



Università degli Studi di Padova

FACOLTÀ DI SCIENZE STATISTICHE
Corso di Laurea Magistrale in Scienze Statistiche

**METODI MONTE CARLO PER LA STIMA
DI MODELLI A VOLATILITÀ STOCASTICA**

Relatore:
Prof. SILVANO BORDIGNON

Correlatore:
Dott. DAVIDE RAGGI

Laureando:
MARCO BAZZI
Matricola: 619529

Anno Accademico 2010-2011

Indice

1	Origini e principali proprietà del modello a Volatilità Stocastica	1
1.1	Attività finanziarie: fatti stilizzati	1
1.2	Modelli per la volatilità	3
1.3	Proprietà probabilistiche dei modelli SV	4
1.4	Inferenza dei modelli a Volatilità Stocastica	8
1.4.1	Rapida rassegna della letteratura	8
2	Modelli <i>state space</i> lineari e Gaussiani e stima QMV del modello SV	11
2.1	Modelli <i>state space</i> lineari e Gaussiani	11
2.1.1	Introduzione	11
2.1.2	Filtro di Kalman	12
2.1.3	<i>Smoothing</i>	14
2.1.4	Stima di Massima Verosimiglianza	15
2.1.5	Il <i>Simulation smoothing</i>	15
2.2	Stima di Quasi Massima Verosimiglianza del modello SV	17
3	Modelli <i>state space</i> parzialmente non Gaussiani e metodi MCMC	21
3.1	Modelli <i>state space</i> parzialmente non Gaussiani	21
3.1.1	Il modello	22
3.2	Teoria alla base dei metodi MCMC	23
3.2.1	Idea base	23
3.2.2	Principi essenziali	23
3.2.3	Algoritmi MCMC	28
3.3	Le densità mistura	32
3.3.1	Introduzione	32
3.3.2	Misure di normali	33
3.4	Applicazione al modello SV	37
3.4.1	Un semplice algoritmo Gibbs <i>sampler</i>	37

3.4.2	Algoritmo applicato al modello linearizzato approssimato da una mistura di Normali	41
4	Modelli <i>state space</i> generalizzati e <i>Importance Sampling</i>	49
4.1	Inferenza di modelli <i>state space</i> generalizzati	49
4.1.1	Introduzione	49
4.1.2	Funzione di verosimiglianza	50
4.1.3	L' <i>Importance sampling</i>	50
4.2	Il metodo di Shephard, Pitt, Durbin e Koopman	52
4.2.1	Presentazione del metodo	52
4.2.2	Stima dei parametri e degli <i>standard errors</i>	56
4.3	Applicazione del metodo SPDK al modello a Volatilità Stocastica	57
4.3.1	Modello non Gaussiano	57
5	L' <i>Efficient Importance Sampling</i>	65
5.1	Presentazione dell'algoritmo	65
5.1.1	Filtraggio	69
5.2	Applicazione dell'EIS al modello SV	71
5.2.1	Il modello	71
5.2.2	Implementazione dell'algoritmo	71
6	Analisi di una serie reale e conclusioni	75
6.1	Verifica empirica	75
6.1.1	Analisi preliminari	76
6.1.2	Stima dei parametri	77
6.1.3	Analisi grafiche	79
6.2	Conclusioni	81
	Bibliografia	87

Introduzione

Le scelte d'investimento in attività finanziarie sono caratterizzate da incertezza: non è possibile conoscere, basandoci su cosa è successo nel passato, cosa capiterà nel futuro. Tuttavia, pur rimanendo i prezzi imprevedibili, concentrarsi sui rendimenti delle attività permette di ottenere preziose informazioni sulla loro rischiosità, intesa come variabilità. E' noto da molto tempo in letteratura, a partire dai lavori di Mandelbrot (1963) e Fama (1965) che i rendimenti presentano una volatilità non costante nel tempo, in un alternarsi di periodi dove le variazioni sono più contenute e periodi dove sono più ampie. Engle (1982) e Bollerslev (1986) hanno proposto di utilizzare, per spiegare l'elevata persistenza che caratterizza l'eteroschedasticità dei rendimenti, una classe di modelli chiamata GARCH, nella quale la varianza condizionale è funzione delle osservazioni e dei suoi valori passati.

Un'attraente alternativa è rappresentata dal considerare la varianza come un processo stocastico latente. Tali modelli a Volatilità Stocastica (SV) prendono inizialmente forma nella letteratura finanziaria per dare una risposta al problema del *pricing* delle opzioni (si veda, tra gli altri, Hull and White (1987)) connesso alla generalizzazione della formula di Black e Scholes. Assumere che i prezzi seguano un moto Browniano con varianza non più costante nel tempo, sembra adattarsi meglio alla realtà. Sebbene le tecniche di *pricing* di opzioni utilizzino processi di diffusione a tempo continuo, i modelli statistici e econometrici a Volatilità Stocastica sono stati inizialmente formulati in tempo discreto.

Il primo capitolo di questo lavoro presenta il modello mostrando le interessanti proprietà probabilistiche che lo caratterizzano. Tuttavia, l'eleganza teorica di pensare che l'evolversi nel tempo delle osservazioni sia guidato da un processo stocastico non osservabile, non è accompagnata da una facilità di stima.

I metodi basati sulla funzione di Verosimiglianza si scontrano con l'impossibilità di risolvere analiticamente o numericamente un integrale di dimensione uguale alla numerosità campionaria. Neppure il filtro di Kalman, grazie al quale la Verosimiglianza dei modelli *state space* è ottenuta dalla scomposizione dell'errore di previsione, è utilizzabile poiché il processo latente entra nel modello in maniera non lineare.

Il secondo capitolo è dedicato alla soluzione suggerita da Harvey et al. (1994) e Ruiz (1994), i quali riescono ad applicare il filtro di Kalman linearizzando il modello originario. Dal momento che in tale modello approssimato viene a cadere l'ipotesi di Normalità, dalla quale deriva l'ottimalità della procedura di Kalman, si rende necessario il ricorrere alla teoria della Quasi Massima Verosimiglianza.

Il terzo capitolo, muovendosi in un'ottica Bayesiana, mostra il metodo proposto da Shephard (1993) e Kim et al. (1998): utilizzando le *Markov Chain Monte Carlo* si ottiene un campione simulato utile per studiare le caratteristiche della densità a posteriori di interesse. La deviazione dalla Gaussianità è gestita dalla combinazione di una mistura di Normali e una variabile ausiliaria.

Nel quinto e nel sesto capitolo, ritornando in un ambito frequentista, un'altra tecnica di simulazione Monte Carlo, l'*Importance Sampling* permette di approssimare la Verosimiglianza accuratamente. Seguendo Shephard and Pitt (1997) e Durbin and Koopman (1997), si utilizza l'*Importance Sampling* per correggere le procedure legate al filtro di Kalman, mentre l'*Efficient Importance Sampling*, proposto da Richard and Zhang (2007), si prefigge di minimizzare direttamente l'imprecisione della stima della funzione di Verosimiglianza.

L'intento principale di questo lavoro è offrire un confronto tra questi diversi metodi, presentando gli strumenti teorici dai quali sono originati e testando la loro capacità inferenziale sia su un campione simulato sia, infine, su una serie reale.

Capitolo 1

Origini e principali proprietà del modello a Volatilità Stocastica

1.1 Attività finanziarie: fatti stilizzati

Prezzi e rendimenti

L'analisi delle serie storiche finanziarie studia l'evolversi nel tempo del valore delle attività finanziarie. L'attenzione non si concentra direttamente sui prezzi ma piuttosto sui rendimenti dal momento che, coerentemente con l'ipotesi di efficienza dei mercati, i prezzi contengono tutta l'informazione disponibile e la loro dinamica non è prevedibile. Da un punto di vista statistico, si assume per una serie di prezzi un processo *Random Walk*¹, grazie al quale si riesce a formalizzare l'incapacità di fare previsioni sui movimenti futuri degli stessi. Il passaggio ai rendimenti, r_t , intesi, d'ora in poi, come *rendimenti a capitalizzazione continua* e definiti da $r_t = p_t - p_{t-1}$, dove p_t e p_{t-1} sono i logaritmi dei prezzi dell'attività di interesse, osservati, rispettivamente, nell'istante temporale t e in quello $t - 1$, consente

¹Un processo Random Walk, $\{p_t\}$, è rappresentabile tramite la seguente espressione

$$p_t = p_{t-1} + a_t,$$

dove a_t è un rumore bianco con varianza uguale a σ_a^2 .
Per ogni orizzonte di previsione l , abbiamo

$$\hat{p}_h(l) = E[p_{h+l} | p_h, \dots, p_1] = E \left[\left(\sum_{j=1}^l a_{h+j} \right) + p_h \mid p_h, \dots, p_1 \right] = p_h,$$

dove $\sum_{j=1}^l a_{h+j}$ rappresenta l'errore di previsione ad l passi in avanti, $e_h(l)$. La stima puntuale della previsione, $\hat{p}_h(l)$, è sempre uguale all'ultimo valore osservato mentre la sua varianza, pari alla varianza dell'errore di previsione, diverge all'infinito, poiché è pari a $l\sigma_a^2$, rendendo la previsione sempre più incerta.

di modellare un processo che gode di caratteristiche statistiche più agevolmente maneggiabili².

Principali caratteristiche dei rendimenti

1. I rendimenti sono serialmente incorrelati nel tempo e quindi, quando presentano una media diversa da zero (che tuttavia rimane tipicamente molto piccola), questa può essere stimata consistentemente dalla media campionaria.
2. I rendimenti non sono indipendenti dal momento che, prendendone i quadrati o i valori assoluti, si ricavano serie con evidenti segnali di dipendenza temporale³.
3. Le varianze dei rendimenti non sono costanti nel tempo; la forte eteroschedasticità è caratterizzata da effetti di *clustering*, ovvero si alternano periodi di bassa e elevata variabilità.
4. La variabilità è negativamente correlata con il movimento dei prezzi, dal momento che, ragionevolmente, un aumento del prezzo ha un effetto negativo sulla variabilità e viceversa. Tale fenomeno prende il nome di *effetto leva*.
5. La variabilità è altamente persistente, in modo particolare per dati osservati ad elevate frequenze.
6. La distribuzione marginale dei rendimenti presenta code più pesanti rispetto a quelle della distribuzione Normale.

Per tener conto delle caratteristiche appena elencate, indicando con $\{r_t\}$ una serie di rendimenti, possiamo ipotizzare il seguente modello:

$$r_t = \sigma_t \epsilon_t, \tag{1.1}$$

dove ϵ_t è una sequenza di variabili indipendenti e identicamente distribuite secondo una Normale Standard o, per enfatizzare la leptocurtosi, secondo distribuzioni come la t -di Student, eventualmente riscalate in modo da avere una varianza unitaria; la formulazione di σ_t determina la dinamica della variabilità della serie, caratterizzata, come abbiamo detto, da una forte persistenza come anche dalla presenza, in alcune

²Applicando una differenza prima alla serie non stazionaria p_t , otteniamo la serie $r_t = p_t - p_{t-1} = a_t$ stazionaria dal momento che a_t è un Rumore Bianco (stazionario per definizione).

³L'indipendenza implica che per qualsiasi trasformazione misurabile $h(\cdot)$ e $g(\cdot)$ valga $E[h(r_i)g(r_{jj})] = 0$ per ogni $i \neq j$. Al contrario, le funzioni di autocorrelazione dei quadrati o dei valori assoluti dei rendimenti mostrano, tipicamente per i primi ritardi, segnali di dipendenza temporale.

situazioni, di effetti leva. Nel contesto delle serie finanziarie, la variabilità assume il nome di *volatilità* e il significato di una misura di rischio associata alle diverse attività finanziarie.

1.2 Modelli per la volatilità

Modelli GARCH

A partire dai pionieristici lavori di Engle (1982) e Bollerslev (1986), si è diffusa nella letteratura e nella pratica una particolare classe di modelli per le serie storiche finanziarie: *Generalized Autoregressive Conditional Heteroscedasticity*, per modellare, come il nome stesso suggerisce, la varianza condizionale. Per rendimenti a media nulla, la varianza condizionale, utilizzando la notazione introdotta nella 1.1, è definita come

$$Var[r_t|I_t] = Var[\sigma_t\epsilon_t|I_t] = E[\sigma_t^2|I_t]E[\epsilon_t^2|I_t] = E[\sigma_t^2|I_t]$$

dove I_t denota l'informazione disponibile al tempo t , le realizzazioni passate degli r_t , e dove le ultime uguaglianze valgono se gli ϵ_t sono assunti essere *White Noise* a varianza unitaria.

Nei modelli GARCH, la varianza condizionale, $E[\sigma_t^2|I_t]$, è una combinazione stazionaria e positiva dei quadrati delle osservazioni passate e delle varianze condizionali passate. La specificazione della volatilità condizionale di un GARCH(1,1) è uguale a

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2,$$

dove per garantire la necessaria positività devono valere i vincoli $\omega, \alpha, \beta > 0$ mentre per la stazionarietà debole deve valere $\alpha + \beta < 1$.

Modelli a Volatilità Stocastica

Un'attraente alternativa ai modelli GARCH appena presentati, è offerta dai modelli a Volatilità Stocastica, i quali considerano per la volatilità un processo latente autoregressivo positivo e stazionario.

Tali modelli sono stati inizialmente suggeriti nel contesto del *pricing* dei derivati, dove la volatilità gioca un ruolo fondamentale. Ricordiamo infatti, senza esplicitarla, che la nota formula di Black e Scholes assume per la dinamica del prezzo, un moto geometrico Browniano con varianza costante, implausibile, ad esempio,

per la presenza di *smile*. Se infatti, in un istante temporale generico, calcoliamo e inseriamo in un sistema cartesiano la *volatilità implicita* (nell'asse delle ordinate) per un insieme di derivati con un comune sottostante ma diversi prezzi di esercizio (nell'asse delle ascisse), non otteniamo una retta come l'ipotesi di varianza costante suggerirebbe, ma una curva, simile ad un sorriso. Sostituire la volatilità costante con una volatilità stocastica sembra spiegare questo fenomeno.

Ricordiamo, infine, che i modelli per il *pricing* delle opzioni sono formulati a tempo continuo e che quindi i processi discreti dei quali ci occupiamo in questo lavoro devono essere visti solo come un'approssimazione di questi.

1.3 Proprietà probabilistiche dei modelli SV

Per analizzare le proprietà probabilistiche del modello a Volatilità Stocastica (si vedano Davis and Mikosch 2000, 2009), riprendiamo la 1.1

$$r_t = \sigma_t \epsilon_t,$$

dove la volatilità σ_t è una sequenza strettamente stazionaria di variabili casuali positive, indipendente da quella dei disturbi IID, $\{\epsilon_t\}$.

Stazionarietà e positività

Se nei modelli GARCH devono essere posti particolari vincoli sui parametri, a garanzia della stazionarietà e della positività, nei modelli a Volatilità Stocastica, l'indipendenza tra il disturbo $\{\epsilon_t\}$ (stazionario per definizione) e la volatilità $\{\sigma_t\}$ implica che la stazionarietà del processo $\{r_t\}$ dipende esclusivamente da quella di $\{\sigma_t\}$.

Per assicurarci, contemporaneamente, che il processo della volatilità sia stazionario e positivo, è utile considerare una generica serie $\{Y_t\}$ stazionaria, e poi prenderne la trasformazione esponenziale, $\sigma_t = \exp(Y_t)$. Grazie alla stazionarietà, possiamo rappresentare $\{Y_t\}$ come una combinazione lineare di disturbi, attraverso la rappresentazione MA(∞),

$$Y_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i}, \tag{1.2}$$

dove $\{\eta_t\}$ è una sequenza di variabili casuali IID a media nulla e varianza finita e $\{\psi_t\}$ una sequenza di coefficienti deterministici con $\psi_0 = 1$. La stazionarietà di $\{Y_t\}$ implica che la serie $\sum_j \psi_j^2$ è finita, condizione sufficiente per la convergenza in

media quadratica della serie infinita 1.2.

Possiamo allora considerare per $\{\sigma_t\}$ un qualsiasi processo non deterministico purché stazionario, scelto, ad esempio, nella classe dei processi ARMA. Per una discussione più approfondita di questo aspetto rimandiamo a Brockwell and Davis (2009). Se, inoltre, assumiamo che gli η_t provengano da una normale a media nulla, la sequenza $\{\sigma_t\}$ avrà marginali log-normali⁴.

Modello log-normale e momenti marginali

Se consideriamo per $\{Y_t\}$ un processo AR(1) normale, possiamo riscrivere la 1.1, nel modo seguente:

$$\begin{aligned} r_t &= e^{\frac{\theta_t}{2}} \epsilon_t, \\ \theta_t &= \gamma + \phi\theta_{t-1} + \nu\eta_t \\ \theta_1 &\sim N\left(\frac{\gamma}{1-\phi}, \frac{\nu^2}{1-\phi^2}\right) \end{aligned} \tag{1.3}$$

dove ϵ_t e η_t sono disturbi serialmente e mutualmente incorrelati, provenienti da Normali a media nulla e varianza unitaria. Per la stazionarietà del processo, dobbiamo imporre $|\phi| < 1$. Tale modello, suggerito da Taylor (1982), costituisce la versione più semplice del modello a Volatilità Stocastica e prende il nome di Log-Normale per quanto detto a conclusione del precedente paragrafo. In accordo con l'ultima equazione della 1.3, la variabile latente al tempo 1, θ_1 , è assunta distribuirsi come una Normale di media e varianza pari ai momenti non condizionali del processo autoregressivo.

La media marginale di $E[r_t]$ è nulla mentre la varianza è data da

$$Var[y_t] = E[y_t^2] = E[e^{\theta_t}] = e^{\mu_\theta + \frac{\sigma_\theta^2}{2}},$$

dove si sono sfruttate ancora una volta le proprietà della log-normale. Le costanti μ_θ e σ_θ^2 rappresentano i momenti marginali del processo latente θ_t e sono date, rispettivamente da, $\gamma/(1-\phi)$ e $\nu^2/(1-\phi^2)$.

Una leggera riparametrizzazione del modello prevede di eliminare la costante γ dal processo autoregressivo e inserire un parametro β nell'equazione dei rendimenti:

$$\begin{aligned} r_t &= \beta e^{\frac{\theta_t}{2}} \epsilon_t, \\ \theta_t &= \phi\theta_{t-1} + \nu\eta_t \end{aligned} \tag{1.4}$$

⁴Ricordiamo infatti che per le note proprietà della log-normale, se $X \sim N(\mu, \sigma^2)$, $Y = \exp(X)$ si distribuisce come una log-normale, la cui media è uguale a $\mu + \frac{\sigma^2}{2}$.

$$\theta_1 \sim N\left(0, \frac{\nu^2}{1-\phi^2}\right)$$

dove β^2 rappresenta un parametro di scala, interpretabile come volatilità media della serie di rendimenti. Vale ovviamente una relazione biunivoca tra i parametri β e γ , facilmente ricavabile dall'espressione per la varianza marginale. Abbiamo infatti che $2 \log \beta = \frac{\gamma}{1-\phi}$. Nei metodi di stima che vedremo nei prossimi capitoli, sarà a volte più agevole trattare β piuttosto che γ .

Modello linearizzato

Nel modello 1.3 (e equivalentemente nel modello 1.4), la variabile latente entra nella specificazione dell'equazione per i rendimenti in modo non lineare e questo, come presto diremo, porterà a delle difficoltà quando vorremo fare inferenza sui parametri del modello. E', allora, possibile considerare una versione linearizzata prendendo il logaritmo dei quadrati dei rendimenti:

$$\begin{aligned} y_t = \log(r_t^2) &= \theta_t + \log(\epsilon_t^2), \\ \theta_{t+1} &= \gamma + \phi\theta_t + \nu\eta_t, & \eta_t &\sim N(0,1), \\ \log(\epsilon_t^2) &\perp \eta_s, & &\forall t, s; \end{aligned} \quad (1.5)$$

dove θ_t entra linearmente nell'equazione per le osservazioni ma dove non sarà più possibile assumere Normale il disturbo $\log(\epsilon_t^2)$.

Inoltre tale modello presenta una complicazione computazionale che sorge dal fatto che il logaritmo del quadrato dei rendimenti si può pericolosamente tradurre in valori molto piccoli. La soluzione suggerita da Fuller (1996) e da noi adottata, considera la seguente trasformazione:

$$y_t = \log r_t^2 \approx \log(r_t^2 + cs_r^2) - cs_r^2/(r_t^2 + cs_r^2) = y_t^*,$$

dove s_r^2 è la varianza campionaria di r_t e c una costante che assume un valore molto piccolo, noi considereremo 0.02. Se non diversamente specificato, quando stimeremo i parametri del modello linearizzato, ricorreremo sempre a questa trasformazione.

Struttura di covarianza

Usualmente, per analizzare la struttura di dipendenza di una generica sequenza stazionaria $\{A_t\}$, l'attenzione si concentra innanzitutto sulla funzione di autocovarianza (ACVF), γ_A , e sulla funzione di autocorrelazione (ACF), ρ_A . Purtroppo, poiché il processo della volatilità stocastica $\{\sigma_t\}$ è fortemente lontano dalla Nor-

malità, la sua struttura di covarianza non è molto informativa. Infatti, come viene mostrato più avanti, $\{r_t\}$ è incorrelato ma ancora dipendente. Maggiori informazioni sulla natura della dipendenza nei processi a Volatilità Stocastica possono essere ottenute studiando le funzioni di correlazione delle potenze del processo dei rendimenti e di quello della volatilità, date rispettivamente da $\{|r_t|^p\}$ e $\{\sigma_t^p\}$ per un qualche $p > 0$. Ricordando che $E(\epsilon_t) = 0$, $Var(\epsilon_t) = 1$ e $E[\exp(2|\epsilon|)] < \infty$, otteniamo

$$\begin{aligned}\gamma_r(h) &= \rho_r(h) = 0, \quad h > 0, \\ Var(r) &= Var(\sigma),\end{aligned}\tag{1.6}$$

$$Var(\sigma) = \prod_{j=0}^{\infty} m_{\eta}(2\psi_j) - \prod_{j=0}^{\infty} m_{\eta}^2(\psi_j),\tag{1.7}$$

dove $m_{\eta}(\epsilon) = E(e^{\epsilon\eta})$ denota la funzione generatrice dei momenti di η . Se gli η_i provengono da una Normale a media nulla e varianza $\nu^2 > 0$, allora abbiamo $m_{\eta}(\epsilon) = \exp(0.5\nu^2\epsilon^2)$. Quindi

$$Var(\sigma) = \exp\{Var(Y)\}(\exp\{var(Y)\} - 1).\tag{1.8}$$

Osserviamo che $\{r_t\}$ è una sequenza di variabili incorrelate. Come abbiamo già detto, questo è in accordo con i rendimenti finanziari reali. Per ottenere informazioni più preziose è diventato comune nell'analisi delle serie storiche finanziarie studiare l'ACVF/ACF dei valori assoluti e dei quadrati. Nel modello a Volatilità Stocastica si può sfruttare l'indipendenza tra $\{\sigma_t\}$ e $\{\epsilon_t\}$ per ottenere un'esplicita espressione per γ_r . In quello che segue, ancora una volta, ci concentriamo sul modello con disturbi IID distribuiti secondo una Normale a media nulla e varianza $\nu^2 > 0$ per ottenere le corrispondenti ACVF e ACF. L'ACVF di $\{Y_t\}$ è data da

$$\gamma_Y(h) = \nu^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}, \quad h \geq 0.\tag{1.9}$$

Calcoli simili a quelli che hanno portato all'equazioni 1.7 e 1.8 permettono di ottenere per ogni $p > 0$ e $h > 0$,

$$\begin{aligned}\gamma_{|r|^p}(h) &= (E(|\epsilon|^p))^2 \gamma_{\sigma^p}(h), \\ \gamma_{\sigma^p}(h) &= E[e^{p(Y_0+Y_h)}] - (E[e^{pY}])^2 \\ &= \exp\left\{(p\nu)^2 \sum_{i=0}^{\infty} \psi_i^2\right\} \left[\exp\left\{(p\nu)^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}\right\} - 1\right] \\ &= e^{p^2 \gamma_Y(0)} \left[e^{p^2 \gamma_Y(h)} - 1\right]\end{aligned}$$

dove assumiamo che $E[|\epsilon|^p] < \infty$. Poiché $\gamma_Y(h) \rightarrow 0$ al tendere di h ad ∞ , possiamo scrivere

$$\gamma_{\sigma^p}(h) = p^2 e^{p^2 \gamma_Y(0) \gamma_Y(h)}, \quad h \rightarrow \infty. \quad (1.10)$$

In maniera analoga

$$\begin{aligned} \text{Var}(\sigma^p) &= e^{p^2 \gamma_Y(0)} (e^{p^2 \gamma_Y(0)} - 1), \\ \text{Var}(|\sigma|^p) &= E(|\epsilon|^{2p}) e^{p^2 \gamma_Y(0)} (e^{p^2 \gamma_Y(0)} - 1) + \text{var}(|\epsilon|^p e^{p^2 \gamma_Y(0)}), \end{aligned}$$

dove abbiamo assunto che $E(|\epsilon|^{2p}) < \infty$. Infine, possiamo calcolare l'ACF di $\{|r_t|^p\}$ e $\{|\sigma_t|^p\}$ per $h > 0$:

$$\rho_{\sigma^p}(h) = \frac{e^{p^2 \gamma_Y(h)} - 1}{e^{p^2 \gamma_Y(0)} - 1} \sim \frac{p^2}{e^{p^2 \gamma_Y(0)} - 1} \gamma_Y(h), \quad (1.11)$$

$$\rho_{|r|^p}(h) = \frac{(E(|\epsilon|^p))^2}{E(|\epsilon|^p) + \text{var}(|\epsilon|^p) (e^{p^2 \gamma_Y(0)} - 1)^{-1}} \rho_{\sigma^p}(h). \quad (1.12)$$

Possiamo quindi concludere che γ_{σ^p} eredita il comportamento asintotico della ACVF di $\{Y_t\}$, γ_Y , e, a sua volta, $\gamma_{|r|^p}$ eredita il comportamento asintotico di γ_{σ^p} . In contrasto con quanto accade per i modelli GARCH, potendo selezionare per $\{Y_t\}$ un processo all'interno di una classe ampia, riusciamo a modellare il comportamento dell'ACVF della serie dei rendimenti assoluti in maniera flessibile.

