

Università degli studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Magistrale in  
Scienze Statistiche



TESI DI LAUREA

**STIMA DEL PARAMETRO A MEMORIA LUNGA VARIABILE  
NEL TEMPO CON UN MODELLO GAS**

Relatore Prof. Luisa Bisaglia

Dipartimento di Scienze Statistiche

Laureanda Valeria Sfragara

Matricola N 1090803

Anno Accademico 2016/2017



# Indice

<b>Introduzione</b>	<b>5</b>
<b>1 I processi a memoria lunga</b>	<b>7</b>
1.1 Introduzione . . . . .	7
1.2 I modelli ARFIMA . . . . .	10
<b>2 Modelli ARFIMA con parametro a memoria lunga variabile nel tempo</b>	<b>17</b>
2.1 Stima del parametro a memoria lunga TV con modelli di tipo TAR . . . . .	19
2.1.1 Modello SETAR . . . . .	20
2.1.2 Modello STAR . . . . .	23
2.2 Stima del parametro a memoria lunga TV con metodo <i>Wavelet</i>	27
<b>3 I modelli GAS</b>	<b>33</b>
3.1 Specificazione del modello e proprietà . . . . .	34
3.2 Riparametrizzazione . . . . .	36
3.3 Stima di massima verosimiglianza . . . . .	38

<b>4 TV-ARFIMA con <math>d</math> stimato con un modello GAS</b>	<b>41</b>
4.1 Specificazione del modello . . . . .	42
4.1.1 Stima dei parametri . . . . .	44
4.2 Simulazioni . . . . .	45
<b>Conclusioni</b>	<b>51</b>
<b>Bibliografia</b>	<b>53</b>

# Introduzione

Molte variabili economiche e finanziarie sono caratterizzate da memoria lunga e negli anni sono state proposte molte classi di modelli che descrivono la dinamica di fenomeni che presentano una struttura di dipendenza che persiste nel tempo. In particolare, uno dei processi a memoria lunga più studiato è l'ARFIMA( $p, d, q$ ) introdotto da Granger e Joyeux (1980) e Hosking (1981).

Il modello ARFIMA è costruito in maniera tale che il parametro di integrazione frazionaria sia costante nel tempo. Questa assunzione sembra essere restrittiva, poiché in letteratura sono stati evidenziati dei casi in cui sembra esservi presenza di *break* strutturali nel parametro di memoria lunga. Sono state, quindi, proposte delle estensioni dei modelli ARFIMA con parametro frazionario variabile nel tempo, tali specificazioni, tuttavia, non danno informazioni sull'evoluzione del parametro. Viene proposto da Dufrénot (2005a), Dufrénot (2005b) e Dufrénot (2006) che il parametro a memoria lunga variabile nel tempo si evolva secondo un modello SETAR; Boutahar (2008) e Boubaker e Sghaier (2014) affermano che il parametro si evolva secondo un modello STAR e stimano il parametro con il metodo della "regressione organizzata", il primo, e con un metodo basato sull'approccio *wavelet*, il

secondo.

Il contributo principale di questo lavoro di tesi è sempre quello di tenere conto di eventuali modifiche strutturali nella memoria dei dati, e per fare ciò viene proposto di stimare il parametro di memoria lunga variabile nel tempo con il modello GAS proposto da Creal (2013) e Harvey (2013).

Questo studio è quindi articolato come segue. Nel primo capitolo si introdurrà il concetto di processo a memoria lunga, nello specifico verranno affrontati i modelli ARFIMA (*AutoRegressive Fractional Integrated Moving Average*).

Successivamente, verranno presentati i modelli ARFIMA con parametro di memoria lunga variabile nel tempo e il tema verrà contestualizzato entro il panorama letterario presente sull'argomento; la formulazione sarà descritta in dettaglio nel capitolo 2.

Nel capitolo 3, viene presentata una classe di modelli per parametri variabili nel tempo che si basa sulla funzione di punteggio della densità del modello al tempo  $t$ ; in particolare si fa riferimento ai modelli GAS (*Generalized Autoregressive Score*).

Ultima, ma certamente non per importanza, viene proposta una stima alternativa per il parametro di memoria lunga variabile nel tempo di un ARFIMA, tramite un modello GAS, ed è stata anche condotta una simulazione per determinare la correttezza del modello proposto, effettuata nel capitolo 4.

# Capitolo 1

## I processi a memoria lunga

### 1.1 Introduzione

In questo capitolo viene posta l'attenzione su una particolare classe di processi stocastici lineari che si chiamano processi a memoria lunga o con dipendenza di lungo periodo. La dipendenza di lungo periodo può essere definita in molti modi, tuttavia, come sottolineato da Hall (1997), la motivazione originale per il concetto di memoria lunga è strettamente correlata alla stima della media di un processo stazionario. Se le autocovarianze di un processo stazionario sono assolutamente sommabili, allora la media del campione è consistente. In generale, si dice che questi processi abbiano una breve memoria. Al contrario, un processo ha una memoria lunga se le sue autocovarianze non sono assolutamente sommabili (Palma (2007)).

Le serie temporali generate da processi a memoria lunga presentano delle

caratteristiche ben precise. In primo luogo le osservazioni tendono ad assumere valori al di sopra o al di sotto della media per periodi di tempo relativamente lunghi: valori elevati tendono ad essere seguiti da valori elevati, mentre valori bassi sono tendenzialmente seguiti da valori bassi. Inoltre, nelle serie sono presenti cicli non periodici in successione casuale, se si osserva la serie in un periodo di tempo breve, può sembrare che la serie mostri un ciclo o un *trend* locale; anche se risulta evidente l'assenza di *trend* persistenti o di cicli se si guarda alla serie intera. Le serie, poi, appaiono complessivamente stazionarie sia in media che in varianza, ma la varianza della media campionaria converge a zero più lentamente di  $1/n$ . Asintoticamente si ha:

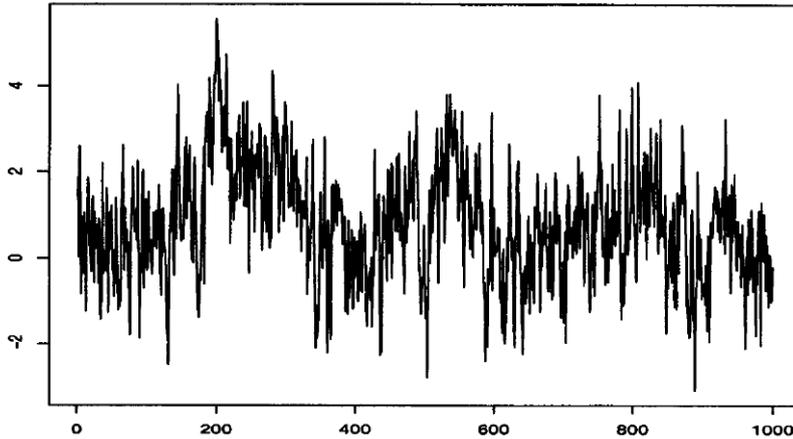
$$\text{Var}(\bar{y}_n) \sim C_{var} n^\alpha \quad \text{quando } n \rightarrow \infty,$$

dove  $\bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i$ ,  $\alpha \in (0, 1)$  e  $C_{var}$  è una costante positiva. Infine, è rilevante notare che la funzione di autocorrelazione (autocovarianza) converge molto lentamente a zero; asintoticamente si ha che:

$$\rho(k) \sim C_\rho k^\alpha \quad \text{quando } k \rightarrow \infty,$$

dove  $C_\rho$  è una costante non nulla e  $\alpha \in (0, 1)$ . Ciò implica che la dipendenza tra osservazioni successive decresce lentamente all'aumentare della distanza temporale tra le stesse, ovvero decresce con un andamento iperbolico e non esponenziale. La figura 1.1 mostra la tipica traiettoria di una serie storica a memoria lunga.

Se si assume che il processo osservato sia stazionario, è possibile definire un processo a memoria lunga o con dipendenza di lungo periodo o dipendente in senso forte, nel seguente modo:



**Figura 1.1:** Andamento di una serie a memoria lunga con  $N = 1000$  e  $H = 0.9$

**Definizione 1.1.** Il processo  $y = (y_t, t \in \mathbb{Z})$ , viene chiamato processo stazionario con memoria lunga o dipendenza di lungo periodo, se è stazionario e se esiste un numero reale  $H \in (1/2, 0)$ , e una funzione  $L_{\rho(k)}$  che varia lentamente all'infinito, tale che:

$$\rho(k) \sim L_{\rho(k)} k^{2H-2}, \quad (1.1)$$

quando  $k$  tende a  $\infty$ .

Si osserva che la funzione di autocorrelazione  $\rho(k)$  converge a zero così lentamente che  $\sum_{k=-\infty}^{\infty} \rho(k)$  diverge.

Bisogna notare che la definizione appena data di processo a memoria lunga è una definizione asintotica e il comportamento asintotico della funzione di autocorrelazione al tendere del ritardo all'infinito è fondamentale. Difatti l'equazione 1.1 determina solamente la convergenza a zero delle correlazioni e non dà informazioni sulla presenza di ritardi specifici per cui  $\rho(k)$  è particolarmente grande, ma soltanto che all'infinito le correlazioni sono ancora

diverse da zero. In teoria, quindi, un processo con dipendenza di lungo periodo potrebbe avere autocorrelazioni nulle fino a un ritardo  $k$  arbitrariamente grande, se, da quel ritardo, la funzione di autocorrelazione si comporta come nella definizione 1.1. Non ci si può limitare allo studio dei primi o di pochi coefficienti di autocorrelazione, ma bisogna cercare di stabilire in che modo la funzione di autocorrelazione converge a zero. Per poter fare inferenza bene, bisogna capire qual è l'effetto combinato di tutte le correlazioni. Dato che è rilevante capire come la funzione di autocorrelazione tende a zero, questo rende complicato definire se una serie è generata da un processo a memoria lunga oppure no (Beran (1994)).

Esistono diverse classi di modelli per cui sono soddisfatte le condizioni date dalla definizione precedente. In quest'elaborato verrà trattata solamente quella dei modelli ARFIMA.

## 1.2 I modelli ARFIMA

Una classe di modelli a memoria lunga sono i processi ARFIMA (*Auto-Regressive Fractionally Integrated Moving-Average*) introdotti da Granger e Joyeux (1980) e Hosking (1981). Questi sono un'estensione dei modelli ARMA (*AutoRegressive Moving-Average*) per descrivere la dinamica di fenomeni che presentano una struttura di dipendenza che persiste nel tempo.

Nella pratica si verifica spesso che molte serie siano non stazionarie, viene quindi proposta un'estensione dei processi ARMA( $p, q$ ) che studiano questo

tipo di serie. Vengono, perciò, considerati dei processi omogenei non stazionari di grado  $d$ , che possono essere resi stazionari attraverso opportune trasformazioni. Se si considera la  $d$ -esima differenza  $(1 - B)^d y_t$ , con  $d \in \mathbb{N}$ , si dirà che  $y_t$  è un processo integrato di ordine  $d$  e lo si denoterà come ARIMA( $p, d, q$ ),  $d$  viene chiamato parametro di integrazione o di differenziazione. L'equazione corrispondente è perciò:

$$\Phi(B)(1 - B)^d y_t = \Theta(B)\varepsilon_t, \quad (1.2)$$

dove  $\varepsilon_t$  è una successione di *white noise* con varianza finita ( $\varepsilon_t \sim w.n(0, \sigma^2)$ ),  $\Phi(B)$  e  $\Theta(B)$  sono due polinomi di grado  $p$  e  $q$  rispettivamente, tali che  $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  e  $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$  e  $B$  è tale che  $B^j y_t = y_{t-j}$ ,  $j = \dots, -1, 0, 1, \dots$ . Si noti che i processi ARMA( $p, q$ ) possono essere pensati come ARIMA( $p, 0, q$ ).

