Factor Modelling
Predint el PIB d'Espanya

2022-2023

Tutor: José García Montalvo

Albert Español, Leonardo Barni, Roger Casas,
Danae Jordan, Lenka Ferrer

Your Website Here

RESUM EXECUTIU

El Producte Interior Brut (PIB) és avui en dia un dels majors indicadors per mesurar la potència d'una economia. És dels primers indicadors amb el que els estudiants d'Economia es topen al primer any de carrera, i el trobem present en milers d'informes econòmics, notícies, estudis i investigacions d'arreu del món.

És una eina necessària no només per mesurar la capacitat productiva d'una nació, sinó també per dur a terme comparacions entre països i, el més important per prendre decisions polítiques. D'aquí neix la motivació del nostre estudi, on ens plantegem si podem elaborar un model economètric que predigui el PIB d'Espanya mensualment amb gran rigor, velocitat i precisió per tal de garantir una millor presa de decisions en l'esfera política.

A través del Factor Modelling hem sigut capaços de crear el nostre model, capaç de superar l'estimació que fa el model Autoregressiu de primer orde (AR (1)), presentant el nostre un R^2 del 97%. A més, és rellevant fer menció de l'avantatge que el nostre model introdueix, consistent en ser capaç de predir el PIB amb dades mensuals que són publicades amb més freqüència.

Per arribar a aquests resultats primer vam haver de construir la nostra base de dades, recopilant aquestes de diverses fonts com l'Institut Nacional d'Estadística o el Ministeri de Treball i Economia Social. Seguidament, vam haver de tractar les dades recopilades per tal que fossin estacionàries i, pel fet que les dades del PIB a Espanya eren trimestrals i les nostres mensuals, vam haver de dur a terme la tècnica del Blocking Approach. A banda del tractament i processament de dades, hem realitzat diverses proves estadístiques per tal d'arribar a resultats més rigorosos com el Test Dickey-Fuller o el Diebold-Mariano.

Finalment, amb els resultats obtinguts és fàcilment observable com hem sigut capaços d'idear el model que buscàvem i superar les nostres expectatives en el camí. Proposant no només un model sinó una nova eina de gran utilitat per prendre decisions polítiques basades en prediccions precises del PIB mensual d'Espanya.

ÍNDEX DE CONTINGUTS

1. INTRODUCCIÓ, HIPÒTESI I OBJECTIUS DE L'ESTUDI **1**

2. METODOLOGIA **1**

2.1. Recopilació de les dades 1

2.2. Tractament de les dades 2

2.2.1. Estacionarietat de les variables: 2

2.2.2. Blocking Approach 6

2.3. Estimació i criteris de selecció de factors 7

2.4. Resultats 9

2.5. Anàlisi del Component cíclic de les sèries generades pels dos models 10

2.5.1. Estimació de la tendència: el Hodrick-Prescott filter 11

2.5.2. Comparació dels cicles resultants 12

2.6. Avaluació pseudo-out-of-sample 15

2.6.1. El contrast Diebold-Mariano per a la diferència d'error de predicció 15

3. CONCLUSIONS **17**

4. BIBLIOGRAFIA CITADA **19**

5. ANNEX **21**

1. INTRODUCCIÓ, HIPÒTESI I OBJECTIUS DE L'ESTUDI

L'objectiu principal d'aquest treball ha estat crear un model economètric predictiu del PIB a Espanya. La nostra motivació neix de la necessitat d'obtenir el PIB el més ràpidament possible i amb la màxima precisió pel fet que aquest poderós indicador és important per mesurar el creixement de l'economia, per fer comparacions entre països, per obtenir informació sobre l'ocupació i la inversió d'un país, però encara més important, per prendre decisions polítiques. Com més aviat es conegui el valor del PIB, millors polítiques podran ser aplicades i seran caracteritzades per la seva major eficiència.

Per aquest motiu, el nostre estudi s'ha basat en la següent hipòtesi: És possible idear un model economètric amb bon ajust que predigui el PIB amb dades mensuals?

Per tal de comprovar i evaluar la nostra hipòtesi inicial, hem decidit elaborar un "Factor Model" que utilitza diverses variables de freqüència mensual que tenen correlació amb el PIB per així poder superar les prediccions que fa el model autoregressiu de primer ordre AR(1).

2. METODOLOGIA

2.1. Recopilació de les dades

Volíem aconseguir que el nostre model fos més dinàmic i efectiu que els més simples com el model autoregressiu de primer ordre AR(1) (Stock i Watson, 2018). Així que el primer que vam dur a terme va ser recopilar i tractar les dades per crear-lo.

Pel que fa a la recopilació de dades, les vam obtenir de diverses fonts; començant per l'INE (Institut Nacional d'Estadística), continuant amb altres com MITMA (Ministeri de Transports, Mobilitat i Agenda Urbana d'Espanya) per variables de la categoria de construcció, MITES (Ministeri de Treball i Economia Social) per les relacionades amb el mercat laboral, o la FRED (Federal Reserve Economic Data) per dades pertanyents a la categoria de l'ús del PIB i preus.

S'han utilitzat 54 variables que tenen relació amb el PIB d'Espanya. Aquestes variables s'han dividit en les següents categories: ús del PIB (5), construcció (7), finances (13), mercat laboral (12), indústria (4), preus (8) i enquestes (5). La freqüència d'aquestes és mensual i el període d'objecte d'estudi és del Gener del 2013 fins al Setembre del 2022.