1.4 Inferenza dei modelli a Volatilità Stocastica

1.4.1 Rapida rassegna della letteratura

La maggiore difficoltà nell'implementazione dei modelli a Volatilità Stocastica e il principale svantaggio rispetto ai modelli GARCH sono provocati dalla difficoltà di stimare tali modelli a causa della non osservabilità della volatilità, assunta essere un processo latente. Nei modelli ARCH, per costruzione, la funzione di Verosimiglianza (o di Quasi Verosimiglianza) è facilmente reperibile. Non avviene così nei modelli SV e questo ha ispirato un'ampia letteratura econometrica nella quale si possono individuare due approcci separati.

Gli stimatori costruiti a partire dai Momenti sono facilmente calcolabili ma hanno il grosso limite di essere altamente inefficienti. Taylor (1982) ha per primo utilizzato questo metodo, Melino and Turnbull (1990), hanno proposto l'applicazione del *Generalized Method of Moments*, (GMM). Andersen and Sørensen (1996) hanno

migliorato la qualità dell'inferenza attraverso una scelta accurata dei Momenti da includere. Rimane la limitazione di dover effettuare una seconda stima per far sul processo latente della volatilità.

Un miglioramento su quest'ultimo aspetto viene dato dall'usare un'approssimazione lineare del sistema basata sul filtro di Kalman, che permette di ottenere prima la stima puntuale dei parametri e poi consente di estrarre la serie *smoothed* della volatilità. Harvey et al. (1994) e Ruiz (1994) utilizzano il filtro di Kalman in un'ottica di Quasi Massima Verosimiglianza per il modello SV log-normale. Tuttavia la non corretta specificazione del modello influenza molto la qualità delle stime.

Si sono sviluppati metodi che sfruttano l'efficienza dell'inferenza basata sulla funzione di Verosimiglianza, non ottenibile analiticamente nei modelli a Volatilità Stocastica, attraverso una sua approssimazione tanto più precisa quanto più costosa in termini computazionali e quanto più raffinata per le tecniche sfruttate. Shephard (1993), Jacquier et al. (1994) e Kim et al. (1998) applicano i metodi Monte Carlo basati sulle Catene di Markov (MCMC) per ottenere, tramite simulazione, un'approssimazione dello stimatore di Massima Verosimiglianza (muovendosi in un paradigma frequentista) o della densità a posteriori (in un paradigma Bayesiano). Questi approcci funzionano bene in circostanze molto generali ma la loro efficacia dipende dall'abilità di generare densità *proposal* appropriate ed efficienti per densità condizionali che sono potenzialmente molto complesse.

Un'alternativa è rappresentata dai metodi che massimizzano la stima della funzione di Verosimiglianza, ottenuta tramite simulazione. Per ridurre la varianza di questa stima, viene adottata la tecnica dell'*Importance Sampling*. Richard and Zhang (2007) propongono l'*Efficient Importance Sampling* (EIS) e Liesenfeld and Richard (2003) lo adottano in questo contesto. Infine Durbin and Koopman (1997) usano l'*Importance Sampling* per trattare modelli *state space* non lineari e/o non Gaussiani, Sandmann and Koopman (1998), stimano con questa tecnica diverse tipologie di modelli a Volatilità Stocastica.

Dati simulati e *software*

Per fornire un'immediata presentazione dei metodi presentati in questo lavoro, abbiamo simulato una serie di $T = 3000$ osservazioni. Basandoci su questa serie presenteremo, attraverso grafici e tabelle, i risultati più significativi degli algoritmi proposti. Per la scelta dei valori dei parametri di questi dati simulati abbiamo seguito un approccio simile a quello di Jacquier et al. (1994) e Sandmann and Koopman (1998) i quali utilizzano il *coefficiente di variazione* per scegliere il valore dei parametri. L'espressione di questo coefficiente, dove le quantità in gioco sono

definite nella 1.3, è data da:

$$CV = \frac{Var(\theta)}{E[\theta^2]} = \exp\left(\frac{\nu^2}{1 - \phi^2}\right) - 1.$$

Possiamo associare ad alti valori di tale indicatore una forza relativa della Volatilità Stocastica mentre, se il CV assume valori bassi, la Volatilità Stocastica è molto vicina ad essere costante, rendendo indistinguibile la serie dei rendimenti da una serie *White Noise*.

Scegliamo quindi per il parametro autoregressivo, ϕ , un valore tipico delle situazioni reali, ovvero 0.95, per γ prendiamo -0.368 . Infine per ottenere un coefficiente di variazione vicino ad 1, dobbiamo fissare per ν il valore 0.26. Nel modello con $\gamma = 0$, β sarà uguale a 0.025.

I risultati di questo lavoro sono ottenuti utilizzando tre diversi *software* statistici, per i metodi esposti nel secondo e nel quarto capitolo si è utilizzato S-PLUS, sfruttando la libreria SsfPack che implementa gli algoritmi legati al filtro di Kalman per i modelli standard. Per il terzo capitolo si è utilizzato OX sfruttando la libreria SvPack che implementa gli algoritmi presentati in questo lavoro mentre per quanto riguarda il metodo esposto nel quinto capitolo si è usato il *free software* R.

Capitolo 2

Modelli *state space* lineari e Gaussiani e stima QMV del modello SV

I modelli *state space* considerano una serie storica come il risultato di un sistema dinamico perturbato da un disturbo casuale. In questo modo è possibile individuare agevolmente le diverse componenti come *trend*, *stagionalità*, *ciclo* e anche introdurre eventuali variabili *esplicative*. Tuttavia, nel modello preso in esame in questo lavoro, assumiamo che sia la variabilità delle nostre osservazioni (i rendimenti delle attività finanziarie depurati dalle loro medie) a essere determinata da un processo stocastico latente.

La Verosimiglianza dei modelli basati su una rappresentazione *state space*, è facilmente ottenibile, sfruttando una serie di risultati legati alla distribuzione Normale Multivariata, attraverso equazioni ricorsive. Presentiamo ora una breve introduzione a questi modelli, usando alla notazione adottata da Durbin and Koopman (2001).

2.1 Modelli *state space* lineari e Gaussiani

2.1.1 Introduzione

Il modello lineare Gaussiano può essere espresso nella seguente forma

$$\begin{aligned} y_t &= Z_t \alpha_t + \epsilon_t, & \epsilon_t &\sim N(0, H_t) \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t), \quad t = 1, \dots, n \end{aligned} \tag{2.1}$$

dove y_t è un vettore $p \times 1$ di osservazioni, α_t è un vettore non osservato $m \times 1$ di stati, R_t è una matrice di selezione composta da r colonne della matrice Identità I_m e le matrici di varianza H_t e Q_t sono non singolari.

Si assume che tutti o alcuni elementi delle matrici H_t , Q_t , Z_t , e T_t dipendano da un vettore dei parametri non noto, ψ .

L'idea sottostante al modello è che lo sviluppo nel tempo del sistema sia determinato da α_t in accordo con la seconda equazione di (2.1), ma poiché α_t non può essere osservato direttamente dobbiamo basare l'analisi sulle osservazioni y_t . La prima equazione di (2.1) è chiamata *equazione delle osservazioni* mentre la seconda *equazione degli stati*. Le matrici Z_t , T_t , R_t , H_t e Q_t sono inizialmente assunte note e i termini d'errore ϵ_t e η_t sono assunti essere indipendenti serialmente e congiuntamente ad ogni istante temporale. Le matrici Z_t e T_{t-1} possono dipendere da y_1, \dots, y_{t-1} . L'iniziale vettore degli stati α_1 è assunto essere estratto dalla distribuzione $N(a_1, P_1)$, dove a_1 e P_1 sono noti.

La prima equazione di (2.1) ha la struttura di un modello lineare di regressione dove il vettore dei coefficienti α_t varia nel tempo mentre la seconda equazione rappresenta un modello vettoriale autoregressivo del primo ordine.

Fondamentale nel contesto dei modelli *state space* è la distinzione tra la *previsione*, il *filtraggio* e il *lisciamento*: se per $s \leq t$ definiamo $y_s^t = (y_s, y_{s+1}, \dots, y_t)$ e in maniera simile il *segnale* $\theta_t = Z_t \alpha_t$, indicando la densità condizionale di Θ_t dato Y_1^s con $f_{t|s}(\theta_t | y_1^s)$, faremo previsione quando $s < t$, filtraggio quando $s = t$ e infine *smoothing* quando $s > t$.

2.1.2 Filtro di Kalman

Il filtro di Kalman, la cui origine si deve al pionieristico contributo di Kalman (1960), è costituito semplicemente da una serie di equazioni che permettono di calcolare in maniera ricorsiva la distribuzione condizionale del processo latente basandosi su proprietà elementari della Normale Multivariata. Dal momento che tutte le distribuzioni in 2.1 sono Normali, lo sono anche le distribuzioni condizionate di un sottoinsieme di variabili ad un altro sottoinsieme di variabili. Inoltre, ancora una volta in virtù della Normalità, conoscere i primi due momenti condizionali del processo latente è sufficiente a specificare l'intera distribuzione condizionale, $P(\alpha_t | Y_t)$. Se indichiamo con $a_{t+1} = E[\alpha_{t+1} | Y_t]$ il valore atteso condizionale della variabile latente α al tempo $t + 1$ e con $P_{t+1} = Var(\alpha_{t+1} | Y_t)$ la sua varianza condizionale, $p(\alpha_{t+1} | Y_t)$ è $N(a_{t+1}, P_{t+1})$. Dal momento che $\alpha_{t+1} = T_t \alpha_t + R_t \eta_t$,

abbiamo

$$\begin{aligned} a_{t+1} &= E(T_t \alpha_t + R_t \eta_t | Y_t) \\ &= T_t E(\alpha_t | Y_t), \end{aligned}$$

$$\begin{aligned} P_{t+1} &= \text{Var}(T_t \alpha_t + R_t \eta_t | Y_t) \\ &= T_t \text{Var}(\alpha_t | Y_t) T_t' + R_t Q_t R_t', \end{aligned}$$

per $t = 1, \dots, n$. Se indichiamo con v_t , l'errore di previsione ad un passo di y_t dato Y_{t-1} , ricaviamo che

$$v_t = y_t - E(y_t | Y_{t-1}) = y_t - E(Z_t \alpha_t + \epsilon_t | Y_{t-1}) = y_t - Z_t \alpha_t.$$

Quando Y_{t-1} e v_t sono fissati, anche Y_t è fissato e viceversa. Quindi $E(\alpha_t | Y_t) = E(\alpha_t | Y_{t-1}, v_t)$. Otteniamo quindi $E(v_t | Y_{t-1}) = E(y_t - Z_t \alpha_t | Y_{t-1}) = E(Z_t \alpha_t + \epsilon_t - Z_t \alpha_t | Y_{t-1}) = 0$. E di conseguenza $E(v_t) = 0$ e $\text{Cov}(y_j, v_t) = E(y_j E(v_t | Y_{t-1}')) = 0$ con $j = 1, \dots, t-1$.

Riusciamo ad ottenere¹

$$\begin{aligned} E(\alpha_t | Y_t) &= E(\alpha_t | Y_{t-1}, v_t) \\ &= E(\alpha_t | Y_{t-1}) + \text{Cov}(\alpha_t, v_t) (\text{Var}(v_t))^{-1} v_t \\ &= a_t + M_t F_t^{-1} v_t, \end{aligned}$$

dove $M_t = \text{Cov}(\alpha_t, v_t)$ e $F_t = \text{Var}(v_t)$. Vale inoltre, per la legge dei valori attesi iterati,

$$\begin{aligned} M_t &= \text{Cov}(\alpha_t, v_t) = E[E(\alpha_t (Z_t \alpha_t + \epsilon_t - Z_t a_t) | Y_{t-1})] \\ &= E[E(\alpha_t (\alpha_t - a_t)' Z_t | Y_{t-1})] = P_t Z_t', \\ F_t &= \text{Var}(Z_t \alpha_t + \epsilon_t - Z_t a_t) = Z_t P_t Z_t' + H_t. \end{aligned}$$

¹In virtù del fatto che se indichiamo con $W = (X, Y, Z)$ un vettore di variabili aleatorie Normali, anche W è normale e valgono i seguenti risultati

$$\begin{aligned} E[x|y, z] &= E[x|y] + \Sigma_{xz} \Sigma_{zz}^{-1}, \\ \text{Var}(x|y, z) &= \text{Var}(x|y) - \Sigma_{xz} \Sigma_{zz}^{-1} \Sigma'_{xz}. \end{aligned}$$

Possiamo quindi ricavare dalla seconda proprietà della Normale Multivariata, enunciata nella precedente nota

$$\begin{aligned}
\text{Var}(\alpha_t|Y_t) &= \text{Var}(\alpha_t|Y_{t-1}, v_t) \\
&= \text{Var}(\alpha_t|Y_{t-1}) - \text{Cov}(\alpha_t, v_t)[\text{Var}(v_t)]^{-1}\text{Cov}(\alpha_t, v_t)' \\
&= P_t - M_t F_t^{-1} M_t' \\
&= P_t - P_t Z_t' F_t^{-1} Z_t P_t.
\end{aligned}$$

E quindi otteniamo

$$P_{t+1} = T_t P_t L_t' + R_t Q_t R_t',$$

per $t = 1, \dots, n$ con $L_t = T_t - K_t Z_t$.

La ricorsività alla base del filtro di Kalman ci permette di aggiornare la nostra conoscenza del sistema ogni volta che una nuova osservazione è disponibile. Il vantaggio principale è che non dobbiamo invertire una matrice ($pt \times pt$) per stimare il modello ogni volta che la t -esima osservazione arriva ma dobbiamo invertire semplicemente la matrice ($p \times p$) F_t . Nel caso da noi preso in esame, abbiamo $p = 1$, ovvero F_t è uno scalare. Per facilità di lettura, riportiamo tutte insieme le equazioni che costituiscono il filtro:

$$\begin{aligned}
v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\
K_t &= T_t P_t Z_t' F_t^{-1}, & L_t &= T_t - K_t Z_t, \\
a_{t+1} &= T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t L_t' + R_t Q_t R_t',
\end{aligned} \tag{2.2}$$

per $t = 1, \dots, n$ e con a_1 media e P_1 varianza dello stato iniziale α_1 .

2.1.3 *Smoothing*

Come già accennato, lo *smoothing* rappresenta la stima del vettore dello stato, α_t , data l'intera serie di osservazioni y_1, \dots, y_n . Tale stima di α_t è ottenuta attraverso la sua media condizionale, $\hat{\alpha}_t = E(\alpha_t|y_1, \dots, y_t)$; inoltre possiamo anche ricavare la matrice di varianza dell'errore, $V_t = \text{Var}(\alpha_t - \hat{\alpha}_t) = \text{Var}(\alpha_t|y_1, \dots, y_t)$. Basandoci ancora una volta sui risultati teorici della Normale Multivariata, possiamo ricorrere ad una serie di equazioni ricorsive per ottenere queste quantità, $\hat{\alpha}_t$ e V_t :

$$\begin{aligned}
r_{t-1} &= Z_t' F_t^{-1} v_t, & N_{t-1} &= Z_t' F_{t-1} Z_t + L_t' N_t L_t, \\
\hat{\alpha}_t &= a_t + P_t r_{t-1}, & V_t &= P_t - P_t - P_t N_{t-1} P_t,
\end{aligned} \tag{2.3}$$

per $t = n, \dots, 1$ e fissando $r_n = 0$ e $N_n = 0$.

2.1.4 Stima di Massima Verosimiglianza

Nei modelli lineari Gaussiani la Verosimiglianza, per un dato insieme di parametri, può essere calcolata con una semplice applicazione del filtro di Kalman e quindi massimizzata numericamente per ottenere la stima di Massima Verosimiglianza dei parametri. La Verosimiglianza quando il vettore degli stati iniziali ha densità $N(a_1, P_1)$ con a_1 e P_1 noti è

$$L(Y_t) = p(y_1, \dots, y_n) = p(y_1) \prod_{t=2}^n p(y_t | Y_{t-1}), \quad (2.4)$$

dove $Y_t = (y_1, \dots, y_t)$. Nella pratica, generalmente si lavora con la Log-Verosimiglianza

$$\log L(Y_t) = \sum_{t=1}^n \log p(y_t | Y_{t-1}), \quad (2.5)$$

dove $p(y_1 | Y_0) = p(y_1)$. Poiché $E(y_t | Y_{t-1}) = Z_t a_t$, ponendo $v_t = y_t - Z_t a_t$, $F_t = \text{Var}(y_t | Y_{t-1})$ e sostituendo $N(Z_t a_t, F_t)$ per $p(y_t | Y_{t-1})$ in 2.5, otteniamo

$$\log L(Y_t) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n (\log |F_t| + v_t' F_t^{-1} v_t). \quad (2.6)$$

Dal momento che le quantità v_t e F_t sono calcolate ricorsivamente utilizzando il filtro di Kalman, $\log L(Y_t)$ è facilmente ottenibile. Poiché, come abbiamo già detto, v_t rappresenta l'errore medio di previsione ad un passo e F_t la sua varianza, si dice che la 2.6 è basata sulla *decomposizione dell'errore di previsione*. Se il vettore dello stato iniziale è ignoto la 2.6 deve essere modificata, per un approfondimento rimandiamo a Durbin and Koopman (2001).

2.1.5 Il *Simulation smoothing*

L'estrazione di campioni degli stati o dei disturbi condizionatamente alle osservazioni è chiamato *simulation smoothing*. Su questi campioni si svilupperà il metodo presentato nel quinto capitolo, dove tuttavia ci limiteremo ad utilizzare la possibilità di simulare i disturbi, sia dell'equazione di misura sia dell'equazione di transizione degli stati. Durbin and Koopman (2002) sviluppano un efficiente algoritmo per estrarre campioni casuali dei disturbi ϵ_t e η_t e degli stati α_t , qui ci limitiamo a presentare quanto ci servirà nel quinto capitolo.

Simulazione dei disturbi

Definiamo $w = (\epsilon'_1, \eta'_1, \dots, \epsilon'_n, \eta'_n)$, $\hat{w} = E(w|y)$ e $W = \text{var}(w|y)$. Dal momento che il modello è lineare e Gaussiano, la densità di $w|y$ è quella di una $N(\hat{w}, W)$. La

matrice W gode dell'importante proprietà di non dipendere da y . Dal momento che y è una esatta funzione lineare degli elementi di w , la matrice W è singolare; tuttavia, questa singolarità non ha effetti sui calcoli.

Per estrarre vettori casuali \tilde{w} da $p(w|y)$, simuliamo in maniera indipendente vettori y da $N(0, W)$ e aggiungiamo a queste estrazioni casuali il vettore noto \hat{w} .

La distribuzione di w è

$$w \sim N(0, \Omega), \quad \Omega = \text{diag}(H_1, Q_1, \dots, H_n, Q_n). \quad (2.7)$$

Sia w' un vettore casuale estratto da $p(w)$ e sia y' il vettore trasporto dei valori di y_t generati dall'estrazione di un vettore α'_1 da $p(\alpha_1)$ e sostituendo α_1 e w nell'equazione generale 2.1 con α'_1 e w' . Si calcoli $\hat{w}' = E(w'|y')$ usando il *disturbance smoother* definito più avanti nella 2.8. Poiché W è indipendente da y , $\text{Var}(w'|y') = W$. Di conseguenza, $w' - \hat{w}'$ è l'estrazione desiderata da $N(0, W)$. Sia $\tilde{w} = \hat{w} + w' - \hat{w}'$. Segue che \tilde{w} è un'estrazione dalla densità $p(w|y)$. In particolare, abbiamo

$$E(\tilde{w}|y) = E(\hat{w} + w' - \hat{w}'|y) = E(w' - \hat{w}'|y) + \hat{w} = \hat{w},$$

$$\text{Var}(\tilde{w}|y) = E\{(\hat{w} - w')(\hat{w} - w')'|y\} = W,$$

dal momento che $w' - \hat{w}'$ è indipendente da y .

Il vettore dei disturbi *smoothed* \hat{w}

$$\hat{w}_t = \begin{bmatrix} H_t F_t^{-1} & -H_t K_t' \\ 0 & Q_t R_t' \end{bmatrix} \begin{pmatrix} v_t \\ r_t \end{pmatrix}, \quad \hat{w} = (\hat{w}'_1, \dots, \hat{w}'_n)', \quad (2.8)$$

dove r_t è valutato grazie all'iterazione

$$r_{t-1} = Z_t F_t^{-1} v_t + L_t' r_t, \quad (2.9)$$

per $t = n, n-1, \dots, 1$ con $r_n = 0$.

2.2 Stima di Quasi Massima Verosimiglianza del modello SV

Consideriamo la versione linearizzata del modello a Volatilità Stocastica Log-Normale presentato nel primo capitolo e riproponiamola per facilitare la lettura di quanto segue:

$$\begin{aligned}\log(r_t^2) &= \theta_t + \log(\epsilon_t^2), \\ \theta_{t+1} &= \gamma + \phi\theta_t + \nu\eta_t.\end{aligned}\tag{2.10}$$

Mentre η_t continua ad essere Normale, poiché non è cambiato nulla nella specificazione del processo latente, $\log(\epsilon_t^2)$ si distribuisce come il logaritmo di un chi-quadro con un grado di libertà, χ_1^2 ; la media e la varianza di questa distribuzione possono essere approssimate rispettivamente da -1.27 e $\pi^2/2$.

Per ottenere un disturbo che sia a media nulla, possiamo considerare la seguente variante di 2.10

$$\begin{aligned}\log(r_t^2) &= -1.27 + \theta_t + \xi_t, \\ \theta_{t+1} &= \gamma + \phi\theta_t + \nu\eta_t,\end{aligned}\tag{2.11}$$

dove $\sigma_\xi^2 = \pi^2/2$.

Poiché ξ_t non è distribuito normalmente, non possiamo evidentemente considerare questo modello come *state space* lineare e Gaussiano e i risultati teorici presentati nella sezione precedente non possono più essere considerati validi. Tuttavia Harvey et al. (1994) e Ruiz (1994) suggeriscono di considerare questo modello, almeno approssimativamente, un modello *state space* Gaussiano, ottenendo dall'applicazione del filtro di Kalman una stima della funzione di Verosimiglianza basata sulla decomposizione dell'errore di previsione. Una simile procedura si inserisce nel contesto della stima di Quasi Massima Verosimiglianza. Usando i risultati ottenuti da Dunsmuir (1979), in Ruiz (1994), si mostra che per tale stimatore vale ancora la Normalità asintotica e viene fornita una stima della matrice di varianze e covarianze robusta alla non corretta specificazione del modello².

Riparametrizzazione del modello e Metodo Delta

Per massimizzare numericamente la Verosimiglianza, è consigliabile riparametrizzare il modello in modo tale che non si debba utilizzare una procedura numerica

²Il *software* da noi utilizzato per stimare tale modello, S-PLUS, implementa molto agevolmente questa versione robusta della matrice di varianza.

vincolata che limiti il movimento del vettore di parametri $\psi = (\gamma, \phi, \sigma)$ nello spazio parametrico $[\mathfrak{R} \times (-1, 1) \times \mathfrak{R}^+]$. Consideriamo perciò la seguente riparametrizzazione $\tilde{\psi} = (p_1, p_2, p_3)$, dove $\gamma = p_1$ dal momento che γ è libero di muoversi in \mathfrak{R} , $\nu = \exp p_3$ e $\phi = \exp(p_2)/(1 + \exp(p_2))$.