I modelli ARFIMA( $p, d, q$ ) generalizzano i modelli ARIMA( $p, d, q$ ) in quanto il parametro  $d$  può assumere un qualsiasi valore reale; è quindi possibile generalizzare l'equazione (1.2) in questo senso. Si può osservare che se  $d$  è intero, allora:

$$(1 - B)^d = \sum_{k=0}^d \binom{d}{k} (-1)^k B^k,$$

dove il coefficiente binomiale è pari a:

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$$

e  $\Gamma(\cdot)$  è la funzione gamma. Dato che la funzione gamma è definita per tutti i valori reali, la precedente definizione del coefficiente binomiale può essere

estesa a  $d \in \mathbb{R}$ , come:

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k = \sum_{k=0}^{\infty} \pi_k B^k,$$

dove

$$\pi_k = \frac{\Gamma(k - d)}{\Gamma(k + 1)\Gamma(-d)} = \prod_{0 < j \leq k} \frac{j - 1 - d}{j}, \quad j = 1, 2, \dots$$

A questo punto si può definire un processo ARFIMA( $p, d, q$ ) come segue:

**Definizione 1.2.**  $Y = (y_t, \quad t \in \mathbb{Z})$  è un processo ARFIMA( $p, d, q$ ) con  $d \in (-1/2, 1/2)$ , se è stazionario<sup>1</sup> e se soddisfa l'equazione alle differenze:

$$\Phi(B)\Delta^d y_t = \Theta(B)\varepsilon_t, \quad (1.3)$$

dove  $\varepsilon_t \sim w.n(0, \sigma^2)$  e  $\Phi(\cdot)$  e  $\Theta(\cdot)$  sono polinomi di grado  $p$  e  $q$  rispettivamente.

Se  $d \in (0, 1/2)$  il processo ARFIMA( $p, d, q$ ) è un processo stazionario con memoria lunga, se  $d = 0$  si ottiene nuovamente il modello ARMA( $p, q$ ) e se  $d \in (-1/2, 0)$  il processo è stazionario con memoria intermedia o dipendenza negativa o antipersistente. Quando  $d < 1/2$  si può dimostrare (Hosking (1981)) che il processo è stazionario. Quando  $d > 1/2$  si ottiene il processo ARFIMA( $p, d, q$ ) con  $d \in (-1/2, 1/2)$  dopo aver opportunamente differenziato la serie. Le figure 1.2 e 1.3 mostrano alcuni esempi di andamento di processi ARFIMA e le corrispondenti funzioni di autocorrelazione.

Se  $p = q = 0$ , il processo  $Y = (y_t, \quad t \in \mathbb{Z})$  è spesso chiamato rumore integrato frazionario (*Fractionally Integrated Noise*) e viene indicato come

<sup>1</sup>Il processo è stazionario se le radici di  $\Phi(B) = 0$  cadono fuori del circolo unitario.

ARFIMA( $\theta, d, \theta$ ). Questo processo è una soluzione stazionaria con media zero dell'equazione alle differenze:

$$\Delta^d y_t = \varepsilon_t \quad (1.4)$$

dove  $\varepsilon_t \sim w.n(0, \sigma^2)$  per ogni  $t \in \mathbb{Z}$ , e  $\Delta^d = (1 - B)^d$  è definito come in (1.3).

Per tale processo vale il seguente teorema:

**Teorema 1.2.1.** *Se  $d \in (-1/2, 1/2)$  allora esiste un'unica soluzione stazionaria  $y_t$  di (1.4) data da:*

$$y_t = \sum_{j=0}^d \psi_j \varepsilon_{t-j} = \Delta^{-d} \varepsilon_t$$

dove  $\psi_0 = 1$  e

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} = \prod_{k=1}^j \frac{k-1+d}{k}, \quad j = 1, 2, \dots$$

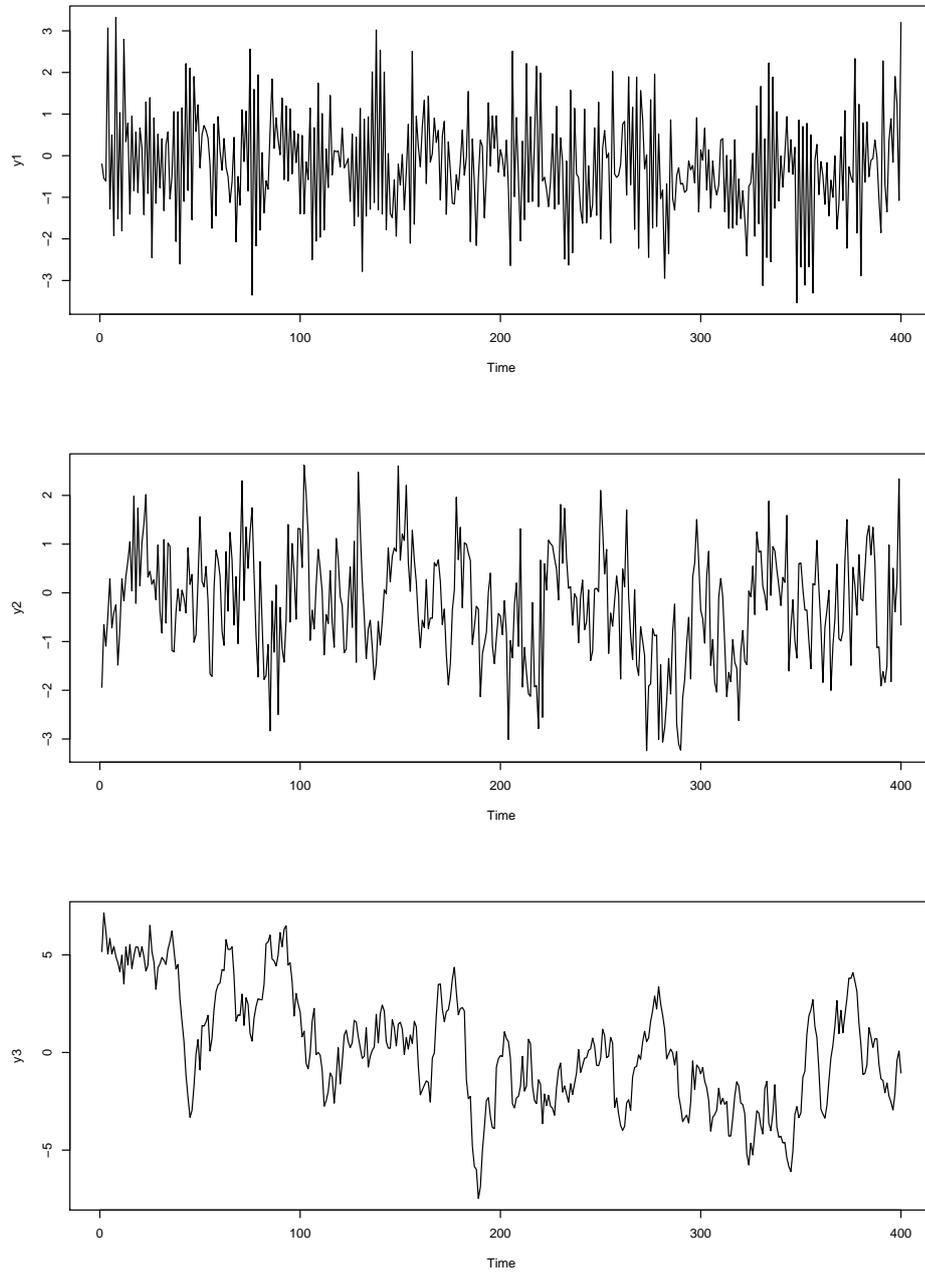
La varianza del processo è:

$$\gamma_0 = \sigma^2 \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}$$

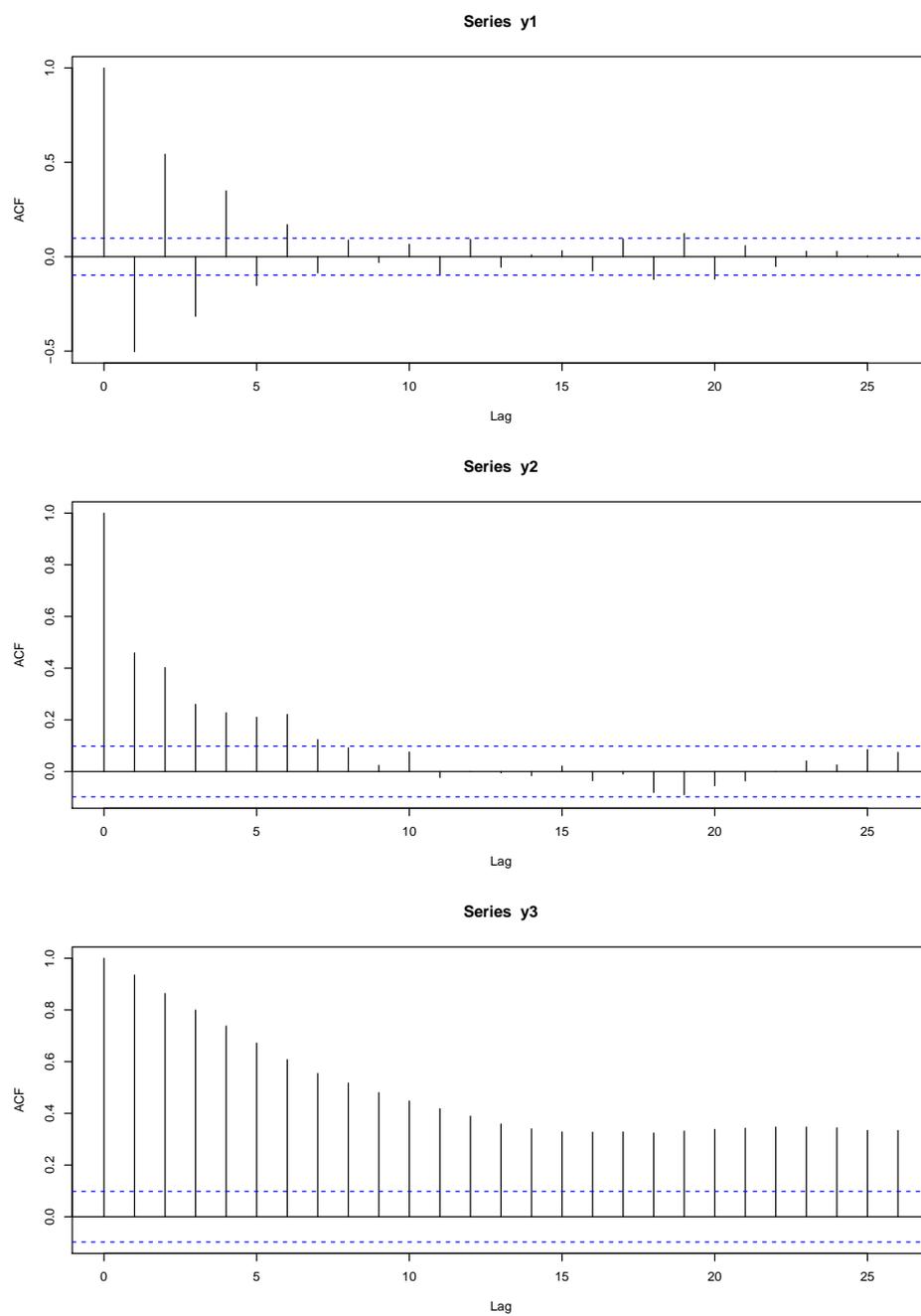
e la funzione di autocorrelazione è:

$$\rho_k = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} = \prod_{j=1}^k \frac{j-1+d}{j-d}, \quad j = 1, 2, \dots$$

Il parametro  $d$ , quindi, misura l'intensità della dipendenza. Per  $d \in (0, 1/2)$  il processo ARFIMA( $\theta, d, \theta$ ) è caratterizzato da una funzione di autocorrelazione che tende molto lentamente a zero, con  $d = 0$  si ottiene una successione di variabili casuali *i.i.d* e per  $d \in (-1/2, 0)$  il processo ha memoria intermedia o antipersistente.



