

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica e Gestione delle Imprese



**RELAZIONE FINALE
CAMBIAMENTO DI REGIME NEI MODELLI INARCH**

Relatore: Prof.ssa Luisa Bisaglia
Dipartimento di Scienze Statistiche

Laureando: Manuel Litti
Matricola N. 620914

Anno Accademico 2012/2013

SOMMARIO

INTRODUZIONE	5
1. PROCESSI AUTOREGRESSIVI A VALORI INTERI (INAR).....	7
1.1 Modello autoregressivo a valori interi di ordine 1, INAR(1)	8
1.1.1 Momenti e proprietà distributive del modello INAR(1)	10
1.1.2 Stima dei parametri del modello INAR(1).....	12
1.2 Modello autoregressivo a valori interi di ordine p, INAR(p)	18
2. PROCESSI GARCH A VALORI INTERI (INGARCH)	21
2.1 Modello INGARCH(p,q)	22
2.1.1 Convergenza e stazionarietà.....	23
2.2 Modello INGARCH(1,1)	26
3. PROCESSI INARCH E TEST PER VERIFICARE UN	
CAMBIAMENTO DEI PARAMETRI IN SERIE STORICHE PER	
DATI DI CONTEGGIO: APPLICAZIONE E SIMULAZIONE	31
3.1 Modello INARCH(1)	31
3.2 Test per la verifica di un cambiamento dei parametri	33
3.2.1 Inferenza sulle stime dei parametri del modello INARCH(1)	36
3.2.2 La distribuzione asintotica della statistica test sotto le ipotesi.....	39
3.3 Simulazione.....	42

APPENDICE.....	53
BIBLIOGRAFIA.....	55

INTRODUZIONE

Questo lavoro trae spunto dall'articolo "Changepoints in times series of counts" di Franke *et al.* (2012). Gli autori propongono un modello particolare per le serie storiche per dati di conteggio chiamato INARCH (*Integer-valued Autoregressive Conditional Heteroscedasticity*) e si soffermano poi sulla messa a punto di un test statistico volto a verificare un eventuale cambiamento di regime in questo tipo di modelli.

L'obiettivo di questa tesi è, dunque, replicare il lavoro di Franke *et al.* simulando dei processi INARCH con diverse numerosità e combinazioni di parametri e verificando l'efficacia del test su di essi.

Poiché gli INARCH sono pensati per serie storiche a valori interi, la tesi è stata ampliata con una presentazione di altri modelli adatti ad analizzare questo tipo di serie storiche. I dati di conteggio, infatti, sono molto diffusi ed hanno diversi campi di applicazione nella realtà, ad esempio:

- assicurazioni (numero di richieste di risarcimento danni),
- processi di controllo (numero di pezzi difettosi nella produzione),
- informatica (numero di intrusioni in un sistema),
- economia (numero di disoccupati),
- epidemiologia (numero di malati).

Pertanto, la tesi risulta così strutturata: nel Capitolo 1 verranno presentati i processi INAR (*Integer-valued Autoregressive*) con particolare attenzione al processo INAR(1) ed una breve analisi degli INAR(p); nel Capitolo 2 vengono descritti i processi INGARCH (*Integer-valued Generalized Autoregressive Conditional Heteroskedastic*)

soffermandosi sul caso INGARCH(1,1) ed infine nel Capitolo 3 vengono presentati i modelli INARCH, il test statistico ed i risultati della simulazione condotta.

1. PROCESSI AUTOREGRESSIVI A VALORI INTERI (INAR)

Prima di prendere in esame il modello INAR (*Integer-valued Autoregressive*) è necessario dare la definizione dell'operatore di *thinning* “ \circ ” così come riportata da Steutel & Van Harn (1979) e da Van Harn (1978, p. 85):

Sia \mathcal{X} una variabile casuale non negativa a valori interi; allora $\forall \alpha \in [0, 1]$ l'operatore “ \circ ” è tale che

$$\alpha \circ \mathcal{X} = \sum_{i=1}^{\mathcal{X}} Y_i \quad (1.1)$$

dove Y_i è una successione di variabili casuali i.i.d., indipendenti da \mathcal{X} , tali che

$$\mathcal{P}_r(Y_i = 1) = 1 - \mathcal{P}_r(Y_i = 0) = \alpha.$$

La variabile casuale $\alpha \circ \mathcal{X}$, quindi, conta il numero di successi in un numero casuale di esperimenti bernoulliani in cui la probabilità di successo α rimane costante durante tutto l'esperimento, cioè:

$$\alpha \circ \mathcal{X} | \mathcal{X} \sim Bi(\mathcal{X}, \alpha).$$

Da questa definizione è immediato che $0 \circ \mathcal{X} = 0$ e $1 \circ \mathcal{X} = \mathcal{X}$. Inoltre, valgono le seguenti proprietà (Kedem, Fokianos, 2002: 179):

$$\beta \circ (\alpha \circ \mathcal{X}) \sim (\beta\alpha) \circ \mathcal{X} \quad (1.2)$$

$$E[\alpha \circ \mathcal{X} | \mathcal{X}] = \alpha \mathcal{X} \quad (1.3)$$

$$E[\alpha \circ \mathcal{X}] = \alpha E[\mathcal{X}] \quad (1.4)$$

$$Var[\alpha \circ \mathcal{X} | \mathcal{X}] = \alpha(1 - \alpha)\mathcal{X} \quad (1.5)$$

$$Var[\alpha \circ \mathcal{X}] = \alpha^2 Var[\mathcal{X}] + \alpha(1 - \alpha)E[\mathcal{X}] \quad (1.6)$$

dove la (1.2) implica l'uguaglianza in distribuzione $\forall \beta \in [0, 1]$.

1.1 Modello autoregressivo a valori interi di ordine 1, INAR(1)

Per semplicità verrà inizialmente presentato il processo autoregressivo a valori interi di ordine 1, mentre ad un breve accenno al modello nella sua forma più generale è dedicato il paragrafo 1.2.

Il processo INAR(1) $\{\mathcal{X}_t\}, t = 1, 2, \dots, N$ è definito (Al-Osh & Alzaid, 1987) come:

$$\mathcal{X}_t = \alpha \circ \mathcal{X}_{t-1} + \varepsilon_t \quad (1.7)$$

dove $\alpha \in [0, 1]$ e ε_t è una successione di variabili casuali i.i.d. non negative a valori interi e tra loro incorrelate, con $E[\varepsilon_t] = \mu$ e $Var[\varepsilon_t] = \sigma^2$.

Dalla (1.7) si può notare che il modello INAR imita la struttura del classico modello autoregressivo AR, con l'operatore di *thinning* al posto del prodotto scalare.

Inoltre, questa definizione implica che il valore assunto dal processo al tempo t , \mathcal{X}_t , deriva:

- dagli elementi presenti nel processo al tempo $t - 1$, \mathcal{X}_{t-1} , “sopravvissuti” con probabilità α ,
- dai nuovi elementi che entrano casualmente nel processo nell'intervallo $(t - 1, t]$ rappresentati dal termine ε_t .

Da qui in poi si assume che la successione dei termini d'errore ε_t abbia distribuzione di Poisson di parametro λ , a differenza dei modelli AR nei quali generalmente si assume che gli errori siano Normali a media nulla e varianza costante σ_ε^2 . La scelta della Poisson come distribuzione di riferimento non è però l'unica possibile e le procedure di stima trattate in seguito sono applicabili anche ad altre distribuzioni, purché siano a valori interi non negativi.

Di seguito sono riportate le serie generate da un INAR(1) e da un AR(1). Per poter fare dei confronti, i due processi sono stati simulati in modo da avere la stessa media e la stessa varianza.

Nella figura 1.1 sono riportate 100 osservazioni generate dal modello (1.7), nel quale la Y_i (1.1) è una Bernoulli di parametro $\alpha = 0.5$ ed ε_t è una Poisson di parametro $\lambda = 1$. Nella figura 1.2, invece, le 100 osservazioni sono generate da un AR(1) del tipo $\mathcal{X}_t = \phi\mathcal{X}_{t-1} + \varepsilon_t$, con ε_t Normale di media $\lambda = 1$ e varianza $\lambda(1 + \alpha) = 1.5$ e $\phi = 0.5$.

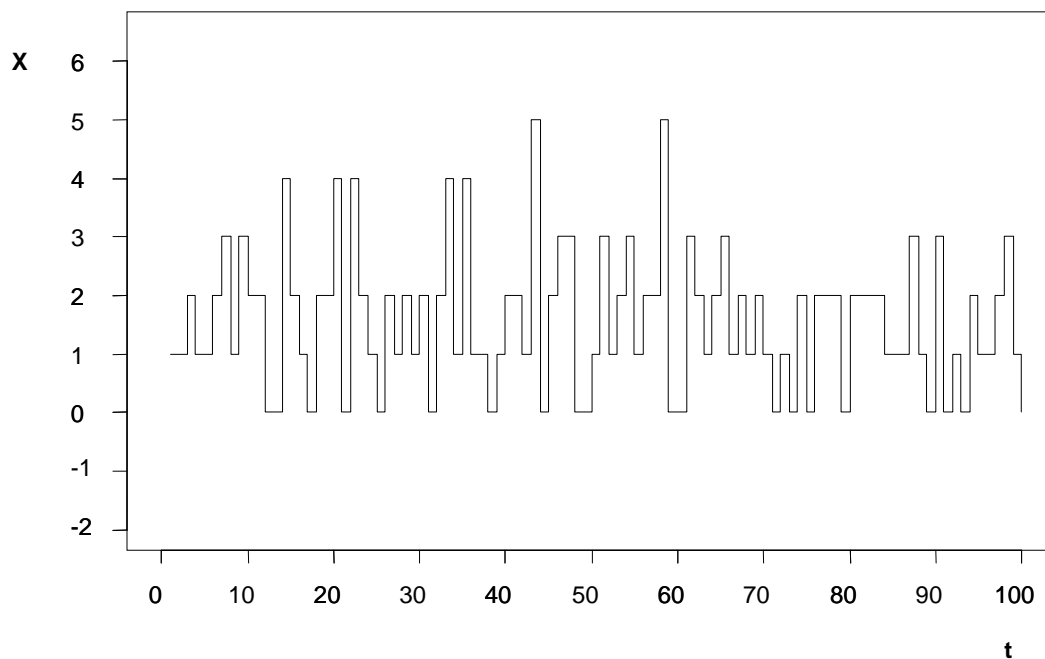


Figura 1.1: INAR(1) generato da una Bernoulli(0.5) e da una Poisson(1)

Come si può vedere, mentre le realizzazioni dell'INAR(1) sono tutte valori interi non negativi quelle dell'AR(1) sono valori reali alcuni dei quali negativi. Inoltre nella figura 1.1 si può notare che l'INAR ha molte realizzazioni uguali o vicine alla media (2), mentre nell'AR questo non accade.

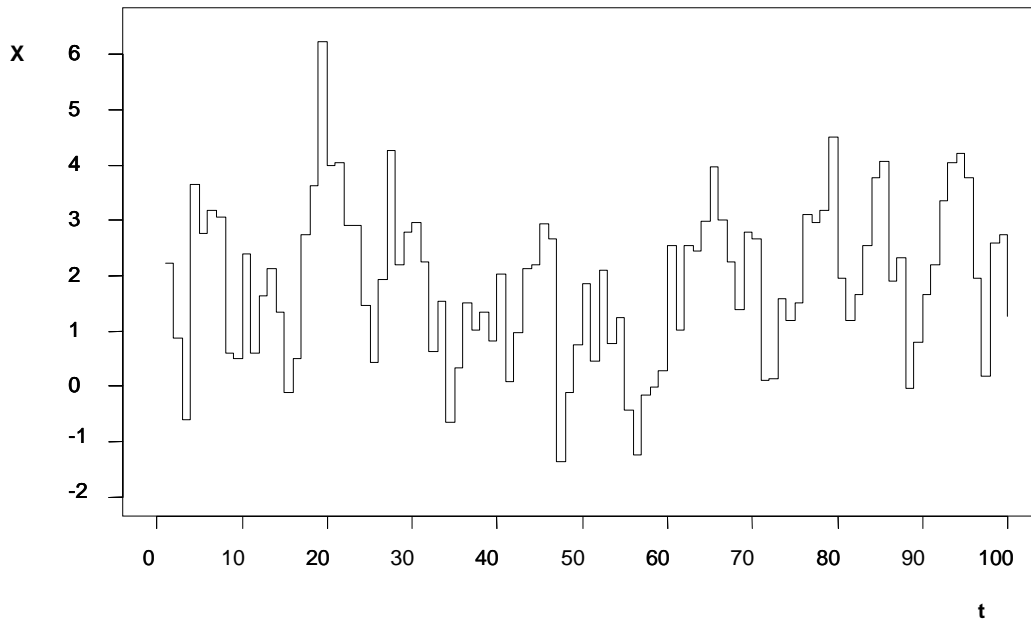


Figura 1.2: AR(1) con $\phi = 0.5$ e $\varepsilon_t \sim N(1, 1.5)$

Il modello (1.7) può anche essere espresso in funzione dei valori presenti e passati di ε_t . Infatti, attraverso sostituzioni successive e grazie alla proprietà (1.2) si ottiene la seguente rappresentazione a media mobile infinita:

$$\mathcal{X}_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j} \quad (1.8)$$

che può essere utilizzata per studiare le proprietà distributive del processo. Ad esempio, questo tipo di rappresentazione mette subito in evidenza che, per $\alpha \in (0, 1)$, la dipendenza di \mathcal{X}_t dalla successione ε_t decresce esponenzialmente al crescere di t .

