s.e.	0.0922	0.1314	0.1224	
sigma^2 estimated as 0.2564: log likelihood=-40.06				
AIC=88.13 AICC=88.93 BIC=96.16				
Training set error measures:				
	ME	RMSE	MAE	MPE
Training set	-0.02449024	0.4713593	0.3185618	0.116403
	MAPE	MASE	ACF1	
Training set	2.172576	0.1592809	0.01138753	

**Taula 7.** Estimació automàtica per la taxa d'atur a nivell nacional.

Partint del output obtingut amb el programa de selecció automàtica (*Taula 7*), l'expressió matemàtica del model dinàmic s'escriuria de la següent manera:

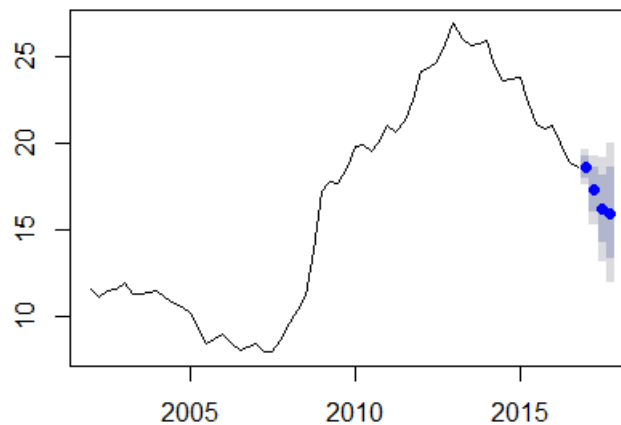
$$(1-0.7418L)(1-0.5232L^4-0.3185L^8)(1-L)(1-L^4)Y_t = u_t$$

Tot i que s'ha utilitzat un programa de selecció automàtica per a l'obtenció del ARIMA òptim, al Gràfic 13 es presenten els correlogrames associats als residus.



**Gràfic 13.** Comportament dels residus del model ARIMA  $(1, 1, 0) \times (2, 1, 0)$  4.

Al Gràfic 14 es mostren les prediccions extra-mostrals generades amb un grau de confiança del 80 i 95% respectivament.



**Gràfic 14.** Representació de les prediccions del model ARIMA  $(1, 1, 0) \times (2, 1, 0)$  4.

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	18.62564	17.97676	19.27452	17.63327	19.61801
2017 Q2	17.29721	15.99397	18.60045	15.30408	19.29034
2017 Q3	16.20642	14.22895	18.18390	13.18213	19.23072
2017 Q4	15.97124	13.32919	18.61329	11.93057	20.01191

**Taula 8.** Prediccions i intervals de confiança del model ARIMA (1,1,0) x (2,1,0) 4.

Un cop obtingudes les prediccions relatives als quatre trimestres de l'any 2017, podem observar a la *Taula 9* com el model ARIMA (1,1,0) x (2,1,0) 4 mostra una molt bona capacitat predictiva, (EPAM del 1.42%).

EAM(H)	EQM(H)	EPAM(H)
0.24	0.10	1.42%

**Taula 9.** Capacitat predictiva del model ARIMA (1,1,0) x (2,1,0) 4.

### 3. Avaluació de la capacitat predictiva

A la *Taula 10* presentem els resultats de l'avaluació de la capacitat predictiva per tots els models implementats en el treball. Arrel dels estadístics s'observa com el model ARIMA seleccionat per la funció auto-ARIMA genera prediccions amb un EAM, EQM i EPAM inferior al model ARIMA seleccionat al subapartat 2.1.. Així mateix aquest model ARIMA també mostra una millor capacitat predictiva que la resta de mètodes deterministes, incloent l'AEH amb un EPAM d'1.79%.

No és cap sorpresa partint de la primícia de que la funció auto-ARIMA processa tots els models possibles (dins de les restriccions proporcionades) mostrant aquell amb el menor AICc i màxima versemblança.

Mètode o model	EAM(H)	EQM(H)	EPAM(H)
TL	8.91	80.86	52.22%
DMM	0.34	0.18	2.00%
AEH	0.30	0.15	1.79%
BJ: ARIMA (0,2,0) x (0,1,1) 4	0.43	0.33	2.56%
R: ARIMA (1,1,0) x (2,1,0) 4	0.24	0.10	1.42%

**Taula 10.** Capacitat predictiva dels mètodes deterministes i dels models ARIMA seleccionats.

Per tant, tenint en compte els resultats obtinguts a la segona part del exercici d'investigació, ens decantaríem per la utilització del programa de selecció automàtica per a les prediccions regionals.



### III. ANÀLISI REGIONAL

En aquest capítol es replica la metodologia automàtica de selecció de models ARIMA explicada al subapartat 2.2. per a cada una de les diferents CCAA de l'Estat espanyol (disset sense incloure Ceuta i Melilla). Un cop obtinguts els respectius models ARIMA s'avalua la bondat predictiva d'aquests, també mitjançant el càlcul dels tres estadístics d'avaluació de la capacitat predictiva descrits a l'Annex 2.

