

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per l'Economia e l'Impresa



RELAZIONE FINALE

Verosimiglianza empirica e processi PoINAR(1)

Relatrice Prof.ssa Luisa Bisaglia
Dipartimento di Scienze Statistiche

Laureanda: Daria Capasso
Matricola N 1173018

Anno Accademico 2021/2022

Indice

1	Introduzione	2
2	Processi INAR(1)	4
2.1	Definizione del modello	4
2.2	Momenti del modello INAR(1)	6
2.3	Distribuzione del termine d'innovazione	8
3	Processi PoINAR(1)	9
3.1	Digressione sulla distribuzione di Poisson	9
3.2	Definizione e momenti del modello PoINAR(1)	10
4	Verosimiglianza empirica	16
4.1	Digressione sulla verosimiglianza	16
4.2	Definizione di verosimiglianza empirica e quantità annesse . .	17
4.3	Verosimiglianza empirica per processi PoINAR(1)	21
5	Altri metodi di stima per processi PoINAR(1)	26
5.1	Minimi quadrati condizionali (CLS)	26
5.2	Modified quasi-likelihood (MQL)	28
5.3	Massima verosimiglianza condizionata (CML)	29
6	Studio di simulazione	31
7	Conclusioni	36
8	Appendice	37
	Bibliografia	39

1 Introduzione

L'analisi delle serie storiche è un ambito della statistica da tempo studiato, in quanto si ha da sempre l'interesse ad analizzare l'andamento dei fenomeni nel tempo. Le serie storiche sono raccolte in diversi campi, come l'economia, la meteorologia, l'informatica, l'epidemiologia, la finanza e in tutti quegli ambiti in cui è possibile raccogliere dati nel tempo per studiarne gli andamenti e ricavarne informazioni utili.

Una serie storica temporale è una successione di osservazioni ordinate sequenzialmente nel tempo e deriva da un processo stocastico X_t generatore dei dati. In altre parole, è possibile dire che la serie storica è una realizzazione finita di un processo stocastico. In letteratura sono presenti molti articoli che trattano serie storiche in cui le variabili casuali facenti parte del processo stocastico generatore dei dati sono distribuite in modo continuo; ovvero, dove i possibili risultati del processo sono numeri reali (X_t ha intervallo \mathbb{R} , dove \mathbb{R} è l'insieme dei numeri reali). Tuttavia, in molte applicazioni l'assunzione di un intervallo di valori continui non è adeguata poiché, per esempio, i dati raccolti sono dati di conteggio, come il numero di persone presenti in un determinato luogo o il numero di eventi che si verificano in uno specifico spazio temporale. Per questo motivo, da decenni, in letteratura si studiano le serie storiche di dati di conteggio.

Se la realizzazione di una variabile aleatoria X nasce da un conteggio, allora viene chiamata variabile aleatoria di conteggio ed è una variabile avente come supporto l'insieme discreto di valori interi non negativi.

Serie storiche di questo tipo hanno bisogno di modelli ad hoc per essere modellate, che tengano conto della caratteristica di conteggio delle osservazioni. Per casi di questo genere vengono introdotti in letteratura i modelli INAR(p), acronimo di *INteger-valued AutoRegressive*, da studiosi come Al-Osh e Al-

zaid (1987). Questo lavoro prende spunto dall'articolo "*A new estimation for INAR(1) process with Poisson distribution*" (2021), dove gli autori Lu e Wang propongono un nuovo stimatore basato sulla verosimiglianza empirica per i processi PoINAR(1), dove Po sta per Poisson.

L'obiettivo di questa tesi è quello di riportare ed approfondire con gli studi presenti in letteratura gli argomenti trattati nell'articolo, in particolare i processi PoINAR(1) e la verosimiglianza empirica, e descrivere gli stimatori utilizzati per stimare i parametri del modello PoINAR(1), confrontandoli e valutandone le performance.

La tesi è strutturata come segue. Il capitolo 2 tratta i processi INAR(1) definendo il modello e le relative caratteristiche; il capitolo 3 presenta un approfondimento sui processi PoINAR(1) e al suo interno viene replicata una piccola simulazione di tale processo utile per avere una visione grafica del modello; il capitolo 4 introduce ed approfondisce la verosimiglianza empirica e nel paragrafo 4.3 viene illustrato il nuovo stimatore proposto, denominato MELE; il capitolo 5 si occupa invece degli altri metodi di stima utilizzati per i processi PoINAR(1); infine, il capitolo 6 propone lo studio di simulazione dove si valutano e si confrontano le performance degli stimatori descritti nell'articolo.

2 Processi INAR(1)

2.1 Definizione del modello

I processi INAR(1), acronimo di “first-order INteger-valued AutoRegressive”, vengono utilizzati per la modellazione di serie temporali per dati di conteggio, cioè a valori interi non negativi. Ci si avvale di essi in molti ambiti come la medicina, l’epidemiologia, l’economia, la statistica attuariale e, in tutti quei fenomeni di studio dove le variabili di interesse rilevate sono a valori interi; il supporto di tali variabili, quindi, non è continuo, bensì discreto.

Prima di introdurre il modello INAR(1) è necessario definire l’operatore di *thinning* “ \circ ”, esposto da Steutel e Van Harn (1979), e Van Harn (1978, p.85). Sia X una variabile casuale a valori interi non negativi; allora per qualsiasi $\phi \in [0,1]$, l’operatore “ \circ ” è definito da:

$$\phi \circ X = \sum_{i=1}^X Y_i$$

dove Y_i è una sequenza di variabili casuali i.i.d. (indipendenti e identicamente distribuite) che seguono una distribuzione di Bernoulli, indipendente da X , tali che:

$$P(Y_i = 1) = 1 - P(Y_i = 0) = \phi$$

Ciò significa che $\phi \circ X$ conta il numero di successi ottenuti in un numero casuale (X) di esperimenti Bernoulliani, dove ϕ , cioè la probabilità di successo, rimane costante nel corso dell’esperimento.

Dalla definizione dell'operatore risultano chiare le seguenti proprietà:

- $0 \circ X = 0$
- $1 \circ X = X$
- $E[\phi \circ X] = \phi E[X]$
- $\beta \circ (\phi \circ X) \stackrel{d}{\rightarrow} (\beta\phi) \circ X \quad \forall \beta \in [0,1]$

Il modello autoregressivo a valori interi di primo ordine (INAR(1)) è definito come segue:

$$X_t = \phi \circ X_{t-1} + \epsilon_t \quad (2.1)$$

dove $\phi \in [0,1]$ e ϵ_t è una sequenza di variabili casuali i.i.d. a valori interi non negativi con media μ_ϵ e varianza σ_ϵ^2 .

Per il modello appena definito le componenti del processo X_t al tempo t sono determinate dagli elementi del processo al tempo precedente X_{t-1} , con probabilità di "sopravvivenza" ϕ , e dagli elementi che entrano nel processo nell'intervallo $(t-1,t]$ come termini di innovazione (ϵ_t).

Esistono situazioni in cui il parametro ϕ nel modello (2.1) può variare nel tempo ed essere casuale. Zheng et al. (2007) ne riportano un esempio per capire l'importanza di questo aspetto: supponendo che X_t indichi il numero di pazienti terminali in un determinato mese, il fenomeno X_t potrebbe essere trattato con modelli INAR, dove X_{t-1} sarebbero i pazienti "sopravvissuti" del mese precedente, mentre ϵ_t indicherebbero i nuovi pazienti nel mese corrente. Tuttavia, il tasso di sopravvivenza ϕ sarebbe affetto da fattori ambientali da tenere in considerazione, per questo motivo il tasso potrebbe variare nel tempo.

Prendendo in considerazione questo aspetto, gli studiosi hanno introdotto i

modelli RCINAR, che si differenziano dai modelli INAR per la caratteristica di variazione del parametro autoregressivo ϕ .

Il modello RCINAR(1) è definito come segue:

$$X_t = \phi_t \circ X_{t-1} + \epsilon_t \quad (2.2)$$

dove ϕ_t sono realizzazioni di variabili casuali i.i.d., che assumono valori nell'intervallo $[0,1)$.