1.1.1 Momenti e proprietà distributive del modello INAR(1)

Il valore atteso del processo \mathcal{X}_t definito nella (1.7) è:

$$E[\mathcal{X}_t] = \alpha E[\mathcal{X}_{t-1}] + \mu = \alpha^t E[\mathcal{X}_0] + \mu \sum_{j=0}^{t-1} \alpha^j \quad (1.9)$$

mentre la varianza è:

$$\begin{aligned} Var[\mathcal{X}_t] &= \alpha^2 Var[\mathcal{X}_{t-1}] + \alpha(1-\alpha)E[\mathcal{X}_{t-1}] + \sigma^2 \\ &= \alpha^{2t} Var[\mathcal{X}_0] + (1-\alpha) \sum_{j=1}^t \alpha^{2j-1} E[\mathcal{X}_{t-j}] + \sigma^2 \sum_{j=1}^t \alpha^{2(j-1)}. \end{aligned} \quad (1.10)$$

In caso di stazionarietà del processo, media e varianza diventano rispettivamente:

$$E[\mathcal{X}_t] = \frac{\mu}{1-\alpha} \quad (1.11)$$

$$Var[\mathcal{X}_t] = \frac{\alpha\mu + \sigma^2}{1-\alpha^2}. \quad (1.12)$$

Per poter ricavare la funzione di autocovarianza si noti che, dalla rappresentazione a media mobile infinita (1.8), si può giungere alla seguente uguaglianza in distribuzione (Al-Osh & Alzaid, 1987):

$$(\mathcal{X}_t, \mathcal{X}_{t-k}) = \left(\alpha^k \circ \mathcal{X}_{t-k} + \sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t-j}, \mathcal{X}_{t-k} \right).$$

Grazie a questo risultato si ottiene che l'autocovarianza al ritardo k , per ogni k intero non negativo, è:

$$\begin{aligned} \gamma[k] &= Cov[\mathcal{X}_{t-k}, \mathcal{X}_t] \\ &= Cov[\mathcal{X}_{t-k}, \alpha^k \circ \mathcal{X}_{t-k}] + Cov \left[\mathcal{X}_{t-k}, \sum_{j=0}^{k-1} \alpha^j \circ \varepsilon_{t-j} \right] \\ &= \alpha^k Var[\mathcal{X}_{t-k}] + \sum_{j=0}^{k-1} \alpha^j Cov[\mathcal{X}_{t-k}, \varepsilon_{t-j}] = \alpha^k \gamma[0]. \end{aligned} \quad (1.13)$$

Infine, poiché la funzione di autocorrelazione per definizione è data da $\rho[k] = \gamma[k]/\gamma[0]$, si ottiene che per un processo INAR(1) essa vale:

$$\rho[k] = \alpha^k. \quad (1.14)$$

Questo dimostra che l'autocorrelazione tende esponenzialmente a zero al crescere del ritardo k ed ha la stessa forma delle equazioni di Yule-Walker di un AR(1), ma a differenza di quest'ultimo, $\rho[k]$ assume sempre valori positivi per un INAR(1).

Anche per quanto riguarda le proprietà distributive vi sono delle similarità con i processi AR. In primo luogo, se \mathcal{X}_t è uguale in distribuzione a \mathcal{X}_{t-1} e \mathcal{X}_{t-1} è indipendente da ε_t allora \mathcal{X}_t ha, per sua stessa definizione (1.7), una distribuzione discreta auto-decomponibile (Steutel & Van Harn, 1979).

Di conseguenza \mathcal{X}_t è anche unimodale e la sua distribuzione dipende unicamente da quella di ε_t ; \mathcal{X}_t ha, quindi, una distribuzione di Poisson se e solo se ε_t ha una distribuzione di Poisson così come accade per l'AR(1) nel quale la distribuzione Normale degli errori garantisce la distribuzione Normale del processo.

1.1.2 Stima dei parametri del modello INAR(1)

Di seguito vengono esposti tre diversi metodi (Al-Osh & Alzaid, 1987) per ottenere le stime dei parametri α e λ : il primo è il parametro di scala della Binomiale risultante dall'operazione di *thinning* nel modello (1.7), mentre il secondo si riferisce alla successione $\{\varepsilon_t\}$ che, come detto in precedenza, si assume avere distribuzione di Poisson di media λ .

A differenza di un processo AR(1), la procedura di stima è più complessa nel caso dell'INAR(1) poiché la distribuzione di \mathcal{X}_t condizionata a \mathcal{X}_{t-1} è la convoluzione tra una distribuzione di Poisson ed una Binomiale.

I. Metodo di Yule-Walker

Il metodo più semplice per ottenere una stima di α è sostituire il $\gamma[k]$ che compare nell'equazione di Yule-Walker con l'autocovarianza campionaria e di esplicitare l'equazione rispetto ad α .

Così facendo si ottiene:

$$\hat{\alpha} = \frac{\sum_{t=0}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=0}^n (x_t - \bar{x})^2} \quad (1.15)$$

dove \bar{x} è la media campionaria.

A questo punto si può calcolare $\hat{\varepsilon}_t = x_t - \hat{\alpha}x_{t-1}$ per $t = 1, 2, \dots, n$ e di conseguenza una stima ragionevole per λ è data da:

$$\hat{\lambda} = \sum_{t=1}^n \frac{\hat{\varepsilon}_t}{n}. \quad (1.16)$$

II. Metodo dei minimi quadrati condizionati

Per ottenere le stime ai minimi quadrati condizionati (CLS) di α e λ è necessario prima ricavare la media condizionata di X_t dato X_{t-1} . Utilizzando la proprietà (1.3) si ha che:

$$E[X_t | X_{t-1}] = \alpha X_{t-1} + \lambda \equiv g(\theta, X_{t-1})$$

dove $\theta = (\alpha, \lambda)^T$.

Una volta specificato questo risultato, la procedura di stima che si utilizza è quella sviluppata da Klimko & Nelson (1978), che si basa sulla minimizzazione della somma dei quadrati degli scarti dalla media condizionata. Pertanto, le stime CLS sono quei valori di α e λ che minimizzano:

$$Q_n(\theta) = \sum_{t=1}^n [X_t - g(\theta, X_{t-1})]^2$$

rispetto a θ . Derivando $Q_n(\theta)$ rispettivamente rispetto ad α e λ si ottiene:

$$\hat{\alpha} = \frac{\sum_{t=1}^n x_t x_{t-1} - (\sum_{t=1}^n x_t \sum_{t=1}^n x_{t-1})/n}{\sum_{t=1}^n x_{t-1}^2 - (\sum_{t=1}^n x_{t-1})^2/n} \quad (1.17)$$

$$\hat{\lambda} = \frac{1}{n} \left(\sum_{t=1}^n x_t - \hat{\alpha} \sum_{t=1}^n x_{t-1} \right). \quad (1.18)$$

Come si può notare, nel caso di stazionarietà del processo, le stime (1.17) e (1.18) sono molto simili, rispettivamente, a quelle in (1.15) e (1.16).

Inoltre, la funzione g e le sue derivate prime e seconde soddisfano le condizioni di regolarità del teorema (3.1) specificate da Klimko & Nelson (1978, p. 634) e dunque le stime CLS sono consistenti in senso forte. E se $E[|\varepsilon_t|^3] < \infty$, condizione sempre verificata nel caso in cui $\varepsilon_t \sim \mathcal{P}(\lambda)$, allora sono rispettate le condizioni anche per il teorema (3.2) di Klimko & Nelson nel quale si afferma che $\hat{\theta} = (\hat{\alpha}, \hat{\lambda})^T$ ha distribuzione asintoticamente Normale:

$$\sqrt{n}(\hat{\theta} - \theta_0) \sim \mathcal{N}_2(\underline{0}, \mathcal{V}^{-1} \mathcal{W} \mathcal{V}^{-1})$$

dove $\theta_0 = (\alpha_0, \lambda_0)^T$ è il vettore che contiene i veri valori dei parametri, \mathcal{V} è una matrice 2×2 i cui elementi sono:

$$\mathcal{V}_{ij} = E \left(\frac{\partial g(\theta_0, \mathcal{X}_{t-1})}{\partial \theta_i} \cdot \frac{\partial g(\theta_0, \mathcal{X}_{t-1})}{\partial \theta_j} \right) \quad i, j = 1, 2$$

e \mathcal{W} è formata da:

$$\mathcal{W}_{ij} = E \left(u_t^2(\theta_0) \frac{\partial g(\theta_0, \mathcal{X}_{t-1})}{\partial \theta_i} \cdot \frac{\partial g(\theta_0, \mathcal{X}_{t-1})}{\partial \theta_j} \right) \quad i, j = 1, 2$$

con:

$$u_t(\theta_0) = \mathcal{X}_t - g(\theta_0, \mathcal{X}_{t-1}).$$

III. Metodo della massima verosimiglianza

Come già accennato all'inizio del paragrafo, la distribuzione del processo \mathcal{X}_t è la convoluzione tra una distribuzione di Poisson ed una Binomiale, dunque la funzione di verosimiglianza può essere scritta come:

$$\mathcal{L}(\mathbf{x}, \alpha, \lambda) = \left(\prod_{t=1}^n \mathcal{P}_t(x_t) \right) \frac{[\lambda/(1-\alpha)]^{x_0} \exp[-\lambda/(1-\alpha)]}{x_0!} \quad (1.19)$$

dove:

$$\mathcal{P}_t(x) = e^{-\lambda} \sum_{i=0}^{\min(x_{t-1}, x_t)} \frac{\lambda^{x-i}}{(x-i)!} \binom{x_{t-1}}{i} \alpha^i (1-\alpha)^{x_{t-1}-i} \quad t = 1, 2, \dots, n$$

e $\mathbf{x} = (x_0, x_1, \dots, x_n)$.

Per semplicità, quindi, si preferisce prendere inizialmente in considerazione la funzione di verosimiglianza condizionata con x_0 noto, da cui ricavare le stime CML di α e λ :

$$\mathcal{L}(\mathbf{x}, \alpha, \lambda | x_0) = \prod_{t=1}^n \mathcal{P}_t(x_t).$$

Per massimizzare $\mathcal{L}(\mathbf{x}, \alpha, \lambda | x_0)$ rispetto ad α e λ si segue il metodo descritto da Sprott (1983), valido nel caso in cui $\mathcal{P}_t(x)$ sia indipendente da t . Le derivate della log-verosimiglianza condizionata, quindi, sono rispettivamente:

$$\mathcal{S}_\lambda = \sum_{t=1}^n \frac{\partial \log \mathcal{P}_t(x_t)}{\partial \lambda} = \sum_{t=1}^n \frac{\mathcal{P}_t(x_t - 1)}{\mathcal{P}_t(x_t)} - n = \sum_{t=1}^n \mathcal{H}(t) - n \quad (1.20)$$

$$\mathcal{S}_\alpha = \sum_{t=1}^n \frac{\partial \log \mathcal{P}_t(x_t)}{\partial \alpha} = \sum_{t=1}^n \frac{(x_t - \alpha x_{t-1}) - \lambda \mathcal{H}(t)}{\alpha(1-\alpha)}. \quad (1.21)$$

Ora, per ottenere le stime CML, si pone $\mathcal{S}_\lambda = 0$ e $\mathcal{S}_\alpha = 0$ e dalla (1.21) si arriva a:

$$\sum x_t = \alpha \sum x_{t-1} + n\lambda. \quad (1.22)$$

Quest'ultima equazione può essere utilizzata per esplicitare uno dei due parametri, ad esempio α e in questo caso si otterrebbe \mathcal{S}_λ in funzione del solo λ . A quel punto, iterando \mathcal{S}_λ si giunge ad una stima del parametro e, per sostituzione, alla stima dell'altro. Se dalla (1.22), invece, si esplicita λ si vede subito che si arriva ad una stima di quest'ultimo nella stessa forma della (1.18).

Adesso è possibile prendere in esame la funzione di verosimiglianza originaria (1.19), considerarne il logaritmo e ricavarne le derivate rispetto a λ ed α , rispettivamente pari a:

$$\mathcal{S}_\lambda^0 = \frac{\partial \log \mathcal{L}}{\partial \lambda} = \mathcal{S}_\lambda + \frac{x_0}{\lambda} - \frac{1}{1-\alpha} \quad (1.23)$$

$$\mathcal{S}_\alpha^0 = \frac{\partial \log \mathcal{L}}{\partial \alpha} = \mathcal{S}_\alpha + \frac{1}{1-\alpha} \left(x_0 - \frac{\lambda}{1-\alpha} \right). \quad (1.24)$$

Ponendo $\mathcal{S}_\lambda^0 = 0$ e $\mathcal{S}_\alpha^0 = 0$ si ottiene:

$$\lambda = \left(\sum_{t=0}^n x_t - \alpha \sum_{t=1}^n x_t \right) / \left(n + \frac{1+\alpha}{1-\alpha} \right). \quad (1.25)$$

Sostituendo quest'ultimo valore di λ nella (1.24) e iterando \mathcal{S}_α^0 si trova la stima ML di α e per sostituzione nella (1.25) anche quella di λ .

Come si può notare, a differenza dei due metodi precedenti che conducevano a delle formule chiuse da cui ricavare con semplicità le stime dei parametri d'interesse, il metodo della massima verosimiglianza permette soltanto il calcolo approssimato tramite analisi numerica e richiede dunque un maggior sforzo computazionale.

Al-Osh & Alzaid (1987) per valutare le performance dei diversi metodi di stima appena descritti hanno effettuato una simulazione con campioni di diversa numerosità e con diversi valori dei parametri, soffermandosi in particolare sulla distorsione e

sull'errore quadratico medio (MSE) delle stime ottenute. Per quanto riguarda la distorsione, i risultati ottenuti hanno evidenziato che:

- sui 20 casi presi in esame, $\hat{\alpha}$ è risultato sottostimato e $\hat{\lambda}$ sovrastimato in 17 casi con le stime ML, in 19 con le stime CLS e in 20 con le stime Yule-Walker,
- la distorsione di entrambi i parametri, in particolare quella di $\hat{\lambda}$, aumenta al crescere del loro valore sia con le stime Yule-Walker che CLS. Ciò non avviene con le stime ML,
- la distorsione di entrambi i parametri è inversamente proporzionale all'ampiezza campionaria, cioè raddoppiando la grandezza del campione la distorsione si dimezza. Ciò avviene in misura minore per le stime ML.

In generale, viene fatto notare che la distorsione di $\hat{\alpha}$ e $\hat{\lambda}$ nelle stime ML è rispettivamente il 60% e 52% di quella nelle stime Yule-Walker quando $\alpha = 0.3$ e queste percentuali decrescono all'aumentare di α sino al 5% e 1% rispettivamente con $\alpha = 0.9$. Lo stesso accade all'aumentare della numerosità campionaria e confrontando la distorsione delle stime ML con quella delle stime CLS.

Invece, dall'analisi del MSE si nota che:

- nelle stime Yule-Walker e CLS, il MSE di $\hat{\lambda}$ aumenta all'aumentare di α , ma non si verifica lo stesso per $\hat{\alpha}$. Invece, nelle stime ML, non esiste nessuna relazione di questo tipo,
- in tutti e tre i metodi, il MSE di entrambi i parametri è inversamente proporzionale all'ampiezza campionaria.

Anche in questo caso, il MSE di $\hat{\alpha}$ e $\hat{\lambda}$ nelle stime ML è il 90% di quello nelle stime Yule-Walker e CLS quando $\alpha = 0.3$ e decresce fino al 5% quando $\alpha = 0.9$.

In conclusione, gli autori affermano che vale la pena utilizzare il metodo della massima verosimiglianza, poiché è il più efficace tra i tre, nonostante richieda più calcoli e in un'ipotetica classifica al secondo posto verrebbe il metodo CLS seguito, infine, da quello di Yule-Walker.

