Università degli Studi di Padova

Dipartimento di Diritto Pubblico, Internazionale e Comunitario

Corso di laurea in Diritto e Tecnologia
a.a. 2023/2024

La previsione di variabili economiche. Il pacchetto forecast di R e due applicazioni: PIL e Indice della produzione industriale

Relatore
Prof. Tommaso Di Fonzo

Studente
Bortolozzo Francesco
Matricola
2040742

Indice

Elenco delle figure

Introduzione

La previsione di variabili economiche riveste un ruolo fondamentale nell'ambito della ricerca statistica. Per decisori politici, investitori ed imprese è di cruciale importanza avere la capacità di anticipare gli andamenti a venire, in modo da poter adottare la strategia migliore per il futuro, evitando i rischi e massimizzando le opportunità. Ecco che quindi la previsione accurata di variabili macroeconomiche come il Prodotto Interno Lordo (PIL) o l'Indice della Produzione Industriale (IPI) diventa uno strumento primario. Le variabili di interesse considerate sono trattate come serie storiche, ossia come una sequenza di osservazioni ordinate nel tempo. Lo scopo diventa quindi quello di studiare l'evoluzione passata del fenomeno rispetto al tempo, cercando di prevedere l'andamento futuro. Questa tesi ha l'obiettivo di analizzare l'efficacia e la semplicità nell'utilizzo del pacchetto **forecast** di R[1], uno strumento ampiamente riconosciuto nell'analisi delle serie storiche, che offre un'ampia gamma di metodi avanzati per la modellazione, l'analisi e la previsione di dati complessi ordinati nel tempo. L'integrazione nell'ambiente di programmazione R facilita l'uso pratico dei risultati e rende le analisi facilmente replicabili, il che è molto prezioso per ricercatori e professionisti del settore economico e non. Il lavoro è quindi strutturato come segue:

- **Capitolo 1:** Viene presentato nello specifico il concetto di serie storica. Successivamente, vengono descritti i dati utilizzati, ossia le fonti e le caratteristiche dei dati, con particolare riguardo alle trasformazioni preliminari e alla gestione della stagionalità.

- **Capitolo 2:** Viene fornita un'introduzione al pacchetto **forecast** di R e vengono descritti i principali modelli utilizzati durante il lavoro.

- **Capitolo 3:** I modelli di previsione vengono applicati alle serie storiche del PIL e dell'IPI, in modo da fornire una visione completa ed accurata delle previsioni, discutendo i risultati e confrontando i modelli.

- **Capitolo 4:** Vengono offerte delle considerazioni finali sull'efficacia dei modelli utilizzati.

[1] indirizzo web della pagina del CRAN dedicata a forecast:https://cran.r-project.org/web/packages/forecast/index.html

Attraverso questa struttura, il lavoro mira a fornire un'analisi approfondita e applicata delle tecniche di previsione delle variabili macroeconomiche, quali PIL e IPI, utilizzando il pacchetto **forecast** di R come strumento principale. L'obiettivo è di valutare la semplicità e l'effiacia del pacchetto per poter avere una migliore comprensione e previsione delle dinamiche economiche, da utilizzare a supporto dei processi di decisione economica.
Tutti i codici utilizzati per ottenere le previsioni sono disponibili nel repository GitHub all'indirizzo https://github.com/THBfra/Tesi-.

Capitolo 1

L'analisi delle storiche

1.1 Le Serie Storiche

Una serie storica può essere considerata come una successione di valori, completa delle informazioni su quando quei valori si sono realizzati (Hyndman, 2018). I valori delle variabili sono osservati per intervalli di tempo che possono essere giornalieri, settimanali, mensili, trimestrali e così via. Vari modelli di previsione sono poi stati applicati alle serie storiche in modo da poter stimare gli andamenti futuri. Nel contesto attuale, caratterizzato da una crescente attenzione alla raccolta e all'analisi dei dati, le serie storiche assumono un'importanza cruciale. Oggi, infatti, esistono grandi banche dati, come quelle fornite dall'ISTAT, dove sono raccolte enormi quantità di serie storiche disponibili senza oneri. Queste risorse permettono a chiunque lo desideri di accedere a dati dettagliati e aggiornati, offrendo l'opportunità di effettuare analisi approfondite in diversi settori. Nel fare riferimento alle serie storiche è necessario introdurre i concetti di Trend, Stagionalità e Ciclo, riferiti a caratteristiche che si possono riscontrare nelle serie storiche.

1.1.1 Trend

Un Trend esiste quando avviene un incremento o un decremento, non necessariamente lineare, dei dati nel lungo termine. Come mostrato nella Figura 1.1, il grafico evidenzia come sia presente un netto trend decrescente.

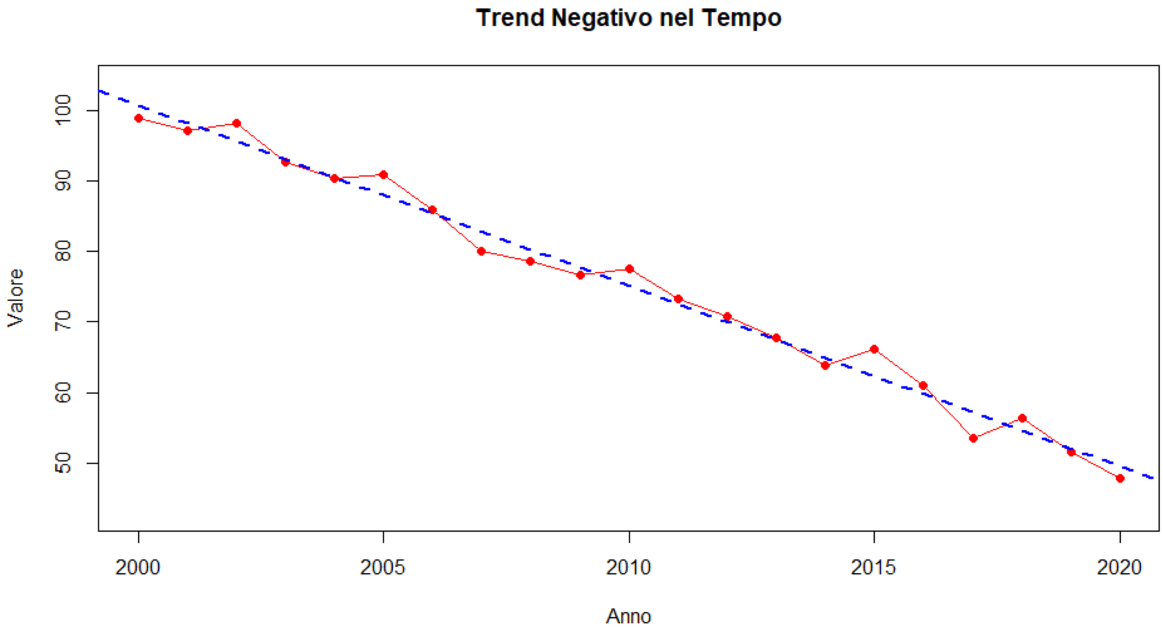

Figura 1.1: Trend Descrescente

1.1.2 Stagionalità

Un andamento Stagionale si verifica quando una serie storica è influenzata da fattori ripetuti nel tempo in modo regolare, come il periodo dell'anno o il giorno della settimana. La stagionalità ha sempre una frequenza fissa e nota. Il grafico nella Figura 1.2, mostra come siano presenti tanto un trend crescente, quanto una marcata stagionalità visibile in forma di andamento ripetuto nel corso di ciascun anno.

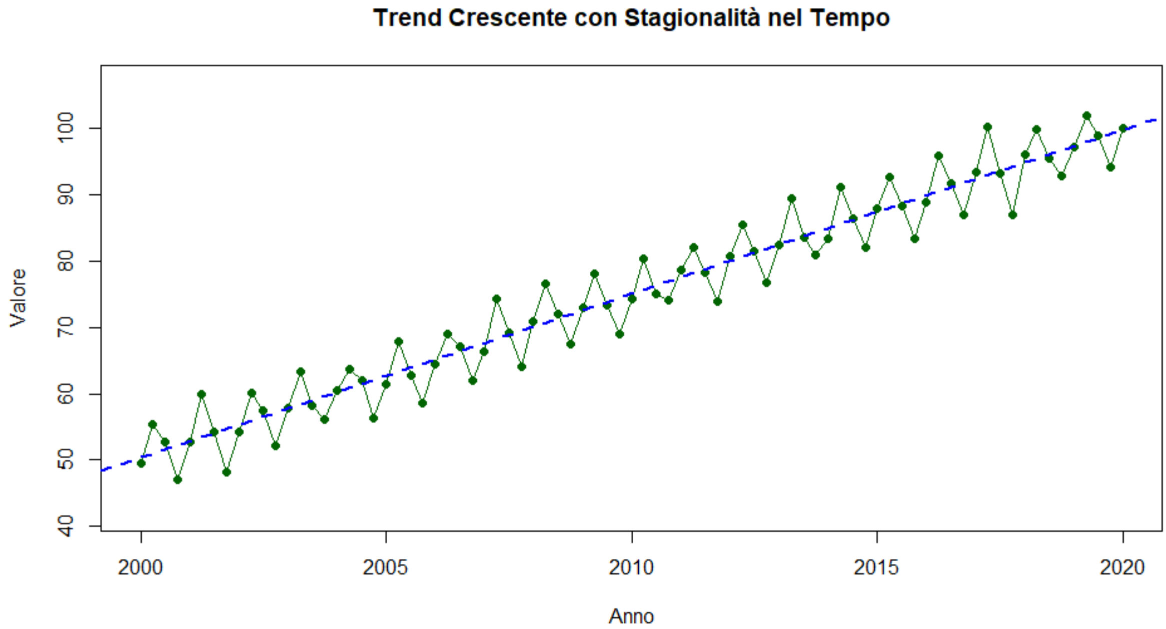

Figura 1.2: Stagionalità

1.1.3 Ciclo

Una componenente ciclica si registra quando i dati mostrano aumenti e diminuzioni che non hanno una frequenza fissa, spesso dovuti a condizioni economiche specifiche. È fondamentale riconoscere la differenza tra stagionalità e ciclicità: se le fluttuazioni non hanno una frequenza fissa, allora sono cicliche; se la frequenza è costante e associata a qualche aspetto del calendario (come giorni, settimane, mesi o anni), allora il pattern è da considerare stagionale. Inoltre la ciclicità è molto più facile da riconoscere quando le serie storiche sono molto lunghe. Come mostrato nella Figura 1.3, il grafico evidenzia una componente ciclica, ossia fluttuazioni ricorrenti a medio-lungo termine che non seguono una periodicità fissa.

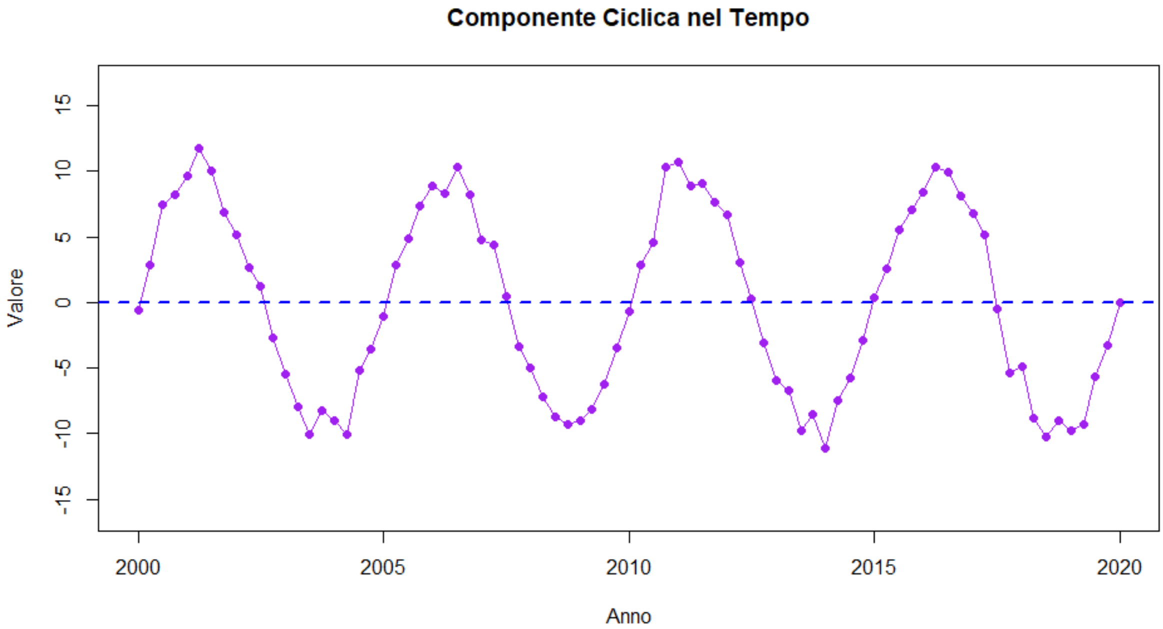

Figura 1.3: Componente ciclica

1.1.4 Componente irregolare

È importante evidenziare come non tutte le serie storiche presentano necessariamente le tre componenti summenzionate, come mostrato in Figura 1.4. È possibile che ciò accada quando il fenomeno osservato è influenzato per la maggior parte da variazioni casuali o irregolari, che rendono difficile l'identificazione di pattern sistematici nel tempo. Di conseguenza, nell'analisi delle serie storiche è fondamentale verificare l'effettiva presenza di tali componenti prima di procedere con modelli che intendono sfruttarle, per evitare conclusioni errate o fuorvianti.