2.2 Momenti del modello INAR(1)

Avendo chiara la definizione del modello e dell'operatore di thinning, si possono ricavare i momenti del processo (Al-Osh e Alzaid,1987), il quale si assume essere debolmente stazionario (stazionarietà di secondo ordine).

Il valore atteso condizionato alla storia passata è:

$$E[X_t|X_{t-1}] = \phi E(X_{t-1}) + \mu = \phi^t E(X_0) + \mu \sum_{j=0}^{t-1} \phi^j \quad (2.3)$$

mentre la varianza condizionata è:

$$\begin{aligned} Var(X_t|X_{t-1}) &= \phi^2 Var(X_{t-1}) + \phi(1 - \phi)E(X_{t-1}) + \sigma^2 \\ &= \phi^{2t} Var(X_0) + (1 - \phi) \sum_{j=1}^t \phi^{2j-1} E(X_t - j) + \sigma^2 \sum_{j=1}^{\phi^{2(j-1)}} \end{aligned} \quad (2.4)$$

Media e varianza marginali del processo assumono, invece, la seguente forma:

$$E[X_t] = \frac{\mu}{1 - \phi} \quad (2.5)$$

$$Var[X_t] = \frac{\phi\mu + \sigma^2}{1 - \phi^2} \quad (2.6)$$

La stazionarietà di secondo ordine citata sopra impone i seguenti vincoli al processo:

- $E(X_t) = \mu < \infty \quad \forall t$
- $V(X_t) = \sigma^2 < \infty \quad \forall t$
- $Cov(X_t, X_{t-k}) = \gamma_k \quad \forall t, \forall k$

per cui media e varianza sono costanti e finite nel tempo, mentre la covarianza dipende solo dal ritardo k .

Per ricavare la funzione di autocovarianza risulta utile esprimere la distribuzione marginale del modello in termini di ϵ_t (Al-Osh e Alzaid, 1987):

$$X_t \stackrel{d}{=} \sum_{j=0}^{\infty} \phi^j \circ \epsilon_{t-j}$$

La funzione di autocovarianza a ritardo k risulta essere quindi:

$$\begin{aligned} \gamma(k) &= cov(X_{t-k}, X_t) \\ &= cov(X_{t-k}, \phi^k \circ X_{t-k}) + cov(X_{t-k}, \sum_{j=0}^{k-1} \phi^j \circ \epsilon_{t-k}) \\ &= \phi^k Var(X_{t-k}) + \sum_{j=0}^{k-1} \phi^j cov(X_{t-k}, \epsilon_{t-j}) \\ &= \phi^k \gamma(0) \end{aligned} \tag{2.7}$$

Infine, la funzione di autocorrelazione ρ_k , che per definizione è data da $\frac{\gamma(k)}{\gamma(0)}$, è:

$$\rho_k = \phi^k \tag{2.8}$$

ciò significa che essa decresce a zero in modo esponenziale al crescere del ritardo k . Lo stesso si verifica per i processi AR(1), tuttavia per i processi INAR(1) l'autocorrelazione risulta essere sempre positiva.

2.3 Distribuzione del termine d'innovazione

Da quanto visto fino ad ora il modello INAR(1) (2.1) prevede che la sequenza dei termini d'innovazione ϵ_t , anche detti termini d'errore, sia una sequenza di variabili casuali indipendenti e identicamente distribuite (i.i.d.) di media μ_ϵ e varianza σ_ϵ^2 ; non ne viene specificata, quindi, la distribuzione.

Per studiare questa classe di processi è necessario fare assunzioni sulla distribuzione dei termini d'innovazione che deve tener conto della tipologia dei dati trattati da tali modelli. Inoltre, generalmente il comportamento di dispersione delle osservazioni è facilmente controllabile da quello delle innovazioni. Il ruolo che riveste la distribuzione del termine d'errore nel determinare la distribuzione di X_t per i processi INAR(1) è simile a quello della distribuzione normale nei processi AR(1).

In letteratura vengono discusse alcune casistiche riguardanti i termini d'innovazione per trattare i modelli INAR(1): la distribuzione Geometrica, la distribuzione Binomiale Negativa e la distribuzione di Poisson sono fra le più utilizzate; sebbene ne esistano altre, queste sono quelle che hanno un maggiore riscontro negli studi.

La distribuzione di Poisson è la più ampiamente utilizzata, in quanto il modello INAR(1) diventa di facile utilizzo se le innovazioni seguono tale distribuzione. Questo delinea la classe di processi PoINAR(1), che sarà oggetto di questo elaborato, per cui nel prossimo capitolo ne saranno riportate le caratteristiche.

3 Processi PoINAR(1)

3.1 Digressione sulla distribuzione di Poisson

Una variabile casuale che ha distribuzione di Poisson è una variabile casuale discreta che può assumere qualsiasi valore intero non negativo. Questo modello probabilistico viene utilizzato in tutte quelle situazioni in cui si effettua un conteggio del numero di occorrenze, anche chiamato numero di successi, che si verificano in un certo dominio di studio, tipicamente limitato nel tempo e/o nello spazio. Alcuni esempi in cui si utilizza questa distribuzione sono: il numero di telefonate giornaliere in un call center, il numero di persone che entrano in un negozio, il numero di interruzioni di energia elettrica in un anno, il numero di pazienti che contraggono una determinata malattia nell'arco di anno, l'ammontare di goal segnati da un giocatore. Come si può notare gli ambiti in cui può essere utilizzata sono molteplici e svariati.

Formalizzando è possibile definire le seguenti condizioni:

- c'è un dominio di studio D , tipicamente limitato nel tempo e/o nello spazio;
- gli eventi accadono casualmente all'interno del dominio;
- il verificarsi di un evento è indipendente dal verificarsi di altri eventi;
- c'è un "tasso" sottostante a cui gli eventi si verificano;
- il tasso non varia nel dominio.

Indicata con Y la variabile casuale che descrive il numero di eventi che si verificano nel dominio di analisi, si dirà che Y ha distribuzione di Poisson di parametro λ , e si indica con:

$$Y \sim \text{Po}(\lambda) \quad ,$$

e la funzione di probabilità di Y è:

$$f(Y) = e^{-\lambda} \frac{\lambda^y}{y!}$$

con supporto $S = \{0,1,2,\dots\}$.

La media e la varianza della distribuzione di Poisson sono rispettivamente:

$$E[Y] = \lambda \text{ e } \text{Var}[Y] = \lambda.$$

Infine, è noto che un modello di Poisson rappresenta il limite a cui tende un modello Binomiale quando n è elevato e p è sufficientemente piccolo (con n e p parametri del modello Binomiale).

3.2 Definizione e momenti del modello PoINAR(1)

I modelli PoINAR(1) sono molto trattati e discussi in letteratura in quanto risultano utili per analizzare serie temporali caratterizzate da osservazioni espresse come piccoli conteggi.

La distribuzione di Poisson dei termini d'innovazione permette che il processo INAR(1) abbia realizzazioni positive, cioè valori interi non negativi, a differenza dei processi AR(1), dove il termine d'errore segue una distribuzione Normale di media 0 e varianza costante e quindi presenta possibili valori negativi.

Nel modello (2.1) si assume che il termine d'innovazione ϵ_t abbia una distribuzione di Poisson di parametro λ , $\epsilon_t \sim \text{Po}(\lambda)$ con $\lambda > 0$.

Supponendo che $\phi < 1$, e che $\{\epsilon_t\}$ sia una sequenza di variabili casuali i.i.d. con distribuzione di Poisson di parametro λ , si ottiene che $\{X_t\}$ è una catena di Markov stazionaria ¹ avente distribuzione marginale di Poisson con media

¹La catena di Markov è un sistema in cui la probabilità di uno stato i al tempo k dipende esclusivamente dal tempo immediatamente precedente $k - 1$.

e varianza:

$$E[X_t] = Var[X_t] = \frac{\lambda}{1 - \phi} = \mu \quad (3.1)$$

Ciò deriva da due importanti proprietà della distribuzione di Poisson:

- l'invarianza rispetto all'operatore di thinning: cioè, se $X \sim Po(\mu)$ allora $\phi \circ X \sim Po(\phi\mu)$,
- l'additività: cioè, se $Z \sim Po(\phi\mu)$ ed $\epsilon \sim Po((1 - \phi)\mu)$, entrambe indipendenti, allora $Z + \epsilon \sim Po(\phi\mu + (1 - \phi)\mu) = Po(\mu)$.