1.2 Modello autoregressivo a valori interi di ordine p , INAR(p)

A volte si ha a che fare con serie storiche per dati di conteggio le cui realizzazioni non dipendono solo dal passato più recente X_{t-1} , ma anche da tutte le altre precedenti realizzazioni. Di conseguenza, vi è la necessità di estendere il modello INAR(1) per poter tenere conto dei dati passati. Tale estensione è simile a quella fatta per i processi AR(p), con l'operatore di *thinning* al posto della moltiplicazione scalare.

Il modello INAR(p) è, dunque, stato definito da Alzaid & Al-Osh (1990) come:

$$X_t = \alpha_1 \circ X_{t-1} + \dots + \alpha_p \circ X_{t-p} + \varepsilon_t \equiv \sum_{i=1}^p \alpha_i \circ X_{t-i} + \varepsilon_t \quad (1.26)$$

con $t = 1, 2, \dots, N$; $\{\varepsilon_t\}$ successione di variabili casuali i.i.d. non negative ed a valori interi con media μ_ε e varianza finita σ_ε^2 ; tutte le operazioni di *thinning* mutuamente indipendenti l'una dall'altra e le costanti $\alpha_i \in [0, 1)$ ($i = 1, 2, \dots, p$) tali che $\sum_{i=1}^p \alpha_i < 1$.

Inoltre, poiché ogni singola operazione di *thinning* ha distribuzione condizionata Binomiale di parametro α_i , la distribuzione condizionata del vettore $(\alpha_1 \circ X_t, \dots, \alpha_p \circ X_t)^T$ dato $X_t = x_t$ è multinomiale di parametri $((\alpha_1, \dots, \alpha_p), x_t)$ ed è indipendente dal passato del processo.

Jin-Guan & Yuan (1991) nel teorema 2.1 affermano che il modello (1.26) è stazionario se le radici di:

$$\lambda^p - \alpha_1 \lambda^{p-1} - \dots - \alpha_{p-1} \lambda - \alpha_p = 0 \quad \text{con } \alpha_p \neq 0$$

sono tutte in modulo < 1 , mentre Alzaid & Al-Osh (1990) nel teorema 2.1 aggiungono che, se vale la condizione appena descritta e se $\sum_{j=0}^{\infty} (j+1)^{-1} r_j < \infty$ con $r_j = \sum_{k=j+1}^{\infty} \mathcal{P}(\varepsilon_1 = k)$, allora il processo (1.26) ha una distribuzione limite con funzione generatrice di probabilità data da:

$$\phi(s) = \prod_{j=0}^{\infty} \psi(1 - w_j + w_j s)$$

dove $\psi(s)$ denota la funzione generatrice di probabilità di ε_t e $\{w_j\}_{j=0}^{\infty}$ è una successione di pesi tali che $w_0 = 1$ e $w_j = \sum_{i=1}^{\min(j,p)} \alpha_i w_{j-i}$, che altro non sono che i coefficienti della rappresentazione a media mobile del processo in esame.

Il valore atteso di un processo INAR(p) sotto l'assunzione di stazionarietà è:

$$E[\mathcal{X}_t] = \frac{\mu_{\varepsilon}}{1 - \sum_{i=1}^p \alpha_i}$$

Per calcolare l'autocovarianza, Jin-Guan & Yuan (1991) utilizzano la rappresentazione matriciale del processo (1.26). Siano:

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

$$\mathbf{X} = (X_1, X_2, \dots, X_p)^T$$

e quindi $A \circ \mathbf{X} = (\sum_{i=1}^p (\alpha_i \circ X_i), X_1, \dots, X_{p-1})$. Allora, il lemma 2.2 afferma che:

$$E[A \circ \mathbf{X}] = AE[\mathbf{X}]$$

$$E[(A \circ \mathbf{X})(A \circ \mathbf{X})^T] = AE[\mathbf{X}\mathbf{X}^T]A^T + C$$

dove:

$$C = \begin{bmatrix} \sum_{i=1}^p \alpha_i(1 - \alpha_i)E[X_i] & 0 \\ 0 & 0 \end{bmatrix}_{p \times p}.$$

Date queste premesse, il processo INAR(p) $\mathbf{X}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})^T$ si può scrivere come $\mathbf{X}_t = A \circ \mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t$, con $\boldsymbol{\varepsilon}_t = (\varepsilon_t, 0, \dots, 0, 0)^T$. Poiché la covarianza campionaria di $\{\mathcal{X}_t\}$ è $\hat{\gamma}_k = (1/N) \sum_{t=1}^{N-k} (\mathcal{X}_t - \bar{\mathcal{X}})(\mathcal{X}_{t+k} - \bar{\mathcal{X}})$ si ha che:

$$\begin{aligned} \boldsymbol{\gamma}_k &= E(\mathbf{X}_t - E(\mathbf{X}_t))(\mathbf{X}_{t-k} - E(\mathbf{X}_{t-k}))^T \\ &= E(\mathbf{X}_t \mathbf{X}_{t-k}^T) - E(\mathbf{X}_t)E(\mathbf{X}_{t-k}^T) \\ &= AE(\mathbf{X}_{t-1} \mathbf{X}_{t-k}^T) + E(\boldsymbol{\varepsilon}_t)E(\mathbf{X}_{t-k}^T) - E(\mathbf{X}_t)E(\mathbf{X}_{t-k}^T) \\ &= A\boldsymbol{\gamma}_{k-1} + (A - I)E(\mathbf{X}_t)E(\mathbf{X}_t^T) + E(\boldsymbol{\varepsilon}_t)E(\mathbf{X}_t^T) \\ &= A\boldsymbol{\gamma}_{k-1} \end{aligned}$$

e quindi:

$$\gamma_k = \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \dots + \alpha_p \gamma_{k-p}$$

$$\rho_k \equiv \gamma_k / \gamma_0 = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \dots + \alpha_p \rho_{k-p}$$

dove $\gamma_k = E(X_t X_{t-k})$.

Quanto scritto finora implica che l'AR(p) e l'INAR(p) sono simili non solo nella forma, ma anche nelle regioni di stazionarietà e nella struttura della correlazione. Anche per quanto riguarda le stime dei parametri non si tratta altro che di un'estensione di quanto già affrontato nel paragrafo 1.1.2.

2. PROCESSI GARCH A VALORI INTERI (INGARCH)

Molti dei modelli usati per analizzare le serie storiche, tra cui gli INAR presi in esame nel capitolo precedente, operano sotto l'assunzione di omoschedasticità. Questa ipotesi, però, è poco realistica in molte situazioni concrete, ad esempio quando si vuole analizzare il corso di un titolo azionario, contraddistinto solitamente da grande volatilità, oppure dei dati medici/epidemiologici.

Per questo motivo, dunque, Engle (1982) introdusse i modelli ARCH (*Autoregressive Conditional Heteroscedasticity*), formulati in modo tale che la varianza condizionata del processo potesse variare nel tempo in funzione dei valori passati della serie secondo una forma autoregressiva. Un modello ARCH(q) può essere così definito:

$$\begin{cases} \mathcal{X}_t = \sigma_t \varepsilon_t \\ \mathcal{X}_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_t^2) \\ \sigma_t^2 = \gamma_0 + \sum_{i=1}^q \gamma_i \mathcal{X}_{t-i}^2 \end{cases} \quad \forall t \in \mathbb{Z} \quad (2.1)$$

con $\varepsilon_t \sim WN(0, 1)$, $\gamma_0 > 0$ e $\gamma_i \geq 0, i = 1, \dots, q$, mentre \mathcal{F}_{t-1} rappresenta l'insieme delle osservazioni della serie storica disponibili fino al tempo $t - 1$ incluso.

In seguito, Bollerslev (1986) propose un'estensione di questi modelli denominata GARCH (*Generalized Autoregressive Conditional Heteroskedastic*) allo scopo di ottenere una struttura della correlazione più flessibile ed una memoria più lunga, nonché una maggior parsimonia in termini di parametri. Infatti si può dire che un ARCH sta ad un GARCH così come un AR sta ad un ARMA (Di Fonzo, Lisi, 2005).

Il processo GARCH(p,q) di Bollerslev è definito come:

$$\begin{cases} \mathcal{X}_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_t^2) \\ \sigma_t^2 = \gamma_0 + \sum_{i=1}^q \gamma_i \mathcal{X}_{t-i}^2 + \sum_{j=1}^p \delta_j \sigma_{t-j}^2 \end{cases} \quad \forall t \in \mathbb{Z} \quad (2.2)$$

con $\gamma_0 > 0, \gamma_i \geq 0, i = 1, \dots, q$ e $\delta_j \geq 0, j = 1, \dots, p$. Per $p = 0$ il processo si riduce ad un ARCH(q), mentre per $p = q = 0$ diventa semplicemente un *white noise*.

2.1 MODELLO INGARCH(p,q)

In seguito, Heinen (2003) e Ferland *et al.* (2006) hanno introdotto la versione a valori interi del modello GARCH, l'INGARCH(p,q). Ferland *et al.* (2006) lo definiscono come un processo $\{\mathcal{X}_t\}_{t \in \mathbb{Z}}$ tale che:

$$\begin{cases} \mathcal{X}_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t) \\ \lambda_t = \gamma_0 + \sum_{i=1}^q \gamma_i \mathcal{X}_{t-i} + \sum_{j=1}^p \delta_j \lambda_{t-j} \end{cases} \quad (2.3)$$

dove $\gamma_0 > 0, \gamma_i \geq 0, i = 1, \dots, q$ e $\delta_j \geq 0, j = 1, \dots, p$.

Dalla formulazione del processo $\{\mathcal{X}_t\}$ è evidente che la varianza condizionata, che coincide con la media condizionata dato che la distribuzione di riferimento è quella di Poisson, dipende sia dai valori passati della serie che dai propri valori passati.

Per poter definire le condizioni di stazionarietà del processo, si devono prima introdurre due polinomi:

$$D(B) = 1 - \delta_1 B - \dots - \delta_p B^p$$

$$G(B) = \gamma_1 B + \dots + \gamma_q B^q$$

dove B è l'operatore ritardo e $K(B) = D(B) - G(B)$.

Sotto l'assunzione che $D(1) = \sum_{j=1}^p \delta_j < 1$, $D(B)$ ha un'inversa $D^{-1}(B)$ e quindi si può scrivere:

$$\lambda_t = D^{-1}(B)(\gamma_0 + G(B)\mathcal{X}_t) = \gamma_0 D^{-1}(1) + H(B)\mathcal{X}_t \quad (2.4)$$

dove

$$H(B) = G(B)D^{-1}(B) = \sum_{j=1}^{\infty} \psi_j B^j.$$

I coefficienti ψ_j sono dati dallo sviluppo in serie di potenze della funzione $G(B)/D(B)$ nell'intorno di 0.

Fatta questa premessa, Ferland *et al.* (2006) affermano che il processo (2.3) è stazionario del secondo ordine se $K(1) = D(1) - G(1) > 0$ o, in modo equivalente, se $0 \leq \delta_1 + \dots + \delta_p + \gamma_1 + \dots + \gamma_q < 1$.

2.1.1 Convergenza e stazionarietà

Sia $\{\mathcal{U}_t\}_{t \in \mathbb{Z}}$ una successione di variabili di Poisson indipendenti con la stessa media $\psi_0 = \gamma_0/D(1)$ e sia $\{\mathcal{Z}_{t,i,j}\}_{i,j \in \mathbb{N}}$ una successione di variabili di Poisson indipendenti con media ψ_i . Inoltre, si assuma che tutte le variabili appena definite siano mutuamente indipendenti. Allora si definisce la successione $\mathcal{X}_t^{(n)}$ come segue:

$$\mathcal{X}_t^{(n)} = \begin{cases} 0 & n < 0 \\ \mathcal{U}_t & n = 0 \\ \mathcal{U}_t + \sum_{i=1}^n \sum_{j=1}^{\mathcal{X}_{t-i}^{(n-i)}} \mathcal{Z}_{t-i,i,j} & n > 0 \end{cases} \quad (2.5)$$

e per $n \rightarrow \infty$ si dimostra che $\mathcal{X}_t^{(n)}$ ammette limite quasi certo \mathcal{X}_t e che il processo che ne deriva, $\{\mathcal{X}_t\}_{t \in \mathbb{Z}}$, coincide con il (2.3). Inoltre, si dimostra che tale processo è strettamente stazionario poiché per ogni n fissato $\{\mathcal{X}_t^{(n)}\}_{t \in \mathbb{Z}}$ è strettamente stazionario.

Per fare questo, Ferland *et al.* (2006) cominciano col dimostrare che $\mathcal{X}_t^{(n)}$ può essere riscritta attraverso una serie di operazioni di *thinning*: data una variabile casuale non negativa ed a valori interi \mathcal{W} e una successione indipendente di Poisson i.i.d. $\{\mathcal{V}_j\}_{j \in \mathbb{N}}$ con la stessa media ψ , allora $\psi \circ \mathcal{W} = \sum_{j=1}^{\mathcal{W}} \mathcal{V}_j$ se $\mathcal{W} > 0$ e $\psi \circ \mathcal{W} = 0$ altrimenti.

In questo modo, dunque, $\mathcal{X}_t^{(n)}$ può essere anche scritta come:

$$\mathcal{X}_t^{(n)} = \mathcal{U}_t + \sum_{i=1}^n \psi_i^{(t-i)} \circ \mathcal{X}_{t-i}^{(n-i)} \quad n > 0 \quad (2.6)$$

e quindi, ad esempio:

$$\begin{aligned} \mathcal{X}_t^{(1)} &= \mathcal{U}_t + \psi_1^{(t-1)} \circ \mathcal{U}_{t-1} \\ \mathcal{X}_t^{(2)} &= \mathcal{U}_t + \psi_2^{(t-2)} \circ \mathcal{U}_{t-2} + \psi_1^{(t-1)} \circ \left(\mathcal{U}_{t-1} + \psi_1^{(t-2)} \circ \mathcal{U}_{t-2} \right) \\ \mathcal{X}_t^{(3)} &= \mathcal{U}_t + \psi_3^{(t-3)} \circ \mathcal{U}_{t-3} + \psi_2^{(t-2)} \circ \left(\mathcal{U}_{t-2} + \psi_1^{(t-3)} \circ \mathcal{U}_{t-3} \right) \\ &\quad + \psi_1^{(t-1)} \circ \left[\mathcal{U}_{t-1} + \psi_2^{(t-3)} \circ \mathcal{U}_{t-3} + \psi_1^{(t-2)} \circ \left(\mathcal{U}_{t-2} + \psi_1^{(t-3)} \circ \mathcal{U}_{t-3} \right) \right] \end{aligned}$$

e via dicendo, per ogni $n \in \mathbb{N}$.

