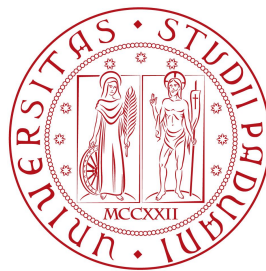


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**Raggruppamento di serie storiche finanziarie in
presenza di cambiamenti di regime: un approccio
basato su modelli a mistura finita**

Relatrice: prof.ssa Giovanna Menardi

Dipartimento di Scienze Statistiche

Correlatore: prof. Massimiliano Caporin

Dipartimento di Scienze Statistiche

Laureanda: Ilenia Franco

Matricola n. 2072677

Anno Accademico 2023/2024

*"Gli uomini passano,
le idee restano
e continuano a camminare
sulle gambe di altri uomini."
~ Giovanni Falcone*

Indice

Introduzione	1
1 Modelli statistici per serie storiche finanziarie	5
1.1 Introduzione alle serie storiche finanziarie	5
1.2 Modelli autoregressivi condizionatamente eteroschedastici	6
1.2.1 Presentazione generale	6
1.2.2 Modelli della classe GARCH	8
1.3 Modelli per serie storiche finanziarie con cambiamenti di regime	11
1.3.1 Serie storiche finanziarie con cambiamenti di regime	11
1.3.2 I modelli ST-GARCH	12
2 Modelli statistici per l'identificazione di gruppi	14
2.1 <i>Clustering</i> basato su modelli a mistura finita	14
2.1.1 Introduzione ai metodi di <i>clustering</i>	14
2.1.2 Specificazione del modello	15
2.1.3 Stima di massima verosimiglianza mediante algoritmo EM	16
2.2 <i>Co-clustering</i> basato su modelli mistura	20
2.2.1 Introduzione al <i>co-clustering</i>	20
2.2.2 Il modello a blocchi latenti	21
2.2.3 Stima di massima verosimiglianza mediante algoritmo LBEM	24
3 Raggruppamento di serie storiche finanziarie in presenza di cambiamenti di regime	27
3.1 Un modello per l'identificazione di gruppi di serie storiche con cambiamenti di regime	27
3.2 Specificazione del modello	28
3.3 Stima di massima verosimiglianza	30
3.3.1 La verosimiglianza completa	30

3.3.2	Algoritmo CEM	32
3.3.3	Algoritmo SEM- <i>Gibbs</i>	36
4	Alcune analisi empiriche	41
4.1	Uno studio di simulazione	41
4.1.1	Obiettivi e scenari di simulazione	41
4.1.2	Dettagli di implementazione	42
4.1.3	Analisi dei risultati	44
4.2	Un'applicazione a dati reali	46
4.2.1	I dati EURO STOXX 50	46
4.2.2	Dettagli di implementazione	50
4.2.3	Analisi dei risultati	51
	Conclusioni	54
	Bibliografia	56

Introduzione

Le serie storiche finanziarie rappresentano una sequenza di dati raccolti nel tempo riguardanti, ad esempio, i prezzi delle azioni, i tassi di cambio, i tassi di interesse o i volumi di scambio di attività finanziarie. Catturando fluttuazioni e tendenze dei mercati, tali dati sono fondamentali per comprendere le dinamiche economiche. In particolare, la loro analisi permette di identificare *pattern* ricorrenti, comportamenti anomali e relazioni tra variabili economiche, offrendo strumenti preziosi per la gestione del rischio e la previsione dei movimenti futuri. L'analisi delle serie storiche finanziarie richiede l'adozione di metodi sofisticati, poiché queste presentano spesso caratteristiche complesse come la non stazionarietà e l'eteroschedasticità, che le differenziano dalle più classiche serie storiche economiche. In base a tali caratteristiche, la variabilità del dato rappresenta uno dei fattori di rischio maggiormente presi in considerazione dagli operatori finanziari, divenendo oggetto di numerosi sviluppi teorici e applicativi negli ultimi decenni.

Tra i tratti caratteristici degli strumenti finanziari, è ampiamente riconosciuto in letteratura che aziende con caratteristiche simili tendono a comportarsi in maniera analoga nel tempo, poiché reagiscono in modo simile agli stessi fattori di mercato. Lo studio approfondito di queste relazioni, positive o negative, è fondamentale in contesti come l'*asset allocation* e il *risk management*. Comprendere le caratteristiche specifiche e le dinamiche di ciascun *asset* all'interno di un portafoglio è particolarmente rilevante in relazione al principio di diversificazione, una strategia chiave per minimizzare il rischio non sistematico. In questo contesto, diversificare significa distribuire il rischio su una varietà di *asset* non perfettamente correlati, riducendo l'impatto di eventuali perdite significative su un singolo strumento (Markowitz, 1959). Le correlazioni possono amplificare o mitigare il rischio, a seconda che gli *asset* si muovano in modo simile od opposto in risposta a eventi di mercato. Analizzare queste interdipendenze consente di identificare potenziali rischi sistemici e di adottare strategie di copertura o diversificazione per

proteggere il portafoglio dalle fluttuazioni negative del mercato. La diversificazione, quindi, non solo riduce il rischio complessivo, ma migliora anche le opportunità di rendimento attraverso una gestione del rischio più efficace e informata.

Nonostante l'ampia letteratura sulle serie storiche finanziarie, si riscontra una generale carenza di strumenti che tengano conto della possibilità che attività simili possano divergere nel tempo in risposta a cambiamenti di regime. I cambiamenti strutturali nel processo della varianza sono un fenomeno ampiamente riconosciuto nella letteratura economica come caratteristico dei mercati finanziari. Tali cambiamenti sono generalmente imputabili a fattori esterni alle singole serie storiche, come l'introduzione di nuove regolamentazioni o l'influenza di fattori sociali, politici ed economici. La mancata considerazione di questi potenziali cambiamenti di regime può influenzare negativamente le *performance* del portafoglio e, in generale, l'efficacia della gestione del rischio.

Da un punto di vista statistico, un possibile approccio per descrivere il contesto delineato consiste nell'identificazione di *cluster* variabili nel tempo, capaci di catturare le complessità e le transizioni caratteristiche dei mercati finanziari a seconda delle diverse fasi del mercato. Inoltre, lavorando su un insieme di serie storiche, il *clustering* consente agli strumenti finanziari simili di reagire in modo differenziato in risposta ad uno *shock* esterno.

L'obiettivo della presente relazione è analizzare le reazioni di diverse serie storiche finanziarie ad uno *shock* comune, raggruppandole in base al comportamento prima e dopo lo *shock*. Il contributo consiste nella formulazione di un modello statistico per l'identificazione di gruppi di serie storiche finanziarie in presenza di cambiamenti di regime. Coerentemente con il contesto in esame, l'attenzione è rivolta, in particolare, al caso di modelli autoregressivi condizionatamente eteroschedastici generalizzati (GARCH). Inoltre, poiché il passaggio da un regime all'altro non avviene necessariamente in modo brusco, ma può attraversare regimi intermedi, risultano particolarmente appropriati a descrivere la volatilità quei modelli i cui parametri possono variare gradualmente nel tempo, come ad esempio i modelli *smooth-transition*.

Ai fini dell'inferenza sui parametri del modello, nella tesi vengono sviluppate due varianti dell'algoritmo EM per la stima di massima verosimiglianza, specificamente formulate per il problema in esame. A questo scopo, data la necessità di identificare *cluster* diversi a seconda del regime del mercato, vengono presi in prestito alcuni strumenti propri dei modelli di *co-clustering*, che pur avendo un diverso obiettivo,

presentano diverse analogie con il contesto d'interesse.

La trattazione si sviluppa come segue.

Nel Capitolo 1, dopo una breve introduzione alle proprietà che caratterizzano le serie storiche dei prezzi e dei rendimenti di strumenti finanziari, verranno presentati i principali modelli statistici impiegati per l'analisi di questi dati, con particolare enfasi sui modelli autoregressivi condizionatamente eteroschedastici. Tra questi, sarà dedicata una sezione all'esplorazione di alcuni modelli che tengono conto della possibile presenza di cambiamenti di regime nella dinamica dei mercati finanziari, con un focus particolare sul modello *Smooth Transition-GARCH*, che consente una transizione graduale tra i diversi regimi.

Nel Capitolo 2 verranno approfonditi i metodi statistici per l'identificazione di gruppi mediante il *clustering* basato su modelli a mistura finita. Il capitolo inizierà con una breve panoramica generale sugli approcci tradizionali all'analisi di raggruppamento, evidenziandone i limiti e le differenze rispetto all'approccio parametrico. Successivamente, si passerà alla specificazione dettagliata del modello a mistura finita e alla discussione sui metodi di stima dei parametri. Infine, verranno esplorati i concetti relativi al *co-clustering* come naturale sviluppo del *clustering*, con particolare enfasi sul modello a blocchi latenti e sulle tecniche di stima di massima verosimiglianza.

Nel Capitolo 3 verranno illustrate le metodologie statistiche adottate per affrontare il problema oggetto di studio, ovvero le tecniche parametriche per l'identificazione di gruppi di serie storiche finanziarie in presenza di cambiamenti di regime. Il modello proposto estenderà quelli discussi nel capitolo precedente, focalizzandosi su due distinti algoritmi per la stima di massima verosimiglianza dei parametri e la suddivisione in gruppi. Gli approcci verranno presentati attraverso l'analisi di serie storiche finanziarie distribuite secondo una delle forme più basilari del modello ST-GARCH, poiché l'obiettivo di quest'analisi è introdurre una nuova procedura di *clustering* piuttosto che individuare il modello più appropriato per i dati.

Nel Capitolo 4 si esporranno, in prima battuta, i risultati relativi ad uno studio di simulazione volto a valutare il comportamento e l'affidabilità dell'approccio proposto al variare di alcune caratteristiche dei dati. L'efficacia sarà misurata sulla base delle capacità dei due algoritmi proposti di ripartire correttamente le serie storiche in gruppi. Nella seconda parte del capitolo verrà, infine, presentata un'applicazione del modello e dei due algoritmi di stima proposti ad un insieme di dati reali relativi all'indice EURO STOXX 50, con l'obiettivo di individuare l'istante

temporale dello *shock* e di studiare il comportamento delle serie storiche prima e dopo lo stesso.

La relazione si conclude con una riflessione generale sui risultati ottenuti, offrendo anche diversi spunti per possibili applicazioni e miglioramenti futuri.

Capitolo 1

Modelli statistici per serie storiche finanziarie

1.1 Introduzione alle serie storiche finanziarie

Quando si parla di serie storiche finanziarie il riferimento è alle serie temporali di dati relativi a variabili economiche o di mercato, come ad esempio prezzi di azioni, tassi di cambio o volumi di scambio, e sono fondamentali per analizzare e prevedere le dinamiche dei mercati finanziari. Si differenziano dalle classiche serie storiche economiche per la presenza di alcuni fatti stilizzati caratteristici che rendono più complessa la gestione del dato, ovvero, secondo la terminologia del settore, dei "*risultati empirici così coerenti tra vari mercati da essere accettati come verità*" (Cont, 2001). Per quanto riguarda i prezzi, solitamente lo studio si concentra sui rendimenti, ossia sulla variazione (logaritmica) dei prezzi stessi. Inoltre, poiché i prezzi non sono costanti e non possono essere previsti deterministicamente, la loro variabilità rappresenta uno dei fattori di rischio maggiormente presi in considerazione dagli operatori finanziari. Nella terminologia economica, tale variabilità viene definita volatilità, il cui termine indica una quantità non misurabile ma quantificabile mediante misure specifiche (come, ad esempio, la varianza *rolling* dei rendimenti, i rendimenti al quadrato, etc.).

Le regolarità empiriche che si osservano nelle serie storiche finanziarie sono, dunque, le seguenti:

- non stazionarietà dei prezzi dovuta alla presenza di un trend stocastico con continui cambi del valore della media nel tempo;

- rendimenti stazionari in media (il più delle volte ≈ 0) ma condizionatamente eteroschedastici;
- distribuzione dei rendimenti approssimativamente asimmetrica e leptocurtica, quindi generalmente non-Normale;
- presenza di *volatility clustering*: periodi con elevata variabilità (in modulo) si alternano a periodi con bassa variabilità;
- rendimenti serialmente incorrelati, ossia con autocorrelazioni generalmente non significative: possono diventare debolmente correlati quando la frequenza di osservazione dei prezzi è elevata;

Oltre a tali fatti stilizzati che si presentano con regolarità, a caratterizzare le serie storiche finanziarie sono generalmente anche:

- presenza di un effetto stagionale quando la frequenza del dato è elevata;
- *leverage effect*: differente impatto degli *shock* negativi sulla volatilità rispetto all'impatto che hanno *shock* positivi della stessa entità;
- aumento della volatilità nella fase di apertura del mercato dovuto all'accumulo di informazioni durante il periodo di chiusura.

Per un approfondimento su questi aspetti si veda Cont (2001) e Sewell (2011).

1.2 Modelli autoregressivi condizionatamente eteroschedastici

1.2.1 Presentazione generale

Un modello soddisfacente per i rendimenti dovrebbe essere in grado di descrivere i principali fatti stilizzati presentati nel paragrafo precedente.

I tradizionali modelli econometrici per serie storiche, come quelli autoregressivi a media mobile (ARMA) introdotti da Box e Jenkins (1976), assumono che la varianza sia costante nel tempo, risultando quindi inadatti per descrivere il comportamento di attività finanziarie. Per consentire questa ipotesi, Engle (1982) ha introdotto una nuova classe di processi stocastici chiamati *modelli autoregressivi condizionatamente*

eteroschedastici (ARCH). Quest'ultimi hanno media zero, sono incorrelati serialmente e presentano varianze non costanti condizionatamente al passato, ma costanti non condizionatamente. Quindi, a differenza degli ARMA, la previsione un passo in avanti della varianza non è costante ma trae informazioni dal recente passato. A partire dal prezzo p_t al tempo t di un dato strumento finanziario, la modellazione dell'ARCH(p) si concentra sul rispettivo rendimento logaritmico

$$y_t = \log(p_t) - \log(p_{t-1})$$

Data la serie storica $\{y_t\}_{t=1,\dots,T}$ dei rendimenti di lunghezza T e l'informazione passata I_{t-1} disponibile fino al tempo $t-1$, il modello ARCH(p) si compone di due parti principali, un'equazione per i rendimenti e una per la varianza condizionata che si suppone seguire un processo AR(p):

$$\begin{aligned} y_t &= \sigma_t \epsilon_t && \text{con } \epsilon_t \sim D.i.d.(0, 1) \\ y_t | I_{t-1} &\sim D(0, \sigma_t^2) \\ \sigma_t^2 &= \omega + \alpha_1 y_{t-1}^2 + \dots + \alpha_p y_{t-p}^2 \end{aligned}$$

dove D è una generica distribuzione per le innovazioni ϵ_t (ad esempio Normale, t di Student, etc.) e $\omega, \alpha_1, \dots, \alpha_p$ sono i parametri del modello stimabili mediante il metodo dei minimi quadrati o la massima verosimiglianza.

Il modello ARCH è sicuramente uno dei più semplici per la modellazione della volatilità in finanza ma presenta alcuni limiti. Infatti, potrebbe non catturare alcune caratteristiche della volatilità come l'asimmetria, la persistenza o la memoria lunga; potrebbe, inoltre, non tenere conto di altre fonti di volatilità, come shock esogeni, rotture strutturali o cambiamenti di regime. Per poter superare queste limitazioni, Bollerslev (1986) ha sviluppato un modello più sofisticato, detto *ARCH-generalizzato* (GARCH), con l'unica differenza che nella specificazione della varianza condizionata si ha un processo autoregressivo a media mobile. Grazie alla presenza dei ritardi passati non solo dei rendimenti y ma anche della varianza condizionata σ^2 , il modello GARCH possiede diverse varianti che possono incorporare le caratteristiche non catturate dal semplice ARCH. Si discuterà più nel dettaglio del modello GARCH e si accennerà ad alcune sue varianti nel prossimo paragrafo.

Un'altra classe di modelli particolarmente adeguata a descrivere serie storiche finanziarie è costituita dai modelli a volatilità stocastica (SV), di cui viene fornito un iniziale lavoro da Taylor (1982, 1994) e una più specifica formulazione da Shephard (1996). A differenza dei modelli

ARCH e GARCH che trattano la volatilità come una funzione deterministica dei dati passati, i modelli SV assumono che la volatilità stessa segua un processo stocastico. Il vantaggio risiede in una rappresentazione più realistica della volatilità rispetto ai GARCH, perché permette una maggiore flessibilità nel catturare fenomeni come la persistenza e i cambiamenti improvvisi nei livelli di rischio. Tuttavia, i modelli SV sono più complessi da stimare e richiedono tecniche avanzate come i metodi bayesiani o il filtro di Kalman (Harvey *et al.*, 1994).

Proprio per questa stima complicata che richiederebbe oneri computazionali difficilmente gestibili, i modelli GARCH sono quelli più diffusamente usati per descrivere il comportamento delle serie storiche finanziarie e rappresentano il punto di partenza di questo lavoro.

1.2.2 Modelli della classe GARCH

Come accennato in precedenza, i modelli GARCH (Bollerslev, 1986) sono una classe di modelli econometrici utilizzati per modellare la volatilità condizionata dei rendimenti finanziari. A differenza degli ARCH, consentono una maggiore flessibilità nella modellazione della volatilità nel tempo, supponendo una struttura di tipo ARMA per la varianza condizionata al passato.