Dalla (3.1) si evince uno dei limiti del modello PoINAR(1) ovvero quello della equidispersione del processo (media pari alla varianza). Tuttavia, anche nei casi in cui la distribuzione del termine d'errore non è di Poisson, bensì è Binomiale o Binomiale Negativa, il modello presenta dei limiti in quanto a dispersione. In particolare, con la Binomiale il processo presenta sottodispersione (media maggiore della varianza) mentre con la Binomiale Negativa si ha una sovradispersione (media minore della varianza).

Per la media e la varianza condizionata si ha invece (Alzaid e Al-Osh, 1988):

$$E[X_t|X_{t-1}] = \phi X_{t-1} + \lambda \quad (3.2)$$

$$Var[X_t|X_{t-1}] = \phi(1 - \phi)X_{t-1} + \lambda \quad (3.3)$$

Una volta conosciute tutte le caratteristiche del processo e, specialmente, la distribuzione, è possibile effettuare una simulazione del processo PoINAR(1). Ne segue un esempio.

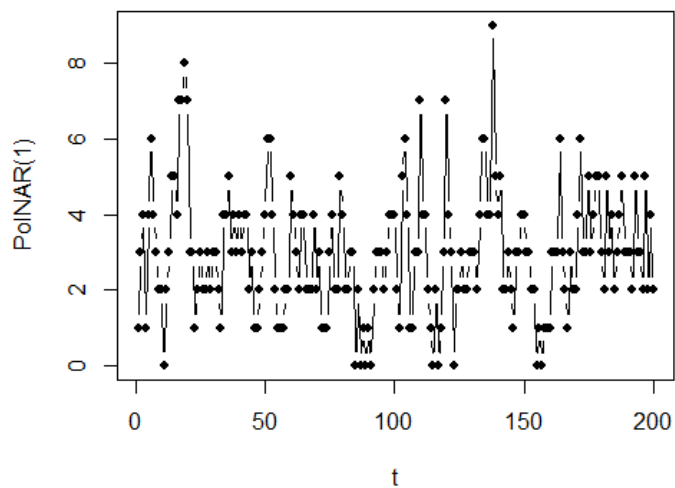


Figura 1: Realizzazione di un processo PoINAR(1) con $\phi = 0.5$ e $\lambda = 1.5$

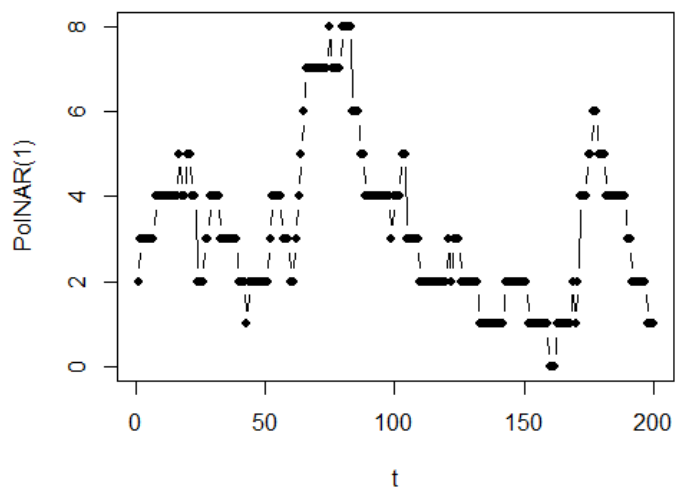


Figura 2: Realizzazione di un processo PoINAR(1) con $\phi = 0.95$ e $\lambda = 0.15$

Nelle figure 1 e 2 sono rappresentate due realizzazioni dei processi PoINAR(1). Per entrambi è fissata una media di $\mu=3$, mentre quello che cambia è il parametro ϕ e quindi anche la media dei termini d'innovazione $\lambda = \mu(1 - \phi)$.

Nella *Figura 1* si ha $\phi = 0.5$ e $\lambda = 1.5$; questo livello moderato del parametro ϕ è visibile dall'andamento oscillante (crescente e decrescente) del modello. Nella *Figura 2* invece, è riportato un caso più estremo dove $\phi = 0.95$ e $\lambda = 0.15$. Il parametro λ così definito implica che l'innovazione generata più raramente è positiva. Inoltre, $\phi = 0.95$ porta $\phi \circ X$ ad essere uguale a X per la maggior parte del tempo o, diversamente, porta ad un comportamento lentamente decrescente.

Se si osserva, infatti, il grafico della funzione di autocorrelazione globale (ACF) si ha che nella *Figura 3*, dove il processo PoINAR(1) ha $\phi = 0.5$ e $\lambda = 1.5$, la funzione decresce più velocemente a zero; mentre, nella *Figura 4*, dove il processo ha $\phi = 0.95$ e $\lambda = 0.15$, la funzione decresce molto più lentamente.

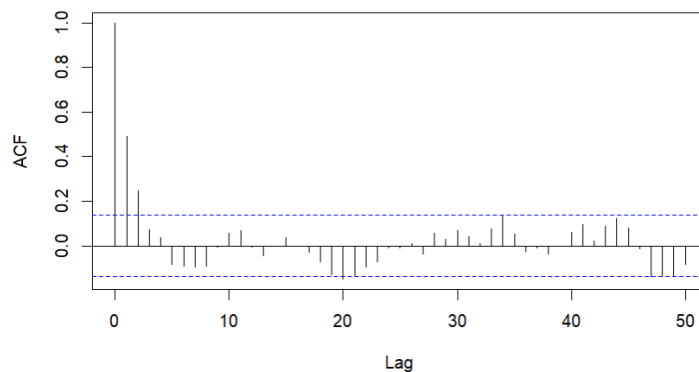


Figura 3: Funzione di autocorrelazione empirica della serie riportata in *Figura 1*

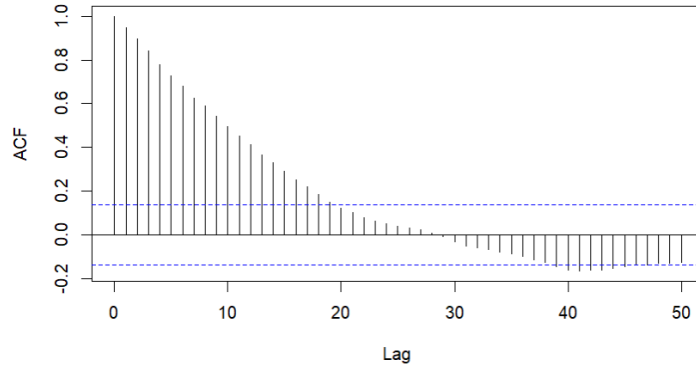


Figura 4: Funzione di autocorrelazione empirica della serie riportata in Figura 2

Di seguito verranno riportati i momenti k -esimi del processo PoINAR(1) descritti da Bourguignon e Vasconcellos (2015).

Il momento di ordine k del processo è dato da:

$$E[X_t^k] = \mu_k = \mu(\mu_{k-1} + \frac{d\mu_{k-1}}{d\mu}) \quad k \geq 1 \quad (3.4)$$

In particolare,

$$E[X_t^2] = \mu_2 = \mu + \mu^2$$

$$E[X_t^3] = \mu_3 = \mu + 3\mu^2 + \mu^3$$

$$E[X_t^4] = \mu_4 = \mu^4 + 6\mu^3 + 7\mu^2 + \mu$$

La stazionarietà del processo PoINAR(1) implica che i momenti congiunti del k -esimo ordine delle variabili casuali $X_t, X_{t+s_1}, X_{t+s_{k-1}}$ esistono e hanno la forma:

$$E[X_t X_{t+s_1} \dots X_{t+s_{k-1}}] = \mu(s_1, \dots, s_{k-1}) \quad (3.5)$$

Le probabilità di transizione del processo sono date da:

$$Pr(X_t = k | X_{t-1} = l) = e^{-\lambda} \sum_{i=0}^{\min(k,l)} \frac{\lambda^{k-i}}{(k-i)!} \binom{l}{i} \phi^i (1-\phi)^{l-i} \quad (3.6)$$

Nella pratica, i veri valori dei parametri ϕ e λ del modello non si conoscono, per cui li si deve stimare dai dati.