Sfruttando il fatto che $\mathcal{X}_t^{(n)}$ per come è stata definita è una successione monotona non decrescente di valori interi non negativi, Ferland *et al.* (2006) nella proposizione 2 affermano che tale successione ha un limite quasi certo se $K(1) > 0$ e questo limite è proprio \mathcal{X}_t definito nella (2.3).

Di conseguenza, poiché inizialmente si è affermato che la condizione $K(1) > 0$ garantisce anche la stazionarietà, basta dimostrare che $\mathcal{X}_t^{(n)}$ è stazionario per tutti gli n per poter affermare che \mathcal{X}_t è strettamente stazionario e per farlo si ricorre alla funzione generatrice di probabilità (Ferland *et al.*, 2006, p. 929).

A questo punto, assumendo il verificarsi della condizione $K(1) > 0$, si ha anche che i primi due momenti del processo (2.3) sono finiti. Per ottenere il valore atteso si utilizza la rappresentazione (2.4) del processo, dalla quale è immediato che:

$$\mu = \gamma_0 D^{-1}(1) + \sum_{j=1}^{\infty} \psi_j \mu \Rightarrow \mu = \frac{\gamma_0 D^{-1}(1)}{1 - \sum_{j=1}^{\infty} \psi_j}$$

e dunque:

$$\begin{aligned} E[\mathcal{X}_t] = \mu &= \frac{\gamma_0 D^{-1}(1)}{1 - G(1)D^{-1}(1)} = \frac{\gamma_0}{D(1) - G(1)} = \frac{\gamma_0}{K(1)} \\ &= \frac{\gamma_0}{1 - \left(\sum_{i=1}^q \gamma_i + \sum_{j=1}^p \delta_j \right)}. \end{aligned} \quad (2.7)$$

Invece, per dimostrare che anche il momento secondo è finito, Ferland *et al.* (2006)

applicano la disuguaglianza di Cauchy-Schwarz a $\{\mathcal{X}_t^{(n)}\}_{n \in \mathbb{N}}$ e si ha che:

$$\begin{aligned} E\left[\left(\mathcal{X}_t^{(n)}\right)^2\right] &\leq \sum_{j=1}^n \psi_j E[\mathcal{X}_t] + \left(\sum_{j=1}^n \psi_j\right)^2 E\left[\left(\mathcal{X}_t^{(n)}\right)^2\right] \\ &\quad + 2\gamma_0 D^{-1}(1) \sum_{j=1}^n \psi_j E[\mathcal{X}_t] + E[\mathcal{U}_t^2]. \end{aligned}$$

Pertanto:

$$\begin{aligned} E\left[\left(\mathcal{X}_t^{(n)}\right)^2\right] &\leq \frac{E[\mathcal{X}_t](1 + 2\gamma_0 D^{-1}(1)) \sum_{j=1}^n \psi_j + E[\mathcal{U}_t^2]}{1 - \left(\sum_{j=1}^n \psi_j\right)^2} \\ &\leq \frac{E[\mathcal{X}_t](1 + 2\gamma_0 D^{-1}(1)) \sum_{j=1}^{\infty} \psi_j + E[\mathcal{U}_t^2]}{1 - \left(\sum_{j=1}^{\infty} \psi_j\right)^2} = \mathcal{C}. \end{aligned}$$

Per il teorema della convergenza dominata di Lebesgue si ha che $E[\mathcal{X}_t^2] \leq \mathcal{C}$ e quindi anche il momento secondo è finito.

Infine, se $K(1) > 0$ la successione $\{\mathcal{X}_t^{(n)}\}_{n \in \mathbb{N}}$ converge anche in media quadratica a \mathcal{X}_t , cioè:

$$\lim_{n \rightarrow \infty} E\left[\left(\mathcal{X}_t^{(n)} - \mathcal{X}_t\right)^2\right] = E\left[\lim_{n \rightarrow \infty} \left(\mathcal{X}_t^{(n)} - \mathcal{X}_t\right)^2\right] = 0.$$

2.2 Modello INGARCH(1,1)

Un caso particolare del modello INGARCH(p,q) analizzato nel paragrafo precedente è quello in cui $p = q = 1$. Il modello che ne deriva è un INGARCH(1,1) ed è definibile come:

$$\begin{cases} \mathcal{X}_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t) & \forall t \in \mathbb{Z} \\ \lambda_t = \gamma_0 + \gamma_1 \mathcal{X}_{t-1} + \delta_1 \lambda_{t-1} \end{cases} \quad (2.8)$$

Se $0 \leq \gamma_1 + \delta_1 < 1$ allora tutti i momenti del processo (2.8) esistono e sono finiti, quindi da qui in poi si assume il verificarsi di questa condizione.

Come prima cosa, analogamente a quanto visto per il processo INGARCH(p,q) nella (2.7), si ottiene facilmente che il valore atteso di un INGARCH(1,1) è:

$$E[\mathcal{X}_t] = \mu = \frac{\gamma_0}{1 - (\gamma_1 + \delta_1)}. \quad (2.9)$$

La varianza, invece, è data da:

$$Var[\mathcal{X}_t] = \frac{\mu(1 - (\gamma_1 + \delta_1)^2 + \gamma_1^2)}{1 - (\gamma_1 + \delta_1)^2}. \quad (2.10)$$

Poiché media e varianza sono diverse, si evince che la distribuzione marginale di \mathcal{X}_t non è Poisson. Infine, la funzione di autocovarianza al ritardo k è:

$$\gamma[k] = \frac{\gamma_1(1 - \delta_1(\gamma_1 + \delta_1))(\gamma_1 + \delta_1)^{k-1}\mu}{1 - (\gamma_1 + \delta_1)^2}, \quad \forall k \geq 1 \quad (2.11)$$

mentre l'autocorrelazione di conseguenza è:

$$\rho[k] = \frac{\gamma[k]}{Var[\mathcal{X}_t]} = \frac{\gamma_1(1 - \delta_1(\gamma_1 + \delta_1))(\gamma_1 + \delta_1)^{k-1}}{\gamma_1^2 - (\gamma_1 + \delta_1)^2 + 1}. \quad (2.12)$$

Per stimare i parametri del processo in esame si potrebbe ricorrere al metodo dei minimi quadrati, ma Ferland *et al.* (2006) a questo scopo prendono in considerazione il metodo della massima verosimiglianza condizionata. La procedura è analoga a quella utilizzata per i più classici modelli GARCH che a sua volta porta a risultati abbastanza

simili a quelli per i modelli ARCH, come evidenziato da Bollerslev (1986). Per semplicità si adotta il condizionamento ai valori pre-campione, che non influenza in alcun modo i risultati asintotici (Weiss, 1984).

La funzione di verosimiglianza delle n osservazioni $\mathcal{X}_1, \dots, \mathcal{X}_n$ condizionata ai valori pre-campione è:

$$\mathcal{L}(\Theta) = \prod_{t=1}^n \frac{e^{-\lambda_t} \lambda_t^{\mathcal{X}_t}}{\mathcal{X}_t!}$$

dove:

$$\Theta = (\gamma_0, \dots, \gamma_q, \delta_1, \dots, \delta_p)^T = (\theta_0, \dots, \theta_{p+q})^T$$

$$\lambda_t = \gamma_0 + \sum_{i=1}^q \gamma_i \mathcal{X}_{t-i} + \sum_{j=1}^p \delta_j \lambda_{t-j}.$$

Naturalmente da questa funzione di verosimiglianza non si può arrivare analiticamente a delle stime calcolando le derivate e ponendole uguali a zero; si ottiene, infatti, un sistema di equazioni non-lineari. Per questo, Ferland *et al.* (2006) suggeriscono di ricorrere ad un metodo numerico per trovare il valore ottimo di Θ .

Si comincia col considerare, come al solito, la funzione di log-verosimiglianza:

$$\ell(\Theta) = \sum_{t=1}^n [\mathcal{X}_t \log \lambda_t - \lambda_t - \log(\mathcal{X}_t!)] = \sum_{t=1}^n \ell_t(\Theta).$$

Le derivate prime di $\ell_t(\Theta)$ rispetto a $\theta_i, i = 0, \dots, p + q$ sono:

$$\frac{\partial \ell_t}{\partial \theta_i} = \left(\frac{\partial \lambda_t}{\partial \theta_i} \right) \left(\frac{\mathcal{X}_t}{\lambda_t} - 1 \right) \quad (2.13)$$

mentre le derivate seconde sono:

$$\frac{\partial^2 \ell_t}{\partial \theta_i \partial \theta_j} = \frac{\partial^2 \lambda_t}{\partial \theta_i \partial \theta_j} \left(\frac{\mathcal{X}_t}{\lambda_t} - 1 \right) - \frac{\mathcal{X}_t}{\lambda_t^2} \frac{\partial \lambda_t}{\partial \theta_i} \frac{\partial \lambda_t}{\partial \theta_j} \quad (2.14)$$

per $0 \leq i, j \leq p + q$. In particolare, si ricavano le seguenti derivate parziali:

$$\begin{aligned}\frac{\partial \lambda_t}{\partial \gamma_0} &= 1 + \sum_{j=1}^p \delta_j \frac{\partial \lambda_{t-j}}{\partial \gamma_0}, \\ \frac{\partial \lambda_t}{\partial \gamma_i} &= x_{t-i} + \sum_{j=1}^p \delta_j \frac{\partial \lambda_{t-j}}{\partial \gamma_i}, \quad i = 1, \dots, q, \\ \frac{\partial \lambda_t}{\partial \delta_j} &= \lambda_{t-j} + \sum_{k=1}^p \delta_k \frac{\partial \lambda_{t-k}}{\partial \delta_j}, \quad j = 1, \dots, p.\end{aligned}$$

Per n abbastanza grande la distribuzione approssimata dello stimatore di massima verosimiglianza $\hat{\Theta}$ è $\hat{\Theta} \sim \mathcal{N}(\Theta_0, n^{-1}\mathfrak{F}(\Theta_0)^{-1})$, dove $\mathfrak{F}(\Theta_0)$ è la matrice d'informazione valutata in Θ_0 , il vero valore del parametro Θ .

Ponendo $\mathcal{W}_t^T = (1, x_{t-1}, \dots, x_{t-q}, \lambda_{t-1}, \dots, \lambda_{t-p})$ e indicando con $\nabla \mathcal{G}$ il gradiente di una generica funzione \mathcal{G} si può scrivere:

$$\nabla \lambda_t = \mathcal{W}_t + \sum_{j=1}^p \delta_j \nabla \lambda_{t-j}.$$

Di conseguenza la (2.13) e la (2.14) possono essere rispettivamente riscritte nel modo seguente:

$$\begin{aligned}\nabla \ell_t &= \left(\frac{x_t}{\lambda_t} - 1 \right) \nabla \lambda_t, \\ \mathbf{H}_t &= \left(\frac{x_t}{\lambda_t} - 1 \right) \nabla [\nabla^T \lambda_t] - \frac{x_t}{\lambda_t^2} \nabla \lambda_t \nabla^T \lambda_t.\end{aligned}$$

A questo punto, le matrici:

$$\begin{aligned}\hat{\mathcal{S}}_n &= \frac{1}{n} \sum_{t=1}^n \nabla \ell_t \nabla^T \ell_t, \\ \hat{\mathcal{D}}_n &= -\frac{1}{n} \sum_{t=1}^n \nabla [\nabla^T \ell_t]\end{aligned}$$

rappresentano delle stime consistenti per la matrice d'informazione \mathfrak{F} , pertanto entrambe possono essere utilizzate per stimare la matrice di covarianza asintotica dello stimatore di massima verosimiglianza $\hat{\Theta}$. Ad esempio, White (1982) suggerisce un approccio di stima di quasi-massima verosimiglianza del tipo:

$$E \left[(\hat{\Theta} - \Theta_0)(\hat{\Theta} - \Theta_0)^T \right] \equiv \frac{1}{n} (\hat{\mathcal{D}}_n \hat{\mathcal{S}}_n^{-1} \hat{\mathcal{D}}_n)^{-1}.$$

All'atto pratico il calcolo di queste matrici può risultare scomodo e difficoltoso, di conseguenza si può fare riferimento al metodo bootstrap di stima suggerito da Efron e Tibsharani (1993) per $E \left[(\hat{\Theta} - \Theta_0)(\hat{\Theta} - \Theta_0)^T \right]$, che inoltre risulta essere tanto più appropriato quanto più n è piccolo.

3. PROCESSI INARCH E TEST PER VERIFICARE UN CAMBIAMENTO DEI PARAMETRI IN SERIE STORICHE PER DATI DI CONTEGGIO: APPLICAZIONE E SIMULAZIONE

I processi INARCH (*Integer-valued Autoregressive Conditional Heteroscedasticity*) che verranno presentati in questo capitolo non sono altro che la variante a valori interi dei processi ARCH di Engle (1982) di cui si è già accennato nel capitolo precedente.

Essi, inoltre, sono anche un caso particolare dei processi INGARCH studiati da Ferland *et al.* (2006) e presentati anch'essi nel Capitolo 2; per la precisione, come si vedrà tra poco, è immediato notare che un INGARCH(0,q) corrisponde ad un INARCH(q).

In particolare, per semplicità, ci si soffermerà sul caso $q = 1$, cioè sull'INARCH(1) che, quindi, coincide con un INGARCH(0,1) come sottolineato anche da Weiß (2010a).

Partendo da questo modello e seguendo il lavoro di Franke *et al.* (2012) verrà poi presentato un test statistico che ha come obiettivo quello di verificare se i parametri di una serie storica per dati di conteggio dopo un certo numero di osservazioni cambino. Infine, verranno presentati i risultati di una simulazione relativa a questo modello svolta per valutare l'efficacia del test.

3.1 MODELLO INARCH(1)

Il modello INARCH(1) proposto da Franke *et al.* (2012) consiste in un processo $\{\mathcal{X}_t\}_{t \in \mathbb{Z}}$ definito nel seguente modo:

$$\begin{cases} \mathcal{X}_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t) \\ \lambda_t = \omega + \alpha \mathcal{X}_{t-1} \end{cases} \quad (3.1)$$

dove \mathcal{F}_{t-1} è l'insieme di tutte le osservazioni antecedenti al tempo t ; in pratica la distribuzione dell'osservazione al tempo t condizionata ai valori passati è una Poisson

di parametro variabile, il quale dipende solo dall'osservazione immediatamente precedente.

Poiché la distribuzione condizionata è una Poisson, e dunque media e varianza condizionata corrispondono entrambe a λ_t , in modo analogo a quanto accade nei modelli ARCH si può osservare che questo processo è costruito in modo tale che la varianza condizionata cambi nel tempo in funzione dei valori passati della serie secondo un modello autoregressivo. Questo aspetto permette di modellare l'eteroschedasticità tipica di certe serie storiche; inoltre tale modello è parsimonioso, avendo solo due parametri, ed è anche in grado di modellare la sovradisersione di certi dati, cioè il fatto che la loro variabilità sia maggiore della media.