#### ANDALUSIA

```
Series: Andalusia
ARIMA(2,1,0)(1,0,0)[4]

Coefficients:
      ar1      ar2      sar1
    0.2363  0.3037  0.3304
s.e.  0.1223  0.1275  0.1304

sigma^2 estimated as 0.8128:  log likelihood=-76.54
AIC=161.08  AICc=161.82  BIC=169.39
```

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	28.04927	26.89387	29.20468	26.28223	29.81631
2017 Q2	27.91250	26.07533	29.74968	25.10278	30.72222
2017 Q3	27.69972	25.09690	30.30255	23.71905	31.68040
2017 Q4	27.62486	24.32720	30.92252	22.58152	32.66819

EAM(H)	EQM(H)	EPAM(H)
2.32	5.96	9.20%

## ARAGÓ

Series: Aragó  
ARIMA(0,1,1)(0,1,2) [4]

Coefficients:

	ma1	sma1	sma2
	0.2757	-0.5834	-0.3312
s.e.	0.1384	0.2662	0.1945

sigma<sup>2</sup> estimated as 0.8879: log likelihood=-76.3  
AIC=160.61 AICc=161.41 BIC=168.64

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	14.32620	13.10573	15.54667	12.45965	16.19275
2017 Q2	14.12749	12.15095	16.10402	11.10464	17.15034
2017 Q3	13.45047	10.93574	15.9652	9.604528	17.29641
2017 Q4	13.42645	10.46993	16.38297	8.904846	17.94805

EAM(H)	EQM(H)	EPAM(H)
2.19	5.36	19.44%

## CANTÀBRIA

Series: Cantàbria  
ARIMA(1,0,0)(0,1,1) [4]

Coefficients:

	ar1	sma1
	0.9450	-0.7012
s.e.	0.0431	0.1481

sigma<sup>2</sup> estimated as 1.683: log likelihood=-94.67  
AIC=195.34 AICc=195.8 BIC=201.42

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	14.30929	12.64671	15.97187	11.76660	16.85199
2017 Q2	13.04692	10.75941	15.33443	9.548480	16.54536
2017 Q3	11.35061	8.623487	14.07774	7.179834	15.52139
2017 Q4	12.49078	9.423870	15.55770	7.800343	17.18123

EAM(H)	EQM(H)	EPAM(H)
0.89	0.94	6.63%

## CASTELLA I LLEÓ

Series: Castella i Lleó  
ARIMA(2,0,1)(1,1,1)[4]

Coefficients:

	ar1	ar2	ma1	sar1	sma1
	1.8588	-0.8774	-0.7042	0.1905	-0.7272
s.e.	0.1373	0.1309	0.2220	0.2378	0.1895

sigma<sup>2</sup> estimated as 0.5272: log likelihood=-60.91  
AIC=133.83 AICc=135.54 BIC=145.98

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	15.50679	14.57625	16.43733	14.08365	16.92993
2017 Q2	13.95212	12.53080	15.37343	11.77840	16.12583
2017 Q3	12.21704	10.36938	14.06469	9.391295	15.04278
2017 Q4	13.20713	10.97539	15.43887	9.793977	16.62028

EAM(H)	EQM(H)	EPAM(H)
0.55	0.32	3.94%

## CASTELLA LA MANXA

Series: Castella la Manxa  
ARIMA(2,0,1)(2,1,0)[4]

Coefficients:

	ar1	ar2	ma1	sar1	sar2
	1.8158	-0.8386	-0.597	-0.5900	-0.3530
s.e.	0.1276	0.1219	0.193	0.1308	0.1397

sigma<sup>2</sup> estimated as 1.017: log likelihood=-79.31  
AIC=170.61 AICc=172.32 BIC=182.76

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	22.88679	21.59420	24.17938	20.90994	24.86363
2017 Q2	21.44512	19.40729	23.48294	18.32854	24.56170
2017 Q3	20.55510	17.85149	23.25871	16.42029	24.68992
2017 Q4	20.70564	17.39822	24.01306	15.64738	25.76391

EAM(H)	EQM(H)	EPAM(H)
1.03	1.39	5.20%

## CATALUNYA

Series: Catalunya  
ARIMA(1,1,0)(2,0,0)[4]

Coefficients:

	ar1	sar1	sar2
	0.4214	0.3188	0.2639
s.e.	0.1199	0.1234	0.1291

sigma<sup>2</sup> estimated as 0.7917: log likelihood=-76.1  
AIC=160.2 AICc=160.94 BIC=168.51

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	14.76976	13.62946	15.91006	13.02582	16.51370
2017 Q2	14.03934	12.05762	16.02106	11.00856	17.07012
2017 Q3	13.18897	10.49613	15.88182	9.070618	17.30733
2017 Q4	13.32048	10.01988	16.62108	8.272644	18.36832

EAM(H)	EQM(H)	EPAM(H)
0.67	0.47	5.09%

## COMUNITAT DE MADRID

Series: Comunitat de Madrid  
ARIMA(0,1,0)(1,0,0)[4]

Coefficients:

	sar1
	0.3002
s.e.	0.1244

sigma<sup>2</sup> estimated as 0.7011: log likelihood=-72.93  
AIC=149.85 AICc=150.07 BIC=154.01

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	14.69005	13.61700	15.76310	13.04896	16.33113
2017 Q2	14.52196	13.00444	16.03948	12.20111	16.84280
2017 Q3	14.20379	12.34521	16.06236	11.36134	17.04623
2017 Q4	14.02669	11.88059	16.17279	10.74452	17.30886

EAM(H)	EQM(H)	EPAM(H)
1.02	1.48	7.91%

## EXTREMADURA

Series: Extremadura  
ARIMA(3,0,0)(2,1,0)[4]