2.2. Tractament de les dades

2.2.1. Estacionarietat de les variables:

Una de les principals condicions per a dur a terme un model de factors és utilitzar variables que siguin estacionàries, és a dir, tals que, al llarg de la sèrie temporal, no s'observi una tendència a l'alça o a la baixa, o tals que l'esperança o la variància de la sèrie no depengui del temp. És per això que s'ha dut a terme un estudi de l'estacionarietat de les variables emprades per tal d'avaluar si és necessari modificar els valors de la sèrie en taxes de creixement en el cas que no sigui estacionària o deixar-la tal com està en cas contrari. Aquest estudi ha seguit el següent procediment:

1. Pas 1: Anàlisi descriptiva

   Per cada variable, es realitza un gràfic de punts on s'avalua l'evolució de la mateixa al llarg del temps considerat, adjuntant-hi la seva línia d'ajust. Si s'observa una línia amb pendent igual a 0, ens indicarà estacionarietat, mentre que si té un pendent positiu o negatiu, no estacionarietat, atès que segueix una tendència determinista. També, s'ha obtingut el coeficient d'autocorrelació de primer ordre de la sèrie a partir de la correlació de les dades de la sèrie amb el lag de la variable en qüestió. Com més alt sigui aquest valor, ens portarà a pensar que la variable no és estacionària.

2. Pas 2: Dickey-Fuller Test

   Es pot continuar valorant estacionarietat a partir de la realització d'un Dickey-Fuller Test (Stock i Watson, 2018).

Donada una sèrie temporal, el DF Test es defineix de la següent forma:

1. Partint del següent model original:
   $X_t = \alpha + \rho X_{t-1} + \mu_t$, on α indica l'ordenada a l'origen i ρ el pendent.
2. Es resta X_{t-1} a cada una de les bandes:
   $X_t - X_{t-1} = \alpha + \rho X_{t-1} - X_{t-1} + \mu_t$
3. Traiem factor comú i substituïm la banda esquerra com a l'increment, ΔX_T
   $\Delta X_T = \alpha + (\rho - 1)X_{t-1} + \mu_t$
4. Fem un canvi de coeficient del pendent:
   $\Delta X_T = \alpha + \gamma X_{t-1} + \mu_t$, on $\gamma = (\rho - 1)$

L'equació de l'apartat 4 permet realitzar el Dickey Fuller Test a partir de la regressió de les primeres diferències de X respecte el primer lag de X i fer el següent contrast d'hipòtesi per determinar si la sèrie temporal és estacionària o no:

$H_0: \gamma = 0 \Leftrightarrow \rho = 1 \rightarrow$ La sèrie temporal no és estacionària.

$H_1: \gamma < 0 \Leftrightarrow \rho < 1 \rightarrow$ La sèrie temporal és estacionària.

Fem el contrast d'hipòtesis a un nivell de confiança del 95%, tenint en compte el valor crític de McKinnon al 5% que és -2,86. Per tant, si el t-estadístic calculat de la regressió de l'apartat 4 és major que aquest valor, no podem rebutjar la hipòtesi nul·la, indicant no estacionarietat. No obstant això, si indica que el valor de t és més petit, la sèrie és estacionària.

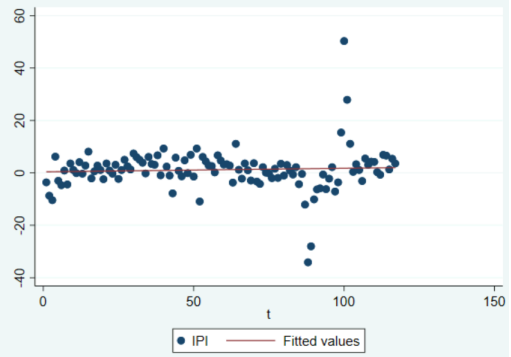
Una vegada realitzat el procediment anterior, en el cas de tenir una sèrie temporal estacionària, mantindrem els valors de la sèrie. No obstant això, si la variable no és estacionària, caldrà generar la taxa de creixement mensual per tal d'obtenir estacionarietat.

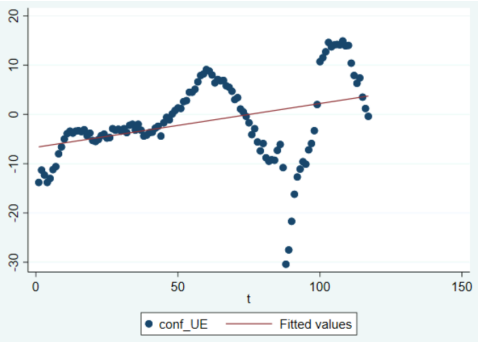
Exemple:

Per a dues variables de la categoria de la Indústria: l'Índex de Producció Industrial (IPI) i el Nivell de Confiança de la Indústria a la Unió Europea (conf_UE).

Pas 1: Anàlisi descriptiva

[1]





En el primer gràfic, de l'Índex de Producció Industrial, s'observa una línia d'ajust completament plana, indicant estacionarietat. No obstant això, per al nivell de confiança de la indústria per a la UE, veiem una tendència positiva al llarg del temps, indicant no estacionarietat.

```
.        corr IPI l.IPI
(obs=116)

                |                 L.
                |     IPI       IPI
----------------+------------------
            IPI |
            --. |  1.0000
            L1. |  0.5379    1.0000
```

```
.        corr conf_UE l.conf_UE
(obs=116)

                |                 L.
                |  conf_UE   conf_UE
----------------+------------------
        conf_UE |
            --. |  1.0000
            L1. |  0.9513    1.0000
```

A partir de les autocorrelacions arribem a les mateixes conclusions. Pel que fa a l'IPI l'autocorrelació no és gaire alta, indicant estacionarietat; mentre per l'altre variable, s'observa una autocorrelació molt elevada, mostrant no estacionarietat.