Il processo (3.1) può anche essere scritto nella forma:

$$\mathcal{X}_t = \omega + \alpha \mathcal{X}_{t-1} + \varepsilon_t, \text{ con } \varepsilon_t = \mathcal{X}_t - \lambda_t,$$

dove i residui ε_t sono *white noise* a media nulla e varianza costante e tra loro a due a due incorrelati, cioè $Cov(\varepsilon_t, \varepsilon_s) = 0 \forall t \neq s$. Inoltre, si ha che $Var[\varepsilon_t] = E[\mathcal{X}_t]$, mentre per quanto riguarda i parametri le condizioni che garantiscono la stazionarietà sono che $\omega > 0$ e $0 \leq \alpha < 1$.

In caso di stazionarietà si verifica facilmente che il valore atteso del processo è:

$$E[\mathcal{X}_t] = \frac{\omega}{1 - \alpha}$$

mentre la varianza è:

$$Var[\mathcal{X}_t] = \frac{\omega}{(1 - \alpha)(1 - \alpha^2)}.$$

Come si può vedere, se $\alpha \neq 0$ il modello presenta sicuramente della sovradisersione in quanto:

$$Var[\mathcal{X}_t] = \frac{\omega}{(1 - \alpha)(1 - \alpha^2)} > \frac{\omega}{1 - \alpha} = E[\mathcal{X}_t]$$

e questa è una caratteristica che può tornare utile, poiché molte serie storiche per dati di conteggio nella pratica presentano questo effetto che può derivare ad esempio da una correlazione positiva tra gli eventi osservati o da una variazione, nel periodo di riferimento, della loro probabilità di verificarsi e via dicendo.

Infine, la funzione di autocorrelazione è la stessa di un processo AR(1) gaussiano con parametro autoregressivo α , cioè:

$$\rho[k] \doteq \text{Corr}(\mathcal{X}_t, \mathcal{X}_{t-k}) = \alpha^k, \quad k \geq 0.$$

Il modello INARCH(1) nella forma proposta da Franke *et al.* (2012) e presa in esame finora rientra in una classe più generale di processi dello stesso tipo in cui la funzione che esprime il parametro della distribuzione condizionata dei dati può assumere altre forme, anche non lineari:

$$\begin{cases} \mathcal{X}_t | \mathcal{F}_{t-1} \sim \mathcal{P}(\lambda_t) \\ \lambda_t = g_\theta(\mathcal{X}_{t-1}) \end{cases} \quad (3.2)$$

dove g_θ è una qualche funzione positiva nota definita in \mathbb{N}_0 che dipende dal parametro $\theta \in \Theta \subset \mathbb{R}^d$. Quest'ultimo processo è possibile scriverlo come $\mathcal{X}_t = g_\theta(\mathcal{X}_{t-1}) + \varepsilon_t$, dove i residui sono *white noise* a media nulla. Inoltre, si assume che la funzione autoregressiva g_θ abbia sempre una costante come primo termine, cioè:

$$g_\theta(x) = \theta_1 + g_{\theta_2, \dots, \theta_d}(x).$$

3.2 TEST PER LA VERIFICA DI UN CAMBIAMENTO DEI PARAMETRI

Il test statistico, proposto da Franke *et al.* (2012), che verrà qua presentato, ha l'obiettivo di verificare se avviene un cambiamento dei parametri che specificano il modello durante il periodo di osservazione dei dati.

Il primo passo nella formulazione del test consiste nel definire due serie storiche stazionarie ed indipendenti tra loro, $\{Y_t\}$ e $\{Y_t^*\}$, che seguano entrambe il modello (3.2)

ma con differenti parametri, rispettivamente $\theta = \theta_0$ per la prima e $\theta = \theta_0^*$ per la seconda. Di conseguenza, la serie storica dei dati osservati X_0, \dots, X_n può essere espressa come:

$$X_t = \begin{cases} Y_t & t \leq k^* \\ Y_t^* & t > k^* \end{cases} \quad t = 0, \dots, n$$

con $1 \leq k^* \leq n$ che rappresenta l'eventuale punto in cui i parametri del modello, per qualche motivo, cambiano.

Se $k^* = n$ naturalmente si ha che $X_t = Y_t \forall t$ e quindi non si presenta alcun cambiamento rilevante nella serie di valori osservati, mentre per $k^* < n$ si ha che il parametro del modello subisce una modifica dal tempo k^* in poi.

Da queste prime considerazioni ne deriva che il sistema d'ipotesi più adatto per il test è del tipo:

$$\begin{cases} H_0: k^* = n \\ H_1: k^* < n \end{cases}$$

cioè sotto l'ipotesi nulla i parametri del modello rimangono gli stessi per tutte le osservazioni, mentre sotto l'ipotesi alternativa c'è un cambiamento di essi.

Per poter fare, in seguito, inferenza sul test è necessario ipotizzare che ci sia un adeguato numero di osservazioni prima e dopo il punto di cambiamento, se ce n'è uno. Pertanto, sotto H_1 si assume che $k^* = \gamma n$ per un qualche $0 < \gamma < 1$.

A questo punto si ha bisogno della stima $\hat{\theta}$ del vettore dei parametri θ , che si ottiene col metodo dei minimi quadrati condizionati:

$$\min_{\theta \in \Theta} Q_n(\theta) = \min_{\theta \in \Theta} \sum_{t=1}^n (X_t - g_\theta(X_{t-1}))^2 = \min_{\theta \in \Theta} \sum_{t=1}^n q_t(\theta).$$

Nel caso specifico del modello (3.1), si tratta semplicemente di risolvere:

$$\min_{0 < \alpha < 1, \omega > 0} \sum_{t=1}^n (\mathcal{X}_t - \omega - \alpha \mathcal{X}_{t-1})^2.$$

Calcolando le derivate parziali rispetto ad α e ω e ponendole uguali a 0 si arriva a:

$$\hat{\alpha} = \frac{\sum_{t=1}^n \mathcal{X}_t \mathcal{X}_{t-1} - \frac{1}{n} \sum_{t=1}^n \mathcal{X}_t \sum_{k=1}^n \mathcal{X}_{k-1}}{\sum_{t=1}^n \mathcal{X}_{t-1}^2 - \frac{1}{n} (\sum_{t=1}^n \mathcal{X}_{t-1})^2}$$

$$\hat{\omega} = \frac{1}{n} \left(\sum_{t=1}^n \mathcal{X}_t - \hat{\alpha} \sum_{t=1}^n \mathcal{X}_{t-1} \right).$$

Dalle espressioni appena ottenute si ricavano infine le stime nella forma più compatta:

$$\hat{\alpha} = \frac{\hat{\gamma}[1]}{\widehat{Var}(\mathcal{X}_t)} = \frac{\hat{\gamma}[1]}{\hat{\gamma}[0]} = \hat{\rho}[1] \quad (3.3)$$

$$\hat{\omega} = (1 - \hat{\alpha}) \bar{\mathcal{X}}_n$$

dove $\hat{\rho}[1]$ è l'autocorrelazione campionaria al ritardo 1, mentre $\bar{\mathcal{X}}_n$ denota la media campionaria delle osservazioni.

Una volta ottenute le stime dei parametri si possono calcolare i residui campionari del modello e la loro somma cumulata, che servirà da base per la statistica test:

$$\hat{\varepsilon}_t = \mathcal{X}_t - g_{\hat{\theta}}(\mathcal{X}_{t-1}) = \mathcal{X}_t - \hat{\lambda}_t = \mathcal{X}_t - (\hat{\omega} + \hat{\alpha} \mathcal{X}_{t-1}) \quad t = 1, \dots, n$$

$$\hat{\mathcal{S}}_n(k) = \sum_{t=1}^k \hat{\varepsilon}_t \quad k = 1, \dots, n. \quad (3.4)$$

La quantità $\hat{\mathcal{S}}_n(k)$ è la somma cumulata (CUSUM) dei residui campionari e, come statistica test, Franke *et al.* (2012) propongono di usare il massimo di questa quantità presa in valore assoluto ed opportunamente normalizzata:

$$\mathcal{J}_n = \max_{1 \leq k < n} \sqrt{\frac{n}{k(n-k)}} |\hat{\mathcal{S}}_n(k)|. \quad (3.5)$$

Questa statistica test converge quasi certamente ad infinito, ma usando fattori di normalizzazione appropriati si può arrivare ad una convergenza alla distribuzione dei valori estremi di Gumbel, come si vedrà più avanti. Essendo, però, la convergenza lenta si possono prendere in considerazione anche altre statistiche test collegate direttamente alla (3.5) come, ad esempio:

$$\mathcal{T}_{n,2} = \max_{\tau n \leq k \leq n - \tau n} \sqrt{\frac{n}{k(n-k)}} |\hat{\mathcal{S}}_n(k)|, \quad (3.6)$$

$$\mathcal{T}_{n,3} = \max_{1 \leq k < n} \left(\frac{n^2}{k(n-k)} \right)^\beta \frac{1}{\sqrt{n}} |\hat{\mathcal{S}}_n(k)| \quad (3.7)$$

con $\tau > 0$ e $0 \leq \beta < 1/2$. Queste due statistiche convergono in distribuzione rispettivamente al $\sup_{\tau < t < 1 - \tau} |B(t)|$ e al $\sup_{0 < t < 1} |B(t)| / (t(1-t))^\beta$, con $\{B(t), 0 \leq t \leq 1\}$ che è un ponte Browniano (Csörgö & Horváth, 1997).

Nel resto del capitolo però, sia per l'inferenza sul test che per svolgere la simulazione, si farà riferimento solo alla statistica \mathcal{T}_n così come presentata nella (3.5).

3.2.1 Inferenza sulle stime dei parametri del modello INARCH(1)

Weiß (2010b) ha dimostrato la normalità asintotica e la consistenza delle stime CLS (3.3) del modello (3.1). Nello specifico, egli afferma che la distribuzione asintotica degli stimatori è data da:

$$\sqrt{n}(\hat{\omega}_n - \omega, \hat{\alpha}_n - \alpha)^T \xrightarrow{D} \mathcal{N}(0, \Sigma_{\omega, \alpha})$$

dove \xrightarrow{D} indica la convergenza in distribuzione e $\Sigma_{\omega, \alpha}$ corrisponde a:

$$\Sigma_{\omega, \alpha} = \begin{pmatrix} \frac{\omega}{1-\alpha} \left(\omega(1+\alpha) + \frac{1+2\alpha^4}{1+\alpha+\alpha^2} \right) & -\omega(1+\alpha) - \frac{(1+2\alpha)\alpha^3}{1+\alpha+\alpha^2} \\ -\omega(1+\alpha) - \frac{(1+2\alpha)\alpha^3}{1+\alpha+\alpha^2} & (1-\alpha^2) \left(1 + \frac{\alpha(1+2\alpha^2)}{\omega(1+\alpha+\alpha^2)} \right) \end{pmatrix}.$$

Questo risultato permette anche di avere un'approssimazione asintotica degli errori standard degli stimatori semplicemente inserendo i valori numerici delle stime di ω e α nella matrice.

Franke *et al.* (2012) hanno generalizzato questi risultati al modello (3.2), estendendoli al caso in cui sia presente un punto di cambiamento nel campione. Infatti, sotto H_1 , la stima $\hat{\theta}$ non converge né a θ_0 né a θ_0^* , bensì ad un $\tilde{\theta}_0$ definito come segue: per prima cosa, la quantità $q_t(\theta) = (\mathcal{X}_t - g_\theta(\mathcal{X}_{t-1}))^2$ di $Q_n(\theta)$ viene leggermente modificata

$$\tilde{q}_t(\theta) \begin{cases} (Y_t - g_\theta(Y_{t-1}))^2 & t \leq [\gamma n] \\ (Y_t^* - g_\theta(Y_{t-1}^*))^2 & t > [\gamma n] \end{cases}$$

e si pone $\tilde{Q}_n(\theta) = \sum_{t=1}^n \tilde{q}_t(\theta)$.

Sotto H_0 , poiché $\mathcal{X}_t = Y_t \forall t$ come già visto in precedenza, si ha che $q_t(\theta) = \tilde{q}_t(\theta)$ per ogni t . Nell'altro caso, invece, $q_t(\theta)$ e $\tilde{q}_t(\theta)$ differiscono solo in corrispondenza del punto di cambiamento k^* cosicché la differenza tra $Q_n(\theta)$ e $\tilde{Q}_n(\theta)$ diventi trascurabile al crescere della numerosità campionaria n .

Fatta questa premessa, $\tilde{\theta}_0$ viene definito da Franke *et al.* (2012) come la quantità che minimizza:

$$e(\theta) = \gamma E[\tilde{q}_1(\theta)] + (1 - \gamma)E[\tilde{q}_n(\theta)].$$

Si assume che il $\tilde{\theta}_0 \in \Theta$ tale per cui $e(\tilde{\theta}_0) = \min_{\theta \in \Theta} e(\theta)$ esista, sia unico e giaccia nella parte interna di Θ .

Per dimostrare la convergenza di $\hat{\theta}_n$ a $\tilde{\theta}_0$ è anche necessario assumere che la funzione autoregressiva $g_\theta(x)$ sia una contrazione, ossia $\forall \theta \in \Theta$ $g_\theta(x)$ è uniformemente lipschitziana in x per una costante $L_\theta < 1$:

$$|g_\theta(x) - g_\theta(x')| \leq L_\theta |x - x'| \quad \forall x, x' \in \mathbb{N}_0.$$

Questa caratteristica garantisce anche l'esistenza di una soluzione stazionaria per il modello (3.2), mentre nel caso più semplice del modello (3.1) è equivalente ad assumere che $\alpha < 1$.

Infine, la funzione $g_\theta(x)$ dev'essere regolare e dunque, indicando con ∇ e ∇^2 rispettivamente il gradiente e l'hessiano rispetto a θ della funzione, si assume che:

- $g_\theta(x)$ sia due volte continua e differenziabile rispetto a $\theta \forall x \in \mathbb{N}_0$,
- se $\{Y_t\}$ è una soluzione stazionaria per (3.2) con $\theta = \theta_0$ o $\theta = \theta_0^*$ allora:

$$E \left[\sup_{\theta \in \Theta} \|\nabla g_\theta(Y_t) \nabla^T g_\theta(Y_t)\| \right] < \infty, \quad E \left[Y_t \sup_{\theta \in \Theta} \|\nabla^2 g_\theta(Y_{t-1})\| \right] < \infty.$$

Si noti che, nel caso del modello (3.1), la prima condizione è sempre verificata – essendo la funzione $g_\theta(x) = \omega + \alpha \mathcal{X}_{t-1}$ un polinomio di primo grado - mentre la seconda è automaticamente soddisfatta se $\alpha < 1$.