Coefficients:

	ar1	ar2	ar3	sar1	sar2
	1.1975	0.0976	-0.3489	-0.9323	-0.5869
s.e.	0.1345	0.2187	0.1334	0.1145	0.1125

sigma<sup>2</sup> estimated as 1.578: log likelihood=-92.99  
AIC=197.98 AICc=199.7 BIC=210.14

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	29.00995	27.39983	30.62007	26.54748	31.47241
2017 Q2	28.19819	25.68622	30.71017	24.35646	32.03992
2017 Q3	27.27940	23.75924	30.79956	21.89578	32.66302
2017 Q4	29.25174	24.88759	33.61589	22.57735	35.92613

EAM(H)	EQM(H)	EPAM(H)
2.33	7.35	9.21%

## GALÍCIA

Series: Galicia  
ARIMA(2,0,1)(2,1,0)[4]

Coefficients:

	ar1	ar2	ma1	sar1	sar2
	1.9291	-0.9469	-0.8491	-0.5816	-0.4769
s.e.	0.0563	0.0519	0.1373	0.1282	0.1344

sigma<sup>2</sup> estimated as 0.64: log likelihood=-66.76  
AIC=145.53 AICc=147.24 BIC=157.68

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	17.12319	16.09798	18.14840	15.55527	18.69111
2017 Q2	16.25694	14.74797	17.76590	13.94917	18.56470
2017 Q3	14.49721	12.59080	16.40362	11.58161	17.41281
2017 Q4	14.88261	12.63046	17.13475	11.43825	18.32696

EAM(H)	EQM(H)	EPAM(H)
0.14	0.03	0.87%

## ILLES CANÀRIES

Series: Illes Canàries  
ARIMA(0,1,1)(0,0,1)[4]

Coefficients:

	ma1	sma1
	0.2285	0.5373
s.e.	0.1223	0.1132

sigma<sup>2</sup> estimated as 1.347: log likelihood=-92.19  
AIC=190.39 AICc=190.82 BIC=196.62

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	24.36493	22.87759	25.85228	22.09023	26.63963
2017 Q2	25.32346	22.96743	27.67949	21.72022	28.92670
2017 Q3	25.12913	22.14760	28.11066	20.56927	29.68899
2017 Q4	24.79012	21.29323	28.28701	19.44209	30.13815

EAM(H)	EQM(H)	EPAM(H)
2.09	5.24	9.18%

## ILLES BALEARS

Series: Illes Balears  
ARIMA(0,1,0)(0,1,1)[4]

Coefficients:

	sma1
	-0.5460
s.e.	0.1442

sigma<sup>2</sup> estimated as 2.744: log likelihood=-106  
AIC=216.01 AICc=216.24 BIC=220.02

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	16.39426	14.27132	18.51719	13.14750	19.64101
2017 Q2	10.46824	7.465947	13.47052	5.876632	15.05984
2017 Q3	7.834338	4.157302	11.51137	2.210797	13.45788
2017 Q4	11.36484	7.118966	15.61072	4.871335	17.85835

EAM(H)	EQM(H)	EPAM(H)
1.01	1.18	9.06%

## LA RIOJA

Series: La Rioja  
ARIMA(0,1,0)

sigma<sup>2</sup> estimated as 1.962: log likelihood=-103.6  
AIC=209.19 AICc=209.26 BIC=211.27

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	10.90	9.105042	12.69496	8.154848	13.64515
2017 Q2	10.90	8.361546	13.43845	7.017769	14.78223
2017 Q3	10.90	7.791041	14.00896	6.145258	15.65474
2017 Q4	10.90	7.310084	14.48992	5.409697	16.39030

EAM(H)	EQM(H)	EPAM(H)
1.09	1.85	8.69%

## NAVARRA

Series: Navarra  
ARIMA(0,1,0)(1,0,1)[4]

Coefficients:

sar1	sma1
0.9066	-0.6980
s.e. 0.1197	0.2224

sigma<sup>2</sup> estimated as 1.184: log likelihood=-88.45  
AIC=182.91 AICc=183.34 BIC=189.14

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	10.76321	9.368925	12.15750	8.630833	12.89559
2017 Q2	9.899489	7.927669	11.87131	6.883851	12.91513
2017 Q3	9.647624	7.232648	12.06260	5.954238	13.34101
2017 Q4	9.211683	6.423110	12.00026	4.946929	13.47644

EAM(H)	EQM(H)	EPAM(H)
0.61	0.41	5.95%

## PAÍS BASC

Series: País Basc  
ARIMA(1,1,0)

Coefficients:

ar1  
0.2667  
s.e. 0.1248

sigma<sup>2</sup> estimated as 0.627: log likelihood=-69.48  
AIC=142.96 AICc=143.18 BIC=147.12

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	12.12865	11.11383	13.14346	10.57662	13.68067
2017 Q2	12.09094	10.45318	13.72871	9.586198	14.59569
2017 Q3	12.08089	9.953566	14.20821	8.827428	15.33435
2017 Q4	12.07821	9.544163	14.61225	8.202720	15.95370

EAM(H)	EQM(H)	EPAM(H)
0.79	0.84	7.18%

## PAÍS VALENCIÀ

Series: País Valencià  
ARIMA(2,1,2)(1,0,1)[4]

Coefficients:

ar1 ar2 ma1 ma2 sar1 sma1  
1.3157 -0.5657 -1.0946 0.6458 0.9541 -0.8189  
s.e. 0.2627 0.3338 0.3159 0.3734 0.1103 0.2419

sigma<sup>2</sup> estimated as 0.8156: log likelihood=-75.91  
AIC=165.81 AICc=168.01 BIC=180.36

