



UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Scienze Economiche e Statistiche

Corso di Laurea Triennale in Economia Aziendale

Tesi in *Statistica*

**MODELLI PER L'ANALISI DEI DATI
TEMPORALI: EVIDENZE
EMPIRICHE SUL PIL**

Relatore:

Ch.mo Prof. Alessandra Amendola

Candidato:

Gabriele Sollai

matr. 0212102806

Anno Accademico 2018/2019

A chi non crede più
A chi non ha mai creduto
A chi si è dato per vinto
A tutti coloro a cui tengo

SOMMARIO

INTRODUZIONE.....	1
I cenni storici e l'analisi del PIL	2
1.1. Il PIL: uno sguardo d'insieme	3
1.1.1. La nascita e lo sviluppo del Prodotto Interno Lordo.....	3
1.1.2. Il PIL come strumento di valutazione del benessere?	4
1.1.3. La non superabilità dell'indicatore	6
1.1.4. Un'opinione condivisa	7
1.2. Analisi dell'aggregato.....	8
1.2.1 La definizione.....	8
1.2.2 PIL reale e nominale	9
1.2.3 Il PIL pro capite.....	9
1.2.4 La condizione economica mondiale	10
1.2.5 La stagnazione del PIL Italiano	11
1.2.6 La ripresa della Grecia	11
2.1 Le serie storiche	13
2.1.1 L'analisi della serie.....	13
2.2 Gli indici di posizione.....	14
2.2.1 La media.....	14
2.2.2 La mediana e i quartili	14
2.3 Gli indici di variabilità.....	15
2.3.1 La varianza	15
2.4 La rappresentazione grafica di una serie storica	16
2.5 Le trasformazioni di una serie.....	16
2.5.1 La serie delle differenze.....	17
2.5.2 Le differenze stagionali.....	17
2.6 I legami tra i caratteri	18
2.6.1 L'autocorrelazione.....	18
2.6.2 L'ACF.....	19
2.7 I processi stocastici.....	20
2.7.1 La stazionarietà del processo.....	20
2.7.2 Il processo White Noise	21
2.7.3 L'invertibilità	21
2.7.4 Il teorema di Wold.....	22
2.7.5 L'ergodicità	22
2.7.6 L'autocovarianza e l'autocorrelazione nel processo stocastico	22
2.7.7 La PACF	23
2.8 I modelli ARMA	24
2.8.1 Il modello Moving Average di ordine q.....	25
2.8.2 Il processo AR(p)	26
2.8.3 L'ARMA(p,q)	27
2.9 Il modello ARIMA	27
2.10 Gli step per l'individuazione del processo stocastico	28
2.10.1 L'analisi preliminare.....	28
2.10.2 L'identificazione del modello.....	29
2.10.3 La verifica del modello	30

2.11 Le previsioni	32
2.11.1 Gli errori.....	33
2.11.2 Il Mean Squared Error	33
2.11.3 Le previsioni nei modelli ARMA, ARIMA e SARIMA	34
2.11.4 Il coefficiente di Theil.....	35
APPLICAZIONE DEI MODELLI ARIMA AL PIL.....	37
3.1 Lo studio dei modelli ARIMA applicati al PIL.....	38
3.2 I dati e il software	38
3.3 Una prima rappresentazione grafica	38
3.4 La stagionalità dell'indicatore.....	40
3.5 La serie destagionalizzata e detrendizzata	41
3.6 La normalità	42
3.7 L'ACF e la PACF	42
3.8 La scelta del modello.....	45
3.9 La verifica del modello.....	46
3.10 Il fitting del modello scelto	49
LE PREVISIONI	51
4.1 Le previsioni dei dati futuri.....	52
4.2 La bontà previsiva del modello	52
4.3 Le previsioni in sample per il modello associato ai dati Italiani	53
4.4 Le previsioni in sample per il modello associato ai dati Greci	53
4.5 Le previsioni out of sample.....	54
CONCLUSIONI.....	57
BIBLIOGRAFIA E SITOGRAFIA	58
RINGRAZIAMENTI.....	59

INTRODUZIONE

Scopo di questo lavoro è quello di andare ad osservare da vicino, tramite il software statistico R, la situazione del PIL Italiano e Greco e di andare a formulare delle previsioni per i restanti due trimestri del 2019 e per, anche se meno precisamente dato l'arco temporale ampio e la numerosità delle variabili geopolitiche che tutt'ora coinvolgono il sistema mondiale, i quattro trimestri del 2020 di tutti e due i paesi soggetti dello studio. Il lavoro sarà svolto applicando la teoria statistica dei modelli ARIMA con la quale si potranno andare a formulare le previsioni che saranno poi riportate graficamente e analiticamente per dare un'immagine dettagliata dei dati rilevati attraverso il software. Interessante sarà visualizzare se le previsioni stimate per il PIL Italiano saranno coincidenti con le stesse dei principali istituti di statistica mondiale e se, altro fattore da non sottovalutare, il modello ARIMA possa ben adattarsi anche a situazioni anomale come quella della Grecia, protagonista di una crisi di fiducia nel 2009 causata anche dalla falsificazione dei bilanci dello Stato per far sì che questo entrasse a far parte della zona Euro.

Lo studio è suddiviso in quattro capitoli volti ad analizzare i vari aspetti del PIL. Nel primo capitolo verranno introdotti i cenni storici dell'aggregato e la situazione mondiale attuale; nel secondo capitolo verrà riportata, in sintesi, la teoria statistica alla base dell'elaborato per rendere chiari i passaggi successivi; il terzo e il quarto capitolo sono invece dedicati al raggiungimento dello scopo del lavoro. Nel terzo, infatti, verranno individuati i modelli adatti ai dati dei due Paesi mentre nel quarto, e ultimo, capitolo, verranno formulate ed esposte le previsioni.

CAPITOLO UNO

I CENNI STORICI E L'ANALISI DEL PIL

SOMMARIO:1.1 Il PIL: uno sguardo d'insieme-1.1.1 La nascita e lo sviluppo del Prodotto Interno Lordo-1.1.2 Il PIL come strumento di valutazione del benessere?-1.1.3 La non superabilità dell'indicatore-1.1.4 Un'opinione condivisa-1.2 Analisi dell'aggregato-1.2.1 La definizione-1.2.2 PIL reale e nominale-1.2.3 Il PIL pro capite-1.2.4 La condizione economica mondiale-1.2.5 La stagnazione del PIL Italiano-1.2.6 La ripresa della Grecia

1.1. Il PIL: uno sguardo d'insieme

«Proprio come un satellite nello spazio può rilevare le condizioni meteorologiche attraverso un intero continente, così il PIL può darci una fotografia complessiva dello stato dell'economia. Permette al Presidente, al Congresso e alla Federal Reserve di giudicare se l'economia si sta contraendo o espandendo, se l'economia ha bisogno di un aiuto o dovrebbe essere un po' frenata, e se c'è la minaccia di una grave recessione o inflazione. Senza misure degli aggregati economici come il PIL, i politici sarebbero come alla deriva in un mare di dati privi di ordine. Il PIL e i relativi dati sono come fari che aiutano i politici a guidare l'economia verso obiettivi economici chiave»¹.

Desidero partire da questa citazione di Paul Samuelson, Premio Nobel per l'Economia, per dare, fin dall'inizio di questo studio, una spiegazione chiara e autorevole di cosa il PIL rappresenti per la società odierna. Il Prodotto Interno Lordo (PIL) è una misura macroeconomica aggregata e definisce, in uno Stato, il valore, a prezzi di mercato, di tutti i beni e servizi finali e che quindi sono strettamente destinati al consumo. Nel PIL, infatti, non vengono calcolati i beni e i servizi intermedi, come ad esempio il valore di mercato dell'acciaio destinato alla costruzione di un macchinario industriale o il valore del legname utilizzato per produrre un tavolo, ma vengono inclusi solo i prodotti destinati al consumo come, ricollegandosi agli esempi precedenti, un macchinario industriale o un tavolo.

1.1.1. La nascita e lo sviluppo del Prodotto Interno Lordo

Per comprendere al meglio le potenzialità e le criticità di questo potente strumento, bisogna andare ad osservare il contesto storico in cui è nato. A seguito della Grande Depressione, una delle più grandi crisi finanziarie della nostra società, sorse la necessità di andare a ricercare nuovi metodi per la valutazione del benessere economico di uno Stato che andassero ad integrare i modestissimi strumenti fino ad allora utilizzati come i dati del

¹ Samuelson, Economics, 15th edition, 1995, McGraw-Hill Companies

gettito fiscale. Fu proprio in questo periodo, nel 1930, che Simon Kuznets si trasferì negli Stati Uniti dove divenne membro del National Bureau of Economic Research (NBER), un ente no-profit americano che si occupava di ricerche in ambiti economici volti allo sviluppo del benessere dello Stato. Nel 1934 Kuznets sviluppò per il Congresso degli Stati Uniti il primo progetto che portava alla nascita del PIL attuale. Lo studio, denominato “National Income, 1929-1932”, fu un tripudio di meticolosità e precisione nei calcoli, in cui si evinceva, in un solo numero, la somma della spesa in beni e servizi degli Stati Uniti, numero che era destinato a crescere nel momento in cui l’economia nazionale era in fase espansiva e a decrescere nel momento in cui vi fosse una fase di regressione economica, dando dunque al Congresso ciò che più aveva ricercato: un nuovo indicatore in grado di guidare le scelte economiche e di verificarne la bontà. In seguito al lavoro svolto, Kuznets fu nominato consigliere nell’amministrazione Roosevelt e il suo principale incarico fu quello di trovare un modo per alimentare la macchina da guerra americana, impegnata nel conflitto Mondiale, senza però andare ad intaccare i consumi interni. Le sue opere furono un successo totale e permisero agli Stati Uniti di rimanere economicamente stabili, a differenza degli altri Paesi coinvolti, anche dopo aver affrontato un durissimo scontro. Ciò permise all’indicatore di Kuznets, in seguito agli accordi Bretton Woods del 1944 e grazie alle linee guida per la contabilità nazionale definite dalla Banca Mondiale, dal FMI e dall’ONU, di essere riconosciuto come principale fonte di informazione sull’economia di un paese e di essere, infine, esportato in tutto il mondo.

1.1.2. Il PIL come strumento di valutazione del benessere?

Non mancarono le critiche a questo nuovo strumento di verifica economica. Già dal 1934, anno della sua nascita, lo stesso Kuznets metteva in guardia: «Il benessere di una nazione difficilmente può essere dedotto da una misura di reddito nazionale come sopra definito».²

² Simon Kuznets, nel suo discorso tenuto al Congresso degli Stati Uniti nel 1934

La critica principale al PIL, infatti, può essere facilmente agganciata all'avvertimento che diede lo stesso fondatore: il nuovo strumento tiene conto solo delle transazioni in denaro che avvengono all'interno di un paese escludendo, dunque, tutte quelle piccole ma significative transazioni che avvengono a titolo gratuito come il volontariato, le prestazioni eseguite nell'ambito familiare. Questa critica trovava e trova tutt'ora i suoi fondamenti nei paesi cosiddetti "in via di sviluppo" che sono caratterizzati principalmente da un'economia su base familiare e dunque escluse dal calcolo suddetto. Una delle critiche più dure ma sicuramente più diretta e significativa è stata data dal sociologo e filosofo Zygmunt Bauman che affermava, con degli esempi molto semplici, ciò che segue: «Se lei fa un incidente in macchina l'economia ci guadagna. I fornitori di medicinali incassano e così il suo meccanico. Se lei invece entra nel cortile del vicino e gli dà una mano a tagliare la siepe compie un gesto antipatriottico perché il PIL non cresce. Se un bene passa da una mano all'altra senza scambio di denaro è uno scandalo.»³ Come si evince da queste illustri parole, il PIL, anche se continuamente aggiornato nei metodi di calcolo, ha la grande pecca di considerare positivi e, quindi, di arricchimento del paese, anche episodi negativi come un incidente d'auto potenzialmente mortale. Molto interessante, da questo punto di vista, è il parere di Ignacio Fernando Lara, Ricercatore presso l'Alta Scuola di Economia e Relazioni Internazionali presso l'Università Cattolica, il quale afferma, nel suo articolo "Quale dibattito sul PIL", che: *«Come conseguenza della sua semplicità e dei buoni risultati che il suo utilizzo ha prodotto in passato, il PIL è divenuto un indicatore ampiamente diffuso e accettato, e viene elaborato da quasi l'intera comunità internazionale. Nonostante siano evidenti i vantaggi delle stime derivate dall'evoluzione della contabilità nazionale e della misurazione del PIL negli ultimi settant'anni, alcuni economisti hanno individuato una serie di punti critici per la costruzione e l'utilizzo dell'indicatore»*⁴

³ Zygmunt Bauman, nell'intervista con Andrea Malaguti per il giornale "La Stampa".

⁴ Ignacio Fernando Lara, Quale dibattito sul PIL?, Impresa&Stato, 2010.

Oltre alle critiche pervenute negli anni di utilizzo di questo indicatore, vi sono stati dei dibattiti accesi e dei tentativi anche istituzionali di sostituzione dello stesso, cercando dei nuovi strumenti che meglio si potessero adattare alla definizione moderna di “benessere”. Il più significativo tra essi è sicuramente la “*Commission on the Measurement of Economic Performance and Social Progress*”, fortemente voluta dall’allora premier francese Nicolas Sarkozy e costituita inizialmente da due Premi Nobel per l’Economia, Joseph Stiglitz e Amartya Sen, unitamente a Jean-Paul Fitoussi dell’Institut d’Etudes Politiques de Paris, poi ampliata con altri 22 membri esperti in economia, scienze sociali, scienze politiche e problematiche ambientali. Gli obiettivi della Commissione erano tre:

- Evincere i limiti fondamentali del PIL quale indicatore di progresso economico e sociale;
- Valutare le informazioni complementari potenzialmente necessarie allo sviluppo di altri indicatori più pertinenti;
- Verificare l’esistenza di strumenti alternativi.

1.1.3. La non superabilità dell’indicatore

È proprio sul secondo punto che, nelle motivazioni alla Proposta di Risoluzione, esposta al Parlamento Europeo, “Non solo il PIL”, nel 2011, si esprime quanto segue: « *Mancano spesso informazioni, dati e indicatori definiti in modo chiaro (per calcolare indici alternativi). A livello di Unione europea e degli Stati membri sono state avviate numerose iniziative per colmare le lacune in questi ambiti. È tuttavia difficile individuare un metodo coerente. Per quanto riguarda gli indicatori, esistono, ad esempio, numerose attività a livello nazionale e internazionale, che sono ormai poco chiare. In particolare, in tutte le discussioni sugli indicatori non bisogna dimenticare che essi da soli non forniscono un contributo sufficiente per il dibattito politico, il processo decisionale e la valutazione*»⁵.

⁵ Relazione di Anna Rosbach su “Non solo il PIL-Misurare il progresso in un mondo in cambiamento”, Parlamento Europeo, 27 Aprile 2011

Possiamo quindi affermare che, pur riconoscendo i limiti del PIL, questo non risulti ancora tecnicamente superabile e che dunque risulti ancora l'indicatore maggiormente utilizzato per le scelte dei policy maker.

Un altro intervento che spiega ancora la non superabilità tecnica del PIL è stato profuso da un contributo del Professor Ignazio Musu dell'Università Ca' Foscari di Venezia: *«Le notevoli difficoltà per arrivare al calcolo di un indicatore di valore che rifletta adeguatamente le diverse componenti del benessere sociale sono dunque legate soprattutto alla necessità di calcolare dei prezzi ombra, che sono per loro natura il prodotto di un lavoro artificiale, e quindi basato su criteri in qualche misura arbitrari, mentre il PIL si basa su dati oggettivi quali i prezzi di mercato. [...] È essenzialmente questo intrinseco grado di artificialità che esige che i criteri per il calcolo di una misura del benessere sociale più adeguata del PIL emergano da una convenzione internazionale, in quanto gli indicatori risultanti dal calcolo debbono poi essere confrontati per i diversi paesi e il confronto deve essere condiviso da tutti questi paesi.»*⁶

1.1.4. Un'opinione condivisa

In conclusione di questa parentesi dedicata alla storia e alle criticità riscontrate nel corso degli anni, si giunge a un pensiero condivisibile dalla maggior parte della comunità economica e politica internazionale: si riconoscono i molteplici limiti del PIL quale indicatore di benessere e si riconosce l'esigenza di ricercare altri indicatori che possano integrarlo per tener conto dell'impatto ambientale e sociale che le politiche economiche hanno nel quotidiano. Si riconosce altresì che *«le statistiche nazionali sul reddito e sulla produzione»* come affermato dall'ex presidente dell'American Economic Association, Robert Eisner *«sono state tra i maggiori contributi di questo secolo alla conoscenza economica.»*⁷

⁶ Musu I. nell'articolo "Perché è così difficile sostituire il PIL" su Italianieuropei.it, 28 Febbraio 2008.

⁷ Eisner R., "The total incomes system of account", University of Chicago Press, 1989

1.2 Analisi dell'aggregato

La formazione del PIL è un processo che richiede uno sforzo non esiguo sia in termini di tempo che in termini di informazioni da incamerare. Questo viene formato a partire dalla formula:

$$Y=C+G+I+(X-M)$$

Dove la Y rappresenta il Prodotto Interno Lordo, la C rappresenta i consumi, la G rappresenta la spesa pubblica, la I rappresenta gli investimenti lordi privati in beni durevoli come l'acquisto di un immobile o di un macchinario aziendale mentre la differenza (X-M) rappresenta il saldo della bilancia commerciale formata dalla differenza tra esportazioni e importazioni del Paese.