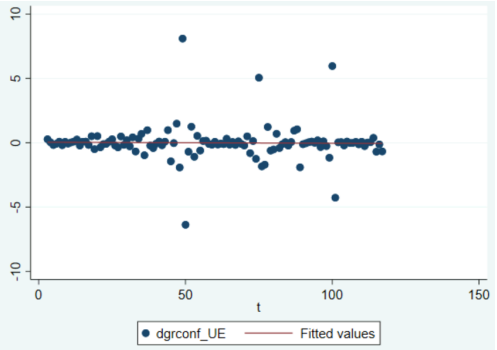
[1] Gràfic 1: Valors de la variable IPI a cada mes amb la seva línia d'ajust.
Gràfic 2: Valors de la variable conf_UE per a cada mes amb la seva línia d'ajust.

Pas 2: Dickey-Fuller Test

Seguint amb el DF test, obtenim els mateixos resultats que s'han comentat abans:
[2]

	T-estadístic	Valor crític
IPI	-4,07	-2,86
Conf_UE	-1,67	-2,86

Per a la variable IPI, rebutgem la hipòtesi nul·la, ja que el valor de t és menor que el valor crític (-4,07<-2,86), de manera que trobem estacionarietat per a aquesta variable, per tant, no cal manipular les dades de la variable en qüestió. Pel que fa al Conf_UE, no podem rebutjar la hipòtesi nul·la (-1,67>-2,86), això indica que la sèrie no és estacionària, en conseqüència, cal fer la taxa de creixement per incorporar-la al model. Si ho realitzem, veiem com la variable es converteix en una estacionària:
[3]




```
.          corr dgrconf_UE l.dgrconf_UE
(obs=114)

                          L.
             dgrcon~E dgrcon~E
-------------+------------------
dgrconf_UE   |
         --. |   1.0000
         L1. |  -0.5504   1.0000
```

	T-estadístic	Valor crític
dgrConf_UE	-6,87	-2,86

[2] Taula: T-estadístic obtingut a l'haver realitzat el Dickey Fuller Test comparat amb el valor crític a considerar.

[3] Gràfic: Taxa de creixement de la variable conf_UE per a cada mes amb la seva línia d'ajust
Taula 1: Autocorrelació de la taxa de creixement de la variable conf_UE.
Taula 2: T-estadístic en haver realitzat el DM test per a la taxa de creixement de conf_UE comparat amb el valor crític a considerar.

En tots casos, s'indica estacionarietat: tenim una línia d'ajust completament plana, una correlació no gaire alta, i rebutgem la hipòtesi nul·la. Amb això queda demostrat com, creant la taxa de creixement, convertim la sèrie en estacionària.

2.2.2. Blocking Approach

Tal com hem comentat anteriorment, les dades de les variables són mensuals mentre que la taxa de creixement del PIB és trimestral, de manera que calia transformar les variables a trimestrals per poder fer el model, a causa de l'absència de dades mensuals del PIB a Espanya. Per realitzar això de forma correcta cal dur a terme un *blocking approach* (Fontenay, 2018), una tècnica utilitzada en el procés de dades per a millorar l'eficiència al realitzar cerques o comparacions de dades grans. El seu objectiu és reduir el número de comparacions necessàries per dividir les dades en blocs o grups més petits per després comparar només dins d'aquests blocs.

En aquest cas, el *blocking approach* el vam utilitzar, com hem esmentat, per passar les dades mensuals a trimestrals i poder dur a terme el model. A les taules adjuntes a continuació s'observa com ha funcionat aquest procés.

[4]

Mesos	Variable X
1	100
2	101
3	102
4	103
5	104
6	105
7	106
8	107
9	108

Trimestres	M1	M2	M3
Q1	100	101	102
Q2	103	104	105
Q3	106	107	108

Aquesta tècnica redueix la quantitat de dades i simplifica l'anàlisi, el que a la seva vegada pot millorar l'eficiència i facilitar la detecció de patrons i tendències a nivell trimestral. A

[4] Taula 1: Variable X per mesos abans d'aplicar el blocking approach
Taula 2: Variable X per trimestres després d'aplicar el blocking approach

l'obtenir dades trimestrals, es poden realitzar comparacions més significatives i extreure conclusions més sòlides sobre les tendències i comportaments al llarg del temps.

2.3. Estimació i criteris de selecció de factors

L'anàlisi factorial, de manera simplificada, consisteix a agrupar variables que estiguin correlades entre elles sota els mateixos factors, latents en principi. Per obtenir l'estimació del Factor Model s'ha utilitzat la fórmula següent, la qual es suggereix en (Stock i Watson, 2002 a), (Stock i Watson 2002 b) , (Schumacher, 2005), (Abdić et al., 2020) :

$$X_t = \Lambda F_t + e_t$$

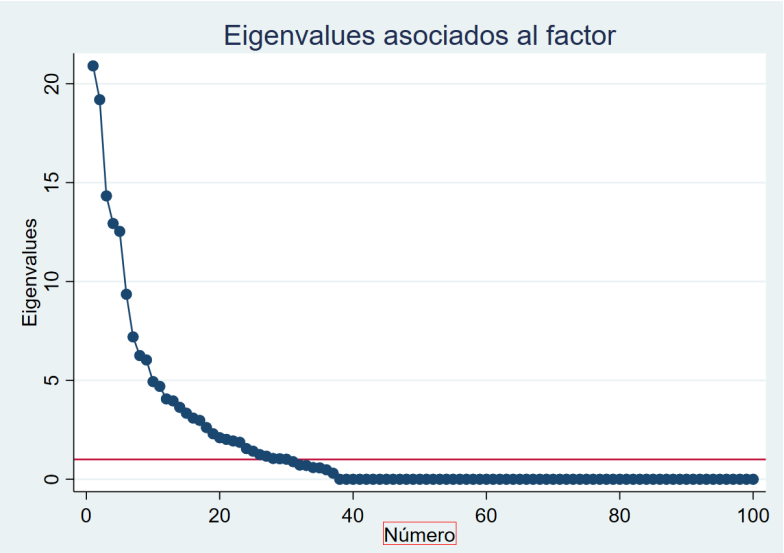
$$Y_t = \sum_{k=0}^{K} \beta_k Y_{t-k} + \beta'_F F_t + \beta'_w W_t + \varepsilon_t$$