Se le assunzioni fatte finora sono tutte soddisfatte, allora, come affermato nel teorema 1 da Franke *et. al* (2012), lo stimatore CLS $\hat{\theta}_n$ di $\tilde{\theta}_0$ è consistente in senso forte:

$$\hat{\theta}_n \xrightarrow{q.c.} \tilde{\theta}_0 \text{ per } n \rightarrow \infty.$$

Inoltre, se si assume che l'hessiano $\nabla^2 e(\tilde{\theta}_0)$ è definito positivo, allora $\hat{\theta}_n$ è anche asintoticamente normale:

$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_0) \xrightarrow{D} \mathcal{N}(0, A^{-1}BA^{-1}) \text{ per } n \rightarrow \infty$$

dove $A = \nabla^2 e(\tilde{\theta}_0)$ e $B = \lim_{n \rightarrow \infty} \frac{1}{n} E[\nabla \tilde{Q}_n(\tilde{\theta}_0) \nabla^T \tilde{Q}_n(\tilde{\theta}_0)]$, con $\nabla \tilde{Q}_n(\theta)$ che denota il gradiente di $\tilde{Q}_n(\theta)$ rispetto a θ .

Anche quest'ultima assunzione è automaticamente soddisfatta se si considera il modello (3.1), infatti in tal caso:

$$A = 2\gamma \begin{pmatrix} 1 & E[Y_0] \\ E[Y_0] & E[(Y_0)^2] \end{pmatrix} + 2(1 - \gamma) \begin{pmatrix} 1 & E[Y_{n-1}^*] \\ E[Y_{n-1}^*] & E[(Y_{n-1}^*)^2] \end{pmatrix}$$

e il determinante è:

$$\begin{aligned} \det(A) &= 2\gamma[E[(Y_0)^2] - (E[Y_0])^2] + 2(1 - \gamma)[E[(Y_{n-1}^*)^2] - (E[Y_{n-1}^*])^2] \\ &= 2\gamma \text{Var}[Y_0] + 2(1 - \gamma) \text{Var}[Y_{n-1}^*] > 0. \end{aligned}$$

Quindi A è definita positiva.

3.2.2 La distribuzione asintotica della statistica test sotto le ipotesi

La distribuzione asintotica della statistica test \mathcal{T}_n (3.5) sotto l'ipotesi nulla di nessun cambiamento dei parametri permette di derivare i valori critici, in corrispondenza di un certo livello di confidenza, coi quali si può decidere se accettare o meno H_0 .

Serve però un'ulteriore condizione e cioè che il momento centrale di ordine $(2 + \nu)$ di $\nabla g_{\theta_0}(Y_t)$ esista per un qualche $\nu > 0$. Nel caso del modello (3.1) con $\alpha < 1$, poiché $\nabla g_{\theta_0}(Y_t) = (1, Y_t)^T$ e tutti i momenti di Y_t esistono, la condizione è verificata.

A questo punto è possibile ottenere la distribuzione asintotica di \mathcal{T}_n . Nel teorema 3, Franke *et al.* (2012) affermano che sotto H_0 , cioè se $\mathcal{X}_t = Y_t \forall t \in [0, n]$, allora:

$$\mathcal{P} \left(a(\log n) \frac{\mathcal{T}_n}{\tau} - b(\log n) \leq x \right) \rightarrow \exp(-2e^{-x}) \quad \text{per } n \rightarrow \infty,$$

dove le quantità $a(\cdot)$ e $b(\cdot)$ in forma estesa corrispondono a:

$$a(u) = \sqrt{2 \log u}, \quad b(u) = 2 \log u + \frac{1}{2} \log \log u - \frac{1}{2} \log \pi$$

e τ^2 è la varianza dei residui. Essendo però ignota, può essere sostituita con la varianza campionaria dei residui campionari:

$$\hat{\tau}^2 = \frac{1}{n - d} \sum_{t=1}^n \hat{\varepsilon}_t^2.$$

Nel caso del modello (3.1) $d = 2$ poiché esso indica il numero di parametri del modello.

Il comportamento asintotico della statistica test sotto l'ipotesi alternativa, invece, permette di valutare la potenza del test, ossia la sua capacità di rigettare H_0 quando essa è falsa.

Bisogna considerare che nel caso in cui si verifichi un cambiamento dei parametri del modello i residui stimati assumeranno inevitabilmente dei valori troppo alti o troppo bassi, che testimoniano un cattivo adattamento delle stime ai valori osservati. In pratica si commette l'errore di stimare il modello da $t = 0$ a $t = n$ con un unico parametro $\hat{\theta}$, quando invece da $t = 0$ a $t = k^*$ il parametro della serie è θ_0 e da $t = k^* + 1$ a $t = n$ è θ_0^* . I grandi scostamenti tra valori reali e stimati si ripercuotono così anche sulla statistica test \mathcal{T}_n che è basata sui residui stimati. Si può quindi giungere alla conclusione che a valori troppo alti di \mathcal{T}_n , oltre una certa soglia C , dovrebbe coincidere il rifiuto dell'ipotesi nulla.

Lo stesso ragionamento può essere applicato alla media del processo, che essendo funzione dei parametri cambia al tempo k^* .

Sotto H_1 infatti i parametri sono $\theta_0 = (\omega, \alpha)$ prima del cambiamento e $\theta_0^* = (\omega^*, \alpha^*)$ dopo e dunque si ha:

$$E[Y_t] = \frac{\omega}{1 - \alpha} \neq \frac{\omega^*}{1 - \alpha^*} = E[Y_t^*]$$

e di conseguenza anche la differenza tra il valore atteso del processo e quello del modello stimato è utile ad evidenziare la presenza di un punto di cambiamento, in quanto tale differenza sarà un valore non nullo come evidenziato da Franke *et al.*

(2012):

$$|E[Y_t] - E[g_{\hat{\theta}_0}(Y_{t-1})]| = |E[\omega_0 + \alpha_0 Y_{t-1}] - E[\tilde{\omega}_0 + \tilde{\alpha}_0 Y_{t-1}]| = C > 0.$$

Franke *et al.* (2012), quindi, nel teorema 4 affermano che sotto H_1 e $\forall c > 0$ si ha che:

$$\mathcal{P}\left(a(\log n)\frac{\mathcal{J}_n}{\tau} - b(\log n) \geq c\right) \rightarrow 1 \quad \text{per } n \rightarrow \infty.$$

Infine, sotto l'ipotesi alternativa, si può ottenere anche la stima del punto k^* ; nel corollario al teorema 4, si sostiene che la stima consistente per k^* corrisponde al punto in cui la somma cumulativa dei residui stimati, in valore assoluto, raggiunge il suo massimo:

$$\frac{\hat{k}^*}{n} \rightarrow \gamma \quad \text{con } \hat{k}^* = \arg \max_{1 \leq k < n} |\hat{\mathcal{S}}_n(k)|, \quad \text{ossia } |\hat{\mathcal{S}}_n(\hat{k}^*)| = \max_{1 \leq k < n} |\hat{\mathcal{S}}_n(k)|.$$

Il valore della soglia C oltre la quale si è portati a rifiutare H_0 è definito come:

$$C = \hat{\tau} \frac{\log 2 - \log\left(\log \frac{1}{1-\varphi}\right) + b(\log n)}{a(\log n)},$$

dove $0 < \varphi < 1$ è il livello di confidenza scelto per il test; cioè si rigetta H_0 ad un livello φ se $\mathcal{J}_n > C$.

Franke *et al.* (2012) hanno applicato il modello (3.1) ed il relativo test (3.5) a dei dati clinici sugli attacchi epilettici e gli effetti di un farmaco contro di essi.

Sono stati presi in considerazione due gruppi di pazienti: nel primo si è tenuto nota del numero di attacchi registrati fino ad un certo giorno a partire dal quale è poi stato impiegato il nuovo farmaco, mentre nel secondo gruppo il farmaco è stato sostituito da un placebo. L'obiettivo dello studio era, quindi, quello di stabilire se l'introduzione della nuova cura avesse un effetto sul numero medio di attacchi epilettici o meno.

Inoltre è noto che gli attacchi epilettici, mettendo sotto tensione alcune aree del cervello, rendono più probabile il ripetersi di crisi in successione. Pertanto è plausibile assumere una correlazione positiva tra i conteggi \mathcal{X}_t e \mathcal{X}_{t-1} e il modello (3.1) soddisfa

questa caratteristica. Poiché il farmaco in prova ha anche lo scopo di accelerare il ristabilimento di queste aree cerebrali, cioè di ridurre questa dipendenza tra le osservazioni, ci si aspetta una riduzione del parametro autoregressivo α .

I risultati dello studio presentati da Franke *et al.* (2012) mostrano che nel caso del gruppo sottoposto al placebo il test porta, correttamente, a non rifiutare l'ipotesi nulla, mentre nel caso dell'altro gruppo la statistica test raggiunge un massimo ben oltre la soglia, in corrispondenza di un certo istante \hat{k}^* , e si osserva una significativa diminuzione delle stime di entrambi i parametri per le osservazioni a partire da quell'istante in poi. La diminuzione di ω indica una riduzione della frequenza complessiva di attacchi epilettici, mentre la diminuzione di α , come già detto, evidenzia la riduzione della probabilità che una crisi ne provochi altre nell'immediato futuro.

Nel caso di pazienti con crisi epilettiche rare fin dall'inizio, però, è più difficile mostrare un'ulteriore riduzione causata dal farmaco, mentre per altri pazienti l'istante \hat{k}^* e dunque anche il cambio dei parametri non coincide esattamente col giorno in cui si è cominciato a somministrare la nuova cura, bensì è spostato in avanti, segno che gli effetti si sono manifestati in ritardo su alcuni individui.

3.3 Simulazione

In questo paragrafo verranno presentati i risultati della simulazione svolta per valutare l'efficacia del test.

Per prima cosa, utilizzando il software statistico R, è stato preparato un programma (riportato in Appendice) che simula un processo INARCH(1) (3.1) ed il test \mathcal{T}_n (3.5) in diverse situazioni. Nello specifico sono stati simulati dei processi sia sotto H_0 , quindi con un'unica coppia di parametri (ω, α) , sia sotto H_1 , quindi con una coppia di

parametri (ω_1, α_1) per $t \leq k^*$ e (ω_2, α_2) per $t > k^*$, $t = 1, \dots, n$ e $1 < k^* < n$. I parametri rispettano i vincoli già specificati all'inizio del capitolo: $\omega > 0, 0 < \alpha < 1$.

In entrambi i casi sono state simulate serie storiche con una diversa numerosità $n = 100, 200, 500, 1.000$ e con diverse combinazioni di parametri. Per ogni coppia di parametri abbinata ad una certa numerosità, la relativa serie storica è stata simulata 5.000 volte sotto H_0 e 2.000 volte sotto H_1 .

Inoltre, sotto l'ipotesi alternativa e per ogni numerosità, sono stati considerati due diversi casi a seconda della posizione del punto k^* nella serie storica, inserito rispettivamente in corrispondenza di $n/2$ e $3n/4$.

Nella seguente tabella sono riepilogati i risultati della simulazione sotto H_0 :

Tabella 3.1: Percentuale delle volte che si rifiuta H_0 quando le serie sono generate sotto H_0 .
Numero di replicazioni: 5000.

(ω, α)	N=100	N=200	N=500	N=1000
(0.5, 0.15)	1,16%	2,10%	2,52%	3,10%
(0.5, 0.4)	1,28%	2,22%	3,10%	3,22%
(0.5, 0.7)	2,10%	3,20%	4,30%	4,86%
(0.5, 0.9)	6,18%	7,18%	9,04%	10,10%
(1.0, 0.15)	0,70%	1,18%	1,40%	1,64%
(1.0, 0.4)	0,78%	1,64%	1,76%	2,30%
(1.0, 0.7)	0,52%	1,06%	2,14%	2,86%
(1.0, 0.9)	4,88%	5,56%	6,72%	7,32%

Il livello di confidenza usato nel test è del 5% e i valori in tabella rappresentano il numero di volte in cui la statistica test \mathcal{T}_n ha superato la soglia \mathcal{C} , in forma percentuale sul totale delle 5.000 serie simulate, cioè il numero di volte in cui il test porterebbe a rifiutare erroneamente H_0 .

In generale, le percentuali tendono a crescere al crescere di n in tutti i casi presi in esame e in particolare si nota che, dato ω , le percentuali crescono al crescere del parametro autoregressivo α .

Inoltre, per $\omega = 1$ si registrano valori più bassi indipendentemente da α rispetto ai casi in cui $\omega = 0.5$.

Infine, è da sottolineare come le uniche due situazioni in cui le percentuali superano il 5% sono quelle in cui $\alpha = 0.9$, dove si arriva ad un massimo del 10%.