Le precedenti espressioni sono da utilizzare per ottenere immediatamente le stime dei parametri nella scala originaria. Per ottenere invece una corretta stima della matrice di varianze e covarianze, si deve ricorrere ad un metodo approssimato, basato sull'espansione in serie di Taylor, noto come Metodo Delta. Abbiamo quindi la seguente distribuzione asintotica per $\psi = h(\tilde{\psi})$:

$$h(\tilde{\psi}) - h(\tilde{\psi}_0) \sim N_3(0, \nabla h(\tilde{\psi})' \hat{\Sigma} \nabla h(\tilde{\psi})),$$

dove $\hat{\Sigma}$ è la matrice di varianze e covarianze robusta all'errata specificazione e

$$\nabla h(\tilde{\psi}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{p_2} & 0 \\ 0 & 0 & \frac{e^{p_3}}{(1+e^{p_3})^2} \end{bmatrix}.$$

Non normalità di ξ_t

Abbiamo visto che la stima di Quasi Massima Verosimiglianza si basa sull'assunzione che il disturbo nell'equazione delle osservazioni sia una Normale di media -1.27 e varianza $\pi^2/2$ e in virtù di questa Normalità approssimata sia ottenibile una stima della Verosimiglianza basata sulla decomposizione dell'errore di previsione ricavata agevolmente con il filtro di Kalman. Da semplici risultati del Calcolo delle Probabilità è possibile ricavare la distribuzione esatta di ξ_t , la cui densità è pari a:

$$f_{\xi_t}(z) = \frac{1}{\sqrt{2\pi}} \exp \left[\frac{z - e^z}{2} \right]. \quad (2.12)$$

Nella figura 2.1 possiamo vedere una rappresentazione grafica della differenza tra la vera distribuzione e l'approssimazione Normale.

Risultati della stima QML sui dati simulati

Nella tabella 2.1 riportiamo la stima dei parametri, degli *standard errors* e degli intervalli di confidenza di livello α pari a 0.95. La matrice di varianze e covarianze è robusta alla non correttezza del modello specificato, inoltre, per avere parametri nella scala originaria è stato utilizzato il Metodo Delta.

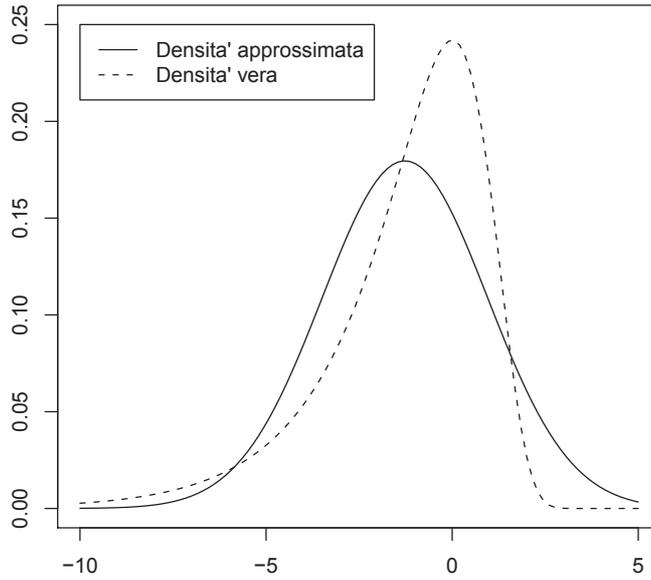


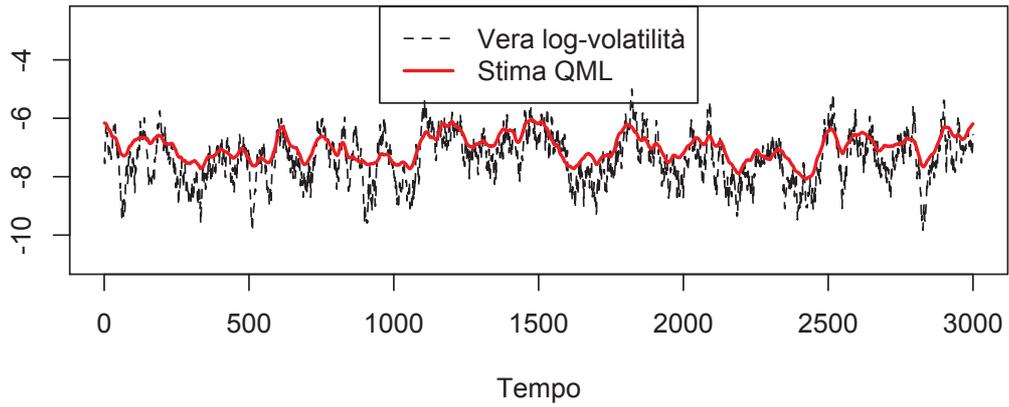
Figura 2.1:
Densità di ξ_t e dell'approssimazione $N(-1.27, \pi^2/2)$

I risultati dell'inferenza sono piuttosto deludenti: la stima puntuale dei parametri γ e ν è distante dai valori veri, nessun intervallo di confidenza ad un livello di significatività $\alpha = 0.95$ contiene il vero valore del parametro. Tali risultati ci fanno supporre che l'errata specificazione del modello abbia avuto conseguenze devastanti. Tuttavia, dalla figura 2.2, la volatilità (o log-volatilità) stimata, se confrontata con quella vera (disponibile perché stiamo usando dati simulati) e con una *proxy* grezza (i rendimenti assoluti), non mostra un andamento assurdo, mostrando che anche con una inferenza di qualità scarsa è possibile ottenere una buona stima *smoothed*.

	Vero θ	$\hat{\theta}_{QML}$	S.E.	Limite Inferiore	Limite Superiore
γ	-0.368	- 0.16267	0.03898	-0.23985	-0.08549
ν	0.26	0.11862	0.01463	0.08995	0.14730
ϕ	0.95	0.97739	0.00544	0.96673	0.96673
Log-Verosimiglianza					-6143.29

Tabella 2.1: Risultati della stima QML nel campione simulato

Vera log-volatilità



Rendimenti assoluti

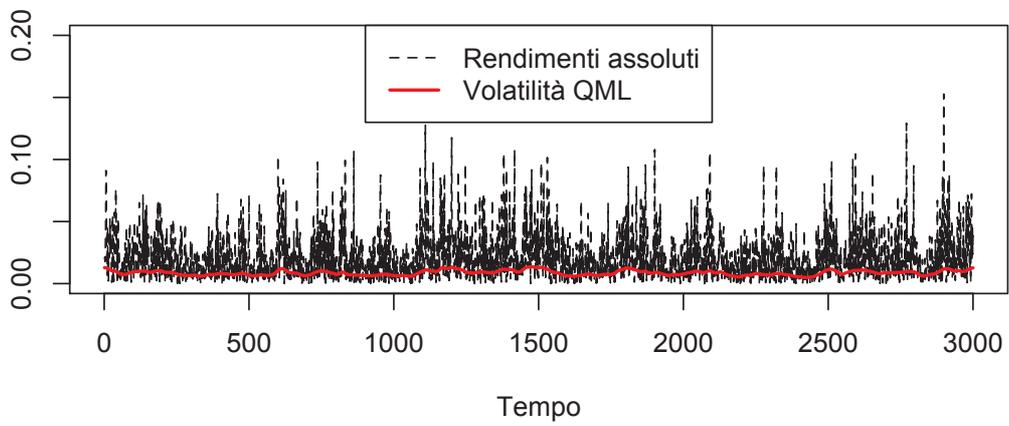


Figura 2.2:
Volatilità *smoothed* ottenuta tramite procedura QML per il campione simulato

Capitolo 3

Modelli *state space* parzialmente non Gaussiani e metodi MCMC

3.1 Modelli *state space* parzialmente non Gaussiani

Nel capitolo precedente abbiamo presentato il modello *state space* Gaussiano e lineare, mostrando come la sua applicazione ad un contesto non Gaussiano portasse ad uno stimatore con scarse capacità inferenziali nonostante l'utilizzo di una correzione derivata da un approccio di Quasi Massima Verosimiglianza della matrice di varianze e covarianze asintotica di tale stimatore.

Tuttavia, benché l'inadeguatezza delle ipotesi di normalità e/o di linearità porti alla perdita delle proprietà di ottimalità godute dal filtro di Kalman, il fatto che alcune di queste qualità continuino a valere in senso *lineare*, hanno portato (si veda Shephard (1994)) a suggerire una modifica degli algoritmi di filtraggio e di *smoothing* che produce risultati più accurati nonché una stima dei parametri consistente. L'idea base di questo approccio è considerare una specificazione del modello che sia, condizionatamente ad una variabile *ausiliaria* ω , ancora Normale; per questa ragione tali modelli sono chiamati *state space parzialmente non Gaussiani*. Presentiamo ora tali modelli, fornendo solo, oltre alla loro formulazione generica, l'idea alla base del *Gibbs sampler*¹ proposto da Kim et al. (1998) per stimare il modello a Volatilità Stocastica.

¹Introdurremo il *Gibbs sampler* nella seconda sezione di questo capitolo.

3.1.1 Il modello

Definiamo nel modo seguente il generico modello *state space* parzialmente non Gaussiano:

$$\begin{aligned} y_t &= Z_t(\omega_t)\alpha_t + d_t(\omega_t) + \epsilon_t(\omega_t), & \epsilon_t(\omega_t) &\sim N(0, H_t(\omega_t)) \\ \alpha_t &= T_t(\omega_t)\alpha_{t-1} + c_t(\omega_t) + \eta_t(\omega_t), & \eta_t(\omega_t) &\sim N(0, Q_t(\omega_t)), \end{aligned} \quad (3.1)$$

dove $\omega = (\omega'_1, \dots, \omega'_n)'$ è un processo di Markov, ovvero

$$f(\omega|\omega_0) = \prod_{t=1}^n f(\omega_t|\omega_{t-1}). \quad (3.2)$$

Come nel modello lineare e Gaussiano, i disturbi $\epsilon_t(\omega_t)$, $\eta_s(\omega_s)$ continuano ad essere assunti indipendenti per tutti i valori di t e s .

Questa formulazione permette di introdurre la non Normalità e la non linearità nel modello, lasciando comunque intatta una certa semplicità di analisi. Dovendo infatti estrarre simulazioni dalla densità $f(\delta|Y_n)$, dove Y_n indica che ci condizioniamo all'intero campione di osservazioni e $\delta = (\alpha', \omega')'$, possiamo utilizzare un semplice algoritmo *Gibbs sampler* che consiste nei seguenti passi:

1. Inizializzare $\delta = (\alpha', \omega')'$.
2. Estrarre α^* dalla densità $f(\alpha|\omega, Y_n)$.
3. Estrarre ω^* dalla densità $f(\omega|\alpha^*, Y_n)$.
4. Salvare $\delta = (\alpha^{*'}, \omega^{*'})'$.
5. Tornare al passo 2 (iterando fino alla convergenza i passi 2-4).

Considerando la seguente decomposizione per $f(\alpha, \omega, Y_n)$:

$$\begin{aligned} f(\alpha|\omega, Y_n) &= f(\alpha_n|\omega, Y_n) \prod_{t=1}^{n-1} f(\alpha_t|\alpha_{t+1}, \alpha_{t+2}, \dots, \alpha_n, \omega, Y_t) \\ &= f(\alpha_n|\omega, Y_n) \prod_{t=1}^{n-1} f(\alpha_t|\alpha_{t+1}, \omega, Y_t) \end{aligned}$$

poiché le $f(\alpha_t|\alpha_{t+1}, \omega, Y_t)$ sono densità condizionali Normali, risulta semplice ottenere estrazioni che possano simulare efficientemente $f(\alpha|\omega, Y_n)$, superando, in questo modo, la non Normalità del modello.

3.2 Teoria alla base dei metodi MCMC

3.2.1 Idea base

Si supponga di volere ottenere un'approssimazione del seguente integrale

$$I = \int h(x)f(x)dx, \quad (3.3)$$

ma di non poter simulare direttamente un campione dalla distribuzione f . Gli algoritmi basati sulle *Monte Carlo Markov Chain* (abbreviato in MCMC) permettono di ottenere un campione X_1, \dots, X_n distribuito approssimativamente secondo f senza simulare direttamente da f . Per ottenere ciò, viene utilizzato un particolare processo stocastico, chiamato *catena di Markov*, che deve godere della particolare proprietà di *ergodicità* e avere come *distribuzione stazionaria* proprio f .

L'utilizzo di una catena di Markov con queste particolari caratteristiche, grazie al Teorema Ergodico, è fondamentalmente identico all'utilizzo di un campione IID da f , poiché tale teorema garantisce la convergenza quasi certa della media campionaria

$$I_T = \frac{1}{T} \sum_{t=1}^T h(X^{(t)}) \quad (3.4)$$

alla quantità $E_f[h(X)]$. Vedremo che le catene di Markov, per costruzione, sono caratterizzate da una forte dipendenza ma questo non genera particolari problemi se lo scopo dello studio di simulazione è esaminare le proprietà della distribuzione f . Non è, perciò, necessario generare n indipendenti catene $\{X_i^{(t)}\}(i = 1, \dots, n)$, prendendo solo il valore terminale $X_i^{(T)}$ di ciascuna. Al contrario, una singola traiettoria (o *path*) della catena di Markov è sufficiente per assicurare un'approssimazione appropriata di I attraverso una stima della funzione h di interesse.

Mostriamo gli elementi teorici fondamentali delle catene di Markov, basandoci essenzialmente sull'ottima trattazione offerta da Robert and Casella (2004).

3.2.2 Principi essenziali

Una catena di Markov è un processo stocastico costruito a partire da un *kernel di transizione*, K , che è una densità di probabilità condizionale tale che $X_{n+1} = K(X_n, X_{n+1})$. Queste catene godono di una proprietà di stabilità molto forte, l'esistenza di una *distribuzione di probabilità stazionaria*, π , tale che se $X_n \sim \pi$, allora $X_{n+1} \sim \pi$. Se il *kernel* K è libero di muoversi liberamente in tutto lo spazio degli stati, la catena sarà chiamata *irriducibile*.

Tale libertà assicura che la maggior parte delle catene di Markov coinvolte negli algoritmi MCMC sia *ricorrente*, ovvero che, *per quasi tutti* i valori iniziali, le catene di Markov abbiano lo stesso comportamento limite. Tale caratteristica è molto utile dal momento che i principali algoritmi prendono avvio da un punto iniziale arbitrario, x_0 .

La distribuzione stazionaria, π , è anche la distribuzione limite di X_{n+1} , indipendentemente dal valore iniziale di X_0 . Una conseguenza molto interessante di questa proprietà di convergenza (già anticipata nel precedente paragrafo) è che la media $\frac{1}{N} \sum_{n=1}^N h(X_n)$ converge quasi certamente al valore atteso $E_\pi[h(X)]$. Quando la catena è *reversibile* (che avviene quando il *kernel di transizione* è simmetrico), per questa media vale anche un Teorema del Limite Centrale.

Catena di Markov

Una catena di Markov è una sequenza di variabili casuali che possono essere pensate evolversi nel tempo, con una probabilità di transizione che dipende unicamente dal particolare insieme in cui si trova la catena e non da come si sia giunti in tale insieme (tale caratteristica sarà formalizzata nella proprietà di Markov, la quale può essere esplicitata sia in una forma debole sia in una forma forte). Perciò, è naturale definire le catene di Markov sulla base del loro *kernel di transizione*, il quale non è altro che la funzione che determina queste transizioni.

Definizione 1 *Un kernel di transizione è una funzione K definita su $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ tale che*

1. $\forall x \in \mathcal{X}, K(x, \cdot)$ è una misura di probabilità;
2. $\forall A \in \mathcal{B}(\mathcal{X}), K(\cdot, A)$ è misurabile.

Nel caso continuo, il *kernel* denota anche la densità condizionale $K(x, x')$ della transizione $K(x, \cdot)$; uguale a $P(X \in A|x) = \int_A K(x, x')dx'$.

La catena $\{X_n\}$ è usualmente definita per $n \in \mathbf{N}$ perciò la distribuzione di X_0 , lo stato iniziale della catena, gioca un ruolo importante. per questo motivo, denotiamo con P_μ la distribuzione di probabilità di $\{X_n\}$, se μ rappresenta la distribuzione dello stato iniziale della catena, mentre utilizziamo la notazione alternativa P_{x_0} , se X_0 è fissato.

Definizione 2 Dato un kernel di transizione K , una sequenza $X_0, X_1, \dots, X_n, \dots$ di variabili casuali è una catena di Markov, indicata da $\{X_n\}$, se, per ogni t , la distribuzione condizionale di X_t dati $x_{t-1}, x_{t-2}, \dots, x_0$ è uguale alla distribuzione di X_t dato x_{t-1} , ovvero se

$$P(X_{k+1} \in A | x_0, x_1, \dots, x_k) = P(X_{k+1} \in A | x_k) = \int_A K(x_k, dx).$$

La catena è *omogenea* se la distribuzione di $(X_{t_1}, \dots, X_{t_k})$ dato x_{t_0} è uguale alla distribuzione di $(X_{t_1-t_0}, \dots, X_{t_k-t_0})$ dato x_0 per ogni k - e $(k+1)$ -pletta $t_0 \leq t_1 \leq \dots \leq t_k$. In parole, la probabilità di transizione dipende unicamente dalla distanza degli istanti temporali, presi in considerazione, e non dall'istante temporale nel quale la catena si trova.

Dalla stessa definizione di catena di Markov, ricaviamo che la sua costruzione, se è nota la distribuzione (o lo stato) iniziale, è interamente determinata dalla *transizione*, ovvero dalla distribuzione di X_n condizionatamente a x_{n-1} .

Se denotiamo $K^1(x, A) = K(x, A)$, il kernel per le n transizioni è dato da ($n > 1$)

$$K^n(x, A) = \int_X K^{n-1}(y, A) K(x, dy)$$

Se $E_\mu(\cdot)$ denota il valore atteso associato alla distribuzione P_μ , possiamo scrivere la proprietà di Markov (debole) nel modo seguente:

Proposizione 1 (proprietà di Markov debole) Per ogni distribuzione iniziale μ e ogni campione (X_0, \dots, X_n) di dimensione $(n+1)$:

$$E_\mu[h(X_{n+1}, X_{n+2}, \dots | x_0, \dots, x_n)] = E_{x_n}[h(X_1, X_2, \dots)]. \quad (3.5)$$

E ciò mostra che i valori attesi esistono.

Tale proprietà formalizza l'*assenza di memoria* dei processi Markoviani: solo lo stato presente (e non la strada per arrivare ad esso) è rilevante per il futuro.

Irriducibilità

La proprietà di *irriducibilità* rappresenta una prima misura della sensibilità della catena di Markov alle condizioni iniziali. Tale proprietà è cruciale per le catene coinvolte negli algoritmi MCMC, perché garantisce immediatamente la convergenza della catena, evitando la necessità di studiare l'operatore di transizione.

Nel caso discreto, la catena è *irriducibile* se tutti gli stati comunicano², ovvero se

$$P_x(\tau_y < \infty) > 0, \quad \forall x, y \in \mathcal{X}$$

dove τ_y indica la prima volta che lo stato y è visitato. L'estensione al caso continuo richiede l'utilizzo di strumenti più sofisticati della teoria degli insiemi e esula dallo scopo di questo lavoro.

Aperiodicità

Nel caso discreto, il *periodo* di uno stato $\omega \in \mathcal{X}$ è definito come

$$d(\omega) = MCD\{m \leq 1; K^m(\omega, \omega) > 0\}$$

dove *MCD* indica Massimo Comune Denominatore. Il valore del periodo è costante su tutti gli stati che comunicano con ω . Nel caso di una catena irriducibile su uno spazio finito \mathcal{X} , la matrice di transizione può essere scritta come una matrice a blocchi

$$P = \begin{pmatrix} 0 & D_1 & 0 & \dots & 0 \\ 0 & 0 & D_2 & & 0 \\ & & \dots & & \\ D_d & 0 & 0 & & 0 \end{pmatrix}$$

dove i blocchi D_i sono matrici stocastiche. Questa rappresentazione mostra chiaramente il passaggio forzato da un gruppo di stati all'altro, con ritorno al gruppo iniziale ogni d -esimo passo. Se la catena è irriducibile, c'è un solo valore per periodo. Una catena irriducibile è *aperiodica* se ha periodo 1. Rimandiamo a Robert and Casella (2004) per l'estensione al caso generale. Qui basti dire che, perché una catena sia aperiodica, il numero minimo di passaggi temporali affinché ci sia probabilità non nulla che la catena torni allo stato in cui si trova, deve essere uguale a uno per ogni stato possibile.

Ricorrenza

Da un punto di vista algoritmico, una catena di Markov deve godere di proprietà di stabilità sufficientemente buone per garantire un'accettabile approssimazione del modello simulato. Nonostante l'irriducibilità garantisca che ogni insieme A sia visitato dalla catena di Markov $\{X_n\}$, questa proprietà è troppo debole per assicurare

²Con parole più semplici, perché una catena sia irriducibile, partendo da un qualsiasi stato, la probabilità di raggiungere ogni altro stato deve essere positiva.

che la traiettoria di $\{X_n\}$ entri in A abbastanza frequentemente.

Considerando per esempio un problema di massimizzazione, la convergenza al massimo globale non può essere garantita senza un' esplorazione approfondita del codominio della funzione. La formalizzazione di questa stabilità delle catene di Markov porta a differenti nozioni di *ricorrenza*. In un contesto discreto, la *ricorrenza di uno stato* è sufficiente per garantire un ritorno sicuro. Questa nozione è quindi necessariamente soddisfatta per le catene irriducibili su uno spazio finito.

Definizione 3 *In uno spazio degli stati finito \mathcal{X} , uno stato $\omega \in \mathcal{X}$ è transitorio se il numero medio di visite in ω , $E_\omega[\eta_\omega]$ è finito, ricorrente se $E_\omega[\eta_\omega] = \infty$.*

Tuttavia è bene sottolineare che per le catene irriducibili, la proprietà di ricorrenza è goduta dalla catena e non da un suo stato particolare.

Stazionarietà

Un livello maggiore di stabilità per la catena $\{X_n\}$ è raggiunto se la distribuzione marginale di X_n è indipendente da n . Più formalmente, questo richiede l'esistenza di una distribuzione di probabilità π per la quale $X_{n+1} \sim \pi$ se $X_n \sim \pi$. I metodi MCMC sono basati sul fatto che questa condizione, che definisce un particolare tipo di ricorrenza positiva, sia rispettata.

Reversibilità

Oltre alla stazionarietà, un'altra garanzia di stabilità per le catene di Markov è data dalla proprietà di *reversibilità*, la quale garantisce che la direzione del tempo non influenza il dinamismo della catena.

Definizione 4 *Una catena di Markov $\{X_n\}$ è reversibile se la distribuzione di X_{n+1} , condizionatamente a $X_n = x$ è uguale a quella di X_{n+1} , condizionatamente a $X_n = x$*

Nella pratica, la reversibilità può essere collegata all'esistenza di una misura stazionaria π se vale una condizione più forte enunciata nella definizione seguente.

Definizione 5 *Una catena di Markov con kernel di transizione K soddisfa la condizione di equilibrio se esiste una funzione f che soddisfi*

$$K(y, x)f(y) = K(x, y)f(x) \quad (3.6)$$

per ogni (x, y) .

Benché questa condizione non sia necessaria affinché f sia una misura di stazionarietà associata al *kernel* di transizione K , fornisce una condizione sufficiente che spesso è facile da verificare e che può essere usata per la maggior parte degli algoritmi MCMC. Questa condizione esprime un equilibrio nel flusso della catena di Markov, ovvero asserisce che la probabilità di essere in x e di muoversi a y è la stessa di essere in y e muoversi in x . Quando f è una densità, implica anche che la catena è reversibile.

Teorema 1 *Si supponga che una catena di Markov con funzione di transizione K soddisfi la condizione di equilibrio con π una funzione di densità di probabilità. Allora:*

1. *La densità π è la densità invariante della catena.*
2. *La catena è reversibile.*

Il legame, enunciato nel precedente teorema, tra reversibilità di una catena e la densità stazionaria, π , di tale catena è cruciale negli algoritmi MCMC, perché offre uno strumento per simulare la traiettoria di una catena che converga a π , la densità di interesse.

3.2.3 Algoritmi MCMC

Dopo aver presentato le principali proprietà teoriche delle catene di Markov ci interroghiamo ora su come poterne costruire una con l'obiettivo, è utile ricordarlo ancora una volta, di ottenere delle estrazioni causali dalla densità obiettivo (la distribuzione invariante della catena, indicata con π), per studiarne le caratteristiche di interesse.

Come abbiamo già brevemente accennato, il legame tra la proprietà di reversibilità e la densità limite della catena, fornisce un metodo pratico per ottenere catene che siano estrazioni casuali, anche se non indipendenti, da tale densità: l'algoritmo Metropolis-Hastings. Una sua versione particolare, il *Gibbs sampler* sarà poi presentata, perché permette di simulare agevolmente da densità multivariate, quali sono quelle di cui ci occupiamo in questo lavoro.

Algoritmo Metropolis-Hastings: idea base

Si consideri la generica funzione di densità $q(x, y)$ che porta la catena di Markov da un punto x a y . Se π è la distribuzione obiettivo, ci aspettiamo uno dei due seguenti risultati:

$$\pi(x)q(x, y) > \pi(y)q(y, x) \quad \text{o} \quad \pi(x)q(x, y) < \pi(y)q(y, x)$$

e quindi, una catena di Markov con generica densità di transizione $q(x, y)$ non ha π come distribuzione invariante. Tuttavia, per recuperare la condizione di reversibilità, necessaria per rendere π stazionaria, è possibile introdurre una quantità, $\alpha(x, y)$, che bilanci le espressioni precedenti, portando a

$$\pi(x) \underbrace{q(x, y)\alpha(x, y)}_{K(x, y)} = \pi(y) \underbrace{q(y, x)\alpha(y, x)}_{K(y, x)}.$$

Può essere provato (si veda Chib and Greenberg (1995)) che la quantità $\alpha(x, y)$ necessaria per ottenere la condizione di reversibilità è

$$\alpha(x, y) = \begin{cases} \min\left(\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right) & \text{se } \pi(x)q(x, y) > 0 \\ 1 & \text{altrimenti} \end{cases}$$

L'algoritmo di *Metropolis-Hastings* (MH) è quindi basato su una densità *proposal* $q(x, Y)$ che fa muovere la catena di Markov da x a y , e sulla *probabilità di accettazione del movimento*, $\alpha(x, y)$, necessaria per assicurare la proprietà di reversibilità. Sotto deboli condizioni può essere mostrato che la catena ottenuta dall'algoritmo MH converge a π .