**Figura 1.2:** Serie simulata di un processo ARFIMA(1, 0.3, 0) con parametro AR  $-0.7$ , un processo ARFIMA(0, 0.3, 0) e un processo ARFIMA(1, 0.3, 0) con parametro AR 0.7



**Figura 1.3:** Funzione di autocorrelazione di un processo ARFIMA(1, 0.3, 0) con parametro AR  $-0.7$ , un processo ARFIMA(0, 0.3, 0) e un processo ARFIMA(1, 0.3, 0) con parametro AR  $0.7$



## Capitolo 2

# Modelli ARFIMA con parametro a memoria lunga variabile nel tempo

I modelli a memoria lunga hanno ricevuto notevole attenzione da ricercatori provenienti da diverse discipline, tra cui l'idrologia, la finanza e altri campi. I libri di Beran (1994) e Palma (2007) forniscono degli aggiornamenti sugli sviluppi in questo argomento. Come già affermato nel paragrafo 1.1, il processo a memoria lunga più studiato è l'ARFIMA( $p, d, q$ ) definito dall'equazione 1.2.

Studi precedenti suppongono che il parametro di integrazione frazionaria,  $d$ , sia costante nel tempo, il che implica che la struttura di dipendenza a lungo raggio del fenomeno sottostante persista nel tempo. Questa ipotesi sembra essere restrittiva a causa della presenza di *break* strutturali nel parametro di integrazione frazionaria (Granger e Hyung (2004), Beltratti e Morana (2006) e Baillie e Morana (2009)). Nello stesso contesto alcuni autori

hanno supposto che il parametro di integrazione frazionaria è variabile nel tempo,  $d_t$ , e che la persistenza agli *shock* varia nel tempo (Jensen (1999a), Jensen (1999b), Whitcher e Jensen (2000), Beran (2009) e Roueff e Von Sachs (2011)). Tuttavia, questi autori non specificano il processo di evoluzione di  $d_t$ .

Se si suppone che  $d$  varia nel tempo, vale a dire  $d_t$ , si ottiene il cosiddetto modello ARFIMA( $p, d, q$ ) variabile nel tempo indicato come TV-ARFIMA( $p, d, q$ ). Questo modello a memoria lunga fa parte dei processi localmente stazionari (Whitcher e Jensen (2000)) ed è dato da:

$$\Phi(B)(1 - B)^{d_t}y_t = \Theta(B)\varepsilon_t. \quad (2.1)$$

dove  $\varepsilon_t$  è una successione di *white noise* con varianza finita  $\varepsilon_t \sim w.n(0, \sigma_\varepsilon^2)$ ,  $\Phi(B)$  e  $\Theta(B)$  sono due polinomi definiti come nella formula 1.2 e  $d_t < 0.5$  è il parametro di integrazione frazionaria che varia nel tempo.

Le definizioni di modello a memoria lunga variabile nel tempo che vengono comunemente usate sono:

**Definizione 2.1.**  $\{y_t\}$  è un processo a memoria lunga localmente stazionario se esiste la funzione di densità spettrale variabile nel tempo,  $sdf(t, \lambda)$ , tale che:

$$sdf(t, \lambda) \sim \lambda^{-2d_t} \quad \text{con } \lambda \rightarrow 0^+.$$

Quindi, se  $d_t > 0$ ,  $sdf(t, \lambda)$  è liscia per le frequenze vicine allo zero, ma non è limitata quando  $\lambda = 0$ . Se  $d_t < 0$ , allora  $sdf(t, \lambda) = 0$  e  $y_t$  è una serie localmente stazionaria anti-persistente. Perciò,  $y_t$  sarà più liscio nei periodi di tempo in cui  $d_t > 0$ , e avrà grandi fluttuazioni quando  $d_t < 0$ .

**Definizione 2.2.**  $\{y_t\}$  è un processo a memoria lunga localmente stazionario se esiste la funzione di autocovarianza locale,  $cov_y(t, g - h)$ , che può essere semplificata come:

$$cov_y(t, g - h) \sim |g - h|^{2d_t - 1} \quad \text{con } |g - h| \rightarrow \infty.$$

Il lento decrescimento iperbolico di  $cov_y(t, g - h)$  è la caratteristica più peculiare dei processi a memoria lunga.

Anche se questo modello indica che il parametro di integrazione frazionaria varia in funzione del tempo, non fornisce informazioni sull'evoluzione di  $d_t$ . In questo capitolo vengono proposti alcuni modelli e metodi di stima per il parametro di memoria lunga che varia nel tempo presenti nella letteratura.

## 2.1 Stima del parametro a memoria lunga TV con modelli di tipo TAR

Nell'economia e nella finanza, la modellazione di regimi multipli sta diventando sempre più importante al fine di tenere conto di fenomeni caratterizzati, ad esempio, da periodi di recessione o espansione o periodi di alta o bassa volatilità. Di conseguenza, nella letteratura sono stati proposti diversi modelli, tra i quali modelli a volatilità stocastica, modelli GARCH non lineari, modelli autoregressivi a soglia (TAR), modelli *Markov switching*, modelli GARCH multifattori, ecc. In questo paragrafo ci si focalizzerà sui modelli TAR che rappresentano una delle classi più importanti dei modelli *regime-switching*.

### 2.1.1 Modello SETAR

I modelli TAR (*Threshold Autoregressive*) sono stati inizialmente introdotti da Tong e Lim (1980) e discussi in Tong (1990). Il concetto su cui si basa questa classe di modelli nonlineari è che l'andamento di un processo possa essere rappresentato da un insieme finito di autoregressioni lineari. Il modello AR che genera il valore della serie in ogni periodo è determinato dalla relazione che intercorre tra una variabile di soglia e i valori della soglia stessa. Quando questa variabile di soglia è la variabile dipendente ritardata  $d$  periodi, come si considera negli articoli di Dufrénot (2005a), Dufrénot (2005b) e Dufrénot (2006), il modello viene chiamato *Self-Exciting Threshold Autoregressive* (SETAR). Inoltre, ogni regime del modello TAR è modellato da un processo autoregressivo e, quindi, a memoria corta. Ma vale la pena notare che i dati economici o finanziari possono presentare a volte un comportamento a lunga memoria in uno o in entrambi i regimi e potrebbe essere interessante considerare dei modelli che presentino un comportamento con lunga memoria almeno in uno dei due regimi. Di conseguenza, in questo paragrafo viene descritto un modello SETAR a due regimi in cui entrambi i regimi sono caratterizzati da un *white noise* frazionario; si considera quindi un processo  $y_t$  che soddisfa il seguente schema:

$$\begin{cases} (1 - B)^{d_1} y_t = \varepsilon_t^{(1)} & \text{se } y_{t-1} \leq c : \text{ regime 1} \\ (1 - B)^{d_2} y_t = \varepsilon_t^{(2)} & \text{se } y_{t-1} > c : \text{ regime 2} \end{cases} \quad (2.2)$$

dove  $d_i \in (0, 1/2)$ ,  $i = 1, 2$  sono i parametri di differenza frazionaria,  $c$  è un parametro di soglia,  $\varepsilon_t^{(i)}$ , sono *white noise* con varianze finite e  $B$  è l'operatore

ritardo. Questo modello include come un caso particolare il seguente sotto modello:

$$\begin{cases} y_t = \varepsilon_t^{(1)} & \text{se } y_{t-1} \leq c : \text{regime 1} \\ (1 - B)^{d_2} y_t = \varepsilon_t^{(2)} & \text{se } y_{t-1} > c : \text{regime 2} \end{cases} \quad (2.3)$$

dove è presente una dinamica persistente in solo uno dei due regimi.

La funzione di autocorrelazione del modello 2.2 è:

$$\gamma_y(\tau) = C(\tau, d_1)N_1(c) + C(\tau, d_2)N_2(c)$$

dove, per  $i = 1, 2$ ,

$$C(\tau, d_i) = \frac{\Gamma(1 - 2d_i)\Gamma(\tau + d_i)}{\Gamma(\tau - d_i + 1)\Gamma(d_i)\Gamma(1 - d_i)}$$

e  $N_1(c)$  e  $N_2(c)$  rappresentano rispettivamente le percentuali di osservazioni nel regime 1 e nel regime 2 e non sono predefinite, ma dipendono dalla soglia  $c$ . La funzione di autocorrelazione teorica è quindi una mistura delle autocorrelazioni dei modelli a memoria lunga in entrambi i regimi e questa presenta una varietà di tassi di decrescimento (da veloce a molto lento) in base al valore di  $c$ .

La stima dei parametri del modello  $(d, c)$  è complicata e l'approccio che viene adottato è quello di: (a) determinare un valore plausibile per la soglia, quindi (b) stimare gli altri parametri del modello condizionatamente a questo valore di soglia. Le diverse fasi utilizzate per stimare il modello sono quindi le seguenti:

- Stimare il valore del parametro di soglia  $c$ .

- Separare le osservazioni in due sottogruppi in base al valore stimato di  $c$  e dedurre  $N_1(c)$  e  $N_2(c)$ .
- Per ogni sottogruppo, stimare il parametro frazionario  $d$ . La stima si basa sull'approccio del log-periodogramma GPH, inizialmente proposto da Geweke e Porter-Hudak (1983) e raffinato da Robinson (1995). Nello specifico, il comportamento della funzione di densità spettrale di un modello ARFIMA quando  $\lambda$  tende a 0 è del tipo:

$$f(\lambda) \sim C|\lambda|^{-2d}, \quad (2.4)$$

dove  $0 < C < \infty$  e  $d \in (-1/2, 1/2)$ . Inoltre, il periodogramma,  $I(\lambda)$  è una stima asintoticamente non distorta della funzione di densità spettrale, ovvero:

$$\lim_{\lambda \rightarrow 0} \mathbb{E}[I(\lambda)] = f(\lambda).$$

Quindi, valutando l'equazione 2.4 alle frequenze di Fourier  $\lambda_{j,T} = (2\pi j/T)$ ,  $j = 1, 2, \dots, T^*$ , dove  $T^*$  è la parte intera di  $(T - 1)/2$ , è possibile considerare l'equazione di regressione:

$$\log(I(\lambda_{j,T})) = \alpha + \beta \log(|\lambda_{j,T}|) + \varepsilon_j, \quad j = 1, \dots, m$$

dove  $\beta = -2d$ , per ottenere uno stimatore consistente e asintoticamente normale di  $d$ , definito come  $\hat{d} = -1/2\hat{b}$ . La stima di  $d$  dipende dal numero di frequenze utilizzate e in questo studio viene posto  $m = T^{0.8}$ , inoltre vengono usate tutte le prime frequenze di Fourier.

