Università degli Studi di Padova

Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in

Statistica per l'Economia e l'Impresa

Modello "ARMA Normale Generalizzato" applicato alle aree dell'economia, dell'idrologia e delle politiche pubbliche

Relatore: Francesco Lisi

Laureando:

Armando Xeka

Matricola: 1127357

Anno Accademico 2024/2025

Indice

1. Introduzione 3

2. Distribuzione Normale Generalizzata……. 4

2.1 Definizione e descrizione della distribuzione 4

2.2 Parametri e la loro influenza 5

3. Modello ARMA Normale Generalizzato 8

3.1 Struttura del modello GN-ARMA 10

3.2 Previsione e criteri di selezione AIC/BIC 11

3.3 Vantaggi rispetto ai Modelli Tradizionali 12

4 Simulazione..……..……..……..……..…13

4.1 Procedura e spiegazione.....13

4.2 Scelta dell'ordine del modello.....17

5. Applicazioni a Dati Reali……..…..…….……..…21

5.1 Inflazione USA(HIPC)……..……..……22

5.2 Flusso idrico(Furnas)……..……..……..……25

5.3 Trasporto Pubblico (Iowa City)……..……..……29

5.4 Impatto del modello GN-ARMA ……..……33

Conclusione e sviluppi futuri 35

Bib1liografia.....36

Capitolo 1

Introduzione

Negli ultimi decenni, sono stati proposti in letteratura vari modelli di serie temporali. Di recente, c'è stato un crescente interesse nello sviluppo di processi non gaussiani a causa della presenza di serie temporali reali che violano l'ipotesi gaussiana. In questo lavoro ho riportato l'analisi svolto nel 2017 da "Eder Milani, Marinho Andrade & Carlos Diniz" dove è stato introdotto per la prima volta il modello ARMA normale generalizzato (GN-ARMA) con idee dal modello GARMA. Le serie temporali forniscono un quadro dinamico dei fenomeni analizzati, permettendo di identificare tendenze, stagionalità e anomalie che possono influenzare le decisioni strategiche in diversi settori. Molti fenomeni reali non seguono una distribuzione gaussiana, mostrando code più pesanti o leggere, asimmetrie o dinamiche non lineari. Ad esempio, eventi estremi come crisi economiche, siccità prolungate o picchi improvvisi nella domanda di trasporto pubblico sfidano i modelli tradizionali, che spesso non riescono a catturare tali caratteristiche. Il modello GN-ARMA rappresenta un'estensione dei modelli ARMA tradizionali, incorporando la flessibilità della distribuzione normale generalizzata per affrontare queste sfide. La capacità di adattarsi a distribuzioni non gaussiane e di gestire dinamiche complesse rende il GN-ARMA uno strumento ideale per analisi più accurate e robuste. Obiettivo della mia tesi è quello di replicare l'analisi sull'efficacia del modello GN-ARMA e come quest'ultimo superi i modelli tradizionali nella gestione di serie temporali non gaussiane. Nelle applicazioni pratiche si analizzano dati reali provenienti da settori economici, idrologici e delle politiche pubbliche per verificare la robustezza e l'utilità del modello. Nell'area dell'economia, è stato adottato la serie temporale dell'inflazione (HICP) degli Stati Uniti. Nell'area dell'idrologia, si è utilizzato la serie temporale sul flusso medio mensile di affluente del piano idroelettrico di Furnas. Nell'area della politica pubblica, è stato scelto l'utenza degli autobus da Iowa City. Le serie che coinvolgono l'utenza dei trasporti pubblici.

Capitolo 2

Distribuzione Normale Generalizzata

2.1 Definizione e descrizione della distribuzione

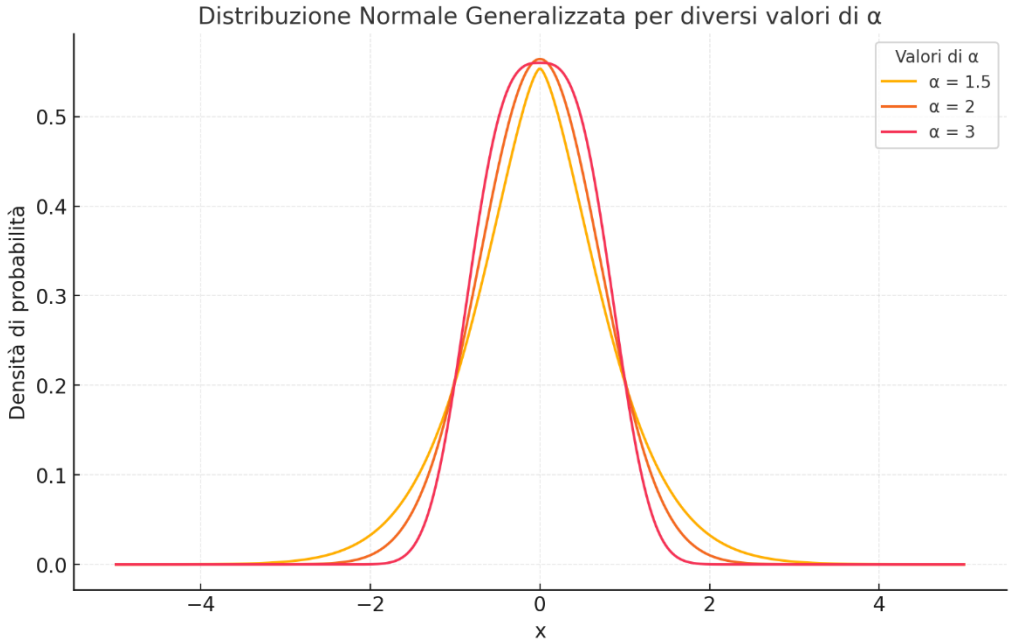
La distribuzione normale generalizzata (GND) rappresenta un'estensione della distribuzione normale classica, progettata per modellare fenomeni con code più pesanti o più leggere rispetto alla normale standard, nonché per gestire dati caratterizzati da asimmetria. Questa distribuzione è stata sviluppata per superare i limiti della distribuzione normale classica, che assume una simmetria perfetta e una determinata forma di curtosi. La distribuzione normale generalizzata è definita tramite una funzione di densità che include parametri che permettono di modificare la forma della distribuzione normale standard. In particolare, la GND consente di modificare la simmetria cioè a differenza della normale standard la distribuzione normale generalizzata può essere asimmetrica. Inoltre può avere code più pesanti o più leggere rispetto alla normale, consentendo di modellare fenomeni con varianze estremamente elevate o con valori estremi più frequenti. La forma della distribuzione GND può adattarsi a distribuzioni che sono più piatte o più piccole rispetto alla normale. Nadarajah (2005) ha proposto una versione della distribuzione normale generalizzata che consente di gestire un'ampia gamma di forme adattabili a molte applicazioni in statistica, finanza, ingegneria.

2.2 Parametri e la loro influenza

La funzione di densità di probabilità (PDF) della distribuzione normale generalizzata è definita come:

$$f(x \mid \mu, \sigma, s) = \frac{s}{2\sigma\Gamma(1/s)} \exp\left(-\left|\frac{x-\mu}{\sigma}\right|^{s}\right),$$

Dove si ha μ=Parametro di posizione (media della distribuzione),$\Gamma(\cdot)$: Funzione Gamma.σ = Parametro di scala (dispersione o deviazione standard) .Per $\sigma > 1$ la distribuzione è più larga e schiacciata invece con σ<1 appare più stretta e appuntita. Una proprietà interessante della distribuzione normale generalizzata è la flessibilità delle code, che possono essere più leggere o più pesanti della distribuzione gaussiana a seconda del valore di s. Per s<2, la distribuzione ha code più pesanti cioè si ha più probabilità di osservare eventi estremi e può essere asimmetrica. Questo implica una maggiore probabilità di osservare eventi estremi come i ritorni di asset finanziari in presenza di cigni neri (eventi estremi imprevedibili). Quando s>2, la distribuzione ha code più leggere. Questo rende la GND adatta per modelli in cui gli eventi estremi sono meno frequenti e la distribuzione tende a concentrarsi più fortemente intorno ai valori centrali.



. Fig 1:Grafico della GND per diversi valori di α con α = s .

La distribuzione normale generalizzata possiede media e varianza ottenuti dal momento n-esimo e momento centrale come viene dimostrato nelle formule sottostanti. La varianza dipende non solo dal parametro σ ma anche da s.

1. **Momento n-esimo della distribuzione normale generalizzata**:

$$E(Y^n) = \mu^n \sum_{k=0}^{n} \left(\frac{\sigma}{\mu}\right)^k [1 + (-1)^k] \frac{\Gamma\left(\frac{k+1}{s}\right)}{\Gamma\left(\frac{1}{s}\right)}$$

   Per $n = 1$, si ottiene $E(Y) = \mu$.