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	19.36907	18.21143	20.52672	17.59861	21.13954
2017 Q2	18.59628	16.76919	20.42338	15.80199	21.39058
2017 Q3	18.02095	15.42573	20.61616	14.05191	21.98998
2017 Q4	17.42541	13.98190	20.86892	12.15901	22.69180

EAM(H)	EQM(H)	EPAM(H)
0.42	0.22	2.36%



## PRINCIPAT D'ASTÚRIES

Series: Principat d'Astúries  
ARIMA(0,1,0)

sigma<sup>2</sup> estimated as 2.21: log likelihood=-107.1  
AIC=216.21 AICc=216.28 BIC=218.29

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	14.59	12.68504	16.49496	11.67661	17.50339
2017 Q2	14.59	11.89597	17.28403	10.46984	18.71016
2017 Q3	14.59	11.29051	17.88949	9.543859	19.63614
2017 Q4	14.59	10.78007	18.39993	8.763219	20.41678

EAM(H)	EQM(H)	EPAM(H)
0.90	1.31	6.88%

## REGIÓ DE MÚRCIA

Series: Regió de Múrcia  
ARIMA(1,1,0)(2,0,0)[4]

Coefficients:

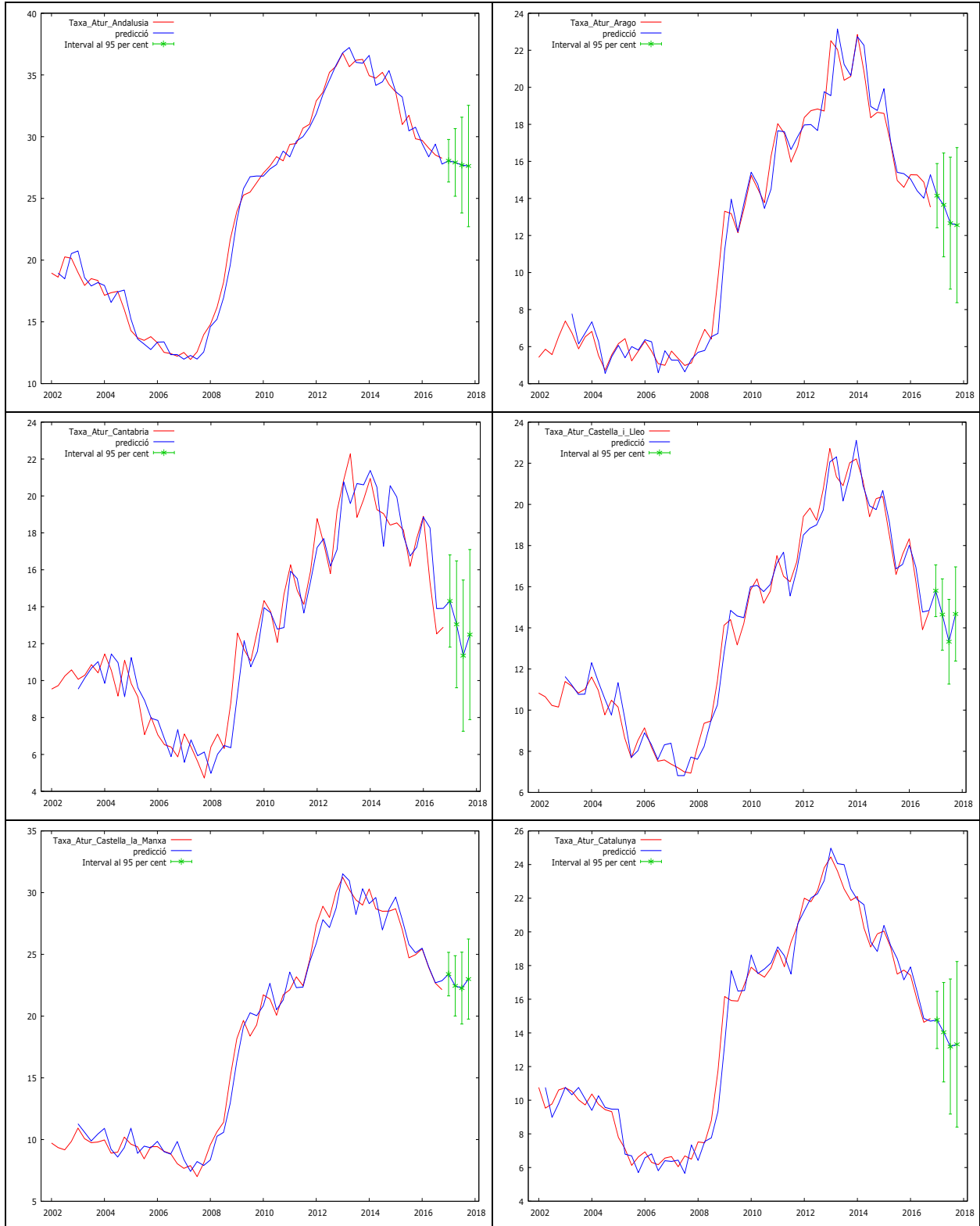
	ar1	sar1	sar2
	0.2371	0.2570	0.3832
s.e.	0.1341	0.1263	0.1362