Per aconseguir els factors fem servir el programa Stata, agrupem totes les variables en una llista i després executem l'ordre "factor $llista, pcf" que ens proporciona una llista de 165 possibles factors ordenats de major a menor eigenvalue. L'opció *pcf* fa referència al fet que hem triat que Stata estimi els factors a partir del mètode *principal components.*

Criteris de selecció de factors

Criteri de Kaiser

El primer criteri per començar a descartar factors innecessaris pel model és el criteri de Kaiser. Aquest consisteix en seleccionar únicament els factors que tinguin un *eigenvalue* superior o igual a 1 i rebutjar tots els que en tinguin un inferior a 1. En estar ordenats de major a menor, en el nostre cas seleccionem fins al factor número 30. Gràficament:



Criteris d'informació

Els següents criteris per veure si és possible reduir més el nombre de factors que expliquen el model són els criteris d'informació. Aquests són l'AIC i el BIC (Stock i Watson, 2002a) i consisteix en obtenir el valor més baix possible d'entre els 30 factors obtinguts amb el primer criteri. Per això realitzem múltiples regressions començant per un model amb 1 sol factor fins a arribar al model amb els 30 factors. Després, amb la comanda "estat ic" observem els valors de l'AIC i BIC per a cada model, que inclouen quantitats diferents de factors, i escollim el model que minimitza aquests valors. Segons el programa, el model amb els AIC i BIC més baixos és el que inclou 30 factors, per la qual cosa arribem a una conclusió equivalent a la del criteri de Kaiser i per això no descartem cap factor addicional. A més observem com el factor número 30 minimitza el Root MSPE (*Mean Squared Prediction Error*).

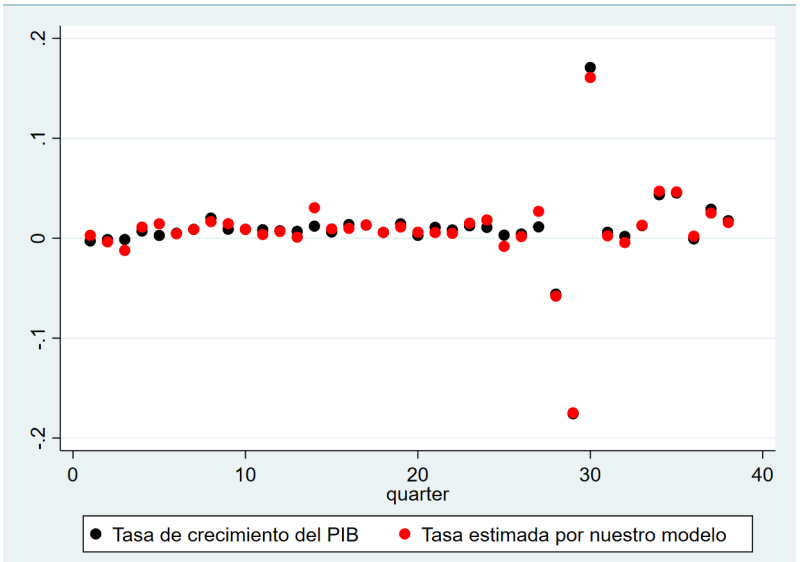
A la taula següent es veuen els valors d'AIC i BIC de cada model a més del Root MSPE.

N. Factors	Root MSPE	AIC	BIC
18	0.02789	-152.5541	-121.44
19	0.02481	-161.4867	-128.735
20	0.02484	-161.5625	-127.1732
21	0.02339	-166.4331	-130.4062

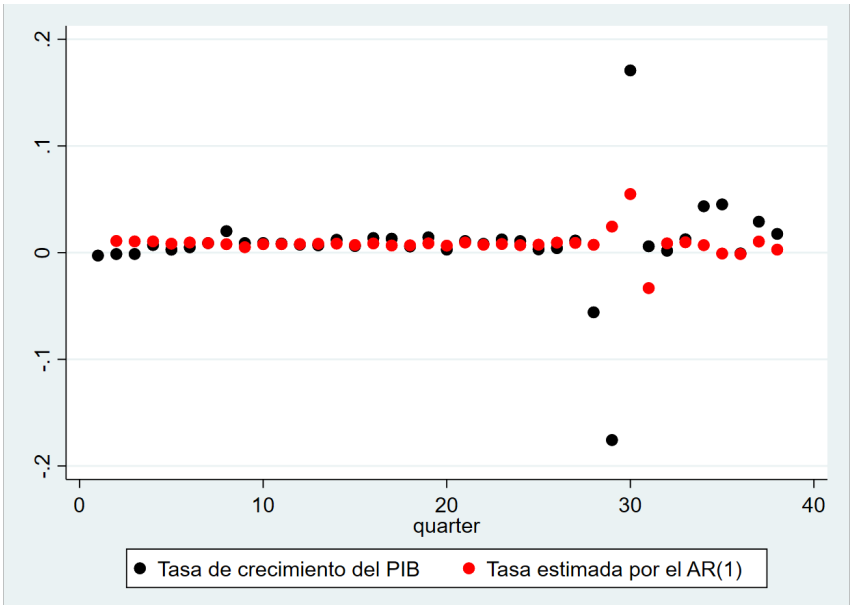
N. Factors	Root MSPE	AIC	BIC
22	0.2167	-172.7168	-135.0524
23	0.01865	-184.7348	-145.4327
24	0.02221	-171.4473	-132.1453
25	0.02126	-175.6021	-134.6624
26	0.01857	-186.9207	-144.3435
27	0.01913	-185.9617	-141.7468
28	0.02004	-184.0442	-138.1918
29	0.02033	-184.9459	-137.4559
30	**0.01448**	**-216.3289**	**-165.5638**