2. **Momento centrale** :

$$E[(Y - \mu)^n] = \sigma^n [1 + (-1)^n] \frac{\Gamma\left(\frac{n+1}{s}\right)}{\Gamma\left(\frac{1}{s}\right)}$$

   Per $n = 2$, otteniamo la varianza:

$$\mathrm{Var}(Y) = \sigma^2 \frac{\Gamma(3/s)}{\Gamma(1/s)}$$

CAPITOLO 3

Modello ARMA Normale Generalizzato

Il modello **Normale Generalizzato ARMA (GN-ARMA)** è un'estensione del classico modello ARMA, progettata per modellare serie temporali che presentano caratteristiche non gaussiane, come code pesanti o leggere e diversa curtosi basato sulla distribuzione normale generalizzata anche sul modello GARMA. Questo lo rende più flessibile nella modellazione di serie temporali con asimmetria e diversa curtosi. Il modello mantiene una struttura di dipendenza lineare con parametri di autoregressione e media mobile. Le proprietà marginali e condizionali del modello sono derivate per garantire un'applicazione robusta.

3.1 Struttura del Modello GN-ARMA

Il modello GN-ARMA si basa sulla relazione tra la media condizionale del processo, μ_t, e il predittore lineare η_t. Seguendo un approccio simile ai modelli GARMA (Benjamin et al., 2003), il modello GN-ARMA utilizza una funzione di link $g(\cdot)$, monotona e due volte differenziabile, per esprimere questa relazione:

$$g(\mu_t) = \eta_t = x_t'\beta + \tau_t$$

Ove $x_t'\beta$ rappresenta il contributo delle variabili esogene,τ_t è la componente ARMA data da :

$$\tau_t = \sum_{j=1}^{p} \phi_j \{g(y_{t-j}) - x_{t-j}'\beta\} + \sum_{j=1}^{q} \theta_j \{g(y_{t-j}) - \eta_{t-j}\}.$$

Nel caso particolare in cui la funzione di link sia l'identità ($g(\mu_t) = \mu_t$), otteniamo la formulazione standard :

$$\mu_t = x_t'\beta + \sum_{j=1}^{p} \phi_j \left(y_{t-j} - x_{t-j}'\beta\right) + \sum_{j=1}^{q} \theta_j \left(y_{t-j} - \eta_{t-j}\right).$$

Nel caso ci fossero le variabili esogene, l'equazione diventa:

$$\mu_t = \beta_0 + \sum_{j=1}^{p} \phi_j \, y_{t-j} + \sum_{i=1}^{q} \theta_j \left(y_{t-j} - \mu_{t-j}\right).$$

La formula teorica per la media marginale del modello GN-ARMA è $E[Y_t] = \frac{\beta_0}{1-\sum_{j=1}^{p}\phi_j}$. Se si includono variabili esogene, diventa $E[Y_t] = x_t'\beta$.

Poi la varianza marginale $Var(Y_t) = \sigma^2 \frac{\Gamma(3/s)}{\Gamma(1/s)} \sum_{i=1}^{\infty} \psi_i^2$, dove il polinomio degli operatori di ritardo $\Psi(B)$ è definito come $\Psi(B) = \frac{1}{\Phi(B)}\Theta(B)$ con $\Phi(B) = 1 - \sum_{j=1}^{p} \phi_j B^j$ (parte autoregressiva) e $\Theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j$ (parte di media mobile).Per $p = q = 1$, la varianza si semplifica in:

$$Var(Y_t) = \sigma^2 \frac{\Gamma\left(\frac{3}{s}\right)}{\Gamma\left(\frac{1}{s}\right)} \left(\frac{1 + \phi_1 + \theta_1}{1 - \phi_1^2}\right).$$

3.2 Previsione e criteri di selezione AIC/BIC

La previsione ad un passo avanti (h = 1) prevede il valore della serie temporale nel momento successivo basandosi sui dati osservati fino all'istante corrente. Formalmente, se Y_t è una serie temporale, la previsione ad un passo avanti $\hat{Y}_{t+1}$ viene calcolata come:

$$\hat{Y}_{t+1} = E(Y_{t+1} \mid F_t)$$

Per poi arrivare alla formula:

$$\hat{Y}_{t+1} = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\phi}_j \, y_{t+1-j} + \sum_{i=1}^{q} \hat{\theta}_i \left(y_{t+1-j} - \hat{\mu}_{t+1-j}\right)$$

dove F_t è l'insieme delle informazioni disponibili fino al tempo t. $\hat{\beta}_0$, $\hat{\phi}_j$ e $\hat{\theta}_j$ per j = 1, . . . , p e i = 1, . . . , q sono gli stimatori di massima verosimiglianza dei parametri β_0, ϕ_j e θi per j = 1, . . . , p e i = 1, . . . , q, rispettivamente, e $\hat{\mu}_{t+1-j}$ si trova assumendo $\hat{\mu}_1 = \cdot \cdot \cdot = \hat{\mu}_q = 0$ e sostituendo gli stimatori dei parametri nell'equazione:

$$\mu_t = \beta_0 + \sum_{j=1}^{p} \phi_j \, y_{t-j} + \sum_{i=1}^{q} \theta_j \left(y_{t-j} - \mu_{t-j}\right)$$

La previsione ad un passo avanti è generalmente più accurata perché utilizza solo valori osservati.

La previsione a multi passo avanti prevede il valore della serie in un istante futuro $t + h$, con $h > 1$.

$$\hat{Y}_{t+h} = \beta_0 + \sum_{j=1}^{p} \phi_j \, \mathbb{E}[Y_{t+h-j} \mid F_t] + \sum_{j=1}^{q} \theta_j \left(\mathbb{E}\left[Y_{t+h-j} \mid F_t \right] - \hat{\mu}_{t+h-j} \right)$$

Per proiezioni a lungo termine, si opta per una previsione multi passo, consapevoli della possibile propagazione dell'errore. L'errore nelle previsioni a multi passo avanti tende ad aumentare con h perché ogni passo dipende dal precedente e introduce un errore aggiuntivo. Per determinare l'ordine ottimale (p,q) del modello GN-ARMA, si utilizzano criteri di selezione **Akaike Information Criterion (AIC)**: AIC=$-2\ell+2k$, dove k è il numero di parametri stimati e **Bayesian Information Criterion (BIC)**: BIC=$-2\ell+k\ln(n)$, dove n è la dimensione del campione. L'idea è scegliere il modello che minimizza il valore di AIC o BIC bilanciando la bontà dell'adattamento del modello con la sua complessità. I parametri del modello (β_0,ϕ,θ,σ,s) sono stimati tramite Massima Verosimiglianza (MLE). La funzione di log-verosimiglianza è:

$$l(\beta_0, \phi, \theta, \sigma, s \mid y_r) = (n-r)\ln\left(\frac{s}{2\sigma\Gamma(1/s)}\right) - \frac{1}{\sigma^s} \sum_{t=r+1}^{n} \mid y_t - \mu_t \mid^s.$$

Gli stimatori di massima verosimiglianza vengono trovati utilizzando metodi iterativi, ad esempio il metodo L-BFGS-B (Byrd et al., 1995).

3.3 Vantaggi riaspetto ai Modelli Tradizionali

I modelli GARMA **(**Modello Autoregressivo a Media Mobile Generalizzato**)** sono un'estensione dei modelli ARMA , progettati per affrontare situazioni in cui le assunzioni di normalità e omoscedasticità dei residui, tipiche dei modelli ARMA, non sono soddisfatte. In altre parole, i modelli GARMA offrono una maggiore flessibilità nell'analisi di serie temporali con errori che non seguono una distribuzione normale o che mostrano variazioni nella loro varianza nel tempo (eteroscedasticità). Combinano componenti autoregressive (AR) e a media mobile (MA), ma con una struttura più generale e flessibile per la modellizzazione dei residui, i cosiddetti errori o innovazioni del modello.Matematicamente, un modello GARMA può essere espresso come:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t$$

Dove μ è la media della serie,$\phi_1, \phi_2, \ldots$ sono i parametri autoregressivi,$\theta_1, \theta_2, \ldots$ sono i parametri della media mobile,ϵ_t sono i residui (errori).Il modello GN-ARMA combina i punti di forza dei modelli ARMA tradizionali e dei modelli GARMA, portando significativi vantaggi in contesti specifici. Il GN-ARMA rispetto agli altri due aggiunge la capacità di gestire relazioni non lineari. Mentre il modello ARMA tradizionale è limitato a relazioni lineari, il GN-ARMA può modellare dinamiche non lineari tra le variabili della serie temporale e i termini di errore. Per fenomeni complessi è ideale, come fluttuazioni non lineari nei mercati finanziari, sistemi biologici, o processi caotici. Comporta maggiore flessibilità cioè permette di catturare comportamenti che i modelli ARMA lineari non possono descrivere, come salti improvvisi, saturazione. effetti di soglia. Migliora la gestione di dinamiche complesse cioè può adattarsi a cambiamenti dinamici nel tempo.

Rispetto al GARMA ha un adattamento specifico a serie temporali cioè mentre questo ultimo si concentra su relazioni tra variabili indipendenti e dipendenti, ma non sono progettati specificamente per catturare la dipendenza temporale intrinseca in una serie temporale ,il GN-ARMA integra esplicitamente relazioni temporali e non lineari, superando questa limitazione. Si nota una maggiore attenzione alla struttura stocastica incorporando componenti autoregressive e di media mobile, il che lo rende più adatto a modellare l'evoluzione stocastica dei dati rispetto al GARMA, che si focalizza più sulla struttura della distribuzione.Il GN-ARMA gestisce meglio l'autocorrelazione residua nei dati e del rumore, modellando anche effetti non lineari nei termini di errore.