Estensione al caso multivariato: il *Gibbs sampler*

Definiamo per un qualche $p > 1$, il vettore casuale $X = (X_1, \dots, X_p)$, dove le X_i sono variabili unidimensionali³. Inoltre, supponiamo di poter simulare dalle corrispondenti *full conditional densità* univariate f_1, \dots, f_p , ovvero, che si possa simulare

$$X_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p \sim f(x_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$$

per $i = 1, \dots, p$. L'idea alla base del *Gibbs sampler* è utilizzare queste densità condizionali come *proposal*, ottenendo di conseguenza una probabilità di accettazione, $\alpha(x, x_i)$, sempre uguale a uno.

Il *Gibbs sampler* è allora dato dal seguente schema di transizione da $X^{(t)}$ a $X^{(t+1)}$:

1. $X_1^{(t+1)} \sim f_1(x_1 | x_2^{(t)}, \dots, x_p^{(t)});$
2. $X_2^{(t+1)} \sim f_2(x_2 | x_1^{(t)}, x_3^{(t)}, \dots, x_p^{(t)});$
3. $X_p^{(t+1)} \sim f_p(x_p | x_1^{(t)}, \dots, x_{p-1}^{(t)}).$

³In realtà le X_i possono essere anche multidimensionali ma senza perdita di generalità, ci limitiamo al caso più semplice.

Le densità f_1, \dots, f_p sono chiamate *full conditional* e sono le sole densità necessarie per ottenere un campione simulato dalla densità multivariata. Infatti, la convergenza del *Gibbs sampler* alla distribuzione invariante multivariata π è assicurata dal noto teorema dovuto a Hammersley e Clifford. Per ragioni di semplicità espositiva qui ci limitiamo a presentarlo nella sua versione per variabili bidimensionali, accennando solo al fatto che l'estensione al caso più generale si basa sulla cosiddetta *condizione di positività*⁴.

Teorema 2 di Hammersley e Clifford per variabili bidimensionali

Prendiamo $p = 2$ e $X_1 = X$ e $X_2 = Y$, la distribuzione congiunta associata alle distribuzioni condizionali $f_{Y|X}(y|x)$ e $f_{X|Y}(x|y)$ ha densità congiunta data da:

$$f(x, y) = \frac{f_{Y|X}(y|x)}{\int [f_{Y|X}(y|x)/f_{X|Y}(x|y)] dy}.$$

Il precedente teorema permette di mostrare che anche MCMC composti convergono, come nel caso di un *Gibbs sampler* in cui alcune *full conditional* vengono simulate tramite MH.

Inferenza Bayesiana

Gli algoritmi MCMC trovano la loro più naturale applicazione nel contesto dell'inferenza Bayesiana. Ricordiamo infatti che in tale paradigma, l'interesse è volto ad ottenere la *densità a posteriori* del vettore di parametri, θ , assunti essere variabili casuali. Tale densità a posteriori è fattorizzata nella seguente espressione

$$\pi(\theta|Y) = \frac{p(\theta)L(Y|\theta)}{\int p(\theta)L(Y|\theta)d\theta}.$$

dove la *densità a priori*, $p(\theta)$, riassume le nostre conoscenze pregresse sul fenomeno in esame e la *verosimiglianza*, $L(Y|\theta)$, contiene l'informazione ricavabile dai dati osservati. L'integrale al denominatore dell'espressione precedente, $\int p(\theta)L(Y|\theta)d\theta$, rappresenta la costante di normalizzazione necessaria perché $\pi(\theta|Y)$ sia una funzione di densità. Tuttavia, tranne per casi molto semplici non è possibile ottenere analiticamente tale costante. E' chiaro allora come possono essere molto utili i metodi MCMC che permettono di ottenere estrazioni casuali da una densità, in questo caso proprio la densità a posteriori, conoscendo la densità a meno di una qualche costante di proporzionalità.

⁴Sia $(Y_1, Y_2, \dots, Y_p) \sim g(y_1, \dots, y_p)$, dove $g^{(i)}$ indica la distribuzione marginale di Y_i . Se $g^{(i)}(y_i) > 0$ per ogni $i = 1, \dots, p$, implica che $g(y_1, \dots, y_p) > 0$, allora g soddisfa la condizione di positività.

Una volta che abbiamo ottenuta un'approssimazione della distribuzione a posteriori, possiamo sintetizzarla attraverso i suoi momenti a posteriori, come ad esempio, la *media a posteriori* di una generica funzione h :

$$E_{\pi}[h(\theta)] = \int h(\theta)\pi(\theta|Y)d\theta.$$

Tale quantità può essere ricavata molto semplicemente attraverso le simulazioni Monte Carlo, ottenute dagli algoritmi MCMC una volta che questi hanno raggiunto la convergenza alla vera distribuzione a posteriori π . Nella pratica, quello che generalmente viene fatto è scartare un certo numero di estrazioni, il cosiddetto periodo di *burn-in* e poi valutare

$$\hat{E}_{\pi}[h(x)] = \frac{1}{T} \sum_{t=1}^T h(x_t).$$

Monitoraggio della convergenza e espressione per la varianza della distribuzione a posteriori

Uno dei principali problemi connessi ai metodi MCMC riguarda come decidere se la catena raggiunge la convergenza e, in caso di risposta affermativa, dopo quale numero si possono ritenere valide le estrazioni. Per rispondere a questi problemi, esistono in letteratura diversi test, tuttavia, ai fini del presente lavoro, consideriamo solo una semplice verifica ottenibile dal calcolo delle medie cumulate delle estrazioni MC e dalla verifica grafica che esse convergano ad un qualche punto stabile. I limiti di questo test ergodico empirico, definito da

$$\hat{E}[\theta] = \frac{1}{\tau} \sum_{t=1}^{\tau} \theta^t$$

per $\tau = 1, 2, \dots$ sono evidenti, basti solo dire che la convergenza della media campionaria non assicura la convergenza di momenti di ordine superiore e quindi, ovviamente, non assicura la convergenza alla distribuzione limite π . Tuttavia, l'interesse primario in questo lavoro è mettere a confronto le medie a posteriori, ottenute tramite i metodi MCMC con le stime di Massima Verosimiglianza (o Quasi Massima Verosimiglianza) ottenute con i metodi esposti negli altri capitoli, benché tale comparazione non sia del tutto corretta da un punto di vista teorico dal momento che le quantità in gioco, medie a posteriori e stime di MV, hanno una natura completamente diversa.

Per dare una misura della precisione di tali medie a posteriori, forniremo anche una stima delle deviazioni standard, ovvero le radici quadrate delle varianze campiona-

rie, le quali tuttavia devono essere calcolate tenendo conto della forte correlazione esistente tra estrazioni Monte Carlo consecutive, vale infatti la seguente relazione:

$$V(h(\Theta)) = \lim_{n \rightarrow \infty} n \text{Var}_\pi(\bar{h}_n) = \text{Var}_\pi[h(\Theta_0)] + 2 \sum_{j=1}^{\infty} \text{Cov}_\pi[h(\Theta_0), h(\Theta_j)],$$

che può essere stimata ricorrendo, tra gli altri, al noto stimatore di Newey-West.

3.3 Le densità mistura

3.3.1 Introduzione

Definizione

Se la distribuzione della variabile X può essere rappresentata da una funzione di densità della forma

$$p(x) = \pi_1 f_1(x) + \dots + \pi_k f_k(x), \quad (3.7)$$

dove i pesi soddisfano

$$\pi > 0, \quad j = 1, \dots, k; \quad \pi_1 + \dots + \pi_k = 1$$

e le densità soddisfano

$$f_j(\cdot) \geq 0, \quad \int f_j(x) dx = 1, \quad j = 1, \dots, k.$$

diciamo che X è una *mistura finita di distribuzioni* con $p(\cdot)$ mistura di funzioni di densità. I parametri π_1, \dots, π_k sono chiamati i pesi della mistura e $f_1(\cdot), \dots, f_k(\cdot)$ le densità delle componenti della mistura. E' immediato verificare che 3.7 definisce una funzione di densità di probabilità.

Poiché, in molte situazioni, le densità $f_1(\cdot), \dots, f_k(\cdot)$ contengono dei parametri specifici, possiamo riscrivere il lato destro della 3.7 con una notazione più precisa

$$\pi_1 f_1(x|\theta_1) + \dots + \pi_k f_k(x|\theta_k), \quad (3.8)$$

dove il vettore θ_j denota i parametri contenuti nella densità $f_j(\cdot)$. Indichiamo con ψ l'insieme completo di tutti i parametri del modello mistura, arrivando a scrivere, se tutte le densità delle componenti della mistura appartengono alla stessa famiglia

parametrica, la funzione di densità della mistura con la seguente forma

$$\pi(x|\psi) = \sum_{j=1}^k \pi_j f(x|\theta_j), \quad (3.9)$$

dove $f(\cdot|\theta)$ denota un membro generico della famiglia parametrica. Poiché in questo caso ogni $\theta_1, \dots, \theta_k$ è un elemento dello stesso spazio parametrico, Θ , gli $\pi = (\pi_1, \dots, \pi_k)$ possono essere pensati come una distribuzione di probabilità in Θ , con $\pi_j = Pr(\theta = \theta_j)$, $j = 1, \dots, k$.

Identificabilità

Sia

$$\mathbf{F} = \{F(x, \theta), \theta \in \Omega, \mathfrak{R}^d\}$$

la classe delle funzioni di distribuzione d -dimensionali dalla quale le misture sono formate. Identifichiamo la classe delle misture finite \mathbf{F} con la classe appropriata delle funzioni di distribuzione, \mathbf{H} , definita da

$$\mathbf{H} = \left\{ H(x) : H(x) = \sum_{j=1}^k \pi_j F(x, \theta_j), \pi_j > 0, \sum_{j=1}^k \pi_j = 1, \right. \\ \left. F(\cdot, \theta_j) \in \mathbf{F}, \forall j, k = 1, 2, \dots, x \in \mathfrak{R}^d \right\},$$

cosicché \mathbf{H} è l'involucro convesso di \mathbf{F} . Allora \mathbf{H} è identificabile se e solo se tutti gli elementi di \mathbf{H} sono distinti.

Definizione formale

Siano H, H' due elementi generici di \mathbf{H} , dati da

$$H = \sum_{j=1}^k \pi_j F_j \quad H' = \sum_{j=1}^{k'} \pi'_j F'_j,$$

e valga $H = H'$ se e solo se $k = k'$ e inoltre si possa ordinare la sommatoria in modo tale che $\pi_j = \pi'_j$, $F_j = F'_j$, $j = 1, \dots, k$. Allora \mathbf{H} è identificabile.

3.3.2 Misture di normali

Mistura di Normali come approssimazione di una generica densità

Dopo aver dato la definizione di una densità mistura, possiamo ricordare l'importante risultato mostrato da Ferguson (1983). Se $f(x)$ indica un'arbitraria densità

definita sull'asse reale, tale densità può essere modellata come una mistura di un numero contabile di distribuzioni Normali nella forma

$$f(x) = \sum_{i=1}^{\infty} \pi_i N(x|\mu_i, \sigma_i^2), \quad (3.10)$$

dove, ovviamente, $N(x|\mu, \sigma^2)$ denota una distribuzione Normale con media μ e varianza σ^2 . Tutterington et al. (1985), basandosi sulla funzione generatrice dei momenti, mostrano che una mistura di Normali univariate gode dell'importante proprietà di identificabilità. Importante perché tale proprietà è necessaria per poter stimare i parametri della mistura e per la validità della teoria asintotica usuale.

Inferenza sui parametri della mistura

Per l'inferenza delle misture, sono stati proposti vari metodi, a partire da quelli basati sui Momenti e sulla Verosimiglianza. Qui, ci limitiamo ad esporre brevemente uno stimatore, proposto da Quandt and Ramsey (1978) basato sulla minimizzazione della distanza tra due funzioni generatrici dei momenti (una empirica e una campionaria).

L'obiettivo è scegliere una stima dei parametri che porti la distribuzione mistura stimata ad essere quanto più vicina possibile ai dati, valutata tramite una misura della distanza tra una trasformazione delle funzioni di densità.

Si supponga, per una qualche variabile ausiliaria $t \in \mathbf{Z}$:

$$G(t|\psi) = E_g(t, X) = \int g(t, x) dF(x|\psi).$$

Valendo $E[(g(t, X))^2] < \infty \forall t \in \mathbf{Z}$, se X_1, \dots, X_n rappresenta un campione casuale da $F(\cdot|\psi)$ e se definiamo

$$\bar{g}_n(t) = n^{-1} \sum_{i=1}^n g(t, X_i),$$

abbiamo, grazie alla legge dei grandi numeri,

$$\bar{g}_n(t) \rightarrow G(t|\psi),$$

quando $n \rightarrow \infty$, per ogni $t \in \mathbf{Z}$.

Per stimare ψ possiamo quindi minimizzare una qualche misura della distanza tra

$\hat{g}_n(\cdot)$ e $G(\cdot|\psi)$, ovvero,

$$\delta[\bar{g}_n(t), G(t|\psi)].$$

Come misura della distanza, è preferibile prendere, per ragioni computazionali, misure di distanza quadratiche come ad esempio la distanza dei Minimi Quadrati Pesati definita come

$$\delta[\bar{g}_n(t), G(t|\psi)] = \int_L |G(t|\psi) - \bar{g}_n(t)|^2 dW(t),$$

dove $W(\cdot)$ è una misura positiva di peso su \mathbf{Z} .

Per dati univariati, Quandt and Ramsey (1978) suggeriscono di utilizzare la funzione generatrice dei momenti per $G(t|\psi)$ e per la sua versione empirica:

$$G_M(t|\psi) = E(e^{tX});$$

$$\hat{g}_n(t) = n^{-1} \sum_{r=1}^n \exp(tx_r).$$

Proprietà asintotiche

Sotto appropriate condizioni di regolarità, gli stimatori ottenuti sono, asintoticamente, non distorti e distribuiti secondo una Normale multivariata con matrice di covarianza calcolabile.

Nel metodo di Quandt and Ramsey (1978), dopo aver scelto un insieme di Q valori distinti t_1, \dots, t_q , si stima il $\hat{\psi}$ che minimizza

$$[G_M(\psi) - g_n]^T [G_M(\psi) - g_n], \quad (3.11)$$

dove $[G_M(\psi)]_i = G_M(t_i|\psi)$, $(g_n)_i = \hat{g}_n(t_i)$, $i = 1, \dots, Q$. Il problema definito nella 3.11 è chiaramente un problema di Minimi Quadrati non lineari, per il quale le equazioni di stazionarietà sono date da

$$A(\psi)^T [G_M(\psi) - g_n] = 0,$$

dove

$$[A(\psi)]_{ij} = \frac{\partial G_M(t_i|\psi)}{\partial \psi_j}$$

Asintoticamente, quando $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\psi} - \psi_0) \rightarrow N(0, V_1),$$

dove $V_1 = (A^T A)^{-1} A^T \Omega_0 A (A^T A)^{-1}$ e

$$\Omega_0 = \Omega(\psi_0) = Cov(g),$$

dove $(g)_i = \exp(t_i, X)$. Seguendo Schmidt (1982) si può ottenere uno stimatore più efficiente, $\tilde{\psi}$, minimizzando, al posto della 3.11:

$$[G_M(\psi) - g_n]^T \Omega(\psi_0) [G_M(\psi) - g_n],$$

per la quale vale la seguente distribuzione asintotica:

$$\sqrt{n}(\tilde{\psi} - \psi_0) \rightarrow N(0, (A^T \Omega_0^{-1} A)^{-1}).$$

Non si è fatto altro che sostituire un criterio dei Minimi Quadrati Ordinari con uno dei Minimi Quadrati Generalizzati anche se, quando la dimensione Q è pari al numero dei parametri da stimare, i due metodi sono equivalenti.

Un esempio estremamente utile di mistura di Normali

Presentiamo ora un esempio di mistura di Normali che sarà alla base di un approccio di stima del modello SV proposto più avanti. Ricordando quanto già visto nel capitolo precedente, se consideriamo la versione linearizzata del modello SV, l'inefficienza derivante dall'applicazione immediata del filtro di Kalman sorgeva essenzialmente dalla distanza tra la vera distribuzione degli errori di tale modello e l'assunta normalità della stessa.

Infatti la funzione di densità per gli errori $\log(\epsilon_t^2)$ è data, come abbiamo già visto, da:

$$f_{\log(\epsilon_t^2)}(z) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{z - e^z}{2}\right].$$

Sfruttando quanto abbiamo appena detto riguardo alle misture di normali, Kim et al. (1998) propongono di approssimare la vera distribuzione con una mistura di $K = 7$ Normali.

3.4 Applicazione al modello SV

Presentiamo ancora una volta, per agevolare la lettura di quello che segue, il modello:

$$\begin{aligned}r_t &= \beta e^{\theta_t/2} \epsilon_t, \\ \theta_{t+1} &= \phi \theta_t + \sigma_\eta \eta_t,\end{aligned}\tag{3.12}$$

L'idea sottostante ai metodi MCMC è quella di ottenere variabili da una densità multivariata (la densità a posteriori in un'ottica Bayesiana) simulando traiettorie da una catena di Markov la cui distribuzione invariante sia la densità obiettivo di interesse. I primi lavori su questi metodi appaiono in Metropolis et al. (1953), Hastings (1970), Ripley (1987).

Nel nostro problema, il punto principale è che la funzione di verosimiglianza $f(r|\theta) = \int f(r|\theta, \psi) f(\theta|\psi) d\theta$, dove ψ indica il vettore di parametri, è intrattabile e questo preclude l'analisi diretta della densità a posteriori $\pi(\psi|r)$ attraverso i metodi MCMC. Questo problema sarà superato concentrandoci invece sulla densità $\pi(\psi, \theta|r)$, dove $\theta = (\theta_1, \dots, \theta_n)$ è il vettore contenente le n volatilità. Le procedure MCMC permettono di campionare da questa densità senza calcolare la verosimiglianza $f(r|\psi)$. Queste estrazioni possono essere usate come base per l'inferenza richiamando il già citato Teorema Ergodico per le catene di Markov. Per esempio, i momenti a posteriori e le densità marginali possono essere stimate con la media delle funzioni di interesse nelle variabili simulate. L'accuratezza di queste stime crescerà con l'aumentare della dimensione del campione simulato.

3.4.1 Un semplice algoritmo Gibbs *sampler*

Come abbiamo già descritto in una precedente sezione di questo capitolo, per il problema di simulare da una densità multivariata $\pi(\psi|y)$, il Gibbs *sampler* è definito da uno schema a blocchi $\psi = (\psi_1, \dots, \psi_d)$ e dalle associate distribuzioni condizionali $\psi_i|\psi_{-i}$, dove ψ_{-i} denota ψ con l'esclusione del blocco i . In maniera analoga definiamo $\theta = (\theta_1, \dots, \theta_T)$ e θ_{-t} . L'algoritmo si basa sull'idea di campionare ogni blocco dalle distribuzioni condizionali dove i valori più recenti dei blocchi condizionanti sono usati nella simulazione.

Per il modello a Volatilità Stocastica, definiamo $\psi^* = (\theta, \psi)$. Per campionare ψ dalla densità a posteriori, una possibilità (suggerita da Jacquier et al. (1994) e Shephard (1993)) è di aggiornare un elemento del vettore ψ^* alla volta.

1. Inizializza θ e ψ .

2. Campiona θ_t da $p(\theta_t|\theta_{-t}, y, \psi)$, $t = 1, \dots, T$.
3. Campiona $\sigma_\eta^2|y, \theta, \phi, \mu, \beta$.
4. Campiona $\phi|\theta, \mu, \beta, \sigma_\eta^2$.
5. Campiona $\mu|\theta, \phi, \sigma_\eta^2$.
6. Torna al punto 2.

Campionamento di θ

La parte più complicata di questo *sampler* è rappresentata dalla necessità di campionare efficientemente da $p(\theta_t|\theta_{-t}, y_t, \psi)$ dal momento che questa operazione deve essere ripetuta molte volte. Inoltre la *full conditional*

$$f(\theta_t|\theta_{-t}, \psi, y) \propto f(\theta_t|\theta_{-t}, \psi)f(y_t|\theta_t, \psi), \quad t = 1, \dots, n,$$

non è nota. Una possibile soluzione è offerta dall'utilizzare una sequenza di algoritmi MH per estrarre da θ_t . Ad esempio, Kim et al. (1998) propongono di simulare da questa densità sviluppando una semplice procedura accettazione/rifiuto⁵. Se con $f_N(t|a, b)$ indichiamo la funzione di densità di una Normale con media a e varianza b , è possibile mostrare che

$$f(\theta_t|\theta_{-t}, \psi, y) = f(\theta_t|\theta_{t-1}, \theta_{t+1}, \psi) = f_N(\theta_t|\theta_t^*, v^2),$$

dove

$$\theta_t^* = \mu + \frac{\phi\{(\theta_{t-1} - \mu) + (\theta_{t+1} - \mu)\}}{1 + \phi^2}$$

e

$$v^2 = \frac{\sigma_\eta^2}{1 + \phi^2}.$$

Inoltre $\exp(-\theta_t)$ è una funzione convessa e può essere quindi delimitata da una funzione lineare in θ_t . Sia $\log f(y_t, \theta_t, \psi) = \text{const} + \log f^*(y_t, \theta_t, \psi)$. Poi

$$\log f^*(y_t, \theta_t, \psi) = -\frac{1}{2}\theta_t - \frac{y_t^2}{2}\{\exp(-\theta_t)\}$$

⁵L'idea chiave dei metodi di accettazione/rifiuto è usare una densità *strumentale* g dalla quale sia semplice simulare, per ottenere estrazioni dalla densità *obiettivo* f ricorrendo anche ad una variabile *ausiliaria* U , distribuita uniformemente. Per un'approfondita discussione del metodo rimandiamo al già citato Robert and Casella (2004).

$$\begin{aligned}
&\leq -\frac{1}{2}\theta_t - \frac{y_t^2}{2} \{ \exp(-\theta_t^*)(1 + \theta_t^*) - \theta_t \exp(-\theta_t^*) \} \\
&= \log g^*(y_t, \theta_t, \psi, \theta_t^*).
\end{aligned}$$

Quindi

$$f(\theta_t|\theta_{-t}, \psi) f^*(y_t, \theta_t, \psi) \leq f_N(\theta_t|\theta_t^*, v^2) g^*(y_t, \theta_t, \psi, \theta_t^*).$$

I termini a destra possono essere combinati e sono proporzionali a $f_N(\theta_t|\mu_t, v^2)$ dove

$$\mu_t = \theta_t^* + \frac{v^2}{2} [r_t^2 \exp(-h_t^*) - 1]. \quad (3.13)$$

Con questi risultati, la procedura di accettazione/rifiuto per campionare θ_t da $f(\theta_t|\theta_{-t}, \psi, y)$ può essere ora implementata nell'ambito di un MH. La *proposal* può basarsi su una Normale con media μ_t e varianza v^2 .

Campionamento di σ_η^2 , ϕ e μ

E' immediato campionare in una sola volta σ_η^2 e ϕ . Se noi assumiamo una priori coniugata $\sigma_\eta^2|\psi, \mu \sim IG(\sigma_r/2, S_\sigma/2)$, σ_η^2 è campionato da

$$\begin{aligned}
&\sigma_\eta^2|y, \theta, \phi, \mu \\
&\sim IG \left[\frac{n+\sigma_r}{2}, \frac{S_\sigma + (\theta_1 - \mu)^2(1 - \phi^2) + \sum_{t=1}^{n-1} ((\theta_{t+1} - \mu) - \phi(\theta_t - \mu))^2}{2} \right], \quad (3.14)
\end{aligned}$$

dove IG denota la distribuzione gamma inversa.