2.4. Resultats

Un cop realitzat el model d'estimació hem comparat les prediccions amb els valors reals del PIB i el que es pot observar és que les prediccions del nostre model són força precises. Els punts vermells representen la taxa de creixement del PIB que estima el nostre model mentre que els punts negres representen la taxa de creixement registrada per a un trimestre determinat. En principi, no s'observen grans diferències entre la predicció i el valor real *in sample*. El model fins i tot és capaç de predir correctament durant períodes amb grans fluctuacions. A més, l'R^2 és de 0.97, per la qual cosa, juntament amb l'anàlisi gràfic, podem concloure que el nostre model és prou bo per a prediccions *in sample*.



Si fem ara una comparació amb el model AR(1), es veu instantàniament que el nostre model és capaç d'estimar el PIB amb moltíssima més precisió dins de la mostra. Un altra vegada, els punts negres representen la taxa de creixement registrada mentre que els vermells ara representen la taxa estimada pel model AR(1). Les diferències més notables es veuen en períodes de gran fluctuació a més de per als darrers períodes de la nostra mostra, on el model tendeix a predir per sota del valor real de manera continuada.



2.5. Anàlisi del Component cíclic de les sèries generades pels dos models

Per tal d'aprofundir en l'avaluació de la capacitat predictiva dels dos models, s'ha acumulat la taxa de creixement predita tant pel model AR(1) com el model de factors obtenint així les prediccions del nivell del PIB per sengles especificacions. D'aquesta manera, per a cada especificació, a la dada del nivell del producte interior brut observada en el primer període, se li ha aplicat la respectiva taxa de creixement predita per aconseguir el primer valor del nivell de producció nacional estimat. Un cop obtinguda aquesta dada, s'ha aplicat la taxa de creixement del següent instant per obtenir el nivell de la sèrie en el segon període. De realitzar aquest procés recursivament, per tant, en resulta la sèrie predita, segons el model autoregressiu i el de factors, del nivell del PIB.

Igualment, si es vol utilitzar el model per al *nowcasting,* pot ser d'interès, no només comparar les sèries en si, sinó també el component cíclic d'aquestes per analitzar el comportament de les prediccions en el curt termini.

2.5.1. Estimació de la tendència: el *Hodrick-Prescott filter*

Per a l'obtenció del component cíclic s'assumeix implícitament que, donada una sèrie y_t:

$$y_t = y_t^g + y_t^c \Leftrightarrow y_t^c = y_t - y_t^g$$

On y_t^g és la tendència de la variable i y_t^c el seu component cíclic.

Existeixen diversos mètodes per obtenir y_t^g. El criteri sobre com estimar la tendència d'una variable, sobretot en casos com el PIB, es discuteix a (Stock i Watson, 1999). En particular, s'argumenta que assumir una simple tendència lineal per a variables com la producció d'un país pot donar lloc a cicles espuris. Es proposen, així, mètodes alternatius, i, si es vol, més sofisticats per a dur a terme un procés de *detrending* com les tècniques de filtratge.

El filtre ideal, segons Stock i Watson, elimina absolutament tant les fluctuacions de molt alta freqüència, possiblement associades a errors de mesura, com les de baixa freqüència relatives a canvis en la tendència. Tot i això, aquest cas és, a la pràctica, impossible d'aconseguir, ja que requeriria un nombre infinit de valors presents i passats de les sèries, és a dir, el filtre ideal és un concepte teòric que s'utilitza com a referència. Existeixen, però, a la pràctica, aproximacions del filtre ideal per a mostres finites com els filtres *bandpass* (Baxter i King, 1999) o el Hodrick-Prescott (Hodrick i Prescott, 1981).

S'ha decidit, per raons de simplicitat i pel fet que la mostra de dades que es posseeix no és excessivament extensa, d'utilitzar el filtre Hodrick-Prescott per aïllar la tendència de les sèries, tant les generades com l'observada del PIB. Tot i que es pot argumentar que el filtre HP elimina l'anomenat *high frequency noise* menys del desitjable, per al propòsit de l'estudi

no s'ha cregut que aquest sigui un problema massa transcendental a l'hora de comparar els resultats de les prediccions *in sample.*

Així, el filtre de Hodrick-Prescott aproxima la tendència d'una variable estimant y_t^g tal que minimitzi l'expressió:

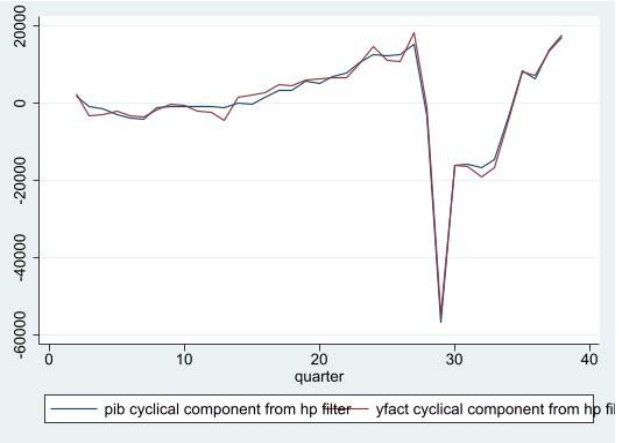
$$\sum_{t=1}^{T}(y_t - y_t^g)^2 + \lambda \sum_{t=2}^{T-1}[(y_{t+1}^g - y_t^g) - (y_t^g - y_{t-1}^g)]^2$$

On λ s'anomena *smoothing parameter*, i pren, convencionalment, el valor de 1600 per a dades macroeconòmiques trimestrals. El primer terme de l'expressió dona pes a la diferència entre la variable observada i la tendència. D'altra banda, el segon terme penalitza els canvis molt freqüents en la tendència estimada. Com més alt sigui λ, més pes cobrarà el segon terme a l'estimació en relació amb el primer i viceversa. En aquest sentit, l'HP filter és una manera analítica de resoldre el dilema existent, en voler estimar la tendència d'una variable, entre ajust a la sèrie observada i filtratge de les variacions d'alta freqüència.