CAPITOLO 4

SIMULAZIONE

4.1 Procedura e spiegazione

L'obiettivo della simulazione è valutare la qualità degli stimatori di massima verosimiglianza (MLE) e analizzare l'influenza del parametro **s**, che controlla la pesantezza delle code nella distribuzione normale generalizzata. La generazione di una serie temporale per il modello GN-ARMA segue diversi passaggi. Come inizio si definiscono i parametri del modello assegnando valori iniziali ai parametri del modello (ad esempio, media, varianza).Poi si assumono condizioni iniziali (ad esempio, valori nulli per alcune osservazioni iniziali) per ridurre effetti indesiderati $\mu_t = 0$ per $t = 1, \dots, r$, dove $r = \max(p, q)$, $Y_t = 0$ per $t = 1, \dots, r - 1$ .Si costruisce una serie di dati simulati utilizzando una distribuzione normale generalizzata con i parametri scelti. Per ogni serie generata, si calcolano le stime dei parametri mediante il metodo della massima verosimiglianza scartando i primi N_0 valori per evitare effetti delle condizioni iniziali(Burn-in).Si è svolto una valutazione delle prestazioni analizzando le metriche di accuratezza come il bias medio assoluto (MAB), l'errore quadratico medio (MSE) e la probabilità di copertura (CP), confrontando le stime con i valori veri dei parametri.Infine l'intero processo viene ripetuto su 1000 campioni con diverse dimensioni (n = 100, 200, 500 e 1000) per ottenere risultati statistici affidabili. Genera Y_t da una distribuzione normale generalizzata dove Γ è la funzione gamma.

$$f(y) = \frac{s}{2\sigma\Gamma(1/s)} \exp\left(-\left|\frac{y-\mu}{\sigma}\right|^s\right)$$

La Tabella 1 riporta i risultati della simulazione per diverse dimensioni del campione e diversi valori del parametro s. Questo parametro regola la forma della distribuzione normale generalizzata.Valori di s più piccoli (ad es, 1.5) indicano una distribuzione con code più pesanti rispetto alla normale standard.Valori più grandi (ad esempio, 4.5) indicano una distribuzione con code più leggere.Sono state calcolate le 4 metriche per ogni parametro cioè Media degli stimatori (EM),

Deviazione assoluta media (MAB) dato da : $MAB(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} \mid \beta_0^* - \widehat{\beta_0} \mid$,

Errore quadratico medio(MSE) : $MSE(\beta_0) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\beta_0^* - \widehat{\beta_0}\right)^2$,

Infine la probabilità di copertura (CP) degli intervalli di confidenza al 95% definite come il quoziente tra il numero degli intervalli di confidenza che contengono il valore del parametro reale e il numero totale di intervalli

Parametro	MAB (s=1.5)	EM (s=1.5)	MSE (s=1.5)	CP (s=1.5)	MAB (s=3)	EM (s=3)	MSE (s=3)	CP (s=3)	MAB (s=4.5)	EM (s=4.5)	MSE (s=4.5)	CP (s=4.5)
n=100												
$\beta 0$	0.1896	0.1036	0.0583	0.8750	0.1379	0.1052	0.0309	0.9320	0.1107	0.1029	0.0204	0.9360
φ_1	0.1582	0.1765	0.0422	0.8450	0.1548	0.1696	0.0391	0.9110	0.1391	0.1754	0.0325	0.8970
θ_1	0.1573	0.3598	0.0413	0.8130	0.1487	0.3706	0.0375	0.8780	0.1337	0.3652	0.0305	0.8870
σ	0.3128	1.9827	0.1628	0.9170	0.1642	1.9853	0.0438	0.9160	0.1175	1.9735	0.0226	0.9330
s	0.3276	1.6216	0.2541	0.9300	1.0910	3.6829	3.3141	0.9570	1.8578	5.7015	8.1732	0.9610
n=200												
$\beta 0$	0.1309	0.1031	0.0279	0.9310	0.0934	0.0992	0.0128	0.9450	0.0748	0.0135	0.0090	0.9390
φ_1	0.1047	0.1888	0.0181	0.9190	0.1073	0.1845	0.0185	0.9350	0.0939	0.1848	0.0140	0.9240
θ_1	0.1017	0.3584	0.0167	0.9230	0.1043	0.3597	0.0171	0.9300	0.0909	0.3594	0.0129	0.9110
σ	0.2114	1.9771	0.0699	0.9440	0.1083	1.9920	0.0190	0.9370	0.0795	1.9957	0.0102	0.9330
s	0.1905	1.5338	0.0627	0.9510	0.5500	3.2591	0.6059	0.9560	1.1963	5.2253	3.3017	0.9570
n=500												
$\beta 0$	0.0831	0.0986	0.0110	0.9490	0.0563	0.1023	0.0050	0.9480	0.0458	0.1021	0.0033	0.9530
φ_1	0.0678	0.1921	0.0075	0.9290	0.0664	0.1923	0.0068	0.9430	0.0565	0.1931	0.0051	0.9390
θ_1	0.0639	0.3557	0.0066	0.9370	0.0604	0.3530	0.0057	0.9490	0.0545	0.3529	0.0047	0.9470
σ	0.1331	1.9940	0.0277	0.9380	0.0655	1.9991	0.0069	0.9410	0.0478	1.9991	0.0036	0.9460
s	0.1175	1.5129	0.0219	0.9410	0.2954	3.0859	0.1456	0.9630	0.5374	4.7123	0.5245	0.9630
n=1000												
$\beta 0$	0.0584	0.0584	0.0054	0.9390	0.0402	0.1017	0.0025	0.9520	0.0316	0.1017	0.0016	0.9490
φ_1	0.0481	0.0481	0.0036	0.9460	0.0465	0.1964	0.0034	0.9300	0.0384	0.1968	0.0024	0.9470
θ_1	0.0453	0.0453	0.0032	0.9470	0.0436	0.3508	0.0030	0.9420	0.0370	0.3508	0.0022	0.9450
σ	0.0946	0.0946	0.0141	0.9390	0.0468	1.9979	0.0034	0.9480	0.0342	1.9989	0.0018	0.9580
s	0.0813	0.0813	0.0105	0.9450	0.1938	3.0304	0.0626	0.9530	0.3473	4.5817	0.1982	0.9600

Tabella 1: MAB, EM, MSE, e CP risultati per i differenti valori di *s*.

Sono stati utilizzati tre valori di s: 1.5, 3.0, e 4.5 nella tabella 1 per valutare la sensibilità delle metriche.Per questo studio di simulazione, si è assunto il modello GN-ARMA(1,1) con parametri impostati a $\beta_0 = 0.1$; $\varphi_1 = 0.2$, $\theta_1 = 0.35$,$\sigma = 2$.I risultati sono presentati nella Tabella 1, in cui si può osservare che quando aumenta la dimensione del campione, si riducono le metriche MAB e MSE, indicando che gli stimatori diventano più accurati e precisi. La probabilità di copertura (CP) e (EM) si avvicina sempre più al valore nominale del 95%, confermando la validità degli intervalli di confidenza. Per valori crescenti di s, MAB e MSE tendono a diminuire per la maggior parte dei parametri (β_0,ϕ_1,θ_1 ,σ), suggerendo una maggiore stabilità nelle stime .Per piccoli campioni (n=100), si osservano deviazioni dalla probabilità di copertura nominale, specialmente per ϕ_1 e θ_1 , indicando una minore affidabilità delle stime in condizioni di scarsità di dati. Questi risultati confermano che il metodo di stima basato sulla massima verosimiglianza per il modello GN-ARMA è efficace e che le sue proprietà asintotiche migliorano con l'aumentare della dimensione del campione. Il parametro s, che controlla la pesantezza delle code della distribuzione normale generalizzata, influisce sulla qualità delle stime.

Parametro	Metrica	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	Osservazione
β_0	MAB	0.1896	0.1309	0.0831	0.0584	Diminuisce con n, alta precisione con $n \geq 500$.
	CP (%)	87.5	93.1	94.9	93.9	Stabilità vicino al 95% per $n \geq 200$.
ϕ_1	MAB	0.1582	0.1047	0.0678	0.0481	Stime sempre più accurate con aumentare di n.
	CP (%)	84.5	91.9	92.9	94.6	Incremento progressivo, vicino al livello nominale con campioni grandi.
θ_1	MAB	0.1573	0.1017	0.0639	0.0453	Comportamento simile a ϕ_1, ma lievemente meno accurato per campioni piccoli.
	CP (%)	81.3	92.3	93.7	94.7	Alta affidabilità per $n \geq 500$.
σ	MAB	0.3128	0.2114	0.1331	0.0946	Precisione accettabile già con $n = 200$.
	CP (%)	91.7	94.4	93.8	93.9	Copertura consistente vicino al 95%.
s	MAB	0.3276	0.1905	0.1175	0.0813	Miglioramento con > di n, ma meno accurato rispetto agli altri parametri.
	CP (%)	93.0	95.1	94.1	94.5	Alta copertura per tutte le dimensioni.

Tabella 1.1 : Riassunto della tab. 1 per le metriche MAB e CP con s = 1.5.

Come si può osservare anche nella Tab 1.1 i valori della probabilità di copertura (CP) sono sempre più vicini al 95% con aumentare del valore di n per tutti gli parametri passando da un minimo di 84.5 % per n=100 ad un massimo di 94.7 per n=1000.I valori di (MAB) migliorano soprattutto da n=500 però già ad n=200 si ha una precisione accettabile per le stime per tutti gli parametri a differenza di σ e s che presentato miglioramenti con aumentare di n ma meno accurati.