Il problema di stima dei modelli PoINAR(1) viene affrontato da Feilong Lu e Dehui Wang (2021) con un approccio diverso, proponendo un nuovo stimatore basato sulla verosimiglianza empirica. Nel prossimo capitolo ci sarà un approfondimento sul tema della verosimiglianza empirica e verrà descritto il procedimento attraverso il quale è possibile definire il nuovo stimatore.

4 Verosimiglianza empirica

4.1 Digressione sulla verosimiglianza

La verosimiglianza è uno degli strumenti statistici più conosciuti ed utilizzati per fare inferenza. La procedura basata sulla verosimiglianza mira a trattare modelli statistici parametrici ed è caratterizzata da una funzione chiamata funzione di verosimiglianza.

Essa fornisce una misura di plausibilità di ciascun valore del parametro alla luce del campione osservato, ovvero fornisce la probabilità che esisteva a priori di osservare il campione che è stato effettivamente osservato.

Formalizzando, si definiscono le seguenti ipotesi:

- il fenomeno d'interesse viene osservato in riferimento ad un campione y_1, y_2, \dots, y_n di osservazioni, le quali sono realizzazioni di una variabile casuale Y con supporto S ;
- si suppone che Y segua una determinata distribuzione, per cui si specifica l'insieme delle possibili alternative di distribuzione. Questo insieme è chiamato modello statistico e si indica con \mathbf{F} ;
- si utilizza il campione osservato per fare inferenza sulla distribuzione.

L'inferenza sarà migliore e più accurata se si delimita \mathbf{F} , il quale sarà definito basandosi sulla natura del fenomeno che ha generato y e su altre informazioni ausiliarie. In funzione delle caratteristiche di \mathbf{F} si determina la natura del modello statistico, parametrico o non parametrico, e le procedure conseguenti per fare inferenza. L'inferenza basata sulla verosimiglianza richiede la specificazione del modello statistico, permettendo così di ottenere la stima di massima verosimiglianza. Quest'ultima è il risultato di massimizzazione della funzione di verosimiglianza.

Indicando con y_{oss} i dati osservati e con \mathbf{F} il modello statistico per y_{oss} , definito come

$$\mathbf{F} = \{Pr_Y(\cdot; \theta), \theta \in \Theta\},$$

è detta funzione di verosimiglianza per θ , basata su y_{oss} la funzione L che si ottiene calcolando Pr_Y in y_{oss} al variare di θ :

$$L(\theta) = L(\theta; y_{oss}) = Pr_Y(y_{oss}; \theta).$$

La funzione di verosimiglianza fornisce procedure per la stima puntuale ed intervallare e per la verifica d'ipotesi. Inoltre, gli stimatori di massima verosimiglianza godono di ottime proprietà formali, come la non distorsione asintotica, la consistenza, l'efficienza asintotica e l'asintoticità in distribuzione alla Normale.

4.2 Definizione di verosimiglianza empirica e quantità annesse

Il metodo della verosimiglianza empirica è stato proposto per la prima volta da Owen (1988) e ulteriormente studiato da egli stesso e da altri ricercatori negli anni a seguire.

L'approccio basato sulla verosimiglianza empirica è un metodo di inferenza non parametrico basato su una funzione del rapporto di verosimiglianza guidata dai dati. I vantaggi di questo approccio sono molteplici: in primo luogo, permette di utilizzare metodi di verosimiglianza senza specificare la famiglia di distribuzione dei dati e, in secondo luogo, le regioni di confidenza basate sulla verosimiglianza empirica non presentano una simmetria predeterminata, in modo tale da avere una corrispondenza migliore con la vera forma della distribuzione sottostante.

La verosimiglianza empirica incorpora direttamente informazioni collaterali

esprese attraverso vincoli o distribuzioni a priori. Inoltre, possiede proprietà asintotiche molto potenti.

La conoscenza che deriva al di fuori dei dati può essere quindi incorporata in essa e assumere la forma di vincoli che limitano il dominio della funzione di verosimiglianza.

I metodi basati sulla verosimiglianza empirica risultano essere molto utili in quanto combinano l'affidabilità dei metodi non parametrici con la flessibilità e l'efficacia dell'approccio di verosimiglianza. Il motivo per cui si chiama "verosimiglianza empirica" è dovuto al fatto che la distribuzione empirica dei dati gioca un ruolo centrale.

Di seguito, verranno riportati dei risultati ottenuti e pubblicati da Owen nel suo libro intitolato "Empirical Likelihood" (2001). Questi risultati saranno la base degli studi di molti ricercatori, come Feilong Lu e Dehui Wang (2021), i quali hanno definito un nuovo stimatore basato sulla verosimiglianza empirica. Si partirà definendo la funzione di ripartizione empirica per arrivare alla definizione di verosimiglianza empirica e del rapporto di verosimiglianza empirica, utilizzato per test d'ipotesi e intervalli di confidenza.

Data una variabile casuale $X \in \mathbb{R}$, si definisce funzione di ripartizione la funzione $F(x) = \Pr(X \leq x)$ con $-\infty < x < \infty$.

Utilizzando la notazione $F(x-)$ per indicare $\Pr(X < x)$ si ha che $\Pr(X = x) = F(x) - F(x-)$.

Definizione 1

Siano $X_1, \dots, X_n \in \mathbb{R}$, allora la funzione di ripartizione empirica di X_1, \dots, X_n è:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{X_i \leq x} \quad (1)$$

per $-\infty < x < \infty$, dove $I_{A(x)}$ è uguale a 1 se $A(x)$ è vera e 0 altrimenti.

Definizione 2

Dato il campione $X_1, \dots, X_n \in \mathbb{R}$, assunto indipendente e con funzione di ripartizione comune, la verosimiglianza non parametrica della funzione di ripartizione F è:

$$L(F) = \prod_{i=1}^n (F(X_i) - F(X_{i-})) \tag{2}$$

Il valore di $L(F)$ è la probabilità di ottenere esattamente i valori campionari osservati X_1, \dots, X_n dalla funzione di ripartizione F .

Inoltre, esiste un teorema (Owen, 2001) che dimostra che la verosimiglianza non parametrica è massimizzata dalla funzione di ripartizione empirica vista poc'anzi. Quindi, la funzione di ripartizione empirica si dimostra essere la stima di massima verosimiglianza non parametrica.

Tra i metodi basati sulla verosimiglianza parametrica, si utilizza comunemente il rapporto di verosimiglianza; per analogia, esiste il rapporto di verosimiglianza non parametrica, utilizzato anch'esso per la verifica d'ipotesi e il calcolo di intervalli di confidenza. Esso si definisce attraverso la verosimiglianza non parametrica (2) come:

$$R(F) = \frac{L(F)}{L(F_n)} \tag{4.1}$$

Si suppone di avere interesse ad un parametro $\theta = T(F)$ per qualche funzione T di distribuzioni. La F fa parte di un gruppo \mathbf{F} di distribuzioni, spesso circoscritto. Owen (2001) definisce quindi la funzione del rapporto di verosimiglianza profilo:

$$R(\theta) = \sup\{R(F) | T(F) = \theta, F \in \mathbf{F}\} \tag{4.2}$$

Questa funzione è utilizzata per la verifica d'ipotesi. In particolare, si rifiuta l'ipotesi nulla $H_0: T(F_0) = \theta_0$, quando $R(\theta_0) < r_0$, per un certo valore soglia r_0 . Mentre, le regioni di confidenza assumono la forma: $\{\theta \mid R(\theta) \geq r_0\}$.

La verosimiglianza empirica viene utilizzata anche nell'ambito delle serie storiche, dove le variabili casuali, che entrano a far parte del processo, sono ordinate rispetto al tempo, per cui risultano correlate (c'è un legame tra di loro).