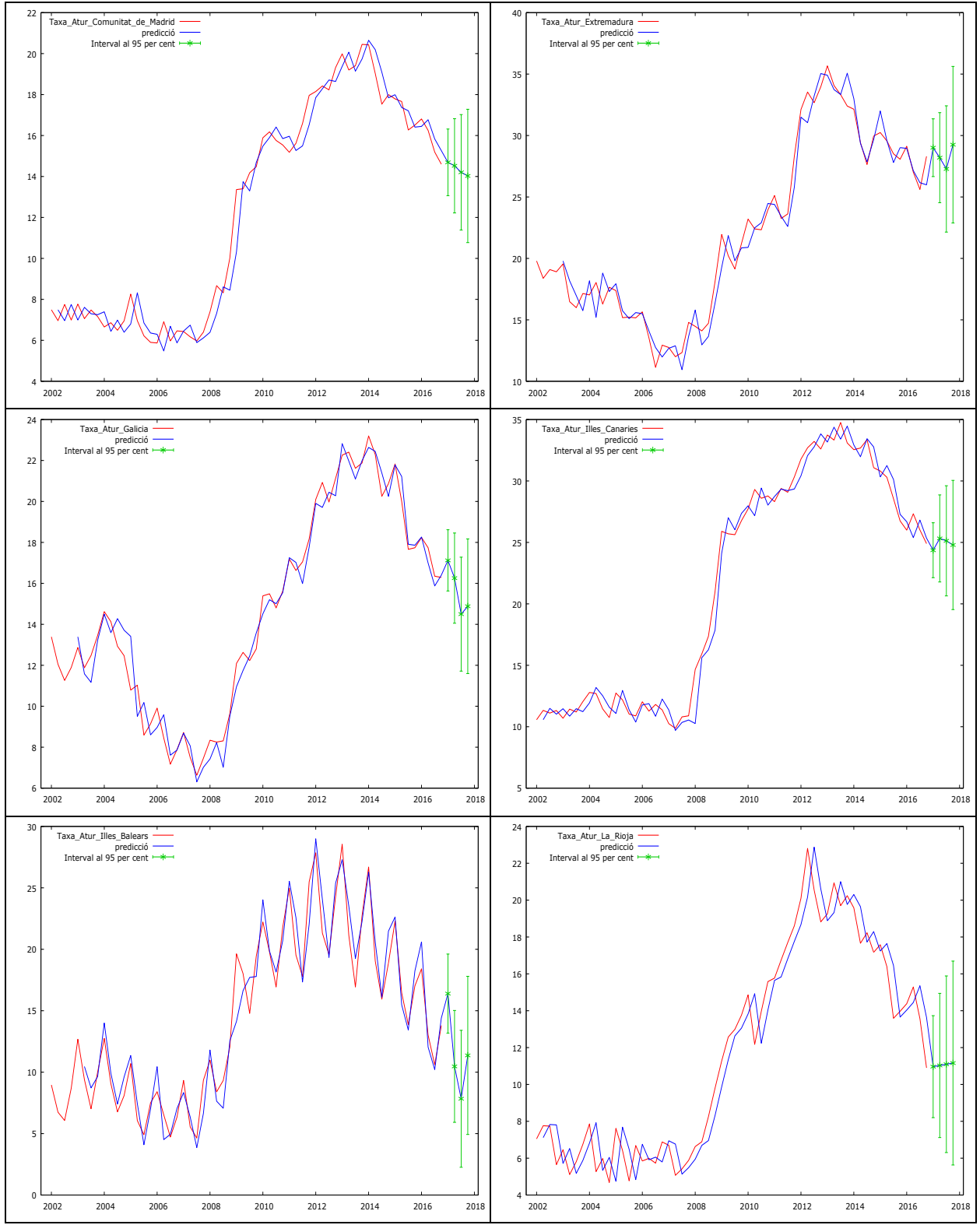
sigma<sup>2</sup> estimated as 1.326: log likelihood=-91.55  
AIC=191.1 AICc=191.84 BIC=199.41