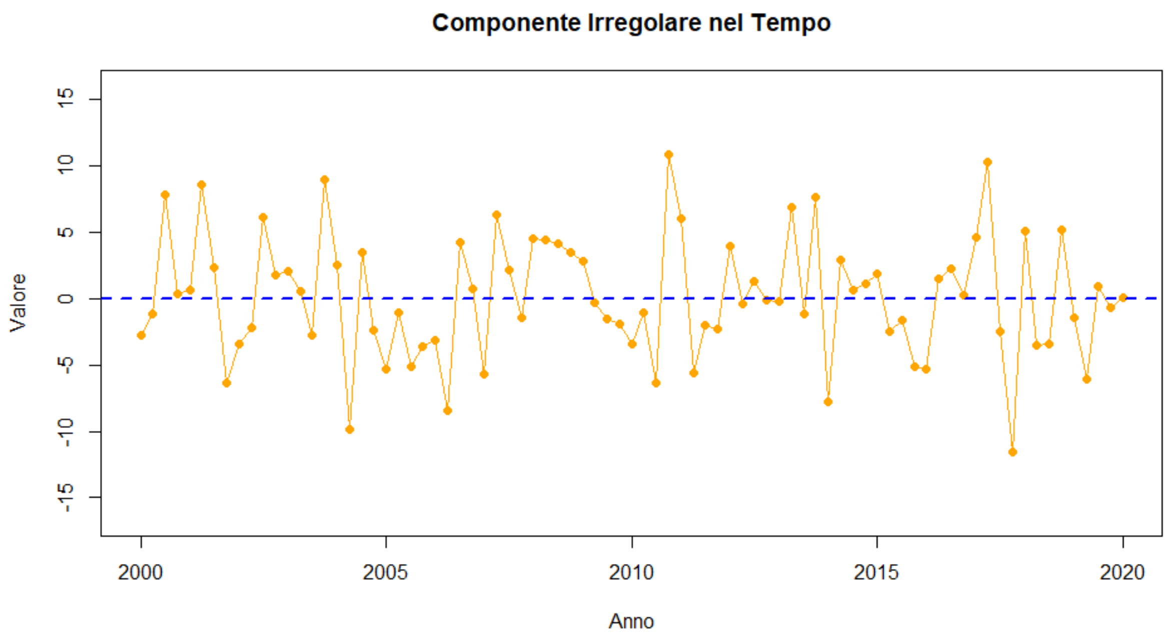

Figura 1.4: Componente irregolare

1.2 Descrizione dei dati

1.2.1 Prodotto Interno Lordo (PIL)

Il Prodotto Interno Lordo (PIL) è una variabile macroeconomica fondamentale che misura il valore monetario totale di tutti i beni e servizi finali prodotti all'interno dei confini di un paese in un periodo specifico (Mankiw, 2019). Il PIL è comunemente utilizzato per valutare la performance economica di una nazione, indicando la sua salute economica e la crescita nel tempo. All'interno di questo lavoro analizzeremo la serie storica trimestrale del PIL italiano con dati destagionalizzati e valori concatenati a prezzi di mercato del 2015 dal primo trimestre del 1996 (1996-Q1) al secondo trimestre del 2024 (2024-Q2), fornendo delle analisi e delle previsioni grazie all'utilizzo di quasi tre decenni di dati ottenuti all'interno della banca dati di serie storiche dell'ISTAT visualizzabile in Figura 1.5. La serie storica può essere così visualizzata su R:

```
# Caricare i pacchetti
library(readxl) #libreria che permette la lettura dei file Excel
library(ggplot2) #libreria che permette la visualizzazione grafica
library(zoo) #libreria che permette di gestire la serie storica

# Leggere il file Excel
```

```r
file_path <- "percorso/del/file/PIL_Trasposta.xlsx"
df <- read_excel(file_path)

# Controllare la struttura del dataframe
str(df)

# Convertire i valori della colonna Tempo
# In formato "1996-Q1", "1996-Q2", etc.
df$Tempo <- as.yearqtr(df$Tempo, format = "%Y-Q%q")

# Creare il grafico
ggplot(df, aes(x = Tempo, y = PIL)) +
    geom_line(color = "blue") +
    labs(title = "Andamento del PIL Trimestrale (1996-Q1 2024-Q2)",
          x = "Anno-Trimestre",
          y = "PIL") +
    theme_minimal()    #Funzione di ggplot2 che permette una visualizzazione semplice
```

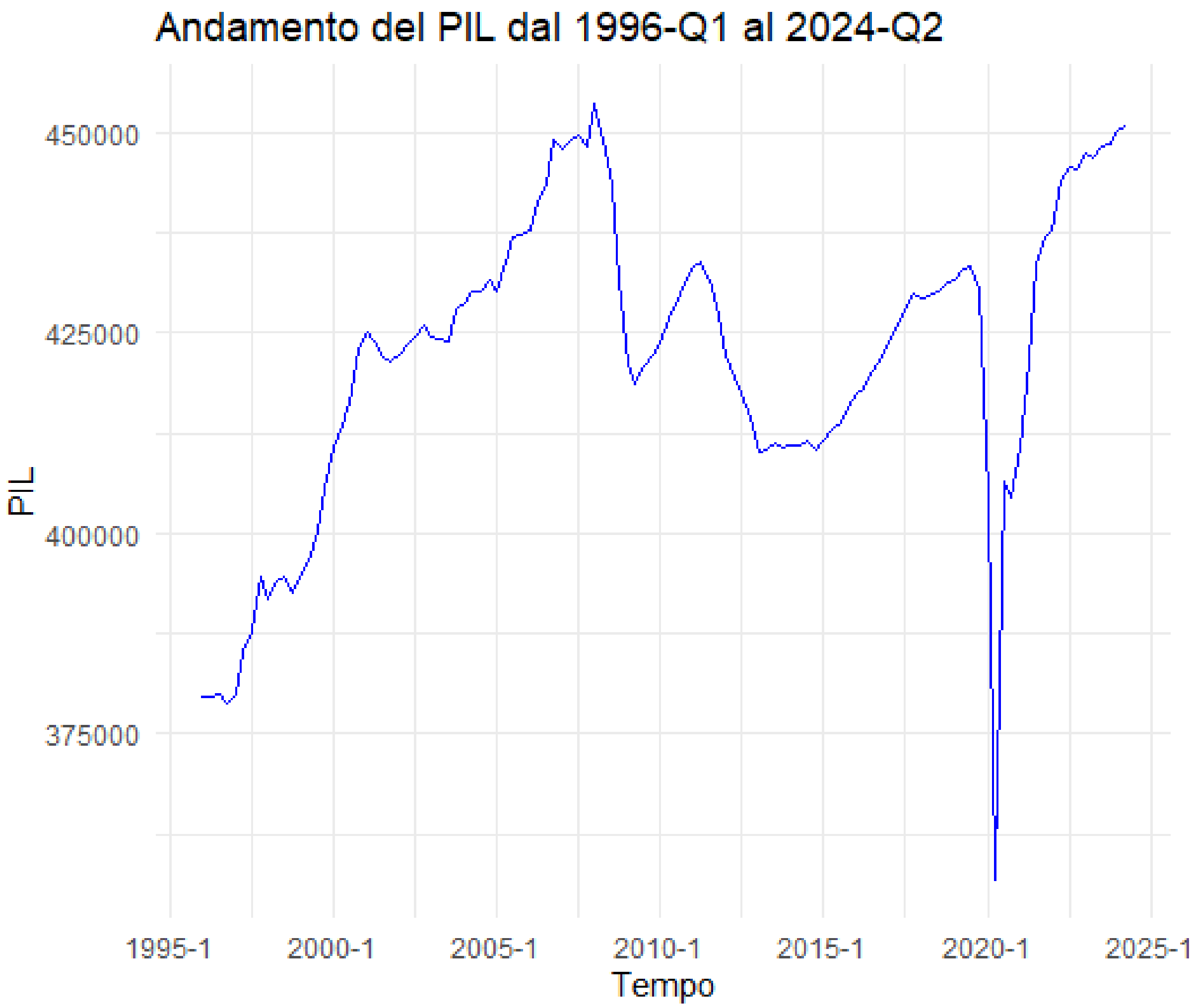

Figura 1.5: Serie Storica del PIL 1996-2024

1.2.2 Indice della Produzione Industriale (IPI)

L'Indice della Produzione Industriale (IPI) è una variabile macroeconomica che misura le variazioni nel tempo della produzione fisica nel settore industriale di un paese, escluse le costruzioni. Durante il lavoro verrà analizzata la serie storica dell'IPI mensile dal gennaio 1990 (1990-01) all'agosto 2024 (2024-08), la quale viene fornita in dati grezzi, come mostrato in Figura 1.6. Come detto precedentemente, il metodo per caricare i file su R è il medesimo se non per il fatto che verrà modificato il formato temporale data la frequenza mensile dei dati. Anche in questo caso i dati sono stati scaricati dal sito dell'ISTAT.

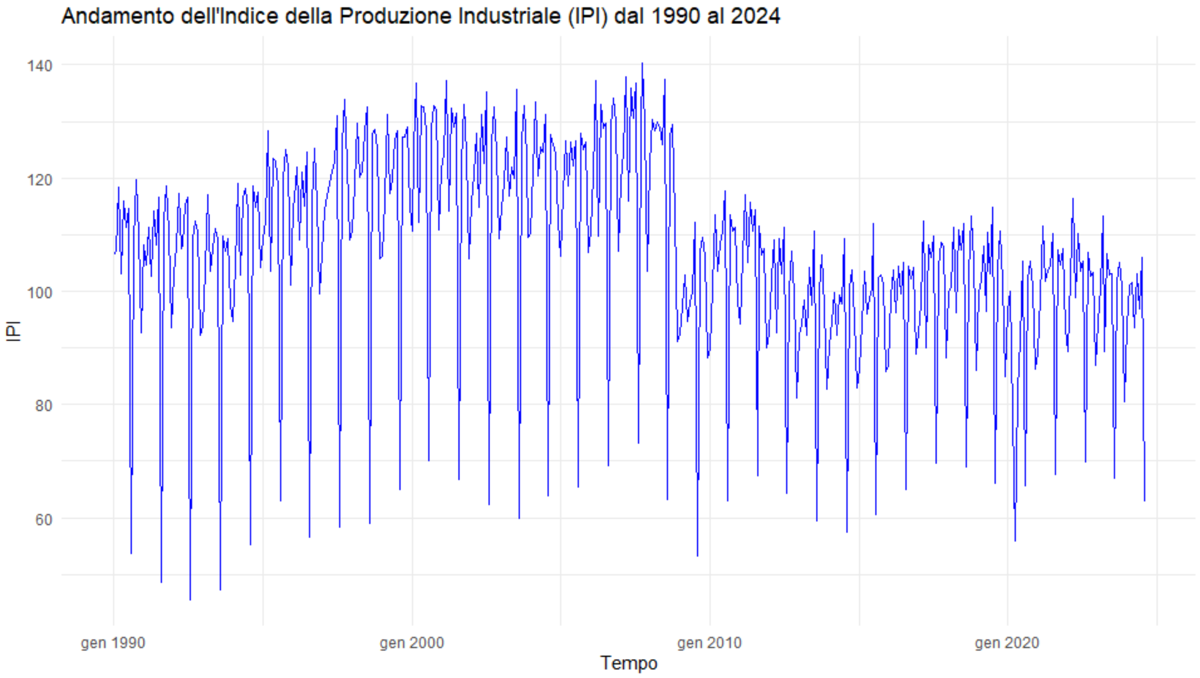

Figura 1.6: Serie Storica dell'IPI 1990-2024

Le serie storiche del PIL e dell'IPI offrono una testimonianza evidente dell'impatto profondo che la pandemia del COVID-19 ha avuto sull'economia nazionale. Fino all'inizio del 2020, la serie storica del PIL mostrava un trend positivo abbastanza costante, riflettendo una fase di crescita economica sostenuta. Tale andamento era il risultato dell'insieme di politiche economiche favorevoli, stabilità dei mercati finanziari e una domanda interna ed esterna in crescita. Tuttavia, con l'espandersi della pandemia, a partire dal primo trimestre del 2020 si nota un improvviso picco negativo del PIL, che interrompe in modo drastico il trend crescente fin lì osservato. Questo calo importante è da ricollegare alle misure restrittive imposte dal Governo nel tentativo di contenere la diffusione del virus, quali ad esempio: i lockdown, la chiusura delle attività commerciali e industriali non essenziali, e le limitazioni di spostamento delle persone. Queste misure hanno portato ad una forte riduzione nella produzione di beni e servizi, ad un calo dei consumi e degli investimenti, e ad un aumento della disoccupazione.
La serie storica dell'IPI evidenzia invece una componente ciclica regolare, con marcate variazioni stagionali e diversi trend che riflettevano le dinamiche specifiche del settore industriale. Prima del 2020, l'IPI rispettava pattern prevedibili influenzati da fattori quali: la domanda di mercato, le innovazioni tecnologiche e le politiche commerciali. Nonostante ciò, con l'inizio della pandemia, si è registrato un brusco stop di questa ciclicità, dovuto all'interruzione delle catene di approvvigionamento globali, alla carenza di materie prime, alla diminuzione della forza lavoro disponibile a causa delle restrizioni sanitarie, e al calo della domanda sia a livello nazionale che internazionale.