Considerando primariamente dati che non hanno legami, se la distribuzione F ha probabilità positive $p_i \geq 0$ con $X_i \in \mathbb{R}$, allora $\sum_{i=1}^n p_i \leq 1$, e $L(F) = \prod_{i=1}^n np_i$, si ha

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{i=1}^n np_i \quad (4.3)$$

Quando i dati non sono indipendenti, l'equazione (4.3) si complica; tuttavia è dimostrato da Owen (2001) che se si usa la (4.3) si ottiene la stessa funzione del rapporto di verosimiglianza profilo $R(\theta)$. Ciò vale per qualsiasi famiglia F di distribuzioni e per qualsiasi funzione $T(F)$ utilizzata per definire θ .

4.3 Verosimiglianza empirica per processi PoINAR(1)

Ritornando ai processi PoINAR(1) e volendo stimare i parametri ignoti del processo, Zhang et al. (2011) mostrano nei loro studi sui processi RCINAR(p) la procedura per ottenere degli stimatori utilizzando il metodo della verosimiglianza empirica. Di seguito andrò a riportarne il contenuto.

Siano X_1, \dots, X_n vettori casuali d-dimensionalmente a valori discreti con funzione di ripartizione comune F e media μ . La verosimiglianza empirica è

$$L(F) = \prod_{t=1}^n dF(X_t) = \prod_{t=1}^n p_t \quad (4.4)$$

e può essere massimizzata dalla funzione di ripartizione empirica (Definizione 1).

Il rapporto di verosimiglianza empirica è definito come

$$R(F) = \frac{L(F)}{L(F_n)} = \prod_{t=1}^n np_t \quad (4.5)$$

Per ottenere la regione di confidenza per la media μ , si definisce la funzione del rapporto di verosimiglianza profilo empirica come:

$$R(\mu) = \max \left\{ \prod_{t=1}^n np_t \mid p_t \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t X_t = \mu \right\} \quad (4.6)$$

Inoltre, Owen (2001) dimostra che la quantità $-2\log R(\mu)$ converge in distribuzione alla distribuzione χ_d^2 .

Successivamente, Zhang et al. (2011) arrivano a specificare le equazioni di stima per il processo RCINAR(p) attraverso la seguente funzione criterio dei minimi quadrati condizionali:

$$S(\beta) = \sum_{t=1}^n \left(X_t - \sum_{i=1}^p \phi X_{t-i} - \lambda \right)^2 \quad (4.7)$$

dove $\beta = (\phi_i, 1 \leq i \leq p, \lambda)'$.

Facendo la derivata di $S(\beta)$ rispetto a β , si ottengono le equazioni di stima

$$-\frac{1}{2} \cdot \frac{\partial S(\beta)}{\partial \beta} = \sum_{t=1}^n m_t = 0 \quad (4.8)$$

in cui

$$\begin{aligned} m_t &= (m_{t1}, m_{t2}, \dots, m_{t(p+1)})' \\ m_{ti} &= (X_t - \sum_{s=1}^p \phi_s X_{t-s} - \lambda) X_{t-i}, \quad 1 \leq i \leq p \\ m_{t(p+1)} &= X_t - \sum_{s=1}^p \phi_s X_{t-s} - \lambda \end{aligned} \quad (4.9)$$

Per processi PoINAR(1), con $\theta = [\phi, \lambda]'$, le equazioni di stima diventano:

$$m_t(\theta) = \begin{pmatrix} m_{1t}(\theta) \\ m_{2t}(\theta) \\ m_{3t}(\theta) \end{pmatrix} = \mathbf{0}$$

dove

$$m_{1t}(\theta) = X_t - \phi X_{t-1} - \lambda \quad (4.10)$$

$$m_{2t}(\theta) = (X_t - \phi X_{t-1} - \lambda) X_{t-1} \quad (4.11)$$

$$m_{3t}(\theta) = (X_t - \phi X_{t-1} - \lambda)^2 - \phi(1 - \phi) X_{t-1} - \lambda \quad (4.12)$$

Dunque per stimare i parametri si procede definendo la funzione del rapporto di verosimiglianza profilo empirica (Owen 1991):

$$R(\theta) = \max \left\{ \prod_{t=1}^n np_t | p_t \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t m_t(\theta) = 0 \right\} \quad (4.13)$$

E' necessario pertanto massimizzare $R(\boldsymbol{\theta})$ tenendo in considerazione i vincoli:

$$p_t \geq 0, \quad \sum_{t=1}^n p_t = 1, \quad \sum_{t=1}^n p_t m_t = 0 \quad (4.14)$$

Per trovare il massimo di questa funzione si utilizzano i moltiplicatori di Lagrange, procedendo come segue.

Si ha

$$H = \sum_{t=1}^n \log(np_t) - n\boldsymbol{\beta}'\left(\sum_{t=1}^n p_t m_t\right) + \gamma\left(1 - \sum_{t=1}^n p_t\right), \quad (4.15)$$

dove $\gamma \in \mathbb{R}$ e $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ sono i moltiplicatori di Lagrange. Ponendo la derivata parziale di H rispetto a p_t uguale a zero, si ha:

$$\frac{\partial H}{\partial p_t} = \frac{1}{p_t} - n\boldsymbol{\beta}'m_t - \gamma = 0 \quad (4.16)$$

da cui si ottiene:

$$p_t = \frac{1}{(\gamma + n\boldsymbol{\beta}'m_t)} \quad t = 1, \dots, n$$

Per trovare γ , invece, si pre-moltiplica l'equazione (4.16) per $\sum_{t=1}^n p_t$:

$$0 = \sum_{t=1}^n p_t \frac{\partial H}{\partial p_t} = n - \gamma \rightarrow \gamma = n$$

Con $\gamma = n$, si esplicita p_t :

$$p_t = \frac{1}{n(1 + \boldsymbol{\beta}'m_t)} \quad t = 1, \dots, n$$

Dal terzo vincolo presente nella (4.14) si ha che:

$$0 = \sum_{t=1}^n p_t m_t(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^n \frac{1}{1 + \boldsymbol{\beta}m_t(\boldsymbol{\theta})} m_t(\boldsymbol{\theta}) \quad (4.17)$$

da cui i β possono essere determinati tramite algoritmi numerici in termini di θ .

Qin e Lawless (1994) dimostrano come ottenere il rapporto di log-verosimiglianza profilo (ELR); di seguito ne riporterò i calcoli.

Poiché è necessario che $0 \leq p_t \leq 1$, si ricava che β e θ devono soddisfare $1 + \beta' m_t(\theta) \geq \frac{1}{n} \quad \forall t$. Per θ fissato, si ha $D_\theta = \{\beta : 1 + \beta' m_t(\theta) \geq \frac{1}{n}\}$; D_θ è convesso e chiuso, ed è limitato se 0 si trova all'interno della convessità.

Inoltre,

$$\frac{\partial}{\partial \beta} \left\{ \frac{1}{n} \sum_{t=1}^n \frac{1}{1 + \beta' m_t(\beta)} m_t(\theta) \right\} = -\frac{1}{n} \sum_{t=1}^n \frac{m_t(\theta) m_t'(\theta)}{\{1 + \beta' m_t(\theta)\}^2} \quad (4.18)$$

è definita negativa per β in D_θ , purché $\sum_{t=1}^n m_t(\theta) m_t'(\theta)$ sia definita positiva. Per il teorema della funzione inversa, $\beta = \beta(\theta)$ è quindi una funzione derivabile continua di θ .

La funzione di verosimiglianza empirica profilo per θ è definibile come:

$$L(\theta) = \prod_{t=1}^n \left\{ \left(\frac{1}{n} \right) \frac{1}{1 + \beta' m_t(\theta)} \right\} \quad (4.19)$$

e la funzione del rapporto di log-verosimiglianza empirica profilo come:

$$l(\theta) = \sum_{t=1}^n \log(1 + \beta' m_t(\theta)) \quad (4.20)$$

La distribuzione asintotica di questa funzione è la distribuzione χ^2 e viene utilizzata per costruire regioni di confidenza per θ . Formalmente si ha quindi che:

Teorema 1

Per $n \rightarrow \infty$, la quantità $2 l(\theta_0) \xrightarrow{d} \chi_2^2$, dove $l(\theta)$ è definita nell'equazione (4.20) ed \xrightarrow{d} indica la convergenza in distribuzione.

