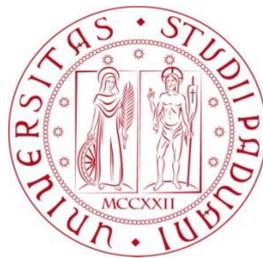


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**Uso delle componenti principali realizzate per
l'analisi del rischio sistemico**

Relatore Prof. Massimiliano Caporin
Dipartimento di Scienze Statistiche

Laureando: Marco Bergamo
Matricola N°2004662

Anno Accademico 2021/2022

Indice

1	Introduzione	9
2	Componenti principali per dati ad alta frequenza	13
2.1	Impostazione dell'ambiente teorico	13
2.1.1	Notazioni e definizioni	13
2.1.2	Autovalori e autovettori	15
2.1.3	Dinamiche della variabile	16
2.1.4	Analisi delle componenti principali ad alta frequenza	17
2.1.5	Funzione spettrale	18
2.2	Teoria asintotica e Stimatori	22
2.2.1	Assunzioni	22
2.2.2	Funzioni spettrali realizzate	23
2.2.3	Autovalori realizzati	26
2.2.4	Componenti Principali realizzate	28
2.2.5	Autovettori realizzati	29
2.2.6	"PCA" sulla matrice di covarianza integrata	30
3	Applicazione ai dati	31
3.1	Formule e descrizione dei dati	31
3.1.1	Componenti principali ad alta frequenza nel settore finanziario	31

3.1.2	Descrizione dei dati	33
3.2	Autovalori e autovettori realizzati sui dati	34
3.2.1	Autovalori realizzati	36
3.2.2	Autovettori realizzati	38
4	Analisi seriale delle componenti principali	43
4.1	Analisi degli autovalori	43
4.1.1	Modelli a effetti moltiplicativi	43
4.1.2	Applicazione alla serie degli autovalori	46
4.2	Analisi degli autovettori	50
5	Spillover	57
5.1	Metodologia	59
5.2	Risultati	61
6	Conclusioni	75
A	Titoli utilizzati	77
B	Output di R	83
	Bibliografia	89

Elenco delle figure

3.1	Percentuale della varianza spiegata da ciascuna delle prime tre componenti principali (rispettivamente in nero, rosso e verde) e somma della varianza spiegata dalle prime tre componenti principali in blu.	36
3.2	Media annuale della varianza spiegata dalle prime tre componenti principali e dalla loro somma	38
3.3	Media annuale del contributo percentuale di ogni settore alla prima componente principale	39
3.4	Composizione della prima componente principale in riferimento solo al settore finanziario, scomposto in banche, assicurazioni e altri servizi finanziari.	40
3.5	Dispersione del contributo percentuale per settore alla prima componente principale per anno.	42
4.1	Densità dell'autovalore della prima componente principale	47
4.2	ACF e PACF della serie dell'autovalore della prima componente principale	47
4.3	Serie dell'autovalore della prima componente principale dal 2004 al 2021 (in arancio) e previsione della serie con modello MEM(1,1) un passo avanti (in blu)	48

4.4	Boxplot dei parametri stimati del MEM(1,1) su finestre mobili nel periodo considerato	49
4.5	Serie dei parametri del modello MEM(1,1) stimati su finestre mobili	50
4.6	Nell'ordine: l'ACF della serie del primo elemento dell'autovettore, la corrispondente PACF e la CCF con il secondo elemento del primo autovettore	52
4.7	Nell'ordine: l'ACF della serie del secondo elemento del primo autovettore, la corrispondente PACF e la CCF con il primo elemento del primo autovettore	53
4.8	Nell'ordine: l'ACF della serie del terzo elemento del primo autovettore, la corrispondente PACF e la CCF con il primo elemento del primo autovettore	54
4.9	Nell'ordine: l'ACF della serie del quarto elemento del primo autovettore, la corrispondente PACF e la CCF con il primo elemento del primo autovettore	55
4.10	Box-Plot di tutti i parametri stimati dal modello VAR sugli autovettori	56
4.11	Box-Plot dei soli parametri significativi stimati dal modello VAR sugli autovettori	56
5.1	<i>Net pairwise spillovers</i> tra la radice di EIGV1 e il VIX, lo SLFSI, il NFCI e l'EPU.	63
5.2	<i>Net pairwise spillovers</i> tra la radice di EIGV2 e il VIX, lo SLFSI, il NFCI e l'EPU.	67
5.3	<i>Net pairwise spillovers</i> tra la radice di EIGV3 e il VIX, lo SLFSI, il NFCI e l'EPU.	69
5.4	Grafico dell'autocorrelazione e autocorrelazione parziale dell'indice VIX.	71

5.5	Grafici dell'autocorrelazione dei residui del modello ARMA(1,0) applicato al VIX senza regressori.	73
5.6	Grafici dell'autocorrelazione dei residui del modello ARMA(1,0) applicato al VIX utilizzando come ulteriore regressore il primo autovalore ritardato.	73