### 2.1.2 Modello STAR

In Boutahar (2008) viene esteso il processo 2.2 considerando una funzione regolare per parametro di memoria lunga  $d_t$ . La motivazione principale per l'utilizzo di tale modello è che, come mostrano Diebold e Inoue (2001) e Granger e Hyung (2004), un cambiamento del parametro di memoria lunga e un cambiamento strutturale può essere facilmente confuso finché si verifica un piccolo cambiamento di regime. Si ha quindi bisogno di un modello che consente di distinguere tra le due proprietà in uno stesso contesto.

Viene quindi considerato il processo TV-ARFIMA 2.1 e si assume che  $d_t$  si evolva come un processo stocastico. Si assume perciò che  $d_t$  sia un processo STAR

$$d_t = d_{(1)}(1 - F_t(\gamma, c)) + d_{(2)}F_t(\gamma, c), \quad (2.5)$$

dove  $F_t(\gamma, c)$  è una funzione di transizione liscia, come per esempio una funzione logistica data da

$$F_t(\gamma, c) = [1 + \exp(-\gamma(y_{t-1} - c))]^{-1}. \quad (2.6)$$

I parametri  $d_{(1)}$  e  $d_{(2)}$  sono i valori del parametro di memoria lunga nei due regimi estremi, cioè quando  $F_t \rightarrow 0$  e  $F_t \rightarrow 1$ . Il parametro  $\gamma$  indica quanto è rapida la transizione tra i due regimi estremi e può essere positivo o negativo a seconda che la curva logistica aumenti o meno. Il parametro  $c$  è il parametro di posizione.

Se  $\gamma > 0$ , i regimi estremi sono ottenuti come segue:

- se  $y_{t-1} \rightarrow -\infty$ , si ha che  $F_t \rightarrow 0$  e quindi  $\Phi(B)(1 - B)^{d_{(1)}}y_t = \Theta(B)\varepsilon_t$ ,

- se  $y_{t-1} \rightarrow +\infty$ , si ha che  $F_t \rightarrow 1$  e quindi  $\Phi(B)(1-B)^{d(2)}y_t = \Theta(B)\varepsilon_t$ ,

Si suppone inoltre che  $\Phi(B) = \Theta(B) = 1$ , cioè si sta considerando il più semplice ARFIMA( $\theta, d_t, \theta$ ), ovvero un I( $d$ ) con un parametro frazionario variabile nel tempo. Sia  $(\Omega, F, P)$  lo spazio di probabilità e si supponga che  $y_t$  e  $d_t$  siano definiti su  $\Omega$  e dati rispettivamente da 1.2, 2.5 - 2.6. Il filtro frazionario stocastico  $(1-B)^{d(t)}$  può essere definito come segue:

$$\forall \omega \in \Omega, (1-B)^{d(t,\omega)} = \sum_{j=0}^{\infty} \frac{\Gamma(j-d(t,\omega))}{\Gamma(j+1)\Gamma(-d(t,\omega))} B^j,$$

dove  $d(t,\omega) := d_{(t)}(\omega)$ . Perciò il processo  $y_t$  può essere scritto come

$$y_t = \sum_{j=0}^{\infty} b_j(y_{t-1})\varepsilon_{t-j},$$

dove

$$b_j(y) = \frac{\Gamma(j+D(y))}{\Gamma(j+1)\Gamma(D(y))}$$

e

$$D(y) = d_{(1)} + (d_{(2)} - d_{(1)}) [1 + \exp(-\gamma(y-c))]^{-1}.$$

Per spiegare come stimare i parametri a memoria lunga  $d_{(1)}$  e  $d_{(2)}$  e il parametro di transizione  $\gamma$  del modello 2.1, 2.5 - 2.6, si assume per semplicità che il parametro di posizione  $c = 0$  e che  $\Phi(B) = \Theta(B) = 1$ . Si considera quindi il modello dato da:

$$\begin{cases} (1-B)^{d_t}y_t = \varepsilon_t \\ d_t = d_{(1)} + (d_{(2)} - d_{(1)}) [1 + \exp(-\gamma(y_{t-1}))]^{-1} \end{cases} \quad (2.7)$$

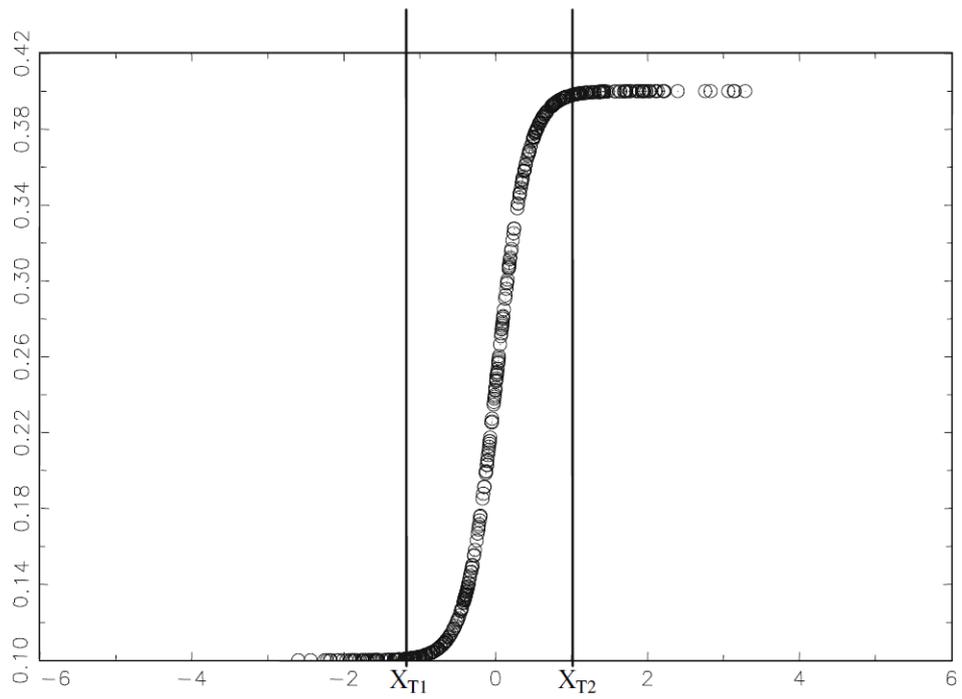
per  $t = 1, \dots, T$

Il metodo di stima proposto utilizza un approccio della letteratura applicato sui modelli non lineari, ovvero quello delle "regressioni organizzate". Una regressione organizzata è una regressione con le variabili endogene ed esogene, vale a dire  $Y$  e  $X$ , riordinate in base ai valori di una variabile particolare, per esempio  $Z$ . Più in particolare, quando la variabile  $Z$  è una variabile di soglia, le osservazioni di  $Y$  e  $X$  sono disposte secondo l'ordine di grandezza ascendente o discendente delle osservazioni di  $Z$ . Le regressioni organizzate in base a variabili ordinate riconoscono in maniera più precisa le soglie o i cambiamenti di regime in una serie. In questo contesto, tale metodologia viene applicata per trovare i valori di  $d_{(1)}$  e  $d_{(2)}$  poiché le variazioni di  $d_t$  tra i due valori estremi sono soggetti a variazioni secondo  $y_{t-1}$ . La funzione logistica viene utilizzata per indicare che le modifiche nel parametro di memoria lunga si verificano in modo continuo. Viene considerato sia l'ordine ascendente che decrescente di  $y_{t-1}$  per scopi di robustezza.

La figura 2.1 mostra il grafico delle osservazioni organizzate di  $d_t$  sulla base dei valori crescenti di  $y_{t-1}$ . La forma della curva è ovviamente una forma logistica, seguendo il modello 2.7. È caratterizzato da un regime più basso con  $d_t \approx d_{(1)}$ , un regime superiore con  $d_t \approx d_{(2)}$  e un regime intermedio con un continuum di punti tra i due regimi estremi. I *break* temporali  $T_1$  e  $T_2$  sono approssimativamente gli istanti temporali in cui cambiano i tre regimi. Il metodo di stima consiste nel trovare prima questi due istanti temporali, da cui dedurre la stima di  $d_{(1)}$  e  $d_{(2)}$ , e quindi la stima di  $\gamma$ .

Il metodo di stima può essere riassunto come segue:

- Per individuare i *break* temporali  $T_1$  e  $T_2$ , si considera una curva co-



**Figura 2.1:** Serie di  $d_t$  ordinati per valori crescenti della variabile di transizione

me quella rappresentata in figura 2.1 per determinarne una prima approssimazione, chiamata  $T_1^A$  e  $T_2^A$ .

- Viene poi considerata la stessa curva, ma in ordine inverso, per determinare una seconda approssimazione, chiamata  $T_1^B$  e  $T_2^B$ .
- Si determina il valore finale di  $T_1$  e  $T_2$  facendo una media delle approssimazioni precedenti. Questa viene calcolata perchè uno di questi primi due passaggi, preso da solo, può produrre dei valori dei *break* che sono al di sopra o al di sotto dei valori corretti di  $T_1$  e  $T_2$ ; usando la procedura doppia e calcolando le medie si risolve questo problema.
- Si è quindi in grado di stimare  $d_{(1)}$  e  $d_{(2)}$  e  $\gamma$ .

## 2.2 Stima del parametro a memoria lunga TV con metodo *Wavelet*

In questo paragrafo viene descritto il modello a memoria lunga in cui il parametro frazionario varia nel tempo proposto da Boubaker e Sghaier (2014). In particolare, seguendo il lavoro di Boutahar (2008) (proposto nel paragrafo 2.1.2), si suppone che  $d_t$  varia secondo due regimi in cui la transizione da un regime all'altro è liscia. La caratteristica interessante del modello è che permette sia la presenza di dipendenza a lungo raggio della serie che l'asimmetria nei gradi di persistenza.