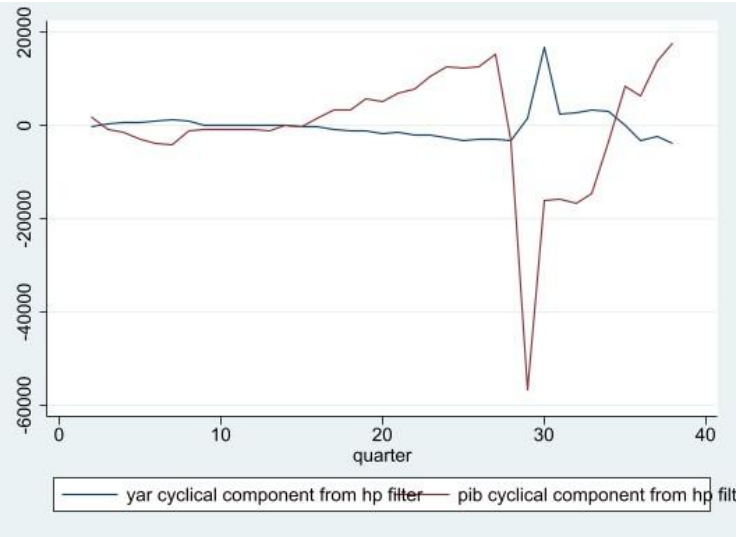
2.5.2. Comparació dels cicles resultants

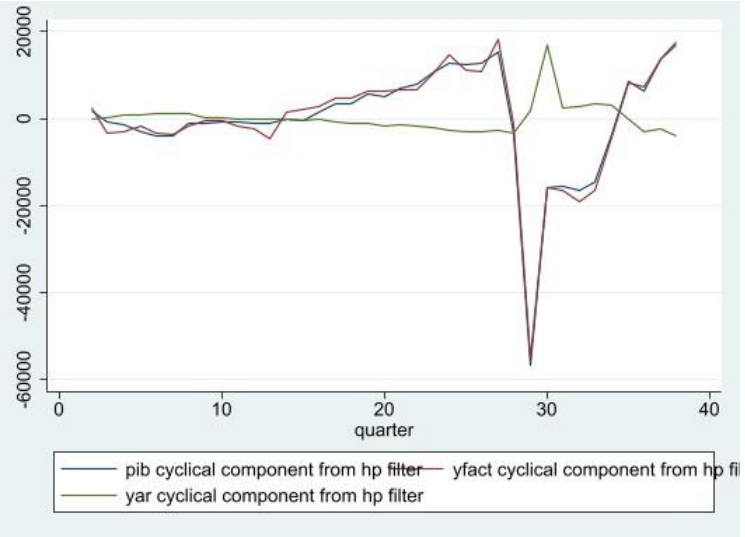
Consegüentment, l'aplicació del filtre HP, d'acord amb les instruccions del manual d'Stata (Stata, 2023), ens ha permès, com s'ha explicat, obtenir la tendència i el cicle del PIB observat, predit pel model AR(1) i pel model de factors. Els resultats es presenten a continuació:

[5]



[5] *Yfact* és el component cíclic de la sèrie generada pel *factor model* i *yar* l'obtingut amb l'AR(1).





Com es pot veure, la sèrie generada pel *factor model* és molt més consistent amb els valors observats que l'obtinguda amb l'AR(1).

La diferència d'ajust esmentada, de fet, es pot mesurar a partir de certs estadístics els quals quantifiquen les relacions considerades gràficament.

[6]

	AR(1)	Factor model	Sèrie observada
$\frac{\sigma_x}{\sigma_y}$	0,2650	1,0077	1
$corr(x_t, y_t)$	0,9932	-0,5444	1
$cor(x_t, x_{t-1})$	0,5427	0,4019	0,5429
$\underset{k}{argmax}\ corr(x_t, y_{t+k})$	0	-7	0

Tal com mostren els resultats, la volatilitat de la sèrie predita amb el model de factors és molt més propera a la del PIB que l'estimada a partir del model autoregressiu, que és marcadament més baixa. En aquest sentit, sembla lògic conjecturar que la diferència de volatilitats relatives es deuria, sobretot, pel millor ajust del *factor model* durant la recessió provocada per la COVID-19.

Addicionalment, cal emfatitzar el fet que la sèrie estimada per l'AR(1) és contracíclica, mentre que la del model de factors és procíclica i guarda una correlació molt elevada amb el PIB. Tant la persistència com la sincronicitat, mesurades en aquest ordre pels estadístics de la taula, presenten molt millor ajust en el cas del *factor model* que en l'AR(1), el qual, plausiblement degut a l'especificació de tal model, genera una sèrie retardada.

En definitiva, es pot concloure que, per a les observacions *in sample,* les dades del PIB aconseguides a partir de les prediccions del *factor analysis* presenten un ajust considerablement millor respecte les dades observades, sigui pel que fa al component cíclic com a la sèrie per se.

[6] S'entén *x* com la variable de predicció en qüestió i *y* com la variable observada. És a dir, en la tercera columna *y* = *x,* però en les altres dues no es dóna això últim.