Elenco delle tabelle

3.1	Tabella degli indici riassuntivi dei rendimenti nel periodo dal 10 gennaio 2003 al 23 febbraio 2021, suddivisi prima e dopo la data di fallimento di Lehman Brothers, 16 settembre 2008.	35
4.1	Indici riassuntivi della serie dell'autovalore della prima componente principale.	46
5.1	<i>Spillover</i> della volatilità tra il primo autovalore (EIGV1) e gli indici di mercato per tutto il periodo.	62
5.2	<i>Spillover</i> della volatilità tra il secondo autovalore (EIGV2) e gli indici di mercato per tutto il periodo.	65
5.3	<i>Spillover</i> della volatilità tra il terzo autovalore (EIGV3) e gli indici di mercato per tutto il periodo.	68
5.4	Parametri stimati, sul primo autovalore, del modello MEM(1,1) con tutti gli indici come regressori al primo ritardo.	70
5.5	Parametri stimati, sul secondo autovalore, del modello MEM(1,1) con tutti gli indici come regressori al primo ritardo	70
5.6	Parametri stimati, sul terzo autovalore, del modello MEM(1,1) con tutti gli indici come regressori al primo ritardo	71
5.7	Output del modello ARMA(1,0) applicato al VIX.	72

5.8	Output del modello ARMA(1,0) applicato al VIX con EIGV1 utilizzato come regressore al primo ritardo.	74
A.1	Lista dei titoli utilizzati nello studio	82

Capitolo 1

Introduzione

La recente crisi finanziaria ha alimentato la ricerca di misure del rischio sistemico. Vari articoli hanno proposto differenti misure di rischio sistemico, come Huang, Zhou e Zhu (2009), Segoviano Basurto e Goodhart (2009), Acharya et al. (2016), Allen, Bali e Tang (2012), Zhou (2010), Brownlees e Engle (2017), Benoit et al. (2016), Billio et al. (2012), Biais et al. (2012), Silva, Kimura e Sobreiro (2017) e l'indagine di Benoit et al. (2017).

Il rischio sistemico presenta più dimensioni; mostra uno schema che varia nel tempo e una struttura cross-sezionale. Gli shock finanziari sono endogeneamente guidati (derivati dal comportamento codipendente e dalle reazioni a catena delle istituzioni finanziarie stesse). Le dimensioni variabili nel tempo e cross-sezionali si sommano assieme durante lo sviluppo di una crisi. Da una parte, il rischio sistemico è correlato con la pro-ciclicità del comportamento degli agenti, è dinamico per natura, e può essere trovato solo attraverso l'osservazione su lunghi periodi; questa dimensione del rischio sistemico è chiamata, appunto, tempo dipendente (Kiyotaki e Moore (1997), Borio e Lowe (2002), Borio, Lowe et al. (2002), Borio e Drehmann (2009), Brunnermeier (2001), Borio (2013)). D'altro canto, il rischio sistemico è correlato anche con la struttura delle interconnessioni nel sistema. Questo aspetto

del rischio sistemico non è adeguatamente catturato dalla dimensione temporale degli aggregati finanziari e per essere descritto richiede una specifica struttura delle interconnessioni finanziarie in ogni momento temporale, che è la dimensione cross-sezionale del rischio sistemico.

Nel seguito si analizzeranno entrambe le dimensioni, guardando all'identificazione del rischio sistemico e dei canali attraverso i quali si trasferisce. La sfida è identificare quando una rete finanziaria è potenzialmente vulnerabile e la natura dell'evento che può innescare un problema. Per fare questo, sono stati considerati due elementi chiave: la combinazione dell'econometria ad alta frequenza e un'analisi vettoriale ad alta dimensionalità. Questi strumenti permettono di considerare la variazione temporale nella struttura fattoriale delle interconnessioni di un grande insieme di *panel data*.

La disponibilità di dati ad alta frequenza (HF) fornisce l'opportunità di analizzare brevi periodi di tempo indipendentemente e di identificare con meno incertezza statistica la natura temporale e la forza (o vulnerabilità) di legami diretti e indiretti nella rete finanziaria.

Lo studio lavora direttamente su grandi sezioni di singoli titoli, e l'analisi delle componenti principali (PCA) risulta esser una delle tecniche più popolari quando si analizzano dataset di grandi dimensioni. Fondamentalmente questo approccio mira a comprimere quanto più contenuto informativo possibile in un numero di componenti principali ridotto, che riassume una gran parte della variazione dei dati. Bai (2003) ha sviluppato una teoria inferenziale per modelli fattoriali su grandi sezioni trasversali e orizzonti lunghi basata sulla PCA. È anche largamente accettato nella letteratura finanziaria che i mercati possono essere descritti da un piccolo numero di fattori derivati dalla PCA. Ad esempio Litterman e Scheinkman (1991) usano la PCA per identificare una struttura a tre fattori nella struttura dei rendimenti dei bond. Egloff, Leippold e Wu (2010) identificano due fattori di volatilità

sulla PCA dei *variance swap*. Vari autori hanno introdotto la PCA per riassumere il contenuto informativo di differenti misure di rischio, si vedano Giglio, Kelly e Pruitt (2016), Nucera et al. (2016), Caporin et al. (2022). Rodríguez-Moreno e Peña (2013) hanno usato la PCA classica per confrontare il rischio sistemico di mercato mensile a livello macro e micro, evidenziando che le misure basate sul *Credit Default Swap* hanno prestazioni migliori rispetto ai tassi interbancari dei prezzi.