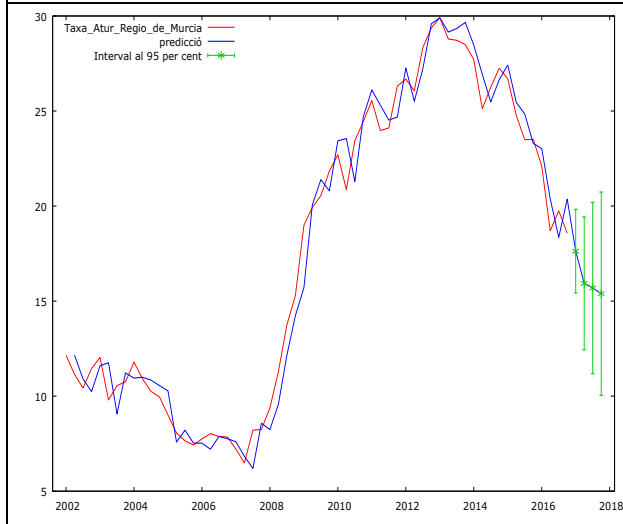
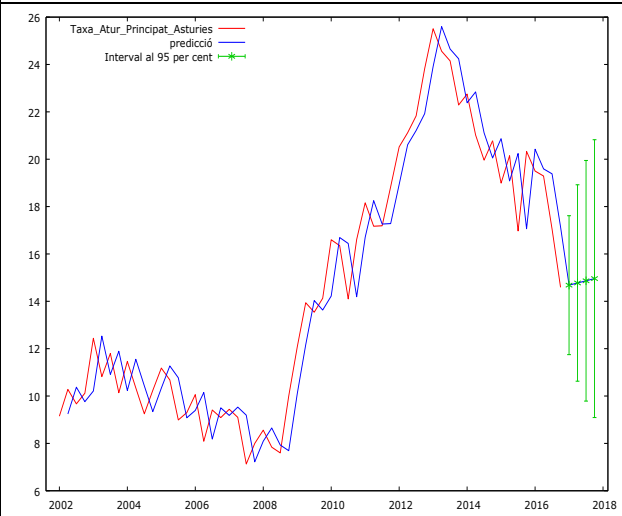
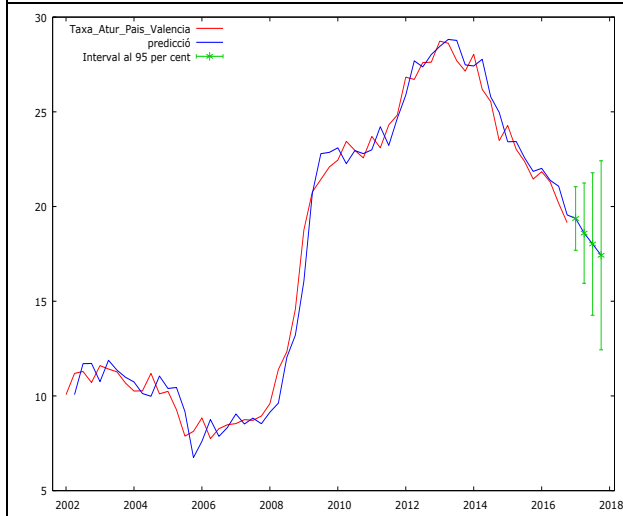
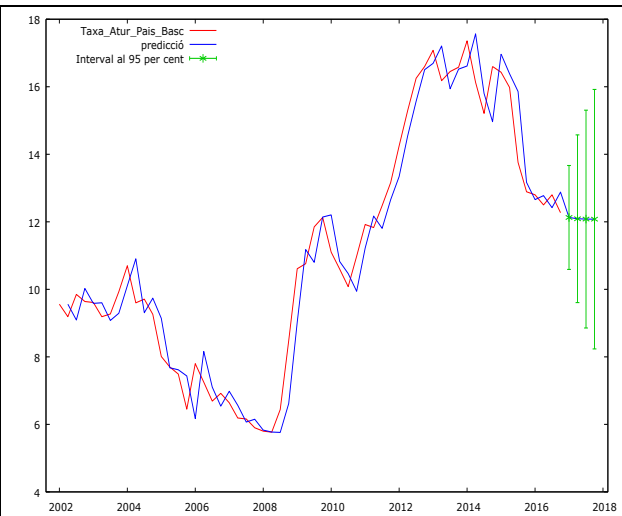
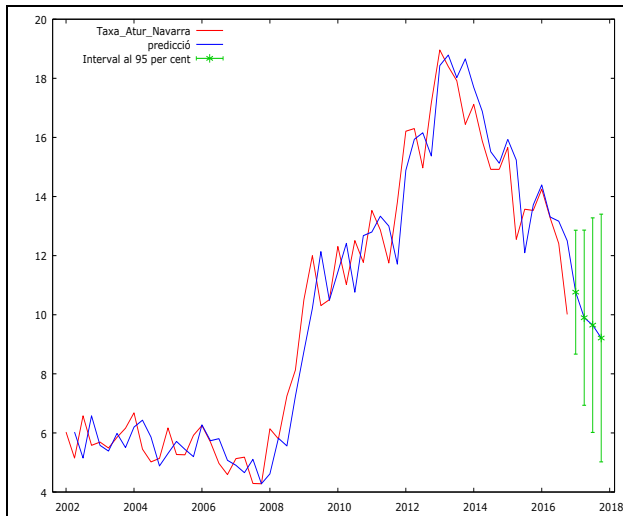
Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017 Q1	17.62886	16.15305	19.10468	15.37180	19.88593
2017 Q2	15.93154	13.58395	18.27914	12.34121	19.52188
2017 Q3	15.69016	12.66458	18.71573	11.06293	20.31738
2017 Q4	15.39217	11.80435	18.98000	9.905067	20.87928

EAM(H)	EQM(H)	EPAM(H)
1.87	3.61	10.39%

Una vegada seleccionat el model ARIMA òptim generem les prediccions extra-mostrals (altre cop els quatre trimestres de l'any 2017). Els resultats es presenten a continuació:







Els resultats exposats a la *Taula 11* indiquen que la millor capacita predictiva s'obté per Galícia i el País Valencià respectivament (ambdós amb un EPAM inferior al 3%). En canvi Aragó, amb un EPAM de casi del 20%, mostra les pitjors estimacions.

També s'observa com rarament les regions de La Rioja, el País Basc i el Principat d'Astúries són les úniques que no mostren patrons estacionals.

Hi ha tres models que es repeteixen: l'ARIMA  $(2,0,1) \times (2,1,0)_4$ , l'ARIMA  $(1,1,0) \times (2,0,0)_4$  i cap darrer l'ARIMA  $(0,1,0)$ . Concretament el primer model és l'òptim per Castella la Manxa i Galícia, el segon per Catalunya i la Regió de Murcia i, l'últim per La Rioja i el Principat d'Astúries.