Capitolo 2

Introduzione al pacchetto forecast e ai modelli ARIMA e ETS

2.1 Introduzione a forecast

Il pacchetto forecast di R, sviluppato da Rob J. Hyndman e collaboratori (Hyndman et al. (2024)), rappresenta uno strumento fondamentale per l'analisi e la previsione di serie storiche, ampiamente utilizzato in ambito economico, finanziario, metereologico e sociale. Esso fornisce metodi e strumenti per la visualizzazione, l'analisi e la previsione di serie storiche univariate tra cui modelli ARIMA (Autoregressive Integrated Moving Average), di lisciamento esponenziale (ETS), modelli di stato spazio e reti neurali. Una delle caratteristiche che contraddistingue il pacchetto forecast è la sua capacità di automatizzare la selezione del modello più appropriato e l'ottimizzazione dei parametri attraverso criteri di informazione come l'AIC (Akaike Information Criterion) e il BIC (Bayesian Information Criterion), che verranno discussi successivamente. Questo aspetto, e il fatto che il pacchetto sia completamente open source, rende possibile l'applicazione di tecniche e modelli statistici complessi anche ad utenti inesperti. Inoltre, esso fornisce strumenti per la visualizzazione avanzata dei dati tramite vati modelli di grafici, la diagnostica dei modelli e la gestione di componenti stagionali e cicliche nelle serie storiche. La flessibilità del pacchetto gli permette di potersi adattare alle diverse caratteristiche dei dati, come trend non lineari, variazioni stagionali e valori anomali, infatti l'efficacia e la versatilità di forecast sono state ampiamente documentate nella letteratura scientifica. Hyndman e Khandakar (2008) hanno illustrato dettagliatamente le capacità del pacchetto nell'automatizzare il processo di previsione, evidenziando come esso possa migliorare l'accuratezza delle previsioni e semplificare l'analisi dei dati temporali. Successivamente lo stesso Hyndman e Athanasopoulos (2018) hanno esteso queste metodologie nel loro libro ”Forecasting: Principles and Practice”, che è diventato un riferimento fondamentale nel campo della previsione di serie storiche.

L'adozione diffusa del pacchetto forecast nella comunità R testimonia la sua importanza come strumento per ricercatori e professionisti che necessitano di implementare modelli di previsione affidabili e accurati. La continua evoluzione del pacchetto, con aggiornamenti che incorporano le ultime scoperte nel campo dell'analisi delle serie storiche, offrendo metodologie di previsione all'avanguardia.

2.2 Descrizione dei principali modelli

Le tecniche per l'analisi e la previsione dei possibili andamenti di serie storiche possono essere suddivise in due categorie differenti: tecniche qualitative e tecniche quantitative. Le tecniche qualitative sono di natura soggettiva e si basano sul giudizio di esperti nel contesto specifico di applicazione della tecnica. Queste tecniche vengono utilizzate quando i dati storici risultano assenti, limitati o incompleti. Inoltre, possono essere impiegate in combinazione con le tecniche quantitative, sia per confrontare i risultati ottenuti dalle due metodologie, sia per affinare l'output di una tecnica quantitativa attraverso un approccio qualitativo. Un esempio tipico è il lancio di un nuovo prodotto, dove esperienza ed intuizione degli esperti sono fondamentali. Le tecniche quantitative, invece, si basano su dati storici e utilizzano modelli matematici e statistici per analizzare tali dati, identificare le caratteristiche chiave e proiettarle nel futuro al fine di stimare l'andamento della variabile di interesse. Tra i metodi fondamentali di questa categoria si annoverano il metodo delle medie mobili (*moving average method*), il metodo naïve e il metodo del random walk, con o senza "deriva" (*drift*). Oltre a questi, esistono modelli più complessi come il modello autoregressivo integrato a media mobile (ARIMA model) e il modello di lisciamento esponenziale (exponential smoothing model, ETS) che saranno approfonditi nelle sezioni successive.

2.2.1 Modelli ARIMA (Autoregressive Integrated Moving Average)

I modelli **ARIMA** (Autoregressive Integrated Moving Average) racchiudono in un solo insieme le caratteristiche dei modelli autoregressivi (AR) e dei modelli a media mobile (MA), aggiungendo una componente di integrazione per gestire l'eventuale non stazionarietà dei dati. Questi modelli sono particolarmente efficaci e semplici da implementare per prevedere serie storiche non stazionarie, registrando sia le dipendenze temporali che la casualità residua nei dati.

Modelli a Media Mobile (MA)

Un modello **MA(**q**)** di ordine q esprime la serie storica come una combinazione lineare degli errori casuali passati. L'equazione generale è:

$$y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q} \quad ,$$

dove ϵ_t è un errore casuale con media zero e varianza costante, μ è la media della serie e θ_i sono i parametri del modello. L'obiettivo è modellare y_t come funzione degli errori casuali dei periodi precedenti.

Modelli Autoregressivi (AR)

Un modello **AR(p)** di ordine p rappresenta y_t come una funzione lineare dei suoi valori passati. L'equazione generale è:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \epsilon_t \quad ,$$

dove ϕ_i sono i parametri autoregressivi e ϵ_t è l'errore residuo. In questo modello, y_t dipende direttamente dai suoi valori passati, al netto della media.

Stazionarietà e Differenziazione

Per poter applicare nel miglior modo i modelli AR e MA, la serie storica deve essere stazionaria, ovvero avere media e varianza costanti nel tempo ed autocorrelazione che dipende sollo dalla ”distanza” nel tempo della serie osservata. Se la serie non è stazionaria, è possibile convertirla attraverso la **differenziazione**, ossia:

$$y'_t = y_t - y_{t-1}$$

Se una differenziazione non è sufficiente, si possono applicare differenze di ordine superiore:

$$y''_t = y'_t - y'_{t-1} = y_t - 2y_{t-1} + y_{t-2}$$

La differenziazione è un passo cruciale per trasformare una serie non stazionaria in stazionaria, permettendo l'applicazione di modelli ARIMA.

Modelli ARMA e ARIMA

Un modello **ARMA(p,q)** combina componenti autoregressive e a media mobile:

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q} \quad ,$$

dove p è l'ordine autoregressivo e q è l'ordine della media mobile. I modelli ARMA sono adatti per serie storiche stazionarie.
Un modello **ARIMA(p,d,q)** estende l'ARMA includendo un componente di differenziazione di ordine d per gestire la non stazionarietà:

$$\Delta^d y_t = y_t^{(d)} = y_t - y_{t-d}$$

dove Δ^d rappresenta l'operatore di differenziazione di ordine d.

Modelli ARIMA Stagionali (SARIMA)

Per serie storiche con stagionalità, i modelli ARIMA possono essere estesi a modelli **SARIMA (Seasonal ARIMA)**, indicati come ARIMA(p,d,q)(P,D,Q)$_s$, dove s è il periodo stagionale:

- P: ordine autoregressivo stagionale
- D: ordine di differenziazione stagionale
- Q: ordine della media mobile stagionale

L'inclusione di componenti stagionali permette al modello di catturare pattern che si ripetono regolarmente nel tempo.

Funzioni di Autocorrelazione (ACF) e Autocorrelazione Parziale (PACF)

Le funzioni **ACF** e **PACF** sono strumenti di diagnostica essenziale per poter identificare gli ordini p e q nei modelli ARIMA. L'ACF misura la correlazione tra una variabile e i suoi ritardi, mentre la PACF isola la correlazione diretta con un ritardo specifico, escludendo le influenze intermedie. Tramite l'analisi dei grafici di ACF e PACF, possiamo determinare gli ordini appropriati del modello.

Previsioni

Una volta selezionato e stimato il modello ARIMA, è possibile effettuare previsioni proiettando i valori osservati nel tempo. Le previsioni puntuali sono ottenute utilizzando i parametri stimati, mentre gli intervalli di confidenza sono calcolati per quantificare l'incertezza associata alle previsioni:

$$\hat{y}_{T+h|T} \pm z_{\alpha/2} \cdot \hat{\sigma}_h \quad ,$$

dove $\hat{y}_{T+h|T}$ indica la previsione puntuale della variabile y per il periodo $T + h$, data l'informazione disponibile fino al periodo T, e $z_{\alpha/2}$ è il quantile della distribuzione normale standard e $\hat{\sigma}_h$ è la deviazione standard dell'errore di previsione per l'orizzonte h.

Conclusione

I modelli ARIMA rappresentano uno strumento potente e flessibile per l'analisi e la previsione di serie storiche non stazionarie. La combinazione di differenziazione, componenti autoregressive e a media mobile permette di modellare efficacemente le dinamiche temporali complesse presenti nei dati. Attraverso l'utilizzo di criteri di informazione per la selezione del modello e strumenti diagnostici come ACF e PACF, è possibile identificare il modello ARIMA più adatto per una vasta gamma di applicazioni.

2.2.2 Modelli di Lisciamento Esponenziale (ETS)

I modelli di **lisciamento esponenziale** rappresentano una famiglia di metodi utilizzati per la previsione di serie storiche, basati sull'idea di assegnare pesi decrescenti esponenzialmente alle osservazioni passate. Ciò significa che le osservazioni più recenti hanno un'influenza maggiore sulle previsioni rispetto a quelle più lontane nel tempo. Questi metodi sono particolarmente efficaci per serie storiche che presentano trend e/o stagionalità.

Componente di Trend

Il metodo più semplice all'interno dei modelli di lisciamento esponenziale è il **Simple Exponential Smoothing** (SES), adatto per serie storiche senza trend né stagionalità. Tuttavia, quando i dati mostrano un chiaro andamento crescente o decrescente, è più appropriato utilizzare il **metodo di Holt**, che estende il SES includendo una componente di trend lineare.
Le equazioni fondamentali del metodo di Holt sono:

1. **L'equazione di previsione**:
$$\hat{y}_{t+h|t} = \ell_t + hb_t$$

2. **L'aggiornamento del livello**:
$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$$

3. **L'aggiornamento del trend**:
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

dove:

- ℓ_t è la stima del livello al tempo t,
- b_t è la stima del trend al tempo t,
- α è il parametro di smoothing per il livello ($0 < \alpha < 1$),
- β^* è il parametro di smoothing per il trend ($0 < \beta^* < 1$).

In questo modello, i pesi assegnati alle osservazioni passate diminuiscono esponenzialmente, dando maggiore importanza ai dati più recenti.

Trend Smorzato

Per evitare che le previsioni crescano o decrescano indefinitamente, il **metodo di Holt con trend smorzato** introduce un parametro di smorzamento ϕ ($0 < \phi < 1$), che riduce gradualmente l'influenza del trend nelle previsioni a lungo termine:

1. **Equazione di previsione con smorzamento**:

$$\hat{y}_{t+h|t} = \ell_t + \left(\frac{1-\phi^h}{1-\phi}\right) b_t$$

2. **Aggiornamento del trend smorzato**:

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\phi b_{t-1}$$

Questo approccio è utile quando si prevede che il trend attuale non continui indefinitamente nel futuro.

Componente Stagionale

Per serie storiche che presentano una stagionalità evidente, il **metodo di Holt-Winters** estende ulteriormente il modello includendo una componente stagionale. La stagionalità può essere modellata in modo **additivo** o **moltiplicativo**, a seconda che l'ampiezza delle fluttuazioni stagionali sia costante o proporzionale al livello della serie storica.

Modello Stagionale Additivo

1. **Equazione di previsione**:
$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+m^*}$$

2. **Aggiornamento del livello**:
$$\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

3. **Aggiornamento del trend**:
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

4. **Aggiornamento della stagionalità**:
$$s_t = \gamma(y_t - \ell_t - b_t) + (1-\gamma)s_{t-m}$$

dove:

- s_t è la stima della componente stagionale al tempo t,
- γ è il parametro di smoothing per la stagionalità ($0 < \gamma < 1$),
- m è la periodicità stagionale (ad esempio, $m = 12$ per dati mensili),
- $m^* = (h-1) \bmod m + 1$.

Modello Stagionale Moltiplicativo

Quando l'ampiezza della stagionalità varia in proporzione al livello della serie, si utilizza il modello moltiplicativo:

1. **Equazione di previsione**:
$$\hat{y}_{t+h|t} = (\ell_t + hb_t) \times s_{t+m^*}$$

2. **Aggiornamento del livello**:
$$\ell_t = \alpha\left(\frac{y_t}{s_{t-m}}\right) + (1-\alpha)(\ell_{t-1} + b_{t-1})$$

3. **Aggiornamento del trend**:

$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$$

4. **Aggiornamento della stagionalità**:

$$s_t = \gamma \left(\frac{y_t}{\ell_t + b_t} \right) + (1 - \gamma)s_{t-m}$$

Intervalli di Previsione

I modelli di lisciamento esponenziale consentono di calcolare intervalli di previsione che riflettono l'incertezza associata alle stime future. Assumendo errori normalmente distribuiti, gli intervalli di previsione al livello di confidenza $(1 - \alpha)$ sono pari a:

$$\hat{y}_{t+h|t} \pm z_{\alpha/2} \cdot \sigma_h \quad ,$$

dove $\hat{y}_{T+h|T}$ indica la previsione puntuale della variabile y per il periodo $T + h$ e:

- $z_{\alpha/2}$ è il quantile della distribuzione normale standard,
- σ_h è la deviazione standard dell'errore di previsione all'orizzonte h.