Nelle figure di seguito vengono riportati come esempio due casi relativi alla combinazione $(1, 0.9)$ con $n = 500$:

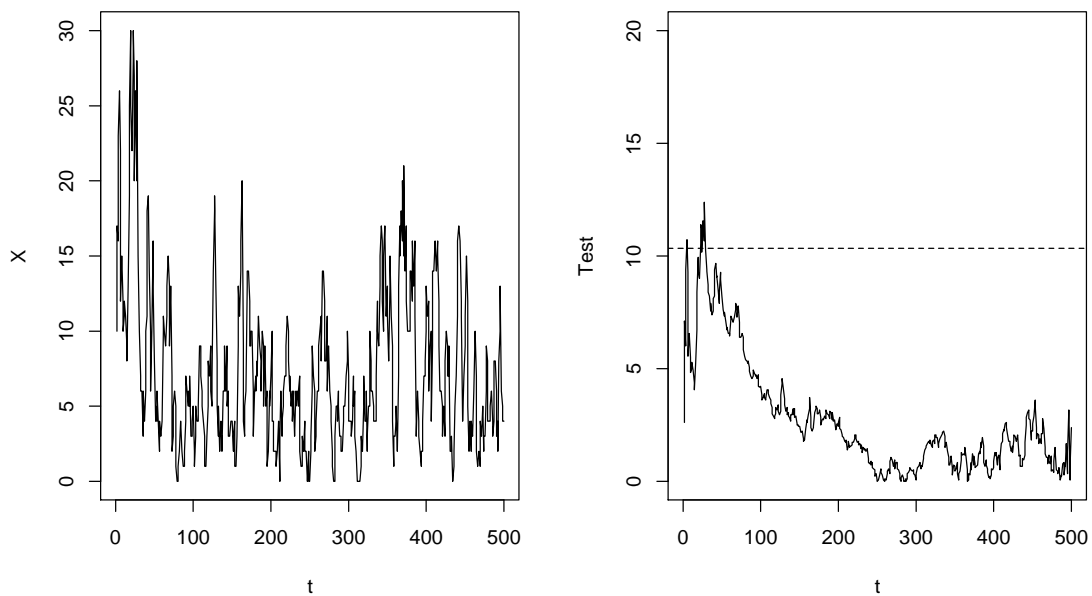


Figura 3.1: INARCH(1) con parametri $(1, 0.9)$

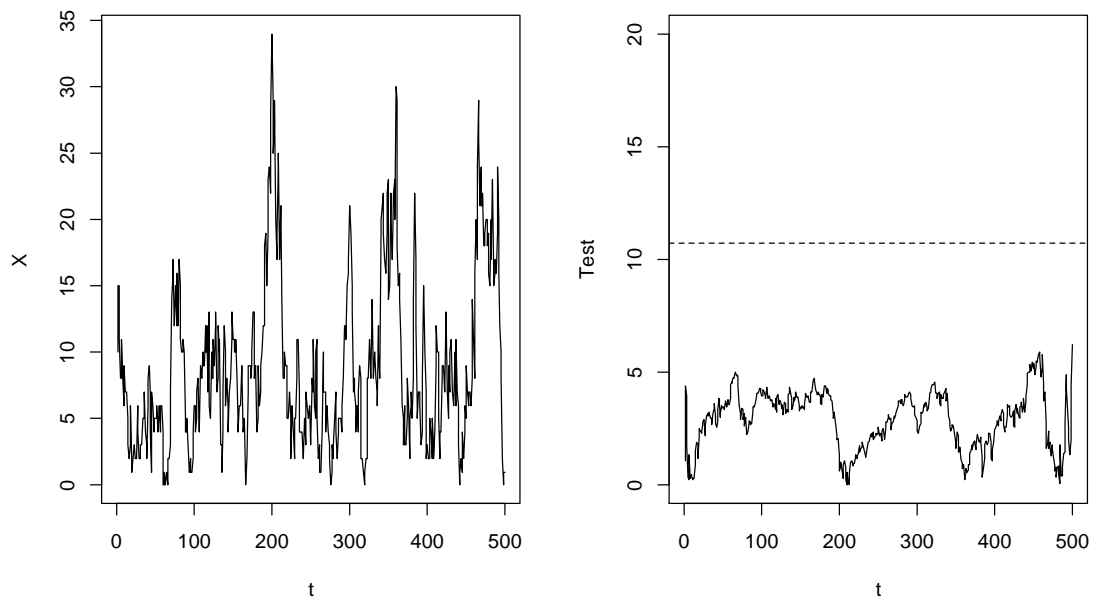


Figura 3.2: INARCH(1) con parametri (1, 0.9)

In entrambe le figure, a sinistra si trova una delle serie storiche simulate mentre a destra è stato rappresentato l'andamento dei relativi valori della statistica test assieme al valore assunto dalla soglia C rappresentato dalla linea tratteggiata orizzontale. Come si può vedere, la figura 3.1 rappresenta uno dei casi in cui la statistica \mathcal{T}_n porta a rifiutare non correttamente H_0 in quanto ha un massimo che oltrepassa la soglia C , mentre la figura 3.2 è relativa ad uno dei casi in cui si è portati giustamente ad accettare H_0 poiché il massimo di \mathcal{T}_n sta sotto la soglia.

Nel complesso, quindi, il test può essere giudicato efficace nella maggior parte dei casi riepilogati nella tabella 3.1, compreso naturalmente il caso (1, 0.15) preso in esame anche da Franke *et al.* (2012) nella loro simulazione. Al momento, però, non può essere ritenuto perfetto in quanto esistono alcune situazioni in cui i risultati ottenuti non sono quelli sperati e inoltre sembra essere fin troppo conservativo in molti casi.

A questo punto, si passa ad analizzare i risultati ottenuti dalla simulazione sotto H_1 riepilogati nella tabella 3.2:

Tabella 3.2: Percentuale delle volte che si rifiuta H_0 quando le serie sono generate sotto H_1 . Numero di replicazioni: 2000.

(ω_1, α_1)	(ω_2, α_2)	N=100		N=200	
		K=50	K=75	K=100	K=150
(0.5, 0.15)	(1.0, 0.7)	2,70%	20,00%	32,00%	75,85%
(0.5, 0.3)	(2.0, 0.4)	19,10%	32,40%	98,95%	99,75%
(0.5, 0.7)	(1.0, 0.4)	1,60%	1,60%	3,80%	3,20%
(0.5, 0.7)	(2.0, 0.4)	1,50%	1,15%	31,65%	14,90%
(1.0, 0.2)	(0.6, 0.8)	2,55%	9,15%	8,10%	38,00%
(1.0, 0.5)	(0.3, 0.15)	19,90%	2,40%	90,85%	39,15%
(1.0, 0.7)	(0.3, 0.4)	5,60%	1,30%	43,70%	6,75%
(1.0, 0.9)	(2.0, 0.3)	10,55%	3,50%	15,70%	4,10%
(1.0, 0.9)	(4.0, 0.3)	5,80%	3,05%	15,20%	3,65%

(ω_1, α_1)	(ω_2, α_2)	N=500		N=1000	
		K=250	K=375	K=500	K=750
(0.5, 0.15)	(1.0, 0.7)	99,75%	99,80%	100,00%	100,00%
(0.5, 0.3)	(2.0, 0.4)	100,00%	100,00%	100,00%	100,00%
(0.5, 0.7)	(1.0, 0.4)	6,65%	4,70%	9,60%	5,45%
(0.5, 0.7)	(2.0, 0.4)	88,90%	65,55%	99,80%	97,60%
(1.0, 0.2)	(0.6, 0.8)	48,55%	86,90%	96,95%	99,40%
(1.0, 0.5)	(0.3, 0.15)	100,00%	100,00%	100,00%	100,00%
(1.0, 0.7)	(0.3, 0.4)	99,80%	91,90%	100,00%	100,00%
(1.0, 0.9)	(2.0, 0.3)	50,50%	8,30%	96,95%	39,00%
(1.0, 0.9)	(4.0, 0.3)	33,15%	6,00%	67,30%	10,15%

Anche in questo caso le percentuali rappresentano il numero di volte in cui la statistica test \mathcal{T}_n ha un massimo che supera la soglia C ; situazione nella quale si è portati a rifiutare H_0 . Dunque, stavolta, ci si aspettano dei valori il più possibile alti, prossimi al 100% poiché le serie storiche sono state simulate con un cambiamento nei parametri dal punto k^* in poi.

Come si può vedere, anche in questo caso le percentuali crescono all'aumentare della numerosità campionaria e per $n = 1.000$ in quasi tutti i casi il test si dimostra affidabile, ma è necessario approfondire l'analisi.

Per prima cosa, come ci si poteva aspettare, è più difficile per il test cogliere il cambiamento dei parametri se questo avviene a $3n/4$ rispetto a $n/2$ per il fatto che nel primo caso ci sono poche osservazioni dopo il punto k^* . In generale, si può dire che la potenza del test dipende dalla posizione del punto k^* .

Infatti, in sei dei nove casi presi in esame, per le numerosità più basse, si nota come le percentuali in corrispondenza di $k^* = n/2$ siano maggiori di quelle di $k^* = 3n/4$. Però le differenze si fanno progressivamente meno marcate ed una volta arrivati a $n = 1.000$, o anche $n = 500$ in alcuni casi, si raggiunge il 100% in entrambe le situazioni.

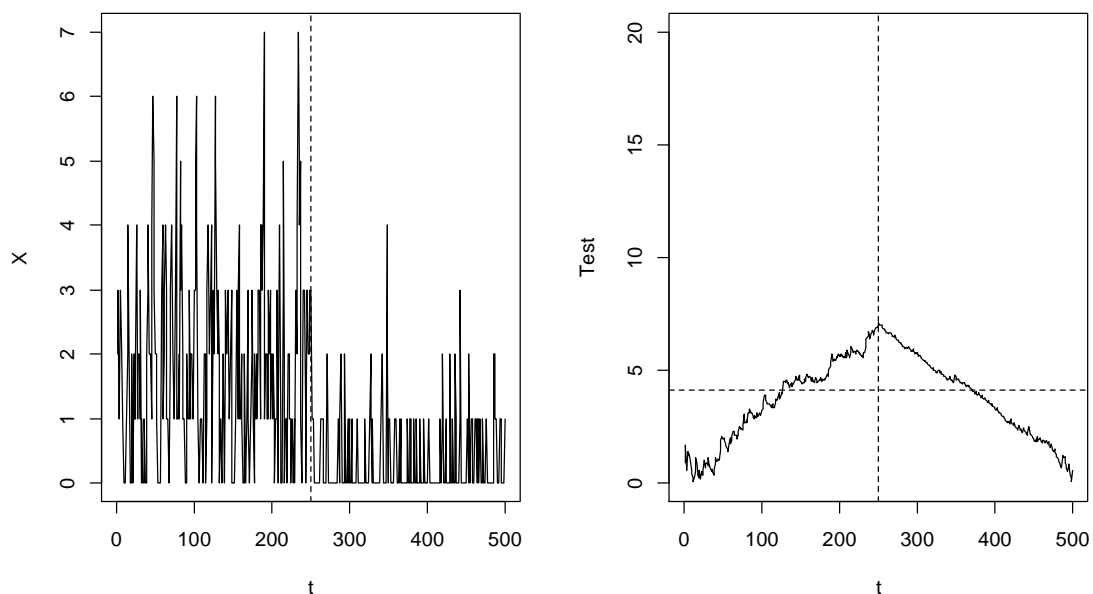


Figura 3.3: INARCH(1) con parametri $(1, 0.5)$, $(0.3, 0.15)$ e $k^* = 250$

Nella figura 3.3 a sinistra è rappresentata una serie simulata da un modello INARCH(1) di parametri $(1, 0.5)$ prima del cambiamento e $(0.3, 0.15)$ dopo il cambiamento avvenuto a metà campione (di grandezza $n = 500$) ed indicato dalla linea tratteggiata verticale, mentre a destra si trova il test \mathcal{T}_n , la soglia C rappresentata dalla linea tratteggiata orizzontale e il punto k^* rappresentato dalla linea tratteggiata verticale. Come si può vedere, il test raggiunge il suo massimo proprio in corrispondenza di $n = 250$ ed ha un valore maggiore della soglia, quindi il test si comporta alla perfezione.

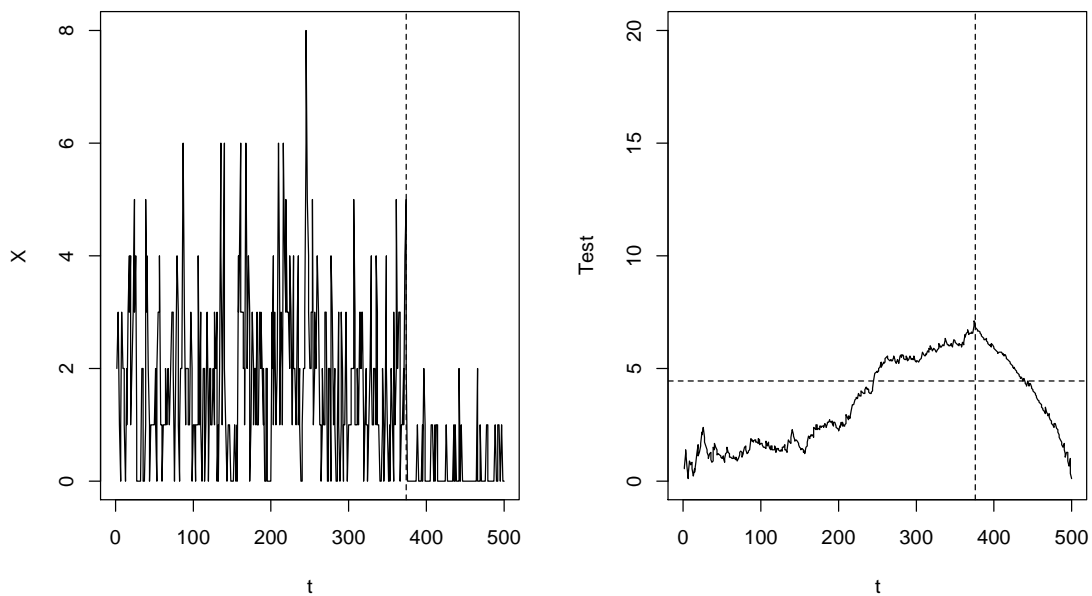


Figura 3.4: INARCH(1) con parametri $(1, 0.5)$, $(0.3, 0.15)$ e $k^* = 375$

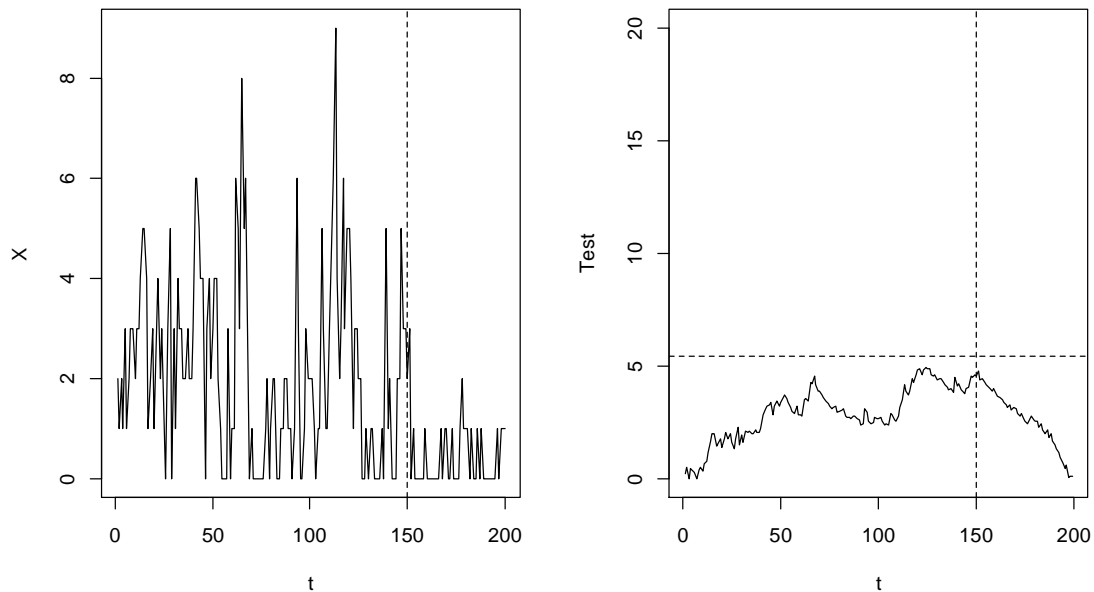


Figura 3.5: INARCH(1) con parametri $(1, 0.5)$, $(0.3, 0.15)$ e $k^* = 150$

Anche nella figura 3.4, in cui il cambiamento avviene a $3/4$ del campione, il test funziona, mentre nella figura 3.5 è rappresentato un caso in cui con $n = 200$ e $k^* = 150$ il test porta erroneamente a rifiutare l'ipotesi alternativa

Negli altri tre casi (per la precisione nel primo, secondo e quinto), in corrispondenza di numerosità basse, si verifica invece il caso opposto, cioè le percentuali per $k^* = 3n/4$ sono maggiori di quelle per $k^* = n/2$, ma una volta raggiunta la numerosità massima il problema sparisce ed entrambe sono al 100% (a parte nel quinto caso in cui si nota ancora una leggera differenza: 96.95% contro 99.40%).