1.2.1 La definizione

Il PIL si definisce “Interno” in quanto è calcolato tenendo conto dei confini politici di un Paese. Si terrà dunque conto di tutti i beni e servizi finali prodotti da agenti che hanno il centro dei propri interessi all'interno dei confini politici, calcolando, inoltre, anche ciò che viene prodotto all'estero da agenti operanti nel perimetro economico di un Paese come, ad esempio, una nave in acque internazionali o un'ambasciata all'estero. Verranno di conseguenza esclusi i beni e servizi prodotti da agenti operanti entro i confini geografici del paese interno ma che hanno il loro centro di interessi situato in un Paese estero o che operano su un territorio politico appartenente a un altro Paese come, appunto, un'ambasciata estera situata nei confini geografici dello Stato soggetto del calcolo.

Si definisce inoltre “Lordo”. Questo termine sta ad indicare che il calcolo verrà effettuato non tenendo conto degli ammortamenti. L'ammortamento è un processo contabile utilizzato per la distribuzione del costo di un bene su più esercizi ed effettuato deducendo al valore iniziale dello stesso la quota che presumibilmente è stata utilizzata negli esercizi precedenti. Si distingue, come dice il termine stesso, dal Prodotto Interno Netto, calcolato appunto tenendo conto degli ammortamenti.

1.2.2 PIL reale e nominale

Si distingue poi il PIL nominale dal PIL reale. La differenza tra i due indicatori sta nel metodo di valutazione monetaria dei beni che entrano a farne parte, infatti:

- Quando si calcola il PIL nominale si utilizza come riferimento il prezzo dei beni e servizi rilevato durante l'anno t di riferimento. Quindi nel 2019 verrà utilizzata una valutazione dei beni e servizi in base ai prezzi correnti del 2019.
- Nel momento in cui è il PIL reale l'oggetto in considerazione la valutazione dei beni e servizi rientranti nel calcolo sarà fatta ricorrendo ai prezzi costanti di un anno base di riferimento.

Il PIL reale, essendo quindi immune dalle variazioni dei prezzi intercorse durante gli anni, permetterà la verifica dell'effettivo andamento dell'economia senza l'interferenza dell'inflazione, positiva o negativa che sia, in quanto i valori dei PIL considerati in anni diversi potranno essere paragonati per valutare se vi sia una crescita, una decrescita, o una fase costante dell'economia del Paese.

Il rapporto tra PIL nominale e reale, detto deflatore del PIL, inoltre, consente di misurare l'indice dei prezzi senza dover osservare, direttamente, nulla degli stessi.

1.2.3 Il PIL pro capite

A partire dal Prodotto Interno Lordo, infine, è possibile derivare uno degli indicatori più diffusi per l'osservazione del benessere di un Paese: il reddito pro capite. Vi è da fare, tenendo conto delle autorevoli critiche prima esposte, l'osservazione che l'indicatore non tiene conto, come già detto, di molte variabili che determinano il reale benessere dei cittadini come la condizione sociale e le condizioni ambientali del luogo di riferimento.

L'Italia, nella classifica mondiale stilata in base al PPA-PIL pro capite, dove PPA sta per Parità del Potere d'Acquisto, un metodo di aggiustamento del valore che consente di tener conto delle diverse valute monetarie adottate negli Stati in classifica, oscilla tra la trentatreesima e la trentanovesima posizione con un reddito, come già ribadito,

approssimativo e non sempre realistico, di 39.472 dollari mentre il più “ricco” paese al mondo è il Qatar con un reddito pro capite di 128.487 dollari.

1.2.4 La condizione economica mondiale

Analizzando il contesto storico in cui la nostra società vive, si può affermare che, almeno politicamente parlando, ci troviamo in un periodo che potrebbe essere definito caotico. Tra i maggiori eventi che preoccupano i mercati mondiali, due hanno la ribalta mediatica per la loro importanza: la Brexit e la guerra commerciale tra Stati Uniti e Cina. Gli eventi che vedono protagonista il nostro Paese, invece, sono legati principalmente ad un’incertezza politica dovuta alla recente caduta del Governo anche se, con la nascita del Governo successivo, i mercati internazionali hanno mostrato fiducia nei confronti del nostro Paese.

L’economia mondiale, come riportato nell’ultimo World Economic Outlook stilato dal Fondo Monetario Internazionale, sta affrontando ancora una crescita lenta e, infatti, le previsioni di crescita del PIL mondiale sono state revisionate e portate al +3.2% per il 2019 e +3.5% per il 2020, diminuite quindi dello 0.1% ciascuno rispetto al WEO di Aprile. Il Fondo ha motivato queste decisioni al ribasso in tre punti:

1. Indebolimento del fronte degli investimenti in beni durevoli;
2. La guerra commerciale tra Cina e Stati Uniti;
3. Una crescita leggermente più lenta del previsto delle cosiddette economie in via di sviluppo.

In particolare sul secondo punto il FMI è stato molto critico, consigliando ai principali attori del panorama internazionale di attuare politiche volte al Multilateralismo, alla non imposizione di dazi per riportare in equilibrio la bilancia commerciale e ha, inoltre, consigliato il potenziamento e la chiarificazione degli accordi commerciali tra gli Stati del continente Americano e tra la Gran Bretagna e l’Unione Europea.

In particolare, nei paesi in via di sviluppo, si può notare dal rapporto un taglio delle previsioni abbastanza importante specialmente per quanto riguarda l’India che dal +7.3% e

+7.5% previsti ad Aprile, rispettivamente per il 2019 e per il 2020, sono passate al +7.0% e al +7.2% a causa di un indebolimento della domanda interna.

Per quanto riguarda invece la previsione di crescita dei paesi europei in via di sviluppo, come la Turchia, si parla di un primo quarto di 2019 migliore di quanto si potesse prevedere che spingerà il PIL della regione per l'anno in corso a un +1.0%, 0.2% in più di quanto si prevedeva ad Aprile ma, per il 2020, viene previsto un progressivo rallentamento della zona che ha portato a tagliare dello 0.5% le aspettative precedenti fino a portarle a un +2.5%.

Nel rapporto del Fondo Monetario, inoltre, emerge che la crescita in alcune economie avanzate è risultata essere più forte del previsto, come negli Stati Uniti e in Giappone, ma, in contrapposizione, più lenta del previsto in alcune economie in via di sviluppo.

1.2.5 La stagnazione del PIL Italiano

Il PIL Italiano, come vedremo anche nei capitoli successivi, è in una fase di stagnazione: l'Italia infatti, insieme alla Germania, rimane fanalino di coda nella classifica europea per la crescita prevista e la situazione politica instabile non aiuta sicuramente i mercati a dare quella spinta necessaria a far ripartire l'economia. Come si evince infatti dal rapporto del FMI le stime di crescita per il 2020 sono state ulteriormente tagliate e portate al +0.8% mentre per il 2019 è rimasta invariata al +0.1%.

1.2.6 La ripresa della Grecia

La Grecia, invece, risulta essere uno dei migliori Paesi Europei in termini di crescita del PIL. Questa viene stimata dal FMI, per il 2019, al +2.4% e ciò dà delle prospettive ottimistiche per il Paese. Il Fondo, tuttavia, prevede che il ritorno ai valori del PIL pre-crisi avverrà solo entro un decennio e consiglia inoltre di tenere basso l'obiettivo fiscale per dare una spinta ai consumi interni, contrariamente a quanto sostenuto dalla Commissione Europea che spinge per non far diminuire questa pressione in modo tale che il gettito possa contribuire al sostenimento della spesa pubblica.

CAPITOLO DUE

MODELLI PER L'ANALISI DELLE SERIE STORICHE

SOMMARIO: 2.1 Le serie storiche-2.1.1 L'analisi della serie-2.2 Gli indici di posizione-2.2.1 La media-2.2.2 La mediana e i quartili-2.3 Gli indici di variabilità-2.3.1 La varianza-2.4 La rappresentazione grafica di una serie storica-2.5 Le trasformazioni di una serie-2.5.1 La serie delle differenze-2.5.2 Le differenze stagionali-2.6 I legami tra i caratteri-2.6.1 L'autocorrelazione-2.6.2 L'ACF-2.7 I processi stocastici-2.7.1 La stazionarietà del processo-2.7.2 Il processo White Noise-2.7.3 L'invertibilità-2.7.4 Il teorema di Wold-2.7.5 L'ergodicità-2.7.6 L'autocovarianza e l'autocorrelazione nel processo stocastico-2.7.7 La PACF-2.8 I modelli ARMA-2.8.1 Il modello Moving Average di ordine q -2.8.2 Il processo $AR(p)$ -2.8.3 L' $ARMA(p,q)$ -2.9 Il modello ARIMA-2.10 Gli step per l'individuazione del processo stocastico-2.10.1 L'analisi preliminare-2.10.2 L'identificazione del modello-2.10.3 La verifica del modello-2.11 Le previsioni-2.11.1 Gli errori-2.11.2 Il Mean Squared Error-2.11.3 Le previsioni nei modelli ARMA, ARIMA e SARIMA-2.11.4 Il coefficiente di Theil

2.1 Le serie storiche

Prima di arrivare alle previsioni e ai confronti che avranno come protagonista il PIL italiano, ci si soffermerà, in questo capitolo, sulla teoria statistica alla base delle serie storiche e dei processi stocastici.

Obiettivo principale di una serie storica è quello di rappresentare l'andamento di un fenomeno, di qualunque natura, e, da questa, cercare di estrapolare il processo stocastico che è alla base della stessa per valutarne e prevederne l'andamento in un futuro più o meno prossimo.

Le serie storiche, durante gli anni, sono state applicate a numerosi campi di studio quali l'economia, la biologia, la medicina e la geologia, dando modo agli studiosi di poter avere numerosi progressi in ciascuno di essi.

Una serie storica è un set di osservazioni x_t , ciascuna di esse raccolte con una cadenza s durante un qualunque periodo temporale t .

2.1.1 L'analisi della serie

Per analizzare al meglio una serie storica bisogna valutare gli indicatori di posizione e variabilità che la contraddistinguono e verificare la presenza delle componenti principali.

Queste ultime sono 3:

- Trend: andamento di lungo periodo di una serie causato da eventi che agiscono in modo sistematico sul fenomeno analizzato;
- Ciclo: rappresenta la situazione in cui la serie ripete i propri andamenti con cadenze superiori all'anno. Si dice in questo caso, come indica lo stesso termine, un andamento ciclico della stessa;
- Stagionalità: si osserva questa componente nel momento in cui vi sono fluttuazioni osservate nel periodo inferiore all'anno, ripetute per ciascun anno di rilevazione.

2.2 Gli indici di posizione

Un indicatore di posizione è utile per la ricerca di un valore rappresentativo della serie, intorno al quale si muove tutta la serie storica e, in genere, questi prescindono dalla condizione di non scambiabilità degli elementi, distruggendo quindi le informazioni relative all'evoluzione del fenomeno.

2.2.1 La media

L'indicatore di posizione per eccellenza che rappresenta il valore intorno al quale si muove la serie storica ottenuto dalla somma dei valori delle osservazioni rilevate, divisa per il numero di osservazioni, è la media. Questa definizione può essere riassunta dalla formula:

$$\overline{X}_t = \frac{1}{t} \sum_{i=1}^t X_i$$

Un'altra formula importante che è possibile ricavare da quella sopracitata è quella della media parziale che consente di conoscere l'evoluzione del fenomeno e capire se questa è stazionaria o meno in media. La formula è la stessa della precedente ma, invece di inserire al suo interno tutti i caratteri della serie, vengono inseriti t termini tali che $t < n$ o al più uguale a n . Qualora si volesse costruire la serie delle medie parziali si andrà a calcolare la media parziale per ogni t della serie di partenza. Da quest'ultima sarà poi possibile osservare se vi è stazionarietà in media o meno, infatti: se i valori delle medie parziali si discostano poco dalla media totale allora si osserva questa presenza viceversa, se i valori si dovessero discostare significativamente dalla media non si riscontrerebbe la stazionarietà in media.

2.2.2 La mediana e i quartili

La mediana è l'indicatore che rappresenta il valore centrale della distribuzione ordinata. Questo valore, infatti, sarà superiore alla metà dei caratteri mentre sarà inferiore all'altra metà. Per calcolarla, dunque, bisognerà innanzitutto costruire una nuova serie andando ad ordinare in modo crescente i caratteri della serie iniziale, individuando in seguito il carattere che la rappresenta tramite il seguente metodo:

- Se la serie ordinata fosse composta da un numero dispari di caratteri allora la mediana sarebbe il carattere che occupa la posizione $(n+1)/2$ e quindi: $x_{(n+1)/2}$
- Se, invece, la serie fosse composta da un numero pari di osservazioni allora la mediana sarà il valore risultante dal seguente calcolo: $(x_{n/2} + x_{1+n/2})/2$

La mediana, dunque, divide in due sottoserie il fenomeno osservato. I quartili non saranno altro che le mediane di queste due nuove serie create e, in particolare, il primo quartile Q_1 sarà la mediana della sottoserie a sinistra della mediana che presenta solo valori inferiori rispetto alla mediana; il secondo quartile Q_2 corrisponderà con la mediana dell'intera serie; Q_3 , il terzo quartile, corrisponderà alla mediana della sottoserie in cui sono presenti solo valori superiori alla mediana. I quartili, inoltre, sono utili per individuare i possibili valori eccezionali che sono elementi caratterizzati da un'elevata differenza, non ripetuta nel tempo, col valore medio della serie. Tramite i quartili si può infatti calcolare i cosiddetti valori cardine, massimo o minimo, entro i quali la maggior parte degli elementi sono collocati. Qualora vi fossero valori inferiori o superiori a questi, si tratterebbe dunque di un elemento eccezionale. Le formule per il calcolo di questi due valori sono:

- $H_1 = Q_1 - 1.5(Q_3 - Q_1)$ per il limite minimo;
- $H_2 = Q_3 + 1.5(Q_3 - Q_1)$ per il limite massimo.

2.3 Gli indici di variabilità

Gli indici di variabilità, invece, ci danno informazioni rispetto alla dispersione dei caratteri della serie rispetto al valore medio, dandoci quindi un'idea della sua distribuzione.

2.3.1 La varianza

Il più utilizzato tra questa tipologia di indicatori è la varianza che rappresenta il valore medio del quadrato degli scarti, un indicatore utile a comprendere la dispersione dei caratteri e la loro distanza dal valore medio, minore sarà dunque il valore della varianza, maggiore sarà la concentrazione degli elementi intorno alla media. La formula della varianza è:

$$\sigma^2 = \frac{1}{N} \sum_{t=1}^N (X_t - \bar{X})^2$$

Altri indici di variabilità sono: la deviazione standard, ricavata semplicemente dalla varianza facendone la radice quadrata e la differenza tra quartili relativa ricavata nel seguente modo: $(Q_3 - Q_1)/2$.

2.4 La rappresentazione grafica di una serie storica

Un elemento positivo delle serie storiche è anche la facilità di rappresentazione grafica che esse hanno. La più usata intuitiva è la rappresentazione su piano cartesiano o time-plot: data una serie X_t con $t=1,2,...,n$, ponendo il tempo sull'asse delle ascisse e i caratteri x_t sulle ordinate e rappresentando con un punto, poi unito da linee spezzate con gli altri, ogni coppia (t, X_t) , si otterrà una riproduzione grafica della serie considerata dando all'osservatore degli elementi chiave per carpirne l'evoluzione.

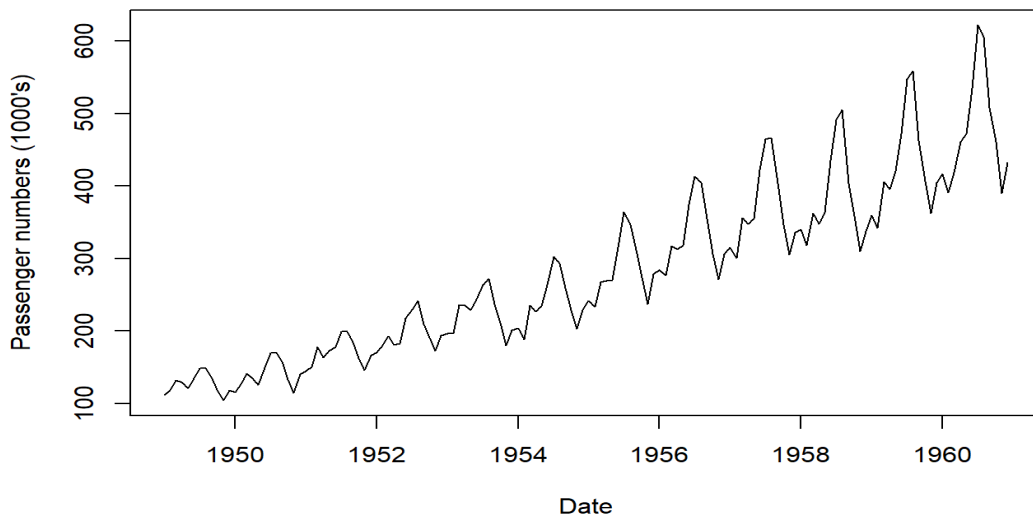


Figura 1. Esempio di rappresentazione grafica di una serie storica. Passeggeri linee aeree 1949-1961.

2.5 Le trasformazioni di una serie

Quando si analizza una serie storica, d'altronde, non si è interessati ad un unico valore ma all'intera evoluzione della stessa. Per analizzare al meglio le serie, dunque, sarà necessario

effettuare delle trasformazioni per eliminare la presenza di eventuali caratteristiche non visibili dalla rappresentazione prodotta in un primo momento.