La deviazione standard σ_h varia a seconda del modello ETS specifico utilizzato e può essere calcolata ricorsivamente.

Previsioni

Le previsioni si generano estendendo le equazioni del modello ai periodi successivi. Nel caso si utilizzi il metodo di Holt, la previsione per h periodi nel futuro è data da:

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$

Per quanto riguarda i modelli che includono stagionalità, la previsione include anche la componente stagionale corrispondente:

- **Modello Additivo**:

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m^*}$$

- **Modello Moltiplicativo**:

$$\hat{y}_{t+h|t} = (\ell_t + hb_t) \times s_{t+h-m^*}$$

Gli intervalli di previsione aumentano con l'orizzonte h a causa dell'accumulo di incertezza associata alle stime dei parametri e agli errori residui.

Conclusione

I modelli di lisciamento esponenziale (ETS) forniscono un approccio flessibile e potente per la previsione di serie storiche con caratteristiche differenti, come trend e stagionalità. Grazie alla loro struttura adattativa e alla capacità di incorporare componenti additive o moltiplicative, questi modelli sono ampiamente utilizzati in vari settori per fornire previsioni accurate ed affidabili.

2.3 Scelta del modello ottimale

Per selezionare il modello più appropriato, è fondamentale introdurre il concetto di massima verosimiglianza, poiché esso compare in alcuni criteri di selezione dei modelli. Questo approccio rappresenta un'alternativa alla scelta dei parametri basata sulla minimizzazione della somma dei quadrati degli errori. Invece di minimizzare gli errori, la massima verosimiglianza mira a trovare i parametri che massimizzano la probabilità che il modello abbia generato i dati osservati, fornendo così una rappresentazione più accurata dei dati stimati (Hyndman e Athanasopoulos, 2018). Per i modelli lineari con errori normalmente distribuiti, questa metodologia porta agli stessi risultati che si possono ottenere con la minimizzazione della somma dei quadrati degli errori. Nonostante ciò, nei modelli più complessi o non lineari, come alcuni modelli ARIMA, si possono ottenere risultati diversi, rendendo la massima verosimiglianza un criterio preferibile in taluni casi. I parametri dei modelli ARIMA sono soggetti a restrizioni per poter garantire la stazionarietà e l'invertibilità del modello, due aspetti fondamentali. Ad esempio, nel modello ARIMA(p, d, q), i parametri autoregressivi devono soddisfare le condizioni di stazionarietà, il che implica che le radici del polinomio autoregressivo devono trovarsi al di fuori del cerchio unitario nel piano complesso. Analogamente, i parametri della media mobile devono soddisfare le condizioni di invertibilità (Hyndman e Athanasopoulos, 2018). Un ulteriore approfondimento sulla scelta dei parametri può essere ottenuto considerando le proprietà matematiche dei modelli ARIMA. L'obiettivo è modellare correttamente le correlazioni temporali nei dati e assicurare che il modello sia adeguato ai dati storici. Ad esempio, nel modello AR(1), il parametro autoregressivo ϕ deve soddisfare la condizione $|\phi| < 1$ per garantire la stazionarietà. In alcuni casi,

potrebbe essere necessario estendere o modificare queste restrizioni per adattarsi meglio ai dati, il che potrebbe comportare l'uso di differenziazioni aggiuntive o l'inclusione di componenti stagionali nel modello ARIMA (Hyndman e Athanasopoulos, 2018).
La selezione del modello avviene secondo tre criteri principali:

- **Akaike's Information Criterion (AIC)**, calcolato come:

$$AIC = -2\log(L) + 2k \tag{2.1}$$

  dove L è la funzione di verosimiglianza del modello, mentre k è il numero totale di parametri stimati nel modello, inclusa la varianza dei residui.

- **Bayesian Information Criterion (BIC)**, definito dalla formula:

$$BIC = AIC + k[\log(T) - 2] \tag{2.2}$$

Per ciascuno di questi criteri, il modello migliore sarà quello che fornirà il valore minimo. È fondamentale, inoltre, specificare che alcuni modelli ARIMA che includono componenti stagionali moltiplicative possono causare instabilità nei calcoli quando nella serie sono presenti dati non strettamente positivi o quando i dati mostrano variazioni stagionali irregolari. In particolare, modelli come l'ARIMA con stagionalità moltiplicativa possono essere sensibili a queste condizioni, rendendo necessario un trattamento speciale dei dati o la considerazione di modelli alternativi (Hyndman e Athanasopoulos, 2018). Diventa quindi necessario un confronto con i modelli ETS (Error, Trend, Seasonality) per poter sottolineare alcune differenze. Nei modelli ETS, i parametri α (livello), β (trend), γ (stagionalità) e ϕ (smorzamento) sono soggetti a restrizioni specifiche per garantire la stabilità del modello e l'ammissibilità delle componenti di stato. Ad esempio, nei modelli ETS con errori moltiplicativi, le restrizioni sui parametri possono essere meno stringenti, ma possono comunque sorgere problemi quando i dati contengono valori zero o negativi, causando instabilità nei calcoli (Hyndman e Athanasopoulos, 2018). Al contrario, nei modelli ARIMA, le restrizioni sui parametri sono legate principalmente alle condizioni di stazionarietà e invertibilità. L'approccio basato sulla differenziazione rende i modelli ARIMA più adatti ed efficienti per serie storiche non stazionarie, ossia senza una struttura stagionale ben definita. Inoltre, mentre i modelli ETS sono particolarmente efficaci nel catturare componenti stagionali e trend strutturati nei dati, i modelli ARIMA offrono maggiore flessibilità nell'adattarsi a diverse dinamiche temporali attraverso la combinazione di componenti autoregressive e a media mobile.

2.4 Indici di Accuratezza

Nell'ambito dell'analisi e della previsione delle serie storiche tramite i modelli ARIMA ed ETS, gli indici di accuratezza sono fondamentali misure statistiche impiegate per valutare le prestazioni di previsione di tali modelli. Questi indici quantificano la differenza tra i valori previsti dal modello e i valori effettivamente osservati nella serie storica, fornendo una misura dell'errore di previsione. Gli indici di accuratezza sono utili per:

- **Valutare la qualità delle previsioni**: misurare l'entità degli errori di previsione.
- **Confrontare modelli diversi**: consentire di identificare quale modello è in grado di offrire le previsioni più accurate.
- **Ottimizzare i modelli**: fornire feedback per migliorare i parametri del modello.

Tra gli indici di accuratezza più comuni troviamo:

Errore Medio (Mean Error, ME)

$$\mathrm{ME} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)$$

Misura la media degli errori di previsione; un valore vicino a zero indica che le previsioni non sono distorte.

Errore Assoluto Medio (Mean Absolute Error, MAE)

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

Calcola la media degli errori assoluti tra i valori osservati y_t e quelli previsti $\hat{y}_t$.

Errore Quadratico Medio (Mean Squared Error, MSE)

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2$$

Misura la media dei quadrati degli errori, penalizzando maggiormente gli errori più grandi.

Radice dell'Errore Quadratico Medio (Root Mean Squared Error, RMSE)

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

Fornisce l'errore in unità comparabili con quelle dei dati originali.

Errore Percentuale Assoluto Medio (Mean Absolute Percentage Error, MAPE)

$$\text{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|$$

Esprime l'errore medio in termini percentuali rispetto ai valori osservati.

Errore Percentuale Assoluto Medio Simmetrico (Symmetric Mean Absolute Percentage Error, sMAPE)

$$\text{sMAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\frac{|y_t - \hat{y}_t|}{(|y_t| + |\hat{y}_t|)/2}$$

Evita problemi associati al MAPE quando i valori osservati sono prossimi a zero.

Errore Assoluto Medio Scalato (Mean Absolute Scaled Error, MASE)

$$\text{MASE} = \frac{\text{MAE}}{\frac{1}{n-1}\sum_{t=2}^{n}|y_t - y_{t-1}|}$$

Fornisce un errore medio assoluto scalato rispetto all'errore medio assoluto di un modello naïve.

Interpretazione degli Indici di Accuratezza

Un valore minore degli indici ME, MAE, MSE, RMSE, MAPE, sMAPE e MASE indica un'accuratezza migliore del modello. È importante utilizzare questi indici insieme in modo da avere una valutazione completa ed ottimale delle prestazioni del modello, poiché ognuno di essi può fornire diverse informazioni.

Applicazione ai Modelli ARIMA ed ETS

Gli indici di accuratezza possono essere applicati sia ai modelli ARIMA che ai modelli ETS per valutare le prestazioni previsionali. Dopo aver sviluppato e stimato i modelli, si confrontano le previsioni con i valori effettivamente osservati utilizzando questi indici. Secondo Hyndman,inoltre, l'uso di indici come il MASE è particolarmente consigliato poiché consente di confrontare l'accuratezza tra serie storiche diverse e non è influenzato dalla scala dei dati.

Capitolo 3

Applicazioni Empiriche

3.1 Rolling Forecasting: Previsioni a Finestra Mobile

Il **rolling forecasting**, in italiano ”previsione a finestra mobile”, è una tecnica fondamentale nell’analisi delle serie storiche, utilizzata per valutare, validare e migliorare le prestazioni dei modelli previsivi. Questa metodologia riproduce il processo di previsione in un contesto reale, in cui il modello viene costantemente aggiornato man mano che nuovi dati diventano disponibili. In altre parole, il rolling forecasting permette di simulare il comportamento di un modello nel tempo, in modo che questo si adatti ai cambiamenti e che fornisca previsioni basate sulle informazioni più recenti. Alla base del rolling forecasting vi è l’idea di dividere la serie storica in due sezioni distinte:

1. **Set di training (o apprendimento)**: Questo insieme di dati storici è impiegato per stimare i parametri del modello. È in questo contesto che quest’ultimo ”impara” dai dati passati, identificando pattern, trend e relazioni che possono rivelarsi utili per le previsioni future.

2. **Set di test (o di validazione)**: Questa parte della serie storica, non utilizzata durante la fase di stima, è di fondamentale importanza per valutare l’accuratezza e l’affidabilità delle previsioni del modello. Confrontando le previsioni generate con i dati effettivamente osservati, è possibile quantificare l’errore di previsione e valutare le prestazioni del modello.

La caratteristica distintiva del rolling forecasting è che il modello viene ricalibrato ad ogni passo temporale. Ciò significa che, man mano che nuovi dati diventano disponibili, questi vengono aggiunti al set di training, il modello viene ristimato e una nuova previsione viene effettuata per il periodo successivo. Questo processo viene ripetuto iterativamente, scorrendo lungo la serie storica, da cui il termine ”rolling”.

Esistono due approcci principali per definire il set di training nel contesto del rolling forecasting, ciascuno con le proprie caratteristiche e implicazioni:

3.1.1 Finestra Espandibile (Expanding Window)

In questo approccio, il set di training si espande ad ogni iterazione, includendo tutti i dati disponibili fino a quel punto come mostrato in Figura 3.1. All'inizio, il modello viene stimato su un insieme iniziale di dati. Poi, ad ogni nuovo passo, il dato più recente viene aggiunto al set di training, aumentando così la dimensione complessiva del dataset utilizzato per la stima. Questo metodo riflette una situazione in cui si desidera sfruttare al massimo tutte le informazioni storiche disponibili, nella convinzione che i dati passati contengano informazioni utili per le previsioni future. L'inclusione continua di nuovi dati può aiutare il modello a catturare cambiamenti strutturali o evoluzioni nei pattern della serie storica.