Data, quindi, la serie storica $\{y_t\}_{t=1,\dots,T}$ dei rendimenti di lunghezza T e l'informazione passata I_{t-1} disponibile fino al tempo $t-1$, il modello GARCH(p,q) suppone che:

$$\begin{aligned} y_t &= \mu_t + \eta_t = \mu_t + \sigma_t \epsilon_t && \text{con } \epsilon_t \sim D.i.d.(0, 1) \\ \eta_t | I_{t-1} &\sim D(0, \sigma_t^2) \\ \sigma_t^2 &= \omega + \alpha_1 \eta_{t-1}^2 + \dots + \alpha_p \eta_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2 \end{aligned}$$

con ω parametro costante, $\alpha_1, \dots, \alpha_p$ coefficienti che misurano l'impatto delle "notizie" (o shock) passate sulla varianza condizionata, β_1, \dots, β_q coefficienti che misurano la persistenza della volatilità, ossia il fenomeno per cui periodi di alta volatilità tendono a essere seguiti da ulteriori periodi ad alta volatilità.

Il termine di media μ_t solitamente consiste in una costante o in un processo ARMA (da cui si origina il cosiddetto modello ARMA-GARCH(p,q)), ma poiché la maggior parte delle volte si riscontra che le serie dei rendimenti sono stazionarie in media con un trend attorno allo zero, si assume che la media del processo $\mu_t = 0 \forall t$. Per tale ipotesi nell'espressione della varianza parlare delle innovazioni η_t o dei rendimenti

y_t è equivalente, giungendo alla seguente formulazione:

$$y_t | I_{t-1} \sim D(0, \sigma_t^2)$$

$$\sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 + \dots + \alpha_p y_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2$$

Per quanto riguarda la distribuzione D delle innovazioni standardizzate scelte comuni sono, ad esempio, la Normale o la t di Student (Engle e Bollerslev, 1986). In questo lavoro ci si concentra su una specificazione gaussiana, che si adatta spesso bene ai dati ma non risulta essere quella più appropriata quando si è in presenza di leptocurtosi. In ogni caso, le successive considerazioni sono facilmente generalizzabili anche all'uso di altre distribuzioni.

Come indicato dal modello, σ_t^2 rappresenta la varianza condizionata del processo per y_t , la quale è dipendente dal tempo. Non condizionatamente, però, i processi stocastici della classe GARCH sono, di norma, debolmente stazionari, ossia caratterizzati da valore atteso e varianza finiti e non dipendenti dal tempo e da autocovarianze (o autocorrelazioni) che possono dipendere dal ritardo temporale ma non dal tempo. Affinché un processo GARCH(p,q) sia debolmente stazionario, occorre che $\sum_{l=1}^p \alpha_l + \sum_{n=1}^q \beta_n < 1$. Sotto tale ipotesi è possibile calcolare la varianza non condizionata del processo. Osservando, infatti, che

$$\mathbb{E}[y_t] = \mathbb{E}[\eta_t] = \mathbb{E}[\mathbb{E}[\eta_t | I_{t-1}]] = \mathbb{E}[0] = 0$$

si ottiene

$$\begin{aligned} \mathbb{V}(y_t) &= \mathbb{V}(\eta_t) \\ &= \mathbb{E}[\eta_t^2] \\ &= \mathbb{E}[\mathbb{E}[\eta_t^2 | I_{t-1}]] \\ &= \mathbb{E}[\sigma_t^2] \\ &= \mathbb{E}[\omega + \alpha_1 \eta_{t-1}^2 + \dots + \alpha_p \eta_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2] \\ &= \omega + \alpha_1 \mathbb{E}[y_{t-1}^2] + \dots + \alpha_p \mathbb{E}[y_{t-p}^2] + \beta_1 \mathbb{E}[\sigma_{t-1}^2] + \beta_q \mathbb{E}[\sigma_{t-q}^2] \\ &= \omega + \sum_{l=1}^p \alpha_l \mathbb{V}(y_{t-l}) + \sum_{n=1}^q \beta_n \mathbb{V}(y_{t-n}) \\ &= \omega + \mathbb{V}(y_t) \sum_{l=1}^p \alpha_l + \mathbb{V}(y_t) \sum_{n=1}^q \beta_n \end{aligned}$$

da cui

$$\mathbb{V}(y_t) = \frac{\omega}{1 - \sum_{l=1}^p \alpha_l - \sum_{n=1}^q \beta_n}. \quad (1.1)$$

Da notare che se $\sum_{l=1}^p \alpha_l + \sum_{n=1}^q \beta_n = 1$ allora si è in presenza di una radice unitaria nella parte autoregressiva del processo, giungendo al modello GARCH(p,q) integrato.

Quindi, mettendo insieme tutti gli elementi, affinché la varianza condizionata sia positiva, la varianza non condizionata esista e sia positiva e il modello sia stazionario occorre che i parametri rispettino i seguenti vincoli:

$$\omega > 0, \quad \alpha_1 \geq 0, \dots, \alpha_p \geq 0, \quad \beta_1 \geq 0, \dots, \beta_q \geq 0, \quad \sum_{l=1}^p \alpha_l + \sum_{n=1}^q \beta_n < 1$$

Per quanto riguarda la stima di tali parametri vi sono diverse vie, come il metodo dei momenti o i metodi bayesiani; tuttavia la procedura più diffusa è tramite massima verosimiglianza (Engle e Bollerslev, 1986). Durante la stima occorre tenere presente che ciascuna osservazione temporale non è indipendente da quelle precedenti. Inoltre, poiché la varianza condizionata entra nel calcolo della densità e, quindi, della funzione di verosimiglianza da massimizzare, si può notare fin da subito che l'ottimizzazione richiede l'utilizzo di algoritmi numerici, come il metodo di Newton-Raphson (Press *et al.*, 2007), Broyden-Fletcher-Goldfarb-Shanno (BFGS) (Broyden *et al.*, 1965) o altri basati sulla discesa del gradiente.

Uno dei vantaggi dei modelli GARCH consiste nel fatto che, a partire dalla specificazione più semplice appena vista, è possibile ottenere facilmente alcune sue varianti che tengano in considerazione anche altre caratteristiche dei dati (come asimmetria, *leverage effect*, stagionalità, etc.). Ad esempio, oltre al sopracitato ARMA-GARCH che consente una dinamica dei rendimenti in media di tipo ARMA, il modello GJR-GARCH permette di catturare il *leverage effect* grazie all'inserimento nella dinamica in varianza di un parametro legato alla negatività del rendimento al tempo precedente:

$$\sigma_t^2 = \omega + \alpha_1 y_{t-1}^2 + \gamma y_{t-1}^2 \mathbf{I}_{y_{t-1} < 0} + \beta_1 \sigma_{t-1}^2$$

con $\mathbf{I}_{y_{t-1} < 0}$ funzione indicatrice che vale 1 se $y_{t-1} < 0$ e 0 altrimenti. Infatti, se $\gamma > 0$ i rendimenti negativi tendono ad aumentare la volatilità più di quanto farebbero shock positivi di uguale entità. Per maggiori approfondimenti sui principali modelli della classe GARCH si rimanda a Tsay (2010).

1.3 Modelli per serie storiche finanziarie con cambiamenti di regime

1.3.1 Serie storiche finanziarie con cambiamenti di regime

La volatilità dei mercati finanziari è stata oggetto di numerosi sviluppi e applicazioni negli ultimi decenni, a livello sia teorico sia empirico. In questo contesto, la classe di modelli più ampiamente utilizzata è certamente quella dei GARCH ma tali modelli spesso indicano un'elevata persistenza della varianza condizionata, ossia possiedono valori dei parametri che si avvicinano ad un processo GARCH quasi integrato ($\sum_{l=1}^p \alpha_l + \sum_{n=1}^q \beta_n \approx 1$). Diebold (1986) e Lamoureux e Lastrapes (1990), tra gli altri, sostengono che il comportamento quasi integrato della varianza condizionata possa in realtà originare da cambiamenti strutturali nel processo della varianza che non sono però considerati dai modelli GARCH standard. A supporto di questa teoria, Mikosch e Stărică (2004) mostrano proprio che stimare un modello GARCH su un campione che presenta cambiamenti strutturali nella varianza condizionata porta effettivamente ad un GARCH integrato. Questi risultati indicano chiaramente che se la forma della varianza condizionata viene mantenuta fissa durante l'intero periodo di osservazione allora si incorre in una potenziale fonte di specificazione errata, dove le stime di un modello GARCH risentono di una distorsione verso l'alto nel parametro di persistenza.

Pertanto, in tali situazioni possono risultare più appropriati per modellare la volatilità modelli i cui parametri possono cambiare nel tempo, come ad esempio quelli basati sull'idea dei cambiamenti di regime. Tra questi, un possibile modello è il Markov-*switching* GARCH (MS-GARCH) proposto da Bauwens *et al.* (2010) in cui la media e la varianza condizionate passano nel tempo da un processo GARCH ad un altro e il cambiamento è governato da una catena di Markov latente. Giustificazione all'utilizzo di tali modelli è data dal fatto che i mercati finanziari tendono a cambiare il loro comportamento in modo brusco, solitamente a causa di nuove regolamentazioni, cambi di politica, fattori sociali ed economici (come la caduta delle Torri Gemelle, il martedì nero della Borsa di New York, etc.). Il cambiamento spesso persiste per diversi periodi dopo che esso è avvenuto ma non è detto che il passaggio sia effettivamente drastico. Per questo occorre allontanarsi dai modelli

a cambiamento di regime di Markov per andare verso modelli *smooth-transition* che consentano il passaggio per regimi intermedi. In tal caso, l'utilizzo del termine "regime" ha un significato diverso da quello di un modello di Markov: la funzione di transizione è continua, non "a soglia", e si preferisce parlare non tanto di un alternarsi quanto di un susseguirsi di regimi in cui la probabilità di passare da uno all'altro è pari a uno. Inoltre, il passaggio tra i regimi non è dettato da una catena di Markov latente ma da una variabile di transizione osservabile che appartiene alla storia del processo stesso.

Come approfondito da González-Rivera (1998), un esempio di applicazione di modelli *smooth-transition* riguarda le serie storiche dei tassi di cambio. Se ai tassi di cambio viene stimato un semplice modello a soglia per catturare l'asimmetria in varianza, l'evidenza di tale caratteristica risulta piuttosto debole, quasi da definire il processo simmetrico. Tuttavia, se viene stimato un modello *smooth-transition* l'evidenza di asimmetria diviene più forte, ad indicare che la transizione da un regime a bassa volatilità ad uno ad alta volatilità è graduale, rifiutando la specificazione a soglia in cui la transizione è invece netta.

Ai fini di questo lavoro verrà presentato il modello *Smooth-Transition GARCH* (ST-GARCH). In particolare, senza perdita di generalità si considera il caso con due soli regimi di tipo GARCH(1,1) avente parametri in varianza condizionata diversi. I ragionamenti sono poi immediatamente estendibili a qualsiasi modello della classe GARCH.

1.3.2 I modelli ST-GARCH

Data la serie storica $\{y_t\}_{t=1,\dots,T}$ dei rendimenti di lunghezza T e l'informazione passata I_{t-1} disponibile fino al tempo $t - 1$, il modello ST-GARCH formato da due regimi GARCH(1,1) diversi suppone che:

$$\begin{aligned}
 y_t &= \mu_t + \eta_t = 0 + \sigma_t \epsilon_t && \text{con } \epsilon_t \sim N.i.d.(0, 1) && (1.2) \\
 y_t | I_{t-1} &\sim N(0, \sigma_t^2) \\
 \sigma_t^2 &= g(z_t, \delta, \lambda) [\omega_1 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2] + \\
 &\quad + (1 - g(z_t, \delta, \lambda)) [\omega_2 + \alpha_2 y_{t-1}^2 + \beta_2 \sigma_{t-1}^2]
 \end{aligned}$$

con $\omega_1 > 0$, $\omega_2 > 0$, $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, $\beta_1 \geq 0$, $\beta_2 \geq 0$. La funzione di transizione $g(t, \lambda)$ può assumere diverse forme (logistica classica e di secondo ordine, esponenziale, etc.). In questo lavoro ci si concentra sulla funzione logistica

$$g(z_t, \delta, \lambda) = g(t, \delta, \lambda) = (1 + \exp\{-\delta(t - \lambda)\})^{-1}$$

con parametro di liscio $\delta \geq 0$ ma le stesse considerazioni possono essere facilmente estese anche a scelte diverse. Come si nota la transizione non dipende da una variabile temporale esogena ma dall'istante temporale t corrente confrontato con la soglia $0 \leq \lambda \leq T$, la quale come detto prima potrebbe rappresentare una data nei pressi di un cambiamento politico, sociale o economico. Da notare che, con le stesse considerazioni viste per il modello GARCH, si sono assunte $\mu_t = 0 \forall t$ e innovazioni normalmente distribuite.

Riassumendo, il modello suppone che fino al tempo λ la serie storica finanziaria segua un GARCH(1,1) con parametri $(\omega_1, \alpha_1, \beta_1)$ mentre da λ in poi un GARCH(1,1) con parametri $(\omega_2, \alpha_2, \beta_2)$; il passaggio tra i due modelli, però, non avviene secondo una soglia netta ma con velocità che dipende dal parametro di liscio δ della funzione di transizione logistica.

Per quanto riguarda la stima dei parametri $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2)$, della soglia λ e del parametro di liscio δ , è possibile procedere mediante massima verosimiglianza (González-Rivera, 1998). Anche in questo caso, data la forma della varianza condizionata, viene però richiesto l'utilizzo di algoritmi di ottimizzazione numerica come Newton-Raphson, Broyden-Fletcher-Goldfarb-Shanno (BFGS) o altri basati sulla discesa del gradiente.

Capitolo 2

Modelli statistici per l'identificazione di gruppi

2.1 *Clustering* basato su modelli a mistura finita

2.1.1 Introduzione ai metodi di *clustering*

Lo scopo principale del *clustering* è quello di ripartire un insieme di dati in gruppi tali che ciascuno di essi contenga osservazioni omogenee tra loro ma differenti da quelle degli altri gruppi, il tutto sulla base delle caratteristiche dei dati stessi.

L'approccio al *clustering* più classico e diffuso si basa sulla valutazione delle similarità tra le unità, e comprende metodi gerarchici (agglomerativi e divisivi, Kaufman e Rousseeuw (1990)) e metodi di partizione (come il *k-means* (MacQueen, 1967)). L'idea alla base di queste procedure euristiche è quella di creare i gruppi minimizzando la distanza tra osservazioni dello stesso gruppo e massimizzando la distanza tra osservazioni di gruppi differenti. A tale scopo, questi approcci necessitano il calcolo di una matrice di dissimilarità il cui numero di elementi cresce con il quadrato della dimensione del campione. Ciò comporta significative limitazioni computazionali all'aumentare della dimensione del campione, sia in termini di elaborazione che di memoria. Inoltre, per quanto possano essere altamente intuitivi, il problema fondamentale è che non si tratta di metodi statistici rigorosi e per questo la validità dei risultati ottenuti è affetta da un certo grado di arbitrarietà, così come non vi sono metodi oggettivi per la scelta del numero di gruppi. A cau-

sa di queste criticità, si sono diffusi negli ultimi decenni approcci più rigorosi basati sulla teoria dell'inferenza.

L'approccio parametrico, o *model-based*, consiste nell'assumere che i dati osservati siano stati generati da un modello probabilistico, noto a meno di un certo numero di parametri, su cui verte l'inferenza. Ogni *cluster* è caratterizzato da uno specifico vettore di parametri e definisce una componente di un modello mistura di riferimento.

Basare l'analisi dei *cluster* su un modello di probabilità ha diversi vantaggi. Innanzitutto, trattandosi di una metodologia statistica standard, consente di effettuare inferenze in modo rigoroso. Inoltre, fornisce una strategia accurata per scegliere il numero dei gruppi, riducendosi ad un semplice problema di selezione del modello in cui si cerca il giusto compromesso tra numero di *cluster* e complessità del modello stesso. Infine, è possibile valutare l'incertezza sul *clustering* e tener conto, mediante un'opportuna specificazione del modello, delle caratteristiche dei dati o, ad esempio, della presenza di valori anomali. In questo capitolo si presenterà una panoramica generale su tale approccio, facendo riferimento principalmente agli aspetti più rilevanti ai fini degli sviluppi di questo elaborato. Si rimanda a McLachlan e Peel (2000) e Bouveyron *et al.* (2019) per maggiori approfondimenti.

2.1.2 Specificazione del modello

Si suppone che l'insieme di dati da analizzare sia costituito da N osservazioni multivariate y_1, y_2, \dots, y_N con $y_i = (y_{i1}, \dots, y_{iT})$, $i = 1, \dots, N$. Tipicamente si assume che tali osservazioni siano copie indipendenti e identicamente distribuite di una variabile casuale multivariata Y distribuita secondo un *modello a mistura finita*

$$p(y) = \sum_{k=1}^K \pi_k p(y|\theta_k) \quad (2.1)$$

ossia una combinazione lineare di K ($< N$) distribuzioni di probabilità o funzioni di densità $p(\cdot|\theta_k)$ definite da uno specifico vettore di parametri θ_k . Il generico peso $\pi_k \geq 0$ viene chiamato *proporzione della mistura* e corrisponde alla probabilità (a priori) che un'osservazione sia stata generata dalla k -esima componente della mistura, con $k = 1, \dots, K$ e $\sum_{k=1}^K \pi_k = 1$.

Ai fini del raggruppamento, ogni *cluster* è descritto da una specifica componente della mistura e la sua identificazione è riconducibile alla stima dei parametri da cui è caratterizzato.

Considerando il numero di componenti K noto, i parametri del modello (2.1) sono $\theta = (\theta_1, \dots, \theta_K)$ e $\pi = (\pi_1, \dots, \pi_K)$ e la stima può essere eseguita tramite massima verosimiglianza.