2.5.1 La serie delle differenze

La trasformazione più utilizzata è la serie delle differenze. Con quest'ultima si forma una nuova serie Z_t formata dalle differenze tra i caratteri x_t e i loro precedenti x_{t-1} , nel caso della differenza prima, che ci permetterà di eliminare il trend della serie qualora questo dovesse essere lineare. Nel caso in cui il trend dovesse essere di natura esponenziale, si utilizzerebbe un nuovo operatore delle differenze applicato alla serie Z_t , formato dunque dalle differenze tra gli elementi z_t e i loro precedenti z_{t-1} . Obiettivo delle serie delle differenze, oltre ad eliminare il trend, è anche andare a ridurre la varianza per rendere più agevole l'analisi. Per valutare se la trasformazione sia necessaria o meno, infatti, bisognerà effettuare una verifica: $\delta = 1 - (\text{var}_z / \text{var}_x)$. Grazie a questo calcolo avremo una rappresentazione precisa di come sono costruite le due serie infatti: qualora δ dovesse essere positivo vorrebbe dire che il rapporto tra le varianze è minore di 1, di conseguenza allora la var_z è minore della var_x raggiungendo lo scopo prefissato di diminuire la variabilità degli elementi, rendendo utile quindi la trasformazione di X_t in Z_t .

2.5.2 Le differenze stagionali

Un'altra trasformazione delle serie storiche, utile specialmente in ambito economico e, infatti, come vedremo nel capitolo successivo, per il PIL, è la differenza stagionale. Quando infatti si riscontrano delle stagionalità all'interno della serie è opportuno adoperare questo genere di trasformazione per eliminarla. La serie delle differenze stagionali, data una s tale che questa sia la cadenza con la quale la serie storica è stata rilevata, sarà dunque formata da: $Z_t = \nabla_s X_t = X_t - X_{t-s}$, per $t = s+1, s+2, \dots, N$. Supponendo che la cadenza sia di 12, vorrà dire osservare le differenze che intercorrono tra il mese dell'anno t e lo stesso mese dell'anno precedente, $t-1$. In modo simile alla serie delle differenze si potrà andare a rilevare l'impatto della stagionalità sulla serie effettuando il seguente calcolo: $1 - (\text{Var}_z / \text{Var}_x)$. Più il risultato di

questa operazione si avvicinerà ad 1 più sarà forte l'incidenza della stagionalità, qualora dovesse essere negativo allora non vi è presenza di stagionalità.

2.6 I legami tra i caratteri

Quando si ha a che fare con una serie storica, un elemento importantissimo e, quindi, da valutare attentamente, è il legame che intercorre tra un carattere e i suoi precedenti. Qualora questo non dovesse esistere vorrebbe dire che i dati sono del tutto casuali e, di conseguenza, non analizzabili da alcun modello. Per valutare questo elemento si ricorre, come primo passo, alle autocovarianze che misurano per l'appunto il legame lineare che potrebbe intercorrere tra X_t e X_{t+k} , prendendo in considerazione tutti gli elementi distanti k unità temporali. La formula per il calcolo delle autocovarianze è:

$$C_x(k) = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad k = 0, 1, 2, \dots$$

Come si può notare, la formula è molto simile a quella della covarianza, l'unica differenza infatti sta nel fatto che qui si va a calcolare sulla stessa X , rilevata in tempi diversi, e non dunque Y come nella covarianza.

2.6.1 L'autocorrelazione

A partire dall'autocovarianza si può calcolare l'autocorrelazione che ci indica quale sia la forza del legame lineare calcolato pocanzi. L'autocorrelazione va a rapportare l'autocovarianza a lag k al numeratore e l'autocovarianza al lag 0 al denominatore. Da notare che quest'ultima sarà uguale alla varianza. La formula dunque sarà:

$$r_x(k) = \frac{\sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sigma_x \sigma_x} = \frac{C_x(k)}{C_x(0)}$$

Il risultato del rapporto sarà compreso tra -1 e 1 ma diventerà sempre meno affidabile all'aumentare di k . Si consiglia infatti di non considerare un valore del lag che vada oltre il 25% del numero dei caratteri contenuti nella serie presa in considerazione.

2.6.2 L'ACF

Una volta calcolate tutte le autocorrelazioni necessarie sarà possibile costruire un grafico che ci permetterà di avere molte informazioni riguardanti la serie: il correlogramma. Quest'ultimo viene formato a partire dalla statistica Q di Ljung-Box che prevede un test delle ipotesi:

$$\begin{cases} H_0: \rho_0 = \rho_1 = \dots = \rho_m = 0 \\ H_1: \rho_0 \neq 0 \end{cases}$$

Ed è valutato calcolando la seguente espressione e riportato graficamente come nella Figura 2:

$$LB_m = n(n+2) \frac{\sum_{k=1}^m \rho_{a_k}^2}{n-k}$$

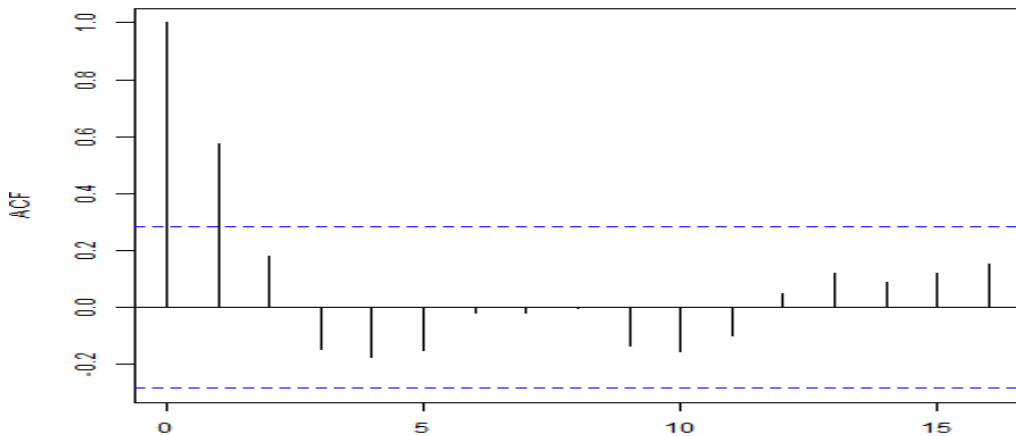


Figura 2. Esempio di ACF.

Sull'asse orizzontale verranno riportati i lag con i quali sono stati calcolati gli indici, sull'asse verticale, invece, il risultato del test Q. In particolare si può dire che:

- Qualora le autocorrelazioni dovessero avere un andamento lineare decrescente, allora si evidenzierà la presenza di un forte trend lineare;
- Se invece il grafico dovesse prendere una forma sinusoidale di periodo s allora si potrebbe affermare che c'è una stagionalità con cadenza s;
- Nel caso in cui dovessero essere presenti entrambi gli andamenti riportati pocanzi, allora vi sarebbe sia un forte trend che una stagionalità;

- Qualora vi fosse una totale assenza di struttura, invece, si direbbe che è una serie puramente residua e che quindi tutte le sue autocorrelazioni non sono significative e, di conseguenza, uguali a 0. Questo si può visualizzare sul grafico quando, per ogni lag, gli indici sono all'interno delle barre di confidenza.

2.7 I processi stocastici

Ciò che si utilizzerà per lo studio e la comprensione di una serie storica sono i processi stocastici. Questi sono una sequenza di variabili casuali scandite da un parametro temporale, in questo caso, discreto, indicato con la lettera t , e determinati dal set di distribuzioni di tutte le variabili al loro interno. La maggior parte delle informazioni dei processi possono però essere descritte attraverso lo studio della media, della varianza e della covarianza. Qualora il processo fosse formato da una serie di variabili casuali normali, inoltre, il primo e il secondo momento determinano completamente l'intero set di distribuzioni.

La serie storica può essere definita come la realizzazione finita di un processo stocastico formato dunque da elementi ordinati in base alla linea temporale come: $x_0, x_1, \dots, x_{t-1}, x_t$.

La serie storica di nostro interesse sarà il punto di partenza dello studio attraverso il quale bisogna ricavare il processo stocastico sottostante che ci aiuterà a descriverne l'andamento e a formulare delle previsioni. Studiare un processo stocastico partendo da una sola delle sue infinite realizzazioni può divenire molto complicato e, perciò, è utile applicare delle restrizioni teoriche per facilitarne la comprensione, la prima di esse è la stazionarietà.

2.7.1 La stazionarietà del processo

Un processo si definisce stazionario in senso forte se la sua distribuzione è indipendente dal parametro t e che quindi la realizzazione finita X_t sarà simile alla realizzazione "time shifted" X_{t+h} . I requisiti impliciti di tale affermazione sono:

- I momenti della distribuzione, qualora questi esistano, devono essere indipendenti dal tempo;

- Le funzioni di densità, sia univariate che multivariate, rimangono invariate quando l'asse temporale viene traslato di un lag k generico.

Questa restrizione in senso forte, d'altronde, è difficilmente verificabile dal momento che dovrebbe essere rispettata una condizione di invarianza rispetto a tutte le variabili casuali che compongono il processo. Ai nostri scopi sarà dunque utile applicare una restrizione in senso debole. Un processo si definisce stazionario in senso debole se vengono rispettate le seguenti proprietà:

- Il momento primo deve essere indipendente da t ;
- Deve rispettare la condizione di omoschedasticità e, cioè, c'è la necessità che la varianza sia costante per ogni t .
- La covarianza deve dipendere solo dal lag temporale k applicato agli elementi, cioè che $E[(X_t - \mu)(X_{t+k} - \mu)] = \gamma(k)$

2.7.2 Il processo White Noise

Il più semplice processo stazionario è il White Noise, indicato con a_t , che è una collezione di variabili casuali incorrelate tra loro, a media 0 e con varianza costante σ^2 . La covarianza, per rispettare l'incorrelazione, del processo dovrà essere uguale a 0 tranne se il lag k preso in considerazione non dovesse essere 0, in questo caso la covarianza sarà uguale a σ^2 .

2.7.3 L'invertibilità

Un processo si definisce inoltre invertibile se dovesse esistere una funzione $h(X_{t-1}, X_{t-2}, \dots)$ tale che si possa riscrivere X_t tramite essa, formata dalle variabili casuali passate, sommata a un processo White Noise.

2.7.4 Il teorema di Wold

Il principale postulato che ci permette di continuare con lo studio è il teorema di Wold che afferma: dato un processo stazionario X_t , questo può essere scomposto in due processi tali che: $X_t = V_t + Z_t$. I due nuovi processi saranno incorrelati tra loro, dunque con una covarianza uguale a 0. Il processo V_t è un processo deterministico e quindi senza alcuna componente stocastica. Z_t , invece, è la componente stocastica formata da una combinazione lineare infinita di White Noise. I due processi seguiranno le seguenti espressioni:

$$V_t = \mu + \sum_{j=1}^{\infty} [\alpha_j \sin(\lambda_j t) + \beta_j \cos(\lambda_j t)]$$

$$Z_t = \sum_{j=0}^{\infty} \Psi_j a_{t-j}$$

2.7.5 L'ergodicità

Un'importante proprietà che i processi stocastici devono possedere è l'ergodicità. Questa è simile alla proprietà di consistenza per gli stimatori che ci dice che all'aumentare delle informazioni a disposizione lo stimatore converge in probabilità al valore del parametro da stimare. Allo stesso modo l'ergodicità ci dà la possibilità di fare inferenza sull'intero processo, grazie ad esse, infatti, sarà possibile effettuare delle stime sui parametri che lo costituiscono come, ad esempio, nel caso della media, al tendere all'infinito della numerosità campionaria, lo stimatore convergerà in media quadratica al valore della media statistica.

2.7.6 L'autocovarianza e l'autocorrelazione nel processo stocastico

Per un processo stocastico, l'autocovarianza sarà data dalla covarianza tra le due variabili casuali X_t e X_{t-k} e dunque sarà: $\text{Cov}(X_t, X_{t-k}) = \gamma(k)$. Se k dovesse essere posto uguale a 0, allora si otterrà la varianza della variabile casuale X_t . Il legame lineare tra le due variabili,

grazie all'autocovarianza, potrà essere calcolata con le autocorrelazioni. Con l'autocorrelazioni globali si tiene conto di tutto ciò che intercorso tra una variabile e l'altra e sarà ricavata grazie al rapporto tra la $Cov(X_t, X_{t-k})$ e il prodotto delle deviazioni standard delle due variabili:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{Cov(X_t, X_{t-k})}{\sqrt{Var(X_t)Var(X_{t-k})}} \quad k = 0, 1, 2, \dots$$

Questo parametro $\rho(k)$ sarà compreso tra -1 ed 1, sarà simmetrico e, qualora k dovesse essere uguale a 0, questo sarà uguale a 1. Date m variabili casuali, si potrà formare una matrice di Toeplitz che contiene al suo interno tutte le autocorrelazioni. La matrice è semi-definita positiva e, cioè, tutti i suoi minori principali sono non negativi ed è simmetrica.

$$P_{(m)} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(m-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(m-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(m-3) \\ \vdots & & & & \vdots \\ \rho(m-1) & \rho(m-2) & \rho(m-3) & \cdots & 1 \end{bmatrix}$$

2.7.7 La PACF

Grazie alla matrice costruita pocanzi, se il processo è stazionario, sarà possibile ricavare anche l'autocorrelazione parziale che rappresenta il legame lineare che intercorre tra le v.c. X_t e X_{t-k} al netto delle $k-1$ variabili che si interpongono tra esse. È indicata con $\pi(k)$ ed è calcolata grazie ai determinanti di una seconda matrice, costruita a partire dalla matrice delle correlazioni e, appunto, della matrice delle correlazioni. La seconda sarà costruita sostituendo alla matrice delle correlazioni solo l'ultima colonna, rimpiazzata con un vettore $[\rho(1), \rho(2), \rho(3), \dots, \rho(k)]$. In seguito sarà possibile individuare la formula:

$$\pi(k) = \frac{|P_{(k)}^*|}{|P_{(k)}|}$$

Una volta calcolate tutte le autocorrelazioni parziali, sarà possibile rappresentarle graficamente in modo simile all'Auto Correlation Function, come riportato nella Figura 3.

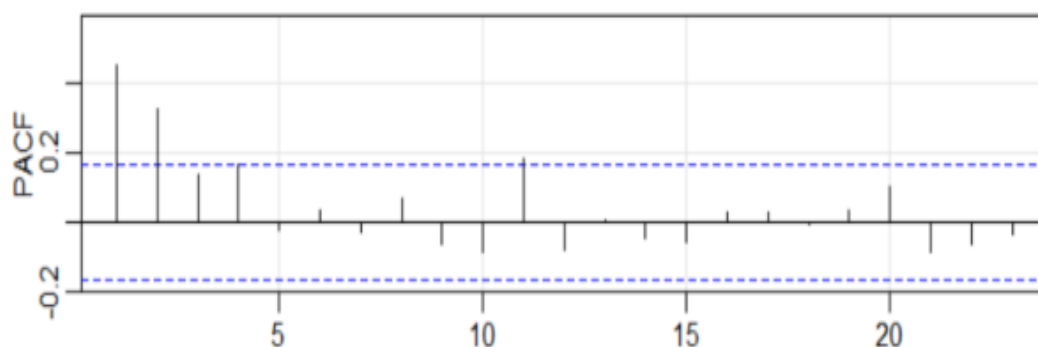


Figura 3. Esempio di PACF.

2.8 I modelli ARMA

Prima di introdurre lo studio dei modelli ARMA, autoregressivi a media mobile, è utile esplicitare la funzione dell'operatore Backward, indicato con la lettera B. Questo ci consente di scrivere un elemento del passato in funzione di una variabile presente, utile per la comprensione è l'esempio: $B(X_t) = X_{t-1}$ oppure $B^j(X_t) = X_{t-j}$ oppure, ancora, può essere utilizzato per differenziare la serie storica semplicemente con $(1-B)X_t$.

Grazie a questo operatore sarà possibile riscrivere la componente stocastica del processo in una forma molto più semplice:

$$Z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} = \sum_{j=0}^{\infty} \psi_j B^j a_t = \left(\sum_{j=0}^{\infty} \psi_j B^j \right) a_t$$

Indicando, infine, il polinomio di grado infinito nell'incognita B con Ψ , si otterrebbe la forma finale che sarà: $Z_t = \Psi(B)a_t$. Sappiamo poi che un polinomio di grado infinito può essere riscritto sottoforma di rapporto di grado finito. Questo dunque ci permetterà di dare una forma ancora più semplice, eliminando definitivamente il grado infinito del polinomio e avremo dunque che:

$$\Psi(B) = \frac{\Theta_q(B)}{\Phi_p(B)}$$

Da questo polinomio avremo finalmente l'equazione che rappresenta un modello autoregressivo a media mobile di ordine (p,q):

$$\Phi_p(B)Z_t = \Theta_q(B)a_t$$

O, in modo equivalente:

$$Z_t = \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

Possiamo dunque notare che il processo è costituito da due elementi. Il primo, a sinistra della prima formula, è la componente autoregressiva mentre il secondo, a destra, è la componente moving average o media mobile. Per comprendere al meglio il processo, sarà dunque utile analizzare questi due elementi separatamente.

2.8.1 Il modello Moving Average di ordine q

Un processo MA(q) si realizza nel momento in cui l'ordine p del processo AR è uguale a 0, portando il processo ad essere riscritto come:

$$Z_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

Andando quindi ad eliminare l'elemento autoregressivo. Il processo risulterà sempre stazionario ma invertibile se e solo se le q radici dell'equazione caratteristica sono in modulo maggiori di 1. Le autocorrelazioni globali di un processo MA(q) mostreranno un punto di troncamento nel momento in cui $k=q$, ciò vuol dire che per $k < q$ le autocorrelazioni globali saranno significativamente diverse da 0, viceversa per $k > q$ queste saranno significativamente uguali a 0. Le sue autocorrelazioni parziali, invece, mostreranno un andamento tendente allo 0 sottoforma di una mistura di esponenziali.