E' altrettanto semplice campionare dalla densità condizionale per ϕ . Sia $\phi = 2\phi^* - 1$ dove ϕ^* è distribuito come una Beta di parametri $(\phi^{(1)}, \phi^{(2)})$. Quindi la distribuzione a priori di ϕ è

$$\pi(\phi) \propto \left\{ \frac{1+\phi}{2} \right\}^{\phi^{(1)}-1} \left\{ \frac{1-\phi}{2} \right\}^{\phi^{(2)}-1}, \quad \phi^{(1)}, \phi^{(2)} > \frac{1}{2}, \quad (3.15)$$

e ha supporto nell'intervallo $(-1, 1)$ con media a priori pari a $\{2\phi^{(1)}/(\phi^{(1)} + \phi^{(2)}) - 1\}$. La densità condizionale di ϕ è quindi proporzionale a

$$\pi(\phi) f(\theta|\mu, \phi, \sigma_\eta^2),$$

dove

$$\log f(\theta|\mu, \phi, \sigma_\eta^2) \propto -\frac{(\theta_1 - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} + \frac{1}{2} \log(1 - \phi^2)$$

$$-\frac{\sum_{t=1}^{n-1} \{(\theta_{t+1} - \mu) - \phi(\theta_t - \mu)\}^2}{2\sigma_\eta^2}, \quad (3.16)$$

questa funzione è concava in ϕ per tutti i valori di $\phi^{(1)}, \phi^{(2)}$. Questo significa che ϕ può essere campionato usando un algoritmo di accettazione. Ricorriamo ad una espansione di Taylor al primo grado della priori attorno a

$$\hat{\phi} = \frac{\sum_{t=1}^{n-1} (\theta_{t+1} - \mu)(\theta_t - \mu)}{\sum_{t=1}^{n-1} (\theta_t - \mu)^2},$$

e combiniamola con $f(\theta|\mu, \phi, \sigma_\eta^2)$

Supponendo di lavorare con una priori diffusa per μ , μ è campionato dalla densità condizionale

$$\mu|\theta, \phi, \sigma_\eta^2 \sim N(\hat{\mu}, \sigma_\mu^2), \quad (3.17)$$

dove

$$\hat{\mu} = \sigma_\mu^2 \left\{ \frac{(1 - \phi^2)}{\sigma_\eta^2} \theta_1 + \frac{(1 - \phi)}{\sigma_\eta^2} \sum_{t=1}^{n-1} (\theta_{t+1} - \phi\theta_t) \right\},$$

e

$$\sigma_\mu^2 = \sigma_\eta^2 \{(n-1)(1-\phi)^2 + (1-\phi^2)\}^{-1}.$$

Tuttavia Kim et al. (1998) mostrano che questo Gibbs *sampler* è inefficiente a causa dell'elevata correlazione tra le componenti di $\theta|y$ e quelle di ψ . Il lento decadimento delle autocorrelazioni è provocato dal fatto che, campionando ogni componente dalle distribuzioni condizionali, si producono piccoli spostamenti nelle estrazioni.

3.4.2 Algoritmo applicato al modello linearizzato approssimato da una mistura di Normali

Presentiamo ancora una volta il modello SV linearizzato

$$\begin{aligned} y_t = \log(r_t^2) &= \theta_t + \log(\epsilon_t^2), \\ \theta_{t+1} &= \gamma + \phi\theta_t + \nu\eta_t. \end{aligned} \quad (3.18)$$

Kim et al. (1998), muovendosi nel contesto della stima MCMC, suggeriscono di considerare per $\log(\epsilon_t^2)$ una mistura di sette normali. Recuperando quanto detto nel paragrafo dedicato alle misture, possiamo scrivere per i rendimenti trasformati.

$$y_t^* = \theta_t + z_t, \quad (3.19)$$

dove y^* indica che abbiamo utilizzato la trasformazione di Fuller, precedentemente definita.

$$f(z_t) = \sum_{i=1}^K q_i f_N(z_t | m_i - 1.2704, v_i^2), \quad (3.20)$$

è una mistura di K densità normali, f_N , con probabilità di ogni componente pari a q_i , medie $m_i - 1.2704$ e varianze v_i^2 . Le costanti $\{q_i, m_i, v_i^2\}$ sono selezionate per approssimare quanto meglio possibile la vera densità di $\log(\epsilon_t^2)$. Possiamo scrivere una mistura di densità anche utilizzando una variabile indicatrice di componente, s_i , come

$$\begin{aligned} z_i | s_i = i &\sim N(m_i - 1.2704, v_i^2) \\ Pr(s_i = i) &= q_i \end{aligned} \quad (3.21)$$

Per selezionare K e $\{m_i, q_i, v_i^2\}$ ($i \leq K$) in modo tale che l'approssimazione sia sufficientemente buona Kim et al. (1998) eguagliano i primi quattro momenti teorici di $f_{\exp(z)}(r)$ (l'implicita distribuzione Log-Normale) e $f(z_t)$ ai momenti teorici rispettivamente di una variabile casuale χ_1^2 e $\log \chi_1^2$ in modo tale che tra le densità vere e quelle approssimate ci sia la minor distanza possibile. Attraverso una procedura di Minimi Quadrati non lineari, i pesi, le medie e le varianze vengono lasciati libere di muoversi. Questa ottimizzazione non lineare deve essere fatta solo una volta perché non coinvolge i parametri dipendenti dal modello e al termine di questa procedura gli autori citati ottengono che il numero ottimale di componenti è 7 con pesi, medie e varianze di ciascuna componente riportati nella Tabella 3.1. La Figura 3.1 mostra una rappresentazione grafica dell'accuratezza di tale scelta.

ω	$Pr(\omega = i)$	m_i	v_i^2
1	0.00730	-10.12999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	-1.08819	1.26261

Tabella 3.1:
Pesi, medie e varianze
per le componenti della mistura di $K = 7$ Normali

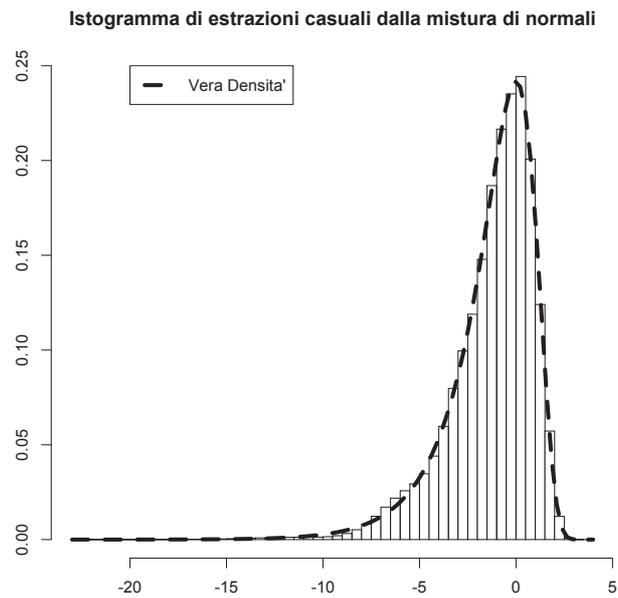


Figura 3.1:
Istogramma di 10 000 000 estrazioni casuali
dalla mistura di normali vs vera densità di ξ_t

Simulatore basato sulla mistura

Nel contesto MCMC, i modelli a mistura sono stimati più efficientemente utilizzando la rappresentazione in 3.21. La densità a posteriori di interesse è $\pi(s, \theta, \phi, \nu, |y^*)$, dove $s = (s_1, \dots, s_n)$. In questo caso, θ e s possono essere campionati separatamente in un unico blocco e il *sampler* segue la seguente procedura.

1. Inizializza s, ϕ, ν e μ .
2. Campiona θ da $\theta|y^*, s, \phi, \nu, \mu$.
3. Campiona s da $s|y^*, \theta$.
4. Aggiorna ϕ, ν, μ secondo (3.16), (3.14) e (3.17).
5. Torna al punto 2.

L'importante miglioramento rispetto al Gibbs *sampler* esposto precedentemente, è rappresentato dal fatto che ora è possibile campionare efficientemente da una distribuzione Gaussiana multivariata $\theta|y^*, s, \phi, \nu, \mu$ poiché $y^*|s, \phi, \nu, \mu$ è una serie storica Gaussiana che può essere rappresentata in una forma *state space* associata al filtro di Kalman. Questa particolare struttura del modello ci consente di campionare dall'intera densità $\theta|y^*, s, \phi, \nu, \mu$ usando il simulation smoother. Le estrazioni s da $s|y^*, \theta$ sono ottenute campionando indipendentemente ogni s_t usando la funzione di probabilità

$$Pr(s_t = i|y_i^*, \theta_t) \propto q_i f_N(y_t^*|\theta_t + m_i - 1.2704, v_i^2), \quad i \geq K.$$

Applicazione ai dati simulati

Provando ad applicare tale algoritmo ai nostri dati simulati, ricaviamo i risultati presentati nella Tabella 3.2 e nelle Figure 3.2 e 3.3. Per il *Gibbs sampler* si sono estratte 350000 osservazioni ed eliminate 50000, per il metodo basato sulla mistura 12000 di cui 2000 gettate. Confrontando i grafici la superiorità del secondo metodo è evidente. Infatti le autocorrelazioni decadono molto più velocemente per tutti e tre i parametri e il test empirico per verificare la convergenza delle medie campio-narie alle medie teoriche mostra chiaramente che nel secondo caso, con un numero decisamente inferiore di osservazioni, si ottengono buoni risultati. Il metodo basa-to sulla mistura di Normali si mostra migliore anche se andiamo a confrontare le stime numeriche. Infatti la media a posteriori è più vicina al vero valore del pa-rametro sia per γ , sia per ϕ mentre non lo è per ν . Anche le Deviazioni Standard sono buone perché, con un trentesimo delle replicazioni usate nel *Gibbs sampler*, sono più elevate ma non in maniera sconsiderata. Nella Figura 3.4 viene mostrata una stima filtrata della log-volatilità, ottenuta tramite un *particle filter*, proposta sempre da Kim et al. (1998), della quale tuttavia non forniamo i dettagli.

	Gibbs	$\hat{E}[\theta]$	Deviazione standard	Quantile 0.025	Quantile 0.975
γ	-0.368	-0.31793	0.00136	-0.46216	-0.20102
ν	0.26	0.22893	0.00045	0.18843	0.27715
ϕ	0.95	0.95744	0.00018	0.93830	0.97302

	Mistura	$\hat{E}[\theta]$	Deviazione standard	Quantile 0.025	Quantile 0.975
γ	-0.368	-0.35052	0.00379	-0.50736	-0.22034
ν	0.26	0.20910	0.00243	0.17239	0.25536
ϕ	0.95	0.95188	0.00052	0.93067	0.96964

Tabella 3.2: Risultati dei metodi MCMC

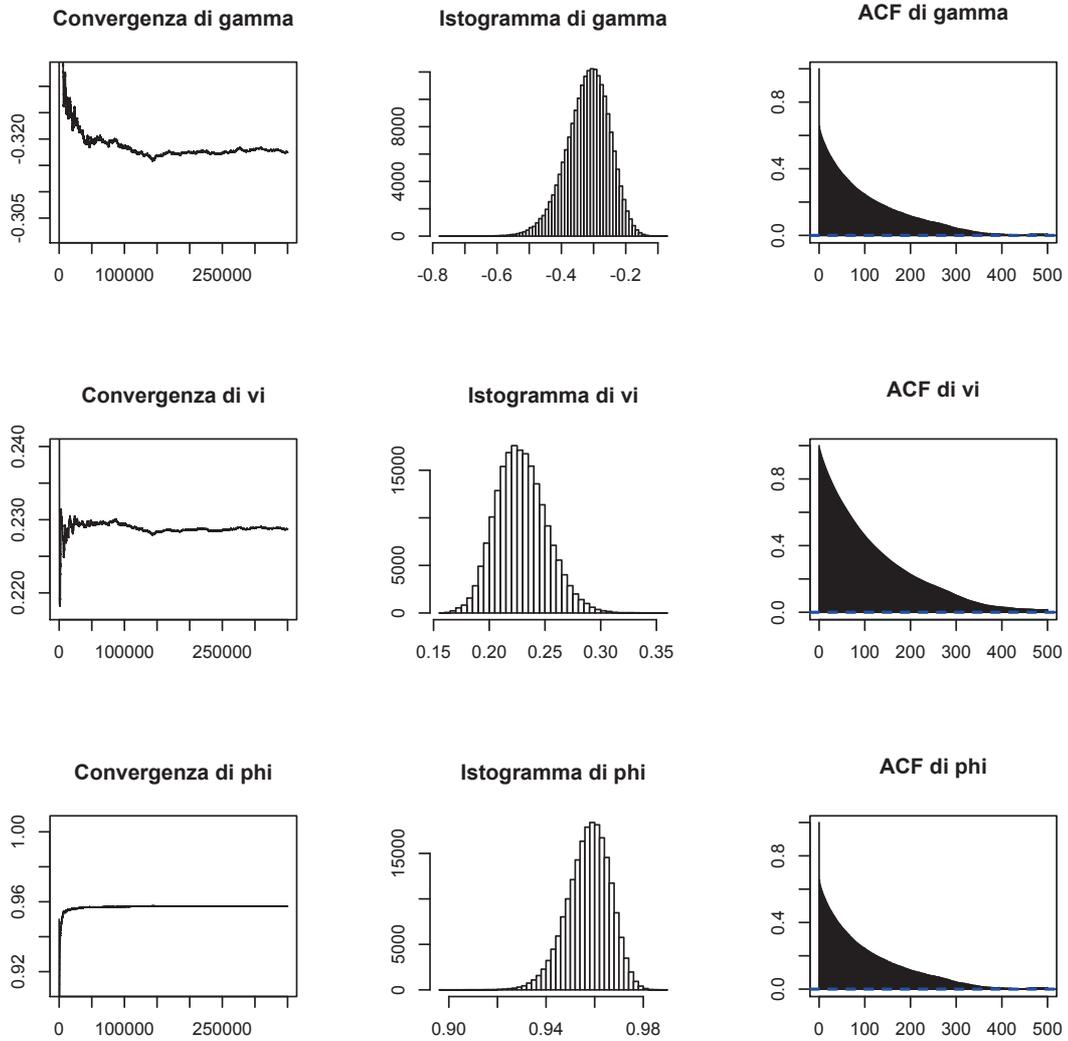


Figura 3.2: Analisi grafica dei risultati ottenuti con il *Gibbs sampler* per il campione simulato

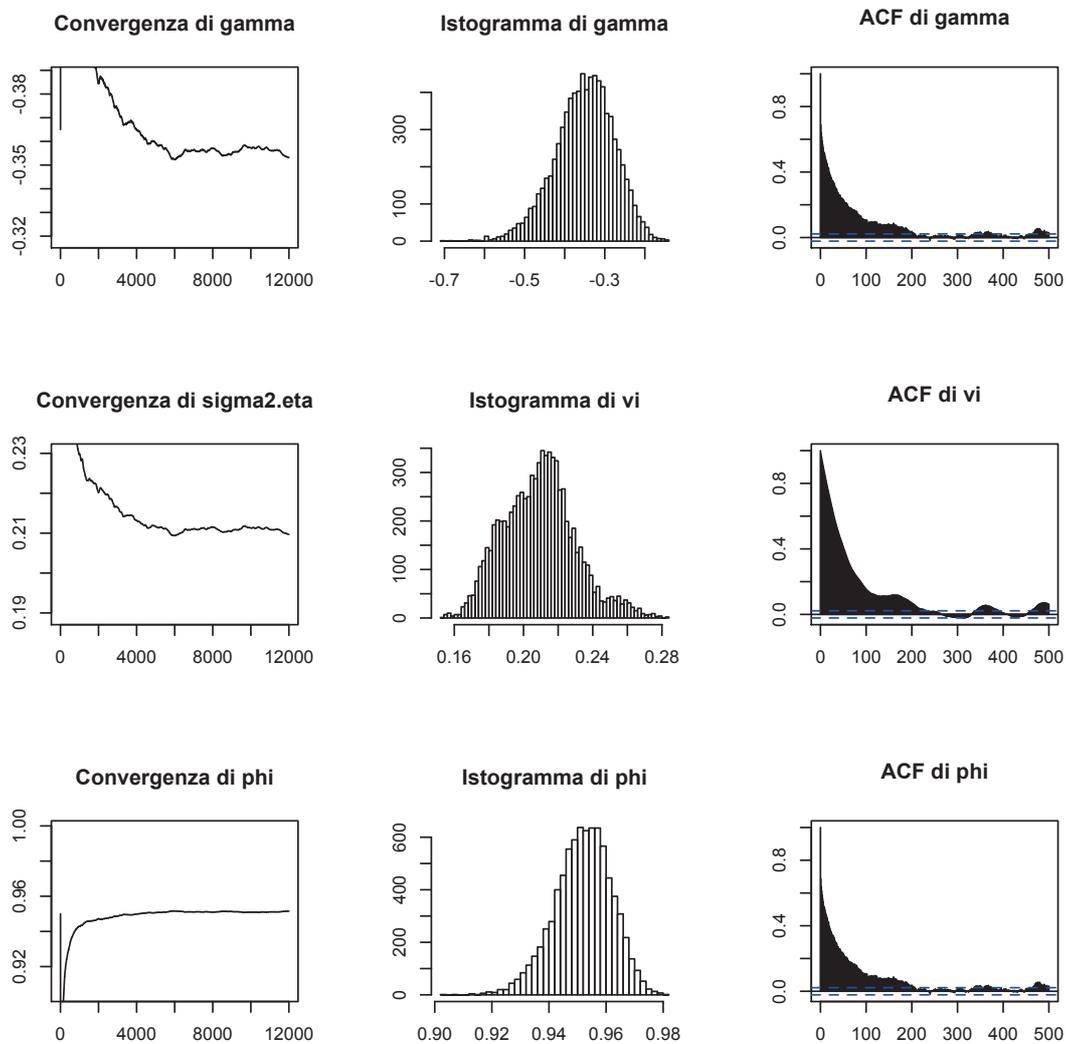
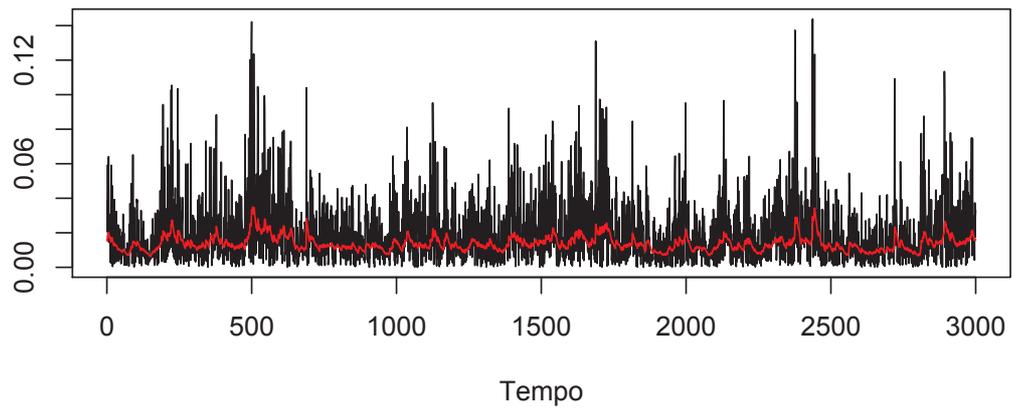


Figura 3.3: Analisi grafica dei risultati ottenuti con *sampler* basato sulla mistura di Normali

Valori assoluti vs. stima filtrata della volatilità'



Log-volatilità campionaria vs. sua stima filtrata

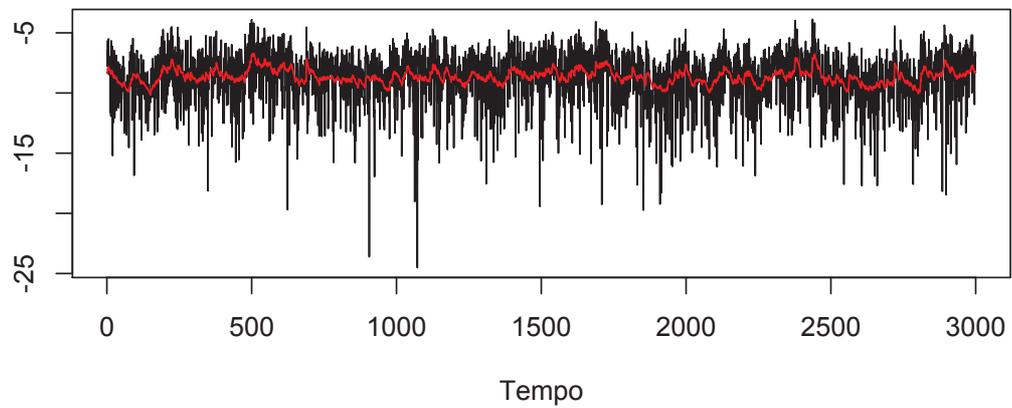


Figura 3.4: Stima filtrata del processo della volatilità
basata sul *particle filter*

Capitolo 4

Modelli *state space* generalizzati e *Importance Sampling*

4.1 Inferenza di modelli *state space* generalizzati

4.1.1 Introduzione

Rappresentiamo il modello *state space* generalizzato (perché rilassiamo le ipotesi di linearità e/o di normalità) con la seguente forma, simile a quella fornita per il modello lineare Gaussiano nella (2.1):

$$\begin{aligned} y_t | \theta_t &\sim p(y_t | \theta_t; \psi), & \theta_t &= Z_t \alpha_t, & t &= 1, \dots, n \\ \alpha_t &= d_t + T_t \alpha_{t-1} + \eta_t, & \alpha_1 &\sim N(a_1, P_1), & \eta_t &\sim N(0, Q_t), \end{aligned} \quad (4.1)$$

dove y_t è un vettore $p \times 1$ di osservazioni, θ_t è un vettore di *segnali*, (non direttamente osservabili) e α_t è un vettore $m \times 1$ di stati. Il vettore $m \times 1$ di costanti d_t , la matrice $m \times m$ di transizione T_t e la matrice di varianze Q_t determinano le proprietà dinamiche del modello. La densità per le osservazioni $p(y_t | \theta_t; \psi)$ può basarsi su una trasformazione non lineare del segnale θ_t (in questo caso si parla di modello *state space non lineare*) oppure può non essere Normale (e in questo caso si parla di modello *state space non Gaussiano*). Tuttavia l'equazione di transizione dello stato continua ad essere nella sua forma originale lineare Gaussiana. Il vettore ignoto di parametri ψ contiene i coefficienti non noti nella densità delle osservazioni come anche i coefficienti non noti dai quali dipendono le matrici del sistema.

Per stimare questi modelli, dal momento che l'usuale procedura abbinata al filtro di Kalman non può essere applicata direttamente, si deve stimare la log-verosimiglianza attraverso la simulazione Monte Carlo e poi massimizzarla nume-

ricamente rispetto a ψ .

4.1.2 Funzione di verosimiglianza

Definiamo $Y = \{y_t\}_{t=1}^T$ e $\Theta = \{\theta_t\}_{t=1}^T$ e consideriamo ψ come un ignoto. La verosimiglianza del modello, $L(y; \psi)$, è data dal seguente integrale, intrattabile analiticamente

$$\begin{aligned} L(y; \psi) = p(y; \psi) &= \int p(y, \Theta; \psi) d\Theta \\ &= \int \prod_{t=1}^T p(y_t, \theta_t | Y_{t-1}, \Theta_{t-1}; \psi) d\Theta. \end{aligned} \quad (4.2)$$

dove $p(y, \theta; \psi)$ è la densità congiunta di y e θ .

Condizionandoci a (θ_t, Y_{t-1}) , y_t è indipendente da Θ_{t-1} e quindi l'espressione all'interno della produttoria può essere ulteriormente fattorizzata:

$$p(y_t, \theta_t | Y_{t-1}, \Theta_{t-1}; \psi) = p(y_t | \theta_t, Y_{t-1}; \psi) p(\theta_t | Y_{t-1}, \Theta_{t-1}; \psi) \quad (4.3)$$

Teoricamente, $L(y; \psi)$ potrebbe essere stimata ricorrendo al cosiddetto *stimatore Monte Carlo naturale*

$$\hat{L}(Y, \psi) = \frac{1}{S} \sum_{s=1}^S \left[\prod_{t=1}^T p(y_t | \theta_t^{(s)}(\psi), Y_{t-1}; \psi) \right] \quad (4.4)$$

dove $\{\theta_t^{(s)}(\psi)\}_{t=1}^T$ denota una traiettoria nella quale ogni $\theta_t^{(s)}$ è estratta dalla densità condizionale $p(\theta_t | Y_{t-1}, \Theta_{t-1}; \psi)$. Poiché questa stima Monte Carlo è basata su una sequenza di densità simulate $p(\theta_t | Y_{t-1}, \Theta_{t-1}; \psi)$ ottenute direttamente dalla specificazione statistica del modello non sfruttando il fatto che le osservazioni Y contengono informazioni sul processo latente sottostante, risulta essere altamente inefficiente. Per aumentare l'efficienza si rivela necessario adottare tecniche standard come l'*importance sampling*, le variabili antitetiche e quelle di controllo.

4.1.3 L'Importance sampling

Per superare l'inefficienza dello stimatore Monte Carlo del paragrafo precedente, possiamo ricorrere alla tecnica nota come *importance sampling*.

L'intuizione alla base di questo metodo è di sostituire le simulazioni provenienti direttamente dalla densità della quale si vuole calcolare la verosimiglianza con simulazioni provenienti da appropriate densità ausiliarie, chiamate *importance den-*

sity che sfruttino efficacemente l'informazione sui θ_t inglobata negli y_t osservati, portando così ad una stima della verosimiglianza più accurata.