4.2 Scelta dell'Ordine del Modello

Per verificare quanti termini autoregressivi e di media mobile dovrebbero essere incorporati nel modello per spiegare meglio le serie temporali e stato svolto uno studio di simulazione. Innanzitutto si genera una serie temporale di dimensione n.Si ottengono gli stimatori di massima verosimiglianza dei parametri del modello GN-ARMA di ordini diversi, come (1,1), (1,2), (2,1) e (2,2) con differenti valori dei parametri per ogni modello generato.Per ciascun modello, è stata testata la capacità degli indicatori AIC e BIC di identificare correttamente il vero ordine del modello. AIC si concentra sull'accuratezza del modello ma penalizza meno la complessità mentre BIC penalizza maggiormente la complessità del modello, privilegiando modelli più semplici con diverse dimensioni del campione (n=100,200,500,1000).

Modello Generato	**Parametri**	**M.Stimato (n = 100)**	**M.Stimato (n = 200)**	**M.Stimato (n = 500)**	**M.Stimato (n = 1000)**
GN-ARMA(1,1)	$\beta_0 = 0.1$, $\phi_1 = 0.9$, $\theta_1 = 0.2$, $\sigma = 2.5$, $s = 3$	64.6/72.6(1,1) 7.2 / 7.6(1,2) 22.4/18.7(2,1) 5.8/ 1.1 (2,2)	72.0/83.7(1,1) 4.7 / 4.5 (1,2) 15.9/11.4(2,1) 7.4 / 0.4 (2,2)	75.8/89.7(1,1) 4.5 / 4.0 (1,2) 8.3 / 5.6 (2,1) 11.4 /0.7 (2,2)	78.5 / 93.1 (1,1) 3.2 / 3.0 (1,2) 6.6 / 3.9 (2,1) 11.7 / 0.0 (2,2)
GN-ARMA(1,2)	$\beta_0 = 0.1$, $\phi_1 = 0.9$, $\theta_1 = 0.6$, $\phi_2 = -0.4$, $s = 3$	2.6/3.5(1,1) 65.2/69.4(1,2) 21.0/22.2(2,1) 11.2 /4.9 (2,2)	0.0 / 0.0(1,1) 79.5/84.3(1,2) 13.1/13.4(2,1) 7.4 / 2.3 (2,2)	0.0 /0.0(1,1) 91.7/96.5(1,2) 1.6 / 2.4 (2,1) 6.7 / 1.1 (2,2)	0.0 / 0.0 (1,1) 93.9 / 98.9 (1,2) 0.1 / 0.3 (2,1) 6.0 / 0.8 (2,2)
GN-ARMA(2,1)	$\beta_0 = 0.1$, $\phi_1 = 0.9$, $\phi_2 = -0.4$, $\theta_1 = 0.6$, $s = 3$	4.0/7.0(1,1) 5.3/5.7(1,2) 69.8/75.5(2,1) 20.9/11.8(2,2)	0.2/1.0(1,1) 3.0/3.4(1,2) 83.1/89.3(2,1) 13.7 /6.3 (2,2)	0.0 / 0.0 (1,1) 0.5 / 0.5 (1,2) 88.5/94.5(2,1) 11.5 /5.0 (2,2)	0.0 / 0.0 (1,1) 0.0 / 0.0 (1,2) 89.4 / 97.1 (2,1) 10.6 / 2.9 (2,2)
GN-ARMA(2,2)	$\beta_0 = 0.1$, $\phi_1 = 0.95$, $\phi_2 = -0.65$, $\theta_1 = 0.8$, $\theta_2 = -0.35$, $s = 3$	0.0/3.5(1,1) 0.5/0.5(1,2) 44.0/54.9(2,1) 55.2/44.1(2,2)	0.0/0.0(1,1) 0.0/0.0(1,2) 24.0/37.3(2,1) 76.0/62.7(2,2)	0.0 / 0.0 (1,1) 0.0 / 0.0 (1,2) 3.5 / 8.0 (2,1) 96.5/92.0(2,2)	0.0 / 0.0 (1,1) 0.0 / 0.0 (1,2) 0.0 / 0.8 (2,1) 100.0 / 99.2 (2,2)

Tabella 2: Percentuale di selezione corretta degli ordini tramite AIC/BIC

Dai risultati ottenuti durante si nota che maggiore è la dimensione del campione, più alta è la percentuale di selezione corretta infatti per il modello GN-ARMA(1,1) la percentuale di selezione corretta va dal 64,6% con 100 osservazioni fino al 78,5% con 1000 osservazioni per il criterio AIC, mentre per il BIC la precisione è ancora maggiore, raggiungendo il 93,1% per 1000 osservazioni. Le prestazioni per GN-ARMA(1,2) vanno dal 65.2% con n = 100 per AIC al 93.9% con n=1000 mentre per BIC si va da 69.4% a 98.9%.Il modello con ordini (2,1) ha valori simili ai due precedenti mentre l'ultimo il (2,2) per n=100 la selezione è meno precisa con AIC: 55.2% invece BIC= 44.1%.Però raggiunge con 1000 osservazioni valori pari a 99.2%.Infatti per modelli complessi come GN-ARMA(2,2), è necessaria una maggiore dimensione del campione per ottenere selezioni affidabili. Confrontando gli risultati AIC è più permissivo nel selezionare modelli complessi al invece per modelli semplici è meno preciso rispetto a BIC. Quindi BIC preferisce gli modelli semplici e diventa estremamente affidabile per grandi dimensioni tipo n=1000.L'analisi degli ordini suggerisce che il GN-ARMA identifica correttamente la struttura della serie temporale anche per campioni di piccole dimensioni (n=100), con una precisione superiore al 60% eccetto nell'ordine p = 2 e q = 2 .La selezione corretta aumenta con la dimensione del campione, superando l'80% per n≥500 ad eccezione del criterio AIC nell'ordine p = q = 1.

I risultati mostrano che la selezione corretta dell'ordine del modello GN-ARMA dipende fortemente dalla dimensione del campione. Per campioni piccoli (n ≤ 200) i modelli più semplici GN-ARMA(1,1) e GN-ARMA(2,1) vengono selezionati correttamente con alta frequenza (superiore all'80% per BIC). Modelli più complessi come GN-ARMA(2,2) sono difficili da identificare correttamente. Per campioni medi (n = 500) il criterio BIC seleziona GN-ARMA(1,1), GN-ARMA(1,2) e GN-ARMA(2,1) con oltre il 90% di precisione. GN-ARMA(2,2) viene riconosciuto con maggiore accuratezza, ma con qualche incertezza residua. Infine campioni grandi (n = 1000 o più): Tutti i modelli vengono selezionati con alta precisione. GN-ARMA(2,2) è identificato correttamente nel 99,2% dei casi dal BIC, rendendolo una scelta valida per dataset estesi. Come si nota anche nella figura 2.

.

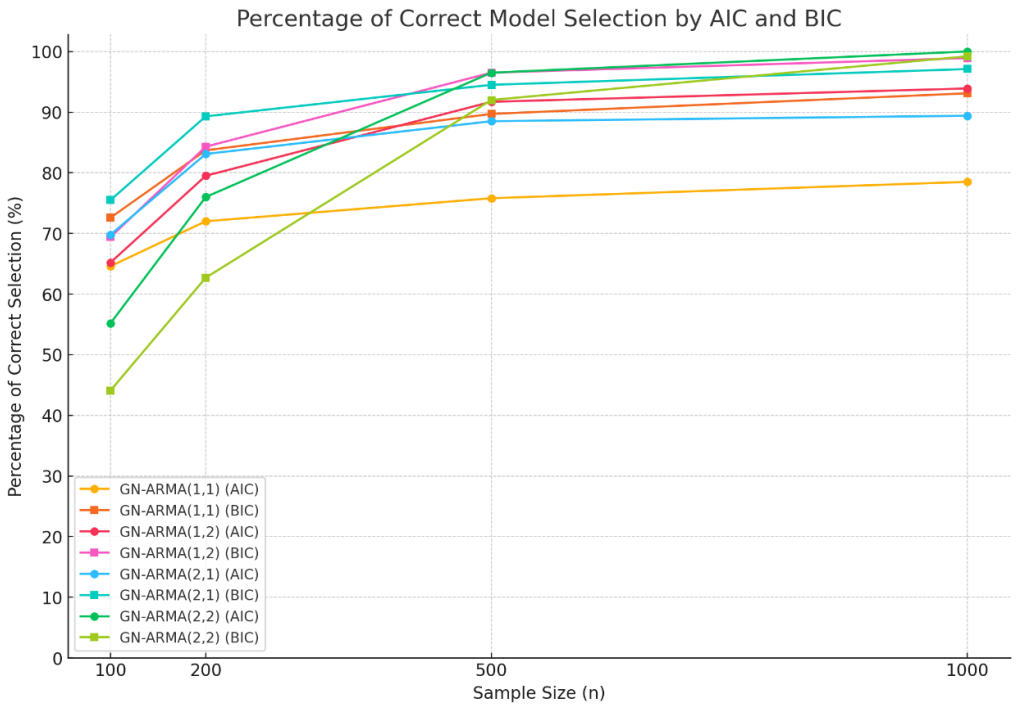


Fig.2 Percentuale di selezione corretta dell'ordine del modello

La Tabella 3 presenta i risultati di tre metriche di errore (MAPE, MSE, e MAE) utilizzate per valutare l'accuratezza delle previsioni del modello GN-ARMA(1,1) in diversi scenari di dimensioni campionarie della serie temporale n = 110, 210, 510 e 1010 con $\beta_0 = 2.0$; $\varphi_1 = 0.8$; $\theta = 0.35$; $\sigma = 1$; s = 1.5. Per ogni campione, sono state effettuate previsioni a 10 passi avanti mostrando una riduzione consistente degli errori all'aumentare della dimensione del campione. Gli intervalli di confidenza per i parametri stimati sono stati calcolati assumendo che le loro distribuzioni asintotiche siano normali, cioè: $\hat{\theta} \sim N\left(\theta, \mathrm{Var}(\hat{\theta})\right)$