Il lavoro presente si propone di riprodurre e ampliare lo studio fatto da Jiménez-Martin, Caporin e Garcia-Jorcano (2022), ricostruendo la misura di rischio sviluppata su dati diversi e ampliandone l'orizzonte conducendo un'analisi delle correlazioni seriali delle quantità sviluppate nell'articolo. Infatti nel caso classico la PCA soffre della maledizione della dimensionalità. Il numero di parametri aumenta molto più velocemente rispetto alla grandezza cross-sezionale, richiedendo anni di serie per la stima, con problemi di distorsione dovuti alla *sopravvivenza* delle aziende quotate, non stazionarietà e consistenza dei parametri. Due articoli, Aït-Sahalia e Xiu (2017) e Aït-Sahalia e Xiu (2019), sviluppano una metodologia per utilizzare la PCA su dati ad alta frequenza (HF), i quali presentano grandi dimensioni sia cross-sezionali che temporali. Ci si aspetta che la HF PCA (*High Frequency Principal Component Analysis*) migliori l'approccio classico, infatti è capace di:

- (i) occuparsi della crescita del numero di parametri quando la dimensione cross-sezionale aumenta;
- (ii) eliminare le condizioni di stazionarietà, permettendo alla volatilità di cambiare nel tempo o permettendo salti nei log-prezzi;
- (iii) infine, è in grado di catturare potenziali relazioni non lineari grazie alla stima locale delle componenti principali.

Nel seguito si procederà a combinare dati ad alta frequenza e analisi fattoriali usando la PCA per approssimare il contenuto informativo di un dataset ad alta frequenza. Per prima cosa si otterrà una misura della percentuale spiegata di rischio sistemico stimando e analizzando il comportamento dinamico degli autovalori e degli autovettori usando il modello sviluppato da Aït-Sahalia e Xiu (2019). Quindi si procederà con un'analisi seriale delle componenti stimate, per mettere in evidenza possibili pattern degli stimatori. Inoltre è d'interesse condurre un'analisi della trasmissione degli shock da (e verso) il mercato a (e da) altri indicatori di rischio sistemico, attraverso la metodologia sviluppata da Diebold e Yilmaz (2012).

Le analisi empiriche sono state fatte su dati rilevati a un minuto di titoli quotati al NYSE e al NASDAQ per gli anni dal 2003 al 2021, mostrando che autovalori realizzati elevati, e quindi picchi nella varianza spiegata dalle prime tre componenti principali, che indicano un aumento del livello di co-movimento tra i titoli e del grado di fragilità del sistema, coincidono con alcuni importanti eventi associati alla politica monetaria statunitense e europea o ad altri eventi di importanza globale.

Il resto del lavoro è organizzato come segue. Nel Capitolo 2 è descritto il contenuto teorico sviluppato da Aït-Sahalia e Xiu (2019) per costruire gli stimatori. Nel Capitolo 3 è presente l'analisi empirica condotta sul dataset della metodologia presa in considerazione. Il Capitolo 4 riporta l'analisi seriale degli stimatori e infine il Capitolo 5 riporta l'analisi dello *spillover* tra le serie degli autovalori stimati e alcuni indici macroeconomici.

Capitolo 2

Componenti principali per dati ad alta frequenza

La costruzione delle componenti principali ad alta frequenza descritta nel seguito è stata sviluppata nei lavori di Aït-Sahalia e Xiu (2017) e Aït-Sahalia e Xiu (2019). Rispettivamente i due lavori sviluppano uno stimatore per il numero di fattori comuni quando la frequenza di campionamento e il numero delle variabili aumenta e una metodologia per condurre l'analisi delle componenti principali ad alta frequenza.