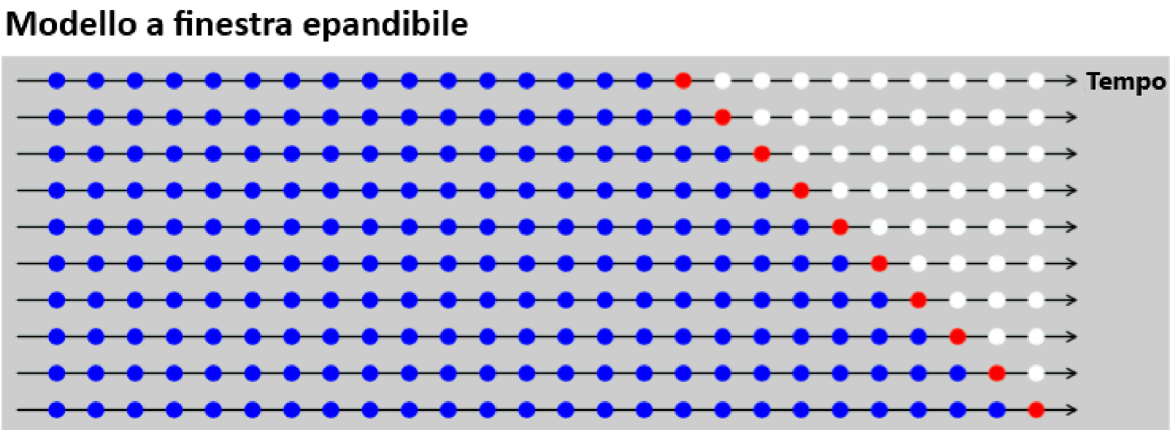

Figura 3.1: Rolling Forecasting: Finestra Espandibile

3.1.2 Finestra a Lunghezza Fissa (Fixed-Length Window)

In questo caso, il set di training mantiene una lunghezza costante durante tutto l'esperimento di rolling forecasting, e ad ogni iterazione la finestra avanza nel tempo. Quando, poi, è disponibile un nuovo dato, esso viene aggiunto al set di training, mentre il dato più vecchio in quel momento viene rimosso, mantenendo così costante la dimensione complessiva del set come mostrato in Figura 3.2. Questo approccio è utile quando si parte dal presupposto che i dati più recenti siano più rilevanti per le previsioni future, e che i dati passati possano non riflettere più le condizioni attuali che caratterizzano la serie storica. Inoltre, mantenendo una dimensione costante del set di training, si riduce il carico computazionale associato alla stima del modello, il che può essere vantaggioso quando si lavora con modelli complessi o dataset molto grandi.

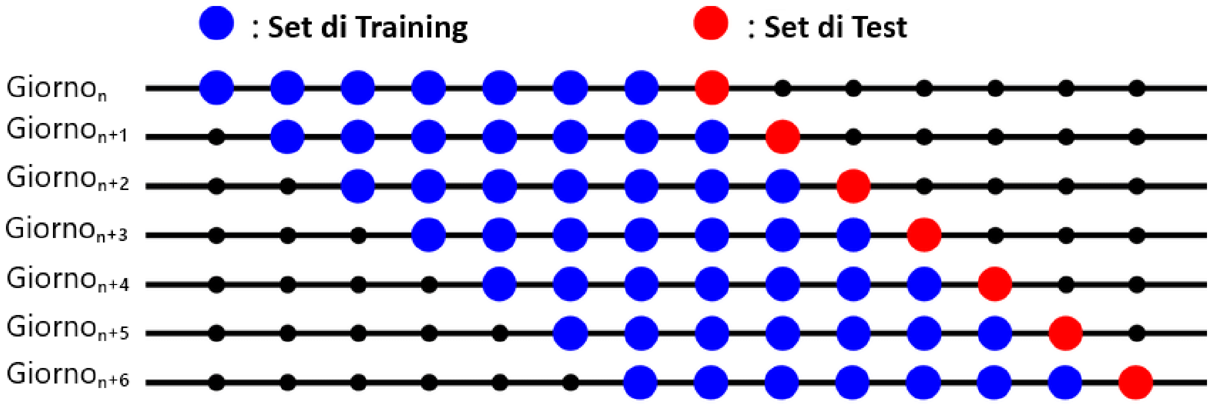

Figura 3.2: Rolling Forecasting: Finestra a Lunghezza Fissa

3.2 Applicazione dei modelli al PIL con ARIMA

3.2.1 Finestra Espandibile ARIMA

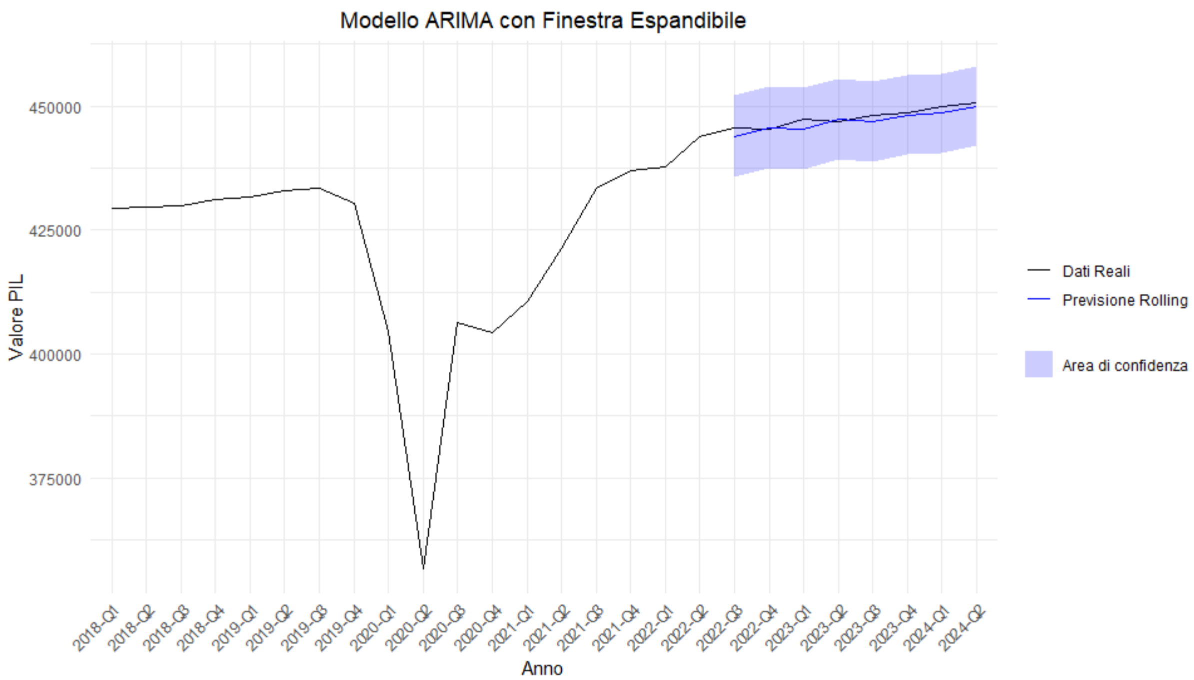

Figura 3.3: PIL: Previsione Finestra Espandibile ARIMA

Per analizzare l'andamento trimestrale del PIL italiano e prevederne le future evoluzioni come mostrato in Figura 3.3, è stato implementato un modello ARIMA utilizzando una finestra espandibile. I dati presi in considerazione coprono il periodo che va dal primo trimestre del 1996 (1996-Q1) al secondo trimestre del 2024 (2024-Q2), fornendo una serie storica estesa e dettagliata. Ad ogni iterazione del rolling forecasting, il modello viene addestrato utilizzando tutti i dati disponibili fino al trimestre immediatamente precedente alla previsione corrente. Questo approccio permette di sfruttare l'intero dataset, garantendo che il modello tenga conto di tutte le informazioni disponibili. L'utilizzo del comando forecast **auto.arima()** è giustificato dalla sua capacità e semplicità nel modellare efficacemente le serie storiche stazionarie e non stazionarie attraverso la differenziazione, integrando componenti autoregressive e di media mobile. Come evidenziato da Hyndman e Athanasopoulos (2018), i modelli ARIMA sono strumenti potenti per la previsione di serie storiche economiche. Per le previsioni, è stato adottato un intervallo di confidenza del 70%, che offre un equilibrio tra la precisione delle stime e la larghezza degli intervalli, consentendo di esprimere l'incertezza associata alle previsioni nel miglior modo possibile.

3.2.2 Finestra a Lunghezza Fissa ARIMA

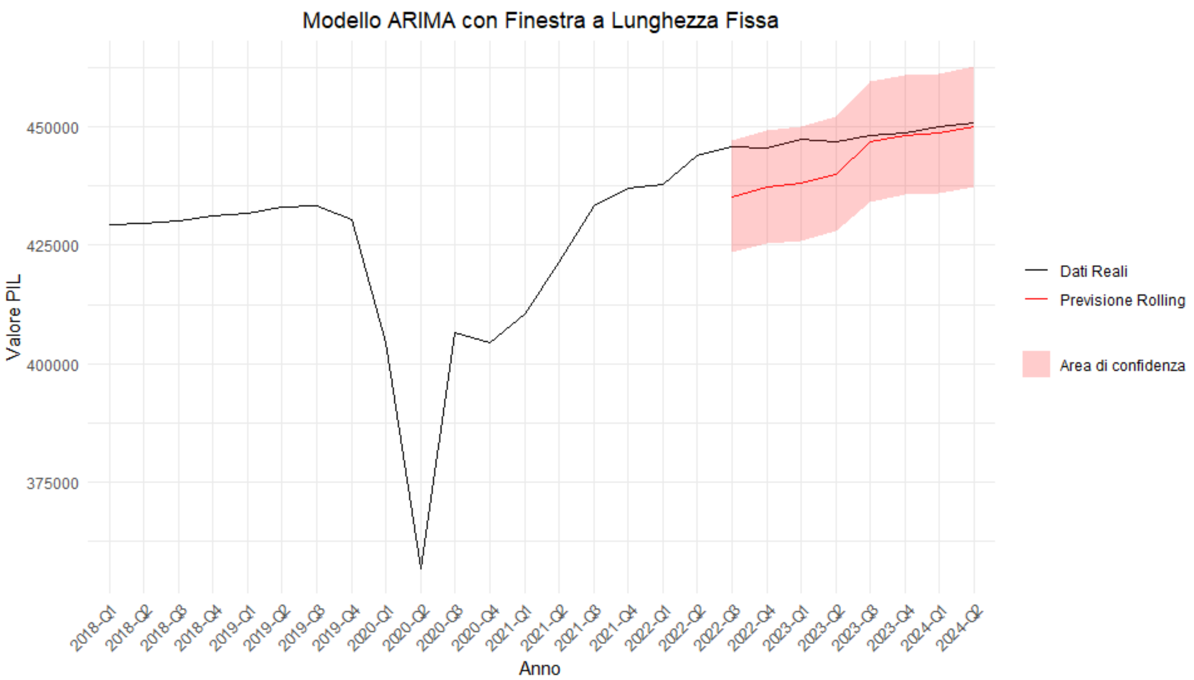

Figura 3.4: PIL: Previsione Finestra a Lunghezza Fissa ARIMA

In parallelo come mostrato nella Figura 3.4, è stato implementato un modello ARIMA utilizzando una finestra a lunghezza fissa. In questo caso, il modello viene addestrato su un sottoinsieme dei dati storici, costituito dagli ultimi 40 trimestri disponibili (equivalenti a 10 anni di dati) al momento di ciascuna previsione. Ad ogni iterazione, la finestra di dati si sposta in avanti di un trimestre, mantenendo costante la dimensione del dataset di training. Questo approccio permette al modello di concentrarsi sulle informazioni più recenti, catturando eventuali cambiamenti strutturali o trend emergenti nella serie storica. Come sottolineato da Hyndman e Athanasopoulos (2018), l'utilizzo di una finestra a lunghezza fissa può essere particolarmente efficace quando si sospetta che i dati più recenti siano più rappresentativi del comportamento futuro della serie rispetto ai dati storici più lontani. Anche in questo caso, le previsioni vengono rappresentate con un intervallo di confidenza del 70%, garantendo coerenza con l'approccio precedente e permettendo un confronto diretto tra i due metodi.

3.3 Applicazione dei modelli al PIL con ETS

3.3.1 Finestra Espandibile ETS

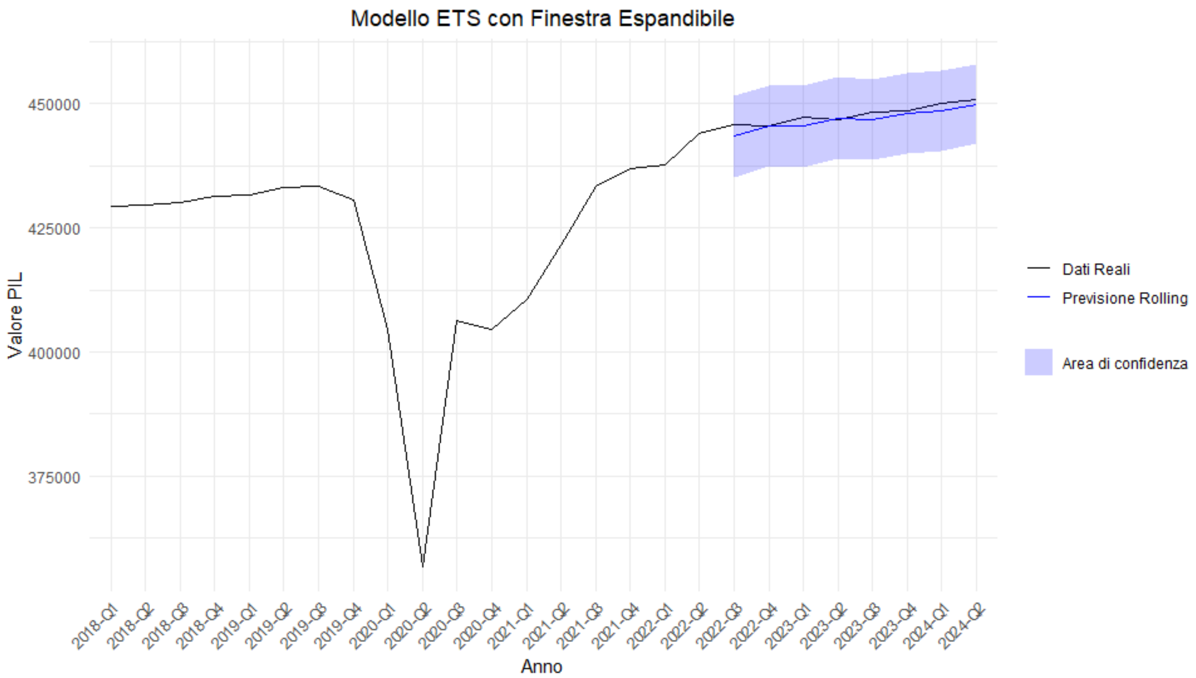

Figura 3.5: PIL: Previsione Finestra Espadibile ETS

Per offrire una prospettiva alternativa come mostrato in Figura 3.5, è stato applicato il modello ETS (Errori, Trend, Stagionalità) **ets()** con finestra espandibile ai dati trimestrali del PIL. Il modello ETS è particolarmente adatto per serie storiche che presentano componenti di trend e stagionalità, sfruttando metodi di lisciamento esponenziale per prevedere valori futuri. Come per l'ARIMA con finestra espandibile, il modello ETS viene addestrato utilizzando tutti i dati disponibili fino al periodo immediatamente precedente alla previsione. Secondo Hyndman et al. (2008), i modelli ETS garantiscono un framework flessibile per la previsione di serie storiche con complesse caratteristiche strutturali. L'implementazione di un intervallo di confidenza dell'70% nelle previsioni permette di comunicare in modo efficace l'incertezza associata alle stime, mantenendo allo stesso tempo intervalli più stretti e focalizzati rispetto ad intervalli con livelli di confidenza più elevati.