Date le osservazioni y_1, \dots, y_N , la *log*-verosimiglianza del modello mistura è

$$\begin{aligned}
 l(\theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K | y_1, \dots, y_N) &= \\
 &= \log p(y_1, \dots, y_N | \theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K) \\
 &= \sum_{i=1}^N \log p(y_i | \theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K) \\
 &= \sum_{i=1}^N \log \left(\sum_{k=1}^K \pi_k p(y_i | \theta_k) \right) \tag{2.2}
 \end{aligned}$$

Senza perdita di generalità, la specificazione consente di trattare qualunque tipo di dato, da continuo a categoriale, e di struttura, da un semplice vettore a una regressione o a una serie storica; il modello è, quindi, molto flessibile ma la diversa specificazione può avere ripercussioni sulla facilità di stima, come si vedrà anche nel seguito.

2.1.3 Stima di massima verosimiglianza mediante algoritmo EM

Osservando la forma della funzione di verosimiglianza (2.2) e della sua derivata prima

$$\frac{\partial l(\theta)}{\partial \theta} = \sum_{i=1}^N \frac{1}{\sum_{k=1}^K \pi_k p(y_i | \theta_k)} \sum_{k=1}^K \pi_k \frac{\partial p(y_i | \theta_k)}{\partial \theta}$$

essa è solitamente complicata e multi-modale, prestandosi raramente a soluzioni analitiche in forma chiusa. Per questo motivo la massimizzazione della verosimiglianza avviene numericamente mediante l'uso dell'algoritmo *Expectation-Maximization* (EM, Dempster *et al.* (1977)) o di qualche sua estensione, un algoritmo versatile per determinare stime di massima verosimiglianza in presenza di dati incompleti.

Ai fini del problema di *clustering* descritto dal modello (2.1), i dati osservati possono considerarsi incompleti in quanto, per ciascuna unità y_i non è noto il gruppo di appartenenza, definito dal vettore $z_i =$

(z_{i1}, \dots, z_{iK}) , con

$$z_{ik} = \begin{cases} 1 & \text{se } y_i \in \text{componente } k \\ 0 & \text{altrimenti.} \end{cases}$$

Si assume, quindi, che z_1, \dots, z_N siano indipendenti e identicamente distribuite secondo una multinomiale con una sola estrazione e con K categorie aventi probabilità associate (π_1, \dots, π_K) :

$$z_i = (z_{i1}, \dots, z_{iK}) \sim \text{Bin}_K((\pi_1, \dots, \pi_K), 1).$$

Data la conoscenza del vettore z_i , la densità dell'osservazione y_i è data da

$$p(y_i | \theta, z_i) = \prod_{k=1}^K p(y_i | \theta_k)^{z_{ik}},$$

da cui deriva che la *log*-verosimiglianza dei dati completi (y_i, z_i) , $i = 1, \dots, N$, è

$$\begin{aligned} l_C(\theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K | y_1, \dots, y_N, z_1, \dots, z_N) &= \log p(y_1, \dots, y_N, z_1, \dots, z_N | \theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K) \\ &= \log \prod_{i=1}^N p(y_i, z_i | \theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K) \\ &= \sum_{i=1}^N \log p(z_i | \pi_1, \dots, \pi_K) p(y_i | z_i, \theta_1, \dots, \theta_K, \pi_1, \dots, \pi_K) \\ &= \sum_{i=1}^N \log \left(\pi_1^{z_{i1}} \dots \pi_K^{z_{iK}} \prod_{k=1}^K p(y_i | z_i, \theta_k)^{z_{ik}} \right) \\ &= \sum_{i=1}^N \log \prod_{k=1}^K \pi_k^{z_{ik}} p(y_i | z_i, \theta_k)^{z_{ik}} \\ &= \sum_{i=1}^N \sum_{k=1}^K \log \left(\pi_k p(y_i | z_i, \theta_k) \right)^{z_{ik}} \\ &= \sum_{i=1}^N \sum_{k=1}^K z_{ik} \log \left(\pi_k p(y_i | z_i, \theta_k) \right) \end{aligned}$$

Data la struttura incompleta dei dati, l'idea alla base dell'algoritmo EM è quella di massimizzare, in luogo della verosimiglianza, il suo valore

atteso al variare degli z_i . A questo scopo, l'algoritmo EM alterna iterativamente due *step*. Il primo, il passo di *Expectation*, consiste nel calcolo del valore atteso della *log-verosimiglianza* completa condizionatamente ai dati osservati e alla stima corrente dei parametri. Nel secondo passo, invece, vengono determinati i parametri che massimizzano il valore atteso calcolato nello *step* precedente, da cui il nome di *Maximization step*.

Indicando con r la generica iterazione dell'algoritmo, i due passi in dettaglio sono:

- *E-step*: fissati $\theta = \theta^{(r)}$ e $\pi = \pi^{(r)}$ si determina il valore atteso della *log-verosimiglianza* completa rispetto ai vettori latenti z_i :

$$\begin{aligned}
\mathbb{E}_z[l_C(\theta, \pi | y_1, \dots, y_N, z_1, \dots, z_N)] &= \\
&= \mathbb{E}_z \left[\sum_{i=1}^N \sum_{k=1}^K z_{ik} \log (\pi_k^{(r)} p(y_i | \theta_k^{(r)})) | y_1, \dots, y_N \right] \\
&= \sum_{i=1}^N \sum_{k=1}^K \log (\pi_k^{(r)} p(y_i | \theta_k^{(r)})) \mathbb{E}_z [z_{ik} | y_1, \dots, y_N] \\
&= \sum_{i=1}^N \sum_{k=1}^K \frac{\pi_k^{(r)} p(y_i | \theta_k^{(r)})}{\sum_{k=1}^K \pi_k^{(r)} p(y_i | \theta_k^{(r)})} \log (\pi_k^{(r)} p(y_i | \theta_k^{(r)}))
\end{aligned}$$

dove nell'ultima uguaglianza si è utilizzato il risultato:

$$\begin{aligned}
\mathbb{E}_z [z_{ik} | y_i] &= \mathbb{P}(z_{ik} = 1 | y_i) \\
&= \frac{\mathbb{P}(z_{ik} = 1, y_i)}{p(y_i)} \\
&= \frac{\mathbb{P}(z_{ik} = 1) p(y_i | z_{ik} = 1)}{p(y_i)} \\
&= \frac{\pi_k p(y_i | \theta_k)}{\sum_{k=1}^K \pi_k p(y_i | \theta_k)} = \tau_{ik}
\end{aligned}$$

Da notare che τ_{ik} corrisponde alla probabilità a posteriori che l'osservazione y_i sia stata generata dalla componente k della mistura e costituisce la parte centrale del calcolo nell'*E-step*.

- *M-step*: fissati $\tau_{ik} = \tau_{ik}^{(r)}$, si stimano $\hat{\theta}^{(r+1)}$ e $\hat{\pi}^{(r+1)}$ come quei

parametri che massimizzano il valore atteso calcolato nell'E-step:

$$\begin{aligned} (\hat{\theta}^{(r+1)}, \hat{\pi}^{(r+1)}) &= \arg \max_{\theta, \pi} \mathbb{E}_z [l_C(\theta, \pi | y_1, \dots, y_N, z_1, \dots, z_N)] \\ &= \arg \max_{\theta, \pi} \sum_{i=1}^N \sum_{k=1}^K \tau_{ik}^{(r)} \log(\pi_k p(y_i | \theta_k)) \end{aligned}$$

In realtà la massimizzazione deve rispettare il vincolo $\sum_{k=1}^K \pi_k = 1$, quindi la funzione obiettivo da massimizzare in questo passo diventa:

$$\sum_{i=1}^N \sum_{k=1}^K \tau_{ik}^{(r)} \log(\pi_k p(y_i | \theta_k)) + \lambda \left(\sum_{k=1}^K \pi_k - 1 \right)$$

Partendo da una soluzione iniziale per il valore dei parametri $\theta^{(0)}$ e $\pi^{(0)}$, sotto condizioni di regolarità piuttosto deboli, l'EM itera i due passi fino a convergere ad un massimo locale della verosimiglianza. Data la natura solitamente non convessa della *log*-verosimiglianza di un modello mistura, a inizializzazioni diverse possono corrispondere risultati diversi in quanto è possibile che l'algoritmo converga ad un massimo locale e non a quello globale. Non vi sono strategie uniformemente più opportune per scegliere $\theta^{(0)}$ e $\pi^{(0)}$: possono essere scelti anche casualmente ma pratica comune è quella di utilizzare metodi gerarchici o di partizione.

La regola di arresto per la convergenza dell'algoritmo viene solitamente impostata sulla verosimiglianza o sui parametri del modello. Ad esempio, l'EM continua fintanto che la variazione della *log*-verosimiglianza o dei parametri da un'iterazione all'altra non è più rilevante.

Una volta ottenute le stime di massima verosimiglianza dei parametri del modello, ciascuna osservazione y_i del campione viene attribuita al gruppo per cui è massima la probabilità a posteriori (*maximum a posteriori rule*, MAP), ossia:

$$y_i \in \text{componente } h \quad \text{con } h \text{ tale che } \hat{\tau}_{ih} = \max_k \hat{\tau}_{ik}$$

Fino ad ora si è considerato noto il numero di componenti K . In realtà, anche questo fa parte di una delle problematiche da risolvere nell'analisi dei gruppi. Grazie al carattere inferenziale che ha la procedura, la scelta del numero di *cluster* presenti nei dati si riduce ad un classico problema di selezione del modello. Tipicamente viene affrontato mediante l'ottimizzazione di criteri quali il *Bayesian Information*

Criterion (BIC) o l'*Integrated Completed Likelihood* (ICL) al variare di K , in quanto per la loro particolare costruzione consentono di trovare il giusto compromesso tra l'adattamento (minimizzazione della funzione di verosimiglianza) e la parsimonia (penalizzazione di forme troppo complesse) del modello.

L'algoritmo EM non è l'unica alternativa possibile per la stima di massima verosimiglianza dei modelli a mistura finita; in particolari situazioni, che dipendono soprattutto dalle caratteristiche dei dati da analizzare, è possibile ricorrere ad alcune varianti, come per esempio *Classification-EM* o *Stochastic-EM* che inseriscono un passo rispettivamente di classificazione o stocastico tra *E-step* e *M-step*. Tali algoritmi verranno maggiormente discussi nel Capitolo 3 in quanto utilizzati nel caso in esame.

2.2 *Co-clustering* basato su modelli mistura

2.2.1 Introduzione al *co-clustering*

Nonostante la comprovata efficacia in molte applicazioni, quando ciascuna unità statistica viene osservata con riferimento ad un elevato numero di variabili, le tecniche di clustering possono incontrare alcune limitazioni, portando a risultati potenzialmente fuorvianti. Infatti, identificare soggetti simili rispetto a un numero elevato di caratteristiche è, anche concettualmente, un obiettivo difficilmente raggiungibile, e aspettarsi che i soggetti all'interno di uno stesso cluster siano simili rispetto a tutte le variabili osservate è spesso poco realistico. Piuttosto, accade spesso che i soggetti si raggruppino insieme solo rispetto a determinate caratteristiche o che le variabili forniscano informazioni simili solo con riferimento a sottoinsiemi di soggetti. La formalizzazione di questa idea dà origine a quella classe di metodi, noti come *bi-clustering* e *co-clustering*, che si rivelano adatti in presenza di dati eterogenei ad alta dimensionalità dove anche le relazioni tra le variabili sono di interesse (Govaert e Nadif, 2013). I metodi di *bi-clustering* mirano a trovare sottoinsiemi di osservazioni descritte da sottoinsiemi, possibilmente sovrapposti, di variabili. Il *co-clustering*, invece, consente il raggruppamento simultaneo sia delle righe che delle colonne di una matrice di dati.

Lo sviluppo dei metodi di *co-clustering* ha trovato applicazione recente in numerosi campi, come la bioinformatica, per raggruppare con-

giuntamente geni e condizioni sperimentali (Hanczar e Nadif, 2010); il *marketing*, per raggruppare utenti e articoli nei sistemi di raccomandazione (Deodhar e Ghosh, 2010); il *text mining*, per raggruppare simultaneamente termini e documenti (Priam *et al.*, 2016).

Analogamente al contesto di *clustering* più tradizionale, anche nell'ambito del *co-clustering* sono stati sviluppati metodi più euristici basati sulla distanza e metodi formalmente più rigorosi, basati sulla densità. Nel secondo caso, affrontato in questo elaborato, il concetto di *cluster* è nuovamente collegato a determinate caratteristiche della distribuzione di probabilità che si presume sottostante ai dati, consentendo la definizione di un obiettivo statistico preciso da perseguire in senso inferenziale classico. Questo permette la valutazione e il confronto di configurazioni alternative di *co-clustering*. Un aspetto degno di nota, che rende particolarmente interessante l'uso di metodi di *co-clustering* su dati ad alta dimensionalità, è che per identificare un *clustering* simultaneo di righe e colonne, è la singola cella della matrice di dati che viene modellata, piuttosto che la riga (o la colonna). Ciò rappresenta una soluzione utile al problema della maledizione della dimensionalità, spesso sofferto dai metodi basati sulla densità, e una caratteristica che consente di catturare dipendenze e interazioni complesse all'interno dei dati.

Sebbene il *co-clustering* non sia direttamente esplicitabile quale obiettivo di questo elaborato, si presenta in questo paragrafo una breve introduzione al modello a blocchi latenti (*Latent Block Model*, LBM, Govaert e Nadif (2013)), il più diffuso tra le formulazioni parametriche del problema, in quanto presenta delle caratteristiche che lo rendono adatto ad essere esteso, con opportuni aggiustamenti, al problema in esame di identificare gruppi di serie storiche in presenza di cambiamenti di regime.

2.2.2 Il modello a blocchi latenti

L'idea di fondo del modello a blocchi latenti è la stessa di quella vista per il *clustering* basato su modello mistura, con la differenza che vi sono dei gruppi contemporaneamente sia sulle righe che sulle colonne della matrice. Per questo, mentre nel modello a mistura finita si assume che ogni osservazione appartenga ad un *cluster* descritto da una variabile casuale latente multinomiale, nel LBM la struttura di gruppo è descritta da una coppia di variabili latenti, che definiscono l'appartenenza ad un gruppo di riga e di colonna. In altre parole, assume che ogni cella (i, c) appartenga congiuntamente ad un *co-cluster* (k, j) di riga e di colonna, chiamato blocco (da cui il nome di modello a blocchi latenti).

Come per il *clustering*, il numero di componenti di riga K e di colonna M può essere noto o ignoto e per identificare i gruppi di appartenenza di una cella occorre definire due vettori latenti, rispettivamente K - e J -dimensionali. In particolare, mettendo insieme i vettori latenti di ciascuna riga e poi colonna si ottiene che:

- la partizione di riga in K *cluster* viene rappresentata dalla matrice di classificazione latente

$$\mathbf{z} = (z_{ik}, i = 1, \dots, N, k = 1, \dots, K)$$

$$\text{con } z_{ik} = \begin{cases} 1 & \text{se } y_i \in \text{componente } k \\ 0 & \text{altrimenti} \end{cases}$$

con probabilità associate

$$\mathbb{P}(z_{ik} = 1) = \pi_k \geq 0 \quad \text{con } \sum_{k=1}^K \pi_k = 1;$$

- la partizione di colonna in J *cluster* viene rappresentata dalla matrice di classificazione latente

$$\mathbf{w} = (w_{cj}, c = 1, \dots, C, j = 1, \dots, J)$$

$$\text{con } w_{cj} = \begin{cases} 1 & \text{se } y_c \in \text{componente } j \\ 0 & \text{altrimenti} \end{cases}$$

con probabilità associate

$$\mathbb{P}(w_{cj} = 1) = \rho_j \geq 0 \quad \text{con } \sum_{j=1}^J \rho_j = 1.$$

Tornando, quindi, alla notazione vettoriale vista nel paragrafo precedente, è possibile esprimere tali risultati in modo più compatto, notando che i vettori riga z_1, \dots, z_N e w_1, \dots, w_C sono nuovamente indipendenti e identicamente distribuiti secondo due multinomiali con una sola estrazione e con rispettivamente K e J categorie aventi probabilità associate (π_1, \dots, π_K) e (ρ_1, \dots, ρ_J) :

$$z_i = (z_{i1}, \dots, z_{iK}) \sim \text{Bin}_K((\pi_1, \dots, \pi_K), 1)$$

$$w_c = (w_{c1}, \dots, w_{cJ}) \sim \text{Bin}_J((\rho_1, \dots, \rho_J), 1).$$

Come già accennato in precedenza, il modello a blocchi latenti parte dall'assunzione che i vettori latenti di riga e di colonna siano variabili casuali indipendenti tra loro, da cui si ottiene che:

$$p(\mathbf{z}, \mathbf{w}) = p(\mathbf{z})p(\mathbf{w})$$

con $p(\mathbf{z}) = \prod_{i=1}^N p(z_i)$ e $p(\mathbf{w}) = \prod_{c=1}^C p(w_c)$.

Inoltre, condizionatamente a \mathbf{z} e \mathbf{w} , le variabili casuali y_{ic} sono tutte indipendenti con densità $p(y_{ic}|\theta_{kj})$, dove θ_{kj} è il parametro della distribuzione del blocco (k, j) .