2.1 Impostazione dell'ambiente teorico

2.1.1 Notazioni e definizioni

Nel seguito si denoterà con \mathbb{R}_d lo spazio euclideo d -dimensionale, e il suo sottoinsieme \mathbb{R}_d^+ contiene tutti i vettori reali non negativi. Sia e^k il vettore unitario in \mathbb{R}_d^+ , con k -esimo elemento pari a uno, $1 = \sum_{k=1}^d e^k$ e sia \mathbb{I} la matrice identità. \mathcal{M}_d denota lo spazio euclideo delle matrici simmetriche $d \times d$ a valori reali. \mathcal{M}_d^+ è

il sottoinsieme di \mathcal{M}_d che include tutte le matrici semidefinite positive. Si userà la notazione \mathcal{M}_d^{++} per indicare lo spazio delle matrici definite positive. Lo spazio euclideo \mathcal{M}_d è dotato del prodotto interno $\langle A, B \rangle = \text{Tr}(A, B)$. Si userà $\|\cdot\|$ per indicare la norma euclidea per vettori o matrici, e l'apice "+" per denotare l'inversione di Moore-Penrose di una matrice reale.

Tutti i vettori sono vettori colonna. Il trasposto di una matrice A è indicato come A^\top e l'operatore "diag" indica il vettore x , diagonale di una matrice.

Sia f una funzione da \mathbb{R}_d^+ a \mathbb{R} . Il gradiente di f è indicato come ∂f e la matrice hessiana come $\partial^2 f$. La derivata di una funzione matriciale $F : \mathcal{M}_d^+ \rightarrow \mathbb{R}$ è indicata da $\partial F \in \mathcal{M}_d$, con ogni elemento scritto come $\partial_{ij} F$, per $1 \leq i, j \leq d$. La funzione di derivazione è definita nel senso usuale, tale che per ogni $A \in \mathcal{M}_d^+$, $\partial_{ij} A = J_{ij}$, dove J_{ij} è una matrice di zeri con il solo elemento (i, j) pari a 1. La matrice hessiana di F è scritta come $\partial^2 F$, dove per ogni elemento si fa riferimento come $\partial_{jk,lm}^2 F$, per $1 \leq j, k, l, m \leq d$. Si utilizza ∂^k per la derivata di ordine k e $\delta_{i,j}$ per indicare la funzione delta di Kronecker che restituisce 1 se $i = j$, 0 altrimenti. Una funzione f è detta di Lipschitz se esiste una costante K tale che $|f(x+h) - f(x)| \leq K\|h\|$, K è una costante generica che può cambiare da riga a riga. Una funzione con k derivate continue è indicata come una funzione C^k .

I dati sono campionati in modo discreto ogni Δ_n unità di tempo. Tutti i limiti sono presi come $\Delta_n \rightarrow 0$. " $\xrightarrow{\text{u.c.p.}}$ " indica uniformità sulla probabilità dei compatti, " \xrightarrow{P} " indica la convergenza in probabilità. Si usa " $\xrightarrow{\mathcal{L}^{-s}}$ " per indicare la legge della convergenza stabile. Si scrive $a_n \asymp b_n$ se per qualche $c \geq 1$, $b_n/c \leq a_n \leq cb_n$ per ogni n . Infine $[\cdot, \cdot]$ indica la covarianza quadratica tra Itô semimartingale, e $[\cdot, \cdot]^c$ la sua parte continua.

2.1.2 Autovalori e autovettori

Si definiscono ora alcuni risultati preliminari riguardanti autovalori e autovettori.

Per ogni vettore $x \in \mathbb{R}_d^+$, \bar{x} indica il vettore con gli stessi elementi di x , ordinati in ordine non decrescente. Si usa $\overline{\mathbb{R}_d^+}$ per indicare il sottoinsieme di \mathbb{R}_d^+ contenente i vettori x che soddisfano $x_1 \geq x_2 \geq \dots \geq x_d \geq 0$.

Per convenzione, per ogni $x \in \overline{\mathbb{R}_d^+}$ si scrive

$$x_1 = \dots = x_{g_1} \geq x_{g_1+1} = \dots = x_{g_2} \geq \dots x_{g_{r-1}} \geq \dots x_{g_{r-1}+1} = \dots = x_{g_r} \geq 0, \quad (2.1)$$

dove $g_r = d$ e r è il numero di elementi distinti. g_1, g_2, \dots, g_r dipende da x . Si definisce una corrispondente partizione di indice $I_j = g_{j-1} + 1, g_{j-1} + 2, \dots, g_j$ per $j = 1, 2, \dots, r$.

Per ogni $A \in \mathcal{M}_d^+$, $\lambda(A) = (\lambda_1(A), \lambda_2(A), \dots, \lambda_d(A))^\top$ è il vettore dei suoi autovalori in ordine non crescente. Questa notazione permette di considerare λ come una mappatura da \mathcal{M}_d^+ a $\overline{\mathbb{R}_d^+}$. Un risultato importante, Tao (2012), stabilisce che la continuità di λ è:

Lemma 1. $\lambda : \mathcal{M}_d^+ \rightarrow \overline{\mathbb{R}_d^+}$ è Lipschitz.