<b>CCAA</b>	<b>Model ARIMA</b>	<b>EPAM(H)</b>
Andalusia	$(2,1,0) \times (1,0,0)_4$	9.20%
Aragó	$(0,1,1) \times (0,1,2)_4$	19.44%
Cantàbria	$(1,0,0) \times (0,1,1)_4$	6.63%
Castella i Lleó	$(2,0,1) \times (1,1,1)_4$	3.94%
Castella la Manxa	$(2,0,1) \times (2,1,0)_4$	5.20%
Catalunya	$(1,1,0) \times (2,0,0)_4$	5.09%
Comunitat de Madrid	$(0,1,0) \times (1,0,0)_4$	7.91%
Extremadura	$(3,0,0) \times (2,1,0)_4$	9.21%
Galícia	$(2,0,1) \times (2,1,0)_4$	0.87%
Illes Canàries	$(0,1,1) \times (0,0,1)_4$	9.18%
Illes Balears	$(0,1,0) \times (0,1,1)_4$	9.06%
La Rioja	$(0,1,0)$	8.69%
Navarra	$(0,1,0) \times (1,0,1)_4$	5.95%
País Basc	$(1,1,0)$	7.18%
País Valencià	$(2,1,2) \times (1,0,1)_4$	2.36%
Principat d'Astúries	$(0,1,0)$	6.88%
Regió de Múrcia	$(1,1,0) \times (2,0,0)_4$	10.39%

**Taula 11.** Model ARIMA òptim i EPAM de cada CCAA.

El model ARIMA escollit a nivell nacional com a òptim (ARIMA  $(1,1,0) \times (2,1,0)_4$ ) no coincideix amb ningun dels obtinguts a nivell regional. En conseqüència, trobem que per tal de millorar les prediccions de la taxa d'atur caldria analitzar de forma independent cada regió que engloba l'Estat espanyol.

## IV. CONCLUSIONS

El present estudi es centra en la modelització de la taxa d'atur a Espanya. Per la qual cosa es combina un anàlisi determinista amb un altre d'estocàstica.

Arran dels resultats obtinguts, es troba que les prediccions més precises han estat generades a través d'un model estocàstic i no pas per un mètode determinista.

A nivell nacional, l'anàlisi determinista no detecta la presència d'un comportament estacional significatiu, la qual cosa sembla contradictòria pel fet de que la sèrie temporal tractada és la taxa d'atur, on el sector turístic té un pes relatiu molt rellevant en l'economia. En canvi, els models estocàstics sí que incorporen patrons estacionals.

El mètode determinista de les DMM i de l'AEH, tot i ser menys precisos, presenten una molt bona capacitat predictiva. En canvi, el mètode determinista de la TL resulta tenir una capacitat predictiva molt fluixa, degut al punt de gir que es produeix a l'any 2013. És per això que els mètodes d'estructura fixa es veuen significativament penalitzats.

El procés estocàstic ARIMA obtingut amb el programa de selecció automàtica (ARIMA  $(1,1,0) \times (2,1,0)_4$ ) presenta una bondat predictiva més elevada en comparació amb el procés estocàstic ARIMA aconseguit a través de la metodologia Box-Jenkins (ARIMA  $(0,2,0) \times (0,1,1)_4$ ). Aquest fet s'explica en part pel caràcter subjectiu que integra aquesta darrera metodologia. En canvi, el programa auto-ARIMA no incorpora cap aspecte subjectiu, ja que de tots els models possibles escull el més acurat, en funció del quocient AICc i màxima versemblança i, de manera totalment automatitzada.

Així és que s'ha utilitzat de manera exclusiva la funció auto-ARIMA en el procés de generació de les prediccions a nivell nacional.

Respecte als resultats obtinguts en l'anàlisi regional, Galícia i el País Valencià presenten les prediccions més acurades de la taxa d'atur respectivament, a diferència de la comunitat d'Aragó, en la qual s'obtenen les prediccions menys precises. Així mateix, de les disset CCAA ninguna comparteix un model ARIMA idèntic al escollit com a òptim nivell nacional i, només Galícia presenta una capacitat predictiva més bona. Finalment hi ha tres CCAA amb un model ARIMA sense estacionalitat, les quals són: La Rioja, el País Basc i el Principat d'Astúries.

Arrel del anàlisi es troba que de cara a la millora de les prediccions sobre l'atur fora necessari realitzar una selecció de model a nivell regional. Això pot ser degut a les diferents característiques del mercat laboral a cada CCAA.