Il più semplice dei processi a media mobile è il MA(1) in cui, per l'appunto, $q=1$ e potrà essere rappresentato come:

$$Z_t = a_t - \theta_1 a_{t-1}$$

Stimare i momenti di questo processo sarà sicuramente meno difficoltoso rispetto a processi con più parametri. Sappiamo infatti che le autocorrelazioni avranno un punto di troncamento nel punto $k=1$ e potremo stimare come segue tutti i momenti:

$$E(Z_t) = E(a_t) - \theta E(a_{t-1}) = 0$$

$$Var(Z_t) = E[a_t - \theta a_{t-1}]^2 = E(a_t)^2 + \theta^2 E(a_{t-1})^2 - 2\theta E(a_t a_{t-1}) = \sigma_a^2(1 + \theta^2)$$

$$\gamma(1) = Cov(Z_t, Z_{t-1}) = -\theta \sigma_a^2$$

$$\gamma(2) = \gamma(3) = \dots = 0$$

$$\rho(k) = \begin{cases} 1 & k = 0 \\ \frac{-\theta}{1 + \theta^2} & k = 1 \\ 0 & \text{altrimenti} \end{cases}$$

2.8.2 Il processo AR(p)

Il processo AR(p), invece, si realizza nel momento in cui l'ordine q del processo MA(q) è uguale a 0. Il processo risulterà sempre invertibile ma stazionario se e solo se le p radici dell'equazione caratteristica associata al processo sono in modulo maggiori di 1. A causa della mancanza del processo MA(q), il processo potrà essere riscritto nel seguente modo:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t$$

Per riconoscere un processo di questo tipo sarà utile visualizzare il grafico ACF e PACF, in modo simile a un processo MA(q). Il processo AR(p), infatti, possiede le autocorrelazioni globali che tendono a 0 sottoforma di una mistura di esponenziali mentre le autocorrelazioni parziali avranno un punto di troncamento nel punto in cui $k=p$.

Il più semplice processo AR è l'AR(1). Questo infatti avrà solo un parametro da stimare e sarà formato nel seguente modo:

$$Z_t = \phi Z_{t-1} + a_t$$

Le autocorrelazioni globali saranno: $\rho(k) = \phi^k$

Mentre le autocorrelazioni parziali, che avranno un punto di troncamento al lag $k=1$, dunque nel momento in cui si avrà $k>1$, queste saranno tutte significativamente uguali a 0:

$$\pi(1) = \rho(1) \quad \pi(2) = \pi(3) = \dots = 0$$

2.8.3 L'ARMA(p,q)

Quando sia l'ordine p , del processo $AR(p)$, che l'ordine q , del processo $MA(q)$, saranno maggiori di 0, si parlerà, infine, di un processo $ARMA(p,q)$. Esso potrà dunque essere scritto come:

$$\Phi_p(B)Z_t = \Theta_q(B)a_t$$

Questo avrà caratteristiche miste dei due processi che lo compongono infatti:

- Sarà stazionario se e solo se le p radici dell'equazione $\Phi_p(B) = 0$ sono in modulo maggiori di 1;
- Sarà invertibile se e solo se le q radici dell'equazione $\Theta_q(B) = 0$ sono in modulo maggiori di 1;
- Le autocorrelazioni globali avranno un punto di troncamento per $k=q$, quando si avrà che $k>q$ allora queste saranno nulle, come nel processo $MA(q)$;
- Le autocorrelazioni parziali avranno un punto di troncamento per $k=p$, quando si avrà che $k>p$ allora queste saranno nulle, come nel processo $AR(p)$.

2.9 Il modello ARIMA

I modelli ARMA finora esposti hanno un momento primo uguale a 0. Nel momento in cui il processo preso in considerazione non dovesse avere una media nulla, si passerebbe ad utilizzare degli operatori tali da tenerne conto.

Supponendo che il processo sia stazionario in varianza ma non in media si avrà che quest'ultima sarà funzione del tempo. A questo punto, data una W_t tale che:

$$\nabla^d Z_t = W_t$$

Avremo dunque un nuovo processo che è costruito tenendo conto delle differenze di ordine d del processo precedente Z_t che potrà, infine, essere rappresentato come un $ARMA(p,q)$.

Grazie all'assunzione appena fatta potremo costruire un nuovo processo: l'ARIMA(p,d,q). Il processo terrà conto della non stazionarietà di Z_t e questo potrà essere dunque riscritto come segue:

$$\Phi_p(B)\nabla^d Z_t = \theta_q(B)a_t$$

Possiamo dunque dire che grazie a questa classe di processi Auto Regressive Integrated Moving Average (p,d,q) avremo la possibilità di eliminare ogni possibile trend di ordine r presente nel processo preso in considerazione originariamente, cioè, però, sarà possibile se e solo se W_t potrà essere rappresentato da un ARMA(p,q).

2.10 Gli step per l'individuazione del processo stocastico

L'obiettivo principale da centrare è identificare il processo stocastico alla base dei dati numerici della serie storica e ciò sarà possibile, in questo studio, grazie ai modelli esposti precedentemente. Per fare ciò bisognerà procedere per gradi attraverso un percorso iterativo composto da più step: analisi preliminare, identificazione del modello, stima dei parametri, verifica del modello e, infine, accettazione o rifiuto dello stesso.

2.10.1 L'analisi preliminare

Il primo step, l'analisi preliminare, sarà un approccio deterministico composto da grafici temporali, identificazione di eventuali outliers, studio dei principali indici statistici e la messa in atto di trasformazioni per circoscrivere i valori anomali emersi.

In prima approssimazione, per costruire un'analisi corretta, sarà necessario effettuare un test di normalità della serie o delle trasformazioni scelte per essa: il test di normalità di Jarque-Bera. La statistica test si costruirà seguendo la seguente formula:

$$JB = \frac{N}{6} (S^2 + \frac{1}{4}(K - 3)^2) \quad \begin{matrix} H_0 \\ \sim \end{matrix} \chi^2_{(2)}$$

Mentre le ipotesi alla base di questo saranno le seguenti:

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

Il risultato del test sarà agevolmente visualizzato grazie al p-value, il livello di probabilità che il risultato della statistica test si verifichi sotto la condizione dell'ipotesi nulla, output del software statistico, in particolare:

- Se il p-value dovesse essere maggiore del livello di significatività scelto, nel nostro caso 0.05, si accetterà l'ipotesi nulla, affermando dunque che la trasformazione si distribuisce normalmente
- Viceversa, se il p-value dovesse essere inferiore ad α , si rifiuterà l'ipotesi nulla, negando la normalità della serie.

Nella pratica non è raro lavorare su serie storiche distribuite, originariamente, non normalmente. In quest'ultimo caso sarà necessario, prima di continuare, l'utilizzo della Box-Cox transformation, un metodo che utilizza una famiglia di trasformazione di potenze seguendo la formula:

$$X_{tr} = \frac{X^\lambda - 1}{\lambda} \text{ se } \lambda \neq 0$$

$$X_{tr} = \log(X) \text{ se } \lambda = 0$$

Per ricercare il parametro λ che renda la distribuzione della serie approssimabile ad una normale sarà necessario l'utilizzo del software R che ricercherà un valore tale da massimizzare la funzione L nota come funzione log-likelihood che viene riportata in questo modo:

$$L = \frac{n-1}{2} \ln S_{tr}^2 + (\lambda - 1) \frac{n-1}{n} \sum \ln X$$

Il valore di λ che il software attraverso individuerà sarà dunque della variabile oggetto di studio che meglio normalizzerà la sua distribuzione, rendendo quindi possibile la continuazione dell'analisi.

2.10.2 L'identificazione del modello

Nella fase successiva, l'identificazione del modello, sarà necessario calcolare tutti quei valori espressi precedentemente per cercare di visualizzare chiaramente il modello corretto

che meglio si adattano alla serie sotto esame. Sarà dunque necessario calcolare le autocorrelazioni parziali e globali, per osservare, se ci fossero, i punti di troncamento che permettono di scegliere tra i vari processi AR, MA o ARMA, confrontando l'andamento teorico di questi processi con quello reale dei dati.

Per la stima dei parametri, essendoci delle minimizzazioni iterative numeriche, sarà necessario l'utilizzo di software statistici come R. Solitamente, quando l'errore si distribuisce normalmente, si utilizza il metodo della Massima Verosimiglianza, esatta o condizionata, che differiscono tra loro principalmente nella stima della costante e delle eventuali variabili esogene.

2.10.3 La verifica del modello

Lo step successivo è quello in cui bisogna prestare la maggiore attenzione: la verifica del modello. In questa fase del nostro studio sarà necessario valutare i principali criteri che danno la possibilità di scelta tra un modello e un altro. Questa fase sarà condotta effettuando una verifica sui coefficienti, visualizzando il valore dei parametri e conducendo delle verifiche ulteriori sui residui. Vi è inoltre da enunciare un altro criterio importante per la verifica che è la parsimonia: l'aggiunta di ulteriori parametri da stimare rende sempre più complessa l'analisi e, dunque, bisognerà verificare che questi siano realmente necessari alla conduzione dell'analisi.

Per la prima sottofase, la verifica dei coefficienti, in seguito alla stima degli stessi, sarà effettuato un test di ipotesi per osservare se questi siano significativi o meno:

$$\begin{cases} H_0: \theta = 0 \\ H_1: \theta \neq 0 \end{cases} \quad \text{o} \quad \begin{cases} H_0: \phi = 0 \\ H_1: \phi \neq 0 \end{cases}$$

La statistica test utilizzata saranno le seguenti:

$$t_c = \frac{\hat{\theta}}{\sqrt{\text{VAR}(\hat{\theta})}} = z \quad \text{oppure} \quad t_c = \frac{\hat{\phi}}{\sqrt{\text{VAR}(\hat{\phi})}} = z$$

Come precedentemente enunciato, il p-value del test ci darà un rapido riscontro, dando modo di dare una risposta veloce ed efficace e scegliere se rifiutare o accettare l'ipotesi

nulla. Per verificare poi la stazionarietà e l'invertibilità, nei casi delle componenti AR e MA, bisognerà anche osservare se le radici delle equazioni caratteristiche siano in modulo maggiori di 1.

Conclusasi la sottofase di verifica dei coefficienti, qualora i risultati fossero positivi, si passerà alla valutazione dei criteri. Questi ci daranno un'idea su quanto il modello descritto sia compatibile e adattato ai dati reali, per verificare quindi se questo sia abbastanza preciso da poter essere utilizzato per l'analisi.

Il più usato tra essi, quando si tratta di modelli lineari, è l'indice della bontà di adattamento R^2 . Questo, nella versione semplificata, tiene conto della somma dei quadrati dei residui e della somma dei quadrati della variabile dipendente mentre, nella versione cosiddetta aggiustata, terrà conto anche della numerosità della serie e del numero dei parametri stimati. Rispettivamente i due criteri sono calcolati come segue:

$$R^2 = 1 - \frac{SSR}{TSS}$$

$$R^2 = 1 - \frac{SSR/(n - k)}{TSS/(n - 1)}$$

Un secondo criterio generale, dal quale poi sono stati derivati altri simili, è il “criterio di informazione di Akaike” calcolato nel seguente modo:

$$AIC = -2l(\hat{\theta}) + 2k$$

Questo terrà conto, dunque, della massima verosimiglianza e dei parametri da stimare dando anche spazio al principio della parsimonia in quanto il modello migliore sarà proprio quello che va a minimizzare il suddetto indice AIC.

Il terzo, e ultimo, step per la verifica del modello è l'osservazione dei residui dello stesso. I residui sono lo scarto che c'è tra il valore reale della serie storica e il valore stimato dal modello. Questi costituiranno una seconda serie dinamica che bisognerà osservare per verificare che non vi siano delle costanti quali trend, stagionalità o ciclicità in quanto, se

una di queste dovesse essere presente, vorrebbe dire che il modello scelto trascurava delle informazioni importanti e pertanto deve essere reimpostato correttamente riconducendo il percorso iterativo svolto fino a questo momento. La prima verifica da fare sui residui è la normalità degli stessi. Per fare ciò ci si affiderà al loro grafico per ottenere un risultato visivo e al test di Jarque-Bera enunciato pocanzi per la normalità della serie o della sua trasformata ma con una modifica sostanziale per tener conto anche dei parametri stimati:

$$JB = \frac{N - k}{6} (S^2 + \frac{1}{4} (K - 3)^2) \underset{\sim}{H_0} \chi^2_{(2)}$$

Anche in questo caso la preferenza sarà di accettare l'ipotesi nulla, dichiarando la normalità dei residui e la non sistematicità degli stessi. In secondo luogo verrà costruito nuovamente il correlogramma con ACF e PACF per verificare la non correlazione tra i residui e i loro precedenti. Quest'ultima verifica sarà effettuata con il test Q di Ljung-Box come precedentemente esposto ma, a differenza del caso precedente, il nostro interesse sarà quello di rifiutare l'ipotesi nulla, dichiarando dunque la non significatività delle correlazioni tra un errore e un suo precedente a lag k.

Conclusasi anche questa parte di verifica con un esito positivo, si potrà affermare che il modello scelto sia conforme a tutte le prerogative imposte, che questo sia corretto e, se previsto, adeguato per la formazione di previsioni sull'andamento futuro del fenomeno oggetto di studio.

2.11 Le previsioni

La previsione dei dati è la stima del valore che un fenomeno può assumere nel futuro tramite la costruzione di un modello che tenga conto dei valori attuali e passati con cui lo stesso si è sviluppato. Vi sono due tipi di previsione:

- Si parla di previsione ex-post o in sample nel momento in cui si forma una previsione su dati già noti per valutare le capacità del modello.

- Nel campo delle previsioni ex-ante o out of sample, si parla di stimare attraverso il modello dei dati di istanti temporali al di fuori del campione utilizzato per la stima del previsore.

2.11.1 Gli errori

La bontà delle previsioni, ovviamente, sarà strettamente legata all'ideoneità del modello. Essendo una stima, tuttavia, il valore della previsione conterrà al suo interno anche un errore, l'obiettivo è di minimizzarlo per avvicinarsi il più possibile alla realtà che si presenterà al momento temporale preso in considerazione dalla previsione.

Considerando $\widehat{Z}_n(h)$ il nostro previsore e indicato con h l'orizzonte temporale entro il quale si vorranno prevedere i valori dei dati corrispondenti, si definirà errore di previsione la differenza che si genererà nel momento in cui si utilizzerà lo stimatore anziché Z_{n+h} . Dunque:

$$e_{n,h} = Z_{n+h} - \widehat{Z}_n(h)$$

2.11.2 Il Mean Squared Error

Come detto pocanzi l'obiettivo sarà minimizzare l'errore dunque lo stimatore migliore sarà quello che minimizza il Mean Squared Error (MSE) calcolato come segue:

$$MSE = \frac{\sum_{t=n}^n (Z_t - \widehat{Z}_t)^2}{h + 1}$$

Dal MSE potranno essere ricavati altri tre indicatori utili all'analisi:

- Bias proportion: indicatore di distanza tra la media delle previsioni e la media dei valori reali:

$$\frac{(\widehat{Z}_{n+h} - Z_n)^2}{MSE}$$

- Variance proportion: indicatore di differenza tra la variabilità dei valori stimati e la variabilità dei valori reali:

$$\frac{(s_{\widehat{Z}_{n+h}} - rs_{Z_{n+h}})^2}{MSE}$$

- Covariance proportion: comprende tutto il restante errore non sistematico e non catturato dai due precedenti indicatori:

$$\frac{(1 - r^2)s_{Z_{n+h}}^2}{MSE}$$

Per valutare positivamente la previsione, i primi due indicatori dovranno essere prossimi allo 0 mentre il terzo dovrà risultare vicino ad 1. In ogni caso la somma dei 3 indicatori sarà pari a 1.

2.11.3 Le previsioni nei modelli ARMA, ARIMA e SARIMA

Alla base della teoria sulle previsioni sui modelli ARMA vi sono 3 supposizioni che vanno fatte per un corretto studio:

- Dal principio è conosciuto tutto il passato del fenomeno preso in analisi;
- Il modello assegnato è adeguato;
- Sono noti i valori dei parametri del modello.

Prima ancora di iniziare vi è da dire anche che le strutture delle previsioni per i modelli ARIMA e seasonal ARIMA sono assimilabili a quella dei modelli ARMA in quanto il primo può essere identificato come un ARMA(p+d,q) mentre il secondo a un ARMA(p+d+Ds,q+Qs).

Gli ordini p e q del modello selezionato rappresenteranno questa volta gli step che si svolgeranno andando a ritroso, indicati con “i”, e verranno messi in contrapposizione all’orizzonte temporale prescelto “h”.

Il teorema portante in questa sezione teorica è stato ideato da Granger e Newbold. Questo afferma che il previsore ottimale $Z_n(h)$ è il risultato della media di Z_{n+h} condizionata dai valori passati tali che: $I_n = \{\dots, Z_{-1}, Z_0, Z_1, \dots, Z_n\}$. Da ciò, formalmente:

$$Z_n(h) = E(Z_{n+h}|I_n) = E(Z_{n+h}|Z_1, \dots, Z_n)$$

Se il modello stimato è un ARMA(p,q) avremo:

$$Z_n(h) = E(Z_{n+h}|I_n) = E\left(\sum_{i=1}^p \phi_i Z_{n+h-i} + a_{n+h} - \sum_{i=1}^q \theta_i a_{n+h-i} | I_n\right)$$

Da cui si ricava:

$$Z_n(h) = \sum_{i=1}^p \phi_i E(Z_{n+h-i}|I_n) + E(a_{n+h}) - \sum_{i=1}^q \theta_i E(a_{n+h-i}|I_n)$$

Avremo dunque ora la distinzione tra la prima parte, deterministica, e la seconda, stocastica, e possiamo osservare, nella sezione deterministica, come al variare di i e h cambia la situazione in cui ci troviamo:

- Se $i \geq h$, intuitivamente gli step a ritroso saranno superiori rispetto all'orizzonte temporale e, di conseguenza, il valore Z_{n-i+h} sarà un valore noto dei nostri dati e, banalmente, stiamo effettuando delle previsioni ex-ante;
- Se invece $i \leq h$ risulterà che gli step siano inferiori all'orizzonte, stiamo effettuando previsioni ex-post in cui $E(Z_{n+h-i}|I_n) = Z_n(h-i)$ è un valore non noto.