Si sta quindi ipotizzando che  $d_t$  evolva secondo il modello STAR avanzato da Teräsvirta (1994) e Teräsvirta (1996), come:

$$\begin{cases} \Phi(B)(1 - B)^{d_t} y_t = \Theta(B)\varepsilon_t, \\ d_t = d_{(1)} [1 - F(s_t; \gamma, c)] + d_{(2)} F(s_t; \gamma, c). \end{cases} \quad (2.8)$$

Dove  $d_{(1)}$  e  $d_{(2)}$  sono i valori del parametro di integrazione frazionaria nel primo e nel secondo regime rispettivamente;  $F(s_t; \gamma, c)$  è la funzione di transizione che è continua tra 0 e 1 con  $s_t$  che indica la variabile di transizione, che può essere una delle variabili endogene ritardate  $s_t = d_{t-i}, \forall t > i; t = 1, \dots, T$  o una variabile esogenea. Il parametro  $\gamma$  misura la velocità della transizione tra i due regimi estremi. Il parametro  $c$  rappresenta la soglia per la variabile di transizione  $s_t$  che definisce i regimi sottostanti: se  $s_t < c$  (o  $s_t > c$ ) significa che il regime sottostante è il primo (o il secondo). In letteratura si trovano due tipi di funzione di transizione:

- la funzione logistica, per il modello logistico STAR (LSTAR):

$$F(s_t; \gamma, c) = (1 + \exp(-\gamma(s_t - c)))^{-1},$$

- la funzione esponenziale, per il modello esponenziale STAR (ESTAR):

$$F(s_t; \gamma, c) = 1 - \exp(-\gamma(s_t - c)^2).$$

Si è detto nel paragrafo precedente che, per stimare il parametro  $d_t$ , Bouthahar (2008) ha considerato la "regressione organizzata", che consiste nell'organizzare le variabili endogene ed esogene secondo i valori crescenti delle

osservazioni di un'altra variabile. Invece, Boubaker e Sghaier (2014) propone un metodo alternativo eseguito in due passi. Al primo passo, ci si concentra sulla stima del parametro di integrazione frazionaria variabile nel tempo. Il vantaggio di stimare prima  $d_t$  è che consente di implementare test per verificare la presenza di nonlinearità e per determinare la funzione di transizione e la variabile di transizione appropriata. Nel secondo passo, si stima il modello STAR su  $d_t$ .

Per stimare  $d_t$  si utilizza un metodo basato sull'approccio *wavelet*. In particolare la tecnica MODWT (*Maximum Overlap Discrete Wavelet Transform*) sviluppata da Percival e Walden (2000) che ci fornisce un'approssimazione della relazione log-lineare tra la varianza variabile nel tempo dei coefficienti MODWT e il parametro a memoria lunga variabile nel tempo  $d_t$ . Viene poi applicato lo stimatore istantaneo ai minimi quadrati (ILSE) per ottenere delle stime locali di  $d_t$ .

### **Stima istantanea ai minimi quadrati**

L'idea di base per stimare il parametro  $d$  tramite una trasformazione *wavelet* della serie storica è la varianza *wavelet*. L'analisi della varianza *wavelet* consiste nel suddividere la varianza di una serie in parti che sono associate a differenti scale temporali. Questo è utile per spiegare quali scale danno un contributo importante alla variabilità complessiva di una serie.

In particolare, si consideri la serie temporale di interesse  $y_1, y_2, \dots, y_T$ , che si assume essere una realizzazione di un processo stazionario con varianza  $\sigma_y^2$ . Se i coefficienti di scala per un generico livello  $j$  sono associati a medie di

lunghezza  $2^j$ , allora i coefficienti di *wavelet* di livello  $j$  sono associati con dei cambiamenti alla scala  $\tau_j = 2^{j-1}\Delta_t$ , dove  $\Delta_t$  è l'intervallo campionario di  $y_t$ . Così, la varianza wavelet  $V_y^2(\tau_j)$  per la scala  $(\tau_j \equiv 2^{j-1})^2$  è definita come:

$$\sigma_y^2 \equiv \sum_{j=1}^J V_y^2(\tau_j).$$

Per la stima con un approccio wavelet del parametro di integrazione frazionaria  $d$  sono stati indicati molti metodi. Possono essere riassunti in tre approcci:

- Il primo approccio è un'approssimazione basata sulle *wavelet* dello stimatore di massima verosimiglianza (MLE) di  $d$  sotto l'assunzione di Gaussianità multivariata (McCoy e Walden (1996), Jensen (1999a) e Whitcher e Jensen (2000)).
- Il secondo fa uso del fatto che la relazione tra la varianza dei coefficienti *wavelet* sulle scale è dettata da  $d$ . In questo quadro, viene costruito uno stimatore ai minimi quadrati (LSE) di  $d$  (Abry e Veitch (1998) e Jensen (1999b)).
- Il terzo utilizza solo determinati coefficienti che sono collocati nel tempo, e vengono chiamati stimatori istantanei ai minimi quadrati (ILSE).

L'idea principale dello stimatore istantaneo ai minimi quadrati è quella di utilizzare un singolo coefficiente di *wavelet* da ogni scala, cioè si userà solo  $\tilde{d}_{j,t_j}$  per stimare  $V_y^2(\tau_j)$ , dove  $t_j$  è l'indice di tempo del  $j$ -esimo livello del coefficiente MODWT associato al tempo  $t$  in  $y_t$ ,  $t = 1, \dots, T$ . Lo stimatore

istantaneo ai minimi quadrati è dato da:

$$\hat{d}_{ILSE,t} = \frac{\Delta_J \sum \ln(\tau_j) X_t(\tau_j) - \sum \ln(\tau_j) \sum X_t(\tau_j)}{2(\Delta_J \sum \ln^2(\tau_j) - (\sum \ln(\tau_j))^2)} + \frac{1}{2} \quad (2.9)$$

Dove  $\Delta_J = J - J_0 + 1$  e tutte le somme sono in  $j = J_0, \dots, J$  e  $X_t(\tau_j) \equiv \ln(d_{j,t}^2) - \Psi(1/2) - \ln(2)$  dove  $\Psi$  è la funzione digamma.

Dopo aver stimato  $d_t$ , viene ufatto un test di ipotesi per scegliere la funzione di transizione appropriata, ovvero per selezionare uno tra i modelli LSTAR o ESTAR. Successivamente si può stimare il modello STAR specificato.



# Capitolo 3

## I modelli GAS

I modelli per serie storiche con parametri variabili nel tempo sono stati classificati da Cox (1981) in due classi: i modelli *observation-driven* e i modelli *parameter-driven*. Nell'approccio *observation-driven*, la variazione temporale dei parametri viene introdotta lasciando i parametri come funzioni di variabili dipendenti ritardate. Anche se i parametri sono stocastici, sono perfettamente prevedibili date le informazioni passate. Alcuni esempi di questi modelli sono i modelli GARCH (*Generalized AutoRegressive Conditional Heteroskedasticity*) di Engle (1982), Bollerslev (1986) e Engle e Russell (1998), i modelli ACD (*AutoRegressive Conditional Duration*) e ACI (*AutoRegressive Conditional Intensity*) di Engle e Russell (1998) e Russell (2001), il modello DCC (*Dynamic Conditional Correlation*) di Engle (2002). Il modello presentato in questo capitolo comprende molti dei modelli *observation-driven* esistenti. Inoltre, consente la formulazione di una vasta gamma di nuovi modelli. Nei modelli *parameter-driven*, i parametri sono processi stocastici con il loro

errore. Date le osservazioni passate e attuali, i parametri non sono perfettamente prevedibili. Esempi tipici di modelli *parameter-driven* sono i modelli SV (*Stochastic Volatility*) di Shephard (2005) e i modelli *Stochastic Intensity* di Bauwens e Hautsch (2006) e Koopman (2008b).

In questo capitolo viene presentata una struttura per parametri variabili nel tempo che si basa sulla funzione di punteggio della densità del modello al tempo  $t$ . Utilizzando determinate funzioni di *score*, è possibile ritornare a dei modelli *observation-driven* standard, come i modelli GARCH, ACD e ACI. L'applicazione di questa struttura ad altri modelli non lineari, non gaussiani, multivariati, porterà alla formulazione di nuovi modelli *observation-driven*.

Si fa riferimento a questo modello basato sulla funzione di punteggio come il modello *Generalized Autoregressive Score* (GAS). Il modello GAS proposto da Koopman (2008a), Creal (2013) e Harvey (2013) ha i vantaggi di altri modelli *observation-driven*. Poiché questo si basa sulla funzione *score*, sfrutta completamente la struttura della densità, piuttosto che soltanto la media e momenti più alti. Tale aspetto è ciò che differenzia il modello GAS dagli altri modelli *observation-driven* nella letteratura, come i modelli generalizzati autoregressivi a media mobile di Shephard (1995) e Benjamin (2003) e i modelli di errore moltiplicativo vettoriale di Cipollini (2006).

### 3.1 Specificazione del modello e proprietà

Si definisce il vettore  $y_t$  che rappresenta la variabile dipendente d'interesse,  $f_t$  è il vettore di parametri che varia nel tempo,  $x_t$  è un vettore di variabili

esogene tutte al tempo  $t$ , e  $\theta$  è un vettore di parametri statici. Si definisce  $Y^t = \{y_1, \dots, y_t\}$ ,  $F^t = \{f_0, f_1, \dots, f_t\}$ , e  $X^t = \{x_1, \dots, x_t\}$ . Le informazioni disponibili al momento  $t$  sono costituite da  $\{f_t, \mathcal{F}_t\}$ , dove

$$\mathcal{F}_t = \{Y^{t-1}, F^{t-1}, X^t\}.$$

Si assume che  $y_t$  sia generato dalla densità

$$y_t \sim p(y_t | f_t, \mathcal{F}_t, \theta). \quad (3.1)$$

Inoltre, si assume che il meccanismo di aggiornamento del parametro variabile nel tempo  $f_t$  sia dato dall'equazione di aggiornamento autoregressivo

$$f_{t+1} = \omega + \sum_{i=1}^p A_i s_{t-i+1} + \sum_{j=1}^q B_j f_{t-j+1} \quad (3.2)$$

dove  $\omega$  è un vettore di costanti, le matrici dei coefficienti  $A_i$  e  $B_j$  hanno dimensioni appropriate con  $i = 1, \dots, p$  e  $j = 1, \dots, q$ , mentre  $s_t$  è una funzione dei dati passati,  $s_t = s_t(y_t, f_t, F_t; \theta)$ . I coefficienti non noti nella formula 3.2 sono funzioni di  $\theta$ ; cioè,  $\omega = \omega(\theta)$ ,  $A_i = A_i(\theta)$  e  $B_j = B_j(\theta)$  per  $i = 1, \dots, p$  e  $j = 1, \dots, q$ .

La caratteristica principale dei modelli GAS è che l'evoluzione del vettore di parametri variabili nel tempo  $f_t$  è guidata dallo *score* della distribuzione nella formula 3.1, insieme a una componente autoregressiva. Quando si realizza un'osservazione  $y_t$ , si aggiorna il parametro variabile nel tempo  $f_t$  al periodo successivo  $t + 1$  usando la formula 3.2 con

$$s_t = S_t \nabla_t, \quad \nabla_t = \frac{\partial \ln p(y_t | f_t, \mathcal{F}_t, \theta)}{\partial f_t}, \quad S_t = S(t, f_t, \mathcal{F}_t, \theta) \quad (3.3)$$

dove  $S(\cdot)$  è una matrice scalare nota al tempo  $t$ . Le equazioni 3.1 - 3.3 definiscono il modello *Generalized Autoregressive Score* di ordine  $p$  e  $q$ , abbreviato come  $\text{GAS}(p, q)$ .