Più formalmente, se il nostro obiettivo è ottenere una stima della verosimiglianza, definita come l'integrale:

$$L(y; \psi) = \int p(y|\theta; \psi)p(\theta; \psi)d\theta, \quad (4.5)$$

intrattabile analiticamente, possiamo sfruttare il *simulation smoother* (descritto nel Capitolo 3) per ottenere valori campionati da una particolare densità Normale, $g(\theta|y; \psi)$. Riscriviamo la (4.5)

$$\begin{aligned} L(y; \psi) &= \int p(\Theta, y; \psi)d\Theta \\ &= \int \frac{p(\Theta, y; \psi)}{g(\Theta|y; \psi)}g(\Theta|y; \psi)d\Theta \\ &= g(y; \psi) \int \frac{p(\Theta, y; \psi)}{g(\Theta, y; \psi)}g(\Theta|y; \psi)d\Theta \\ &= L_g(y; \psi) \int \omega(\Theta, y; \psi)g(\Theta|y; \psi)d\Theta \end{aligned} \quad (4.6)$$

dove $L_g(y; \psi)$ è la funzione di verosimiglianza del modello Gaussiano *importance* e

$$\omega(\Theta, y; \psi) = \frac{p(y|\Theta; \psi)}{g(y|\Theta; \psi)}, \quad (4.7)$$

rappresentano i pesi della funzione *importance*; l'ultima uguaglianza nella (4.6) è valida perché abbiamo ipotizzato che la dinamica del processo latente anche in questi modelli *state space* generalizzati rimanga la stessa di quelli lineari e Gaussiani, ovvero $p(\Theta; \psi) = g(\Theta; \psi)$.

Generando S traiettorie indipendenti $\theta^{(1)}, \dots, \theta^{(N)}$ da $g(\Theta|y; \psi)$, la verosimiglianza può essere stimata con la media campionaria

$$\hat{L}(y; \psi) = \hat{L}_g(y; \psi)\bar{\omega}, \quad \bar{\omega} = \frac{1}{S} \sum_{s=1}^S \omega_s \quad \omega_s = \omega(\theta^{(s)}, y; \psi), \quad (4.8)$$

dove ω_s sono i pesi realizzati della funzione *importance* per $\theta^{(s)}$. Sotto deboli condizioni di regolarità, la Legge dei Grandi Numeri assicura che

$$\hat{L}(y; \psi) \xrightarrow{p} L(y; \psi). \quad (4.9)$$

Scelta della densità *importance*

L'efficienza del metodo dell'*Importance sampling* è strettamente connessa, come è facile immaginare, alla scelta appropriata di $g(y, \Theta; \psi)$. In questo capitolo e in quello successivo, presentiamo due differenti approcci a questa scelta; per ora ci limitiamo a esporre molto rapidamente i loro principi ispiratori:

1. Nell'approccio di Shephard, Pitt, Durbin e Koopman, $g(y_t, \theta_t; \psi)$ viene scelta in modo da essere un'approssimazione Gaussiana locale di $p(y_t, \theta_t; \psi)$, locale perché questa approssimazione si basa su uno sviluppo di Taylor della densità $p(y, \theta; \psi)$ intorno alla moda della distribuzione, $\hat{\theta}$.
2. Nell'approccio di Richard e Zang, la scelta della densità *importance* viene direttamente connessa alla varianza dei log-pesi $\log \omega, y; \psi$, con l'obiettivo di ricavare la $g(y, \Theta; \psi)$ che minimizzi la varianza della stima di $L(y; \psi)$.

4.2 Il metodo di Shephard, Pitt, Durbin e Koopman

4.2.1 Presentazione del metodo

Il metodo *importance sampling* di Shephard and Pitt (1997) e Durbin and Koopman (1997), nel seguito abbreviato in SPDK, ricorre ad una approssimazione lineare Gaussiana della log-densità $\log p(y|\theta)$, ricorrendo ad una sua espansione di Taylor del secondo ordine nell'intorno della moda di $p(\theta_t; \psi)$, detta $\hat{\theta}_t^1$. L'obiettivo è minimizzare

$$l(\theta_t) = \log p(y_t|\theta_t) - \log g(y_t|\theta_t), \quad (4.10)$$

dove l'eliminazione della dipendenza da y è giustificata dal fatto che d'ora in poi gli y_t sono considerati come costanti; per garantire una certa flessibilità al modello approssimato, $g(y_t|\theta_t)$, si considera una semplice estensione del modello lineare Gaussiano assumendo che la sua densità sia

$$\log g(y_t|\theta_t) = \text{costante} - \frac{1}{2}(y_t - \theta_t - \mu_t)' H_t^{-1} (y_t - \theta_t - \mu_t).$$

¹Il valore $\hat{\theta}_t$ è ottenuto applicando il filtro e lo *smoother* di Kalman nell'ipotesi che l'usuale modello lineare Gaussiano sia corretto.

Definendo poi $y_t^* = y_t - \mu_t$, possiamo riscrivere la densità precedente con la seguente notazione:

$$\log g(y_t^* | \theta_t) = \text{costante} - \frac{1}{2} (y_t^* - \theta_t)' H_t^{-1} (y_t^* - \theta_t),$$

dobbiamo quindi scegliere i valori per y_t^* e H_t che annullino le prime due derivate della 4.10 nella moda di $p(\theta_t; \psi)$, $\hat{\theta}_t$, ovvero $\partial l(\theta) / \partial \theta = 0$ e $\partial^2 l(\theta) / \partial \theta \partial \theta' = 0$ in $\theta = \hat{\theta}$. Otteniamo in questo modo il seguente sistema di equazioni vettoriali

$$\begin{cases} \frac{\partial \log p(y_t | \theta_t)}{\partial \theta_t} \Big|_{\theta_t = \hat{\theta}_t} - \tilde{H}_t^{-1} (\tilde{y}_t^* - \hat{\theta}_t) = 0, \\ \frac{\partial^2 \log p(y_t | \theta_t)}{\partial \theta_t \partial \theta_t'} \Big|_{\theta_t = \hat{\theta}_t} + \tilde{H}_t^{-1} = 0. \end{cases} \quad (4.11)$$

Queste equazioni si risolvono iterativamente partendo da $\tilde{y}_t^* = y_t$ e \tilde{H}_t arbitrario. Ricorrendo all'usuale *smoother* di Kalman, per ogni nuova coppia (y_t^*, H_t) , otteniamo un nuovo valore di $\hat{\theta}_t$; tale valore viene poi utilizzato nel sistema (4.11) per ottenere nuovi valori di \tilde{y}_t^* e \tilde{H}_t . La convergenza di tale procedura iterativa si mostra essere rapida, scegliendo infatti un'opportuna metrica per indicare i cambiamenti nei diversi passi dell'iterazione, gli autori dicono essere sufficienti tra le 5 e le 15 iterazioni.

Una volta che la procedura iterativa ha raggiunto la convergenza, abbiamo trovato la densità *importance* ottimale che può essere utilizzata, nei passi successivi, per ottenere la stima della verosimiglianza.

Stima Monte Carlo della log-verosimiglianza e correzione della distorsione

Se $\theta^{(s)}$ rappresenta un'estrazione casuale dalla densità *importance* $g(\theta | y)$, densità ricavata come indicato nel paragrafo precedente, mentre l'estrazione è ottenuta con il *simulation smoother* (dal momento che la densità approssimante è ancora Gaussiana, possono essere utilizzati gli strumenti presentati nel Capitolo 3), definiamo $\omega_i = \omega(\theta^{(s)})$, per $s = 1, \dots, S$, come

$$\omega(\theta) = \frac{p(y | \theta)}{g(y | \theta)}. \quad (4.12)$$

Possiamo quindi ricavare una stima basata solamente sull' *importance sampling* della verosimiglianza, $L(y; \psi)$, data da:

$$\hat{L}_{IS}(\psi) = L_g(\psi) \bar{\omega} \quad (4.13)$$

dove $\bar{\omega} = S^{-1} \sum_{s=1}^S \omega_s$. Benché $\hat{L}_{IS}(\psi)$ sia uno stimatore non distorto di $L(\psi)$, per ragioni computazionali (perché i numeri coinvolti sono più facilmente maneggiabili) è preferibile lavorare con la log-verosimiglianza, $\log L(\psi)$, piuttosto che direttamente con la verosimiglianza. Prendendo il logaritmo della (4.13) otteniamo

$$\log \hat{L}_{IS}(\psi) = \log L_g(\psi) + \log \bar{\omega}. \quad (4.14)$$

Purtroppo², $E[\log \bar{\omega}] \neq \log[E(\bar{\omega})]$, perciò il passaggio alla log-verosimiglianza ha introdotto una leggera distorsione, la quale, tuttavia, può essere agevolmente corretta (almeno asintoticamente) notando che

$$\begin{aligned} \log \bar{\omega} - \log \mu_\omega &= \log \left(1 + \frac{\bar{\omega} - \mu_\omega}{\mu_\omega} \right) \\ &= \frac{\bar{\omega} - \mu_\omega}{\mu_\omega} - \frac{1}{2} \left(\frac{\bar{\omega} - \mu_\omega}{\mu_\omega} \right)^2 + O_p(N^{-3/2}), \end{aligned} \quad (4.15)$$

dove $\mu_\omega = E_g(\omega_s)$. Possiamo perciò scrivere

$$E_g(\log \bar{\omega}) = \log \mu_\omega + \frac{\sigma_\omega^2}{2S\mu_\omega^2} + O(N^{-3/2}) \quad (4.16)$$

dove $\sigma_\omega^2 = Var(\omega_s)$.

Sostituendo i momenti teorici con i rispettivi momenti campionari, ovvero μ_ω con $\bar{\omega}$ e σ_ω^2 con s_ω^2 , dove

$$s_\omega^2 = \frac{1}{S-1} \sum_{s=1}^S (\omega_s - \bar{\omega})^2, \quad (4.17)$$

possiamo ottenere uno stimatore approssimativamente non distorto di $\log L(\psi)$, dato da

$$\log \hat{L}_{IS^*}(\psi) = \log L_g(\psi) = \log L_g(\psi) + \log \bar{\omega} + \frac{s_\omega^2}{2S\bar{\omega}^2}. \quad (4.18)$$

Inoltre è possibile mostrare che la varianza asintotica stimata di $\hat{L}_{IS^*}(\psi)$ è pari $s_\omega^2/(S\bar{\omega}^2)$.

Variabile antitetica

Per migliorare ulteriormente l'accuratezza della stima Monte Carlo basata sull'*importance sampling*, Durbin and Koopman (1997) propongono di utilizzare variabili

²In virtù della nota disuguaglianza di Jensen

antitetiche e variabili di controllo. Qui ci limitiamo a mostrare la più semplice variabile antitetica che sarà anche quella utilizzata nell'implementazione pratica dell'algoritmo e rimandiamo all'articolo precedentemente citato per un approfondimento ulteriore.

In Robert and Casella (2004), viene mostrato dettagliatamente come l'utilizzo delle variabili antitetiche possa migliorare l'efficienza della stima Monte Carlo di un integrale. Infatti, per dare un'idea generale di questa tecnica, se due campioni (X_1, \dots, X_m) e (Y_1, \dots, Y_m) provenienti dalla densità f , sono usati per la stima³ di

$$\mathcal{I} = \int_{\mathfrak{R}} h(x)f(x)dx,$$

lo stimatore

$$\frac{1}{2m} \sum_{i=1}^m [h(X_i) + h(Y_i)]$$

è più efficiente di uno stimatore basato su un campione IID di dimensione $2m$ a condizione che le variabili $h(X_i)$ e $h(Y_i)$ siano negativamente correlate e le variabili Y_i prendono il nome di *variabili antitetiche*.

Per ottenere in maniera semplice queste variabili antitetiche Geweke (1989) suggerisce di invertire le X_i e prendere $Y_i = 2\mu - X_i$, quando f è simmetrica attorno a μ .

La prima variabile antitetica proposta da Durbin and Koopman (1997) segue questo suggerimento, infatti è $\tilde{\theta}_t = 2\hat{\theta}_t - \theta_t$, per $t = 1, \dots, T$. Infatti θ_t è distribuita normalmente attorno a $\hat{\theta}_t$ e $\tilde{\theta}_t - \hat{\theta}_t = -(\theta_t - \hat{\theta}_t)$. E' bene sottolineare che l'efficienza guadagnata con questa variabile antitetica deriva anche dal fatto che basta una singola iterazione del *simulation smoother* per ottenere i due campioni $\theta^{(s)}$ e $\tilde{\theta}^{(s)}$. Indicando con $\tilde{\omega}_s$ il valore di ω_s ottenuto da $\tilde{\theta}^{(s)}$, possiamo ricavare

$$\omega_s^* = \frac{1}{2}(\omega_s + \tilde{\omega}_s), \quad (4.19)$$

per $s = 1, \dots, S$ e stimare $\log L(\psi)$ con $\log \hat{L}_{IS^{**}}(\psi)$ che è data dalla (4.18) con ω_s sostituito da ω_s^* .

Riepilogo dell'algoritmo SPDK

1. Identificare una densità Gaussiana $g(y^*|\theta)$, che approssimi $p(y|\theta)$, basata su $y_t^* \sim N(\theta_t H_t)$, dove y_t^* e H_t sono ottenuti come soluzione del sistema di

³La quale, evidentemente, non ricorre all'*importance sampling*.

equazioni 4.11.

2. Calcolare $\log L_g(\psi)$ usando l'usuale filtro di Kalman.
3. Estrarre campioni simulati $\theta^{(s)}$ da $g(\theta|y^*)$ usando il *simulation smoother*, per $s = 1, \dots, S$.
4. Per ogni s calcolare ω_s^* usando la 4.19 e calcolare $\log \hat{L}_{IS^{**}}$ usando la formula (4.18) con ω_s sostituito da ω_s^* .

4.2.2 Stima dei parametri e degli *standard errors*

Stima di Massima verosimiglianza dei parametri

Ora che abbiamo trovato una stima della verosimiglianza per un dato vettore dei parametri ψ , possiamo, attraverso una procedura di ottimizzazione numerica, ottenere la stima di Massima Verosimiglianza di tale vettore, questa stima sarà indicata con $\hat{\psi}$.

Per assicurare stabilità al processo di ottimizzazione numerica, è fortemente auspicabile ricorrere alla tecnica dei *Common Random Numbers*, abbreviato in CRN. Come il nome stesso suggerisce, tale tecnica prevede di utilizzare esattamente gli stessi numeri casuali ad ogni iterazione della procedura di ottimizzazione. Dal momento che l'implementazione pratica di tale tecnica è molto semplice, non ci dilunghiamo qui sulle sue giustificazioni teoriche (rimandando a Gourieroux et al. (1997)); ci limitiamo a notare che se diversi numeri casuali fossero usati per calcolare $\log \hat{L}(\psi)$ in valori vicini di ψ e si provasse a calcolare il gradiente, il risultato potrebbe essere estremamente instabile numericamente.

Matrice di varianza delle stime di Massima verosimiglianza

Assumendo che siano soddisfatte le appropriate condizioni di regolarità, la stima della matrice di varianza degli $\hat{\psi}$ è ricavata dalla formula standard:

$$\hat{\Omega} = \left[-\frac{\partial^2 \log L(\psi)}{\partial \psi \partial \psi'} \right] \Big|_{\psi=\hat{\psi}}, \quad (4.20)$$

dove le derivate della $\log L(\psi)$ possono essere calcolate numericamente tramite i valori di $\hat{\psi}$ nell'intorno di ψ .

4.3 Applicazione del metodo SPDK al modello a Volatilità Stocastica

4.3.1 Modello non Gaussiano

Ricordiamo la versione del modello SV canonico, stimata nel Capitolo 2 attraverso una procedura di Quasi Massima Verosimiglianza e nel Capitolo 3 attraverso un *Gibbs sampler* e una mistura di normali:

$$\begin{aligned}\log(r_t^2) &= \theta_t + \xi_t, \\ \theta_{t+1} &= \gamma + \phi\theta_t + \nu\eta_t.\end{aligned}\tag{4.21}$$

Sandmann and Koopman (1998) suggeriscono di sviluppare il metodo dell' *Importance Sampling* SPDK concentrandosi innanzitutto sulla scelta di un' approssimazione per la distribuzione degli ξ_t migliore di quella offerta da una Normale con media -1.27 e varianza $\pi^2/2$ per tutti i t .

Risulta più efficiente, dopo aver sottratto le medie, usare distribuzioni Normali con varianze che cambiano nel tempo.

Distribuzione approssimata dei disturbi

Per fare ciò scriviamo la densità approssimata $g(y|\theta)$ nella seguente semplice forma:

$$y_t = \mu_t + \theta_t + \xi_t$$

dove ξ_t è una Normale di media nulla e varianza H_t dove chiaramente le H_t non sono costanti nel tempo.

Attraverso l'usuale filtro di Kalman si può ottenere una stima lisciata del vettore degli errori, $\tilde{\xi}$ che insieme ad un valore iniziale arbitrario per gli \tilde{H}_t (per esempio pari a $\pi/2 \forall t$) può dare avvio alla procedura iterativa ispirata al sistema 4.11. Ad ogni passo si ottengono nuovi valori per \tilde{H}_t i quali vengono usati per estrarre attraverso il filtro e lo *smoother* di Kalman un nuovo vettore di disturbi $\tilde{\xi}_t$. Questa procedura viene ripetuta fino alla convergenza che Sandmann and Koopman (1998) e Durbin and Koopman (2001) mostrano essere raggiunta entro massimo dieci iterazioni.

Ad ogni passo dell' iterazione si dovrà risolvere la seguente equazione vettoriale:

$$\frac{1}{2}(1 - e^\xi) \Big|_{\xi=\tilde{\xi}} = - \frac{\xi}{\tilde{H}} \Big|_{\xi=\tilde{\xi}}$$

dalla quale si ottengono le seguenti T equazioni:

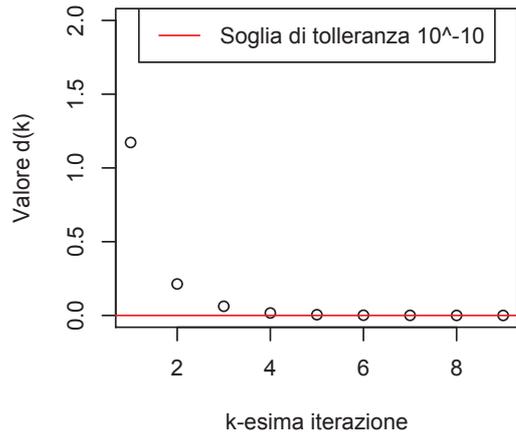
$$\tilde{H}_t = \frac{2\tilde{\xi}_t}{e^{\tilde{\xi}_t} - 1}, \quad (4.22)$$

per $t = 1, \dots, T$. Seguendo Sandmann and Koopman (1998), utilizziamo la metrica $d(k) = \frac{1}{N} \sum \left| \tilde{H}_t^{(k)} - \tilde{H}_t^{(k-1)} \right|$ per descrivere i cambiamenti successivi nelle varianze e verifichiamo che effettivamente bastano dieci iterazioni perché il valore $d(k)$ diventi molto piccolo: questo significa che le varianze \tilde{H}_t si sono stabilizzate e che quindi le Normali a media nulla e varianze \tilde{H}_t rappresentano una buona approssimazione della distribuzione del disturbo ξ_t . Nel primo grafico della Figura 4.1 viene mostrata la rapidità della convergenza di questa procedura. Il secondo grafico di tale Figura nonché le statistiche riassuntive presentate nella Tabella 4.1 evidenziano che le varianze \tilde{H}_t libere di muoversi liberamente in $[0, +\infty]$, non assumono valori esagerati e la loro media e la loro mediana sono abbastanza vicine al valore $\pi^2/2$ ipotizzato nel contesto QML. Inoltre dai due restanti grafici della Figura 4.1 possiamo notare che utilizzare varianze diverse per ogni istante temporale, anziché costanti, migliora leggermente anche la stima *smoothed* della log-volatilità senza essere ancora ricorsi alla procedura di *Importance Sampling*. Tali risultati intermedi possono fornire un'ulteriore prova della bontà di questa approssimazione.

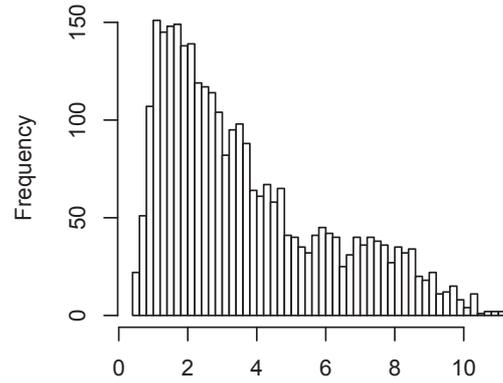
Minimo	Primo quantile	Mediana	Media	Terzo quantile	Massimo
0.4626	1.7300	2.8720	3.5870	4.8820	10.9500

Tabella 4.1: Statistiche riassuntive della varianze approssimate per il campione simulato

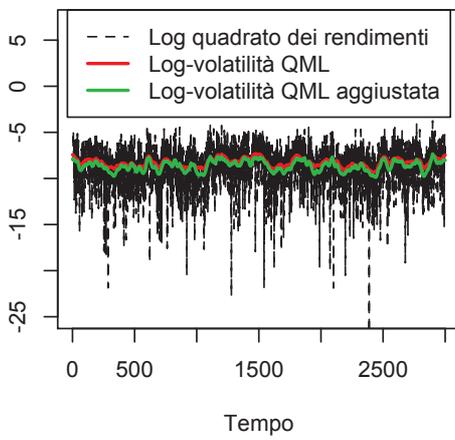
Verifica grafica della convergenza



Istogramma della varianze stimate



Log quadrato dei rendimenti



Rendimenti assoluti

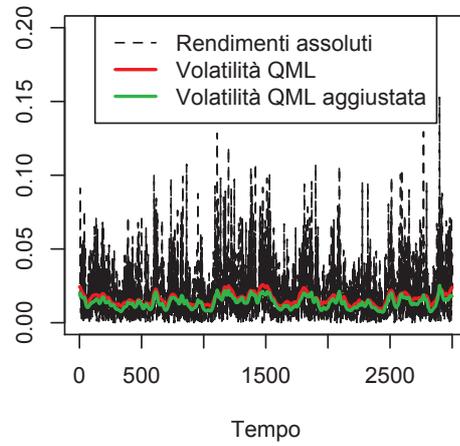


Figura 4.1:
Varianze approssimate e volatilità smoothed corretta per il campione simulato

Ulteriori passi per ottenere al stima della log-verosimiglianza

Avendo ottenuto tramite la procedura iterativa le stime delle varianze degli errori, ξ_t , possiamo calcolare la Log-Verosimiglianza del modello Normale approssimato, $L_G(y|\psi)$ e, attraverso il filtro di Kalman, ottenere le stime *smoothed* degli errori, ovvero $\hat{\xi}$.

La non Normalità del modello può essere a questo punto affrontata attraverso l'*Importance Sampling* SPDK, generando S campioni simulati, $\{\xi^{(i)}\}_{i=1}^S$, dalla densità $p_G(\xi|y, \psi)$, costruendo il campione di variabili antitetiche $\tilde{\xi}^{(i)} = 2\hat{\xi} - \xi^{(i)}$ e calcolando i pesi dell'*importance sampling* dati da

$$\omega^{(i)} = \frac{\omega(\xi^{(i)}) + \omega(\tilde{\xi}^{(i)})}{2}$$

dove $\omega(\xi) = \frac{p_{\log \chi_1^2(\xi|\psi)}}{p_G(\xi|\psi)}$. Otteniamo quindi la stima della Log-Verosimiglianza, attraverso

$$\log \hat{L}(\psi) = \log L_G(y|\psi) + \log \hat{\omega} + \frac{s_\omega^2}{2S\hat{\omega}^2},$$

dove $\hat{\omega}$ e s_ω^2 sono rispettivamente le medie e le varianze campionarie dei pesi $\omega^{(i)}$. Come già specificato, attraverso una procedura di ottimizzazione numerica massimizziamo tale quantità per ottenere la stima di Massima Verosimiglianza del vettore dei parametri, ψ .

Smoothing del processo della volatilità

Una volta che i parametri del modello sono stati stimati, l'interesse si concentra sull'ottenere le stime del processo latente della volatilità nel campione. A differenza dei metodi MCMC basati su un approccio Bayesiano, dove si ottiene una densità congiunta delle volatilità e dei parametri del modello, in un approccio frequentista i parametri devono essere fissi per ottenere una stima della volatilità attraverso lo *smoother* di Kalman.