Per la parte stocastica, cambiano leggermente le cose:

- Se $i \geq h$, effettuando quindi previsioni ex-post, l'errore di previsione sarà noto e valutabile: $E(a_{n+h-i}|I_n) = a_{n+h-i} = Z_{n+h-i} - Z_{n+h-i-1}$;
- Per $i \leq h$, invece, non essendo noto il valore dell'errore, bisogna farne il valore atteso ma, come sappiamo, essendo l'errore distribuito come un White Noise, il valore atteso $E(a_{n+h-i}|I_n) = 0$ evidenziando dunque che l'orizzonte temporale non potrà essere superiore all'ordine q della componente MA altrimenti le previsioni risulteranno pari a 0.

2.11.4 Il coefficiente di Theil

Svolti i passaggi che ci portano ai valori delle previsioni, dovremo cercare di valutare la capacità previsiva del modello. Ciò sarà possibile grazie ad un indicatore sintetico applicato alle previsioni in sample: il coefficiente di Theil. Questo è sempre compreso tra 0 e 1 dove 0 indica delle previsioni ottime mentre 1 indica che queste sono completamente errate. Il

coefficiente viene calcolato col rapporto tra la varianza totale e la somma delle varianze dei dati reali e delle previsioni, formalmente:

$$U = \frac{\sqrt{\frac{\sum_{t=n}^h (Z_t - \hat{Z}_t)^2}{h+1}}}{\sqrt{\frac{\sum_{t=n}^h (\hat{Z}_t)^2}{h+1}} + \sqrt{\frac{\sum_{t=n}^h (Z_t)^2}{h+1}}}$$

CAPITOLO TRE

APPLICAZIONE DEI MODELLI ARIMA AL PIL

SOMMARIO: 3.1 Lo studio dei modelli ARIMA applicati al PIL -3.2 I dati e il software-
3.3 Una prima rappresentazione grafica-3.4 La stagionalità dell'indicatore-3.5 La serie
destagionalizzata e detrendizzata-3.6 La normalità-3.7 L'ACF e la PACF-3.8 La scelta del
modello-3.9 La verifica del modello-3.10 Il fitting del modello scelto

3.1 Lo studio dei modelli ARIMA applicati al PIL

Una volta immagazzinati tutti gli strumenti necessari alla comprensione dell'analisi delle serie storiche, si può iniziare lo studio dei modelli ARIMA applicati agli aggregati macroeconomici come il PIL.

3.2 I dati e il software

Le serie storiche che seguiranno sono state rilevate dall'Eurostat, istituto statistico dipendente dalla Commissione Europea, che ha come obiettivo principale la raccolta di dati e la conseguente valutazione, effettuata sulla base degli stessi, volta a provvedere all'informazione statistica delle principali istituzioni dell'Unione Europea. I dati verranno dunque importati sul software statistico grazie anche all'apposito package "eurostat" contenuto nell'open library di R.

3.3 Una prima rappresentazione grafica

In prima analisi, dopo che i dati sono stati caricati, è opportuno visualizzare l'evoluzione del PIL Italiano e Greco negli ultimi 24 anni per dare un'idea grafica di ciò che è accaduto in passato.

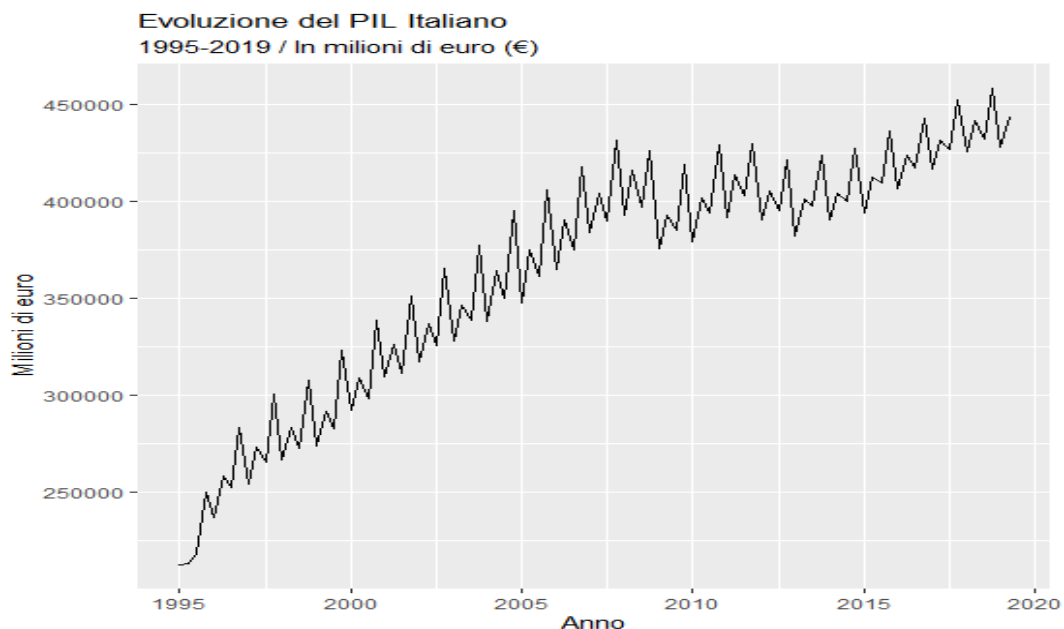


Figura 4 Grafico della serie storica del PIL Italiano

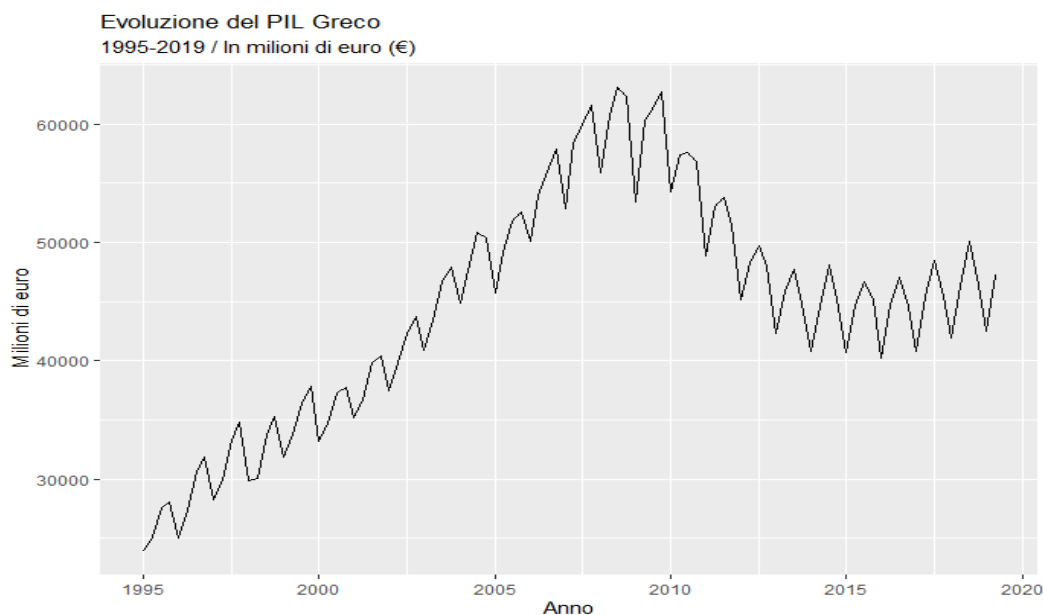


Figura 5 Grafico della serie storica del PIL Greco

Il primo dato che si nota, per quanto riguarda l'Italia, dalla figura 4 è l'improvvisa diminuzione che vi è stata in corrispondenza del 2008 causata dalla crisi finanziaria del mercato immobiliare statunitense. Una seconda osservazione potrebbe inoltre essere riferita al cambio di passo che purtroppo ci ha visti protagonisti: com'è possibile osservare infatti, la crescita del PIL fino al 2007 è caratterizzata da un trend crescente costante ma, nel dopocrisi, c'è stato un rallentamento significativo sia nel ritorno a livelli del pre-crisi che nella crescita. Il PIL Italiano, nel secondo trimestre del 2019 si è attestato intorno ai 443.407 milioni di euro, affermandosi come quarto più grande d'Europa, includendo nella classifica anche il Regno Unito, prossimo all'uscita dall'Unione.

Osservando invece la figura 5 il grafico mostra quanto fosse drammatica la situazione economica e sociale nel post-crisi del 2009: il PIL fino a quel momento, come l'Italia, aveva un buon trend crescente e costante ma, a causa delle falsificazioni dei bilanci e della sfiducia degli investitori esteri, si è verificata un'inversione del trend arrivando a un minimo di 40.716,3 milioni di Euro nel Q1 del 2015 per poi iniziare una lenta ripresa. Nell'ultimo dato rilevato dall'Eurostat, il Q2 del 2019, il PIL Greco si è attestato intorno ai 47216.2 rilevando dunque un divario di circa 13.000 milioni di Euro con i livelli pre-crisi.

3.4 La stagionalità dell'indicatore

Quando si parla del PIL, non è anomalo riscontrare delle stagionalità legate ai periodi dell'anno, nelle figure 6 e 7 questo sarà reso evidente grazie alla rappresentazione grafica annua dell'andamento dell'aggregato.

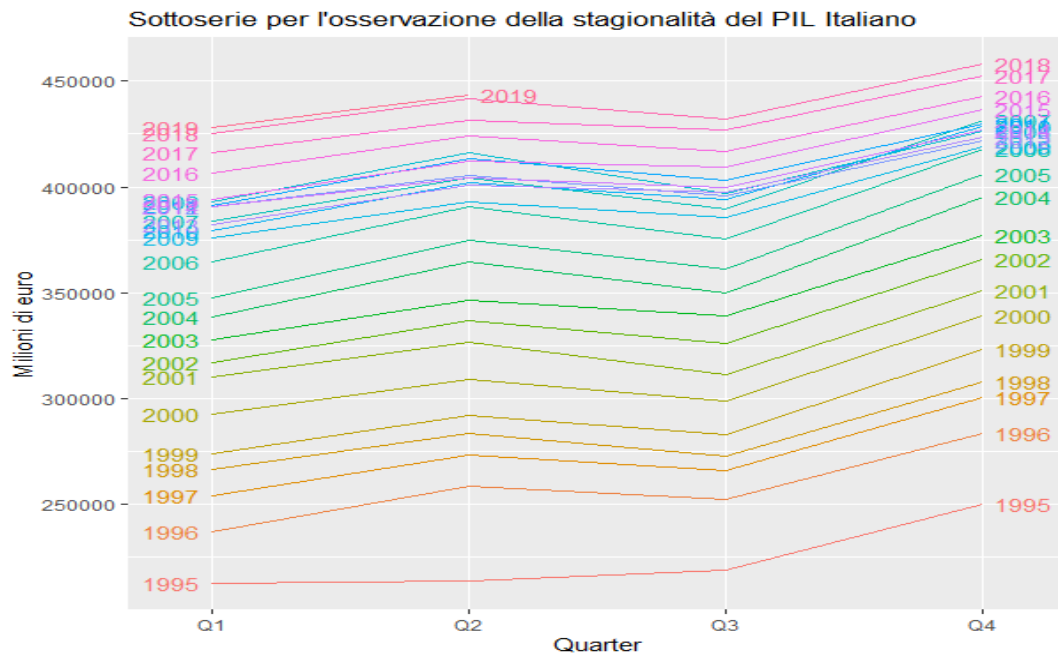


Figura 6 Sottoserie annua del PIL Italiano

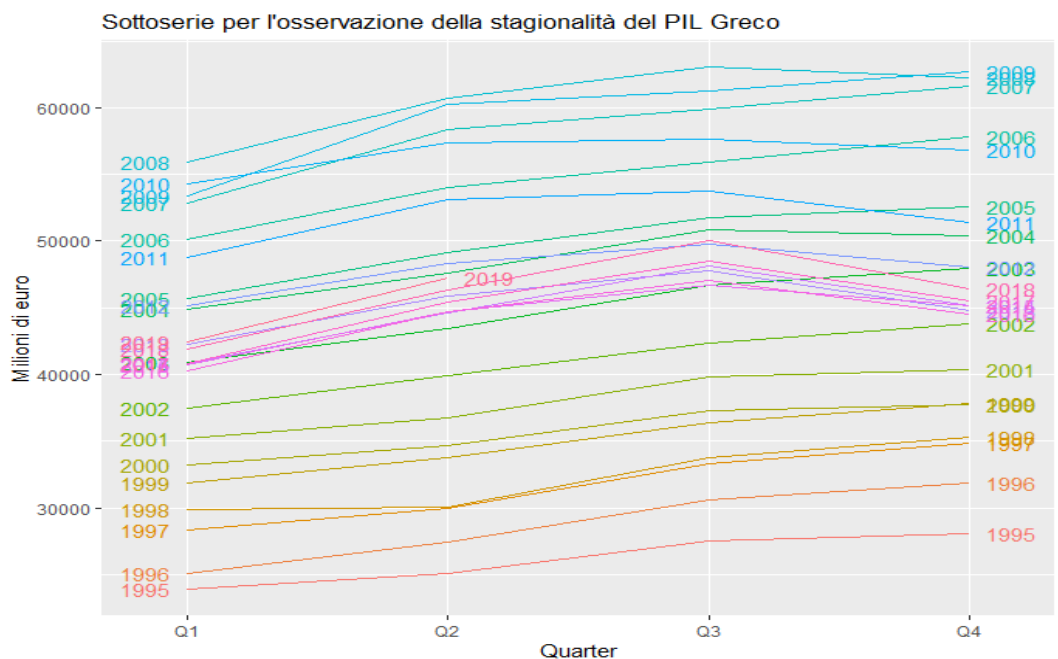


Figura 7 Sottoserie annua del PIL Greco

Da questa rappresentazione grafica si evincono subito gli andamenti stagionali al quale il PIL è sottoposto. Questo andamento ondulare, nella figura 6, che si ripresenta in tutti gli anni di osservazione è, appunto, una stagionalità della serie storica al quale bisognerà prestare attenzione nel corso dello studio per non incappare in errori di valutazione.

Per quanto riguarda la sottoserie del PIL Greco, la situazione è sicuramente più confusionaria a causa dell'inversione del trend ma, negli anni che precedono e succedono la crisi del 2009, si possono comunque notare delle variazioni riscontrabili ogni anno evidenziando quindi la stagionalità.

3.5 Le serie destagionalizzate e detrendizzate

In seguito alla destagionalizzazione e all'estrazione del trend dalla serie, effettuata grazie alla rilevazione, tramite il comando “decompose” di R, che va a estrarre ogni elemento ricorrente per poi riportarlo in una nuova serie, e alla successiva sottrazione di questa alla serie originale, si otterrà un nuovo fenomeno che sarà rappresentato, allo stesso modo sia per l'Italia che per la Grecia, alla figura 8 e 9.