Sulla base delle precedenti premesse, il modello a blocchi latenti per la matrice dei dati \mathbf{y} è così formulato:

$$\begin{aligned}
p(\mathbf{y}) &= \prod_{i=1}^N \prod_{c=1}^C p(y_{ic}) \\
&= \prod_{i,c} p(y_{ic} \cap (z_{i1} = 1 \cup \dots \cup z_{iK} = 1) \cap \\
&\hspace{15em} \cap (w_{c1} = 1 \cup \dots \cup w_{cJ} = 1)) \\
&= \prod_{i,c} \sum_{k=1}^K p(y_{ic} \cap z_{ik} = 1 \cap (w_{c1} = 1 \cup \dots \cup w_{cJ} = 1)) \\
&= \prod_{i,c} \sum_{k=1}^K \sum_{j=1}^J p(y_{ic} \cap z_{ik} = 1 \cap w_{cj} = 1) \\
&= \prod_{i,c} \sum_{k,j} p(z_{ik} = 1 \cap w_{cj} = 1) p(y_{ic} | z_{ik} = 1 \cap w_{cj} = 1) \\
&= \prod_{i,c} \sum_{k,j} \pi_k \rho_j p(y_{ic} | \theta_{kj}) \tag{2.3}
\end{aligned}$$

Il modello presentato è molto flessibile in quanto consente di trattare qualunque tipo di distribuzione, con l'unica accortezza che a specificazioni diverse possono susseguire difficoltà computazionali diverse. Inoltre, mentre nel modello mistura classico si definisce un modello per ogni gruppo, l'LBM ha l'ulteriore vantaggio di essere molto più parsimonioso perché richiede la stima di un unico parametro per ogni blocco.

Dunque, considerando il numero di componenti K e J noto, i parametri del LBM (2.3) sono $\theta_{11}, \dots, \theta_{1J}, \dots, \theta_{KJ}, \pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J$ e la stima può essere eseguita tramite massima verosimiglianza.

2.2.3 Stima di massima verosimiglianza mediante algoritmo LBEM

Per la stima dei parametri, data la matrice dei dati \mathbf{y} la *log*-verosimiglianza del modello (2.3) è:

$$\begin{aligned}
 l(\theta_{11}, \dots, \theta_{1J}, \dots, \theta_{KJ}, \pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J | \mathbf{y}) &= \\
 &= \log p(\mathbf{y} | \theta_{11}, \dots, \theta_{KJ}, \pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J) \\
 &= \sum_{i,c} \log p(y_{ic} | \theta_{11}, \dots, \theta_{KJ}, \pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J) \\
 &= \sum_{i,c} \log \sum_{k,j} \pi_k \rho_j p(y_{ic} | z_{ik} = 1, w_{cj} = 1, \theta_{kj})
 \end{aligned}$$

e, dato $\Theta = (\theta_{11}, \dots, \theta_{1J}, \dots, \theta_{KJ}, \pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J)$, la derivata prima è:

$$\begin{aligned}
 \frac{\partial l(\Theta)}{\partial \Theta} &= \sum_{i,c} \frac{1}{\sum_{k,j} \pi_k \rho_j p(y_{ic} | z_{ik} = 1, w_{cj} = 1, \theta_{kj})} \cdot \\
 &\quad \cdot \sum_{k,j} \pi_k \rho_j \frac{\partial p(y_{ic} | z_{ik} = 1, w_{cj} = 1, \theta_{kj})}{\partial \Theta}
 \end{aligned}$$

Uguagliandola a zero, la derivata prima è solitamente impossibile da risolvere analiticamente perché i parametri non sono esplicitabili. Per questo motivo la massimizzazione della verosimiglianza avviene, come nel *clustering*, numericamente mediante l'uso dell'algoritmo EM o di qualche sua estensione definita per tener conto della doppia struttura latente (\mathbf{z}, \mathbf{w}) .

La *log*-verosimiglianza completa in cui si assumono osservati sia la

matrice dei dati \mathbf{y} sia le matrici di classificazione latenti \mathbf{z} e \mathbf{w} è, quindi:

$$\begin{aligned}
l_C(\theta_{11}, \dots, \theta_{1J}, \dots, \theta_{KJ}, \pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J | \mathbf{y}, \mathbf{z}, \mathbf{w}) \\
&= \log p(\mathbf{y}, \mathbf{z}, \mathbf{w} | \theta_{11}, \dots, \theta_{KJ}, \pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J) \\
&= \log p(\mathbf{z}, \mathbf{w} | \pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J) p(\mathbf{y} | \mathbf{z}, \mathbf{w}, \theta_{11}, \dots, \theta_{KJ}) \\
&= \log p(\mathbf{z} | \pi_1, \dots, \pi_K) p(\mathbf{w} | \rho_1, \dots, \rho_J) p(\mathbf{y} | \mathbf{z}, \mathbf{w}, \theta_{11}, \dots, \theta_{KJ}) \\
&= \log \left(\prod_{i=1}^N p(z_i) \prod_{c=1}^C p(w_c) p(\mathbf{y} | \mathbf{z}, \mathbf{w}, \theta_{11}, \dots, \theta_{KJ}) \right) \\
&= \sum_{i,c} \log \left(\prod_{k=1}^K \pi_k^{z_{ik}} \prod_{j=1}^J \rho_j^{w_{cj}} \prod_{k=1}^K \prod_{j=1}^J p(y_{ic} | z_i, w_c, \theta_{kj})^{z_{ik} w_{cj}} \right) \\
&= \sum_{i,c,k,j} \log \left((\pi_k \rho_j)^{z_{ik} w_{cj}} p(y_{ic} | z_i, w_c, \theta_{kj})^{z_{ik} w_{cj}} \right) \\
&= \sum_{i,c,k,j} \log \left(\pi_k \rho_j p(y_{ic} | z_i, w_c, \theta_{kj}) \right)^{z_{ik} w_{cj}} \\
&= \sum_{i,c,k,j} z_{ik} w_{cj} \log \left(\pi_k \rho_j p(y_{ic} | z_i, w_c, \theta_{kj}) \right) \\
&= \sum_{i,k} z_{ik} \log \pi_k + \sum_{c,j} w_{cj} \log \rho_j + \sum_{i,c,k,j} z_{ik} w_{cj} \log p(y_{ic} | \theta_{kj})
\end{aligned}$$

in cui il primo termine dipende dalle proporzioni dei gruppi di riga, il secondo da quelle dei gruppi di colonna e il terzo dalle densità di ogni blocco.

A causa della presenza di due quantità latenti, l'inferenza non risulta così semplice come nel caso del modello mistura classico. Infatti, provando a scrivere il valore atteso della *log*-verosimiglianza completa condizionatamente ai dati osservati e alla stima corrente dei parametri ($\theta = (\theta_{11}, \dots, \theta_{KJ}) = \theta^{(r)}$, $\pi = (\pi_1, \dots, \pi_K) = \pi^{(r)}$, $\rho = (\rho_1, \dots, \rho_J) = \rho^{(r)}$), si ottiene

$$\begin{aligned}
\mathbb{E}_{\mathbf{z}, \mathbf{w}} [l_C(\theta, \pi, \rho | \mathbf{y}, \mathbf{z}, \mathbf{w})] &= \\
&= \mathbb{E}_{\mathbf{z}, \mathbf{w}} \left[\sum_{i,c,k,j} z_{ik} w_{cj} \log \left(\pi_k^{(r)} \rho_j^{(r)} p(y_{ic} | \theta_{kj}^{(r)}) \right) | \mathbf{y} \right] \\
&= \sum_{i,c,k,j} \log \left(\pi_k^{(r)} \rho_j^{(r)} p(y_{ic} | \theta_{kj}^{(r)}) \right) \mathbb{E}_{\mathbf{z}, \mathbf{w}} [z_{ik} w_{cj} | \mathbf{y}]
\end{aligned}$$

difficile da calcolare in quanto $\mathbb{E}_{\mathbf{z}, \mathbf{w}} [z_{ik} w_{cj} | \mathbf{y}] = \mathbb{P}(z_{ik} w_{cj} = 1 | \mathbf{y})$ richiede il calcolo della distribuzione congiunta condizionata delle latenti. Infatti,

sebbene le etichette latenti \mathbf{z} e \mathbf{w} siano indipendenti, non lo sono condizionatamente ai dati \mathbf{y} e quindi l'*E-step* richiede il calcolo di termini che non possono essere fattorizzati come in una mistura standard.

Per ovviare a questo problema vengono utilizzate altre forme dell'algoritmo EM, come *Variational-EM* (VEM), *Classification-EM* (CEM) e *Stochastic-EM* (SEM). Per approfondimenti su questi algoritmi si rimanda a Govaert e Nadif (2013), mentre nel Capitolo 3 verranno approfonditi CEM e SEM-*Gibbs* per la formulazione oggetto di studio.

Capitolo 3

Raggruppamento di serie storiche finanziarie in presenza di cambiamenti di regime

3.1 Un modello per l'identificazione di gruppi di serie storiche con cambiamenti di regime

Obiettivo ultimo di questo lavoro è la formulazione di un modello statistico per l'identificazione di gruppi di serie storiche finanziarie in presenza di cambiamenti di regime. Coerentemente con la definizione di *cluster* considerata dall'approccio *model-based*, e seguendo la notazione introdotta nel Capitolo 1, si assumerà che ogni regime sia descritto da un modello autoregressivo condizionatamente eteroschedastico e, in particolare, da un modello GARCH(1,1).

Il modello che si propone, estendibile senza troppe difficoltà al caso in cui ciascuna serie sia descritta da M regimi, si concentra sulla più semplice ipotesi che $M = 2$ e si assume che la transizione tra il primo e il secondo regime sia graduale, secondo un modello ST-GARCH.

Formalmente, si immagini di osservare N serie storiche di rendimenti $\{y_{it}\}_{t=1,\dots,T}$, $i = 1, \dots, N$. Seguendo una logica analoga a quella introdotta nei modelli di *co-clustering*, una doppia struttura latente (z_i, w_i) , $i = 1, \dots, N$, definisce l'appartenenza di ciascuna osservazione y_i ad un *cluster* z_i per il primo regime ed un *cluster* w_i per il secondo regime.

3.2 Specificazione del modello

$$\begin{aligned}
y_{it}|z_{ik}, w_{ij} &= \sigma_{ikjt}\epsilon_{it} \quad \text{con } \epsilon_{it} \sim N.i.d.(0, 1) \\
y_{it}|I_{it-1}, z_{ik}, w_{ij} &\sim N(0, \sigma_{ikjt}^2) \\
\sigma_{ikjt}^2 &= g(t, \delta, \lambda)[\omega_k + \alpha_k y_{it-1} + \beta_k \sigma_{ikjt-1}^2] + \\
&\quad + (1 - g(t, \delta, \lambda))[\omega_j + \alpha_j y_{it-1} + \beta_j \sigma_{ikjt-1}^2]
\end{aligned}$$

con $\omega_k > 0$ $\alpha_k \geq 0$ $\beta_k \geq 0$, $k = 1, \dots, K$ e $\omega_j > 0$ $\alpha_j \geq 0$ $\beta_j \geq 0$, $j = 1, \dots, J$ e funzione di transizione logistica

$$g(t, \delta, \lambda) = (1 - \exp\{-\delta(t - \lambda)\})^{-1}$$

con parametro di lisciamiento $\delta \geq 0$ e soglia $0 < \lambda < T$. Da notare che per semplicità si sta assumendo lo stesso parametro di lisciamiento $\delta_i = \delta \forall i = 1, \dots, N$ e, per quanto riguarda la soglia, si assume $\lambda_i = \lambda \forall i = 1, \dots, N$ perché interessati alla reazione ad uno *shock* comune. Si specifica che gli stessi ragionamenti sono immediatamente estendibili anche ad altri modelli della classe GARCH e ad altre specificazioni della funzione di transizione, così come al caso in cui il parametro di lisciamiento e la soglia cambiano tra le serie.

Dato il precedente modello e posto $\theta_{kj} = (\omega_k, \alpha_k, \beta_k, \omega_j, \alpha_j, \beta_j)$, la funzione di densità $p(\cdot|\theta_{kj}, \delta, \lambda)$ per la serie y_i è:

$$\begin{aligned}
p(y_i|\theta_{kj}, \delta, \lambda) &= p(y_{i1}, \dots, y_{ip}|\theta_{kj}, \delta, \lambda)p(y_{ip+1}|y_{ip}, \dots, y_{i1}; \theta_{kj}, \delta, \lambda) \cdot \\
&\quad \cdot \dots \cdot p(y_{iT}|y_{iT-1}, \dots, y_{i1}; \theta_{kj}, \delta, \lambda)
\end{aligned}$$

dove la fattorizzazione mediante densità condizionate è dovuta al fatto che le osservazioni temporali non sono indipendenti tra loro. Pertanto

$$\begin{aligned}
p(y_i|\theta_{kj}, \delta, \lambda) &= p(y_{i1}, \dots, y_{ip}|\theta_{kj}, \delta, \lambda) \prod_{t=p+1}^T p(y_{it}|I_{it-1}; \theta_{kj}, \delta, \lambda) \\
&= p(y_{i1}|\theta_{kj}, \delta, \lambda) \prod_{t=2}^T p(y_{it}|I_{it-1}; \theta_{kj}, \delta, \lambda)
\end{aligned}$$

dove l'ultima uguaglianza deriva dall'aver posto il ritardo massimo p , per poter esprimere la varianza condizionate, pari a 1. Mentre la distribuzione condizionate è nota e pari a $N(0, \sigma_{ikjt}^2)$, quella non condizionate è pratica comune approssimarla con quella condizionate avente come parametri la media e la varianza non condizionate. Trattandosi di un modello ST-GARCH a media nulla dove il primo regime è un GARCH(1,1),

la distribuzione approssimata per la prima osservazione è $N(0, \frac{\omega_k}{1-\alpha_k-\beta_k})$, dove $\frac{\omega_k}{1-\alpha_k-\beta_k}$ è la varianza non condizionata di un GARCH(1,1) calcolata mediante la formula (1.1). Dunque

$$p(y_i|\theta_{kj}, \delta, \lambda) \approx (2\pi \frac{\omega_k}{1-\alpha_k-\beta_k})^{-\frac{1}{2}} \exp\{-\frac{y_{i1}^2}{2\frac{\omega_k}{1-\alpha_k-\beta_k}}\} \cdot \prod_{t=2}^T (2\pi \sigma_{ikjt}^2)^{-\frac{1}{2}} \exp\{-\frac{y_{it}^2}{2\sigma_{ikjt}^2}\} \quad (3.1)$$

con $\sigma_{ikjt}^2 = g(t, \delta, \lambda)[\omega_k + \alpha_k y_{it-1} + \beta_k \sigma_{ikjt-1}^2] + (1 - g(t, \delta, \lambda))[\omega_j + \alpha_j y_{it-1} + \beta_j \sigma_{ikjt-1}^2]$.

Marginalizzando la distribuzione di y rispetto alle etichette di gruppo, e indicate rispettivamente con π_k e ρ_j le probabilità a priori del primo e rispettivamente secondo regime, si ha che

$$p(y) = \sum_{k=1}^K \sum_{j=1}^J \pi_k \rho_j p(y|\theta_{kj}, \delta, \lambda) \quad (3.2)$$

dove $p(y|\theta_{kj}, \delta, \lambda)$ segue la (3.1), $\pi_k \geq 0$, $\rho_j \geq 0$ e $\sum_{k=1}^K \pi_k = \sum_{j=1}^J \rho_j = 1$.

Considerando il numero di componenti K e J noto e assumendo per semplicità che anche δ sia noto, i parametri del modello (3.2) sono:

$$(\lambda, \theta_{11}, \dots, \theta_{1J}, \dots, \theta_{KJ}, \pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J) = \Theta.$$

Date le osservazioni y_1, \dots, y_N , la *log*-verosimiglianza del modello è allora

$$\begin{aligned} l(\Theta|y_1, \dots, y_N) &= \log p(y_1, \dots, y_N|\Theta) \\ &= \sum_{i=1}^N \log p(y_i|\Theta) \\ &= \sum_{i=1}^N \log \left(\sum_{k=1}^K \sum_{j=1}^J \pi_k \rho_j p(y_i|\theta_{kj}, \lambda) \right). \end{aligned} \quad (3.3)$$

3.3 Stima di massima verosimiglianza

3.3.1 La verosimiglianza completa

Osservando la derivata prima della funzione di verosimiglianza (3.3)

$$\frac{\partial l(\Theta)}{\partial \Theta} = \sum_{i=1}^N \frac{1}{\sum_{k=1}^K \sum_{j=1}^J \pi_k \rho_j p(y_i | \theta_{kj}, \lambda)} \sum_{k=1}^K \sum_{j=1}^J \pi_k \rho_j \frac{\partial p(y_i | \theta_{kj}, \lambda)}{\partial \Theta}$$

si nota che, uguagliandola a zero, è solitamente impossibile da risolvere analiticamente perché i parametri non sono esplicitabili. Per questo motivo la massimizzazione della verosimiglianza avviene, come nel *co-clustering*, numericamente mediante l'uso dell'algoritmo EM o di qualche sua estensione definita per tener conto della doppia struttura latente (z, w) .

Ai fini del problema di *clustering* descritto dal modello (3.2), la doppia struttura latente (z_i, w_i) , $i = 1, \dots, N$, che definisce l'appartenenza di ciascuna osservazione y_i ad un *cluster* z_i per il primo regime ed un *cluster* w_i per il secondo regime, ha la seguente forma:

$$\begin{aligned} z_i &= (z_{i1}, \dots, z_{iK}) \\ \text{con } z_{ik} &= \begin{cases} 1 & \text{se } y_i^{(1)} \in \text{componente } k \\ 0 & \text{altrimenti} \end{cases} \\ w_i &= (w_{i1}, \dots, w_{iJ}) \\ \text{con } w_{ij} &= \begin{cases} 1 & \text{se } y_i^{(2)} \in \text{componente } j \\ 0 & \text{altrimenti} \end{cases} \end{aligned}$$

dove $y_i^{(l)}$ indica il regime l -esimo dell'unità y_i .