Tab. 3:Risultati delle metriche di errore con i relativi intervalli di confidenza (IC)

Dimensione campionaria (n)	**MAPE (%)**		**MSE**		**MAE**	
	Media	**IC**	**Media**	**IC**	**Media**	**IC**
$n = 100$	14.12	(4.31-37.56)	2.95	(0.29 - 11.15)	1.33	(0.44 - 3.01)
$n = 200$	13.74	(4.69 - 36.14)	2.84	(0.33 - 11.18)	1.30	(0.48 - 2.99)
$n = 500$	13.82	(4.51 - 34.59)	2.69	(0.33 - 9.39)	1.27	(0.46 - 2.79)
$n = 1000$	13.83	(4.85 - 36.33)	2.73	(0.39 - 10.25)	1.28	(0.49 - 2.80)

Dove $\hat{\theta}$ è la stima del parametro (ad esempio $\hat{\beta}_0, \hat{\phi}_j, \hat{\theta}_j, \hat{\sigma}, \hat{s}$).$\mathrm{Var}(\hat{\theta})$ è la varianza della stima, ottenuta dall'inversa della matrice dell'informazione osservata. L'intervallo di confidenza al 95% è stato calcolato con la formula:

$$IC_{95\%} = \hat{\theta} \pm Z_{\alpha/2} \cdot \mathrm{SE}(\hat{\theta})$$

$Z_{\alpha/2}$ è il valore critico della distribuzione normale standard, pari a 1.96,

$\mathrm{SE}(\hat{\theta}) = \sqrt{\mathrm{Var}(\hat{\theta})}$ è l'errore standard della stima del parametro.

Attraverso l'analisi sui risultati si nota che i valori medi del MAPE diminuiscono leggermente con l'aumentare della dimensione campionaria. Questo suggerisce che le previsioni diventano più accurate man mano che la dimensione del campione aumenta e i relativi intervalli di confidenza (***CI***) si restringono con dimensioni campionarie più grandi, indicando maggiore stabilità e affidabilità delle previsioni.

Anche MSE segue una tendenza decrescente con l'aumento del campione, confermando che previsioni basate su campioni più grandi sono più precise.Gli intervalli di confidenza si riducono con l'aumentare della dimensione campionaria, il che indica che i valori osservati sono più vicini alle previsioni quando il campione è più ampio.Infine MAE segue lo stesso andamento delle altre due metriche.Come in figura 3.

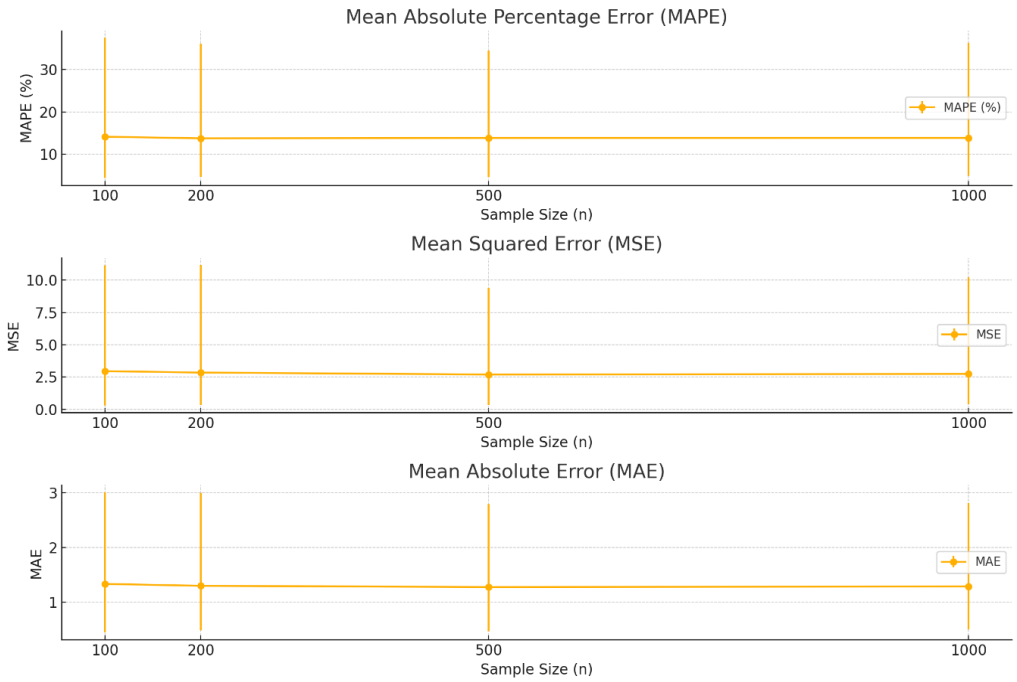


Fig.3 : Grafici delle metriche di errore con i relativi intervalli di confidenza (IC).

Capitolo 5

Applicazioni a Dati Reali

L'applicazione di modelli statistici a dati reali è un passaggio fondamentale nell'analisi delle serie storiche, in quanto consente di verificare la capacità del modello di adattarsi a dati concreti e di migliorare la previsione di fenomeni economici, ambientali o di politica pubblica. Nel caso specifico del modello GN-ARMA l'utilizzo di dati reali consente di testare la sua robustezza rispetto ai modelli classici ARMA e GARMA, valutando le sue prestazioni in termini di accuratezza della previsione e adattabilità a serie con code più pesanti o leggere rispetto alla distribuzione normale. Quindi l'obiettivo dell'applicazione ai dati reali è duplice cioè il primo è quello di validare il modello GN-ARMA confrontando i suoi risultati con quelli di modelli tradizionali ARMA e poi dimostrare la flessibilità del modello nell'adattarsi a diversi contesti. Per testare l'applicabilità del modello, sono state selezionate tre serie temporali appartenenti a differenti ambiti. Il primo ambito è "Economia" con argomento l'indice armonizzato dei prezzi al consumo (HICP) degli Stati Uniti. Il secondo è "Idrologia" attraverso la portata media mensile affluente della centrale idroelettrica di Furnas. Infine le "Politiche pubbliche" con il numero medio mensile di passeggeri del trasporto pubblico a Iowa City. Queste serie temporali presentano caratteristiche diverse in termini di distribuzione dei dati, stagionalità e volatilità, rendendole ideali per valutare l'efficacia del modello GN-ARMA. L'analisi delle serie temporali segue una procedura standard articolata in diversi passaggi. Si comincia con un pretrattamento dei dati attraverso una analisi esplorativa delle serie per identificare trend e stagionalità con trasformazioni necessarie (differenziazione, destagionalizzazione).Si sceglie il modello più adatto confrontando i modelli ARMA, GN-ARMA. Sono stati utilizzati i criteri informativi (AIC, BIC) per selezionare la migliore configurazione del modello. Gli autori hanno stimato dei parametri del modello utilizzando il metodo della massima verosimiglianza. Valutazione della bontà dell'adattamento ed infine previsione e interpretazione dei risultati confrontando le previsioni con i valori osservati attraverso metriche di errore (MAPE, MSE, MAE).

5.1 Indice Armonizzato dei Prezzi al Consumo (HICP)

Nel contesto economico, l'articolo applica il modello GN-ARMA alla serie temporale dell'Indice Armonizzato dei Prezzi al Consumo (HICP) degli Stati Uniti. L'obiettivo è migliorare l'analisi e la previsione di questa serie temporale rispetto ai modelli ARMA standard basati su una distribuzione gaussiana.La serie temporale utilizzata comprende osservazioni mensili dal dicembre 1998 al maggio 2014.È stata applicata una differenziazione ai dati per rimuovere tendenze stazionarie.L'analisi preliminare ha condotto a funzioni di autocorrelazione (ACF) e autocorrelazione parziale (PACF) con correlazioni nelle prime lag e attorno al dodicesimo lag come nella figura 5, suggerendo la presenza di componenti stagionali. Le funzioni ACF e autocorrelazione parziale PACF dei residui del modello hanno mostrato l'assenza di correlazione, confermando un buon adattamento ai dati. Questo ha portato alla necessità di includere una componente stagionale nel modello. Dopo aver differenziato i dati per rimuovere le tendenze, sono stati considerati vari modelli con componenti stagionali (es. lag di 12 mesi).