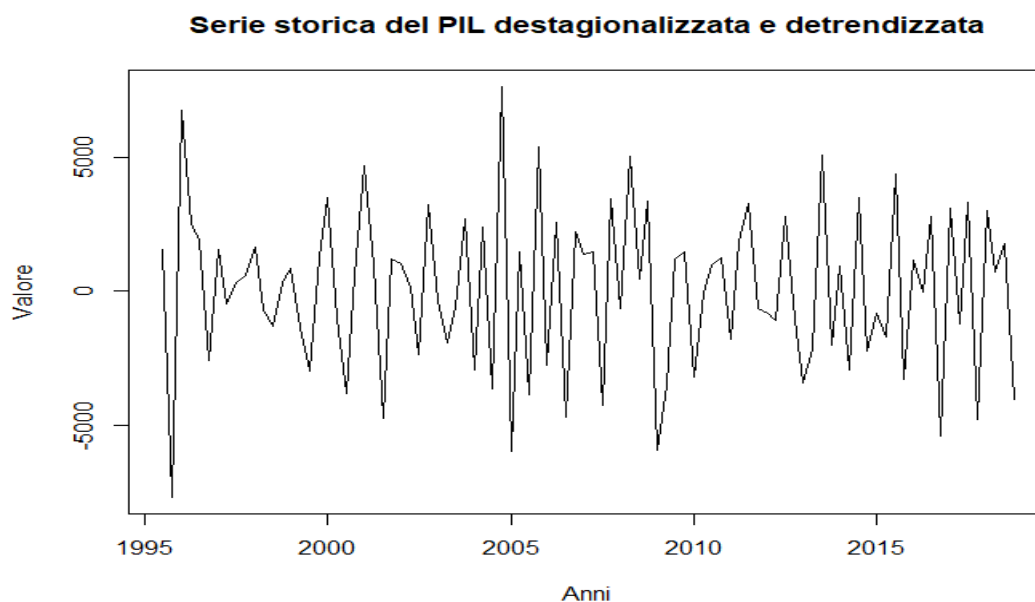


Figura 8 Rappresentazione della serie destagionalizzata e detrendizzata Italiana

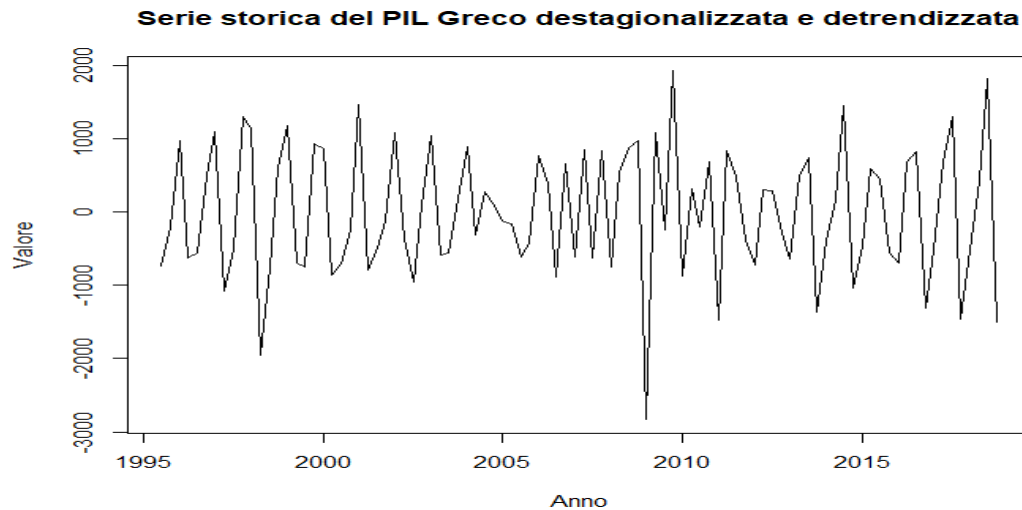


Figura 9 Rappresentazione della serie destagionalizzata e detrendizzata Greca

3.6 La normalità

Il primo step da seguire sarà la verifica della normalità delle serie attraverso il test di Jarque-Bera applicato a entrambe, che ci dà il seguente risultato per l'Italia:

```
> jarque.bera.test(piladj)

Jarque Bera Test

data: piladj
X-squared = 0.41338, df = 2, p-value = 0.8133
```

E il seguente per la Grecia:

```
> jarque.bera.test(piladjGR)

Jarque Bera Test

data: piladjGR
X-squared = 0.67576, df = 2, p-value = 0.7133
```

Essendo il p-value maggiore del livello di significatività scelto di 0.05 per entrambi i Paesi, è possibile accettare l'ipotesi nulla, affermando dunque che le trasformazioni delle due serie si distribuiscono normalmente.

3.7 L'ACF e la PACF

In questa fase dell'analisi si passa poi ad osservare e valutare l'ACF e la PACF per verificare se vi siano elementi per ricorrere ad un modello adatto. Come è possibile notare dai due grafici alle Figure 10 e 11, vi è ancora molta autocorrelazione, specialmente nell'ACF. In

particolare, in quest'ultimo, si nota un andamento sinusoidale che ci suggerisce di utilizzare un modello stagionale come il seasonal ARIMA per la completa cattura dell'informazione. Osservando poi la PACF si nota che non vi è più una correlazione parziale notevole tra un trimestre e lo stesso dell'anno precedente ma continua ad esserci con i trimestri dello stesso anno.

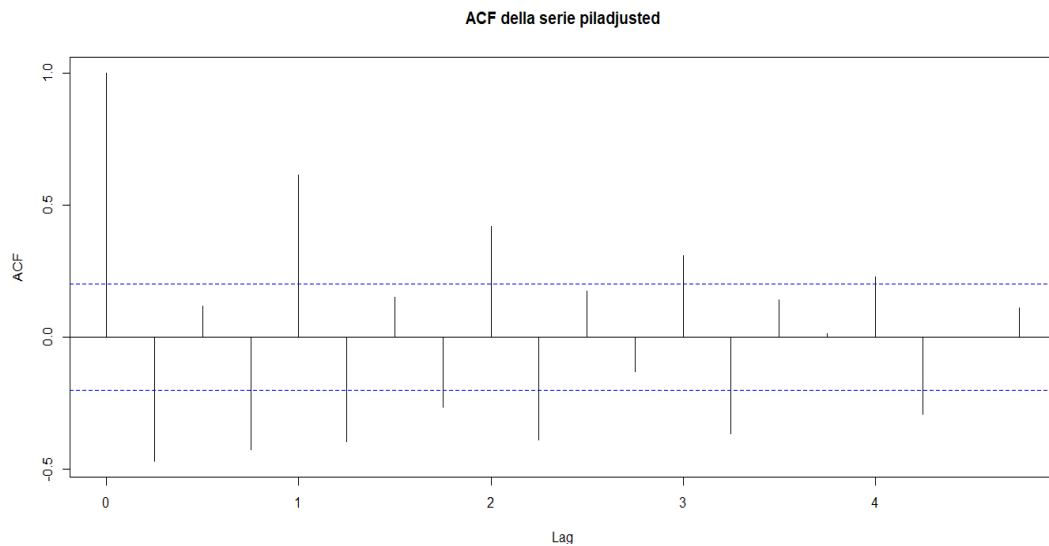


Figura 10 AutoCorrelationFunction della serie

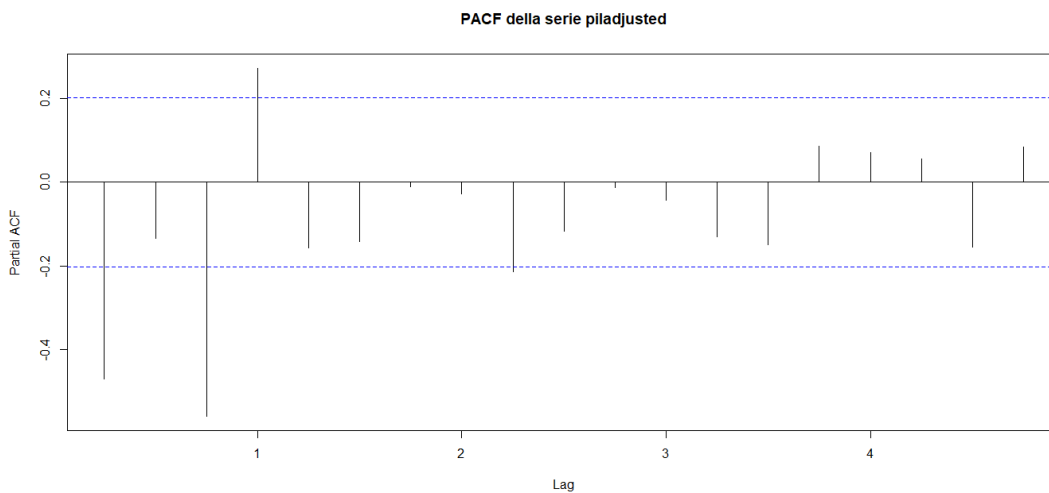


Figura 11 Partial AutoCorrelationFunction della serie piladj

Le autocorrelazioni globali e parziali della serie trasformata riguardante la Grecia, invece, sono rappresentate dalle figure 12 e 13 che seguono.

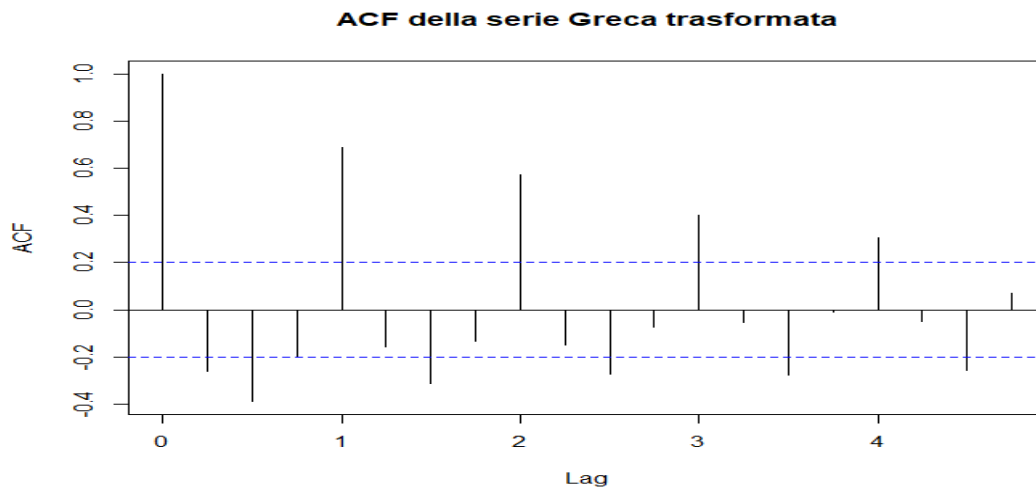


Figura 12 AutoCorrelation Function della serie trasformata Greca

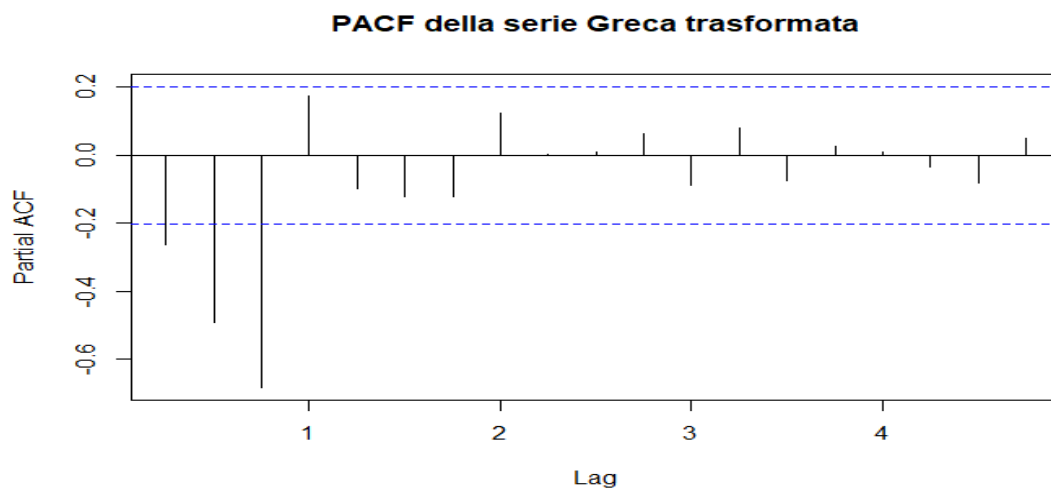


Figura 13 Partial AutoCorrelationFunction della serie trasformata Greca

A differenza delle funzioni delle autocorrelazioni per l'Italia, qui abbiamo una situazione abbastanza chiara del modello che le figure suggeriscono:

- Guardando la ACF nella figura 12 si notano ancora correlazioni che intercorrono tra un trimestre e lo stesso dell'anno prima evidenziando il bisogno di un modello stagionale
- La PACF invece è chiaramente il suggerimento di un modello con l'ordine p della parte autoregressiva pari a 3 dato che le autocorrelazioni parziali si annullano nel momento in cui $k > 3$.

3.8 La scelta del modello

Per rendere il processo di scelta del modello più preciso e veloce, il software R mette a disposizione il comando `auto.arima` grazie al quale sarà in grado di selezionare tramite un percorso iterativo il modello che minimizza il criterio AIC tra i molteplici testati dando, per la serie Italiana, il seguente output:

```
> modelloPIL <- auto.arima(piladj,
+                           trace = FALSE,
+                           ic = "aicc",
+                           approximation = FALSE,
+                           stepwise = FALSE)
> modelloPIL
Series: piladj
ARIMA(0,0,3) (1,0,0) [4] with zero mean

Coefficients:
      ma1      ma2      ma3      sar1
    -0.3982 -0.1722 -0.3837  0.6563
s.e.   0.0985  0.1028  0.1003  0.0889

sigma^2 estimated as 3863631:  log likelihood=-845.62
AIC=1701.24  AICc=1701.92  BIC=1713.96
```

Osservando l'output del software, è possibile osservare che il modello scelto è un seasonal ARIMA(0,0,3)(1,0,0), un modello adatto alle serie stagionali con ulteriori componenti autoregressive e medie mobili (P,D,Q), con un AIC di 1701.24. Qualora vi fosse anche la necessità di dimostrare che questo sia davvero il modello con l'AIC minore, basterebbe impostare il comando “`trace=TRUE`”: in questo caso l'output, sarà composto non solo dal modello che, secondo il software, è il migliore, ma anche tutti i modelli che sono stati iterativamente provati con i relativi AIC.

```
> modelloPILGR
Series: piladjGR
ARIMA(3,0,0) (1,1,1) [4]
Box Cox transformation: lambda= 1.034018

Coefficients:
      ar1      ar2      ar3      sar1      smal
    -0.5630 -0.4826 -0.5756 -0.3468 -0.5039
s.e.   0.1109  0.1181  0.1167  0.1390  0.1689

sigma^2 estimated as 421950:  log likelihood=-709.12
AIC=1430.25  AICc=1431.26  BIC=1445.25
```

Vediamo dunque dal modello stimato dal software che le osservazioni riguardanti le autocorrelazioni erano corrette in quanto è stato ipotizzato che il miglior modello sia un seasonal ARIMA(3,0,0)(1,1,1) con un Akaike Information Criterion di 1430.25, leggermente più basso rispetto al modello scelto per rappresentare la serie Italiana.

3.9 La verifica del modello

Per effettuare poi la verifica dei coefficienti stimati, R mette a disposizione, nel package “lmtest” il comando “coefstest” da cui risulta che per l’Italia abbiamo:

```
> coefstest(modelloPIL)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1  -0.398160   0.098472  -4.0434 5.269e-05 ***
ma2  -0.172174   0.102838  -1.6742 0.094086 .
ma3  -0.383741   0.100338  -3.8245 0.000131 ***
sar1  0.656319   0.088912   7.3817 1.563e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Dall’output mostrato risulta che tutti i coefficienti sono significativi in modo sufficientemente elevato, rilevando infatti che solo la componente MA2 è stata verificata con un livello di significatività α al di sopra dello 0. Per la verifica delle radici delle equazioni caratteristiche, in seguito, si utilizzerà il comando “plot” che darà come output due cerchi in cui saranno rappresentati i reciproci delle radici delle equazioni caratteristiche. Questi, in quanto reciproci, dovranno essere compresi tra -1 ed 1 per dimostrare la stazionarietà e l’invertibilità del processo. Il risultato è rappresentato dalla figura 9.

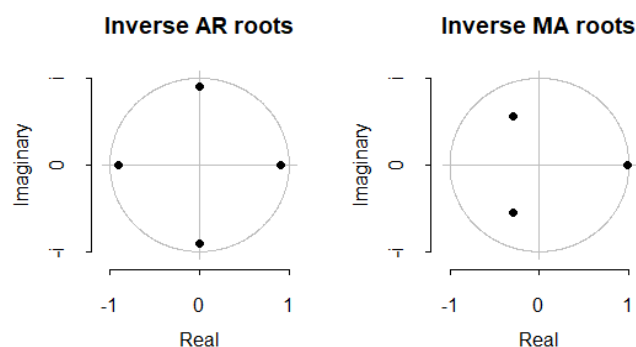


Figura 14 Rappresentazione grafica dei reciproci delle radici per la serie Italiana

Come è possibile osservare, tutti i reciproci delle radici delle equazioni caratteristiche sono all'interno dei due cerchi e ciò sta a significare che le radici, in modulo, sono tutte maggiori di 1 verificando sia l'invertibilità che la stazionarietà del processo.

Per quanto riguarda la verifica dei coefficienti del modello Greco, R ci restituisce il seguente output:

```
> coeftest(modelloPILGR)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  -0.56304    0.11089  -5.0776 3.821e-07 ***
ar2  -0.48264    0.11813  -4.0858 4.392e-05 ***
ar3  -0.57561    0.11667  -4.9337 8.069e-07 ***
sar1 -0.34682    0.13900  -2.4952 0.012590 *
smal -0.50391    0.16887  -2.9839 0.002846 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Come possiamo notare tutti i coefficienti della parte non stagionale sono verificati con significatività pari a 0 mentre, per la parte stagionale, queste sono verificate con significatività, rispettivamente, dello 0.05 e dello 0.01. Possiamo dunque affermare che tutti i coefficienti sono significativi e che sia possibile proseguire con l'analisi delle radici delle equazioni caratteristiche per assumere anche l'invertibilità e la stazionarietà.

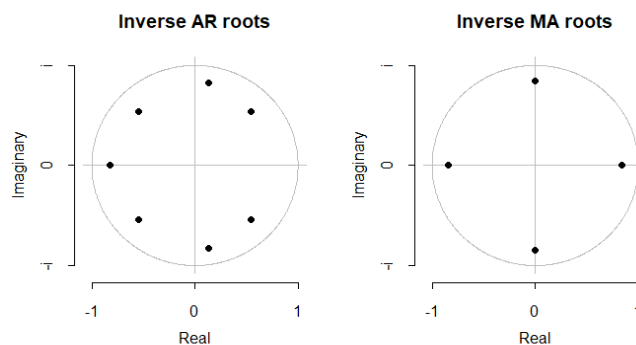


Figura 15 Rappresentazione grafica dei reciproci delle radici per la serie Greca

Dalla figura 15 si può notare che anche i reciproci delle radici delle equazioni caratteristiche associate al modello Greco sono tutti all'interno dei cerchi e dunque possiamo affermare che il processo è sia stazionario che invertibile.