Associato a ogni autovalore λ_g di $A \in \mathcal{M}_d^+$, si indica con γ_g l'autovettore che soddisfa $A\gamma_g = \lambda_g\gamma_g$ e $\gamma_g^\top\gamma_g = 1$. L'autovettore, tranne che per il segno, è univocamente definito quando λ_g è semplice. In tal caso senza perdita di generalità, si richiede che il primo elemento non nullo dell'autovettore sia positivo. In presenza di autovalori ripetuti l'autovettore è determinato da una trasformazione ortogonale. In ogni caso si possono scegliere autovettori tali che per ogni $g \neq h$, $\gamma_g^\top\gamma_h = 0$, si veda Anderson (1962). Quando λ_g è una radice semplice, si indica γ_g come un'altra funzione vettore-valore di A . In tal caso sia $\lambda_g(\cdot)$ e $\gamma_g(\cdot)$ sono lisciatori infinitesimali di A , si veda Magnus e Neudecker (1999).

Lemma 2. *Si supponga λ_g sia una radice semplice di $A \in \mathcal{M}_d^+$, allora $\lambda_g: \mathcal{M}_d^+ \rightarrow \overline{\mathbb{R}}_d^+$ e $\gamma_g: \mathcal{M}_d^+ \rightarrow \overline{\mathbb{R}}_d^+$ sono C^∞ su A . Inoltre, si ha*

$$\partial_{jk}\lambda_g(A) = \gamma_{gj}(A)\gamma_{gk}(A) \quad e \quad \partial_{jk}\gamma_g(A) = (\lambda_g\mathbb{I} - A)_{\cdot,j}^+ \gamma_{gk}(A),$$

dove γ_{gk} è il k -esimo elemento di γ_g . Se anche tutti gli autovalori di A sono semplici, con $(\gamma_1, \gamma_2, \dots, \gamma_g)$ corrispondenti autovettori, allora

$$\begin{aligned} \partial_{jk,lm}^2 \gamma_{gh} = & - \sum_{p \neq g} \frac{1}{(\lambda_g - \lambda_p)^2} (\gamma_{gl}\gamma_{gm}\gamma_{ph}\gamma_{pj}\gamma_{gk} - \gamma_{pl}\gamma_{pm}\gamma_{ph}\gamma_{pj}\gamma_{gk}) \\ & + \sum_{p \neq g} \sum_{q \neq p} \frac{1}{(\lambda_g - \lambda_p)(\lambda_p - \lambda_q)} \gamma_{ql}\gamma_{pm}\gamma_{qh}\gamma_{pj}\gamma_{gk} \\ & + \sum_{p \neq g} \sum_{q \neq p} \frac{1}{(\lambda_g - \lambda_p)(\lambda_p - \lambda_q)} \gamma_{ql}\gamma_{pm}\gamma_{qj}\gamma_{ph}\gamma_{gk} \\ & + \sum_{p \neq g} \sum_{q \neq g} \frac{1}{(\lambda_g - \lambda_p)(\lambda_p - \lambda_q)} \gamma_{qk}\gamma_{ql}\gamma_{gm}\gamma_{ph}\gamma_{pj}. \end{aligned}$$

In generale, mentre gli autovalori sono sempre una funzione continua, gli autovettori associati a radici ripetute non sono necessariamente continui. Nel seguito si considera solamente la stima di autovettori associati ad autovalori semplici.

2.1.3 Dinamiche della variabile

Si analizza una Itô semimartingala d -dimensionale, definita su uno spazio di probabilità filtrato $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t>0}, \mathbb{P})$ con la seguente rappresentazione di Grigelionis:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + (\delta 1_{\|\delta\| \leq 1}) * (\mu - \nu)_t + (\delta 1_{\|\delta\| > 1}) * \mu_t \quad (2.2)$$

dove W è un moto Browniano d -dimensionale, μ una misura casuale di Poisson su $\mathbb{R}^+ \times \mathbb{R}_d$ con la compensazione $\nu(dt, dx) = dt \otimes \bar{\nu}(dx)$, e $\bar{\nu}$ una σ -misura finita. Il modello così strutturato ammette la presenza di eventuali salti nella volatilità e non impone restrizioni sulla dipendenza tra le varie componenti del processo,

permettendo, quindi, di accomodare l'effetto leva (Black (1976)). Come dimostrato da Delbaen e Schachermayer (1994), una semimartingala è appropriata per modellare i rendimenti, in quanto esclude la possibilità di arbitraggio.

Il processo di volatilità σ_s è càdlàg, $c_s = (\sigma\sigma^\top)_s \in \mathcal{M}_d^+$ per ogni $0 \leq s \leq t$. Si indicano d autovettori non negativi di c_s da $\lambda_{1,s} \geq \lambda_{2,s} \geq \dots \geq \lambda_{d,s}$, riassumendoli nel vettore λ_s . Si scrive $\lambda_s = \lambda(c_s)$, indicando con $\lambda(\cdot)$ come una funzione di c_s . Per il Lemma 1, $\lambda(\cdot)$ è càdlàg.