La regione di confidenza di livello $1 - \alpha$ è $C_{\alpha,n} = \{\boldsymbol{\theta} | 2l(\boldsymbol{\theta}) \leq \chi_{1-\alpha,2}^2\}$, dove $\chi_{1-\alpha,2}^2$ è il quantile di ordine $1 - \alpha$ della distribuzione Chi-quadro con due gradi di libertà.

Una volta ottenuti questi risultati, si può trovare lo stimatore di massima verosimiglianza empirica (MELE- "*Maximum Empirical Likelihood Estimator*") $\hat{\boldsymbol{\theta}}$ per il parametro $\boldsymbol{\theta}$ minimizzando $l(\boldsymbol{\theta})$ su $\boldsymbol{\theta}$.

Definizione 3

Per lo stimatore MELE $\hat{\boldsymbol{\theta}}$, si ha:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, V) \quad (3)$$

dove $V = [\boldsymbol{\Omega}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}]^{-1}$ con $\boldsymbol{\Omega} = E\left[\frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]$ e $\boldsymbol{\Sigma} = [m_t(\boldsymbol{\theta})m_t(\boldsymbol{\theta})']$.

Nello studio di simulazione fatto da Lu e Wang (2021), che verrà riportato nell'ultimo capitolo, emergono le buone proprietà di questo stimatore.

Inoltre, se non si considera la terza equazione di stima, ovvero la (4.12), e si usano solamente $m_{1t} = 0$ e $m_{2t} = 0$ come equazioni di stima, lo stimatore MELE è coerente con lo stimatore dei minimi quadrati condizionali.

Oltre allo stimatore citato sopra, esistono anche altri stimatori utilizzati per i processi PoINAR(1); nel capitolo seguente ne verranno illustrati alcuni.

5 Altri metodi di stima per processi PoINAR(1)

In questo capitolo vengono brevemente descritti altri stimatori sfruttati per stimare i parametri ignoti del processo PoINAR(1); questi stimatori sono stati descritti da Al-Osh e Alzaid (1987) e da Freeland e McCabe (2004,2005).

5.1 Minimi quadrati condizionali (CLS)

Sia

$$S(\boldsymbol{\theta}) = \sum_{t=1}^n (X_t - \phi X_{t-1} - \lambda)^2, \quad (5.1)$$

dove $\boldsymbol{\theta} = (\phi, \lambda)'$, la funzione criterio dei minimi quadrati condizionali. Gli stimatori dei minimi quadrati condizionali per ϕ e λ sono ottenuti minimizzando S su $\boldsymbol{\theta} \in \{0 \leq \phi \leq 1, \lambda > 0\}$ e, come mostrato da Klimbo e Nelson (1978) e da Al-Osh e Alzaid (1987), hanno la seguente espressione esplicita:

$$\begin{aligned} \hat{\phi} &= \frac{n \sum_{t=1}^n X_t X_{t-1} - (\sum_{t=1}^n X_{t-1})(\sum_{t=1}^n X_t)}{n \sum_{t=1}^n X_{t-1}^2 - (\sum_{t=1}^n X_{t-1})^2} \\ \hat{\lambda} &= n^{-1} \left(\sum_{t=1}^n X_t - \hat{\phi} \sum_{t=1}^n X_{t-1} \right) \end{aligned} \quad (5.2)$$

Questi stimatori risultano consistenti e asintoticamente normali come mostra il seguente teorema (Zheng et al. 2007).

Teorema 2

Per gli stimatori dei minimi quadrati condizionali $(\hat{\phi}, \hat{\lambda})$ si ha

$$\sqrt{n} \begin{pmatrix} \hat{\phi} - \phi \\ \hat{\lambda} - \lambda \end{pmatrix} \xrightarrow{d} N(0, V^{-1} W V^{-1})$$

$$\text{dove } W = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix},$$

$$\sigma_1^2 = E((X_1 - \phi X_0 - \lambda)^2),$$

$$\sigma_2^2 = E(X_0^2(X_1 - \phi X_0 - \lambda)^2),$$

$$\sigma_{12} = E(X_0(X_1 - \phi X_0 - \lambda)^2),$$

$$V^{-1} = (m_2 - m_1)^{-1} \begin{pmatrix} 1 & -m_1 \\ -m_1 & m_2 \end{pmatrix},$$

$$m_2 = m_1^2 + \sigma_x^2, \quad m_j = E(X_1^j), \quad j = 1, 2, \quad \sigma_x^2 = \text{Var}(X_1) = m_2 - m_1^2 > 0,$$

in cui E indica il valore atteso rispetto alla distribuzione stazionaria.

Freeland e McCabe (2005) trovano un'espressione esplicita, per i processi PoINAR(1), della matrice di covarianza della distribuzione asintotica dello stimatore dei minimi quadrati condizionali; indicandola con la notazione j^{-1} , la matrice è la seguente:

$$j^{-1} = \begin{pmatrix} \frac{\phi(1-\phi)^2}{\lambda} + (1+\phi)(1-\phi) & -(1+\phi)\lambda \\ -(1+\phi)\lambda & \lambda + \frac{1+\phi}{1-\phi}\lambda^2 \end{pmatrix}$$

5.2 Modified quasi-likelihood (MQL)

Ricordando che la varianza condizionata è:

$$V_{\boldsymbol{\theta}}(X_t|X_{t-1}) := \text{Var}(X_t|X_{t-1}) = \phi(1 - \phi)X_{t-1} + \lambda \quad (5.3)$$

allora lo stimatore MQL per i parametri del processo è definito come:

$$\begin{pmatrix} \hat{\phi} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^n X_{t-1}^2 V_{\hat{\boldsymbol{\theta}}}^{-1}(X_t|X_{t-1}) & \sum_{t=1}^n X_{t-1} V_{\hat{\boldsymbol{\theta}}}^{-1}(X_t|X_{t-1}) \\ \sum_{t=1}^n X_{t-1} V_{\hat{\boldsymbol{\theta}}}^{-1}(X_t|X_{t-1}) & \sum_{t=1}^n V_{\hat{\boldsymbol{\theta}}}^{-1}(X_t|X_{t-1}) \end{pmatrix}^{-1} \\ \times \begin{pmatrix} \sum_{t=1}^n X_{t-1} X_t V_{\hat{\boldsymbol{\theta}}}^{-1}(X_t|X_{t-1}) \\ \sum_{t=1}^n X_t V_{\hat{\boldsymbol{\theta}}}^{-1}(X_t|X_{t-1}) \end{pmatrix}$$

dove $\hat{\boldsymbol{\theta}}$ è uno stimatore consistente e, nella pratica, si può utilizzare lo stimatore dei minimi quadrati condizionali di $\boldsymbol{\theta}$.

La distribuzione asintotica dello stimatore è data dal seguente enunciato (Zheng et al. 2007).

Teorema 3

La distribuzione asintotica congiunta dello stimatore MQL è:

$$\sqrt{n} \begin{pmatrix} \hat{\phi} - \phi \\ \hat{\lambda} - \lambda \end{pmatrix} \xrightarrow{d} N(\mathbf{0}, T^{-1}(\boldsymbol{\theta})' Q(\boldsymbol{\theta}) T^{-1}(\boldsymbol{\theta}))$$

dove

$$Q(\boldsymbol{\theta}) = \begin{pmatrix} T_1(\boldsymbol{\theta}) & T_3(\boldsymbol{\theta}) \\ T_3(\boldsymbol{\theta}) & T_2(\boldsymbol{\theta}) \end{pmatrix},$$

$$T^{-1}(\boldsymbol{\theta}) = (T_1(\boldsymbol{\theta})T_2(\boldsymbol{\theta}) - T_3^2(\boldsymbol{\theta}))^{-1} \begin{pmatrix} T_1(\boldsymbol{\theta}) & -T_3(\boldsymbol{\theta}) \\ -T_3(\boldsymbol{\theta}) & T_2(\boldsymbol{\theta}) \end{pmatrix},$$

in cui $T_1(\boldsymbol{\theta}) = E[V_{\boldsymbol{\theta}}^{-1}(X_1|X_0)]$, $T_2(\boldsymbol{\theta}) = E[X_0^2 V_{\boldsymbol{\theta}}^{-1}(X_1|X_0)]$ e $T_3(\boldsymbol{\theta}) = E[X_0 V_{\boldsymbol{\theta}}^{-1}(X_1|X_0)]$.