In base alla scelta della matrice di scala  $S_t$ , il modello GAS consente un'ulteriore flessibilità nel modo in cui lo *score* viene utilizzato per l'aggiornamento di  $f_t$ . È importante notare che ogni scelta diversa per la matrice di scala  $S_t$  è un modello GAS differente. Le proprietà statistiche ed empiriche di ciascuno di questi modelli possono essere diverse. In Creal (2013) viene suggerito di fissare la matrice di scala  $S_t$  a una potenza  $\gamma > 0$  dell'inversa della matrice d'informazione di  $f_t$  per spiegare la varianza di  $\nabla_t$ . Più precisamente:

$$S_t \equiv \mathcal{I}_t^{-\gamma}$$

con

$$\mathcal{I}_t = \mathbb{E}_{t-1} [\nabla_t \nabla_t^T],$$

dove il valore atteso è preso rispetto alla distribuzione condizionata  $p(y_t | f_t, \mathcal{F}_t, \theta)$ .

Il parametro aggiuntivo  $\gamma$  viene fissato dal ricercatore e di solito assume un valore tra  $\{0, \frac{1}{2}, 1\}$ . Quando  $\gamma = 0$ ,  $S_t = \mathbf{I}$  e non c'è la matrice di scala.

## 3.2 Riparametrizzazione

Nella 3.2 il vettore di parametri  $f_t$  ha una specificazione lineare e quindi non è limitato. Nella pratica, lo spazio parametrico di  $f_t$  è spesso limitato ( $\mathcal{F}_t \subset \mathbb{R}^J$ ); ad esempio, quando si modella il parametro di scala di una distribuzione  $t$ -Student, bisogna assicurarsi della sua positività. Anche se questo

problema può essere risolto imponendo dei vincoli su  $\theta$  (come avviene nel modello GARCH), la soluzione standard per una struttura di tipo GAS è quella di utilizzare una funzione di collegamento (possibilmente non lineare)  $\Lambda(\cdot)$  che associa  $\tilde{f}_t \in \mathbb{R}^J$  a  $f_t$  e dove  $\tilde{f}_t \in \mathbb{R}^J$  ha la specificazione dinamica lineare della 3.2. In particolare, sia  $\Lambda : \mathbb{R}^J \rightarrow \mathcal{F}_t$  una funzione di associazione differenziabile due volte in modo tale che  $\Lambda(\tilde{f}_t) = f_t$ . L'equazione di aggiornamento per  $f_t$  è quindi data da:

$$\begin{aligned} f_t &\equiv \Lambda(\tilde{f}_t) \\ \tilde{f}_t &\equiv \omega + A\tilde{s}_t + \tilde{f}_{t-1}, \end{aligned} \tag{3.4}$$

dove

$$\tilde{s}_t \equiv \tilde{S}_t(\tilde{f}_t)\tilde{\nabla}_t(y_t, \tilde{f}_t)$$

e  $\tilde{\nabla}_t(y_t, \tilde{f}_t)$  rappresenta lo score della densità definita in 3.1 rispetto a  $\tilde{f}_t$  e, di conseguenza,  $\tilde{S}_t(\tilde{f}_t)$  può dipendere dalla matrice d'informazione di  $\tilde{f}_t$  data da  $\tilde{\mathcal{I}}_t(\tilde{f}_t)$ . Si indica la matrice Jacobiana di  $\Lambda(\cdot)$  valutata su  $\tilde{f}_t$  come segue:

$$\mathcal{J}(\tilde{f}_t) \equiv \frac{\partial \Lambda(\tilde{f}_t)}{\partial \tilde{f}_t}.$$

Allora, valgono le seguenti relazioni:

$$\begin{aligned} \tilde{\nabla}_t(y_t, \tilde{f}_t) &= \mathcal{J}(\tilde{f}_t)^T \nabla_t(y_t, f_t) \\ \tilde{\mathcal{I}}_t(\tilde{f}_t) &= \mathcal{J}(\tilde{f}_t)^T \mathcal{I}_t(f_t) \mathcal{J}(\tilde{f}_t). \end{aligned}$$

In questo modo quasi tutti i vincoli non lineari possono essere facilmente gestiti tramite la definizione di una corretta funzione di associazione  $\Lambda(\cdot)$  e la relativa matrice Jacobiana associata  $\mathcal{J}(\cdot)$ . I coefficienti da stimare sono raccolti in  $\theta \equiv (\omega, A, B)$  e stimati massimizzando la funzione di log-verosimiglianza come specificato nel successivo paragrafo 3.3.

### 3.3 Stima di massima verosimiglianza

Una proprietà vantaggiosa dei modelli GAS è che, date le informazioni passate e il vettore di parametri statici  $\theta$ , il vettore di parametri variabili nel tempo,  $f_t$ , è perfettamente prevedibile e la funzione di log-verosimiglianza può essere facilmente valutata mediante la decomposizione dell'errore di previsione. Più precisamente, per una serie temporale composta dalle osservazioni  $(y_1, \dots, y_t)$  si può esprimere il problema di massimizzazione come:

$$\hat{\theta} = \arg \max_{\theta} \sum_{t=1}^n l_t \quad (3.5)$$

dove

$$l_t = \ln p(y_t | f_t, \mathcal{F}_t, \theta)$$

per una realizzazione di  $y_t$ . La valutazione della funzione di log-verosimiglianza del modello GAS è particolarmente semplice; richiede solo l'implementazione dell'equazione di aggiornamento GAS 3.2 e la valutazione di  $l_t$  per un particolare valore  $\theta^*$  di  $\theta$ .

Ci sono due importanti precisazioni da fare sulla stima di massima verosimiglianza dei modelli GAS. La prima è che, da una prospettiva teorica, la stima di massima verosimiglianza dei modelli GAS è un argomento di ricerca ancora in corso. I risultati generali sono riportati da Harvey (2013), Blasques (2014a) e Blasques (2014b), mentre risultati validi per alcuni modelli specifici sono stati derivati da Andres (2014) e da Blasques (2016).

La seconda è che, anche quando lo stimatore di massima verosimiglianza è consistente e asintoticamente normale, la massimizzazione numerica della

---

funzione di log-verosimiglianza in 3.5 può essere laboriosa, a causa delle non linearità indotte da  $\Lambda(\cdot)$  e dal modo in cui  $y_t$  entra in  $s_t$ . Di conseguenza è necessario selezionare dei buoni valori iniziali per i modelli GAS.



## Capitolo 4

# TV-ARFIMA con $d$ stimato con un modello GAS

Come è già stato introdotto nel capitolo 2, l'assunzione che in un modello ARFIMA il parametro di integrazione frazionaria sia costante nel tempo è limitativa, data la possibile presenza di *break* strutturali in  $d$ . Quindi il parametro di integrazione frazionaria può essere variabile nel tempo e la formula 2.1 definisce il modello TV-ARFIMA( $p, d, q$ ).

Anche se questo modello indica che il parametro di integrazione frazionaria varia in funzione del tempo, non fornisce informazioni sull'evoluzione di  $d_t$ . Nel capitolo 2, per stimare il parametro di memoria lunga variabile nel tempo, è stato proposto prima il modello SETAR e poi un modello STAR, dove  $d_t$  viene stimato in un caso con la "regressione organizzata" (Boutahar (2008)) e nell'altro con un metodo basato sull'approccio *wavelet* (Boubaker e Sghaier (2014)).

L'idea di base di questo capitolo è sempre quella di tenere conto di eventuali modifiche strutturali nella memoria dei dati. Viene perciò proposta un'alternativa per stimare il parametro di memoria lunga variabile nel tempo e lo si fa utilizzando il modello GAS proposto da Creal (2013) e Harvey (2013).

Sono state, quindi, condotte alcune simulazioni per determinare la correttezza del modello proposto.

## 4.1 Specificazione del modello

In questo paragrafo viene estesa la classe dei modelli ARFIMA( $p, d, q$ ) in 1.3 permettendo che il parametro di memoria lunga  $d$  cambi nel tempo. Le dinamiche del parametro variabile nel tempo  $d_t$  sono specificate sulla base della struttura di *score* di Creal (2013) e Harvey (2013). Il modello TV-ARFIMA con parametro di memoria stimato da un modello GAS è descritto dalle seguenti equazioni:

$$\Phi(B)(1 - B)^{d_t}y_t = \Theta(B)\varepsilon_t,$$

$$d_{t+1} = \omega + \beta d_t + \alpha s_t$$

dove  $\varepsilon_t$  è una successione di *white noise* con varianza finita ( $\varepsilon_t \sim w.n(0, \sigma_\varepsilon^2)$ ),  $\theta = (\omega, \beta, \alpha, \sigma_\varepsilon^2)$  è un vettore di parametri, di dimensione  $(k + 3)$ , da stimare e  $s_t = s_t(d_t, \sigma_\varepsilon^2) = S_t \nabla_t$ , con  $S_t$  matrice scalare definita al paragrafo 3.1,  $\nabla_t$  esprime lo *score* della log-verosimiglianza, ossia  $\partial \log p(y_t | d_t, y_{t-1}, \sigma_\varepsilon^2) / \partial d_t$ .

Per semplicità si suppone che  $\Phi(B) = \Theta(B) = 1$ , cioè si sta considerando il più semplice ARFIMA( $\theta, d_t, \theta$ ), ovvero un I( $d$ ) con un parametro frazionario variabile nel tempo. Si considera quindi il modello:

$$\begin{cases} (1 - B)^{d_t} y_t = \varepsilon_t, \\ d_{t+1} = \omega + \beta d_t + \alpha s_t \end{cases} \quad (4.1)$$

Inoltre, al fine di ottenere lo *score* della funzione di log-verosimiglianza, si è deciso di scrivere il modello ARFIMA( $\theta, d_t, \theta$ ) citato all'equazione 4.1 come un modello AR( $\infty$ ) (Palma (2007)). Ovvero

$$(1 - B)^{d_t} y_t = \varepsilon_t = y_t + \sum_{j=1}^{\infty} \pi_j y_{t-j},$$

dove

$$\pi_j = \prod_{0 < k \leq j} \frac{k - 1 - d_t}{k} \quad \text{oppure} \quad \pi_j = -\frac{d_t \Gamma(j - d_t)}{\Gamma(1 - d_t) \Gamma(j + 1)}. \quad (4.2)$$

Quindi per calcolare  $\nabla_t$  va derivata rispetto a  $d_t$  la funzione di log-verosimiglianza

$$l_t(d_t) = -\frac{1}{2\sigma_\varepsilon^2} \left( y_t + \sum_{j=1}^{t-1} y_{t-j} \pi_j \right)^2.$$

Se si considera la prima uguaglianza dell'equazione 4.2, si ottiene pertanto:

$$\begin{aligned} \nabla_t &= -\frac{1}{\sigma_\varepsilon^2} \left( y_t + \sum_{j=1}^{t-1} y_{t-j} \prod_{0 < k \leq j} \frac{k - 1 - d_t}{k} \right) \cdot \\ &\quad \left( \sum_{j=1}^{t-1} y_{t-j} \sum_{k=1}^j \left( -\frac{1}{k} \right) \prod_{i=1, i \neq k}^j \frac{i - 1 - d_t}{i} \right) \end{aligned} \quad (4.3)$$

Se si considera la seconda uguaglianza dell'equazione 4.2, è possibile scrivere una forma equivalente dello *score* della log-verosimiglianza corrispondente a:

$$\nabla_t = -\frac{1}{\sigma_\varepsilon^2} \left( y_t + \sum_{j=1}^{t-1} y_{t-j} \pi_j \right) \left( \sum_{j=1}^{t-1} y_{t-j} \pi_j \left( -\frac{1}{d_t} \Psi(j - d_t) + \Psi(1 - d_t) \right) \right),$$

dove  $\Psi(\cdot)$  è la funzione digamma.