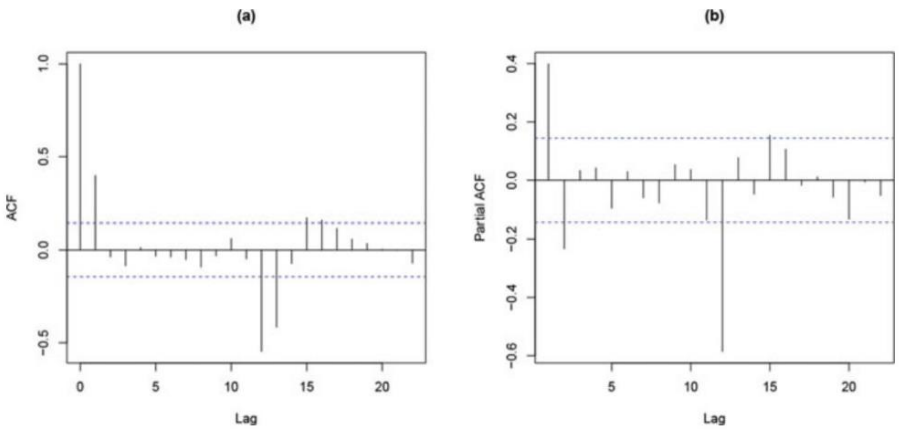


Fig. 4: Funzioni di autocorrelazione e autocorrelazione parziale

Il modello ottimale selezionato tramite i criteri AIC e BIC è stato "GN-ARMA(2,0)S(1,1) dove(2,0) sono i due termini autoregressivi e nessun termine di media mobile, S(1,1) significa un componente stagionale autoregressive e uno di media mobile. I risultati dei criteri AIC e BIC hanno mostrato i valori più bassi del modello scelto, indicando il miglior compromesso tra complessità e qualità di adattamento.

Modello		AIC	BIC
GN-ARMA(2,0)S(1,1)		**182.04**	**200.96**
GN-ARMA(2,0)S(0,1)		188.27	204.03
ARMA(2,0)S(1,1)		193.77	209.54
ARMA(2,0)S(0,1)		200.26	212.87

Tab. 4: I modelli testati con i valori AIC/BIC

Come si vede nella Tab.5 tutti i parametri stimati erano statisticamente significativi al livello del 5%.Una caratteristica spesso osservata nei dati economici.

Parametro	Stima	Intervallo di Confidenza (95%)
ϕ_1	0.4430	(0.2944,0.5915)
ϕ_2	-0.2240	$(-0.3488, -0.0993)$
ϕ_{12} (stagionale)	-0.2370	$(-0.2980, -0.1761)$
θ_{12} (stagionale)	-0.7289	$(-0.8250, -0.6327)$
σ (scarto tipo)	0.3631	(0.2196,0.5067)
s (forma)	1.1233	(0.7795,1.4671)

Tab. 5:Risultati del modello GN-ARMA(2,0)S(1,1)

Il modello **GN-ARMA(2,0)S(1,1)** ha catturato sia le dipendenze di breve termine sia le dinamiche stagionali.Per quanto riguarda la robustezza del Modello GN-ARMA rispetto all'ARMA si nota che il parametro s, stimato a **1.1233**, è significativamente diverso da 2, confermando che i dati presentano code più pesanti rispetto alla distribuzione gaussiana.Una caratteristica spesso osservata nei dati economici.La distribuzione normale generalizzata (GN) ha quindi migliorato la capacità del modello di adattarsi alle caratteristiche

specifiche della serie temporale. Questo approccio è particolarmente utile per serie economiche come l'HICP, dove previsioni accurate sono fondamentali per la politica monetaria e la pianificazione economica. . Le metriche di errore previsivo hanno quantificato la qualità delle previsioni con **MAPE** = 12.22% , **MSE** = 0.03 e **MAE** = 0.12.Le previsioni a 12 mesi sono state calcolate in questa analisi per la serie originale utilizzando la relazione $Y_t(1) = \mu_t + Y_{t-1}$, dove μ_t è la previsione per la serie differenziata.I valori previsti si sono avvicinati bene ai valori osservati, sebbene ci siano state alcune discrepanze nei mesi di maggio, novembre e dicembre come si vede nel grafico in figura 5.

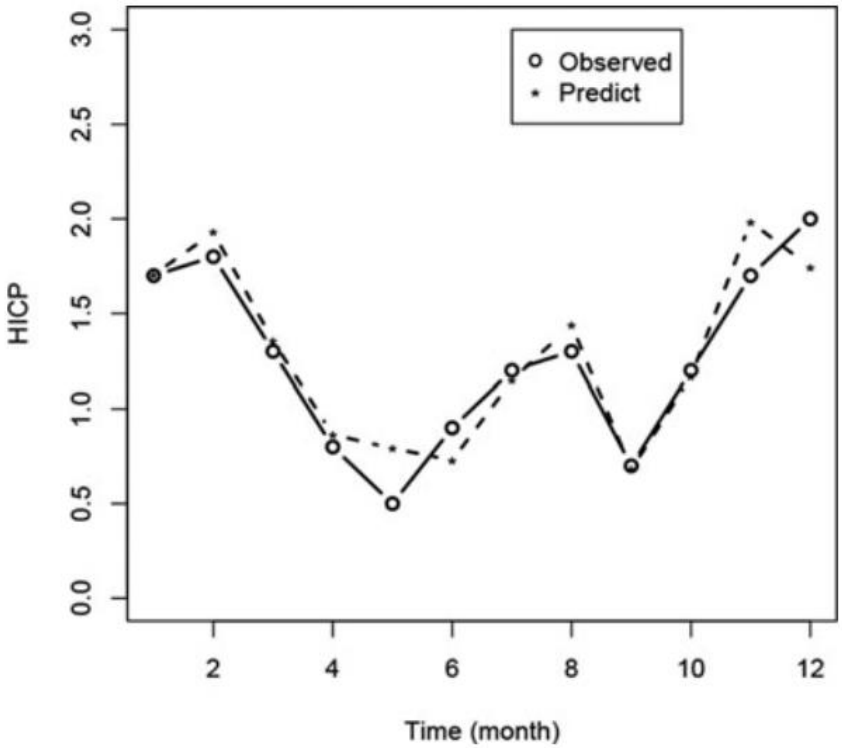


Fig. 5: Previsioni per la serie HICP

5.2 Flusso Idrico (Furnas)

La serie temporale analizzata rappresenta la portata media mensile in m^3 del flusso idrico affluente alla centrale idroelettrica di Furnas, con osservazioni registrate da gennaio 1931 a dicembre 2012. L'analisi si concentra sull'identificazione di modelli che possano descrivere e prevedere la dinamica del flusso idrico. Questa serie temporale presenta caratteristiche peculiari che la rendono un caso di studio interessante infatti la portata media segue un ciclo stagionale legato alle precipitazioni e alla gestione del bacino idrografico. La serie mostra variazioni significative tra mesi e anni, con picchi e minimi influenzati da fattori climatici e antropici. I dati presentano code più pesanti rispetto alla distribuzione normale, suggerendo la necessità di modelli più flessibili come il GN-ARMA. L'analisi delle serie temporali idrologiche è un campo di studio fondamentale per la gestione e la pianificazione delle risorse idriche, in particolare per il funzionamento delle centrali idroelettriche. La previsione della portata media mensile di un corso d'acqua affluente è essenziale per ottimizzare la produzione di energia elettrica, mitigare il rischio di siccità o inondazioni e migliorare la gestione dell'infrastruttura.

L'analisi preliminare dei dati ha evidenziato la necessità di destagionalizzare e normalizzare la serie prima di procedere alla modellazione. La serie è stata destagionalizzata mediante la seguente trasformazione:

$$Z_t = \frac{Y_t(r,m) - \mu_m}{\sigma_m}$$

Dove $Y_t(r,m)$ è il valore della portata media osservata nel mese m dell'anno r,μ_m e σ_m rappresentano la media e la deviazione standard mensili calcolate per ogni mese m sull'intera serie. Le funzioni di autocorrelazione (ACF) e autocorrelazione parziale (PACF) della serie destagionalizzata hanno indicato una correlazione nei primi lag, suggerendo modelli ARMA(1,1) e ARMA(3,0) come potenziali candidati. I grafici nella figura 6 hanno mostrato queste correlazioni significative, suggerendo la necessità di un modello autoregressivo (AR) di ordine basso. Una stagionalità evidente è stata rilevata attorno ai lag corrispondenti a 12 mesi, indicando la necessità di includere componenti stagionali nel modello. Per garantire la stazionarietà della serie, è stata applicata la differenziazione dei dati destagionalizzati.