5.3 Massima verosimiglianza condizionata (CML)

Lo stimatore di massima verosimiglianza condizionata (CML - "Conditional Maximum Likelihood") di $\boldsymbol{\theta} = (\phi, \lambda)'$ è il valore $\hat{\boldsymbol{\theta}}$ che massimizza la funzione di log-verosimiglianza condizionata $l(\phi, \lambda)$.

La funzione di log-verosimiglianza condizionata per un processo PoINAR(1) è:

$$l(\phi, \lambda) = \log \left[\prod_{t=1}^n Pr(x_t | x_{t-1}) \right] = \sum_{t=1}^n \log [Pr_{x_t | x_{t-1}}(\boldsymbol{\theta})] \quad (5.4)$$

dove

$$Pr_{x_t | x_{t-1}}(\boldsymbol{\theta}) := Pr(X_t = x_t | X_{t-1} = x_{t-1}) = \sum_{k=0}^{\min(x_t, x_{t-1})} \binom{x_{t-1}}{k} \phi^k (1-\phi)^{x_{t-1}-k} \frac{\lambda^{x_t-k}}{(x_t - k)!}$$

In generale, questo stimatore si può ottenere usando metodi numerici poichè uguagliare a zero le derivate della verosimiglianza del primo ordine porta a un complicato sistema di equazioni non lineari.

E' noto dagli studi in letteratura che questo stimatore risulti essere uno dei migliori stimatori utilizzati nel campo delle serie storiche a valori interi non negativi, in quanto presenta le migliori performance in termini di distorsione.

Lo stimatore CML ha distribuzione asintoticamente normale:

$$\sqrt{n}(\hat{\phi} - \phi, \hat{\lambda} - \lambda) \xrightarrow{d} N(0, \mathbf{K}^{-1})$$

dove \mathbf{K} è la matrice d'informazione di Fisher.

Tuttavia, gli stimatori CML non hanno un'espressione in forma chiusa e anche la matrice d'informazione di Fisher non è disponibile; per cui è possibile definirli solo tramite algoritmi numerici come, ad esempio, metodi di ottimizzazione non lineare.

Gli stimatori alternativi proposti in letteratura, come gli stimatori di Yule-Walker e il CLS visto in precedenza, al contrario sono facilmente calcolabili e

le procedure sono semplici. Tuttavia, questi stimatori risultano avere performance peggiori in termini di distorsione e di errore quadratico medio rispetto allo stimatore CML.

6 Studio di simulazione

In questo capitolo viene riportato lo studio di simulazione fatto da Lu e Wang (2021), in cui si confrontano gli stimatori descritti finora (MELE, CLS, MQL, CML), con una particolare attenzione al nuovo stimatore proposto MELE per valutarne la performance.

E' stata calcolata la distorsione empirica ("empirical bias") insieme alla radice quadrata dell'errore quadratico medio su 1000 replicazioni per ogni combinazione dei parametri.

L'errore quadratico medio viene solitamente utilizzato per giudicare la qualità di uno stimatore in termini di variazione e distorsione ed è definito nel seguente modo:

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

Uno stimatore $\hat{\theta}$ del parametro θ si dice non distorto se:

$$E(\hat{\theta}) = \theta$$

e si chiama distorsione l'errore sistematico di stima, rappresentato dalla quantità:

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Nella Tabella 1 vengono riportati i valori della distorsione empirica (EB) e della radice dell'errore quadratico medio (RMSE) nel seguente formato: (EB,RMSE).

Dai risultati ottenuti, emerge che lo stimatore MELE è un buono stimatore in quanto produce stime con distorsione empirica e RMSE del tutto equiparabili al metodo CML. Il metodo CML ha performance migliori in quanto

presenta la minor distorsione e minor RMSE, tuttavia non c'è una differenza significativa tra lo stimatore MELE e lo stimatore CML.

Per quanto riguarda i metodi CLS ed MQL, essi hanno le medesime performance e risultano essere gli stimatori con maggior distorsione rispetto agli altri metodi.

Ponendo attenzione sulla variazione della numerosità campionaria, si evince che quando il campione ha una numerosità bassa, lo stimatore MELE risulta migliore rispetto agli stimatori CLS ed MQL. Ciò si verifica anche quando il valore dei parametri è alto.

La tabella riportata in seguito corrisponde alla Tabella 1 presente nell'articolo "*A new estimation for INAR(1) process with Poisson distribution*" di Lu e Wang (2021).

Tabella 1: Stime dei parametri ϕ e $\lambda = 1$

	CLS	MQL	MELE	CML
ϕ				
		$\phi = 0.1$		
$n = 50$	(-0.0235, 0.1416)	(-0.0205, 0.1459)	(0.0142, 0.1204)	(0.0112, 0.1143)
$n = 100$	(-0.0138, 0.0969)	(-0.0138, 0.0969)	(0.0013, 0.0856)	(-0.0011, 0.0836)
$n = 200$	(-0.0091, 0.0726)	(-0.0090, 0.0719)	(-0.0029, 0.0676)	(-0.0036, 0.0656)
$n = 500$	(-0.0017, 0.0451)	(-0.0018, 0.0449)	(-0.0008, 0.0450)	(-0.0012, 0.0441)
λ				
$n = 50$	(0.0252, 0.2112)	(0.0216, 0.2126)	(-0.0150, 0.1986)	(-0.0126, 0.1891)
$n = 100$	(0.0109, 0.1495)	(0.0110, 0.1501)	(-0.0117, 0.1434)	(-0.0028, 0.1403)
$n = 200$	(0.0072, 0.1054)	(0.0072, 0.1047)	(-0.0021, 0.1013)	(0.0012, 0.1003)
$n = 500$	(0.0041, 0.0671)	(0.0042, 0.0671)	(0.0024, 0.0672)	(0.0034, 0.0663)
		$\phi = 0.3$		
ϕ				
$n = 50$	(-0.0414, 0.1387)	(-0.0406, 0.1376)	(-0.0255, 0.1369)	(-0.0289, 0.1300)
$n = 100$	(-0.0164, 0.1030)	(-0.0157, 0.1017)	(-0.0096, 0.0997)	(-0.0097, 0.0945)
$n = 200$	(-0.0087, 0.0727)	(-0.0074, 0.0710)	(-0.0054, 0.0704)	(-0.0033, 0.0669)
$n = 500$	(-0.0040, 0.0453)	(-0.0030, 0.0446)	(-0.0021, 0.0419)	(-0.0012, 0.0404)
λ				
$n = 50$	(0.0589, 0.2436)	(0.0575, 0.2411)	(0.0205, 0.2446)	(0.0408, 0.2305)
$n = 100$	(0.0109, 0.1639)	(0.0177, 0.1636)	(0.0027, 0.1628)	(0.0095, 0.1570)
$n = 200$	(0.0072, 0.1213)	(0.0080, 0.1195)	(0.0036, 0.1177)	(0.0019, 0.1125)
$n = 500$	(0.0041, 0.0752)	(0.0039, 0.0744)	(0.0023, 0.0704)	(0.0017, 0.0686)