In terza battuta, osservando la pratica teorica descritta nel capitolo precedente, si passa all'analisi dei residui che devono essere tra loro incorrelati ed essere distribuiti

normalmente. Per fare ciò R mette a disposizione il comando “checkresiduals” il quale, applicato al modello preso in considerazione, dà il seguente risultato:

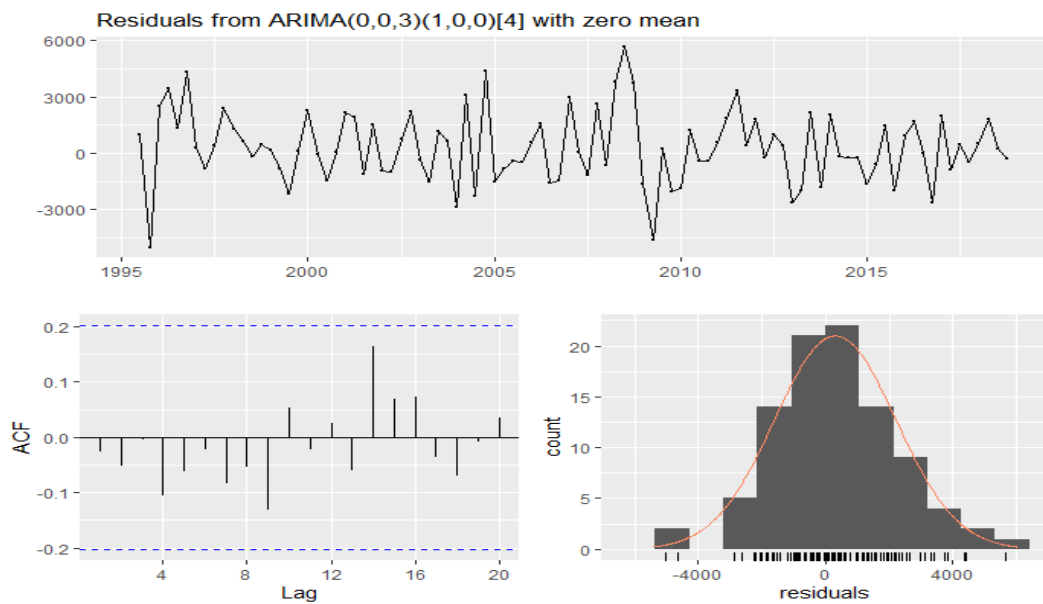


Figura 16 Analisi dei residui del modello Italiano

Dall’analisi dei residui notiamo che la media della distribuzione è intorno allo 0 e che, come evidenziato dall’ACF, i residui sono tutti incorrelati tra loro dato che nessuna delle barre supera l’intervallo di confidenza descritto dalla linea blu. Per quanto riguarda la normalità della distribuzione sarà necessario effettuare nuovamente il test JB:

```
> jarque.bera.test(res)

Jarque Bera Test

data:  res
X-squared = 0.64528, df = 2, p-value = 0.7242
```

Il p-value in questo caso è di 0.7242, sicuramente superiore al livello di significatività di 0.05, ci permette di accettare l’ipotesi nulla e, pertanto, affermando la normalità della distribuzione dei residui.

Svolti tutti i passaggi per la scelta e la verifica del modello con esito positivo, si può affermare che il modello scelto per lo studio del PIL è un $ARIMA(0,0,3)(1,0,0)$.

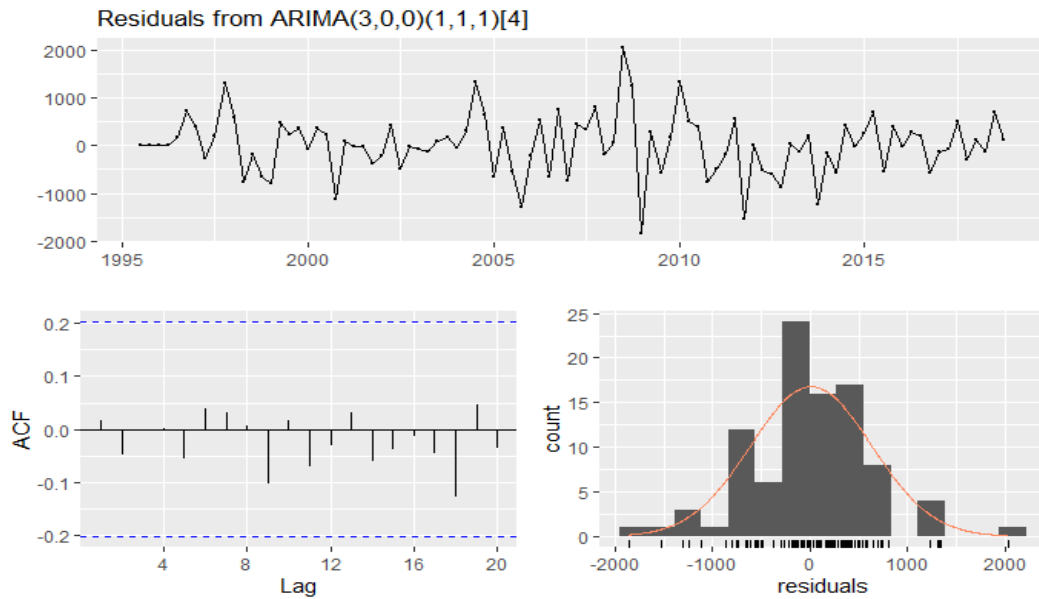


Figura 17 Analisi dei residui del modello Greco

Per la verifica dei residui associati al modello stimato per la serie storica trasformata Greca, possiamo notare che questi hanno media 0 e sono tutti totalmente incorrelati tra loro.

Per verificare la normalità dei residui ci affidiamo nuovamente al Jarque-Bera Test:

```
> jarque.bera.test(modelloPILGR$residuals)

Jarque Bera Test

data: modelloPILGR$residuals
X-squared = 6.4874, df = 2, p-value = 0.03902
```

Notiamo che questa volta per accettare l'ipotesi nulla che verifica la normalità degli errori, dovremmo impostare un livello di significatività di 0.03. Con questa assunzione è possibile accettarla e verificare come il modello si adatta ai dati e come si comporta al momento di prevedere dati futuri.

3.10 Il fitting del modello scelto

Per visualizzare anche graficamente quanto i modelli scelti siano precisi, è possibile rappresentarli insieme alle serie storiche trasformate: ciò verrà illustrato con le figure 18 e 19, utilizzando il comando “plot”, dove i modelli sono rappresentati dalle linee di colore blu mentre le serie storiche originali sono di colore rosso.

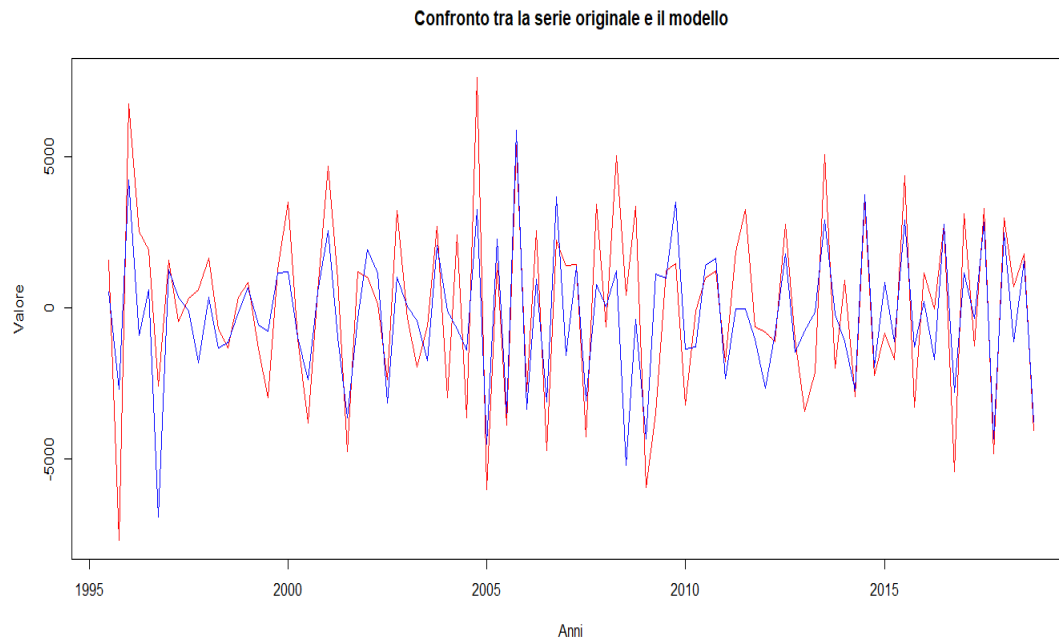


Figura 18 Confronto tra il modello e la serie originaria per l'Italia

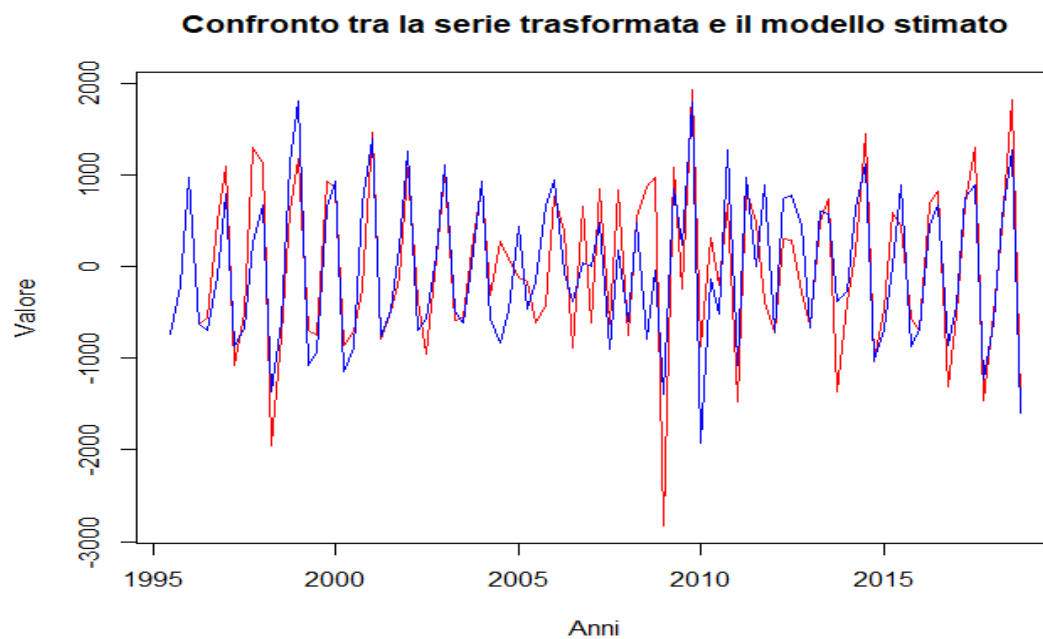


Figura 19 Confronto tra il modello e la serie originaria per la Grecia

Come confermato anche dalle rappresentazioni, i modelli, e in particolare il modello applicato alla serie Greca, esclusi i picchi rappresentanti i valori eccezionali, presenti specialmente nel periodo della crisi finanziaria, per l'Italia, e di fiducia, per la Grecia, hanno ben catturato tutte le informazioni dei dati. Questi modelli quindi si confermano adeguati e verranno utilizzati nel prossimo capitolo per formare le previsioni.

CAPITOLO QUATTRO

LE PREVISIONI

SOMMARIO: 4.1 Le previsioni dei dati futuri-4.2 La bontà previsiva del modello-4.3 Le previsioni in sample per il modello associato ai dati Italiani-4.4 Le previsioni in sample per il modello associato ai dati Greci-4.5 Le previsioni out of sample

4.1 Le previsioni dei dati futuri

L'ultima fase dello studio qui proposto è la formulazione delle previsioni dell'andamento del Prodotto Interno Lordo Italiano e Greco. Ciò sarà possibile grazie all'analisi svolta nel capitolo precedente. A questo scopo, infatti, sono stato ricercati e trovati i modelli ARIMA più adatti per la descrizione dei nostri dati, un seasonal ARIMA(0,0,3)(1,0,0) per i dati della serie storica trasformata Italiana e un seasonal ARIMA(3,0,0)(1,1,1) per i dati della serie Greca.

4.2 La bontà previsiva del modello

In prima battuta vedremo come i modelli analizzati si adattano ai dati originari utilizzando le previsioni ex-post o in sample. Valuteremo quindi la capacità previsiva degli stessi testandoli, mettendoli quindi in condizione di prevedere dei dati già noti così da poterli metterli a confronto. Si procede con i seguenti passaggi:

- Si seleziona un campione dai dati di partenza, ponendo come fine non più il secondo trimestre del 2019 ma il quarto trimestre del 2016;
- Si produrranno poi le previsioni con un modello riadattato alla nuova serie che sarà comunque un SARIMA(0,0,3)(1,0,0) o un SARIMA(3,0,0)(1,1,1).
- Si valuteranno poi gli indici ottenuti valutando quindi nel complesso la capacità del modello preso in considerazione per la formulazione delle previsioni.

Con R, dunque, andiamo ad eseguire i passaggi descritti. Per produrre i train set si utilizzerà il comando “window” rispettivamente per l'Italia, a sinistra, e la Grecia, a destra:

```
> train_set <- window(piladj, end = c(2016,4))
> train_setGR <- window(data[, "PIL"], end = c(2017, 4))
> train_setGR
```

	Qtr1	Qtr2	Qtr3	Qtr4		Qtr1	Qtr2	Qtr3	Qtr4
1995			1568.1496	-7702.3926	1995	23873.5	25099.8	27560.8	28040.2
1996	6770.0444	2529.6862	1931.7496	-2575.7926	1996	25083.5	27442.7	30607.6	31863.2
1997	1567.7569	-460.7513	308.7371	585.9324	1997	28304.2	29935.1	33287.7	34800.0
1998	1643.4819	-695.0013	-1325.1254	343.0449	1998	29891.5	30027.0	33755.1	35315.0
1999	858.5319	-1369.2138	-2971.6254	1266.0199	1999	31888.7	33760.6	36396.0	37845.3
2000	3506.4069	-1018.7138	-3821.1129	569.4199	2000	33198.8	34675.8	37285.0	37751.2
2001	4708.3444	961.1737	-4760.4879	1197.9824	2001	35236.6	36726.4	39842.8	40388.1
2002	1012.8194	161.5737	-2387.3754	3243.4199	2002	37497.1	39873.9	42337.7	43752.0
2003	-326.9556	-1942.2388	-596.5879	2715.8949	2003	40880.6	43418.8	46681.7	47923.8
2004	-2966.3431	2421.9737	-3637.9129	7652.7949	2004	44878.1	47605.0	50815.4	50417.3
2005	-6003.0306	1466.0487	-3892.6004	5410.3574	2005	45727.0	49176.8	51750.2	52588.3
2006	-2761.3306	2571.3487	-4711.5379	2232.3699	2006	50093.1	53992.3	55923.2	57852.9
2007	1401.4944	1457.2612	-4270.9629	3442.8824	2007	52830.7	58358.9	59929.0	61575.9
2008	-631.3181	5056.3612	440.9121	3361.5949	2008	55877.2	60744.8	63078.4	62290.0
2009	-5950.5806	-3499.2513	1223.2871	1470.5699	2009	53386.0	60220.5	61255.1	62672.7
2010	-3222.6056	-82.0763	998.0121	1249.1574	2010	54252.7	57353.1	57630.0	56795.6
2011	-1796.2306	1854.7487	3269.5996	-625.7926	2011	48813.8	53066.7	53779.1	51369.3
2012	-811.1431	-1106.2138	2788.3121	-1048.8426	2012	45124.8	48329.7	49747.5	48001.9
2013	-3415.7806	-2145.6513	5074.3871	-1996.2051	2013	42277.4	45866.2	47731.9	44778.8
2014	924.1069	-2941.1138	3498.5621	-2219.0676	2014	40832.0	44595.3	48115.3	45114.0
2015	-835.6931	-1720.8513	4374.0246	-3277.6426	2015	40716.3	44697.3	46710.0	45134.7
2016	1169.9694	-12.5513	2789.4996	-5420.9801	2016	40225.6	44678.7	47032.0	44551.7
					2017	40813.0	45424.5	48497.5	45482.6

4.3 Le previsioni in sample per il modello associato ai dati Italiani

Stimando nuovamente il modello, iniziando con quello associato ai dati Italiani, riadattato ai nuovi dati, con lo stesso comando usato in precedenza si otterrà:

```
> arima_train <- auto.arima(train_set,
+                           trace = FALSE,
+                           ic = "aicc",
+                           approximation = FALSE,
+                           stepwise = FALSE)
> arima_train
Series: train_set
ARIMA(0,0,3) (1,0,0) [4] with zero mean

Coefficients:
      ma1      ma2      ma3      sar1
    -0.3815  -0.1871  -0.392   0.6480
s.e.    0.1057   0.1074   0.109   0.0967

sigma^2 estimated as 4126025:  log likelihood=-776.47
AIC=1562.93  AICC=1563.68  BIC=1575.21
```

Questo, come possiamo vedere, è lo stesso modello ma con coefficienti leggermente diversi a causa della diversa dimensione del campione.

Si procederà poi con la formulazione delle previsioni in sample e la valutazione delle stesse con un unico comando di R, “accuracy”, creato appositamente:

```
> accuracy(forecast(arima_train, h = 4), piladj)
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 309.4773 1983.461 1548.638 -125.32404 269.60743 0.8063306
Test set    -122.9149 1502.713 1385.923   61.35926  61.35926 0.7216101
      ACF1 Theil's U
Training set -0.0274758      NA
Test set    -0.5720002 0.1690322
```

L’output di R ci mostra che le nostre previsioni, pur avendo un Root Mean Squared Error, banalmente la radice quadrata del MSE esposto nel secondo capitolo, abbastanza elevato, possiede un Coefficiente di Theil particolarmente basso, 0.1690322, e dunque può essere utilizzato per le previsioni out-of-sample.