Si assume, quindi, che z_1, \dots, z_N e w_1, \dots, w_N siano indipendenti e identicamente distribuite secondo due multinomiali con una sola estrazione e con rispettivamente K e J categorie aventi probabilità associate (π_1, \dots, π_K) e (ρ_1, \dots, ρ_J) :

$$\begin{aligned} z_i &= (z_{i1}, \dots, z_{iK}) \sim \text{Bin}_K((\pi_1, \dots, \pi_K), 1) \\ w_i &= (w_{i1}, \dots, w_{iJ}) \sim \text{Bin}_J((\rho_1, \dots, \rho_J), 1). \end{aligned}$$

Il modello (3.2) parte dall'assunzione che, in seguito ad uno shock, ogni serie temporale possa reagire in modo indipendente rispetto al regime precedentemente seguito, da cui si ottiene che:

$$p(\mathbf{z}, \mathbf{w}) = p(\mathbf{z})p(\mathbf{w})$$

con $p(\mathbf{z}) = \prod_{i=1}^N p(z_i)$ e $p(\mathbf{w}) = \prod_{i=1}^N p(w_i)$.

Data la conoscenza dei vettori z_i e w_i , la densità dell'osservazione y_i è data da

$$p(y_i|\theta, \lambda, z_i, w_i) = \prod_{k=1}^K \prod_{j=1}^J p(y_i|\theta_{kj}, \lambda)^{z_{ik}w_{ij}}$$

da cui deriva che la *log*-verosimiglianza completa in cui si assumono osservati sia le serie storiche y_i sia i corrispondenti vettori latenti (z_i, w_i) , $i = 1, \dots, N$, è

$$\begin{aligned} l_C(\Theta|y_1, \dots, y_N, z_1, \dots, z_N, w_1, \dots, w_N) &= \\ &= \log p(y_1, \dots, y_N, z_1, \dots, z_N, w_1, \dots, w_N|\Theta) \\ &= \log (p(z_1, \dots, z_N, w_1, \dots, w_N|\pi_1, \dots, \pi_K, \rho_1, \dots, \rho_J) \cdot \\ &\quad \cdot p(y_1, \dots, y_N|z_1, \dots, z_N, w_1, \dots, w_N, \theta_{11}, \dots, \theta_{KJ}, \lambda)) \\ &= \log (p(z_1, \dots, z_N|\pi_1, \dots, \pi_K)p(w_1, \dots, w_N|\rho_1, \dots, \rho_J) \cdot \\ &\quad \cdot p(y_1, \dots, y_N|z_1, \dots, z_N, w_1, \dots, w_N, \theta_{11}, \dots, \theta_{KJ}, \lambda)) \\ &= \log \left(\prod_{i=1}^N p(z_i)p(w_i)p(y_i|z_i, w_i, \theta_{11}, \dots, \theta_{KJ}, \lambda) \right) \\ &= \sum_{i=1}^N \log \left(\prod_{k=1}^K \pi_k^{z_{ik}} \prod_{j=1}^J \rho_j^{w_{ij}} \prod_{k=1}^K \prod_{j=1}^J p(y_i|z_i, w_i, \theta_{kj}, \lambda)^{z_{ik}w_{ij}} \right) \\ &= \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J \log (\pi_k \rho_j p(y_i|z_i, w_i, \theta_{kj}, \lambda))^{z_{ik}w_{ij}} \\ &= \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik}w_{ij} \log (\pi_k \rho_j p(y_i|z_i, w_i, \theta_{kj}, \lambda)) \\ &= \sum_{i=1}^N \sum_{k=1}^K z_{ik} \log \pi_k + \sum_{i=1}^N \sum_{j=1}^J w_{ij} \log \rho_j + \\ &\quad + \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik}w_{ij} \log p(y_i|z_i, w_i, \theta_{kj}, \lambda) \quad (3.4) \end{aligned}$$

in cui il primo termine dipende dalle proporzioni dei gruppi del primo regime, il secondo da quelle dei gruppi del secondo regime e il terzo dalle densità di ogni serie storica.

A causa della presenza di due quantità latenti, l'inferenza non risulta così semplice come nel caso del modello mistura classico. Infatti,

provando a scrivere il valore atteso della *log*-verosimiglianza completa condizionatamente ai dati osservati e alla stima corrente dei parametri ($\lambda = \lambda^{(r)}, \theta = (\theta_{11}, \dots, \theta_{KJ}) = \theta^{(r)}, \pi = (\pi_1, \dots, \pi_K) = \pi^{(r)}, \rho = (\rho_1, \dots, \rho_J) = \rho^{(r)}$), si ottiene

$$\begin{aligned} \mathbb{E}_{z,w} [l_C(\Theta | y_1, \dots, y_N, z_1, \dots, z_N, w_1, \dots, w_N)] &= \\ &= \mathbb{E}_{z,w} \left[\sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik} w_{ij} \log (\pi_k^{(r)} \rho_j^{(r)} p(y_i | \theta_{kj}^{(r)}, \lambda^{(r)})) | y_1, \dots, y_N \right] \\ &= \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J \log (\pi_k^{(r)} \rho_j^{(r)} p(y_i | \theta_{kj}^{(r)}, \lambda^{(r)})) \mathbb{E}_{z,w} [z_{ik} w_{ij} | y_1, \dots, y_N] \end{aligned}$$

difficile da calcolare in quanto $\mathbb{E}_{z,w} [z_{ik} w_{ij} | y_1, \dots, y_N] = \mathbb{P}(z_{ik} w_{ij} = 1 | y_1, \dots, y_N)$ richiede il calcolo della distribuzione congiunta condizionata delle latenti. Infatti, sebbene le etichette latenti z_i e w_i siano indipendenti $\forall i = 1, \dots, N$, non lo sono condizionatamente ai dati y_i e quindi l'*E-step* richiede il calcolo di termini che non possono essere fattorizzati come in una mistura standard.

Per ovviare a questo problema, estendendo la letteratura presente per il *co-clustering*, vengono utilizzate altre forme dell'algoritmo EM, in particolare il *Classification-EM* (CEM) e lo *Stochastic-EM* con campionamento *Gibbs* (SEM-*Gibbs*).

3.3.2 Algoritmo CEM

L'algoritmo di *Classification-EM* integra un passo di classificazione tra *E-step* e *M-step*. In particolare, data la doppia struttura latente, si propone di eseguire i passi E- e C- separatamente per ciascun regime, concludendo con un unico passo M- finale.

Indicando con r la generica iterazione, l'algoritmo alterna iterativamente i seguenti cinque *step*:

1. *E-step*: fissati $\theta = \theta^{(r)}, \lambda = \lambda^{(r)}, \pi = \pi^{(r)}$ e $w_i = w_i^{(r)}, i = 1, \dots, N$, si determina il valore atteso della *log*-verosimiglianza completa (3.4)

rispetto ai vettori latenti z_i :

$$\begin{aligned}
& \mathbb{E}_z [l_C(\Theta | y_1, \dots, y_N, z_1, \dots, z_N, w_1, \dots, w_N)] = \\
& = \mathbb{E}_z \left[\sum_{i=1}^N \sum_{k=1}^K z_{ik} \log \pi_k^{(r)} + \sum_{i=1}^N \sum_{j=1}^J w_{ij}^{(r)} \log \rho_j^{(r)} + \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik} \cdot \right. \\
& \quad \left. \cdot w_{ij}^{(r)} \log p(y_i | w_{ij}^{(r)}, \theta_{kj}^{(r)}, \lambda^{(r)}) | y_1, \dots, y_N, w_1^{(r)}, \dots, w_N^{(r)} \right] \\
& = \sum_{i=1}^N \sum_{k=1}^K \log \pi_k^{(r)} \mathbb{E}_z [z_{ik} | y_i, w_i^{(r)}] + \sum_{i=1}^N \sum_{j=1}^J w_{ij}^{(r)} \log \rho_j^{(r)} + \\
& \quad + \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J w_{ij}^{(r)} \log p(y_i | w_{ij}^{(r)}, \theta_{kj}^{(r)}, \lambda^{(r)}) \mathbb{E}_z [z_{ik} | y_i, w_i^{(r)}]
\end{aligned}$$

con

$$\begin{aligned}
\mathbb{E}_z [z_{ik} | y_i, w_i^{(r)}] &= \mathbb{P}(z_{ik} = 1 | y_i, w_i^{(r)}) \\
&= \frac{\mathbb{P}(z_{ik} = 1, y_i, w_i^{(r)})}{p(y_i, w_i^{(r)})} \\
&= \frac{p(y_i | z_{ik} = 1, w_i^{(r)}) \mathbb{P}(z_{ik} = 1, w_i^{(r)})}{p(y_i | w_i^{(r)}) p(w_i^{(r)})} \\
&= \frac{p(y_i | z_{ik} = 1, w_i^{(r)}) \mathbb{P}(z_{ik} = 1 | w_i^{(r)})}{p(y_i | w_i^{(r)})} \\
&= \frac{p(y_i | z_{ik} = 1, w_i^{(r)}) \mathbb{P}(z_{ik} = 1)}{p(y_i | w_i^{(r)})} \\
&= \frac{\pi_k^{(r)} p(y_i | \theta_{kj}^{(r)})}{\sum_{k=1}^K \pi_k^{(r)} p(y_i | \theta_{kj}^{(r)})} = \tau_{ik}^{(r)} \quad k = 1, \dots, K.
\end{aligned}$$

Da notare che τ_{ik} corrisponde alla probabilità a posteriori che il primo regime $y_i^{(1)}$ dell'osservazione y_i segua il k -esimo modello GARCH(1,1) e costituisce la parte centrale del calcolo nel primo E-step.

2. C-step: ciascuna serie storica y_i viene attribuita al gruppo per cui è massima la probabilità a posteriori $\tau_{ik}^{(r)}$, ossia:

$$z_{ik}^{(r+1)} = \begin{cases} 1 & \text{se } \tau_{ik}^{(r)} = \arg \max_k \tau_{ik}^{(r)} \\ 0 & \text{altrimenti} \end{cases}$$

ottenendo la partizione $\mathcal{P}^{(r+1)} = (\mathcal{P}_1^{(r+1)}, \dots, \mathcal{P}_K^{(r+1)})$ con

$$\mathcal{P}_k^{(r+1)} = \{y_i : \tau_{ik}^{(r)} = \arg \max_k \tau_{ik}^{(r)}\}.$$

3. *E-step*: fissati $\theta = \theta^{(r)}$, $\lambda = \lambda^{(r)}$, $\rho = \rho^{(r)}$ e $z_i = z_i^{(r+1)}$, $i = 1, \dots, N$, si determina il valore atteso della *log-verosimiglianza* completa (3.4) rispetto ai vettori latenti w_i :

$$\begin{aligned} \mathbb{E}_w [l_C(\Theta|y_1, \dots, y_N, z_1, \dots, z_N, w_1, \dots, w_N)] &= \\ &= \mathbb{E}_w \left[\sum_{i=1}^N \sum_{k=1}^K z_{ik}^{(r+1)} \log \pi_k^{(r)} + \sum_{i=1}^N \sum_{j=1}^J w_{ij} \log \rho_j^{(r)} + \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J \right. \\ &\quad \left. z_{ik}^{(r+1)} w_{ij} \log p(y_i | z_{ik}^{(r+1)}, \theta_{kj}^{(r)}, \lambda^{(r)}) | y_1, \dots, y_N, z_1^{(r+1)}, \dots, z_N^{(r+1)} \right] \\ &= \sum_{i=1}^N \sum_{k=1}^K z_{ik}^{(r+1)} \log \pi_k^{(r)} + \sum_{i=1}^N \sum_{j=1}^J \log \rho_j^{(r)} \mathbb{E}_w [w_{ij} | y_i, z_i^{(r+1)}] + \\ &\quad + \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik}^{(r+1)} \log p(y_i | z_{ik}^{(r+1)}, \theta_{kj}^{(r)}, \lambda^{(r)}) \mathbb{E}_w [w_{ij} | y_i, z_i^{(r+1)}] \end{aligned}$$

con

$$\begin{aligned} \mathbb{E}_w [w_{ij} | y_i, z_i^{(r+1)}] &= \mathbb{P}(w_{ij} = 1 | y_i, z_i^{(r+1)}) \\ &= \frac{\mathbb{P}(w_{ij} = 1, y_i, z_i^{(r+1)})}{p(y_i, z_i^{(r+1)})} \\ &= \frac{p(y_i | w_{ij} = 1, z_i^{(r+1)}) \mathbb{P}(w_{ij} = 1, z_i^{(r+1)})}{p(y_i | z_i^{(r+1)}) p(z_i^{(r+1)})} \\ &= \frac{p(y_i | w_{ij} = 1, z_i^{(r+1)}) \mathbb{P}(w_{ij} = 1 | z_i^{(r+1)})}{p(y_i | z_i^{(r+1)})} \\ &= \frac{p(y_i | w_{ij} = 1, z_i^{(r+1)}) \mathbb{P}(w_{ij} = 1)}{p(y_i | z_i^{(r+1)})} \\ &= \frac{\rho_j^{(r)} p(y_i | \theta_{kj}^{(r)})}{\sum_{j=1}^J \rho_j^{(r)} p(y_i | \theta_{kj}^{(r)})} = \eta_{ij}^{(r)} \quad j = 1, \dots, J. \end{aligned}$$

Da notare che η_{ij} corrisponde alla probabilità a posteriori che il secondo regime $y_i^{(2)}$ dell'osservazione y_i segua il j -esimo modello GARCH(1,1) e costituisce la parte centrale del calcolo nel secondo *E-step*.

4. C-step: ciascuna serie osservata y_i viene attribuita al gruppo per cui è massima la probabilità a posteriori $\eta_{ij}^{(r)}$, ossia:

$$w_{ij}^{(r+1)} = \begin{cases} 1 & \text{se } \eta_{ij}^{(r)} = \arg \max_j \eta_{ij}^{(r)} \\ 0 & \text{altrimenti} \end{cases}$$

ottenendo la partizione $\mathcal{P}'^{(r+1)} = (\mathcal{P}'_1^{(r+1)}, \dots, \mathcal{P}'_J^{(r+1)})$ con

$$\mathcal{P}'_j^{(r+1)} = \{y_i : \eta_{ij}^{(r)} = \arg \max_j \eta_{ij}^{(r)}\}.$$

5. M-step: fissati, quindi, $z_i = z_i^{(r+1)}$ e $w_i = w_i^{(r+1)}$, $i = 1, \dots, N$, si stimano i parametri $\hat{\pi}^{(r+1)}$, $\hat{\rho}^{(r+1)}$, $\hat{\lambda}^{(r+1)}$ e $\hat{\theta}^{(r+1)}$ usando le partizioni appena ottenute:

- $\hat{\pi}_k^{(r+1)} = \frac{|\mathcal{P}_k^{(r+1)}|}{N}$ $k = 1, \dots, K$ con $|\mathcal{P}_k^{(r+1)}|$ cardinalità di $\mathcal{P}_k^{(r+1)}$
- $\hat{\rho}_j^{(r+1)} = \frac{|\mathcal{P}'_j^{(r+1)}|}{N}$ $j = 1, \dots, J$ con $|\mathcal{P}'_j^{(r+1)}|$ cardinalità di $\mathcal{P}'_j^{(r+1)}$
- $\hat{\lambda}^{(r+1)}$ e $\hat{\theta}^{(r+1)}$ valori che massimizzano la *log*-verosimiglianza completa (3.4), ossia

$$\begin{aligned} (\hat{\theta}^{(r+1)}, \hat{\lambda}^{(r+1)}) = \arg \max_{\theta, \lambda} & \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik}^{(r+1)} w_{ij}^{(r+1)} \\ & \cdot \log p(y_i | z_i^{(r+1)}, w_i^{(r+1)}, \theta_{kj}, \lambda). \end{aligned}$$

Per derivare rispetto a λ e θ la funzione obiettivo

$$\begin{aligned} & \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik}^{(r+1)} w_{ij}^{(r+1)} \log p(y_i | z_i^{(r+1)}, w_i^{(r+1)}, \theta_{kj}, \lambda) = \\ & = \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik}^{(r+1)} w_{ij}^{(r+1)} \log \left(2\pi \frac{\omega_k}{1 - \alpha_k - \beta_k} \right)^{-\frac{1}{2}} \\ & \quad \cdot \exp \left\{ -\frac{y_{i1}^2}{2 \frac{\omega_k}{1 - \alpha_k - \beta_k}} \right\} \prod_{t=2}^T (2\pi \sigma_{ikjt}^2)^{-\frac{1}{2}} \exp \left\{ -\frac{y_{it}^2}{2 \sigma_{ikjt}^2} \right\} \end{aligned}$$

occorre calcolare la derivata della varianza condizionata

$$\begin{aligned} \sigma_{ikjt}^2 = g(t, \delta, \lambda) & [\omega_k + \alpha_k y_{it-1} + \beta_k \sigma_{ikjt-1}^2] + \\ & + (1 - g(t, \delta, \lambda)) [\omega_j + \alpha_j y_{it-1} + \beta_j \sigma_{ikjt-1}^2]. \end{aligned}$$

Osservando, però, l'espressione della varianza, si nota che la derivata prima non consente di ottenere stime esplicite dei parametri. Pertanto, per quanto riguarda la stima $\hat{\lambda}^{(r+1)}$ e $\hat{\theta}^{(r+1)}$, è necessario ricorrere ad un algoritmo di ottimizzazione numerica, come ad esempio il *Broyden-Fletcher-Goldfarb-Shanno* (BFGS).

Partendo da una soluzione iniziale per il valore dei parametri, sotto condizioni di regolarità piuttosto deboli, il CEM itera i cinque passi fino a convergere ad un massimo locale della verosimiglianza. Data la natura solitamente non convessa della *log*-verosimiglianza, a inizializzazioni diverse possono corrispondere risultati diversi in quanto è possibile che l'algoritmo converga a un massimo locale e non a quello globale. Non vi sono strategie uniformemente più opportune, ma pratica comune è quella di scegliere un'inizializzazione migliore di quella casuale.