Benché il filtro di Kalman non tenga conto esplicitamente della non Normalità nell'equazione delle osservazioni, quando è applicato al modello Normale approssimato, calcola le stime della moda a posteriori della volatilità (come mostrato da Shephard and Pitt (1997)). Quindi se la media a posteriori è richiesta, un efficiente algoritmo è quello suggerito da Durbin and Koopman (1998) dato da

$$\theta_{t|T} = y_t - \sum_{i=1}^S f_t \xi_t^{(i)}, \quad (4.23)$$

con $f_t = \frac{\omega(\xi_t^{(i)})}{S\hat{\omega}_t}$, dove $\theta_t|T$ è la stima della volatilità e $\xi_t^{(i)}$ è un'estrazione dall'*importance density* ottenuta grazie al *simulation smoother*. Infine Harvey and Shephard (1993) considerano la seguente stima del processo della volatilità:

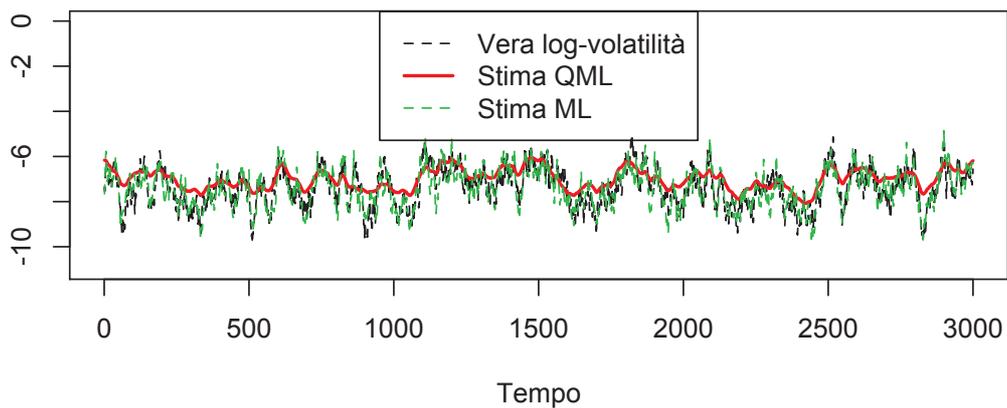
$$\tilde{\sigma}_t^2 = \hat{\sigma}_T^2 e^{\theta_t|T}$$

dove $\hat{\sigma}_t^2 = T^{-1} \sum_{t=1}^T r_t^2 e^{-\theta_t|T}$.

Applicazione del metodo ai dati simulati

Applicando il metodo appena esposto ai dati simulati, utilizzando $S = 50$ repliche e confrontando i risultati che si ottengono con quelli derivati dalla procedura di Quasi Massima Verosimiglianza si nota un indiscutibile miglioramento. Infatti le stime di Massima Verosimiglianza, come emerge chiaramente dalla Tabella 4.2, sono più precise (sebbene gli *standard errors* siano più grandi e gli intervalli di confidenza allo stesso livello 0.95 più larghi). Notiamo che la stima del parametro γ è più vicina al vero valore, come anche le stime di ϕ e di ν . Anche la stima della Log-Verosimiglianza ottenuta tramite questa procedura risulta in termini assoluti più bassa e ciò fornisce un'altra indicazione del fatto che sia preferibile tenere in considerazione la non Normalità delle osservazioni per stimare più efficacemente il modello. Diversamente da quanto fatto per la stima di Quasi Massima Verosimiglianza, adottiamo una procedura di ottimizzazione vincolata perché sembra funzionare meglio di una non vincolata. Come mostrato dalla Figura 4.2, anche la stima *smoothed* della volatilità è indubbiamente più precisa per il metodo SPDK rispetto al metodo QML.

Log-volatilità vera vs sue stime



Rendimenti assoluti

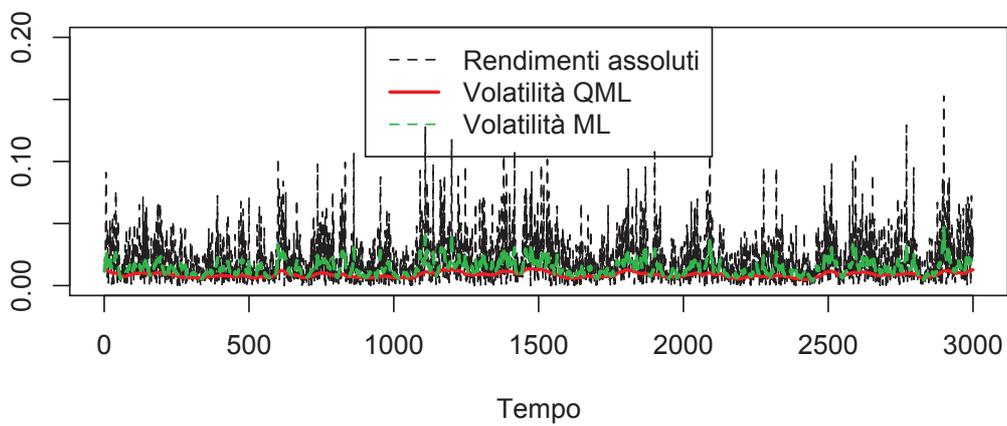


Figura 4.2:
Volatilità *smoothed* tramite procedura SPDK-ML per dati simulati

SPDK θ		$\hat{\theta}$	S.E.	Limite Inferiore	Limite Superiore
γ	-0.368	-0.45914	0.084747	-0.62524	-0.29303
ν	0.26	0.27621	0.02508	0.22703	0.32538
ϕ	0.95	0.93826	0.01131	0.91609	0.96043
Log-Verosimiglianza		-5970.366		S.E.	0.1935

QMV θ		$\hat{\theta}$	S.E.	Limite Inferiore	Limite Superiore
γ	-0.368	- 0.16267	0.03898	-0.23985	-0.08549
ν	0.26	0.11862	0.01463	0.08995	0.14730
ϕ	0.95	0.97739	0.00544	0.96673	0.96673
Log-Verosimiglianza		-6143.29			

Tabella 4.2: Risultati della stima SPDK-ML per dati simulati

Capitolo 5

L' *Efficient Importance Sampling*

A differenza del metodo SPDK, esposto nel precedente capitolo, dove la densità Gaussiana *importance* è scelta in maniera tale che sia un'approssimazione locale¹ della vera densità del modello in esame, l'alternativa suggerita da Richard and Zhang (2007), si focalizza direttamente sugli *importance weights*, ω . Tale metodo chiamato dagli autori *Efficient Importance Sampling*, d'ora in poi sarà abbreviato in EIS.

5.1 Presentazione dell'algoritmo

Principio generale

Se y_t , $t = 1, \dots, T$ indica un vettore n -dimensionale di variabili casuali osservate e θ_t un vettore q -dimensionale di variabili latenti, possiamo scrivere la densità congiunta di $Y = \{y_t\}_{t=1}^T$ e $\Theta = \{\theta_t\}_{t=1}^T$ come $f(Y, \Theta; \psi)$, dove ψ rappresenta il vettore dei parametri ignoti. La verosimiglianza associata alle variabili osservabili Y è data allora dall'integrale $(T \times q)$ -dimensionale

$$L(\psi; Y) = \int f(Y, \Theta; \psi) d\Theta.$$

Definendo $Y_t = \{y_\tau\}_{\tau=1}^t$ e $\Theta_t = \{\theta_\tau\}_{\tau=1}^t$, l'integrale dell'equazione precedente può essere fattorizzato in una sequenza di funzioni di densità condizionali al dato Y_{t-1}, Θ_{t-1} e, basandoci su questa fattorizzazione, possiamo riscrivere la verosimiglianza come segue:

$$L(\psi; Y) = \int \prod_{t=1}^T f(y_t, \theta_t | \Theta_{t-1}, R_{t-1}, \psi) d\Theta, \quad (5.1)$$

¹Come spiegato nel capitolo precedente, l'approssimazione è locale perché centrata nella moda del processo latente θ_t .

dove, per facilità di notazione, assumiamo note le condizioni iniziali $(\theta_0, \theta_{-1}, \dots)$ e (y_0, y_{-1}, \dots) .

Inoltre le assunzioni relative alla struttura dinamica della variabile latente possono essere tipicamente formulate nei termini della fattorizzazione aggiuntiva seguente:

$$f(y_t, \theta_t | \Theta_{t-1}, Y_{t-1}, \psi) = g(y_t | \theta_t, Y_{t-1}, \psi) p(\theta_t | \Theta_{t-1}, Y_{t-1}, \psi),$$

dove $g(\cdot)$ indica la densità condizionale di y_t dato (θ_t, Y_{t-1}) e $p(\cdot)$ indica la densità condizionale di θ_t dato (Θ_{t-1}, Y_{t-1}) ; questa fattorizzazione implica anche che y_t è indipendente da Θ_{t-1} , dato θ_t, Y_{t-1} , indipendenza rispettata nel modello a Volatilità Stocastica, oggetto di questo lavoro.

Possiamo allora ottenere una stima *naturale* Monte Carlo di $L(\psi; Y)$ basata su questa fattorizzazione, analoga a quella vista nel capitolo precedente; tale stima è data da

$$L_S(\psi; Y) = \frac{1}{S} \sum_{i=1}^S \left[\prod_{t=1}^T g(y_t | \tilde{\theta}_t^{(i)}, Y_{t-1}, \psi) \right],$$

dove $\{\tilde{\theta}_t^{(i)}\}_{t=1}^T$ indica una traiettoria della volatilità latente estratta dalla sequenza delle densità condizionali $p(\theta_t | \tilde{\Theta}_{t-1}^{(i)}(\psi), Y_{t-1}, \psi)$. L'inefficienza di questo stimatore Monte Carlo *naturale* è già stata discussa da un punto di vista teorico nel capitolo precedente, qui aggiungiamo solo che le densità p sono direttamente ricavate dalla specificazione statistica del modello, nella formulazione di queste densità, viene completamente ignorato che le osservazioni di Y contengono informazioni essenziali sul processo latente sottostante. In altre parole, le traiettorie estratte dal processo p , virtualmente, non hanno alcuna relazione con la sequenza reale delle variabili latenti θ_t che hanno generato le variabili osservate, y_t .

Inserendosi, come il nome stesso suggerisce, nel contesto delle tecniche di *Importance Sampling*, l'EIS ricerca una sequenza di *importance sampler* che sfrutti efficientemente l'informazione riguardante il processo latente, θ_t , contenuta nelle osservazioni r_t . Possiamo indicare questa sequenza di *sampler ausiliari* con $\{m(\theta_t | \Theta_{t-1}, a_t)\}_{t=1}^T$, dove i parametri ausiliari sono indicati con la sequenza $A = \{a_t\}_{t=1}^T$.

Queste densità ausiliarie sono costruite come un'estensione immediata di quelle naturali $\{p(\theta_t | \Theta_{t-1}, Y_{t-1}, \psi)\}_{t=1}^T$, dove però una particolare scelta dei parametri ausiliari ne garantirà la maggiore efficienza. Possiamo iniziare riscrivendo la verosimiglianza della 5.1, evidenziando la componente di *Importance Sampling*

come

$$L(\psi; Y) = \int \prod_{t=1}^T \left[\frac{f(y_t, \theta_t | \Theta_{t-1}, Y_{t-1}, \psi)}{m(\theta_t | \Theta_{t-1}, a_t)} \right] \prod_{t=1}^T m(\theta_t | \Theta_{t-1}, a_t) d\Theta. \quad (5.2)$$

La corrispondente stima della verosimiglianza basata sull'*Importance Sampling* Monte Carlo è data da

$$\hat{L}_S(\psi; Y, A) = \frac{1}{S} \sum_{i=1}^S \left\{ \prod_{t=1}^T \left[\frac{f(y_t, \tilde{\theta}_t^{(i)}(a_t) | \tilde{\Theta}_{t-1}^{(i)}(a_{t-1}), Y_{t-1}, \psi)}{m(\tilde{\theta}_t^{(i)}(a_t) | \tilde{\Theta}_{t-1}^{(i)}(a_{t-1}), a_t)} \right] \right\}, \quad (5.3)$$

dove $\{\tilde{\theta}_t^{(i)}(a_t)\}_{t=1}^T$ indica una traiettoria estratta dalla sequenza di densità ausiliarie (*importance*) m . Se si trovasse una sequenza delle m densità tale che $\prod_t m(\lambda_t | \Theta_{t-1}, a_t)$ fosse proporzionale a $\prod_t f(y_t, \theta_t | \Theta_{t-1}, Y_{t-1}, \psi)$, la stima Monte Carlo \hat{L}_S coinciderebbe con la vera funzione di Verosimiglianza per tutti gli S e, di conseguenza, non ci sarebbe variabilità dovuta alla simulazione.

L'EIS cerca allora di selezionare i valori dei parametri ausiliari $\{a_t\}_{t=1}^T$ tali da minimizzare la variabilità Monte Carlo. L'elevata dimensione di questo problema di minimizzazione, che deve essere risolto per ogni valore rilevante del vettore dei parametri θ , può essere affrontata suddividendo il problema in una sequenza più facilmente affrontabile. Questa sequenza di sotto-problemi viene suggerita dalla forma fattorizzata nella 5.1, nella quale essenzialmente le osservazioni sono aggiunte una alla volta.

Ci rendiamo conto che, dal momento che per ogni valore, preso come dato, di Y_t e ψ , l'integrale di $f(y_t, \theta_t | \Theta_{t-1}, Y_{t-1}, \psi)$ rispetto a θ_t dipende da Θ_{t-1} mentre quello di $m(\theta_t | \Theta_{t-1}, a_t)$ è uguale a uno per definizione, non è possibile ottenere una scelta ottimale per i parametri ausiliari della densità m separatamente periodo per periodo.

Al contrario, per ogni (Y, ψ) dato, l'EIS richiede che si costruisca un'approssimazione, $k(\Theta_t; a_t)$, della densità $f(y_t, \theta_t | \Theta_{t-1}, Y_{t-1}, \psi)$ che sia però analiticamente integrabile rispetto a θ_t . Nella terminologia Bayesiana, $k(\Theta_t; a_t)$ rappresenta il *kernel* della densità $m(\theta_t | \Theta_{t-1}, a_t)$ la quale può essere quindi vista come

$$m(\theta_t | \Theta_{t-1}, a_t) = \frac{k(\Theta_t; a_t)}{\chi(\Theta_{t-1}, a_t)},$$

dove $\chi(\Theta_{t-1}, a_t) = \int k(\Theta_t; a_t) d\theta_t$.

In altre parole, concentrarsi solo su $f(y_t, \theta_t | \Theta_{t-1}, Y_{t-1}, \psi)$ e $k(\Theta_t; a_t)$, porterebbe a trascurare $\chi(\Theta_{t-1}; a_t)$. Dal momento che, tuttavia, $\chi(\Theta_{t-1}; a_t)$ non dipende da θ_t , può essere portato indietro al sotto-problema di minimizzazione del periodo $t - 1$.

In virtù di quanto detto, l'EIS richiede di risolvere una semplice sequenza ricorsiva (partendo dall'ultimo periodo e andando a ritroso), attraverso una serie di problemi di minimi quadrati a bassa dimensione della forma

$$\hat{a}_t(\psi) = \min_{a_t} \sum_{i=1}^S \left\{ \log \left[f(y_t, \tilde{\theta}_t^{(i)}(\psi) | \tilde{\Theta}_{t-1}^{(i)}(\psi), Y_{t-1}, \psi) \cdot \chi(\tilde{\Theta}_t^{(i)}; \hat{a}_{t+1}(\psi)) \right] \right. \quad (5.4)$$

$$\left. - c_t - \log k(\tilde{\Theta}_t^{(i)}; a_t) \right\}^2 \quad (5.5)$$

per $t = T, \dots, 1$ e con $\chi(\Theta_T; a_{T+1})$ posto uguale a 1. Ancora una volta, $\{\tilde{\theta}_t^{(i)}(\psi)\}_{t=1}^T$ indica una traiettoria estratta dalle densità p mentre le costanti di normalizzazione ignote, c_t , devono essere stimate congiuntamente agli a_t .

Possiamo quindi ottenere la stima EIS della funzione di verosimiglianza per un dato valore di ψ sostituendo $\{\hat{a}_t(\psi)\}_{t=1}^T$ agli $\{a_t(\psi)\}_{t=1}^T$ nella 5.3.

Iterazione dell'algoritmo e massimizzazione della verosimiglianza

Il problema di Minimi Quadrati Ordinari appena presentato, sfrutta (anche se tra breve vedremo un'alternativa più efficiente) traiettorie della variabile latente che sono estratte dal *sampler* naturali. Come abbiamo già più volte ripetuto, l'utilizzo di tali traiettorie per la stima della verosimiglianza stessa, risulta altamente inefficiente ma produce una sequenza di parametri ausiliari $\{\hat{a}_t(\theta)\}_{t=1}^T$ che possono poi essere impiegati per costruire delle densità *importance* molto più efficienti. E' allora assolutamente ragionevole reiterare questa procedura per un piccolo numero di volte, sostituendo all'inizio di ogni passo il *sampler* naturale p con l'*importance sampler* ottenuto al passo precedente per generare le traiettorie del processo latente, $\{\theta_t^{(i)}\}_{t=1}^T$.

Sulla base del principio già suggerito nel precedente capitolo, per assicurare la convergenza dei valori dei parametri ausiliari \hat{a}_t , i quali dovrebbero produrre gli *importance sampler* ottimali, è necessario applicare la tecnica dei CRN.

Nel lavoro Richard and Zhang (2007), se da una parte vengono mostrati i risultati di una procedura iterativa dove si impone un criterio di arresto (ovvero la differenza tra valori successivi dei parametri ausiliari deve essere inferiore ad una fissata, estremamente piccola, soglia di tolleranza) senza quindi pre-fissare un certo numero di passi, dall'altra parte viene anche sconsigliato tale approccio ma si raccomanda di ripetere la procedura 5 volte. Noi, nell'implementazione pratica dell'algoritmo, ci atterremo a questa indicazione.

Dopo aver ottenuto i valori ottimali dei parametri ausiliari potremo simulare gli *importance weights* e, grazie a questi, stimare la funzione di Verosimiglianza per un dato valore del vettore dei parametri, ψ . Come abbiamo fatto nel metodo SPDK,

possiamo, attraverso un ottimizzatore numerico iterativo, ottenere una massimizzazione di questa stima per ricavare la stima di Massima Verosimiglianza dei parametri stessi. Ancora una volta e specialmente per i metodi basati sul gradiente numerico, è necessario utilizzare i CRN, in modo tale da assicurare che $\hat{L}_S(\psi; Y, A)$ sia una funzione sufficientemente liscia in ψ .

Miglioramento del *sampler* iniziale

Come abbiamo più volte ripetuto, l'inefficienza del *sampler* iniziale p , risiede nel fatto che non sfrutta le informazioni sul processo latente presenti nelle osservazioni. Un'interessante alternativa a p , suggerita da Richard and Zhang (2007), consiste nell'utilizzare nella prima iterazione dell'EIS, un *sampler* ricavato da un'espansione di secondo ordine in serie di Taylor della log-densità $\log g(y_t|\theta_t, \psi)$. Tale densità, infatti, sfrutta l'informazione campionaria sui θ_t contenuta negli y_t . Inoltre, come abbiamo visto nel capitolo precedente, anche il metodo SPDK sfruttava una simile approssimazione; differenza fondamentale è che in questo contesto, la serie di Taylor non è centrata in $\hat{\theta}$ (la moda della distribuzione del processo latente) ma in 0, ovvero la media non condizionale dei rendimenti.

5.1.1 Filtraggio

Una volta che i parametri del modello sono stati stimati, l'EIS può anche essere usato per produrre una sequenza appropriata di stime filtrate della variabile latente θ_t o funzioni di questa che potranno servire come base per il controllo diagnostico. Se $h(\theta_t)$ indica una trasformazione del processo latente, la sequenza di valori attesi di $h(\theta_t)$ condizionatamente ai valori passati delle variabili osservate, Y_{t-1} costituisce una sequenza di stime filtrate di $h(\theta_t)$. E questi valori attesi assumo la forma di un rapporto di integrali:

$$E[h(\theta_t|Y_{t-1})] = \frac{\int h(\theta_t)p(\theta_t|\Theta_{t-1}, R_{t-1}, \psi)f(Y_{t-1}, \Theta_{t-1}; \psi)d\Theta_t}{\int f(Y_{t-1}, \Theta_{t-1}; \psi)d\Theta_{t-1}}. \quad (5.6)$$

Trattando il vettore dei parametri ψ come se fosse noto, ponendolo uguale, ad esempio, alla stima di Massima Verosimiglianza, l'integrale nel denominatore corrisponde alla funzione di Verosimiglianza $L(\psi; Y_{t-1})$ associata alle prime $t - 1$ osservazioni e può essere valutato per mezzo dell'integrazione Monte Carlo usando la sequenza di densità $\{m(\theta_t|\Theta_{t-1}, \hat{a}_\tau^{t-1})\}_{\tau=1}^{t-1}$, dove $\{\hat{a}_\tau^{t-1}\}_{\tau=1}^{t-1}$ indica i valori dei parametri ausiliari che caratterizzano i *samplers* EIS per la verosimiglianza $L(\psi; Y_{t-1})$. Una stima Monte Carlo ottenuta tramite l'*Importance Sampling* dell'integrale nel

numeratore è ottenuta da:

$$\frac{1}{S} \sum_{i=1}^S \left\{ h[\tilde{\theta}_t^{(i)}(\psi)] \prod_{\tau=1}^{t-1} \left[\frac{f(y_t, \tilde{\theta}_{\tau-1}^{(i)}) |\tilde{\Lambda}_{\tau-1}^{(i)}(\hat{a}_{\tau-1}^{t-1}, Y_{\tau-1}, \psi)}{m(\tilde{\theta}_{\tau-1}^{(i)}) |\tilde{\Lambda}_{\tau-1}^{(i)}(\hat{a}_{\tau-1}^{t-1}, \hat{a}_{\tau-1}^{t-1})} \right] \right\},$$

dove $\{\tilde{\theta}_\tau^{(i)}(\hat{a}_\tau^{t-1})\}_{\tau=1}^{t-1}$ indica una traiettoria estratta dalla sequenza dei *samplers* EIS associati a $L(\psi; Y_{t-1})$ e $\tilde{\theta}_\tau^{(i)}(\psi)$ un'estrazione dalla densità $p(\theta_t | \tilde{\Theta}_{t-1}^{(i)}(\hat{a}_{t-1}^{t-1}, Y_{t-1}, \psi))$. Ricordiamo che usare le stesse estrazioni causali nella valutazione del numeratore e del denominatore della 5.6 è consigliabile perché può ridurre la varianza dovuta alla simulazione.

5.2 Applicazione dell'EIS al modello SV

5.2.1 Il modello

Ripresentiamo, ancora una volta, il modello SV Log-Normale,

$$\begin{aligned} r_t &= e^{\theta_t/2} \epsilon_t, & \epsilon_t &\sim IN(0, 1) \\ \theta_{t+1} &= \gamma + \phi\theta_t + \nu\eta_t, & \eta_t &\sim IN(0, 1), \\ \epsilon_t &\perp \eta_s, & \forall t, s. \end{aligned}$$

La densità dei rendimenti, r_t , condizionatamente alla log-volatilità θ_t e quella della log-volatilità θ_t condizionatamente al suo valore passato θ_{t-1} , a meno di fattori moltiplicativi che non dipendono da θ_t sono proporzionali a:

$$g(r_t|\theta_t) \propto \exp\left\{-\frac{1}{2}\left[r_t^2 \exp(-\theta_t) + \theta_t\right]\right\}$$

e

$$p(\theta_t|\theta_{t-1}) \propto \exp\left\{-\frac{1}{2\nu^2}(\theta_t - \gamma - \phi\theta_{t-1})^2\right\}.$$

5.2.2 Implementazione dell'algoritmo

L'implementazione dell'EIS richiede, come prima cosa, la selezione all'interno della classe di *importance samplers*, m , di una densità *kernel* k , capace di fornire una buona approssimazione funzionale dei prodotti $f \cdot \chi$. Come abbiamo detto nella precedente sezione, una scelta naturale per m è costituita dall'utilizzo di un'estensione parametrica del *sampler* naturale p . Inoltre, il fatto che la densità Normale sia chiusa rispetto a trasformazioni di scala, suggerisce immediatamente la seguente parametrizzazione per k :

$$k(\Theta_t; a_t) = p(\theta_t|\theta_{t-1}; \psi)\zeta(\theta_t, a_t),$$

dove la funzione ausiliaria $\zeta(\theta_t, a_t)$ è essa stessa il *kernel* di una densità Gaussiana. Sotto questa parametrizzazione, il *sampler* naturale p viene eliminato dal problema dei Minimi Quadrati definito nella 5.4 con la conseguenza che $\log \zeta(\theta_t, a_t)$ serve ad approssimare $\log(r_t|\theta_t, R_{t-1}, \psi) + \log \chi(\Theta_t, a_{t+1})$. Più precisamente, la funzione ausiliaria approssimata per il modello SV canonico è data da $\zeta(\theta_t, a_t) = \exp(a_{1,t}\theta_t + a_{2,t}\theta_t^2)$, con $a_t = (a_{1,t}, a_{2,t})$. I *kernels* dell'*importance samplers* hanno la seguente

forma

$$k(\Theta_t; a_t) \propto \exp \left\{ -\frac{1}{2} \left[\left(\frac{\gamma - \phi\theta_{t-1}}{\nu} \right)^2 - 2 \left(\frac{\gamma + \phi\theta_{t-1}}{\nu^2} + a_{1,t} \right) \theta_t + \left(\frac{1}{\nu^2} - 2a_{2,t} \right) \theta_t^2 \right] \right\}.$$

Conseguentemente, la media e la varianza condizionale di θ_t in m sono date, rispettivamente, da

$$\mu_t = \sigma_t^2 \left(\frac{\gamma + \phi\theta_{t-1}}{\nu^2} + a_{1,t} \right) \quad (5.7)$$

e

$$\sigma_t^2 = \frac{\nu^2}{1 - 2\nu^2 a_{2,t}}. \quad (5.8)$$

Integrando $k(\Theta_t; a_t)$ rispetto a θ_t , otteniamo l'espressione seguente per la costante di integrazione (nella quale omettiamo i fattori moltiplicativi non rilevanti):

$$\chi(\theta_{t-1}, a_t) \propto \exp \left\{ \frac{\mu_t^2}{2\sigma_t^2} - \frac{(\gamma + \phi\theta_{t-1})^2}{2\nu^2} \right\},$$

che costituisce ancora una volta un *kernel* Gaussiano della densità di θ_{t-1} .