## BIBLIOGRAFIA

- Aznar, A., & Trivez, F. J. (1993). *Métodos de predicción en economía I; Fundamentos, Input-Output, Modelos econométricos y métodos no paramétricos de series temporales*. Barcelona: Ariel Economía.
- Box, G. E., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, 26, 211-252. Recollit de <http://links.jstor.org/sici?sici=0035-9246%281964%2926%3A2%3C211%3AAAOT%3E2.0.CO%3B2-6>
- Brockwell, P. J., & Davis, R. A. (2002). *Introduction to Time Series and Forecasting* (2nd ed.). United States of America: Springer.
- Chatfield, C. (2004). *The Analysis of Time Series; An Introduction* (6th ed.). United States of America: CHAMPMAN & HALL/CRC.
- Claveria, O., & Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, 36, 220-228.
- Diebold, F. X. (2007). *Elements of Forecasting* (4th ed.). United States of America: Thomson South-Western.
- Figuerola, A. (2018). Forecasting with R. Auto-ARIMA Step by Step. *LinkedIn Publisher*. Recollit de <https://www.linkedin.com/pulse/forecasting-r-auto-arima-step-anton-figuerola-wischke?articleId=6385568602233413633#comments-6385568602233413633&trk=prof-post>
- Hassani, H., Sirimal Silva, E., Antonakakis, N., Filis, G., & Gupta, R. (2017). Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research*, 63, 112-127.
- Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast. *Journal of Statistical Software*, 27(3). Recollit de <http://www.jstatsoft.org/>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.
- Manjón, P. L. (2013). *Corporación de Radio y Televisión Española (rtve)*. Obtenido de El Paro en España: <http://www.rtve.es/noticias/20130425/paro-llega-6202700-personas-tasa-2716-primer-trimestre-2013/649182.shtml>
- Peña, D. (2005). *Análisis de series temporales*. Madrid, Spain: Alianza Editorial.
- Pérez Ramírez, F. O. (2006). Concepto de proceso estocástico. A *Introducción a las series de tiempo; Métodos paramétricos* (p. 13-17). Universidad de Medellín .
- Snyder, R. D., Koehler, A. B., Hyndman, R. J., & Ord, J. K. (2004). Exponential smoothing models: Means and variances for lead-time demand. *European Journal of Operational Research*, 158, 444-455.
- Trivez, F. J. (1994). Efectos de los distintos tipos de outliers en las predicciones de los Modelos ARIMA. *Estadística Española*, 36, 21-58.

## ANNEXOS

### Annex 1. Contrastos estadístics

#### Contrast de Daniel:

$$\begin{cases} H_A; \text{sèrie sense tendència} \\ H_0; \text{sèrie amb tendència} \end{cases}$$

$$Z = \sqrt{T-1} \left( 1 - \frac{6 \sum_{t=1}^T d_t^2}{T(T^2-1)} \right)$$

on  $Z$  és l'estadístic de prova,  $T$  la grandària mostral i,  $d_t$  la diferència entre el rang ascendent temporal i el rang que resultat d'una ordenació de menor a major de la variable.

#### Contrast de Kruskal-Wallis:

$$\begin{cases} H_A; \text{sèrie sense component estacional} \\ H_0; \text{sèrie amb component estacional} \end{cases}$$

$$H = \frac{12}{T(T+1)} \left[ \sum_{i=1}^s \frac{R_i^2}{T_i} \right] - 3(T+1)$$

on  $H$  és l'estadístic de prova,  $s$  el nombre de períodes estacionals anuals,  $T_i$  el nombre d'observacions corresponents a l'estació  $i$ -èsima i,  $R_i$  la suma dels rangs per a l'estació  $i$ -èsima que resulten d'una ordenació de menor a major de la variable.

### Annex 2. Mesures de qualitat relativa

#### EAM:

$$EAM(H) = \frac{1}{H} \sum_{m=1}^H |e_T(m)| = \frac{1}{H} \sum_{m=1}^H |Y_{T+m} - \hat{Y}_T(m)|$$

on  $H$  és el nombre de períodes extra-mostrals,  $T$  la grandària mostral,  $e$  l'error de predicció que s'obté com  $e_T = Y_T - \hat{Y}_T$  i,  $m$  fa referència a un període extra-mostrals concret.

#### EPAM:

$$EPAM(H) = \frac{100}{H} \sum_{m=1}^H \left| \frac{e_T(m)}{Y_{T+m}} \right|$$

#### EQM:

$$EQM(H) = \frac{1}{H} \sum_{m=1}^H e_T^2(m) = \frac{1}{H} \sum_{m=1}^H (Y_{T+m} - \hat{Y}_T(m))^2$$



### Annex 3. Mètodes deterministes

#### TL:

$$\hat{Y}_t = \hat{T}_t + u_t$$

on  $\hat{T}_t$  és la tendència,  $\hat{T}_t = \hat{\beta}_0 + \hat{\beta}_1(t)$

partint dels paràmetres  $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{t}$  i,  $\hat{\beta}_1 = \frac{\sum_{t=1}^T t Y_t - \bar{Y} \sum_{t=1}^T t}{\sum_{t=1}^T t^2 - t \sum_{t=1}^T t}$

on  $u_t$  és el terme de error o pertorbació aleatori/a

#### DMM:

$$\hat{Y}_t = \hat{T}_t + \hat{\beta}_1(t)$$

on  $\hat{T}_t$  és la tendència,  $\hat{T}_t = 2MM_t - MM'_t$

partint de la mitjana mòbil  $MM_t = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k}$  i,

de la doble mitjana mòbil  $MM'_t = \frac{MM_t + MM_{t-1} + \dots + MM_{t-k}}{k}$

on  $\hat{\beta}_1$  és la pendent,  $\hat{\beta}_1(t) = \frac{2MM_t - 2MM'_t}{K-1}$

on  $t$  fa referència a un període mostral concret

#### AEH:

$$\hat{Y}_t = \hat{T}_t + \hat{\beta}_1(t)$$

on  $\hat{T}_t$  és la tendència,  $\hat{T}_t = \alpha Y_t + (1 - \alpha) \hat{Y}_{t-1}(1)$

on  $\hat{\beta}_1$  és la pendent,  $\hat{\beta}_1(t) = \gamma(\hat{T}_t - \hat{T}_{t-1}) + (1 - \gamma)\hat{\beta}_1(T - 1)$