3.3.2 Finestra a Lunghezza Fissa ETS

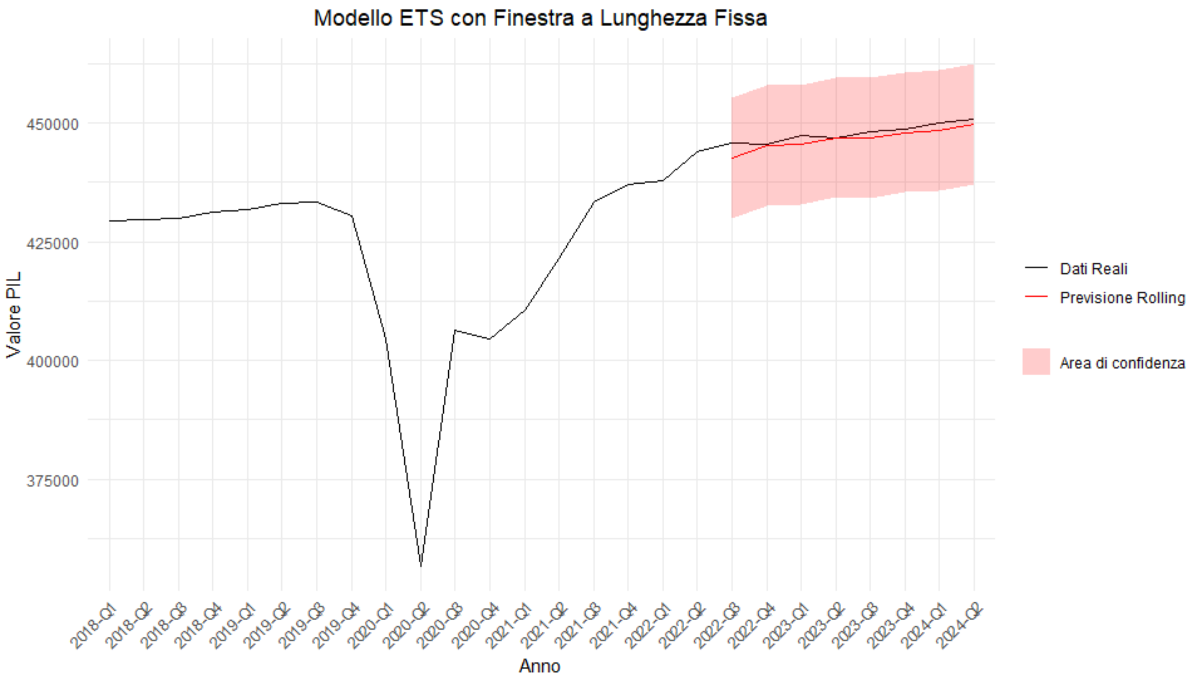

Figura 3.6: PIL: Previsione Finestra a Lunghezza Fissa ETS

Infine come illustrato in Figura 3.6, è stato implementato il modello ETS utilizzando una finestra a lunghezza fissa di 40 trimestri. Questo approccio consente di addestrare il modello su delle osservazioni rilevanti, che potrebbero riflettere meglio le dinamiche attuali del PIL, specialmente in presenza di cambiamenti strutturali o eventi economici significativi come per esempio il Covid. L'ETS con finestra a lunghezza fissa può adattarsi rapidamente a tali cambiamenti, come suggerito da Hyndman e Athanasopoulos (2018). Anche in questo caso, le previsioni sono state effettuate con un intervallo di confidenza del 70%, permettendo un confronto coerente con gli altri modelli e garantendo che le conclusioni tratte siano basate su presupposti simili riguardo all'incertezza delle previsioni.

3.4 Analisi dei modelli ARIMA ed ETS applicati alla serie storica del PIL

Nell'applicazione dei modelli ARIMA ed ETS alla serie storica del PIL, si osservano significative differenze tra l'approccio con finestra espandibile e quello con finestra a lunghezza fissa.

Utilizzando la finestra espandibile, il modello ARIMA mostra prestazioni leggermente superiori rispetto al modello ETS. In particolare, l'ARIMA presenta un errore medio (ME) di 850.875, un errore quadratico medio (RMSE) di 1213.765 e un errore percentuale assoluto medio (MAPE) di 0.2324, mentre l'ETS registra un ME di 973.3287, un RMSE di 1310.363 e un MAPE di 0.2431. Questi risultati indicano che, con l'aggiornamento continuo dei dati, il modello ARIMA riesce a catturare meglio le dinamiche del PIL, probabilmente grazie alla sua capacità di adattarsi a nuove informazioni e pattern emergenti.
Al contrario, con l'utilizzo di una finestra a lunghezza fissa, il modello ETS risulta qualitativamente superiore rispetto all'ARIMA in termini di accuratezza. L'ETS ottiene un ME di 1227.602, un RMSE di 1549.574 e un MAPE di 0.2809, mentre l'ARIMA mostra valori di errore molto più elevati, con un ME di 4819.629, un RMSE di 6246.27 e un MAPE di 1.0791. Questo suggerisce che, quando si limita l'analisi a un set di dati di dimensione costante, l'ARIMA perde efficacia nel prevedere il PIL, forse a causa di una minore capacità di adattamento ai cambiamenti recenti. L'ETS, grazie alla sua struttura basata sul lisciamento esponenziale, sembra invece mantenere una maggiore stabilità e accuratezza nelle previsioni in questo contesto.

3.5 Applicazione dei modelli all'IPI con ARIMA

3.5.1 Finestra Espandibile ARIMA

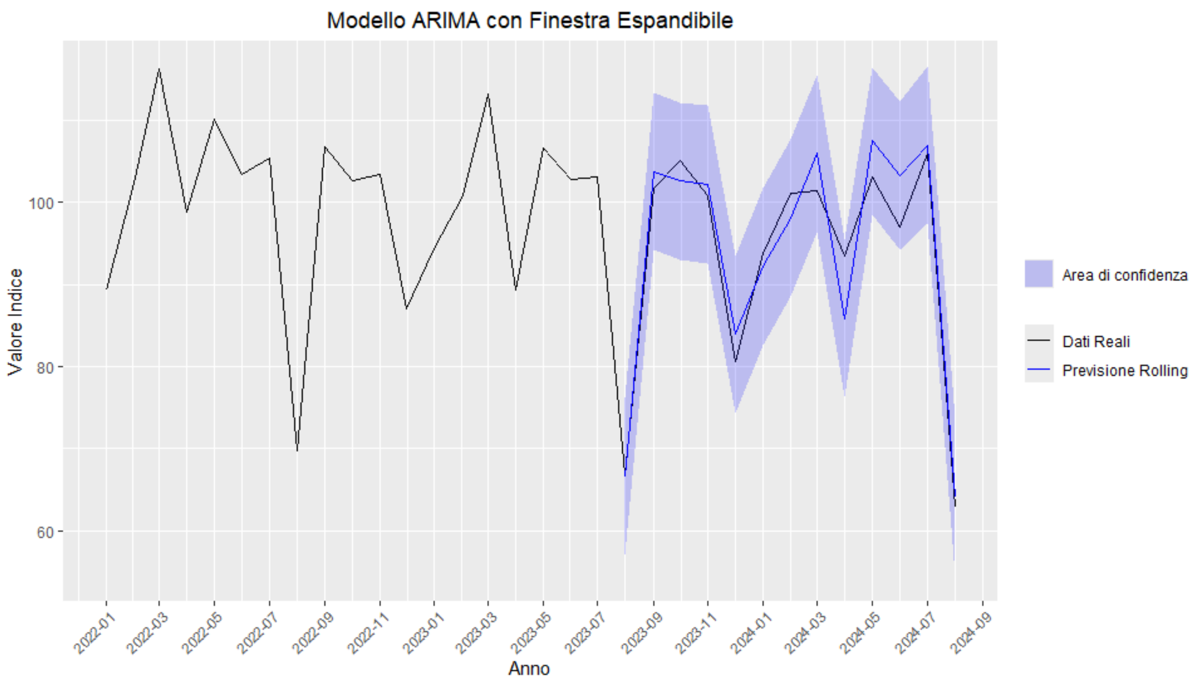

Figura 3.7: IPI: Previsione Finestra Espandibile ARIMA

L’implementazione del modello ARIMA con finestra espandibile come in Figura 3.7 è stata realizzata per catturare efficacemente le dinamiche temporali dell’indice di produzione industriale, seguendo le metodologie proposte da Hyndman e Athanasopoulos (2018). Questo approccio prevede che ad ogni passo di previsione la finestra di training si espanda per includere tutte le osservazioni disponibili fino a quel momento. Ciò consente al modello di sfruttare l’intera storia dei dati, migliorando la stima dei parametri e potenzialmente aumentando l’accuratezza delle previsioni. L’utilizzo del comando forecast **auto.arima()** facilita la selezione automatica del modello ottimale, tenendo conto delle componenti autoregressive, di integrazione e di media mobile, sia in termini non stagionali che stagionali. Questo processo automatizzato è basato sull’algoritmo proposto da Hyndman e Khandakar (2008), che combina l’identificazione del differenziamento necessario per rendere la serie stazionaria con la selezione dei termini AR e MA ottimali. La scelta di una finestra espandibile è motivata dalla necessità di catturare eventuali cambiamenti strutturali nella serie storica, beneficiando dell’intera gamma di dati storici disponibili. Questo è particolarmente rilevante in contesti economici, dove eventi passati possono influenzare le dinamiche future. Inoltre, l’approccio a finestra espandibile permette al modello di potersi adattare in modo progressivo, integrando nuove informazioni e affinando le previsioni man mano che nuovi dati diventano disponibili. La visualizzazione dei risultati è curata in modo che possa garantire una rappresentazione chiara e intuitiva delle previsioni rispetto ai dati osservati. L’inclusione delle bande di confidenza fornisce una misura visiva dell’incertezza associata alle previsioni, elemento cruciale per una corretta interpretazione e per supportare decisioni informate.