Tuttavia, come abbiamo anche precedentemente menzionato, possiamo incrementare l'efficienza dell'algoritmo EIS utilizzando al primo passo, in sostituzione del *sampler* naturale p , un *importance sampler* basato su una espansione del secondo ordine in serie di Taylor di $\log g(r_t|\theta_t, \psi)$. Per il modello SV canonico, tale espansione di Taylor ² centrata in 0, è data da

$$\log g(r_t|\theta_t, \psi) \propto -\frac{1}{2}[r_t^2 \exp(-\theta_t) + \theta_t] \approx \text{costante} + \frac{1}{2}[r_t^2 - 1]\theta_t - \frac{1}{4}r_t^2\theta_t^2. \quad (5.9)$$

Tale approssimazione implica i valori seguenti per i coefficienti della funzione ausiliaria ζ : $a_{1,t} = (r_t^2 - 1)/2$ e $a_{2,t} = -(r_t^2)/4$.

Questi valori sono inseriti nella 5.7 e nella 5.8 per ottenere i momenti dei corrispondenti *sampler* Gaussiani basati sulla TSE. Otteniamo in questo modo, dopo

²Ricordiamo la formula generica dell'espansione in serie di Taylor in un generico intorno a :

$$T(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n$$

alcuni passaggi algebrici,

$$\mu_t = \sigma_t^2 \left(\frac{\gamma + \phi\theta_{t-1}}{\nu^2} + \frac{r_t^2 - 1}{2} \right) \quad (5.10)$$

e

$$\sigma_t^2 = \frac{2\nu^2}{2 + \nu^2 r_t^2}. \quad (5.11)$$

Trovati i momenti del *sampler* Gaussiano basato sulla TSE, possiamo estrarre S traiettorie della variabile latente $\{\tilde{\theta}_t^{(i)}(\psi)\}_{t=1}^T$.

Queste estrazione casuali sono poi usate per risolvere la sequenza ricorsiva (all'indietro) dei problemi di minimi quadrati definita nella 5.5. Per ogni t , questo problema di minimi quadrati è caratterizzato dalla regressione lineare ausiliaria seguente:

$$\begin{aligned} & - \frac{1}{2} \left[r_t^2 \exp(-\tilde{\theta}_t^{(i)}(\psi) + \tilde{\theta}_t^{(i)}(\psi)) \right] + \log \chi(\tilde{\theta}_t^{(i)}(\psi), \hat{a}_{t+1}(\psi)) \\ & = \text{costante} + a_{1,t} \tilde{\theta}_t^{(i)}(\psi) + a_{2,t} [\tilde{\theta}_t^{(i)}(\psi)]^2 + \eta_i^{(i)}, \end{aligned}$$

dove con $\eta_i^{(i)}$ indichiamo il termine d'errore nella regressione. Ricordiamo che l'iniziale condizione per la costante di integrazione è data da $\chi(\theta_T, \cdot) = 1$.

I *samplers* EIS $\{m(\theta_t | \Theta_{t-1}, \hat{a}_t(\psi))\}_{t=1}^T$, caratterizzati dalla media e dalla varianza condizionale espresse nelle 5.10 e 5.11 sono usati per estrarre S traiettorie del processo latente $\{\tilde{\theta}_t^{(i)}\}_{t=1}^T$. Dopo tre iterazioni dell'algorithm EIS, le S traiettorie sono utilizzate per costruire gli *importance weights* necessari per la stima della verosimiglianza, in accordo con l'5.3.

Applicazione dell'EIS ai dati simulati

	Vero θ	$\hat{\theta}_{EIS}$	S.E.	Limite Inferiore	Limite Superiore
γ	-0.368	- 0.3689	0.0279	-0.4236	-0.3142
ν^2	0.0676	0.1038	0.0054	0.093	0.1146
ϕ	0.95	0.9854	0.0148	0.9563	1.0144
Log-verosimiglianza				-5988.788	S=50

Tabella 5.1: Risultati della stima basata sull'EIS

Capitolo 6

Analisi di una serie reale e conclusioni

6.1 Verifica empirica

Per l'analisi di dati reali, si è scelto di utilizzare un campione dell'indice americano S&P500 di $T = 2022$ osservazioni, distribuite tra il 2 Gennaio 1980 e il 30 Dicembre 1987, già studiato in precedenti lavori sui modelli a Volatilità Stocastica, tra i quali Sandmann and Koopman (1998). Questo ha reso possibile una validazione degli algoritmi che si sono dovuti implementare per stimare i modelli proposti in questo lavoro. Le indagini grafiche preliminari riportate nella Figura 6.1, mostrano che la volatilità della serie si presta ad essere modellata dal momento che i primi ritardi dell'ACF dei quadrati dei rendimenti sono significativi.

6.1.1 Analisi preliminari

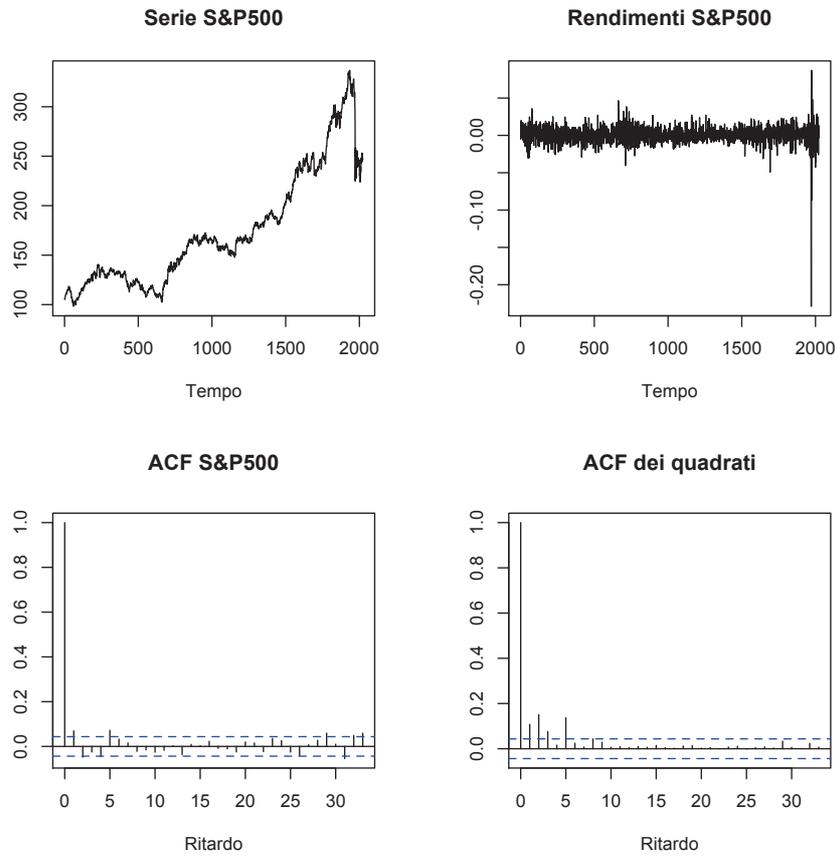


Figura 6.1:
Analisi grafiche preliminari per la serie S&P500

6.1.2 Stima dei parametri

Come avevamo fatto per il campione simulato, anche qui utilizziamo 50 simulazioni per il metodo SPDK e per il metodo EIS. Per quanto riguarda i metodi Bayesiani MCMC, per il metodo che utilizza il *Gibbs sampler* sono state generate 350000 replicazioni e scartate le prime 50000 per il metodo basato sulla mistura di Normali 100000 di cui 10000 scartate. Per questo motivo possiamo osservare nella Tabella 6.1 Deviazioni Standard più basse per il secondo metodo MCMC. Le stime di QMV si discostano molto da quelle ottenute con gli altri metodi, ad eccezione, del parametro di persistenza per cui tutti i metodi stimano un valore molto elevato. Le stime ottenute con il metodo SPDK sono molto vicine alle stime delle medie a posteriori ottenute con il *Gibbs sampler*. Possiamo notare che per quanto riguarda il parametro della persistenza del processo latente i diversi metodi giungono a stime molto vicine.

QMV	$\hat{\theta}$	S.E.	Limite Inferiore	Limite Superiore
γ	-0.04162	0.03348	-0.10724	0.02401
ν	0.04291	0.01154	0.02029	0.06553
ϕ	0.99547	0.00359	0.98843	1.00252
Log- like		-4117.28		
Gibbs	$\hat{E}[\theta]$	Deviazione standard	Quantile 0.025	Quantile 0.975
γ	-0.25838	0.00211	-0.45718	-0.11452
ν	0.15303	0.00051	0.11613	0.20021
ϕ	0.97261	0.00022	0.95165	0.98785
Mistura	$\hat{E}[\theta]$	Deviazione standard	Quantile 0.025	Quantile 0.975
γ	-0.15943	0.00087	-0.30322	-0.06031
ν	0.07573	0.00025	0.05682	0.09974
ϕ	0.98201	0.00040	0.96581	0.99316
SPDK	$\hat{\theta}$	S.E.	Limite Inferiore	Limite Superiore
γ	-0.2785	0.0877	-0.4506	-0.1065
ν	0.1607	0.0221	0.1173	0.2042
ϕ	0.9705	0.0092	0.9523	0.9886
Log-like		-4093.362	S.E.	0.06992201
EIS	$\hat{\theta}$	S.E.	Limite Inferiore	Limite Superiore
γ	-0.1844	0.1274	-0.43410	0.065304
ν	0.1179	0.0578	0.00461	0.231188
ϕ	0.9830	0.0215	0.94086	1.02514
Log-like		-4530.432	S.E.	6.795384

Tabella 6.1: Risultati dell'inferenza nei diversi metodi proposti per la serie S&P500

6.1.3 Analisi grafiche

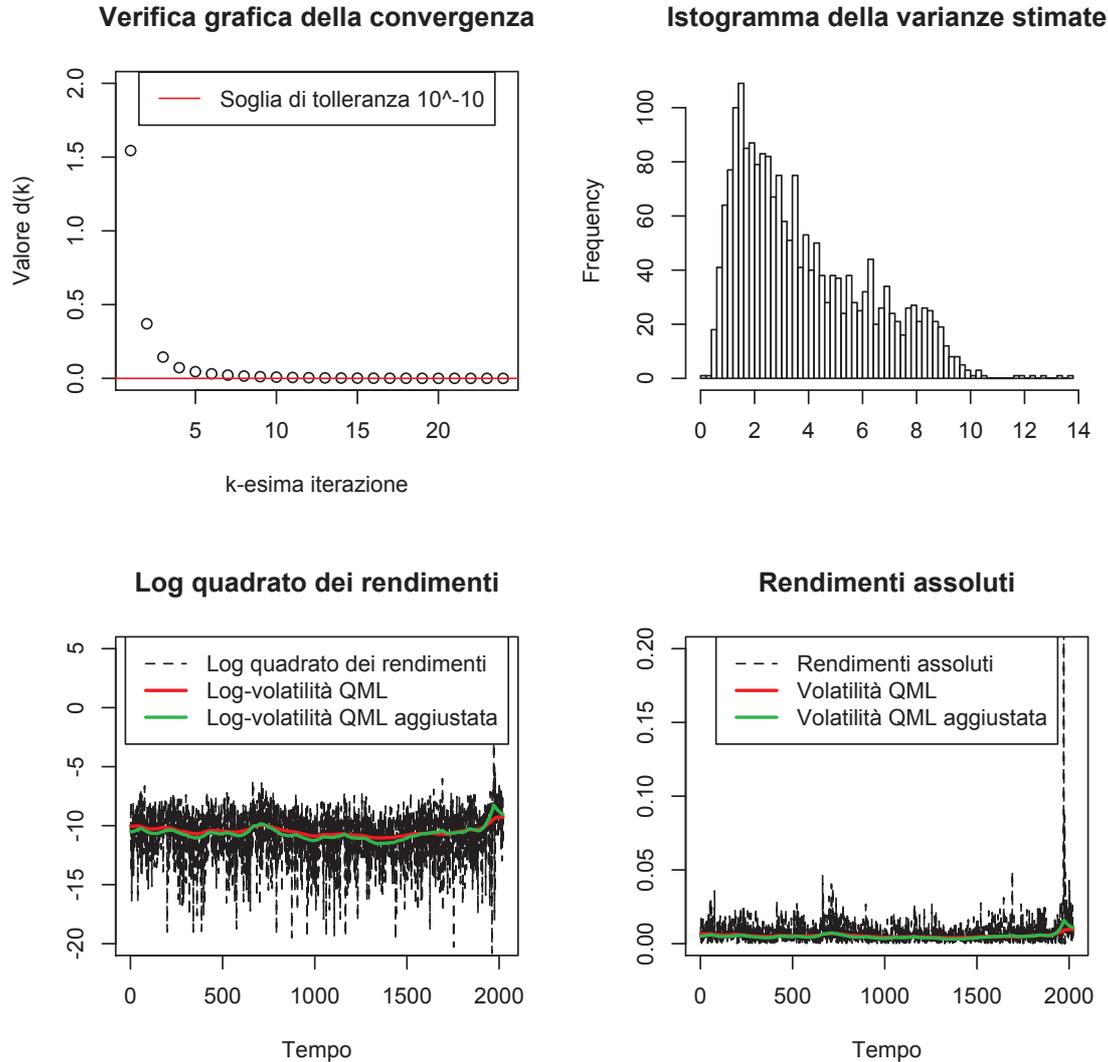
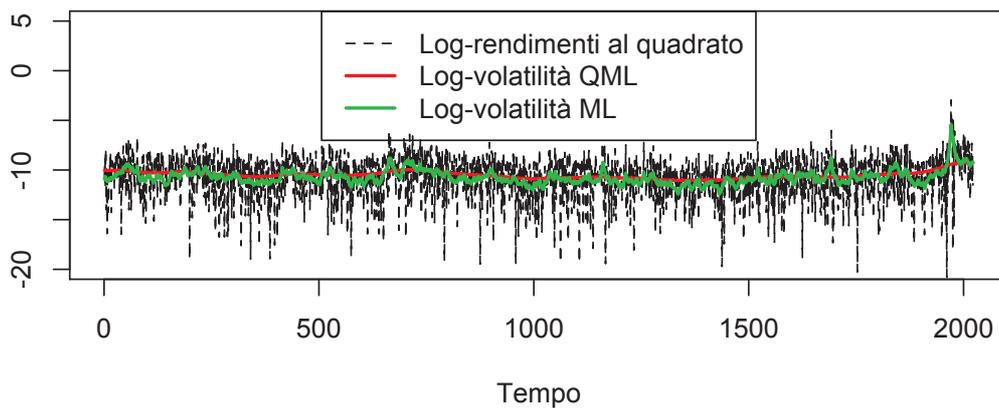


Figura 6.2:
Varianze approssimate e volatilità *smoothed* per la serie S&P500

I risultati grafici riportati nelle Tabelle 6.2 e 6.3 mostrano chiaramente il miglioramento offerto dal metodo SPDK: la stima *smoothed* della volatilità riesce a seguire molto più fedelmente la sua versione empirica di quanto non faccia il modello di QMV, specialmente nel rendimento anomalo che si è verificato verso la fine della serie. I grafici relativi ai metodi MCMC sono stati omessi perché concordano con quelli presentati per il campione simulato nel suggerire che nel modello con mistura di Normali la convergenza della media a posteriori venga ottenuta con un numero minore di simulazioni.

Log-rendimenti al quadrato vs log-volatilità stimata



Rendimenti assoluti vs volatilità stimata

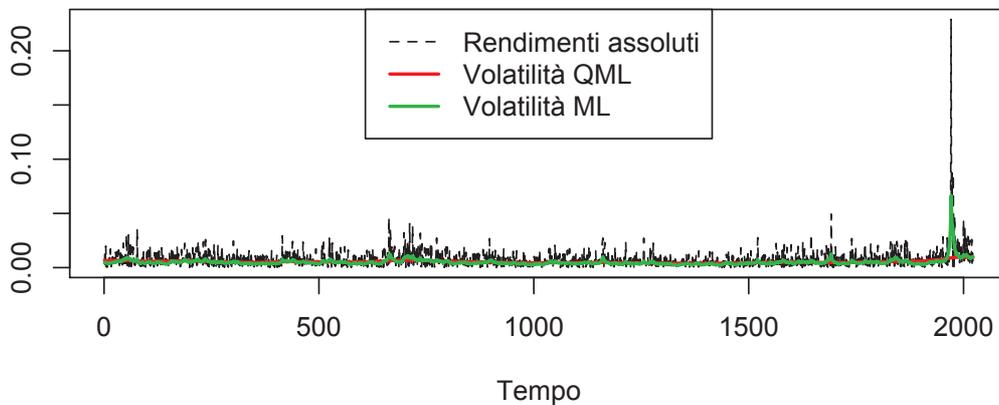


Figura 6.3:
Volatilità *smoothed* tramite procedura SPDK-ML per la serie S&P500

6.2 Conclusioni

Questo lavoro ha confermato le interessanti proprietà godute dal modello a Volatilità Stocastica. Sembra appropriato ipotizzare che la dinamica della variabilità delle serie storiche finanziarie sia guidata da un processo stocastico non osservabile. L'attenzione volta a comprendere e implementare metodi sofisticati per stimare i parametri del modello sembra essere stata ripagata. Se da un lato, infatti, abbiamo trovato conferma delle scarse qualità inferenziali del metodo di Quasi Massima Verosimiglianza, dove la deviazione dall'ipotesi di Normalità influenza negativamente la correttezza delle stime, nonché l'accuratezza della stima *smoothed* del segnale, i metodi presentati nei capitoli successivi hanno mostrato di riuscire a ottenere, a costo di un qualche sforzo computazionale, buoni risultati.

In modo particolarmente interessante, a nostro avviso, il metodo chiamato SPDK riesce a sfruttare le procedure derivate dal filtro di Kalman, correggendo efficientemente la lontananza dall'assunzione di Normalità. Invece l'*Efficient Importance Sampling* non ha del tutto confermato le sue attese poiché, benché nel suo sviluppo teorico si focalizzi sul'ottenere una stima quanto più precisa possibile della Verosimiglianza, a questo sforzo non sono corrisposti apprezzabili risultati, i quali non si discostano troppo da quelli del già citato metodo SPDK e da quelli ottenuti tramite l'utilizzo di particolari algoritmi MCMC. Stimolante è risultato anche l'approccio Bayesiano dove, come nel metodo SPDK, lo sforzo di recuperare parte dei risultati dei modelli *state space* è stato ampiamente compensato da un guadagno in efficienza verificabile se confrontato ad un metodo basato sul *Gibbs sampler*.

Elenco delle figure

2.1	Densità di ξ_t e dell'approssimazione $N(-1.27, \pi^2/2)$	19
2.2	Volatilità <i>smoothed</i> ottenuta tramite procedura QML per il campione simulato	20
3.1	Istogramma di 10 000 000 estrazioni casuali dalla mistura di normali vs vera densità di ξ_t	42
3.2	Analisi grafica dei risultati ottenuti con il <i>Gibbs sampler</i> per il campione simulato	45
3.3	Analisi grafica dei risultati ottenuti con <i>sampler</i> basato sulla mistura di Normali	46
3.4	Stima filtrata del processo della volatilità basata sul <i>particle filter</i> .	47
4.1	Varianze approssimate e volatilità <i>smoothed</i> corretta per il campione simulato	59
4.2	Volatilità <i>smoothed</i> tramite procedura SPDK-ML per dati simulati .	62
6.1	Analisi grafiche preliminari per la serie S&P500	76
6.2	Varianze approssimate e volatilità <i>smoothed</i> per la serie S&P500 . .	79
6.3	Volatilità <i>smoothed</i> tramite procedura SPDK-ML per la serie S&P500	80

Elenco delle tabelle

2.1	Risultati della stima QML nel campione simulato	19
3.1	Pesi, medie e varianze per le componenti della mistura di $K = 7$ Normali	42
3.2	Risultati dei metodi MCMC	44
4.1	Statistiche riassuntive della varianze approssimate per il campione simulato	58
4.2	Risultati della stima SPDK-ML per dati simulati	63
5.1	Risultati della stima basata sull'EIS	73
6.1	Risultati dell'inferenza nei diversi metodi proposti per la serie S&P500	78

Bibliografia

- Andersen, T., Sørensen, B., 1996. GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study. *Journal of Business & Economic Statistics*.
- Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31 (3).
- Brockwell, P., Davis, R., 2009. *Time series: Theory and Methods*. Springer.
- Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings Algorithm. *American Statistician*.
- Davis, R., Mikosch, T., 2000. The Sample Autocorrelations of Financial Time Series Models. *Nonlinear and Nonstationary Signal Processing*, 247–274.
- Davis, R., Mikosch, T., 2009. Probabilistic Properties of Stochastic Volatility Models. In: *Handbook of Financial Time Series*. Springer, pp. 255–267.
- Dunsmuir, W., 1979. A Central Limit Theorem for Parameter Estimation in Stationary Vector time series and its application to models for a signal observed with noise. *The Annals of Statistics*, 490–506.
- Durbin, J., Koopman, S., 1997. Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models. *Biometrika* 84 (3).
- Durbin, J., Koopman, S., 1998. *Time Series Analysis for non-Gaussian Observations based on State Space Models*, mimeo, London School of Economics.
- Durbin, J., Koopman, S., 2001. *Time Series Analysis by State Space Methods*. Oxford University Press.
- Durbin, J., Koopman, S., 2002. A Simple and Efficient Simulation Smoother for State Space Time Series Analysis. *Biometrika* 89 (3).
- Engle, R., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*.

- Fama, E., 1965. The Behavior of Stock-Market Prices. *The Journal of Business* 38 (1).
- Ferguson, T., 1983. Bayesian Density Estimation by Mixtures of Normal Distributions. *Recent Advances in Statistics*.
- Fuller, W., 1996. *Introduction to Statistical Time Series*. Wiley.
- Geweke, J., 1989. Bayesian Inference in Econometric Models using Monte Carlo Integration. *Econometrica*.
- Gourieroux, M., Gourieroux, C., Monfort, A., 1997. *Simulation Based Econometric Methods*. Cambridge University Press.
- Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate Stochastic Variance Models. *The Review of Economic Studies* 61 (2).
- Harvey, A., Shephard, N., 1993. *Estimation and Testing of Stochastic Variance Models*, mimeo, London School of Economics.
- Hastings, W., 1970. Monte Carlo Sampling Methods using Markov Chains and their Applications. *Biometrika* 57 (1).
- Hull, J., White, A., 1987. The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance*.
- Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics* 12 (4).
- Kalman, R., 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering* 82 (1).
- Kim, S., Shephard, N., Chib, S., 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies* 65 (3).
- Liesenfeld, R., Richard, J., 2003. Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics. *Journal of Empirical Finance* 10 (4).
- Mandelbrot, B., 1963. The Variation of Certain Speculative Prices. *The Journal of Business* 36 (4).
- Melino, A., Turnbull, S., 1990. Pricing Foreign Currency Options with Stochastic Volatility. *Journal of Econometrics* 45.

- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., et al., 1953. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* 21 (6).
- Quandt, R., Ramsey, J., 1978. Estimating Mixtures of Normal Distributions and Switching Regressions. *Journal of the American Statistical Association*.
- Richard, J., Zhang, W., 2007. Efficient high-dimensional Importance Sampling. *Journal of Econometrics* 141 (2).
- Ripley, B., 1987. *Stochastic Simulation*. Wiley Online Library.
- Robert, C., Casella, G., 2004. *Monte Carlo Statistical Methods*. Springer.
- Ruiz, E., 1994. Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of Econometrics* 63 (1).
- Sandmann, G., Koopman, S., 1998. Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood. *Journal of Econometrics* 87 (2).
- Schmidt, P., 1982. An Improved Version of the Quandt-Ramsey MGF Estimator for Mixtures of Normal Distributions and Switching Regressions. *Econometrica*.
- Shephard, N., 1993. Fitting Nonlinear Time-Series Models with Applications to Stochastic Variance Models. *Journal of Applied Econometrics* 8 (S1).
- Shephard, N., 1994. Partial non-Gaussian State Space. *Biometrika* 81 (1).
- Shephard, N., Pitt, M. K., 1997. Likelihood Analysis of Non-Gaussian Measurement Time Series. *Biometrika* 84 (3).
- Taylor, S., 1982. Financial Returns Modelled by the Product of Two Stochastic Processes, a Study of Daily Sugar Prices 1961-1979. In: *Time Series Analysis: Theory and Practice*. North-Holland: Amsterdam.
- Tutterington, D., Smith, A., Makov, U., 1985. *Statistical Analysis of Finite Mixture Distributions*. Wiley.

Siamo come nani sulle spalle di giganti, così che possiamo vedere più cose di loro e più lontane, non certo per l'altezza del nostro corpo, ma perché siamo sollevati e portati in alto dalla statura dei giganti.

Bernardo di Chartres