La regola di arresto per la convergenza dell'algoritmo viene solitamente impostata sulla verosimiglianza o sui parametri del modello. Ad esempio, l'EM continua fintanto che la variazione della *log*-verosimiglianza o dei parametri da un'iterazione all'altra non è più rilevante.

Una volta ottenute le stime di massima verosimiglianza dei parametri del modello, ciascuna serie storica osservata y_i del campione viene attribuita al gruppo per cui è massima la probabilità a posteriori (*MAP rule*), così come avviene nei passi di classificazione dell'algoritmo.

3.3.3 Algoritmo SEM-*Gibbs*

L'algoritmo di *Stochastic-EM* integra un passo stocastico tra *E-step* e *M-step*, in cui i dati mancanti vengono simulati secondo la loro distribuzione condizionata, date y_i e le stime correnti dei parametri θ e λ . Tuttavia, sorge il problema dell'impossibilità di simulare simultaneamente z e w in quanto, come visto in precedenza, non è agevole calcolare $\mathbb{P}(z_{ik}w_{ij} = 1 | y_1, \dots, y_N)$. Per simulare le etichette (z_i, w_i) si adotta uno schema di campionamento *Gibbs* (Geman e Geman, 1984), da cui deriva il nome SEM-*Gibbs*. In particolare, data la doppia struttura latente, si propone di eseguire i passi E- e S- separatamente per ciascun regime, concludendo con un unico passo M- finale. È importante sottolineare che non viene effettuata alcuna approssimazione, ma solo simulazioni.

Indicando con r la generica iterazione, l'algoritmo alterna iterativamente i seguenti cinque *step*:

1. *E-step*: fissati $\theta = \theta^{(r)}$, $\lambda = \lambda^{(r)}$, $\pi = \pi^{(r)}$ e $w_i = w_i^{(r)}$, $i = 1, \dots, N$, si determina il valore atteso della *log-verosimiglianza* completa (3.4) rispetto ai vettori latenti z_i :

$$\begin{aligned}
& \mathbb{E}_z [l_C(\Theta | y_1, \dots, y_N, z_1, \dots, z_N, w_1, \dots, w_N)] = \\
& = \mathbb{E}_z \left[\sum_{i=1}^N \sum_{k=1}^K z_{ik} \log \pi_k^{(r)} + \sum_{i=1}^N \sum_{j=1}^J w_{ij}^{(r)} \log \rho_j^{(r)} + \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik} \cdot \right. \\
& \quad \left. \cdot w_{ij}^{(r)} \log p(y_i | w_{ij}^{(r)}, \theta_{kj}^{(r)}, \lambda^{(r)}) | y_1, \dots, y_N, w_1^{(r)}, \dots, w_N^{(r)} \right] \\
& = \sum_{i=1}^N \sum_{k=1}^K \log \pi_k^{(r)} \mathbb{E}_z [z_{ik} | y_i, w_i^{(r)}] + \sum_{i=1}^N \sum_{j=1}^J w_{ij}^{(r)} \log \rho_j^{(r)} + \\
& \quad + \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J w_{ij}^{(r)} \log p(y_i | w_{ij}^{(r)}, \theta_{kj}^{(r)}, \lambda^{(r)}) \mathbb{E}_z [z_{ik} | y_i, w_i^{(r)}]
\end{aligned}$$

con

$$\begin{aligned}
\mathbb{E}_z [z_{ik} | y_i, w_i^{(r)}] &= \mathbb{P}(z_{ik} = 1 | y_i, w_i^{(r)}) \\
&= \frac{\mathbb{P}(z_{ik} = 1, y_i, w_i^{(r)})}{p(y_i, w_i^{(r)})} \\
&= \frac{p(y_i | z_{ik} = 1, w_i^{(r)}) \mathbb{P}(z_{ik} = 1, w_i^{(r)})}{p(y_i | w_i^{(r)}) p(w_i^{(r)})} \\
&= \frac{p(y_i | z_{ik} = 1, w_i^{(r)}) \mathbb{P}(z_{ik} = 1 | w_i^{(r)})}{p(y_i | w_i^{(r)})} \\
&= \frac{p(y_i | z_{ik} = 1, w_i^{(r)}) \mathbb{P}(z_{ik} = 1)}{p(y_i | w_i^{(r)})} \\
&= \frac{\pi_k^{(r)} p(y_i | \theta_{kj}^{(r)})}{\sum_{k=1}^K \pi_k^{(r)} p(y_i | \theta_{kj}^{(r)})} = \tilde{z}_{ik}^{(r)} \quad k = 1, \dots, K.
\end{aligned}$$

Da notare che \tilde{z}_{ik} corrisponde alla probabilità a posteriori che il primo regime $y_i^{(1)}$ dell'osservazione y_i segua il k -esimo modello GARCH(1,1) e costituisce la parte centrale del calcolo nel primo *E-step*.

2. *S-step*: si genera la partizione del primo regime

$$z_i^{(r+1)} = (z_{i1}^{(r+1)}, \dots, z_{iK}^{(r+1)}) \quad i = 1, \dots, N$$

secondo la distribuzione multinomiale

$$z_i^{(r+1)} \sim \text{Bin}_K((\tilde{z}_{i1}^{(r)}, \dots, \tilde{z}_{iK}^{(r)}), 1).$$

3. *E-step*: fissati $\theta = \theta^{(r)}$, $\lambda = \lambda^{(r)}$, $\rho = \rho^{(r)}$ e $z_i = z_i^{(r+1)}$, $i = 1, \dots, N$, si determina il valore atteso della *log-verosimiglianza* completa (3.4) rispetto ai vettori latenti w_i :

$$\begin{aligned} \mathbb{E}_w [l_C(\Theta | y_1, \dots, y_N, z_1, \dots, z_N, w_1, \dots, w_N)] &= \\ &= \mathbb{E}_w \left[\sum_{i=1}^N \sum_{k=1}^K z_{ik}^{(r+1)} \log \pi_k^{(r)} + \sum_{i=1}^N \sum_{j=1}^J w_{ij} \log \rho_j^{(r)} + \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J \right. \\ &\quad \left. z_{ik}^{(r+1)} w_{ij} \log p(y_i | z_{ik}^{(r+1)}, \theta_{kj}^{(r)}, \lambda^{(r)}) | y_1, \dots, y_N, z_1^{(r+1)}, \dots, z_N^{(r+1)} \right] \\ &= \sum_{i=1}^N \sum_{k=1}^K z_{ik}^{(r+1)} \log \pi_k^{(r)} + \sum_{i=1}^N \sum_{j=1}^J \log \rho_j^{(r)} \mathbb{E}_w [w_{ij} | y_i, z_i^{(r+1)}] + \\ &\quad + \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik}^{(r+1)} \log p(y_i | z_{ik}^{(r+1)}, \theta_{kj}^{(r)}, \lambda^{(r)}) \mathbb{E}_w [w_{ij} | y_i, z_i^{(r+1)}] \end{aligned}$$

con

$$\begin{aligned} \mathbb{E}_w [w_{ij} | y_i, z_i^{(r+1)}] &= \mathbb{P}(w_{ij} = 1 | y_i, z_i^{(r+1)}) \\ &= \frac{\mathbb{P}(w_{ij} = 1, y_i, z_i^{(r+1)})}{p(y_i, z_i^{(r+1)})} \\ &= \frac{p(y_i | w_{ij} = 1, z_i^{(r+1)}) \mathbb{P}(w_{ij} = 1, z_i^{(r+1)})}{p(y_i | z_i^{(r+1)}) p(z_i^{(r+1)})} \\ &= \frac{p(y_i | w_{ij} = 1, z_i^{(r+1)}) \mathbb{P}(w_{ij} = 1 | z_i^{(r+1)})}{p(y_i | z_i^{(r+1)})} \\ &= \frac{p(y_i | w_{ij} = 1, z_i^{(r+1)}) \mathbb{P}(w_{ij} = 1)}{p(y_i | z_i^{(r+1)})} \\ &= \frac{\rho_j^{(r)} p(y_i | \theta_{kj}^{(r)})}{\sum_{j=1}^J \rho_j^{(r)} p(y_i | \theta_{kj}^{(r)})} = \tilde{w}_{ij}^{(r)} \quad j = 1, \dots, J. \end{aligned}$$

Da notare che \tilde{w}_{ij} corrisponde alla probabilità a posteriori che il secondo regime $y_i^{(2)}$ dell'osservazione y_i segua il j -esimo modello GARCH(1,1) e costituisce la parte centrale del calcolo nel secondo *E-step*.

4. S-*step*: si genera la partizione del secondo regime

$$w_i^{(r+1)} = (w_{i1}^{(r+1)}, \dots, w_{iJ}^{(r+1)}) \quad i = 1, \dots, N$$

secondo la distribuzione multinomiale

$$w_i^{(r+1)} \sim \text{Bin}_J((\tilde{w}_{i1}^{(r)}, \dots, \tilde{w}_{iJ}^{(r)}), 1).$$

5. M-*step*: fissati, quindi, $z_i = z_i^{(r+1)}$ e $w_i = w_i^{(r+1)}$, $i = 1, \dots, N$, si stimano i parametri $\hat{\pi}^{(r+1)}$, $\hat{\rho}^{(r+1)}$, $\hat{\lambda}^{(r+1)}$ e $\hat{\theta}^{(r+1)}$ usando le partizioni ottenute con i due passi di simulazione:

- $\hat{\pi}_k^{(r+1)} = \frac{\sum_{i=1}^N z_{ik}^{(r+1)}}{N} \quad k = 1, \dots, K$
- $\hat{\rho}_j^{(r+1)} = \frac{\sum_{i=1}^N w_{ij}^{(r+1)}}{N} \quad j = 1, \dots, J$
- $\hat{\lambda}^{(r+1)}$ e $\hat{\theta}^{(r+1)}$ valori che massimizzano la *log*-verosimiglianza completa (3.4), ossia

$$\begin{aligned} (\hat{\theta}^{(r+1)}, \hat{\lambda}^{(r+1)}) = \arg \max_{\theta, \lambda} & \sum_{i=1}^N \sum_{k=1}^K \sum_{j=1}^J z_{ik}^{(r+1)} w_{ij}^{(r+1)} \cdot \\ & \cdot \log p(y_i | z_i^{(r+1)}, w_i^{(r+1)}, \theta_{kj}, \lambda). \end{aligned}$$

Come già osservato nell'esposizione dell'M-*step* dell'algoritmo CEM, per quanto riguarda la stima $\hat{\lambda}^{(r+1)}$ e $\hat{\theta}^{(r+1)}$, è necessario ricorrere ad un algoritmo di ottimizzazione numerica, come ad esempio il BFGS.

A differenza dell'algoritmo CEM, il SEM-*Gibbs* non garantisce un incremento della *log*-verosimiglianza ad ogni iterazione ma genera, invece, una catena di Markov irriducibile con un'unica distribuzione stazionaria, la quale si prevede sia concentrata attorno alla stima di massima verosimiglianza dei parametri. Utilizzando le classiche regole di arresto già discusse per la convergenza, si rischia che l'algoritmo converga dopo un numero di iterazioni significativamente superiore al necessario, oppure che raggiunga la convergenza prima di avvicinarsi sufficientemente al massimo della verosimiglianza. Per questa ragione, è opportuno adottare criteri alternativi, come l'imposizione di una soglia sulla variazione relativa (anziché assoluta) della *log*-verosimiglianza o dei parametri tra iterazioni successive.

Inoltre, poiché l'ultimo valore di verosimiglianza registrato non rappresenta necessariamente il massimo, non è sempre opportuno adottare l'ultima stima dei parametri come definitiva. In alternativa, si

può considerare come stima finale di θ e λ la media rispettivamente delle stime $\{\theta^{(c)} : c = B + 1, B + 2, \dots, B + C\}$ e $\{\lambda^{(c)} : c = B + 1, B + 2, \dots, B + C\}$, determinate dopo un periodo iniziale di lunghezza B . Il valore di B può essere scelto, ad esempio, in base alle variazioni relative della verosimiglianza ad ogni iterazione.

Sebbene non vi sia un incremento progressivo della verosimiglianza, data la sua natura non convessa, inizializzazioni diverse possono comunque condurre a risultati differenti a causa della presenza di massimi locali. Non esistono strategie uniformemente ottimali per affrontare questo problema, ma è prassi comune adottare un'inizializzazione migliore rispetto a quella casuale.

Una volta ottenute le stime di massima verosimiglianza dei parametri del modello, le partizioni z_i e w_i , $i = 1, \dots, N$, vengono generate secondo le due distribuzioni multinomiali, seguendo lo stesso procedimento descritto nei primi quattro passi dell'algoritmo *SEM-Gibbs*.

Capitolo 4

Alcune analisi empiriche

4.1 Uno studio di simulazione

4.1.1 Obiettivi e scenari di simulazione

In questo paragrafo verrà presentato uno studio di simulazione condotto con l'obiettivo di analizzare le prestazioni del metodo di *clustering* proposto, valutandone l'adeguatezza delle partizioni ottenute in contesti di serie storiche finanziarie soggette a un cambiamento di regime indotto da uno *shock* esogeno. Oltre a una valutazione in termini assoluti, lo studio mira a confrontare i due algoritmi di stima, CEM e SEM-*Gibbs*, trattati rispettivamente nei paragrafi 3.3.2 e 3.3.3.

Inglobando una pluralità di elementi, la procedura si presta ad applicazioni in una molteplicità di scenari possibili. Per tale ragione, e tenendo conto dei tempi computazionali, sono state compiute delle scelte mirate, limitando la variazione a caratteristiche specifiche dei dati per valutarne l'efficacia. Nello specifico, sono state analizzate diverse configurazioni dei regimi in termini di distanza tra i gruppi e grado di bilanciamento. Per ciascuna configurazione, la stima di massima verosimiglianza è stata eseguita utilizzando entrambi gli algoritmi CEM e SEM-*Gibbs*, consentendo un successivo confronto.

Per ciascuno scenario considerato, $B = 60$ campioni composti da $N = 50$ serie storiche di lunghezza $T = 1000$ sono state generati dal modello (1.2), ovvero uno ST-GARCH composto da due regimi GARCH(1,1) con funzione di transizione logistica. La soglia è stata fissata al tempo $\lambda = 500$ e si è posto il parametro di liscio $\delta = 0.1$. Il numero di *cluster* è stato mantenuto costante, con $K = J = 2$ nel corso di tutto lo studio. Gli scenari di simulazione differiscono nella specificazione dei parametri $(\omega_1, \alpha_1, \beta_1, \dots, \omega_K, \alpha_K, \beta_K)$ per il primo regime e

$(\omega_1, \alpha_1, \beta_1, \dots, \omega_J, \alpha_J, \beta_J)$ per il secondo, nonché nel grado di bilanciamento dei gruppi (π_1, \dots, π_K) per il primo regime e (ρ_1, \dots, ρ_J) per il secondo.

Entrando maggiormente nel dettaglio, i quattro scenari su cui si è basato lo studio di simulazione sono così definiti:

1. Scenario 1: i gruppi sono separati per il primo regime, con parametri $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2) = (0.1, 0.1, 0.8, 0.3, 0.2, 0.6)$. Per il secondo regime, i gruppi sono differenti rispetto al primo, ma meno distinti tra loro, con parametri $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2) = (0.1, 0.2, 0.7, 0.2, 0.2, 0.5)$. Le proporzioni sono in entrambi i regimi bilanciate, con $(\pi_1, \pi_2) = (\rho_1, \rho_2) = (0.5, 0.5)$;
2. Scenario 2: i gruppi sono separati per il primo regime, con parametri $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2) = (0.1, 0.1, 0.8, 0.3, 0.2, 0.6)$. Per il secondo regime, i gruppi sono differenti rispetto al primo, ma meno distinti tra loro, con parametri $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2) = (0.1, 0.2, 0.7, 0.2, 0.2, 0.5)$. Le proporzioni sono in entrambi i regimi sbilanciate, con $(\pi_1, \pi_2) = (\rho_1, \rho_2) = (0.2, 0.8)$;
3. Scenario 3: i gruppi non sono ben separati per il primo regime, con parametri $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2) = (0.1, 0.1, 0.75, 0.15, 0.15, 0.7)$. Per il secondo regime, i gruppi sono poco differenti rispetto al primo, ma ben separati tra loro, con parametri $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2) = (0.3, 0.1, 0.75, 0.15, 0.2, 0.6)$. Le proporzioni sono in entrambi i regimi bilanciate, con $(\pi_1, \pi_2) = (\rho_1, \rho_2) = (0.5, 0.5)$;
4. Scenario 4: i gruppi non sono ben separati per il primo regime, con parametri $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2) = (0.1, 0.1, 0.75, 0.15, 0.15, 0.7)$. Per il secondo regime, i gruppi sono poco differenti rispetto al primo, ma ben separati tra loro, con parametri $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2) = (0.3, 0.1, 0.75, 0.15, 0.2, 0.6)$. Le proporzioni sono in entrambi i regimi sbilanciate, con $(\pi_1, \pi_2) = (\rho_1, \rho_2) = (0.2, 0.8)$.

4.1.2 Dettagli di implementazione

Per ogni campione simulato è stato stimato il modello (3.2), assumendo noti il numero di gruppi $K = J = 2$ e il parametro di liscio $\delta = 0.1$. La stima del modello è stata effettuata utilizzando sia l'algoritmo CEM sia il SEM-*Gibbs*, come descritto nei paragrafi precedenti.