Tabella 1: continuo

	CLS	SQL	MELE	CML
ϕ				
$n = 50$	(-0.0512, 0.1367)	(-0.0478, 0.1382)	(-0.0155, 0.1219)	(-0.0172, 0.1133)
$n = 100$	(-0.0267, 0.0970)	(-0.0256, 0.0950)	(-0.0092, 0.0830)	(-0.0113, 0.0777)
$n = 200$	(-0.0163, 0.0627)	(-0.0148, 0.0615)	(-0.0071, 0.0556)	(-0.0074, 0.0527)
$n = 500$	(-0.0029, 0.0419)	(-0.0026, 0.0404)	(-0.0001, 0.0346)	(-0.0001, 0.0334)
λ				
$n = 50$	(0.0882, 0.2997)	(0.0817, 0.3022)	(-0.0014, 0.2568)	(0.0195, 0.2451)
$n = 100$	(0.0456, 0.2074)	(0.0437, 0.2046)	(0.0070, 0.1771)	(0.0149, 0.1685)
$n = 200$	(0.0256, 0.1339)	(0.0227, 0.1324)	(0.0055, 0.1219)	(0.0079, 0.1152)
$n = 500$	(0.0055, 0.0909)	(0.0050, 0.0881)	(-0.0005, 0.0759)	(-0.0002, 0.0735)
ϕ				
$n = 50$	(-0.0686, 0.1145)	(-0.0657, 0.1134)	(-0.0089, 0.0787)	(-0.0111, 0.0719)
$n = 100$	(-0.0330, 0.0756)	(-0.0296, 0.0747)	(-0.0050, 0.0495)	(-0.0059, 0.0474)
$n = 200$	(-0.0155, 0.0540)	(-0.0146, 0.0526)	(-0.0033, 0.0350)	(-0.0036, 0.0331)
$n = 500$	(-0.0054, 0.0337)	(-0.0053, 0.0325)	(-0.0004, 0.0223)	(-0.0007, 0.0215)
λ				
$n = 50$	(0.2214, 0.4130)	(0.2119, 0.4093)	(0.0104, 0.2816)	(0.0270, 0.2543)
$n = 100$	(0.0993, 0.2704)	(0.0882, 0.2690)	(0.0003, 0.1744)	(0.0079, 0.1667)
$n = 200$	(0.0424, 0.1853)	(0.0396, 0.1821)	(0.0008, 0.1215)	(0.0029, 0.1163)
$n = 500$	(0.0198, 0.1185)	(0.0192, 0.1137)	(0.0029, 0.0793)	(0.0035, 0.0752)

Tabella 1: continuo

	CLS	SQL	MELE	CML
ϕ				
$n = 50$	(-0.0850, 0.1002)	(-0.0846, 0.1000)	(-0.0013, 0.0568)	(-0.0047, 0.0249)
$n = 100$	(-0.0392, 0.0578)	(-0.0390, 0.0578)	(-0.0045, 0.0308)	(-0.0015, 0.0171)
$n = 200$	(-0.0203, 0.0352)	(-0.0119, 0.0349)	(-0.0016, 0.0137)	(-0.0013, 0.0119)
$n = 500$	(-0.0086, 0.0212)	(-0.0082, 0.0210)	(-0.0009, 0.0076)	(-0.0008, 0.0072)
λ				
$n = 50$	(0.8427, 1.0328)	(0.8376, 1.0298)	(0.1175, 0.5893)	(0.0368, 0.2521)
$n = 100$	(0.3885, 0.5873)	(0.3859, 0.5873)	(0.0375, 0.3121)	(0.0090, 0.1703)
$n = 200$	(0.2018, 0.3630)	(0.1970, 0.3594)	(0.0122, 0.1443)	(0.0096, 0.1192)
$n = 500$	(0.0832, 0.2170)	(0.0796, 0.2150)	(0.0057, 0.0779)	(0.0051, 0.0744)

7 Conclusioni

Questa relazione è basata sul lavoro di Lu e Wang (2021) in cui si dimostra attraverso uno studio di simulazione che lo stimatore proposto MELE risulta essere un buon stimatore al pari dello stimatore di massima verosimiglianza condizionata (CML). Come si può notare dai risultati, passare attraverso il metodo di verosimiglianza empirica permette di ottenere stimatori che hanno buone performance, sebbene non abbiano una definizione in forma chiusa, per cui la loro elaborazione richiede il supporto di algoritmi numerici. Ciò è dovuto dall'espressione articolata che assume la funzione di verosimiglianza empirica per i processi PoINAR, per cui la relativa massimizzazione, attuata per ottenere le stime, non è facilmente calcolabile. Tuttavia, il vantaggio di questi stimatori risulta notevole in termini di distorsione, permettendo di ottenere un'inferenza più accurata.

In letteratura, negli ultimi decenni, sono stati studiati e proposti nuovi stimatori per questo tipo di processi con lo scopo di ottenere un miglioramento delle performance sempre più significativo. L'obiettivo dei ricercatori sarà inoltre quello di estendere e ricavare buoni risultati applicando tali stimatori ad altri modelli.

8 Appendice

In questa sezione riporto il codice del programma (software R) utilizzato nel capitolo 3 per simulare il processo PoINAR(1) ed i relativi grafici.

Simulazione del processo PoINAR(1), Figura 1

```
T <- 200
```

```
mu <- 3
```

```
phi <- 0.5
```

```
lambda <- -mu*(1-phi)
```

```
set.seed(1)
```

Inizializzo X con la distribuzione marginale:

```
X <- rpois(1,mu)
```

```
dati <- c()
```

```
eps <- rpois(T,lambda)
```

```
X1,...,XT:
```

```
for(r in c(1:T)) {
```

```
X <- rbinom(1,X,phi)+eps[r]
```

```
dati <- c(dati,X)
```

```
}
```

Grafici

```
plot(dati, type="b", pch=16, cex=0.8, xlab="t", ylab="PoINAR(1)")
```

```
acf(dati, lag.max= T/4)
```

```
pacf(dati,lag.max= T/4)
```

Simulazione del processo PoINAR(1), Figura 2

```
phi <- 0.95
```

```
lambda <- mu*(1-phi)
```

```
set.seed(5)
```

Inizializzo X con la distribuzione marginale

```
X <- rpois(1,mu)
```

```
dati <- c()
```

```
tabeps <- rpois(T,lambda)
```

```
X1,...,XT:
```

```
for(r in c(1:T)) {
```

```
X <- rbinom(1,X,phi)+tabeps[r]
```

```
dati <- c(dati,X)
```

```
}
```

Grafici

```
plot(dati, type="b", pch=16, cex=0.8, xlab="t", ylab="PoINAR(1)")
```

```
acf(dati, lag.max = T/4)
```

```
pacf(dati, lag.max = T/4)
```

Riferimenti bibliografici

- [1] MA Al-Osh and AA Alzaid. First-order integer-valued autoregressive (inar (1)) process. *Journal of Time Series Analysis*, 8(3):261–275, 1987.
- [2] MA Al-Osh and AA Alzaid. First-order integer-valued autoregressive (inar (1)) process: distributional and regression properties. *Statistica Neerlandica*, 42(1):53–61, 1988.
- [3] M Bourguignon and KLP Vasconcellos. Improved estimation for poisson inar (1) models. *Journal of Statistical Computation and Simulation*, 85(12):2425–2441, 2015.
- [4] T Di Fonzo and F Lisi. *Serie storiche economiche: analisi statistiche e applicazioni*. Carocci, 2005.
- [5] RK Freeland and BPM McCabe. Analysis of low count time series data by poisson autoregression. *Journal of time series analysis*, 25(5):701–722, 2004.
- [6] RK Freeland and BPM McCabe. Asymptotic properties of cls estimators in the poisson ar (1) model. *Statistics & probability letters*, 73(2):147–153, 2005.
- [7] P Gorgi. Stima e previsione per modelli inar (1) con distribuzione dell’errore binomiale e binomiale negativa.
- [8] LA Klimko and PI Nelson. On conditional least squares estimation for stochastic processes. *The Annals of statistics*, pages 629–642, 1978.
- [9] F Lu and D Wang. A new estimation for inar (1) process with poisson distribution. *Computational Statistics*, pages 1–17, 2021.

- [10] AB Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.
- [11] AB Owen. Empirical likelihood for linear models. *The Annals of Statistics*, pages 1725–1747, 1991.
- [12] AB Owen. *Empirical likelihood*. Chapman and Hall/CRC, 2001.
- [13] L Pace and A Salvan. *Introduzione alla statistica: Inferenza, verosimiglianza, modelli.-2001.-xvi, 422 p.* Cedam, 1996.
- [14] J Qin and J Lawless. Empirical likelihood and general estimating equations. *the Annals of Statistics*, 22(1):300–325, 1994.
- [15] J Qin and J Lawless. Estimating equations, empirical likelihood and constraints on parameters. *Canadian Journal of Statistics*, 23(2):145–159, 1995.
- [16] CH Weiß. *An introduction to discrete-valued time series*. John Wiley & Sons, 2018.
- [17] H Zhang, D Wang, and F Zhu. Empirical likelihood inference for random coefficient inar (p) process. *Journal of Time Series Analysis*, 32(3):195–203, 2011.
- [18] H Zheng, IV Basawa, and S Datta. First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*, 137(1):212–229, 2007.