2.1.4 Analisi delle componenti principali ad alta frequenza

Simile alla classica PCA, l'analisi delle componenti principali in questo ambiente consiste nel cercare ripetutamente per combinazioni lineari istantanee di X , cioè le componenti principali, quale massimizza certe misure di variazione, mentre è ortogonale alle componenti principali già costruite, ad ogni tempo tra 0 e t . Al contrario del caso classico, dove si massimizza la varianza della combinazione, Anderson (1958), il criterio qui utilizzato è la parte continua della variazione quadratica.

Lemma 3. *Si supponga che X sia un processo vettore-valore d -dimensionale descritto in (2). Allora esiste una sequenza di $\{\lambda_{g,s}, \gamma_{g,s}\}_{1 \leq g \leq d, 0 \leq s \leq t}$, tale che*

$$c_s \gamma_{g,s} = \lambda_{g,s} \gamma_{g,s}, \quad \gamma_{g,s}^\top \gamma_{g,s} = 1 \quad e \quad \gamma_{h,s}^\top c_s \gamma_{g,s} = 0,$$

dove $\lambda_{1,s} \geq \lambda_{2,s} \geq \dots \geq \lambda_{d,s} \geq 0$. Inoltre per ogni càdlàg e processo vettore-valore γ_s , tale che $\gamma_s^\top \gamma_s = 1$ e $\gamma_s^\top c_s \gamma_{h,s} = 0$, $1 \leq h \leq g-1$,

$$\int_0^u \lambda_{g,s} ds \geq \left[\int_0^u \gamma_{s-}^\top dX_s, \int_0^u \gamma_{s-}^\top dX_s \right]^c, \quad \text{per ogni } 0 \leq u \leq t.$$

Quando $\lambda_{g,s}$ è una radice semplice di c_s tra 0 e t , allora $\gamma_{g,s}$ è un adattamento di un processo càdlàg, a causa della continuità di $\gamma_g(\cdot)$ per il Lemma2, in modo che si possa costruire la corrispondente componente principale $\int_0^t \gamma_{g,s-}^\top dX_s$. Si può

anche investigare la media dei loading della componente principale, ovvero l'autovettore realizzato, che è $\int_0^t \gamma_{g,s} ds$. Essendo la definizione di queste quantità basata sugli integrali, esse sono indicate come autovalori integrati, autovettori integrati e componenti principali integrate.

Di conseguenza si è portati a considerare l'inferenza per $\int_0^t \lambda(c_s) ds$, per il quale la differenziabilità di $\lambda(\cdot)$ è fondamentale: si parte dalla convergenza di uno stimatore \hat{c}_s per c_s , dalla quale occorre un metodo delta per fare inferenza sulla convergenza dell'autovettore integrato stimato. Un autovalore semplice è C^∞ -differenziabile, ma per autovalori ripetuti non è necessariamente vero. Per questa ragione è necessario introdurre la funzione spettrale.

2.1.5 Funzione spettrale

Una funzione a valori reali F definita sul sottoinsieme \mathcal{M}_d^+ è detta funzione spettrale, Friedland (1981), se per ogni matrice ortogonale O in \mathcal{M}_d e X in \mathcal{M}_d^+ , $F(X) = F(O^\top X O)$. Si descrivono gli insiemi in \mathbb{R}_d^+ e le funzioni da \mathbb{R}_d^+ a \mathbb{R}^+ come asimmetriche se sono invarianti sotto permutazioni delle coordinate. Quindi, per ogni funzione simmetrica f con un dominio simmetrico in \mathbb{R}_d^+ abbiamo $f(x) = f(Px)$ per ogni matrice di permutazione $P \in \mathcal{M}_d$. Associata a ogni funzione spettrale F , si definisce la funzione f su \mathbb{R}_d^+ , tale che $f(x) = F(\text{diag}(x))$. Essendo le matrici di permutazione ortogonali, f è simmetrica, e $f \circ \lambda = F$, dove \circ denota la funzione composta.

Diventa necessario differenziare la funzione spettrale F . Si introduce una funzione matriciale associata alla corrispondente funzione simmetrica f di F , che, per

ogni $x \in \overline{\mathbb{R}}_d^+$ nella forma definita in 2.1, è data da

$$\mathcal{A}_{p,q}^f(x) = \begin{cases} 0 & \text{se } p = q; \\ \partial_{pp}^2 f(x) - \partial_{qq}^2 f(x) & \text{se } p \neq q, \text{ e } p, q \in I_l; \\ (\partial_p f(x) - \partial_q f(x)) / (x_p - x_q) & \text{altrimenti} \end{cases}$$

per qualche $l = 1, 2, \dots, r$.