Come già accennato, la funzione di verosimiglianza da massimizzare è generalmente non convessa, il che implica la presenza di possibili

massimi locali che possono condurre a risultati differenti a seconda dell'inizializzazione. Inoltre, è possibile che, per alcuni punti vicini alla frontiera parametrica, la verosimiglianza tenda all'infinito, generando stime degeneri e prive di buone proprietà statistiche. Per questo motivo, è generalmente preferibile inizializzare i parametri a una certa distanza dalla frontiera, al fine di ridurre la probabilità di ottenere una verosimiglianza infinita.

L'inizializzazione dei parametri (ω, α, β) di ciascun gruppo di ogni regime, così come della soglia λ , è stata effettuata in modo casuale, seguendo le accortezze precedentemente descritte e rispettando i vincoli imposti dal modello ST-GARCH. Tra questi vincoli, particolare attenzione è stata posta nell'evitare situazioni di regime GARCH(1,1) non stazionario, poiché per $\alpha + \beta \approx 1$ la varianza non condizionata del processo tende all'infinito, generando stime degeneri dei parametri. Per ciascuno scenario sono stati utilizzati gli stessi valori iniziali, indipendentemente dalla loro vicinanza o lontananza rispetto ai veri parametri. In merito alle proporzioni, si è deciso di adottare un'inizializzazione bilanciata, che corrisponde esattamente ai valori reali per il primo e terzo scenario, ma risulta significativamente diversa per il secondo e quarto scenario. È importante notare che, man mano che i parametri iniziali si avvicinano ai loro valori reali, si riduce il rischio che la procedura converga a massimi locali. Pertanto, per studi futuri, potrebbe essere utile esplorare ulteriori scenari di campionamento per valutare la sensibilità della procedura all'inizializzazione dei parametri. Entrambi gli algoritmi di stima richiedono anche un valore iniziale per i vettori latenti w_i^0 relativi al secondo regime. Poiché in letteratura non esiste una regola universalmente accettata, si è optato per un'inizializzazione più robusta rispetto a quella casuale, utilizzando una tecnica di *clustering* specifica per dati dinamici, come le serie storiche. La procedura seguita è quella proposta da Pérez-Espinosa (2019) mediante l'uso della distanza *Shape-Based Distance* (SBD), una misura di dissomiglianza focalizzata sulla forma delle serie temporali, indipendentemente dalle eventuali variazioni di scala. Inoltre, per il calcolo dei centroidi dei *cluster*, si è scelto un metodo coerente, basato anch'esso sulla forma delle serie storiche, come descritto nell'articolo di riferimento.

L'efficacia della procedura studiata è stata valutata sia in termini assoluti, per ciascuna specificazione, sia in termini relativi, confrontando i diversi scenari e i due algoritmi di stima. La valutazione si è basata sulle stime dei parametri, della soglia e delle proporzioni del modello ST-GARCH, oltre che sull'adeguatezza delle partizioni ottenute. Per

quest'ultimo aspetto, si è utilizzato l'*Adjusted Rand Index* (ARI, (Hubert e Arabie, 1985)), una misura statistica utilizzata in letteratura per valutare la similarità tra due partizioni di un insieme di oggetti, invariante rispetto a permutazioni delle etichette. L'ARI assume valori tra -1 e 1 . Un valore pari a 1 indica una perfetta concordanza tra le partizioni; 0 indica che la similarità non è migliore di quella che ci si aspetterebbe con un raggruppamento casuale; valori negativi indicano una similarità inferiore a quella casuale.

Prima di procedere all'analisi dei risultati, è necessario evidenziare un comportamento critico osservato negli algoritmi impiegati. Entrambi gli algoritmi tendono a entrare in una situazione di stallo qualora, nel corso della stima, una delle proporzioni assuma valori particolarmente piccoli (inferiori a 0.05). Un'analisi più approfondita ha rivelato che questa situazione si manifesta prevalentemente nelle fasi iniziali dell'algoritmo ed è fortemente influenzata dai valori di inizializzazione. In particolare, nel caso in cui nessuno dei singoli regimi sia un modello GARCH(1,1) quasi integrato, una configurazione di parametri tale che ($\alpha + \beta \approx 1$) genera varianze molto elevate, compromettendo la stima delle probabilità a posteriori durante gli *step* di *Expectation*. Per ovviare a tale problema, è stata introdotta una condizione secondo la quale, nel caso in cui le proporzioni stimate scendano sotto il valore di 0.05 , l'algoritmo venga interrotto e si proceda con una nuova inizializzazione casuale dei parametri. Tuttavia, dato l'elevato carico computazionale richiesto dagli algoritmi e il tempo limitato a disposizione per questo studio, è stato imposto un limite massimo di 6 reinizializzazioni per ciascun algoritmo. A causa di questa limitazione, per alcuni campioni generati non sono disponibili i risultati, che si riferiscono, mediamente a circa 50 campioni, a seconda dello scenario.

Lo studio di simulazione è stato eseguito nell'ambiente di programmazione R (R Core Team, 2022), utilizzato anche per l'implementazione degli algoritmi CEM e SEM-*Gibbs*. Durante l'analisi, è stata impiegata la libreria *pdfCluster* (Azzalini e Menardi, 2014) per il calcolo dell'ARI, la libreria *Brobdingnag* (Hankin, 2013) per la gestione di numeri infinitamente piccoli coinvolti nei calcoli, e la libreria *dtwclust* (Sardà-Espinosa, 2019) per l'inizializzazione delle etichette.

4.1.3 Analisi dei risultati

Complessivamente, i risultati mostrano una generale appropriatezza della procedura proposta nel ripartire i dati in gruppi. Analizzando la Ta-

	ARI algoritmo CEM		ARI algoritmo SEM- <i>Gibbs</i>	
	Regime 1	Regime 2	Regime 1	Regime 2
Scenario 1	0.850 [0.112]	0.622 [0.313]	0.835 [0.117]	0.732 [0.098]
Scenario 2	0.754 [0.350]	0.733 [0.263]	0.770 [0.259]	0.740 [0.151]
Scenario 3	0.665 [0.300]	1.000 [0.000]	0.731 [0.133]	1.000 [0.000]
Scenario 4	0.193 [0.404]	1.000 [0.000]	0.406 [0.410]	1.000 [0.000]

Tabella 4.1: Media (deviazione standard) dell'ARI ottenuto mediante gli algoritmi CEM e SEM-*Gibbs* nei diversi scenari di simulazione.

bella 4.1, infatti, emerge che, in presenza di *cluster* ben distinti, entrambi gli algoritmi CEM e SEM-*Gibbs* risultano efficaci nella separazione dei gruppi, indipendentemente dal bilanciamento delle proporzioni e dalla similitudine dei *cluster* nel primo regime. Questo è confermato dai valori medi dell'ARI pari a 1 con una deviazione *standard* nulla. Non sorprendentemente, la configurazione più difficile da ricostruire, in termini di *clustering*, si riferisce a gruppi molto vicini e sbilanciati, con l'algoritmo CEM che mostra un ARI medio di 0.193 e uno *standard error* di 0.404, il che indica un intervallo di confidenza che scende fino a valori negativi. Considerando che la separazione tra i parametri dei gruppi nel primo (rispettivamente secondo) regime è inferiore (risp. superiore) nei primi due scenari, e superiore (risp. inferiore) negli ultimi due, si osserva dalla Tabella 4.1 che, diminuendo il livello di separazione, l'ARI medio aumenta e lo *standard error* diminuisce. Questo comportamento è evidente per entrambi gli algoritmi in tutti gli scenari, indicando che essi incontrano maggiori difficoltà e instabilità nel raggruppamento dei dati quando i gruppi sono più simili tra loro. Per quanto riguarda il bilanciamento dei gruppi, non si riscontra un comportamento uniforme in tutti i regimi. Per il primo regime, a parità di parametri, proporzioni sbilanciate riducono l'ARI, mentre per il secondo regime avviene l'opposto: proporzioni sbilanciate portano a un aumento dell'ARI, ma solo in presenza di *cluster* non perfettamente separati.

Confrontando CEM e SEM-*Gibbs* in tutti gli scenari, il SEM-*Gibbs* mostra una *performance* generalmente superiore. In sette degli otto confronti possibili, il SEM-*Gibbs* registra valori medi dell'ARI superiori rispetto al CEM. Le deviazioni *standard* sono generalmente inferiori per il SEM-*Gibbs*. E' utile tenere presente, tuttavia, che a fronte di una maggior accuratezza, i tempi computazionali, e il numero di iterazioni per raggiungere la convergenza dell'algoritmo SEM-*Gibbs* sono generalmente superiori.

		Soglia			
	Vero	CEM		SEM- <i>Gibbs</i>	
λ	500	501.32 [5.11]		500.54 [4.93]	
		Primo regime		Secondo regime	
	Vero	CEM	SEM- <i>Gibbs</i>	CEM	SEM- <i>Gibbs</i>
ω_1	0.10	0.100 [0.013]	0.100 [0.013]	0.103 [0.016]	0.102 [0.010]
α_1	0.10	0.097 [0.008]	0.099 [0.009]	0.208 [0.017]	0.202 [0.011]
β_1	0.80	0.802 [0.019]	0.801 [0.019]	0.695 [0.022]	0.698 [0.017]
ω_2	0.30	0.302 [0.033]	0.304 [0.034]	0.194 [0.037]	0.208 [0.021]
α_2	0.20	0.197 [0.014]	0.198 [0.014]	0.203 [0.019]	0.203 [0.013]
β_2	0.60	0.601 [0.030]	0.599 [0.030]	0.511 [0.070]	0.484 [0.040]
π_1	0.50	0.475 [0.072]	0.483 [0.075]	0.409 [0.173]	0.481 [0.087]
π_2	0.50	0.525 [0.072]	0.517 [0.075]	0.591 [0.173]	0.519 [0.087]

Tabella 4.2: Risultati per lo Scenario 1: media [deviazione standard] dei parametri stimati mediante algoritmo CEM e SEM-*Gibbs*.

Esaminando le stime dei parametri riportate nelle Tabelle 4.2-4.5, si riscontra che i risultati per la soglia del cambiamento di regime sono soddisfacenti in tutti gli scenari. Le stime dei parametri del modello GARCH(1,1) per i singoli regimi sono coerenti con le osservazioni sull'ARI, mostrando peggiori performance e maggiore instabilità quando i *cluster* sono più vicini e le proporzioni sono sbilanciate, con stime lontane dai valori reali e *standard error* più alti.

Le stime delle proporzioni dei *cluster* risultano tendenzialmente instabili, fortemente influenzate dal livello di separazione dei gruppi e dall'algoritmo utilizzato. In generale, le stime sono accettabili, ma una maggiore separazione tra i gruppi migliora l'accuratezza e la stabilità delle stime, con il SEM-*Gibbs* che tende a produrre valori più vicini ai veri parametri rispetto al CEM. Tuttavia, non si rilevano differenze significative nella stima di π e ρ in funzione del bilanciamento delle proporzioni, a parità di parametri e algoritmo.

4.2 Un'applicazione a dati reali

4.2.1 I dati EURO STOXX 50

Nel contesto finanziario, viene definita *società blue chip* un'azienda grande, consolidata e finanziariamente solida, con una lunga storia di *per-*

		Soglia			
	Vero	CEM		SEM-<i>Gibbs</i>	
λ	500	499.39 [4.51]		499.62 [3.99]	
		Primo regime		Secondo regime	
	Vero	CEM	SEM-<i>Gibbs</i>	CEM	SEM-<i>Gibbs</i>
ω_1	0.10	0.113 [0.046]	0.104 [0.041]	0.107 [0.023]	0.100 [0.016]
α_1	0.10	0.106 [0.028]	0.102 [0.028]	0.207 [0.017]	0.198 [0.015]
β_1	0.80	0.786 [0.055]	0.796 [0.052]	0.683 [0.048]	0.701 [0.028]
ω_2	0.30	0.310 [0.039]	0.313 [0.048]	0.217 [0.038]	0.210 [0.016]
α_2	0.20	0.205 [0.026]	0.202 [0.020]	0.205 [0.019]	0.205 [0.009]
β_2	0.60	0.592 [0.029]	0.588 [0.044]	0.470 [0.076]	0.480 [0.031]
π_1	0.20	0.266 [0.234]	0.229 [0.181]	0.248 [0.204]	0.213 [0.077]
π_2	0.80	0.734 [0.234]	0.771 [0.181]	0.752 [0.204]	0.787 [0.077]

Tabella 4.3: Risultati per lo Scenario 2: media [deviazione standard] dei parametri stimati mediante algoritmo CEM e SEM-*Gibbs*.

		Soglia			
	Vero	CEM		SEM-<i>Gibbs</i>	
λ	500	498.51 [3.54]		498.69 [3.47]	
		Primo regime		Secondo regime	
	Vero	CEM	SEM-<i>Gibbs</i>	CEM	SEM-<i>Gibbs</i>
ω_1	0.10	0.102 [0.011]	0.101 [0.012]	0.341 [0.047]	0.342 [0.048]
α_1	0.10	0.106 [0.014]	0.101 [0.010]	0.105 [0.009]	0.105 [0.009]
β_1	0.75	0.750 [0.022]	0.750 [0.024]	0.726 [0.028]	0.726 [0.029]
ω_2	0.15	0.163 [0.038]	0.152 [0.018]	0.154 [0.013]	0.154 [0.013]
α_2	0.15	0.149 [0.017]	0.151 [0.010]	0.203 [0.012]	0.203 [0.012]
β_2	0.70	0.692 [0.044]	0.698 [0.023]	0.591 [0.024]	0.591 [0.024]
π_1	0.50	0.555 [0.167]	0.496 [0.090]	0.483 [0.070]	0.485 [0.072]
π_2	0.50	0.445 [0.167]	0.504 [0.090]	0.517 [0.070]	0.515 [0.072]

Tabella 4.4: Risultati per lo Scenario 3: media [deviazione standard] dei parametri stimati mediante algoritmo CEM e SEM-*Gibbs*.

		Soglia			
	Vero	CEM		SEM- <i>Gibbs</i>	
λ	500	499.76 [6.13]		499.44 [5.99]	
		Primo regime		Secondo regime	
	Vero	CEM	SEM- <i>Gibbs</i>	CEM	SEM- <i>Gibbs</i>
ω_1	0.10	0.111 [0.021]	0.103 [0.019]	0.340 [0.067]	0.338 [0.067]
α_1	0.10	0.126 [0.023]	0.114 [0.024]	0.104 [0.015]	0.105 [0.014]
β_1	0.75	0.742 [0.038]	0.753 [0.032]	0.726 [0.041]	0.725 [0.042]
ω_2	0.15	0.201 [0.064]	0.171 [0.038]	0.153 [0.011]	0.154 [0.011]
α_2	0.15	0.151 [0.029]	0.151 [0.017]	0.202 [0.009]	0.203 [0.009]
β_2	0.70	0.660 [0.068]	0.680 [0.041]	0.594 [0.019]	0.593 [0.020]
π_1	0.20	0.391 [0.309]	0.395 [0.267]	0.194 [0.060]	0.195 [0.058]
π_2	0.80	0.609 [0.309]	0.605 [0.267]	0.806 [0.060]	0.805 [0.058]

Tabella 4.5: Risultati per lo Scenario 4: media [deviazione standard] dei parametri stimati mediante algoritmo CEM e SEM-*Gibbs*.

formance stabile e affidabile. Generalmente queste aziende sono *leader* nei loro settori e sono note per la loro capacità di operare in modo profittevole durante diversi cicli economici, inclusi i periodi di recessione.

L'EURO STOXX 50 è uno degli indici azionari più significativi in Europa, rappresentando le 50 principali società *blue-chip* dei paesi dell'Eurozona. Lanciato nel 1998 da *STOXX Limited*, una filiale del *Gruppo Deutsche Borse*, l'indice è ampiamente utilizzato come indicatore di riferimento per il mercato azionario europeo e per fondi e derivati. La composizione dell'indice è rivista annualmente a settembre e si basa su criteri di capitalizzazione di mercato, liquidità e settore, per garantire che rifletta adeguatamente la *performance* delle principali società dell'Eurozona. Aziende con una capitalizzazione di mercato più elevata e una maggiore quota di azioni disponibili al pubblico hanno un'influenza maggiore sull'andamento dell'indice, e data la presenza di aziende che operano a livello internazionale, l'indice costituisce un barometro importante non solo per la salute economica dell'Eurozona, ma anche per capire le tendenze dei mercati globali.

Dal punto di vista storico, l'indice ha mostrato una crescita significativa nei periodi di espansione economica, ma è anche vulnerabile a fluttuazioni in risposta a crisi economiche e politiche all'interno dell'Eurozona, come la crisi del debito sovrano europeo del 2010 e la recente pandemia di COVID-19.

Settore	Aziende
Aziende Cicliche	
Automotive:	<i>Volkswagen, Daimler, Stellantis, BMW, Ferrari, Mercedes-Benz Group</i>
Industria:	<i>Siemens, Schneider Electric, Airbus, Safran, Deutsche Post</i>
Tecnologia:	<i>ASML, SAP, Infineon Technologies, Prosus, Adyen, Nokia, Wolters Kluwer</i>
Energia:	<i>TotalEnergies, Eni</i>
Materiali:	<i>BASF, Saint-Gobain</i>
Finanza:	<i>Banco Santander, BNP Paribas, ING Group, Allianz, Axa, Intesa Sanpaolo, Deutsche Börse, Münchener Rück, Nordea Bank, BBVA</i>
Beni di consumo secondari:	<i>LVMH, Kering, Adidas, EssilorLuxottica, Inditex, Hermes International</i>
Aziende Difensive	
Sanità:	<i>Sanofi, Bayer</i>
Telecomunicazioni:	<i>Deutsche Telekom</i>
Utility:	<i>Enel, Iberdrola</i>
Beni di consumo primari:	<i>Danone, Anheuser-Busch InBev, Ahold Delhaize, L'Oréal, Pernod-Ricard</i>
Servizi pubblici:	<i>Vinci</i>

Tabella 4.6: Settori delle aziende, divise per cicliche e difensive.