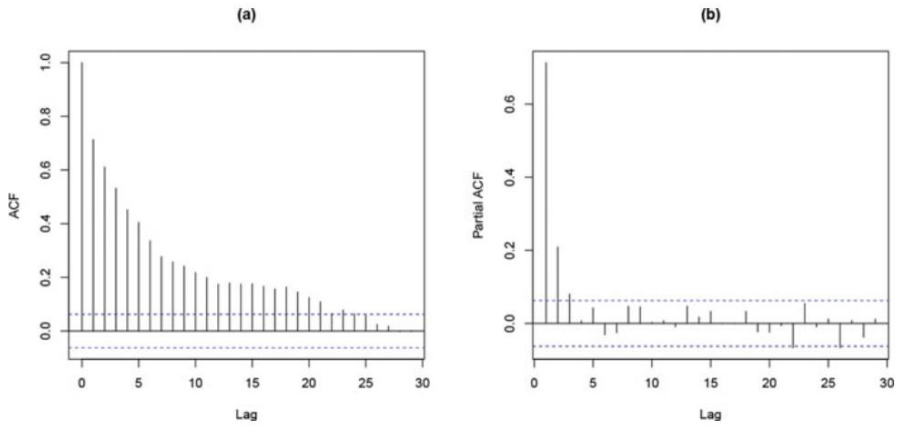


Fig. 6 Le ACF e PACF della serie destagionalizzata

Per scegliere il modello adeguato sono stati confrontati modelli ARMA gaussiani e GN-ARMA.ll modello **GN-ARMA(1,1)** è stato scelto perché ha fornito il miglior adattamento ai dati della serie temporale della portata media mensile del corso d'acqua affluente alla centrale idroelettrica di Furnas**.** Questa scelta è stata basata su diversi criteri di selezione e analisi statistica. I risultati per i criteri di selezione (AIC e BIC) sono riportati nella tabella seguente.

Modello		AIC	BIC
GN-ARMA(3,0)		1926.48	1950.88
GN-ARMA(4,0)		1931.84	1961.11
GN-ARMA(1,1)		**1921.33**	**1940.85**
ARMA(3,0)		1994.71	2014.23
ARMA(2,0)		1999.94	2014.58
ARMA(1,1)		2007.04	2021.67

I risultati dei criteri AIC **e** BIC hanno mostrato che il modello **GN-ARMA(1,1)** aveva i valori più bassi, indicando che è il modello più parsimonioso e accurato.La stima dei parametri del modello GN-ARMA(1,1) è stata effettuata tramite il metodo della massima verosimiglianza. I risultati sono riportati nella tabella 6 sottostante.

Parametro	Stima ($\hat{\beta}$)	Intervallo di Confidenza (95%)
ϕ_1	0.877	(0.830, 0.924)
θ_1	-0.358	(-0.456, -0.261)
σ	0.625	(0.543, 0.708)
s	1.212	(1.078, 1.346)

Tab. 6 Valori di parametri stimati

Tutti i parametri sono risultati significativi al livello del 95%, con intervalli di confidenza che escludono lo zero. Il valore di s è 1.212, inferiore a 2, confermando che la distribuzione dei residui ha code più pesanti rispetto alla distribuzione normale. Per valutare la bontà dell'adattamento del modello GN-ARMA(1,1), sono stati analizzati i residui i quali non seguono una distribuzione normale, confermando che l'ipotesi di normalità dell'ARMA classico non è adeguata. Non hanno notato nessuna autocorrelazione significativa, indicando che il modello ha catturato bene la struttura della serie.

Le previsioni sono state calcolate utilizzando il modello GN-ARMA(1,1) per la serie destagionalizzata e successivamente trasformate nella scala originale della serie. La formula per la previsione è:

$$Y_t(r,m)(1) = \mu_{m+1} + \sigma_{m+1}\hat{Z}_{t+1}$$

Le previsioni per i successivi 12 mesi sono state confrontate con i valori osservati. I risultati sono visualizzati nel grafico sottostante e riassunti in termini di metriche di errore dove **MAPE** =31.4%, **MSE** = 134,174 , **MAE** = 254.2. . Le previsioni si discostano dal valore reale del 31.4% e il modello scelto si adatta bene ai dati.

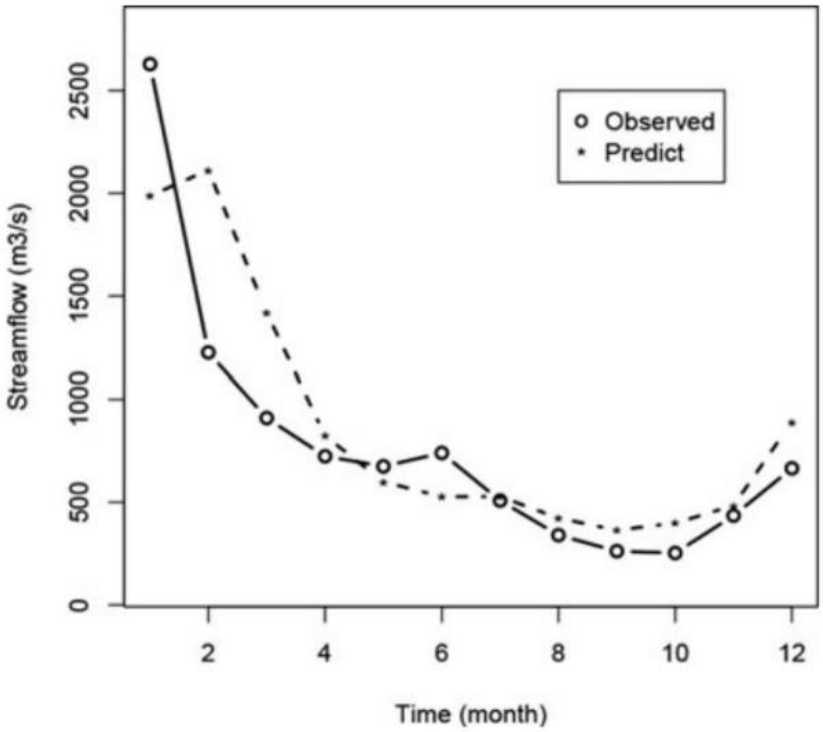


Fig. 5 Previsione della portata media mensile del corso d'acqua a Furnas

5.3 Trasporto Pubblico (Iowa City)

L'applicazione relativa al trasporto pubblico di Iowa City riguarda l'analisi e la previsione della domanda di trasporto pubblico. L'obiettivo dello studio è fornire un modello predittivo accurato che possa essere utilizzato per supportare le decisioni nel settore dei trasporti, ad esempio nella pianificazione delle rotte o nella gestione delle risorse. La serie temporale rappresenta le medie mensili del numero di passeggeri nei sistemi di trasporto pubblico di Iowa City, Stati Uniti, con dati osservati da settembre 1971 a dicembre 1982.La serie presenta stagionalità e trend, il che significa che vi sono fluttuazioni periodiche legate ai mesi dell'anno e una tendenza generale (crescente o decrescente) nel tempo. Il numero di passeggeri segue un andamento ciclico tipico dei sistemi di trasporto pubblico, dove si osservano picchi in determinati mesi dell'anno. Per rendere i dati adatti alla modellizzazione, sono stati applicati due passaggi fondamentali. Innanzitutto la serie è stata destagionalizzata per permettere di rimuovere l'effetto della stagionalità e ottenere una serie più stabile. Infine è stato applicato la differenziazione per rimuovere il trend cioè rimuovere la componente di lungo periodo, lasciando solo le variazioni locali.

Dopo la trasformazione dei dati, si analizzano funzioni di autocorrelazione (ACF) e autocorrelazione parziale (PACF) per determinare l'ordine del modello ARMA più adatto.L'ACF mostra un picco nel primo lag cosi come la PACF suggerendo che un modello ARMA(1,0) o ARMA(0,1) possa essere appropriato come si vede in figura 8.

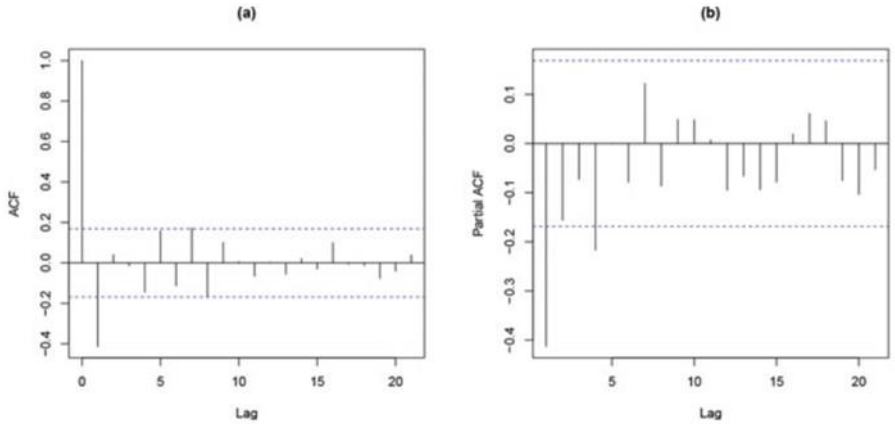


Fig. 8 Funzioni di autocorrelazione e autocorrelazione parziale

Basandosi sull'analisi ACF/PACF e su criteri di selezione come AIC e BIC sono stati testati diversi modelli.I criteri di selezione per i modelli testati sono riportati nella tabella 7 seguente:

Modello		AIC	BIC
GN-ARMA(1,0)		-25.44	-17.01
GN-ARMA(0,1)		**-30.19**	**-21.75**
GN-ARMA(1,1)		-28.19	-16.94
ARMA(1,0)		-22.62	-17.00
ARMA(0,1)		-27.15	-21.53
ARMA(1,1)		-25.16	-16.67

Tab. 7: I modelli testati con i valori AIC/BIC

I risultati dei criteri di selezione hanno indicato che il GN-ARMA(0,1) è il modello migliore per descrivere i dati, poiché ha fornito i valori di AIC e BIC più bassi.I parametri stimati del modello GN-ARMA(0,1) stimati tramite il metodo di massima verosimiglianza. sono riportati nella tabella 8 seguente:

Parametro	Stima ($\hat{\beta}$)	Intervallo di Confidenza (95%)
θ_1	-0.516	(-0.668, -0.363)
σ	0.216	(0.129, 0.302)
s	1.299	(0.806, 1.792)

Tab. 8 Valori di parametri stimati

Il valore di s è inferiore a 2, indicando che la distribuzione dei residui ha code più pesanti della normale (più simile a una distribuzione di Laplace).Il parametro θ_1 è significativo, confermando la necessità della componente MA(1). Una volta stimato il modello, sono state generate previsioni per i 12 mesi successivi. Le previsioni per la serie destagionalizzata e detrendizzata sono state calcolate utilizzando:

$$Z_{t+1} = \mu + \theta_1 \varepsilon_t$$

dove ε_t è il residuo stimato al tempo t.Le previsioni nella scala originale sono state ottenute con:

$$Y_t(1) = \mu_m + \sigma_m Z_{t+1}.$$

Le previsioni sono state visivamente confrontate con i valori reali e mostrano un buon adattamento. Come si vede in figura 7 alcuni scostamenti sono presenti nei mesi di punta, probabilmente a causa di eventi anomali o politiche di trasporto pubblico che hanno influenzato la domanda. In termini di metriche di errore hanno ottenuto MSE = 1,37 , MAE = 1.01 , MAPE del 11.47%. Ciò indica che, in media, le previsioni si discostano dal valore reale del 11.47%. Questo è un buon risultato per un modello di serie storiche. MSE e MAE bassi indicano una buona capacità del modello di adattarsi ai dati.

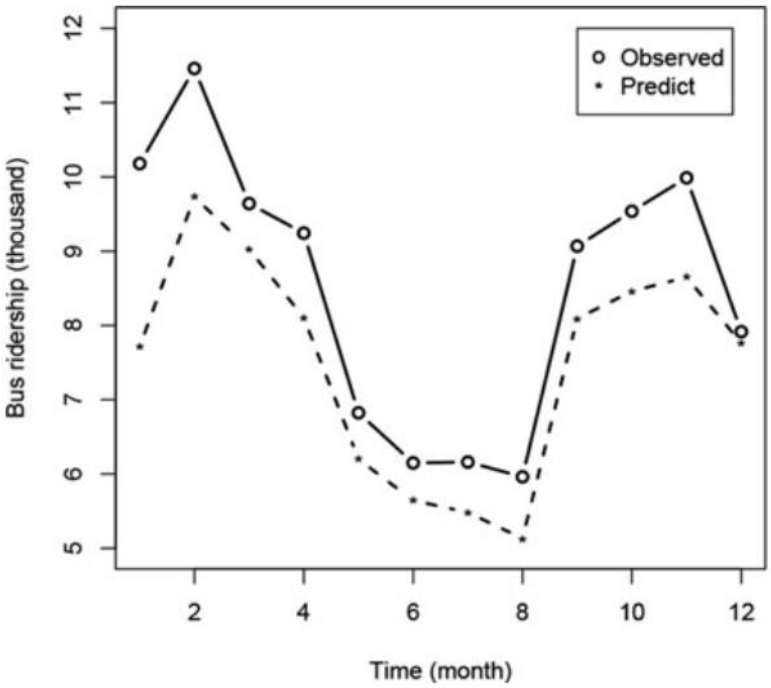


Fig. 7 Previsioni per le serie di utenti degli autobus

Importante per il miglioramento del trasporto pubblico, la pianificazione di nuove rotte e la gestione della domanda stagionale. La flessibilità del GN-ARMA nel trattare variazioni estreme può supportare decisioni in contesti di eventi speciali o cambiamenti demografici. Il modello GN-ARMA(0,1) ha fornito una buona rappresentazione della dinamica della serie sul trasporto pubblico di Iowa City, dimostrando che la distribuzione normale generalizzata migliora la modellazione rispetto ai modelli ARMA gaussiani.

5.4 Impatto del Modello GN-ARMA

L'applicazione del modello GN-ARMA sui tre dataset reali ha avuto un impatto significativo in termini di miglioramento dell'accuratezza e della capacità di modellare fenomeni complessi rispetto ai modelli Gaussiani ARMA. Nella Serie HICP (Indice Armonizzato dei Prezzi al Consumo) il Modello GN-ARMA Ha permesso di catturare la distribuzione dei dati con code più pesanti rispetto alla distribuzione normale ($s = 1.123$).I criteri AIC e BIC hanno mostrato un netto vantaggio del GN-ARMA rispetto all'ARMA, indicando un migliore adattamento del modello. Le previsioni sono risultate accurate con un errore percentuale assoluto medio (MAPE) pari a 12.22%, dimostrando che il modello gestisce bene le fluttuazioni economiche. Quindi Il GN-ARMA ha migliorato la capacità di rappresentare la volatilità economica, un aspetto cruciale per previsioni accurate di fenomeni come l'inflazione. Sul flusso mensile della Centrale di Furnas il nuovo modello ha elaborato con successo la serie temporale con forti componenti stagionali e trend non gaussiani ($s = 1.212$). Rispetto all'ARMA, il GN-ARMA ha prodotto previsioni più precise con metriche dell'errore inferiori. La serie ha mostrato una maggiore variabilità rispetto alla normale standard. Per concludere si è fornito uno strumento robusto per modellare flussi idrologici stagionali, migliorando l'analisi e la previsione dei flussi idrici per scopi di gestione e pianificazione.

Nell'ultimo dataset sui Passeggeri Autobus di Iowa City il Modello GN-ARMA Ha catturato dinamiche stagionali e di trend, con code più pesanti nei dati ($s = 1.299$).Il modello ha superato l'ARMA nei criteri AIC e BIC, dimostrando una migliore rappresentazione della struttura dei dati.Le previsioni hanno mostrato un errore MAPE molto basso, dimostrando l'adeguatezza del GN-ARMA per i dati di trasporto. Il GN-ARMA ha migliorato la precisione delle previsioni per la pianificazione dei trasporti pubblici, supportando decisioni basate su analisi più accurate. Come si vede nella tabella 9.

Dataset	Modello	s	AIC	BIC	MAPE	MSE	MAE
HICP Inflazione...	GN-ARMA(2,0)S(1,1)	1.123	182.04	200.96	12.22%	0.03	0.12
Flusso Idrico (Furnas)	GN-ARMA(1,1)	1.212	1921.33	1940.85	31.4%	134,2	254.2
Trasporto Pubblico (Iowa)	GN-ARMA(0,1)	1.299	-30.19	-21.75	11.47%	1,37	1.01

Tab. 9 Riassunto risultati dell'applicazione sui 3 dataset reali.

Conclusione e sviluppi futuri

Il modello GN-ARMA si è dimostrato superiore ai modelli Gaussiani ARMA tradizionali in tutti e tre i contesti analizzati. Ha migliorato la capacità di catturare fenomeni con distribuzioni non gaussiane e di fornire previsioni più accurate, rendendolo uno strumento ideale per applicazioni in economia, idrologia e politica pubblica. Si è notato che le metriche di errore quadratico medio e di bias assoluto medio sono state influenzate dal valore del parametro s. Gli sviluppi futuri del modello GN-ARMA possono essere le estensioni metodologiche cioè includere variabili esogene nel modello per ampliare il campo di applicazione. Si possono sviluppare pacchetti software per la stima del GN-ARMA, rendendolo più accessibile a ricercatori e professionisti. Ci sono possibilità di definire nuovi contesti applicativi in ambiti come la meteorologia, l'energia e le scienze sociali. Infine effettuare studi approfonditi su dataset di grandi dimensioni per confermare l'efficacia del modello su scala più ampia. In conclusione, il GN-ARMA rappresenta un passo avanti significativo nell'analisi delle serie temporali, offrendo strumenti potenti per affrontare le sfide analitiche in un mondo sempre più complesso e interconnesso.

Bibliografia

Eder Milani, Marinho Andrade & Carlos Diniz (2017) Generalized normal ARMA model applied to the areas of economy, hydrology, and public policy, Communications in Statistics - Simulation and Computation, 46:7, 5819-5835

Jose, K. K., Tomy, L., Sreekumar, J. (2008). Autoregressive processes with normal-Laplace marginals. *Statistics and Probability Letters* 78:2456–2462.

Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics* 32:685–694.

Benjamin, M. A., Rigby, R. A., Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98:214–223.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments. *Scandinavian Journal of Statistics, Series B* 20(2):215–242.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50(4):987–1007.

Nayak, P., Sudheer, K., Rangan, D. M., Ramasastri, K. (2004). A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology* 291:52–66.

Tomy, L. and Jose, K. K. (2009). Generalized normal-Laplace AR process. *Statistics and Probability Letters* 79:1615–1620.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31:307–327.

Kyte, M., Stoner, J., Cryer, J. (1988). Un'analisi di serie temporali del numero di passeggeri del trasporto pubblico a Portland, Oregon, 1971-1982. Transportation Research Part A 22:345-359

Bos, C. S., Franses, P. H., Ooms, M. (2002). Inflation, forecast intervals and long memory regression models. *International Journal of Forecasting* 18:243–264.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Hubrich, K. (2005). Forecasting euro area inflation: Does aggregating forecasts by Hicp component improve forecast accuracy? *International Journal of Forecasting* 21:119–136.