Nella figura 3.6 è rappresentato proprio il quinto caso, cioè quello con parametri $(1, 0.2)$ prima e $(0.6, 0.8)$ dopo il punto k^* a metà campione e si vede che porta a rifiutare l'ipotesi alternativa.

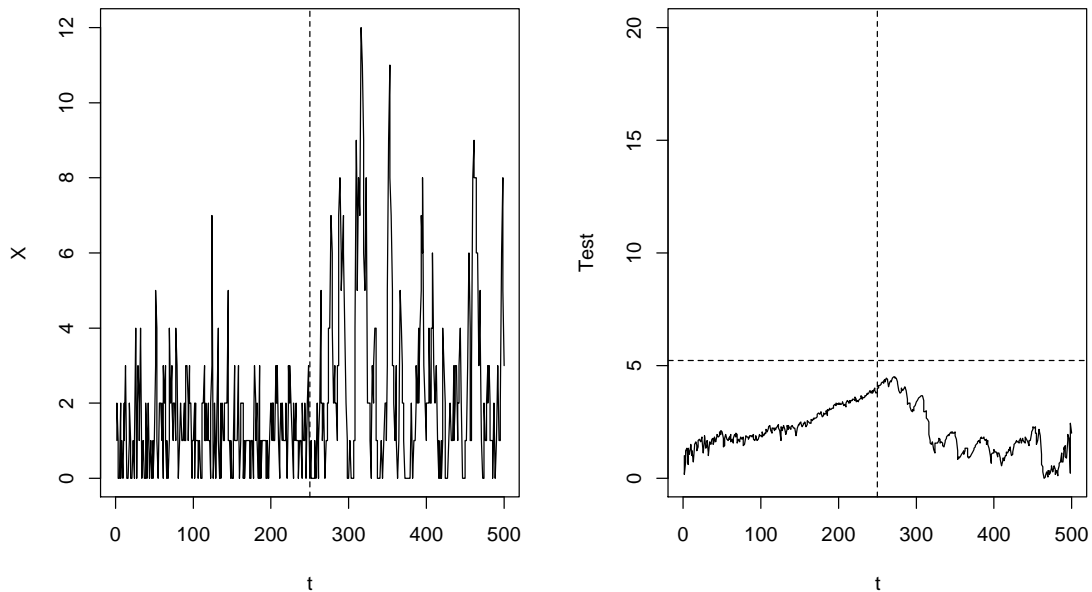


Figura 3.6: INARCH(1) con parametri $(1, 0.2)$, $(0.6, 0.8)$ e $k^* = 250$

L'altro aspetto ancora più particolare è che nel terzo e nono caso (per entrambi i k^*) e nell'ottavo (solo per $k^* = 3n/4$) il test converge a 1 molto lentamente, infatti in corrispondenza di una numerosità pari a 1.000 le percentuali sono ancora molto basse, soprattutto nel terzo caso. Questa situazione potrebbe essere dovuta alla persistenza, cioè al fatto che, con un α_1 già abbastanza alto e vicino a ω_2 , la serie storica presenta un trend crescente che nelle vicinanze di k^* l'ha già portata vicino al nuovo livello e, dunque, il test non riesce a cogliere il cambio dei parametri.

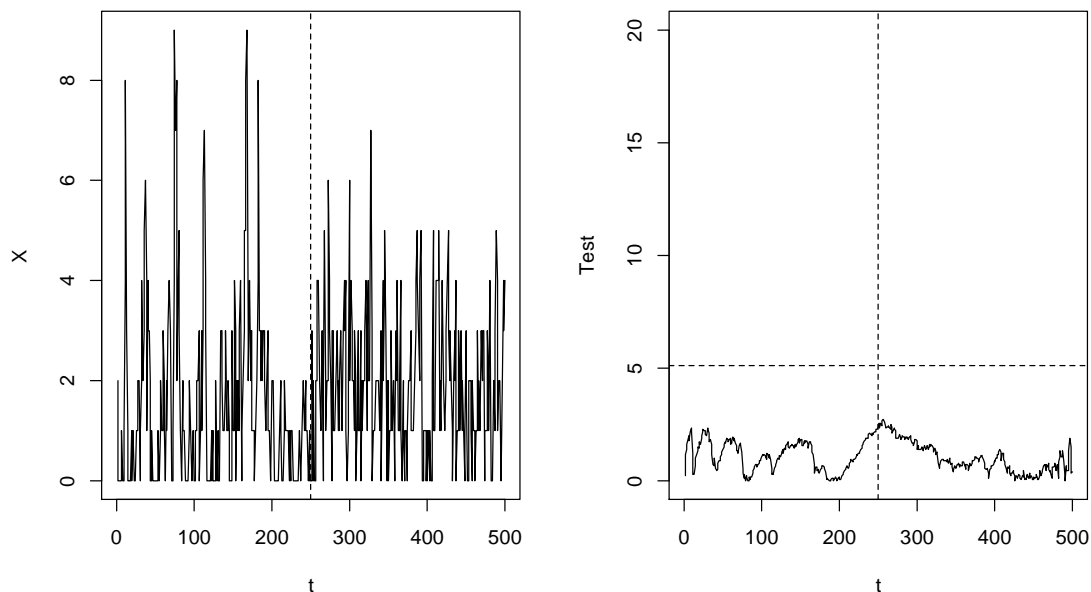


Figura 3.7: INARCH(1) con parametri $(0.5, 0.7)$, $(1, 0.4)$ e $k^* = 250$

Nella figura 3.7 viene rappresentato proprio il terzo caso, quello dalle percentuali più basse, con $n = 500$ e cambiamento a metà campione. Il test porta a rifiutare l'ipotesi alternativa.

Per studiare meglio quest'aspetto, poiché il caso più particolare è $(0.5, 0.7)$, $(1, 0.4)$, è stata simulata una serie storica di parametri $(0.5, 0.7)$, $(2, 0.4)$, ossia col solo parametro ω_2 diverso dal caso precedente e nello specifico più "distante" da α_1 . Il risultato ottenuto, riportato nella quarta riga della tabella 3.2, pare confermare la tesi esposta sopra poiché solo per $n = 100$ i risultati sono pressoché simili, mentre per $n = 1.000$ il test si dimostra efficace, con percentuali prossime al 100%.

Nel complesso, anche sotto H_1 , il test risulta essere efficace nella maggior parte dei casi presi in esame, in particolare quando la numerosità è abbastanza alta. Inoltre, i

risultati del caso $(1, 0.5)$, $(0.3, 0.15)$ sono in linea con quelli ottenuti da Franke *et al.* (2012) nella loro simulazione.

Si noti che i risultati più distanti da quelli attesi, sia sotto H_0 che sotto H_1 , si hanno per le combinazioni che presentano un α elevato, mentre esclusivamente sotto l'ipotesi alternativa il problema delle percentuali più alte per $k^* = 3n/4$ rispetto a $k^* = n/2$ si presenta nei casi con gli α più bassi.

Per concludere, gli autori stessi del test, in chiusura del loro articolo, affermano che eventuali risultati non conformi alle aspettative non sono per nulla inaspettati e sono da imputare principalmente a due aspetti:

- la statistica test (3.5) utilizzata nelle simulazioni è basata sui valori estremi e converge quasi certamente ad infinito (anche se i fattori di normalizzazione utilizzati garantiscono una convergenza alla distribuzione a valori estremi di Gumbel), mentre le due statistiche test alternative menzionate in precedenza, (3.6) e (3.7), hanno distribuzioni asintotiche basate su funzionali di ponti Browniani, pertanto dovrebbero presumibilmente avere un comportamento migliore anche se sono più difficili da applicare,
- le stime ai minimi quadrati usate per calcolare i residui campionari, su cui è basato il test, non rendono tanto quanto quelle di massima verosimiglianza; pertanto l'utilizzo del metodo della massima verosimiglianza per stimare i parametri del modello dovrebbe garantire un miglioramento delle prestazioni del test, però tali stime possono essere ottenute solo con approssimazioni tramite calcolo numerico a differenza di quelle CLS che hanno una forma chiusa (Franke *et al.*, 2012).

APPENDICE

Di seguito è riportato il codice del programma utilizzato sul software statistico R per simulare le serie storiche ed il test. Laddove devono essere inseriti i valori specifici della serie che si vuole simulare sono state lasciate indicazioni generiche contrassegnate da un asterisco (*):

```

inarch<-function(omega1,alfa1,omega2,alfa2,N,k){
  x<-NULL #serie storica
  lambda<-NULL
  media<-omega1/(1-alfa1)
  x[1]<-round(media,0)
  if(k<N){
    for(t in 2:k){
      lambda[t]<-omega1+alfa1*x[t-1]
      x[t]<-rpois(1,lambda[t])
    }
    for(t in (k+1):N){
      lambda[t]<-omega2+alfa2*x[t-1]
      x[t]<-rpois(1,lambda[t])
    }
  }
  if(k==N){
    for(t in 2:N){
      lambda[t]<-omega1+alfa1*x[t-1]
      x[t]<-rpois(1,lambda[t])
    }
  }
  return(ts(x))
}

s=numero di simulazioni* #simulazione
N=numerosità*
I=0
for(i in 1:s){
  xstima<-NULL
  lambdastima<-NULL
  res<-NULL
  x=inarch(omega1,alfa1,omega2,alfa2,N,k)*

  alfastima<-acf(x,plot=F)[[1]][2] #stime
  omegastima<-(1-alfastima)*mean(x)
}

```

```

xstima[1]<-omegastima/(1-alfastima) #residui
res[1]<-x[1]-xstima[1]
lambdastima[1] <- ome gastima+alfastima*xstima[1]
for(t in 2:N){
  lambdastima[t]<-omegastima+alfastima*x[t-1]
  res[t]<-x[t]-lambdastima[t]
}

tau2<-1/(N-2)*sum(res^2) #valore critico
n<-log(N)
a<-sqrt(2*log(n))
b<-2*log(n)+1/2*log(log(n))-1/2*log(pi)
C<-sqrt(tau2)*((log(2)-log(log(1/0.95))+b)/a)

cumsum_res<-cumsum(res) #test
A = NULL
for (k in 1:(N-1)){
  A[k] = sqrt(N/(k*(N-k)))
}
TEST<-max(A*(abs(cumsum_res[1:(N-1)])))

if(TEST>C){ #confronto
  I<-I+1
}
}
perc<-(I/s)*100
print(perc)

```

BIBLIOGRAFIA

Al-Osh, M.A., Alzaid, A.A. (1987) “First-order Integer-valued Autoregressive (INAR(1)) Process”, *Journal of Time Series Analysis*, 8 (3), 261-275.

Alzaid, A.A., Al-Osh, M.A. (1990) “An Integer-valued p th-order Autoregressive Structure (INAR(p)) Process”, *Journal of Applied Probability*, 27 (2), 314-324.

Bollerslev, T. (1986) “Generalized autoregressive conditional heteroskedasticity”, *Journal of Econometrics*, 31, 307-327.

Brockwell, P., Davis, R. (1991) *Time Series: Theory and Methods*, 2nd ed., Springer Verlag, New York.

Csörgö, M., Horváth, L. (1997) *Limit Theorems in Changepoint Analysis*, Wiley, New York.

Di Fonzo, T., Lisi, F. (2005) *Serie Storiche Economiche, Analisi Statistiche e Applicazioni*, Carocci Editore, Roma.

Efron, B., Tibsharani, R. (1993) *An Introduction to Bootstrap*, Chapman & Hall, New York.

Engle, R.F. (1982) “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”, *Econometrica*, 50 (4), 987-1007.

Ferland, R., Latour, A., Oraichi, D. (2006) “Integer-valued GARCH process”, *Journal of Time Series Analysis*, 27 (6), 923-942.

Franke, J., Kirch, C., Kamgaing, J.T. (2012) “Changepoints in times series of counts”, *Journal of Time Series Analysis*, 33 (5), 757-770.

Gouriéroux, C. (1997) *ARCH Models and Financial Applications*, Springer Series in Statistics, New York.

Heinen, A. (2003) *Modeling Time Series Count Data: The Autoregressive Conditional Poisson Model*, PhD thesis, San Diego, CA: Department of Economics, University of California.

Jin-Guan, D., Yuan, L. (1991) “The Integer-valued Autoregressive (INAR(p)) Model”, *Journal of Time Series Analysis*, 12 (2), 129-142.

Kedem, B., Fokianos, K. (2002) *Regression Models for Time Series Analysis*, Wiley, Hoboken.

Klimko, L.A., Nelson, P.I. (1978) “On Conditional Least Squares Estimation for Stochastic Processes”, *Annals of Statistics*, 6 (3), 629-642.

Sprott, D.A. (1983) “Estimating the Parameters of a Convolution by Maximum Likelihood”, *Journal of the American Statistical Association*, 78 (382), 457-460.

Steutel, F.W., Van Harn, K. (1979) “Discrete Analogues of Self-Decomposability and Stability”, *The Annals of Probability*, 7 (5), 893-899.

Van Harn, K. (1978) *Classifying Infinitely Divisible Distributions by Functional Equations*, Amsterdam: Mathematisch Centrum.

Weiss, A.A. (1984) “ARMA models with ARCH errors”, *Journal of Time Series Analysis*, 5 (2), 129-143.

Wei, C.H. (2010a) “INARCH(1) processes: higher-order moments and jumps”, *Statistics and Probability Letters*, 803, 1771-1780.

Weiß, C.H. (2010b) “The INARCH(1) model for overdispersed time series of counts”,
Communication in Statistics – Simulation and Computation, 39, 1269-1291.

White, H. (1982) “Maximum likelihood estimation of misspecified models”,
Econometrica, 50 (1), 1-25.