Si fa notare che lo *score*  $s_t$  può essere visto come l'innovazione del sistema dinamico in 4.1 in quanto fornisce le nuove informazioni che diventano disponibili al tempo  $t$  osservando  $y_t$ . L'interpretazione di  $s_t$  come innovazione è ulteriormente giustificata dal fatto che il suo valore atteso condizionato  $\mathbb{E}[s_t|y_{t-1}, d_t]$  è uguale a zero.

### 4.1.1 Stima dei parametri

Il vettore di parametri statici  $\theta$  del modello appena proposto può essere stimato con la massima verosimiglianza. La funzione di log-verosimiglianza da massimizzare è la seguente:

$$\begin{aligned} \hat{l}(\theta) &= \sum_{t=1}^n \log p(y_t | \hat{d}_t(\theta), y_{t-1}, \sigma_\varepsilon^2) = \\ &= -\frac{n}{2} \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n \left( y_t + \sum_{j=1}^{t-1} y_{t-j} \pi_j \right)^2, \end{aligned}$$

dove  $\hat{d}_t(\theta)$  è ottenuto ricorsivamente usando i dati osservati  $\{y_t\}_{t=0}^n$  come

$$\hat{d}_{t+1}(\theta) = \omega + \beta \hat{d}_t(\theta) + \alpha s_t(\hat{d}_t(\theta), \sigma_\varepsilon^2), \quad (4.4)$$

dove la ricorsione viene inizializzata a un valore fissato  $\hat{d}_0(\theta)$ . Perciò lo stimatore di massima verosimiglianza è dato da:

$$\hat{\theta} = \arg \sup_{\theta \in \Theta} \hat{l}(\theta), \quad (4.5)$$

dove  $\Theta$  è un insieme di parametri definito in  $\mathbb{R} \times (-1, 1) \times \mathbb{R} \times \mathbb{R}$ .

## 4.2 Simulazioni

Per valutare la correttezza del modello proposto, abbiamo effettuato alcune simulazioni. L'analisi è stata eseguita in due differenti scenari che valutano la presenza di persistenza o meno della serie storica. Nel primo scenario sono state simulate 30 serie storiche di lunghezza con  $n = 500$  osservazioni, dove le prime 250 presentano memoria lunga e le restanti non presentano più persistenza. Invece nel secondo scenario sono state simulate 30 serie storiche con  $n = 400$  osservazioni, dove viene mostrata la situazione opposta; ovvero nella prima metà dei dati si ha memoria breve e nell'ultima parte c'è persistenza.

Per costruire il modello definito in 4.1, per semplicità si è deciso di assumere che  $S_t$  sia pari alla matrice identità. Si procede, quindi, definendo il valore iniziale di  $\hat{d}_0(\theta)$  come lo stimatore di massima verosimiglianza della serie generata da  $\{y_t\}_{t=0}^n$  e  $\nabla_0 = 0$ . Si passa, poi, all'aggiornamento di  $d_t$  definito dalla formula 4.1 e di  $\nabla_t$  definito in 4.3, dove per ogni  $t$  viene calcolata la stima di massima verosimiglianza del vettore di parametri  $\theta = (\omega, \beta, \alpha, \sigma_\varepsilon^2)$ , che si comportano perciò come filtri.

Tutti i calcoli e le simulazioni sono state effettuate utilizzando il software **R**.

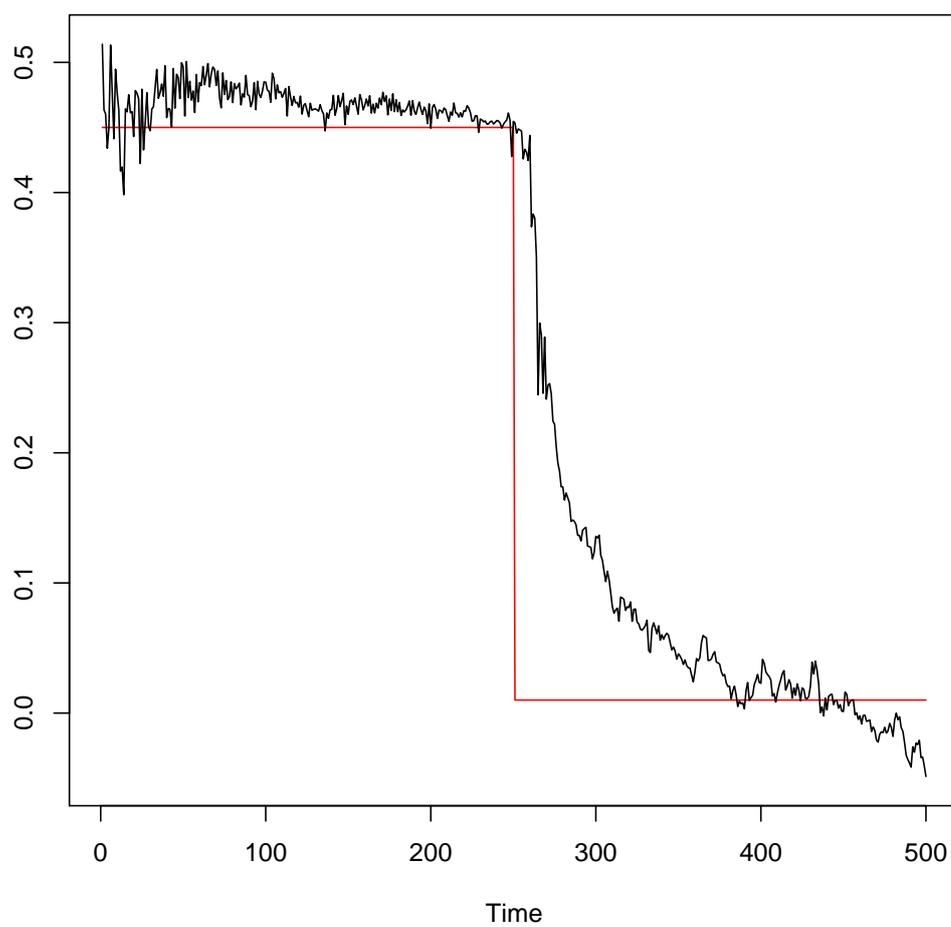
Nel grafico 4.1 la linea rossa indica il vero andamento di  $d$  della serie storica simulata, dove nella prima metà dei dati il parametro a memoria lunga è stato posto pari a 0.45 (presenza di persistenza) e nell'ultima parte  $d = 0$  (assenza di persistenza). La linea nera, invece, rappresenta l'andamento medio di  $d$ , calcolato su 30 replicazioni, stimato col metodo proposto in questo capitolo. Si può notare che la curva di  $d_t$  non si adatta benissimo al vero valore

di  $d$ , infatti il metodo ci mette un po' a identificare il cambiamento di regime.

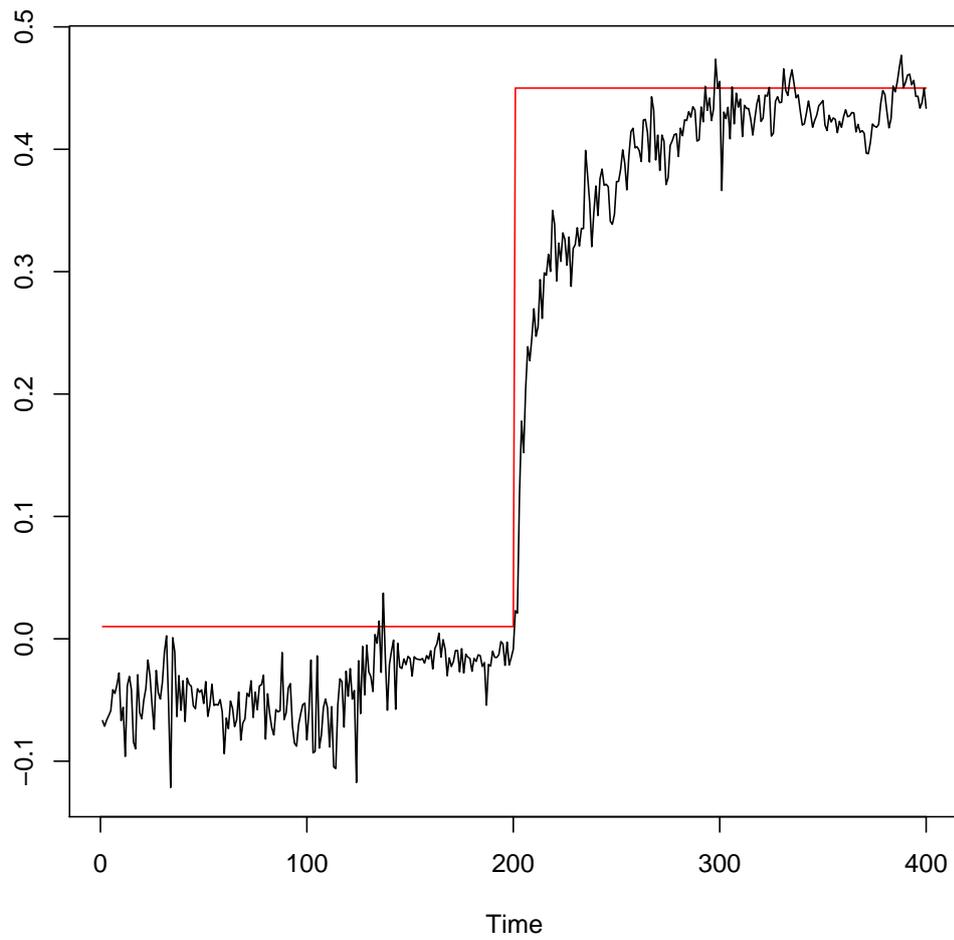
La linea rossa del grafico 4.2 mostra sempre il vero andamento di  $d$  della serie generata e in questo caso si passa da una serie che non presenta persistenza, con  $d$  posto pari a zero, a una serie a memoria lunga, dove  $d = 0.45$ . Anche qui la linea nera rappresenta l'andamento medio di  $d$ , sempre calcolato su 30 repliche, stimato con un modello GAS. Si può osservare dal grafico che la curva simulata si adatta meglio, rispetto al caso precedente, al vero andamento di  $d_t$ . Questa, difatti, individua praticamente subito il cambiamento di regime del parametro a memoria lunga.

Si vede quindi che il metodo proposto funziona meglio quando si passa dall'assenza di persistenza alla presenza di memoria lunga. Inoltre, nel secondo scenario proposto, è necessaria anche una numerosità meno elevata per rilevare la variazione nell'andamento di  $d_t$ .