2.6. Avaluació *pseudo-out-of-sample*

Tot i la contrastada bondat d'ajust dels resultats dintre de la mostra, per tal de validar l'efectivitat del model de factors és precís d'estudiar la seva capacitat de predicció fora d'aquesta, també en relació amb el model autoregressiu de primer ordre.

Per aquest propòsit es mira la bondat d'ajust *out-of-sampl*e dels coeficients a partir de l'anàlisi pseudo-out-of-sample, el qual consisteix a estimar el model en qüestió per a un nombre de períodes inferior al total disponible en la mostra. D'aquesta manera es poden predir les dades yt per als instants que no s'han utilitzat per estimar el model i comparar el resultat amb els valors que de facto pren la variable en aquests períodes. Això permet quantificar la qualitat de predicció *out of sample* d'un model determinat a partir d'estadístics com el Mean Squared Forecasting Error o de contrastos d'hipòtesi com el Diebold-Mariano (Diebold i Mariano, 1995).

2.6.1. El contrast Diebold-Mariano per a la diferència d'error de predicció

Donada una sèrie de valors observats y_t i dues de valors predits $\hat{y}_{it}$, $i \in \{1, 2\}$, es defineix l'error de predicció d'i com:

$$e_{it} = y_t - \hat{y}_{it}$$

I una funció de pèrdua del tipus:

$$g(e_{it}) = e_{it}^2 \text{ o bé } g(e_{it}) = |e_{it}|$$

D'aquesta manera, el diferencial de pèrdua dels models 1 i 2 és:

$$d_t = g(e_{1t}) - g(e_{2t})$$

I es contrasta la hipòtesi:

$$H_0: E(d_t) = 0 \ \forall t$$

$$H_1: E(d_t) \neq 0 \ \forall t$$

Si la sèrie del diferencial de pèrdua presenta memòria curta i estacionarietat en la covariància, llavors l'estadístic del contrast és:

$$\sqrt{T}(\bar{d} - \mu) \sim N(0, 2\pi f_d(0))$$

On:

$f_d(0) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_d(k)$ és la densitat espectral del diferencial de pèrdua en freqüència 0.

$\gamma_d(k)$ és l'autocovariància d'ordre k del diferencial de pèrdua.

$\bar{d} = \sum_{t=0}^{T} d_t$ és la mitjana mostral del diferencial de pèrdua.

μ és l'esperança poblacional del diferencial de pèrdua.

Així, considerant les sèries predites (obtingudes amb el *factor model* i l'AR(1) respectivament) i la sèrie de valors observats, s'ha efectuat el contrast de Diebold-Mariano per a la diferència d'error de predicció entre el *factor model* i l'AR(1). El p-valor resultant és de 0,1574, fet que impedeix de rebutjar la hipòtesi nul·la a qualsevol nivell de significació estàndard.

Possiblement, això últim sigui causat per la manca de grandària de la mostra *pseudo-out-of-sample* i l'especificació del retard k definit per a l'autocovariància en la funció de densitat espectral, el qual possiblement no pugui ser prou elevat, precisament per la manca de períodes d'avaluació.

3. CONCLUSIONS

Després d'haver realitzat el model de factors per predir el PIB d'Espanya i haver valorat els resultats obtinguts, cal destacar quins han estat els punts forts del nostre treball i tenir en compte aquells aspectes que haguessin millorat el model i haguessin permès obtenir unes millors prediccions tant *in sample* com *pseudo-out-of-sample*.

Pel que fa als punts forts del treball i del model portat a cap, és important destacar la gran predicció *in sample* que fa el nostre model, tant a curt termini com a llarg termini. Si comparem la sèrie temporal de la taxa de creixement del nostre model amb la dels valors reals del PIB, es veu clarament com s'ajusten gairebé perfectament. Mentre que si el comparem amb el model AR(1), el nostre model mostra una predicció summament millor a la d'aquest model més simple, sobretot es veu aquesta diferència a partir de la pandèmia de la Covid-19. Això es deu a causa de l'aparició del virus de forma inesperada, de manera que el model AR(1), és incapaç de predir-ho. També, un altre dels punts més rellevants és la capacitat que té el nostre model de predir el PIB a partir d'indicadors que són publicats de forma mensual, en comparació amb el AR(1), que no ho permet, ja que simplement té en compte les dades del PIB real que apareixen de forma trimestral. Òbviament, el nostre model ens permet realitzar prediccions del PIB trimestrals amb uns indicadors que són publicats amb més freqüència.

Tenint en compte aquells punts que haguessin millorat el nostre model, és destacable posar èmfasis en la mostra considerada. Si s'haguessin utilitzat una major quantitat de variables, una millora en la predicció *in sample* com la *pseudo-out-of-sample*, hauria sigut esperada, sobretot aquesta última, ja que com s'ha vist anteriorment, no hi ha diferències entre la *POOS* realitzada per l'AR(1) amb la nostra. Aquest problema, tal com s'ha comentat anteriorment, es podria haver solucionat si més períodes haguessin sigut avaluats. També, el fet d'usar variables que estiguessin més correlades entre elles permetria al nostre model predir, de les dues maneres, millor. Aquest últim aspecte hauria ajudat a reduir el nombre de factors que s'han utilitzat en el nostre model, baixant dels 30 que s'han considerat.

És necessari destacar que el nostre model no avalua la possibilitat que la càrrega dels factors respecte al PIB sigui dinàmica, és a dir, que canviï en diferents instants del temps. Haver usat un Dynamic Factor Model (Stock i Watson, 2016), (Forni et al., 2005), hauria pogut ajudar a millorar ambdues prediccions.

4. BIBLIOGRAFIA CITADA

- Abdić, A., Resić, E., & Abdić, A. (2020). *Modelling and forecasting GDP using factor model: An empirical study from Bosnia and Herzegovina*. Croatian Review of Economic Business and Social Statistics, *6*(1), 10–26.