3.5.2 Finestra a Lunghezza Fissa ARIMA

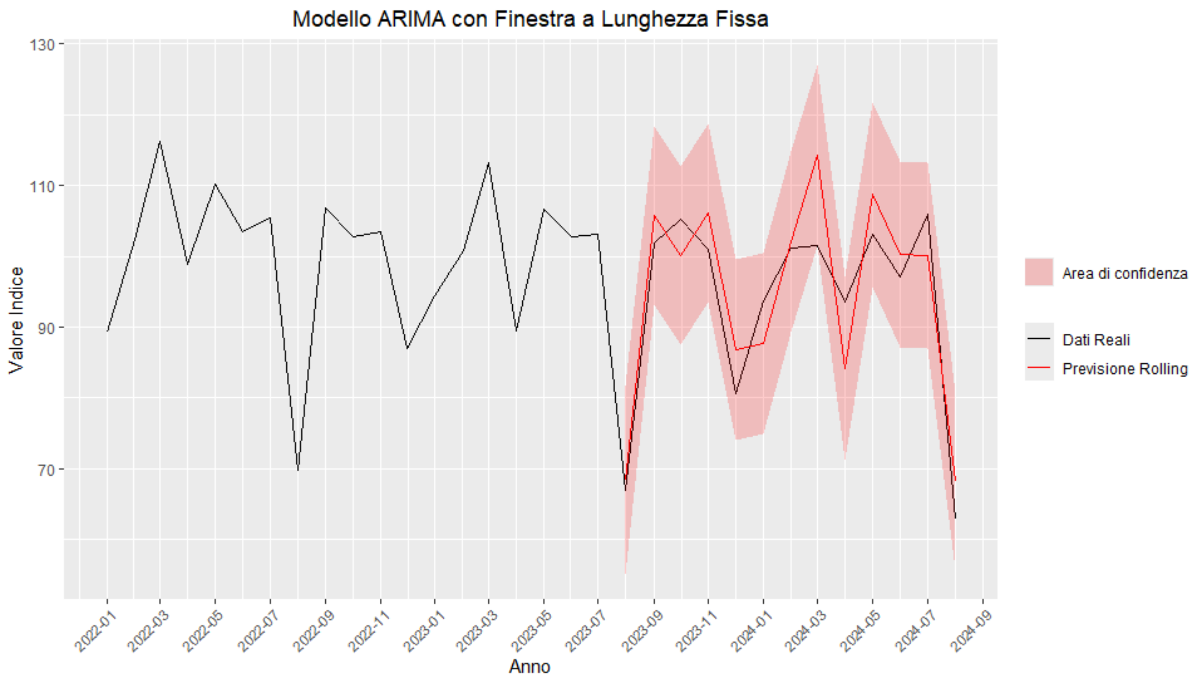

Figura 3.8: IPI: Previsione Finestra a Lunghezza Fissa ARIMA

L'approccio con finestra a lunghezza fissa nel modello ARIMA come mostrato in Figura 3.8 è stato adottato per enfatizzare le informazioni più recenti nella serie storica dell'indice di produzione industriale, in linea con le raccomandazioni di Hyndman e Athanasopoulos (2018). Limitando la finestra di training agli ultimi 10 anni, il modello è in grado di catturare meglio le dinamiche attuali, riducendo l'influenza di dati storici che potrebbero non riflettere le condizioni economiche contemporanee. Questo metodo è particolarmente utile in presenza di cambiamenti strutturali o evoluzioni nel comportamento della serie storica. Ad esempio, l'industria potrebbe aver subito trasformazioni significative negli ultimi anni a causa di innovazioni tecnologiche, cambiamenti normativi o eventi economici globali. Una finestra a lunghezza fissa permette al modello di focalizzarsi su questi cambiamenti recenti, migliorando potenzialmente l'accuratezza delle previsioni. La funzione **auto.arima()** viene utilizzata ad ogni iterazione per identificare il modello ottimale sulla finestra corrente. Ciò significa che il modello può cambiare nel tempo, adattandosi alle nuove caratteristiche dei dati. Questo approccio dinamico è essenziale per mantenere la rilevanza del modello predittivo in contesti in cui le condizioni possono variare rapidamente. La visualizzazione dei risultati fornisce un'importante strumento per valutare l'efficacia del modello e la plausibilità delle previsioni. Le bande di confidenza aiutano a comprendere l'incertezza associata alle stime, elemento cruciale per l'analisi del rischio e per supportare decisioni informate basate sulle previsioni generate.

3.6 Applicazione dei modelli all'IPI con ETS

3.6.1 Finestra Espandibile ETS

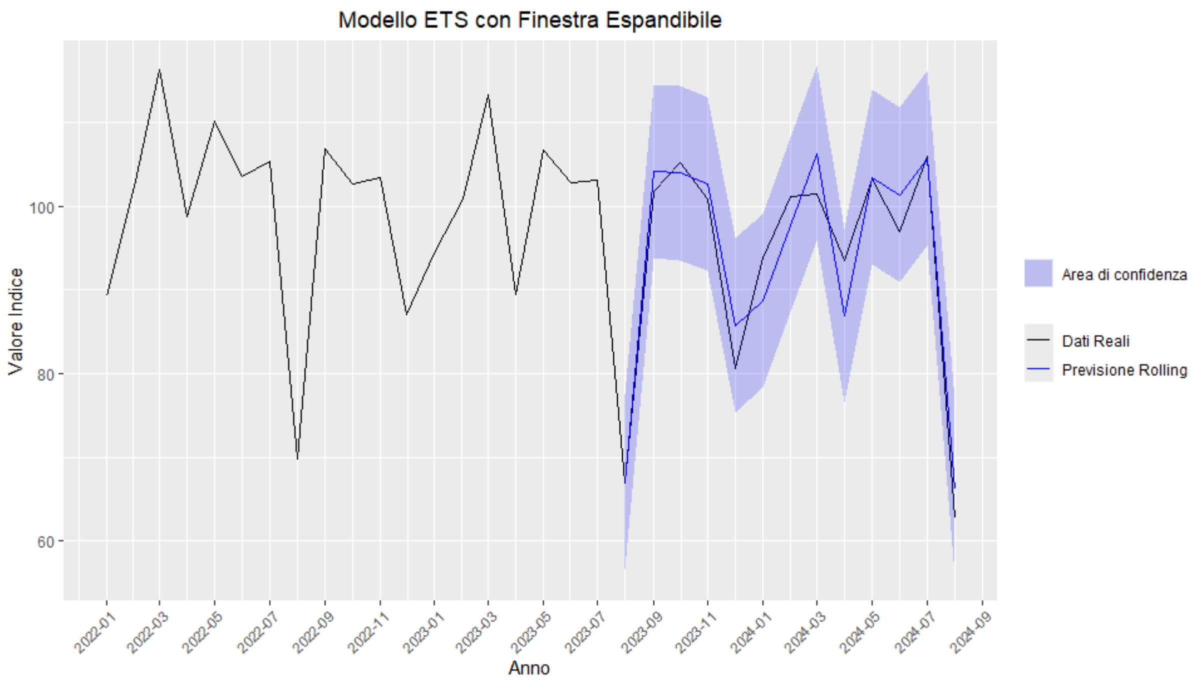

Figura 3.9: IPI: Previsione Finestra Espadibile ETS

L'applicazione del modello ETS con finestra espandibile come mostrato in Figura 3.9 consente di catturare le componenti di trend e stagionalità presenti nella serie storica dell'indice di produzione industriale, sfruttando l'intera storia dei dati disponibili. Come evidenziato da Hyndman e Athanasopoulos (2018), i modelli di smoothing esponenziale sono particolarmente adatti per serie storiche con pattern stagionali e di trend ben definiti. L'utilizzo della funzione **ets()** permette di selezionare automaticamente la forma più appropriata del modello tra le 30 possibili combinazioni di errori, trend e stagionalità (sia additivi che moltiplicativi). Questo processo automatizzato è basato su un framework di spazio di stato esposto da Hyndman et al. (2002), che garantisce una stima ottimale dei parametri e una gestione coerente dell'incertezza associata alle previsioni. La scelta di una finestra espandibile consente al modello di incorporare tutte le informazioni storiche, il che può essere vantaggioso quando le componenti di trend e stagionalità sono stabili nel tempo. Tuttavia, l'inclusione di dati più vecchi potrebbe diluire l'influenza di cambiamenti recenti nella serie. Pertanto, è importante monitorare la performance del modello nel tempo e considerare l'eventuale necessità di alternative come la finestra a lunghezza fissa. La visualizzazione delle previsioni in relazione ai dati reali garantisce una rappresentazione chiara ed efficace di come il modello ETS sia in grado di catturare le dinamiche della serie storica.

Le bande di confidenza forniscono inoltre un'indicazione visiva dell'incertezza associata alle previsioni, elemento fondamentale per l'interpretazione dei risultati e per supportare decisioni basate sui dati.

3.6.2 Finestra a Lunghezza Fissa ETS

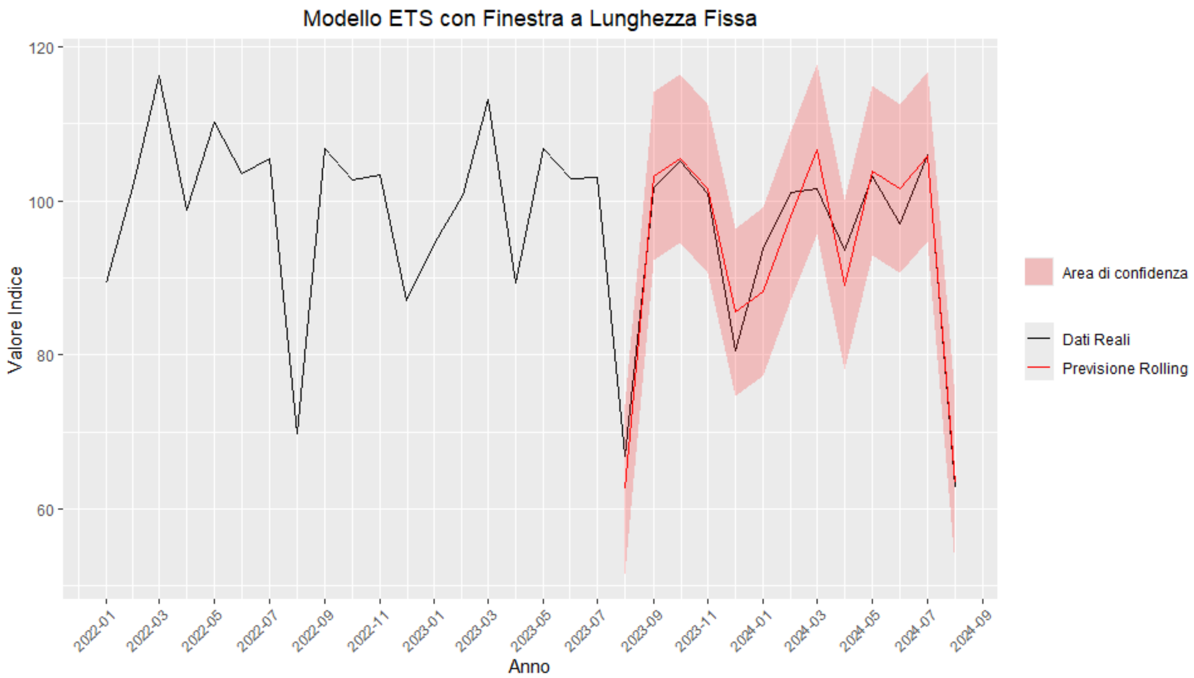

Figura 3.10: IPI: Previsione Finestra a Lunghezza Fissa ETS

L'approccio con finestra a lunghezza fissa nel modello ETS come illustrato in Figura 3.10 è stato adottato per focalizzare l'analisi sulle dinamiche più recenti dell'indice di produzione industriale, in linea con le indicazioni di Hyndman e Athanasopoulos (2018). Limitando la finestra di training agli ultimi 10 anni, il modello può adattarsi più rapidamente a eventuali cambiamenti nelle componenti di trend e stagionalità che potrebbero non essere riflessi nei dati storici più vecchi. Questo è particolarmente rilevante in contesti dove l'economia o il settore industriale subiscono rapidi cambiamenti, ad esempio a causa di innovazioni tecnologiche, mutamenti nelle politiche economiche o eventi imprevisti come crisi finanziarie. Una finestra a lunghezza fissa permette al modello ETS di enfatizzare queste nuove tendenze, potenzialmente migliorando l'accuratezza delle previsioni a breve termine. La funzione **ets()** viene applicata ad ogni iterazione sulla finestra corrente, in modo da consentire al modello di essere ricalibrato in maniera costante. Questo approccio dinamico risulta essenziale per conservare la rilevanza del modello in un ambiente in evoluzione continua. Inoltre, l'uso di criteri informativi per la selezione del modello garantisce il bilanciamento della complessità del modello in relazione alla bontà di adattamento ai dati. La visualizzazione dei risultati facilita l'interpretazione delle previsioni nel contesto delle osservazioni recenti. Le bande di confidenza aiutano a comprendere l'incertezza associata alle stime, fornendo informazioni essenziali per la gestione del rischio e per supportare decisioni strategiche basate sulle previsioni.

3.7 Analisi dei modelli ARIMA ed ETS applicati alla serie storica del IPI

Per quanto riguarda l'IPI, i modelli ARIMA ed ETS mostrano comportamenti molto differenti tra l'approccio con finestra espandibile e quello con finestra a lunghezza fissa. Con la finestra espandibile, entrambi i modelli presentano performance simili, ma l'ETS offre prestazioni leggermente superiori. L'ARIMA registra un ME di -0.6860, un RMSE di 3.6625 e un MAPE di 3.1664%, mentre l'ETS mostra un ME di -0.4232, un RMSE di 3.6296 e un MAPE di 3.2906%. Nonostante le differenze siano marginali, l'ETS potrebbe beneficiare della sua capacità di modellare efficacemente componenti stagionali e trend quando si dispone di dati in crescita.
Con l'adozione della finestra a lunghezza fissa, il modello ETS dimostra una performance nettamente migliore all'ARIMA. L'ETS raggiunge un ME di -0.0422, un RMSE di 3.4546 e un MAPE di 3.1100%, mentre l'ARIMA mostra un ME di -1.4134, un RMSE di 6.2132 e un MAPE di 5.9004%.