Il seguente lemma riassume alcuni risultati riguardanti la differenziabilità continua e la convessità delle funzioni spettrali, che verranno usate nel seguito:

Lemma 4. *La funzione simmetrica f è doppiamente continuamente differenziabile in un punto $\lambda(A) \in \overline{\mathbb{R}}_d^+$ se e solo se la funzione spettrale $F = f \circ \lambda$ è doppiamente continuamente differenziabile nel punto $A \in \mathcal{M}_d^+$. Il gradiente e la matrice hessiana sono date come:*

$$\begin{aligned} \partial_{jk}(f \circ \lambda)(A) &= \sum_{p=1}^d O_{pj} \partial_p f(\lambda(A)) O_{pk}, \\ \partial_{jk,lm}^2(f \circ \lambda)(A) &= \sum_{p,q=1}^d \partial_{pq}^2 f(\lambda(A)) O_{pl} O_{pm} O_{qj} O_{qk} \\ &\quad + \sum_{p,q=1}^d \mathcal{A}_{pq}^f(\lambda(A)) O_{pl} O_{pj} O_{qk} O_{qm}, \end{aligned}$$

dove O è una qualsiasi matrice ortogonale che soddisfa $A = O^\top \text{diag}(\lambda(A))O$. Più in generale, f è una funzione C^k su $\lambda(A)$ se e solo se F è C^k su A , per ogni $k = 0, 1, \dots, \infty$. Inoltre f è una funzione convessa se e solo se F è convessa.

Si può notare che sia autovalori semplici che ripetuti possono essere visti come casi speciali di funzioni spettrali.

Esempio 2.1.1 (Autovalore semplice). Si supponga che il k -esimo autovalore di $A \in \mathcal{M}_d^+$ sia semplice, cioè, $\lambda_{k-1}(A) > \lambda_k(A) > \lambda_{k+1}(A)$. Si definisce per ogni $x \in \mathbb{R}_d^+$,

$$f(x) = \text{il } k\text{-esimo valore più grande in } x = \bar{x}_k.$$

Apparentemente f è una funzione simmetrica ed è C^∞ su ogni punto $y \in \mathbb{R}_d^+$, con $y_{k-1} > y_k > y_{k+1}$. Infatti, $\partial f(y) = e^k$ e $\partial^l f(y) = 0$ per $l \geq 2$. Per il Lemma 4, $\lambda_k(A) = (f \circ \lambda)(A)$ è C^∞ su A .

Esempio 2.1.2 (Autovalore ripetuto). Si supponga che gli autovalori di $A \in \mathcal{M}_d^+$ soddisfino

$$\lambda_{g_{l-1}}(A) > \lambda_{g_{l-1}+1}(A) \geq \dots \geq \lambda_{g_l}(A) > \lambda_{g_{l+1}}(A),$$

per un qualche $1 \leq g_{l-1} \leq g_l \leq d$. Per convenzione, quando $g_l = d$, l'ultimo " $>$ " non si usa. Si consideri la seguente funzione f , stimata su $x \in \mathbb{R}_d^+$, che è data da

$$f(x) = \frac{1}{g_l - g_{l-1}} \sum_{j=g_{l-1}+1}^{g_l} \bar{x}_j.$$

Si può verificare che f è simmetrica e C^∞ su ogni punto $y \in \overline{\mathbb{R}_d^+}$ che soddisfa $y_{g_{l-1}} \geq y_{g_{l-1}+1} \geq \dots \geq y_{g_l} \geq y_{g_{l+1}}$. Inoltre, $\partial f(y) = \frac{1}{g_l - g_{l-1}} \sum_{k=g_{l-1}+1}^{g_l} e^k$, e $\partial^l f(y) = 0$, per ogni $l \geq 2$. Come risulta dal Lemma 4, la corrispondente funzione spettrale

$$F(A) = (f \circ \lambda)(A) = \frac{1}{g_l - g_{l-1}} \sum_{j=g_{l-1}+1}^{g_l} \lambda_j(A),$$

è C^∞ su A . Nel caso in cui

$$\lambda_{g_{l-1}}(A) > \lambda_{g_{l-1}+1}(A) = \dots = \lambda_{g_l}(A) > \lambda_{g_{l+1}}(A),$$

ovvero vi sia un autovalore ripetuto, e $F(A) = \lambda_{g_{l-1}+1}(A) = \dots = \lambda_{g_l}(A)$. Di contro, $\lambda_j(A)$ è non differenziabile, per ogni $g_{l-1} + 1 \leq j \leq g_l$.

Esempio 2.1.3 (Traccia e determinante). Per ogni $x \in \mathbb{R}_d^+$, si definiscono

$$f_1(x) = \sum_{j=1}^d x_j \quad \text{e} \quad f_2(x) = \prod_{j=1}^d x_j.$$

Sia f_1 che f_2 sono simmetriche e C^∞ . Quindi, per ogni $A \in \mathcal{M}_d^+$, $\text{Tr}(A) = (f_1 \cdot \lambda)(A)$ e $\det(A) = (f_2 \cdot \lambda)(A)$ sono C^∞ su A .