4.4 Le previsioni in sample per il modello associato ai dati Greci

Passando poi all’analisi delle previsioni in sample utilizzando gli stessi step eseguiti in precedenza ma con il modello associato alla serie Greca si otterrà:

```

> arima_train <- auto.arima(train_set,
+                             trace = FALSE,
+                             ic = "aicc",
+                             approximation = FALSE,
+                             stepwise = FALSE,
+                             lambda = "auto")
> arima_train
Series: train_set
ARIMA(3,0,0) (1,1,1) [4]
Box Cox transformation: lambda= 1.034737

Coefficients:
      ar1      ar2      ar3      sar1      smal
    -0.5596 -0.4815 -0.5643 -0.3483 -0.5020
s.e.   0.1137   0.1201   0.1189   0.1412   0.1695

sigma^2 estimated as 440113:  log likelihood=-679.35
AIC=1370.7   AICc=1371.76   BIC=1385.42

```

Anche per quanto riguarda questi dati, il modello stimato è lo stesso, differendo da quello stimato precedentemente, di pochissimo, solo per i valori dei coefficienti.

Utilizzando nuovamente il comando “accuracy” vediamo poi come il modello si comporta con le previsioni interne alla serie, analizzandone in seguito i risultati:

```

> accuracy(forecast(arima_train, h = 5), piladjGR)
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set  3.233831 508.6213 375.8432 38.610159 82.85282 0.7581021
Test set      76.641289 331.5851 269.6231 -2.694608 26.31383 0.5438486
      ACF1 Theil's U
Training set  0.0188400      NA
Test set      -0.6840544  0.344914

```

Fin dalla prima analisi si può subito notare che il Root Mean Squared Error è decisamente più basso rispetto al modello utilizzato per i dati Italiani, vediamo che anche il Mean Absolute Percentage Error (MAPE) è molto basso denotando quindi una buona precisione nella previsione in sample. L’unico dato più alto, ma comunque ottimo, è il Coefficiente di Theil, stimato a 0.344914 che denota comunque un’ottima bontà previsiva del modello. Possiamo quindi dire che il modello è adatto per formulare le previsioni out of sample esplicate nel prossimo paragrafo.

4.5 Le previsioni out of sample

Per fare ciò si utilizza il pacchetto “forecast” di R e il comando omonimo, l’output sarà rappresentato dalla Figura 20 dove le bande blu rappresentano la probabilità che il dato si troverà in quella posizione, più ci si allontana dalla linea di previsione più la banda diventa

di un colore più chiaro, ciò per indicare la scarsa probabilità che il valore futuro si trovi in quella posizione.

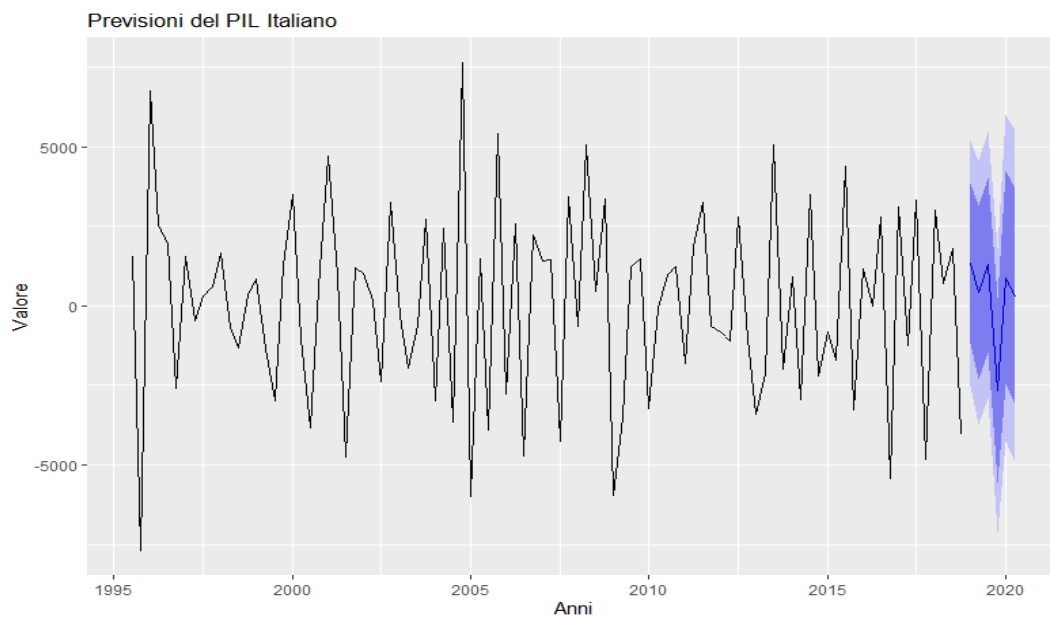


Figura 20 Previsioni formulate col modello scelto

Nella figura 21 vengono poi riportati i dati analitici risultanti dalle precedenti previsioni. C'è comunque da tener presente che il set di dati originali è stato protagonista di trasformazioni volte all'eliminazione del trend e della stagionalità e che dunque, per poter effettuare dei paragoni, sarebbe necessario ricorrere ad ulteriori trasformazioni. Vi è comunque da osservare che la bassa differenza rispetto al passato tra il valore dei trimestri e il loro precedente è particolarmente basso e ciò a conferma delle previsioni di stagnazione economica effettuate dal Fondo Monetario Internazionale che ci riguardano.

Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95	Indicator
1335.0070	-1170.475	3858.2490	-2504.375	5197.306	PIL
416.0069	-2284.380	3126.4581	-3723.065	4566.658	PIL
1292.8816	-1439.136	4043.5206	-2894.143	5503.446	PIL
-2660.1484	-5582.138	242.0931	-7131.882	1779.183	PIL
875.4057	-2460.489	4228.4771	-4238.058	6009.259	PIL
273.0312	-3131.222	3685.4085	-4944.262	5499.441	PIL

Figura 21 Le previsioni per la serie trasformata Italiana

Si passa, infine, alle previsioni out-of-sample per il secondo modello, rappresentato graficamente nella Figura 22.

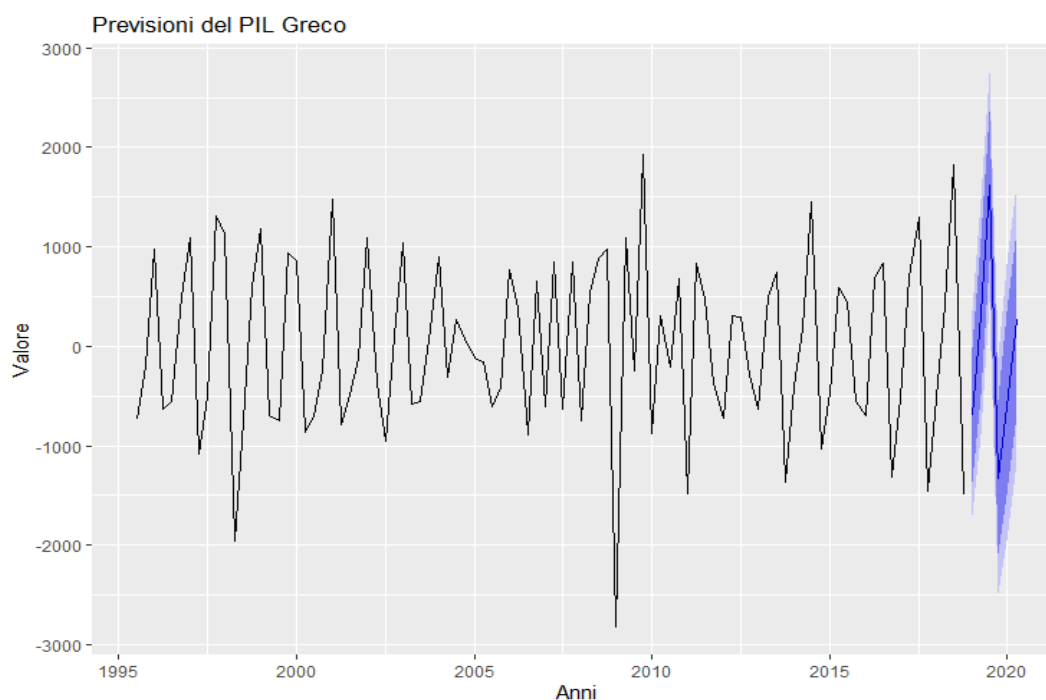


Figura 22 Rappresentazione grafica delle previsioni della serie trasformata Greca

Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95	Indicator
-702.9291	-1360.7372	-16.14459	-1704.1324	358.5099	PIL
315.1131	-489.7501	1081.07741	-894.8001	1477.6150	PIL
1618.1254	860.1658	2363.70870	450.1217	2754.9755	PIL
-1322.4133	-2084.2915	-544.21681	-2483.5073	-116.7546	PIL
-605.9019	-1498.0788	339.74485	-1962.0060	821.6560	PIL
266.3621	-685.4746	1179.36922	-1163.7415	1650.6589	PIL

Figura 23 Risultati analitici delle previsioni sul modello associato ai dati Greci

Come ultima analisi possiamo osservare dalla Figura 23 i valori futuri stimati dal modello. Possiamo osservare come, rispetto all'Italia, ricordando naturalmente la grandezza dei due aggregati, la Grecia sia in una fase di leggera ripresa in quanto i valori sembrano essere abbastanza elevati positivamente, confermando dunque le previsioni inserite nel World Economic Outlook del FMI.

CONCLUSIONI

L'elaborato che va concludendosi aveva come obiettivo la stima di modelli che fossero in grado di adattarsi ai dati in modo tale da formulare delle previsioni dei valori futuri.

Nel primo capitolo è stata esposta la storia della nascita del PIL con l'annessione, inoltre, delle critiche e dei pareri favorevoli e di non superabilità di questo strumento. Nello stesso capitolo è stata poi esplicitata la composizione dell'aggregato e, attraverso l'esposto del Fondo Monetario Internazionale, è stata introdotta una panoramica mondiale e nazionale della situazione economica.

Nel secondo capitolo è stata poi sintetizzata la teoria che è alla base delle rilevazioni e del processo di scelta del modello adatto alla rappresentazione dei dati e, attraverso l'ausilio del software statistico R, uno dei programmi più diffusi, utilizzati e in continua espansione, e dall'analisi delle ACF e PACF, nel terzo capitolo, si sono stimati i modelli migliori che, dopo un'accurata fase di validazione, sono stati utilizzati per riprodurre le serie storiche detrendizzate e destagionalizzate. In particolare, per la serie dei dati Italiani è stato utilizzato un modello SARIMA(0,0,3)(1,0,0), mentre per i valori dei caratteri Greci è stato scelto un modello SARIMA(3,0,0)(1,1,1).

Una volta svolta l'analisi dei modelli si è giunti alla conclusione con la stima delle previsioni, obiettivo principale dell'elaborato. Sono prima state effettuate delle previsioni ex post in modo tale da verificare anche la bontà previsiva del modello oltre che la sua capacità di adattarsi ai dati originari, dopodiché sono state formulate anche le previsioni ex ante per gli ultimi due trimestri del 2019 e i quattro trimestri del 2020.

Possiamo concludere affermando che, pur non essendo certi, i valori riscontrati sono compatibili con le previsioni effettuate dai principali istituti statistici mondiali e che quindi hanno una buona probabilità di rendersi veritieri, sostenuti anche dai Coefficienti di Theil particolarmente vicini allo 0 e, specialmente nel caso dell'analisi dei dati Greci, un ottimo MAPE e RMSE.

BIBLIOGRAFIA E SITOGRAFIA

1. VITALE C., *Introduzione alla statistica per le applicazioni economiche Volume I e II*, E.S.I., 2002
2. ESTELA BEE DAGUM, *Analisi delle serie storiche*, Springer Verlag Italia, 2002
3. BROCKWELL, DAVIS, *Introduction to time series and forecasting*, Springer Verlag, 2002
4. CRYER, CHAN, *Time series analysis with applications in R*, Springer Verlag, 2008
5. SHUMWAY, STOFFER, *Time series analysis and its applications with R examples*, Springer Verlag, 2005
6. BLANCHARD, AMIGHINI, GIAVAZZI, *Macroeconomia*, Il Mulino, 2016
7. SAMUELSON, *Economics*, McGraw-Hill Companies, 1995
8. IGNACIO FERNANDO LARA, *Quale dibattito sul PIL?*, Impresa&Stato, 2010
9. ILSOLE24ORE, *Allarme Fmi: ripresa precaria nel 2020. Per l'Italia crescita allo 0,1%*, 23 Luglio 2019
10. EISNER R., *The total incomes system of account*, University of Chicago Press, 1989
11. www.italianieuropei.it
12. www.milomb.camcom.it
13. www.imf.org
14. www.eur-lex.europa.eu
15. www.proversi.it
16. www.wikipedia.it
17. www.treccani.it
18. www.okpedia.it
19. www.treccani.it
20. www.nber.org
21. www.foreignpolicy.com
22. www.you-ng.it
23. www.st.ilssole24ore.it
24. www.ilpost.it

RINGRAZIAMENTI

Giunto al termine dello studio, vorrei ringraziare le persone che più mi sono state vicine in questo percorso e che mi hanno supportato e sopportato in questi tre anni.

Grazie, mamma e papà, che nonostante i mille problemi, le mille incomprensioni e l'iniziale, giustificata, sfiducia, mi siete stati comunque vicino e, cosa più importante, mi avete cresciuto, mi avete dato un'educazione, una casa, il piatto caldo a tavola, un letto e l'affetto, vi voglio bene.

Grazie Agostino, sempre severo con tutti anche se, in fondo in fondo, sappiamo tutti che hai un cuore d'oro e che ogni volta che ti è possibile aiuti chi ti è caro dando tutto ciò che puoi, mi spiace solo che le nostre strade siano state separate dalla distanza fin da quando ero un bambino ma so che nel momento della necessità ci sarai sempre e lo stesso varrà per te, ricorda il nostro patto. Ringrazio ovviamente anche te, Dalila, per la tua immensa pazienza nei miei confronti quando, da piccolo, proprio non ne volevo sapere di fare quei compiti, per la tua bontà, per i tuoi consigli. Ti ringrazio, inoltre, anche per sopportare mio fratello.

Grazie Serena, anche se i chilometri ci hanno divisi siamo e rimarremo sempre legati da ciò che siamo stati e ciò che siamo, sei la mia sorellina, sei stata la mia spalla quando tutto andava a rotoli, uno dei miei punti fermi della mia adolescenza. Grazie della premura che hai dimostrato nei miei confronti, grazie Teresa. Giovanni voglio ringraziare anche te, che hai sempre trattato con i guanti la mia famiglia un po' complicata, ti ringrazio per essermi stato vicino e per avermi consigliato sempre in modo saggio e ponderato e, anche a te ovviamente, per aver sopportato la mia sorellina.

Grazie Martina, in questo poco tempo sei riuscita a portare nella mia vita la tranquillità che tanto avevo cercato, a darmi la parvenza di ciò che pensavo che non potesse esistere, a farmi pensare che forse l'umanità ha ancora una speranza. Hai sopportato anzi, scusa altrimenti ti arrabbi, hai supportato ogni mia scelta, mi hai tirato su ogni volta che la mia mente andava a spasso cercando un problema per ogni soluzione, ad ogni mio crollo, ad ogni pensiero irrazionale, ad ogni ansia e timore. Grazie per avermi riportato sempre con i piedi a terra, grazie di regalarmi i tuoi sorrisi pieni di vita, ricorda che sei e sarai importante per me, e, anche se non te l'ho mai detto a parole ma solo con i fatti e le dimostrazioni, ti amo. Mi fermo qui altrimenti diventi troppo rossa in viso. Grazie davvero.

Elia, Domenico e Orlando, eccoci, è il vostro turno. Ragazzi sul serio, non so come avrei fatto senza voi, quelli di sempre, quelli che per me sono come fratelli ormai da 20 anni. Non penso che basterebbe un intero libro per descrivere ciò che abbiamo passato insieme. Vi ringrazio perché mi avete sempre accettato così come sono, anche se a volte, ammetto, non

lo meritavo. Vi ringrazio perché grazie a voi so che la vera amicizia esiste, che anche i chilometri non contano nulla quando c'è qualcosa che lega e avvicina. Non saremo perfetti ma siamo veri, ci vogliamo bene e qualunque cosa accada saremo sempre e comunque qui, a casa mia, davanti ad una goccia di caffè, a tirarci su a vicenda da qualunque cosa.

Giorgio, sì, ne ho anche per te. Sei la dimostrazione che all'Università, oltre che colleghi, si possono trovare anche veri amici. Questi tre anni sono volati ma rimarranno sempre nel mio cuore. Sappi che quando avrai bisogno di un consiglio o, molto più probabile, una sgridata per farti ritornare sulla Terra, io ci sarò. Sai già quale domanda ci sarebbe qui ora, se solo potessi scriverla.

Grazie Concetta e Chiara, nonostante tutto mi avete aiutato, specialmente nella mia adolescenza, a superare gli ostacoli che si ponevano sul mio percorso. Grazie per avermi sostenuto e aver trovato sempre una parola di conforto.

Grazie a Aleksandra, Mena, Angela, Marco, Vincenzo, Antonella e Valentina. Ragazzi mi mancherete, sono stati tre anni di studio intenso ma, grazie anche a voi, sono stati anche i più belli.

Grazie alla Professoressa Alessandra Amendola, la mia relatrice, per essere stata sempre disponibile durante il percorso di stesura di questo elaborato dandomi, nel momento del bisogno, l'indicazione giusta. La ringrazio, inoltre, per avermi fatto appassionare, con la sua determinazione e chiarezza, alla Statistica, ambito nel quale voglio specializzarmi in futuro.