Le società incluse nell'EURO STOXX 50 appartengono a diversi settori definiti secondo la classificazione settoriale globale (Global Industry Classification Standard, GICS, Inc. (2023)), uno standard ampiamente utilizzato per la suddivisione delle aziende in base alla loro principale attività produttiva. Dividendo le società in due macro-gruppi, è possibile distinguere le aziende operanti in settori ciclici da quelle in settori difensivi. Le aziende cicliche sono quelle che tendono ad avere migliori risultati durante le fasi di espansione economica e sono più vulnerabili nei periodi di recessione. Appartengono a settori che includono beni non necessari, quali industria, finanza e materie prime. Le aziende difensive, invece, tendono a mantenere *performance* stabili anche durante le fasi di recessione economica, poiché operano in settori che offrono beni e servizi essenziali. La Tabella 4.6 riporta in dettaglio la ripartizione delle società cicliche e difensive.

4.2.2 Dettagli di implementazione

Come discusso ampiamente nell'Introduzione di questa tesi, i cambiamenti strutturali nel processo della varianza rappresentano un fenomeno ricorrente nella maggior parte delle serie storiche finanziarie. La mancata considerazione di tali cambiamenti può compromettere significativamente l'efficacia delle strategie di gestione del rischio. A partire dalle serie storiche dei prezzi delle aziende incluse nell'indice EURO STOXX 50, si rivela quindi di grande interesse identificare *cluster* variabili nel tempo, capaci di catturare le complessità e le transizioni che caratterizzano questo mercato finanziario nelle sue diverse fasi cicliche.

Pertanto, il modello descritto nel Capitolo 3 è stato applicato a questi dati, con l'obiettivo di individuare la soglia attorno alla quale si verificano gradualmente i cambiamenti di regime, e analizzare le reazioni delle diverse serie storiche ad uno *shock* comune, raggruppandole in base al comportamento manifestato prima e dopo lo *shock* stesso.

A causa dell'assenza di dati per un determinato periodo temporale riguardante la società tecnologica *Prosus*, le analisi sono state condotte su un campione di $N = 49$ serie storiche di rendimenti logaritmici, ciascuna della lunghezza di $T = 785$ osservazioni. Al fine di rendere più plausibile l'ipotesi che, durante il periodo considerato, le serie storiche siano soggette ad un unico cambiamento di regime, si sono presi in esame gli intervalli temporali compresi tra il 1/1/2019 e il 31/12/2021.

Analogamente allo studio di simulazione, si è assunto per semplicità che il parametro di lisciamiento sia noto e pari a $\delta = 0.1$. Inoltre, con l'intento di esaminare se le aziende si distinguono in cicliche e difensive in base alla loro reazione a uno *shock* esterno, si suppone noto anche il numero di gruppi possibili per ciascun regime, ovvero $K = J = 2$. L'inizializzazione delle proporzioni dei gruppi per ciascun regime è stata posta pari a $(\pi_1^{(0)}, \pi_2^{(0)}) = (\rho_1^{(0)}, \rho_2^{(0)}) = (0.5, 0.5)$.

Per quanto riguarda, invece, l'inizializzazione dei parametri (ω, α, β) di ciascun gruppo di ogni regime, inizialmente si è proceduto in modo casuale. Tuttavia, ciò ha comportato che, al primo passo di *Expectation* di entrambi gli algoritmi, tutte le serie storiche fossero classificate in un unico gruppo, conducendo alla situazione di stallo descritta nel paragrafo 4.1.3. Pertanto, adottando un criterio più ragionato, i parametri sono stati inizializzati mediante una stima pilota di un modello GARCH(1,1) a media nulla effettuata, convenzionalmente, sulla prima metà e sulla seconda metà delle sequenze temporali di riferimento.

	Soglia			
	CEM		SEM- <i>Gibbs</i>	
λ	346 (29aprile2020)		347 (30aprile2020)	
	Primo regime		Secondo regime	
	CEM	SEM- <i>Gibbs</i>	CEM	SEM- <i>Gibbs</i>
π_1	0.531	0.531	0.816	0.837
π_2	0.469	0.469	0.184	0.163
ω_1	0.00001	0.00001	0.00001	0.00001
α_1	0.04406	0.04506	0.14793	0.14606
β_1	0.91340	0.91078	0.84541	0.84899
ω_2	0.00004	0.00004	0.00005	0.00006
α_2	0.07459	0.07498	0.11968	0.12161
β_2	0.84153	0.83985	0.79762	0.78676

Tabella 4.7: Stime dei parametri ottenute mediante gli algoritmi CEM e SEM-*Gibbs* nei dati EURO STOXX50.

4.2.3 Analisi dei risultati

I risultati ottenuti sono riportati nelle Tabelle 4.7 e 4.8, evidenziando una coerenza soddisfacente con le dinamiche reali osservate.

I parametri stimati relativi al modello GARCH(1,1) per ciascun gruppo di ogni regime sono coerenti con i valori tipicamente osservati nelle serie storiche finanziarie, ossia molto piccoli e positivi per ω , relativamente piccoli per α e più elevati per β . Non emergono differenze significative tra le stime effettuate utilizzando l'algoritmo CEM e quelle derivate dall'algoritmo SEM-*Gibbs*.

Un'attenzione particolare è da rivolgere alla stima della soglia del cambiamento di regime. Entrambi gli algoritmi hanno rilevato una transizione graduale del regime intorno agli ultimi giorni di aprile 2020, in linea con gli sviluppi della pandemia di COVID-19. Il cambiamento di regime non si è verificato esattamente nei primi giorni di marzo, coincidenti con l'inizio del *lockdown*, poiché non tutti gli Stati a cui si riferiscono gli strumenti finanziari hanno adottato misure restrittive in quel periodo. Inoltre, la reazione dei mercati finanziari è stata caratterizzata da un'evoluzione graduale piuttosto che da una risposta immediata.

La Tabella 4.8 riporta il raggruppamento degli strumenti finanziari dell'indice EURO STOXX 50 in base al gruppo di appartenenza del primo e del secondo regime, ottenuto mediante entrambi gli algoritmi CEM e SEM-*Gibbs*. I due algoritmi conducono allo stesso partizionamento,

		Secondo regime	
		Gruppo 1	Gruppo 2
Primo regime	Gruppo 1	AHOLD DELHAIZE, ALLIANZ, AXA, BASF, DANONE, DEUTSCHE BORSE, DEUTSCHE POST, DEUTSCHE TELEKOM, ENEL, ENI, ESSILORLUXOTTICA, FERRARI, HERMES INTERNATIONAL, IBERDROLA, INTESA SANPAOLO, L'OREAL, LVMH, MUENCHENER RUCK., NORDEA BANK, PERNOD-RICARD, SAINT-GOBAIN, SANOFI, SCHNEIDER ELECTRIC, TOTALENERGIES, WOLTERS KLUWER	
	Gruppo 2	ADYEN, AIRBUS, ASML HOLDING, BBV.ARGENTARIA, BANCO SANTANDER, BMW, BNP PARIBAS, INDITEX, ING GROUP, MERCEDES-BENZ GROUP, SAFRAN, SIEMENS, UNICREDIT (SEM), VINCI, VOLKSWAGEN	ADIDAS, ANHEUSER-BUSCH INBEV, BAYER, INFINEON TECHS., KERING, NOKIA, SAP, STELLANTIS, UNICREDIT (CEM)

Tabella 4.8: Strumenti finanziari dell'indice EURO STOXX 50, raggruppati in base al cluster di appartenenza nel primo e secondo regime dagli algoritmi CEM e SEM. L'unica differenza tra i due algoritmi è riferita all'azienda UNICREDIT.

con l'unica differenza che la società *Unicredit* viene classificata in due gruppi differenti sulla base del secondo regime.

Esaminando i settori di appartenenza delle diverse serie storiche, si osserva che, con l'eccezione dei settori dell'energia e dei materiali, la maggior parte delle aziende cicliche è assegnata al medesimo cluster. Tuttavia, anche all'interno dello stesso settore, si riscontrano aziende che, in risposta allo *shock* causato dalla pandemia di COVID-19, manifestano comportamenti divergenti nel secondo regime, seguendo alternativamente il comportamento descritto dai due gruppi. Il risultato è particolarmente interessante perché tale comportamento non è mai stato osservato in letteratura da un punto di vista empirico. Una situazione di minore variabilità si osserva, invece, nel caso delle aziende difensive. Ad eccezione della società *Vinci*, tutte le serie storiche reagiscono allo *shock* in maniera simile, mantenendo una coerenza con le altre che presentano il primo regime distribuito secondo lo stesso modello.

La differenza riscontrata tra le aziende cicliche e difensive è coerente con la definizione stessa di queste categorie. Le aziende difensive, essendo in grado di mantenere buone performance anche durante le fasi di recessione, tendono a mostrare una reazione relativamente uniforme in risposta a uno *shock*. Al contrario, le aziende cicliche, che sono mag-

giormente vulnerabili nei periodi di recessione (come nel caso dell'inizio della pandemia), manifestano reazioni più diversificate per ciascun *asset*. Meno intuitiva è la differente reazione delle società all'interno dello stesso settore; tuttavia si conferma che, pur presentando caratteristiche simili, le aziende che si comportano in modo simile per un periodo di tempo possono comportarsi in maniera diversa a seguito di uno *shock* esterno.

Riconosciuta l'importanza di considerare che serie storiche simili possono reagire in maniera differente, un naturale sviluppo dell'analisi condotta sulle aziende dell'indice EURO STOXX 50 consisterebbe nello studio di uno scenario di *asset allocation*. Questo permetterebbe di valutare l'importanza di tale assunzione in termini di gestione del rischio di portafoglio, motivo per cui tale analisi è lasciata come possibile importante sviluppo futuro.

Conclusioni

Da un punto di vista statistico, il presente lavoro ha esplorato l'applicazione di modelli a mistura finita per l'identificazione di gruppi di serie storiche finanziarie in presenza di cambiamenti di regime, con lo sviluppo di due varianti dell'algoritmo EM ai fini dell'inferenza sui parametri del modello. La metodologia proposta si dimostra uno strumento utile a tale scopo, combinando semplicità costruttiva, poiché derivabile da una diretta estensione degli strumenti propri dei modelli di *co-clustering*, con una notevole flessibilità modellistica, poiché applicabile a qualsiasi tipo di serie storica soggetta a cambiamenti di regime, siano essi gradualisti o più drastici.

I risultati dello studio di simulazione, condotto per valutare la robustezza e l'efficacia dell'approccio proposto e dei due algoritmi di stima CEM e SEM-*Gibbs*, hanno confermato la capacità complessiva di entrambi gli algoritmi di identificare i cambiamenti di regime, di stimare correttamente i parametri del modello e di distinguere tra i diversi gruppi di serie storiche. In particolare, l'algoritmo SEM-*Gibbs* ha mostrato una maggiore stabilità dei risultati, soprattutto in situazioni di elevata complessità, sebbene le differenze rispetto all'algoritmo CEM siano risultate complessivamente minime. Questo rafforza la validità delle metodologie adottate e suggerisce la loro applicabilità in contesti reali di analisi finanziaria.

Partendo dall'analisi delle serie storiche delle aziende facenti parte dell'indice EURO STOXX 50, è stato poi proposto un approccio basato sull'uso di modelli ST-GARCH(1,1) con due regimi e funzione di transizione logistica. L'analisi delle serie storiche delle aziende dell'indice EURO STOXX 50 ha rivelato come, nonostante le aziende appartenenti a settori ciclici e difensivi tendano a seguire comportamenti simili all'interno dello stesso regime, la reazione agli shock esterni, come quello causato dalla pandemia da COVID-19, può variare significativamente anche tra aziende dello stesso settore. Questa osservazione sottolinea l'importanza di considerare la possibilità che strumenti finanziari simili

possano divergere nei loro comportamenti in risposta a eventi di mercato significativi.

I risultati ottenuti confermano, nuovamente, la validità del modello proposto nell'identificare *cluster* dinamici e nel cogliere le complessità della volatilità nei mercati finanziari. Inoltre, l'efficacia dei due algoritmi di stima utilizzati è stata valutata mostrando una coerenza dei risultati con le caratteristiche attese per i parametri stimati nella realtà.

Questa relazione pone le basi per ulteriori sviluppi, come l'implementazione di scenari di *asset allocation* che possano sfruttare l'assunzione che serie storiche simili possano reagire in maniera differente a *shock* esterni, contribuendo così a migliorare le strategie di gestione del rischio di un portafoglio.

Dal punto di vista strettamente modellistico, si suggeriscono come possibili sviluppi futuri l'assunzione che la soglia di cambiamento di regime possa variare per ciascuna serie storica, nonché la combinazione di questa procedura con l'identificazione del modello della classe GARCH(1,1) più appropriato per descrivere l'andamento dei rendimenti nei diversi regimi. Inoltre, in questa analisi si è presupposto noto il numero di gruppi presenti in ciascun regime; tuttavia, potrebbe essere di grande interesse esplorare tecniche di selezione del modello adeguate al caso in esame. Infine, per avvicinarsi a un contesto più realistico, si potrebbe rilassare l'assunzione di indipendenza tra le serie, tenendo conto anche della struttura di correlazione.

Concludendo, le metodologie e i modelli discussi offrono interessanti prospettive di applicazione non solo in ambito finanziario, ma anche in altri contesti dove la gestione del rischio e l'identificazione di cambiamenti strutturali sono di cruciale importanza.

Bibliografia

- Azzalini A.; Menardi G. (2014). Density-based clustering of Gaussian mixture models. *Statistics and Computing*, **24**(5), 753–767.
- Bauwens L.; Preminger A.; Rombouts J. V. (2010). Theory and inference for a markov switching garch model. *The Econometrics Journal*, **13**(2), 218–244.
- Bollerslev T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, **31**(3), 307–327.
- Bouveyron C.; Celeux G.; Murphy T. B.; Raftery A. E. (2019). *Model-based clustering and classification for data science: with applications in R*, volume 50. Cambridge University Press.
- Box G. E.; Jenkins G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Broyden C. G.; Fletcher R.; Goldfarb D.; Shanno D. F. (1965). A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, **19**(92), 577–593.
- Cont R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, **1**(2), 223–236.
- Dempster A. P.; Laird N. M.; Rubin D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, **39**(1), 1–22.
- Deodhar N.; Ghosh J. (2010). Co-clustering through optimal transport. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pp. 101–112. SIAM.
- Diebold F. X. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Reviews*, **5**(1), 51–56.
- Engle R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pp. 987–1007.

- Engle R. F.; Bollerslev T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, **5**(1), 1–50.
- Geman S.; Geman D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**(6), 721–741.
- González-Rivera G. (1998). Smooth-transition garch models. *Studies in Nonlinear Dynamics & Econometrics*, **3**(2).
- Govaert G.; Nadif M. (2013). *Co-clustering: models, algorithms and applications*. John Wiley & Sons.
- Hanczar B.; Nadif M. (2010). Using random forest for co-clustering-based prediction of class membership: application to gene expression data. *Artificial Intelligence in Medicine*, **48**(1), 11–20.
- Hankin R. K. S. (2013). *Brobdignag: Very Large Numbers in R*. R package version 1.2-6.
- Harvey A. C.; Ruiz E.; Shephard N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies*, **61**(2), 247–264.
- Hubert L.; Arabie P. (1985). Comparing partitions. *Journal of Classification*, **2**(1), 193–218.
- Inc. M. (2023). Global industry classification standard (gics) - sector classifications. <https://www.msci.com/gics>. Accessed: 2024-08-28.
- Kaufman L.; Rousseeuw P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- Lamoureux C. G.; Lastrapes W. D. (1990). Heteroskedasticity in stock return data: Volume versus garch effects. *The journal of finance*, **45**(1), 221–229.
- MacQueen J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pp. 281–297. University of California Press.
- Markowitz H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press.
- McLachlan G. J.; Peel D. (2000). *Finite mixture models*, volume 299. John Wiley & Sons.

- Mikosch T.; Stărică C. (2004). Nonstationarities in financial time series, the long-range dependence, and the igarch effects. *Review of Economics and Statistics*, **86**(1), 378–390.
- Press W. H.; Teukolsky S. A.; Vetterling W. T.; Flannery B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3rd edizione.
- Priam R. *et al.* (2016). Simultaneous co-clustering of terms and documents for large-scale data sets. In *Proceedings of the 2016 IEEE International Conference on Big Data*, pp. 50–55. IEEE.
- Pérez-Espinoza A. J. (2019). Time-series clustering in r using the dtwclust package. *R Journal*, **11**(1), 22–43.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sardà-Espinoza A. (2019). dtwclust: Time series clustering along with optimizations for the dynamic time warping distance. *R Journal*, **11**(1), 22–43.
- Sewell M. (2011). Characterization of financial time series. *Rn*, **11**(01), 01.
- Shephard N. (1996). Statistical aspects of arch and stochastic volatility In *Likelihood, Time Series with Econometric and Other Applications: Essays in Honor of Sir David Cox*. A cura di Cox D., Hinkley D., Barndorff-Nielsen O., pp. 1–67. Chapman & Hall.
- Taylor S. J. (1982). Financial returns modelled by the product of two stochastic processes—a study of daily sugar prices 1961–79. *Time Series Analysis: Theory and Practice*, **1**, 203–226.
- Taylor S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, **4**(2), 183–204.
- Tsay R. S. (2010). *Analysis of Financial Time Series*, volume 543. John Wiley & Sons.