- Baxter, M. King, R. (1999). *Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series*. NBER. Review of Economics and Statistics, v81(4,Nov) [575-593].

- Diebold, F. X. Mariano, R. (1995). *Comparing predictive accuracy*. NBER Working Paper Series.

- (2013). *Factor analysis*. Stata. Recuperat de: <https://www.stata.com/manuals13/mvfactor.pdf>

- Fontenay, S. (2018). *sdmxuse: Command to Import Data from Statistical Agencies using the SDMX Standard*. The Stata Journal, *18*(4), 863–870.

- Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). *The generalized dynamic factor model: One-sided estimation and forecasting.* Journal of the American Statistical Association, 100(471), 830–840.

- Hodrick, R. Prescott E. (1981). *Post-war U.S. business cycles: an empirical investigation*.Working Paper, Carnegie-Mellon University; Journal of Money, Credit and Banking 29 [1-16].

- Schumacher, C. (2005). *Forecasting German GDP using alternative factor models based on large datasets*. SSRN Electronic Journal.

- Stock, J. Watson, M. (1999). *Business cycle fluctuations in US Macroeconomic Time Series*. NBER. Handbook of Macroeconomics, in: J. B. Taylor & M. Woodford (ed.), Handbook of Macroeconomics, edition 1, volume 1, chapter 1, pages 3-64 Elsevier.

- Stock, J. H., & Watson, M. W. (2002a). *Forecasting using principal components from a large number of predictors*. Journal of the American Statistical Association, *97*(460), 1167–1179.

- Stock, J. H., & Watson, M. W. (2002b). *Macroeconomic forecasting using diffusion indexes*. Journal of business & economic statistics: a publication of the American Statistical Association, 20(2), 147–162.

- Stock, J. H., & Watson, M. W. (2016). *Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics*. Handbook of Macroeconomics (Vol. 2, pp. 415–525). Elsevier.

- Stock, J. H., & Watson, M. W. (2018). *Introduction to Econometrics* (4ª ed.). Pearson.

- (2023). *Tsfilter: filter a time series for cyclical components.* Stata. Recuperat de: <https://www.stata.com/manuals/tstsfilter.pdf>.

5. ANNEX

Variables considerades per dur a terme el treball:

Categoria	Nom de la variable	Font
Enquestes	Índex sobre estocs i existències. Comerç total.	INE
	Índex sobre estocs i existències. Venta i reparació de vehicles de motor i motocicletes	INE
	Índex sobre estocs i existències. Comerç al per major.	INE
	Índex sobre estocs i existències. Comerç al per menor.	INE
	Índex de confiança del consumidor	CIS

Categoria	Nom de la variable	Font
Construcció	Certificacions fi d'obra: N.º de certificats	MITMA
	N.º d'edificis segons classe de promotor (total)	MITMA
	N.º d'edificis de promotor privat (total)	MITMA
	N.º d'edificis d'administració pública	MITMA
	Tipus d'interès mitjà a l'inici de les hipoteques constituïdes (total finques)	INE

Construcció	Tipus d'interès mitjà a l'inici de les hipoteques constituïdes (vivendes)	INE
	Índex nacional de la mà d'obra en construcció	INE

Categoria	Nom de la variable	Font
Mercat laboral	Nombre total d'aturats a Espanya	MITES
	Taxa de despesa en pensions	INE
	Variació al nombre de pernoctacions en hotels	INE
	Afiliació de treballadors al Sistema de la Seguretat Social	MITES
	Contractes a temps indefinit registrats	MITES
	Contractes temporals registrats	MITES
	Empreses inscrites a la Seguretat Social (empreses)	MITES
	Empreses inscrites a la Seguretat Social (treballadors)	MITES
	Cooperatives afegides	MITES
	Societats laborals registrades	MITES
	Autoritzacions de treball concedides a estrangers	MITES
	Treballadors estrangers afiliats a la Seguretat Social en alta laboral	MITES

Categoria	Nom de la variable	Font
Indústria	Índex General de Producció Industrial	INE
	Índex General de la Xifra de Negoci de la Indústria	INE
	Índex General de la Xifra de Negocis empresarials	INE
	Índex de Confiança en la Indústria a la Unió Europea	Unió Europea

Categoria	Nom de la variable	Font
Financers	Cotització IBEX-35	IBEX-35
	Cotització NASDAQ	NASDAQ
	Cotització S&P 500	S&P 500
	Cotització EuroStoxx	EuroStoxx
	Cotització NYSE	NYSE
	Agregats monetaris (M3)	INE
	Crédit a les Administracions Públiques de la Zona Euro	INE
	Crédit a altres sectors residents de la Zona Euro	INE
	Crédit del sector exterior	INE
	Moneda en circulació i dipòsit de les institucions bancàries residents a Espanya	INE
	Actius estrangers	INE
	Passius exteriors	INE
	Índex de la Borsa	INE

Categoria	Nom de la variable	Font
Ús del PIB i Valor Afegit Brut	PIB normalitzat	FRED
	Exportacions	FRED
	Importacions (Valor béns)	FRED
	Producció: Total Industria excloent construcció	FRED
	Reserves totals, excloent or	FRED

Categoria	Nom de la variable	Font
Preus	IPC	FRED
	Taxes d'interès: Rendiment dels bons governamentals a l/t: 10 anys	FRED
	Taxes d'interès: Valors governamentals, bons governamentals	FRED
	Taxes d'interès: Taxes a 3 mesos o 90 dies i rendiments: Taxes Interbancàries	FRED
	Taxes d'interès: Taxes immediates (<24 Hrs): Call Money/Taxa Interbancària	FRED
	Tipus de canvi efectiu real ampli	FRED
	Taxes d'interès real a 10 anys	FRED
	IPC: Preus harmonitzats	FRED