3.8 Scelta dei modelli

La selezione dei modelli di previsione si basa sul confronto degli RMSE. Secondo Hyndman nel suo libro (Forecasting: Principles and Practice 2nd Ed., 2018) , l'Errore Quadratico Medio (RMSE) è una delle metriche più comuni per valutare l'accuratezza dei modelli predittivi. L'RMSE calcola la media delle differenze quadrate tra i valori previsti e quelli effettivamente osservati, penalizzando maggiormente gli errori più grandi a causa dell'uso di una funzione di perdita quadratica. Questo lo rende particolarmente vantaggioso quando si desidera minimizzare l'impatto degli errori significativi nelle previsioni. Utilizzare l'RMSE come criterio di scelta permette di confrontare l'efficacia di diversi modelli e di selezionare quello che offre le previsioni più precise. Un valore di RMSE più basso indica che il modello ha una migliore capacità predittiva, riducendo l'errore medio nelle stime. Pertanto, basare la selezione del modello sull'RMSE contribuisce a migliorare l'affidabilità delle previsioni e a prendere decisioni più informate basate sui dati.

Modello	PIL RMSE	IPI RMSE
ARIMA EL	**1213.765**	3.662501
ARIMA FL	6246.270	6.213159
ETS EL	1310.363	3.629564
ETS FL	1549.574	**3.454643**

Tabella 3.1: Confronto degli RMSE per i modelli sul PIL e sull'IPI

Capitolo 4

Conclusioni

L'analisi e la previsione dei dati economici rappresentano oggi elementi fondamentali nel panorama economico-finanziario. L'utilizzo di tecniche computazionali avanzate permette di implementare metodologie previsionali sempre più sofisticate, superando i limiti delle tecniche tradizionali. In questo lavoro, sono stati applicati e confrontati i modelli ARIMA ed ETS per la previsione del Prodotto Interno Lordo (PIL) e dell'Indice di Produzione Industriale (IPI), utilizzando diverse configurazioni di finestre temporali. L'analisi comparativa ha evidenziato in modo chiaro come la complessità del modello non sia obbligatoriamente sinonimo di accuratezza previsionale maggiore. Per il PIL, il modello ARIMA con finestra espandibile ha presentato performance superiori, con errori significativamente più contenuti rispetto alle altre configurazioni. Questo fornisce un'analisi evidente sul fatto che per questa variabile macroeconomica, l'inclusione di informazioni storiche più estese contribuisce a catturare meglio i trend e i cicli economici di lungo periodo. Al contrario, per l'IPI, il modello ETS con finestra a lunghezza fissa si è rivelato più efficace, evidenziando come per alcuni indicatori economici una memoria storica più contenuta possa portare a previsioni più accurate. Questa divergenza nelle performance sottolinea l'importanza di una selezione accurata del modello previsionale in base alle caratteristiche specifiche della serie temporale analizzata. I risultati ottenuti evidenziano la complessità intrinseca nella previsione di variabili economiche, influenzate da molteplici fattori socio-economici e caratterizzate da elevata incertezza. Nonostante queste sfide, le metodologie esplorate forniscono strumenti robusti per l'analisi previsionale in ambito economico-finanziario. In considerazione di questi risultati, verranno presentate due previsioni per gli anni futuri: una per il PIL, utilizzando il modello ARIMA con finestra espandibile, e una per l'IPI, basata sul modello ETS con finestra fissa, dimostrando l'applicazione pratica delle metodologie identificate come ottimali per ciascuna serie storica.

I risultati ottenuti dalle previsioni sull'andamento futuro delle serie storiche analizzate forniscono spunti interessanti. Per quanto riguarda il PIL, l'ultimo dato osservato al momento della

stesura di questa tesi è quello del secondo trimestre del 2024, con un valore di 450.759,4 milioni di euro. Le previsioni per i trimestri successivi indicano una crescita zero nel breve periodo. Tuttavia, considerando gli estremi dell'intervallo di confidenza al 70% , sono ugualmente probabili sia tassi di crescita positivi che negativi, indicando quindi un'incertezza sull'andamento della variabile macroeconomica.
Per quanto riguarda l'IPI invece, l'ultimo dato osservato al momento della stesura è relativo ad agosto 2024, con un valore di 62,8. Avendo analizzato i dati storici precedenti, l'IPI mostra una chiara componente ciclica. Questo pattern è evidente durante l'andamento di tutta la serie storica che, ad esempio, nell'ultimo anno, registra massimi a marzo 2023 (113,2) e luglio 2023 (103,1), e minimi ad agosto 2023 (66,9) e dicembre 2023 (80,6). Le previsioni ottenute con il modello ETS a finestra fissa riflettono questo andamento. Dopo il minimo di agosto 2024, si prevede un aumento dell'IPI a partire da settembre 2024 (con valore previsto di 100,13), seguito da ulteriori fluttuazioni che ripercorrono il comportamento della serie storica. Il modello prevede picchi nei mesi di marzo 2025 (102,71) e luglio 2025 (105,21), e un nuovo minimo ad agosto 2025 (63,20).

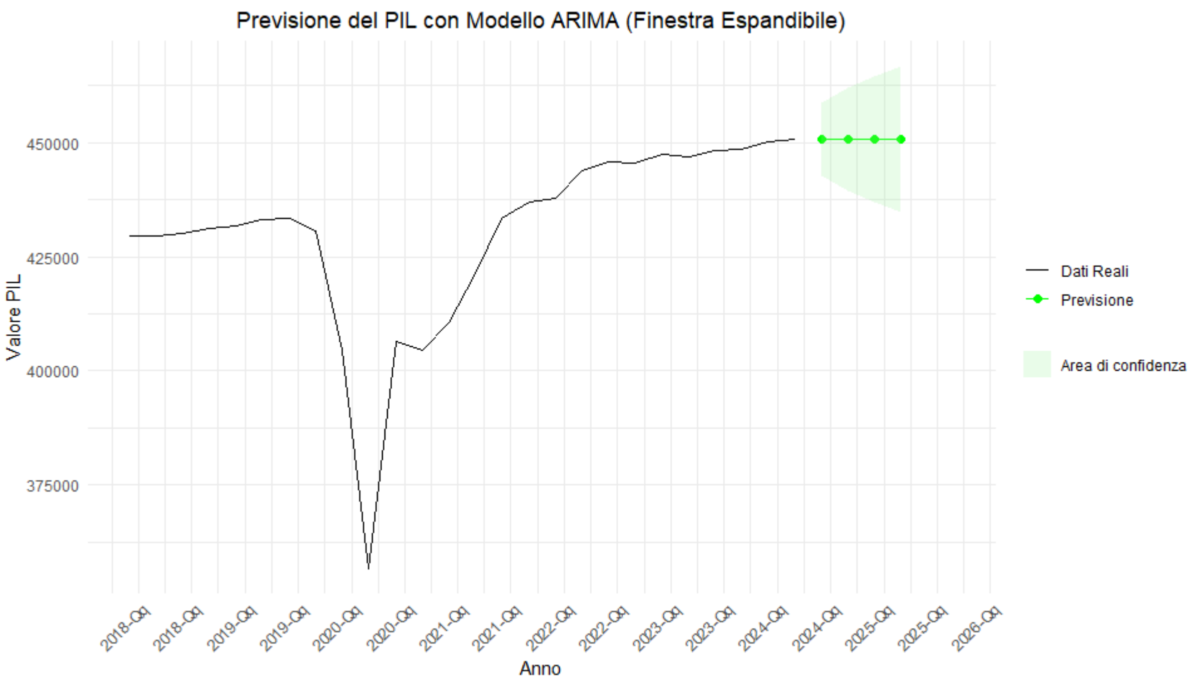

Figura 4.1: Previsione Finestra Espandibile ARIMA del PIL

Tabella 4.1: Previsioni PIL: 2024-Q3 – 2025-Q2

Trimestre	Valore Previsto	Limite Inferiore	Limite Superiore
2024-Q3	450.759,4	442.813,0	458.705,8
2024-Q4	450.759,4	439.521,4	461.997,4
2025-Q1	450.759,4	436.995,8	464.523,0
2025-Q2	450.759,4	434.866,5	466.652,3

Nota: Valori espressi in milioni di euro. Intervalli di confidenza al 70%.

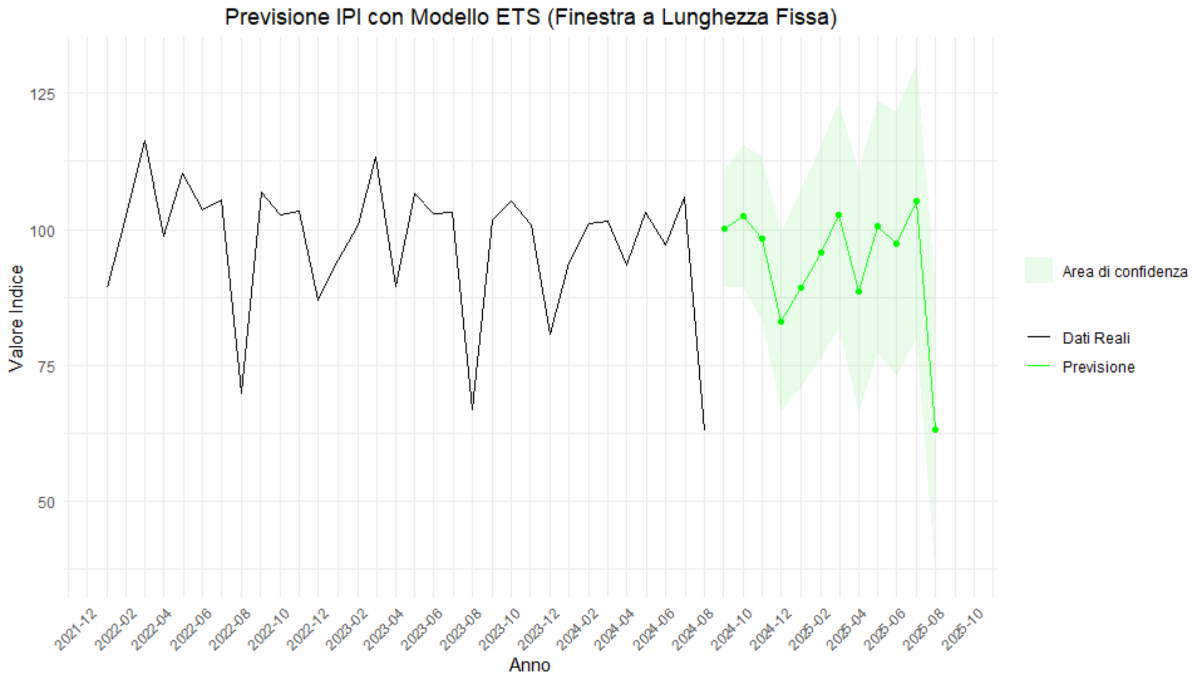

Figura 4.2: Previsione Finestra a Lunghezza Fissa dell'IPI

Tabella 4.2: Previsioni IPI: Settembre 2024 – Agosto 2025

Mese	**Valore Previsto**	**Limite Inferiore**	**Limite Superiore**
2024-09	100,13	89,20	111,06
2024-10	102,41	89,31	115,52
2024-11	98,32	83,35	113,29
2024-12	82,91	66,28	99,53
2025-01	89,20	71,08	107,33
2025-02	95,79	76,27	115,31
2025-03	102,71	81,89	123,52
2025-04	88,47	66,44	110,51
2025-05	100,60	77,41	123,79
2025-06	97,32	73,03	121,62
2025-07	105,21	79,86	130,55
2025-08	63,20	36,85	89,56

Nota: Intervalli di confidenza al 95%.

Bibliografia

[1] E. Gardner e E. McKenzie, *Forecasting trends in time series.* Management Science, 31(10), 1237–1246., 1985.

[2] R. Hyndman e G. Athanasopoulos, *Forecasting: Principles and Practice 2nd edition*. OTexts: Melbourne, Australia, 2018.

[3] R. Hyndman, G. Athanasopoulos, C. Bergmeir et al., *forecast: Forecasting Functions for Time Series and Linear Models*, version 8.23.0, 2024. indirizzo: `https://cran.r-project.org/web/packages/forecast/index.html`.

[4] R. Hyndman e Y. Khandakar, *Automatic time series forecasting: The forecast package for R. Journal of Statistical Software*. 27(3), 1-22, 2008.

[5] R. Hyndman, A. Koehler, J. Ord e R. Snyder, *Forecasting with Exponential Smoothing: The State Space Approach.* Springer, 2008.

[6] ISTAT, *Indice della produzione industriale - Mensili (base 2021)*.

[7] ISTAT, *Prodotto interno lordo e principali componenti - Trimestrali (base 2015)*.

[8] N. G. Mankiw, *Macroeconomics (10ª ed.)* Worth Publishers, 2019.

Ringraziamenti

Alla fine di questo percorso mi sento in dovere di dedicare qualche parola alle persone che mi hanno accompagnato.
Un ringraziamento sentito va al Professore Tommaso Di Fonzo, per il prezioso supporto e per la disponbilità che mi è stato offerta durante la stesura e per la passione con cui mi è stato trasmesso l'interesse per la materia.
Un pensiero va alla mia famiglia per tutto il loro sostegno incondizionato e per tutte le volte che mi hanno dimostrato il loro supporto.
Ad amici e colleghi senza i quali non riuscirei a stare e che sono stati compagni di questo lungo viaggio.
Infine un ringraziamento speciale ad Azzurra, capace di sostenermi sempre con costanza ed amore in ogni momento di incertezza e rimanendo sempre al mio fianco durante ogni scelta importante.
Grazie di cuore a tutti.

Francesco Bortolozzo