I risultati ottenuti in queste simulazioni mostrano che c'è ancora molto lavoro da fare per migliorare il modello proposto in questo lavoro di tesi. Prima di tutto, si potrebbe aumentare il numero di repliche per ogni scenario in maniera da ottenere un andamento di  $d_t$  più liscio. Poi, ci si aspetta anche che migliorino i risultati aumentando la numerosità della serie storica simulata. Nonostante i due casi trattati siano simmetrici, si è deciso di porre  $n = 500$  nel primo scenario e  $n = 400$  nel secondo, perchè nel secondo, quando si passa da non persistenza a persistenza, si riesce a individuare abbastanza bene il cambiamento di regime anche con un numero di osservazioni non troppo elevato, invece si è visto che il primo scenario fa più fatica sebbene sia stata posta una numerosità più elevata. Perciò si è ridotto il numero di osservazioni



**Figura 4.1:** Andamento medio di 30 serie storiche simulate di lunghezza 500, con  $d_t = 0.45$  nelle prime 250 osservazioni e  $d_t = 0$  nelle restanti 250.



**Figura 4.2:** Andamento medio di 30 serie storiche simulate di lunghezza 400, con  $d_t = 0$  nelle prime 200 osservazioni e  $d_t = 0.45$  nelle restanti 200.

del secondo, perché l'algoritmo di stima ha un costo computazionale molto elevato, dato che per ogni  $n$  va massimizzata la funzione di verosimiglianza, e impiega molto tempo a stimare il modello. Infatti un'altro sviluppo futuro è senz'altro quello di ottimizzare quest'algoritmo di stima. Infine, si è detto nel capitolo 3 che in base alla scelta della matrice di scala  $S_t$  il modello GAS consente un'ulteriore flessibilità nel modo in cui lo score viene utilizzato per l'aggiornamento di  $d_t$ .  $S_t$  va posta a una potenza  $\gamma > 0$  dell'inversa della matrice d'informazione di  $d_t$ , ovvero  $S_t \equiv \mathcal{I}_t^{-\gamma}$ , dove  $\gamma$  solitamente assume un valore tra  $\{0, \frac{1}{2}, 1\}$ .

Ci si aspetta, quindi, che ponendo il grado di  $S_t$  diverso da 0 (come è stato fatto in questo elaborato), l'andamento di  $d_t$  stimato si avvicini di più a quello vero. La matrice di informazione calcolata rispetto a  $d_t$  è  $\mathcal{I}_t = \mathbb{E}[\mathcal{J}_t]$ , dove

$$\begin{aligned} \mathcal{J}_t = & \left( \sum_{j=1}^{t-1} y_{t-j} \sum_{k=1}^j \left( -\frac{1}{k} \right) \prod_{i=1, i \neq k}^j \frac{i-1-d_t}{i} \right)^2 + \\ & + \left( y_t + \sum_{j=1}^{t-1} y_{t-j} \prod_{0 < k \leq j} \frac{k-1-d_t}{k} \right). \end{aligned} \quad (4.6)$$

$$\left( \sum_{j=1}^{t-1} y_{t-j} \sum_{k=1}^j \left( -\frac{1}{k} \right) \left( \sum_{i=1, i \neq k}^j \left( -\frac{1}{i} \right) \right) \prod_{h=1, h \neq i, h \neq k}^j \frac{h-1-d_t}{h} \right).$$



# Conclusioni

Il questo lavoro è stata avanzata una nuova proposta di stima del parametro a memoria lunga variabile nel tempo di un modello ARFIMA, supponendo che questo si evolva secondo un modello GAS. Dopo aver definito le caratteristiche del modello è stato condotto uno studio di simulazione per determinare l'adeguatezza del modello proposto. In particolare sono stati creati due grafici che riassumono i risultati ottenuti dallo studio di simulazione.

I risultati ottenuti mostrano che il metodo proposto funziona meglio quando si passa dall'assenza di persistenza alla presenza di memoria lunga, piuttosto che il contrario.

C'è ancora molto lavoro da fare per migliorare il modello proposto in questa tesi. Prima di tutto, si potrebbe aumentare la numerosità della serie storica simulata e per far ciò, va sicuramente ottimizzato l'algoritmo di stima. Questo infatti ha un costo computazionale molto elevato, dato che per ogni osservazione della serie storica va massimizzata la funzione di verosimiglianza.

Inoltre, si potrebbe assumere che la matrice di scala  $S_t$  sia diversa dalla matrice identità (come è stato posto nel seguente lavoro). In questo modo ci si aspetta una maggiore flessibilità nel modo in cui lo *score* viene utilizzato

per l'aggiornamento di  $d_t$  e che l'andamento del parametro stimato si avvicini di più a quello vero.

Infine, un'estensione che si potrebbe fare per portare a una nuova formulazione del modello proposto, potrebbe essere quella di includere anche i parametri autoregressivi e a media mobile nel modello ARFIMA con parametro frazionario variabile nel tempo.

# Bibliografia

- Abry, P. e D. Veitch (1998). «Wavelet analysis of long-range-dependent traffic». In: *IEEE transactions on information theory* 44.1, pp. 2–15.
- Andres, P. (2014). «Maximum likelihood estimates for positive valued dynamic score models; The DySco package». In: *Computational Statistics & Data Analysis* 76, pp. 34–42.
- Baillie, R. T. e C. Morana (2009). «Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach». In: *Journal of Economic Dynamics and Control* 33.8, pp. 1577–1592.
- Bauwens, L. e N. Hautsch (2006). «Stochastic conditional intensity processes». In: *Journal of Financial Econometrics* 4.3, pp. 450–493.
- Beltratti, A. e C. Morana (2006). «Breaks and persistency: macroeconomic causes of stock market volatility». In: *Journal of econometrics* 131.1, pp. 151–177.
- Benjamin, M. A. et al. (2003). «Generalized autoregressive moving average models». In: *Journal of the American Statistical association* 98.461, pp. 214–223.
- Beran, J. (1994). *Statistics for Long-Memory Processes*. Taylor & Francis.

- Beran, J. (2009). «On parameter estimation for locally stationary long-memory processes». In: *Journal of Statistical Planning and Inference* 139.3, pp. 900–915.
- Blasques, F. et al. (2014a). *Maximum Likelihood Estimation for correctly specified Generalized Autoregressive Score Models: Feedback Effects, Contraction Conditions and Asymptotic Properties*. Rapp. tecn. Tinbergen Institute Discussion Paper.
- (2014b). *Maximum likelihood estimation for generalized autoregressive score models*.
- (2016). «Spillover dynamics for systemic risk measurement using spatial financial time series models». In: *Journal of Econometrics* 195.2, pp. 211–223.
- Bollerslev, T. (1986). «Generalized autoregressive conditional heteroskedasticity». In: *Journal of econometrics* 31.3, pp. 307–327.
- Boubaker, H., N. Sghaier et al. (2014). *Wavelet based Estimation of Time-Varying Long Memory Model with Nonlinear Fractional Integration Parameter*. Rapp. tecn.
- Boutahar, M. et al. (2008). «A Simple Fractionally Integrated Model with a Time-varying Long Memory Parameter  $dt$ ». In: *Computational Economics* 31.
- Cipollini, F. et al. (2006). *Vector multiplicative error models: representation and inference*. Rapp. tecn. National Bureau of Economic Research.

- Cox, D. R. (1981). «Statistical analysis of time series: Some recent developments [with discussion and reply]». In: *Scandinavian Journal of Statistics*, pp. 93–115.
- Creal, D. et al. (2013). «Generalized autoregressive score models with applications». In: *Journal of Applied Econometrics* 28.5, pp. 777–795.
- Diebold, F. X. e A. Inoue (2001). «Long memory and regime switching». In: *Journal of econometrics* 105.1, pp. 131–159.
- Dufrénot, G. et al. (2005a). «Modelling squared returns using a SETAR model with long-memory dynamics». In: *Economics Letters* 86.2, pp. 237–243.
- (2005b). «Long-memory dynamics in a SETAR model – applications to stock markets». In: *Journal of International Financial Markets, Institutions and Money* 15.
- (2006). «Changing-regime volatility: a fractionally integrated SETAR model». In: *Applied Financial Economics* 18.
- Engle, R. (2002). «Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models». In: *Journal of Business & Economic Statistics* 20.3, pp. 339–350.
- Engle, R. F. (1982). «Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation». In: *Econometrica: Journal of the Econometric Society*, pp. 987–1007.
- Engle, R. F. e J. R. Russell (1998). «Autoregressive conditional duration: a new model for irregularly spaced transaction data». In: *Econometrica*, pp. 1127–1162.

- Geweke, J. e S. Porter-Hudak (1983). «The estimation and application of long memory time series models». In: *Journal of time series analysis* 4.4, pp. 221–238.
- Granger, C. W. J e R. Joyeux (1980). «An introduction to long-memory time series models and fractional differencing». In: *Journal of Time Series Analysis* 1.1, pp. 15–29.
- Granger, C. WJ e N. Hyung (2004). «Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns». In: *Journal of empirical finance* 11.3, pp. 399–421.
- Hall, P. (1997). *Defining and measuring long-range dependence*. In *Nonlinear Dynamics and Time Series (Montreal, PQ, 1995)*.
- Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*. Vol. 52. Cambridge University Press.
- Hosking, J. R. M. (1981). «Fractional Differencing». In: *Biometrika* 68.1, pp. 165–176.
- Jensen, M. J. (1999a). «An approximate wavelet MLE of short-and long-memory parameters». In: *Studies in Nonlinear Dynamics & Econometrics* 3.4.
- (1999b). «Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter». In: *Journal of Forecasting* 18.1, pp. 17–32.
- Koopman, S. J. et al. (2008a). *A General Framework for Observation Driven Time-Varying Parameter Models*.

- (2008b). «The multi-state latent factor intensity model for credit rating transitions». In: *Journal of Econometrics* 142.1, pp. 399–424.
- McCoy, EJ e AT Walden (1996). «Wavelet analysis and synthesis of stationary long-memory processes». In: *Journal of computational and Graphical statistics* 5.1, pp. 26–56.
- Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*. 1<sup>a</sup> ed. Wiley Series in Probability and Statistics. Wiley-Interscience.
- Percival, D.B. e A.T. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press.
- Robinson, P. M. (1995). «Log-Periodogram Regression of Time Series with Long Range Dependence». In: *Ann. Statist.* 23.3, pp. 1048–1072.
- Roueff, F. e R. Von Sachs (2011). «Locally stationary long memory estimation». In: *Stochastic Processes and their Applications* 121.4, pp. 813–844.
- Russell, J. R. (2001). «Econometric modeling of multivariate irregularly-spaced high-frequency data». In: *Manuscript, GSB, University of Chicago*.
- Shephard, N. (1995). «Generalized linear autoregressions.» In:
- (2005). *Stochastic volatility: selected readings*. Oxford University Press on Demand.
- Teräsvirta, T. (1994). «Specification, estimation, and evaluation of smooth transition autoregressive models». In: *Journal of the American Statistical Association* 89.425, pp. 208–218.

- 
- Teräsvirta, T. (1996). *Modelling Economic Relationships with Smooth Transition Regressions*. Working paper series in economics and finance. EFI.
- Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Clarendon Press.
- Tong, H. e K. S. Lim (1980). «Threshold autoregression, limit cycles and cyclical data». In: *Journal of the Royal Statistical Society, Series B: Methodological* 42, pp. 245–292.
- Whitcher, B. e M. J. Jensen (2000). «Wavelet estimation of a local long memory parameter». In: *Exploration Geophysics* 31.1/2, pp. 94–103.