Gli esempi precedenti mettono in evidenza che gli autovalori sono casi speciali di funzioni spettarali, siano essi semplici o ripetuti. Il Lemma 4 unisce la differenziabilità di una funzione spettrale F , che è necessaria per fare inferenza statistica, a quella della funzione simmetrica f associata. Il punto fondamentale è che la differenziabilità di f è più facile da stabilire.

Si procede dimostrando un risultato utile per caratterizzare la topologia dell'insieme di matrici con strutture speciali degli autovalori. Il dominio delle funzioni spettarali che verrà considerato sarà ristretto a questo insieme dove le funzioni spettarali sono lisce.

Lemma 5. *Per ogni $1 \leq g_1 < g_2 < \dots < g_r \leq d$, l'insieme*

$$\mathcal{M}(g_1, g_2, \dots, g_r) = \{A \in \mathcal{M}_d^{++} | \lambda_{g_l} > \lambda_{g_{l+1}}, \text{ per ogni } l = 1, 2, \dots, r-1\} \quad (2.3)$$

è denso e aperto in \mathcal{M}_d^{++} . In particolare l'insieme delle matrici definite positive con autovalori distinti, $\mathcal{M}(1, 2, \dots, d)$ è denso e aperto in \mathcal{M}_d^{++} .

Si aggiungono alcune notazioni. Si introduce un sottoinsieme aperto e simmetrico di $\overline{\mathbb{R}}^+ / \{0\}$, l'immagine di $\mathcal{M}(g_1, g_2, \dots, g_r)$ sotto $\lambda(\cdot)$:

$$\mathcal{D}(g_1, g_2, \dots, g_r) = \{x \in \overline{\mathbb{R}}_d^+ / 0 | \bar{x}_{g_l} > \bar{x}_{g_{l+1}}, \text{ per ogni } l = 1, 2, \dots, r-1\} \quad (2.4)$$

Quindi si introduce il sottoinsieme $\mathcal{M}(g_1, g_2, \dots, g_r)$, dove la matrice di covarianza istantanea c_t prenderà valori.

$$\begin{aligned} \mathcal{M}^*(g_1, g_2, \dots, g_r) &= \{A \in \mathcal{M}_d^{++} | \lambda_1 = \dots = \lambda_{g_1}(A) > \lambda_{g_1+1}(A) = \dots \\ &= \lambda_{g_2}(A) > \dots \lambda_{g_{r-1}}(A) > \lambda_{g_{r-1}+1} = \dots = \lambda_{g_r}(A)\} \end{aligned}$$

Infine si introduce il seguente insieme, che diventerà rilevante se si è solamente interessati a un singolo autovalore λ_g :

$$\mathcal{M}(g) = \{A \in \mathcal{M}_d^{++} | \lambda_{g-1}(A) > \lambda_g(A) > \lambda_{g+1}(A)\}, \text{ e } \mathcal{D}(g) = \lambda(\mathcal{M}(g)).$$

2.2 Teoria asintotica e Stimatori

2.2.1 Assunzioni

Si parte con l'assunzione standard sul processo X , Jacod e Protter (2012):

Assunzione 1. Il parametro di deriva b_t è progressivamente misurabile e localmente delimitato. La matrice di covarianza istantanea $c_t = (\sigma\sigma^\top)_t$ è una Itô semimartingala. Inoltre, per qualche $\gamma \in [0, 1)$, è presente una sequenza di *stopping times* (τ_n) tendente a ∞ , e una funzione deterministica $\bar{\delta}_n$ tale che $\int_{\mathbb{R}^d} \bar{\delta}_n(x)^\gamma \bar{\nu}(dx) < \infty$ e che $\|\delta(\omega, t, x)\| \wedge 1 \leq \bar{\delta}_n(x)$, per ogni (ω, t, x) con $t \leq \tau_n(\omega)$.

La costante γ serve come limite superiore per l'indice Blumenthal–Gettoor generalizzato dei salti dei prezzi. La covarianza istantanea è un'altra Itô semimartingala, quindi permette la forma generale della volatilità di volatilità e i salti di volatilità.

Lo sviluppo teorico procede facendo inferenza statistica su $\int_0^t F(c_s) ds$. Grazie al Lemma 4 è possibile fare assunzioni direttamente su f invece che F , che sono molto più semplici da verificare.

Assunzione 2. Si supponga F sia una funzione spettrale vettore-valore, e f sia la corrispondente funzione simmetrica vettore-valore tale che $F = f \circ \lambda$. f è una funzione continua, e soddisfa $\|f(x)\| \leq K(1 + \|x\|^\zeta)$ per qualche $\zeta > 0$.

L'assunzione 2 appena enunciata è sufficiente per assicurare la consistenza dello stimatore che verrà proposto. Sono invece necessarie ulteriori assunzioni per stabilire la distribuzione asintotica.

Assunzione 3. Esiste un qualche insieme aperto e convesso \mathcal{C} , tale che il suo insieme chiuso $\bar{\mathcal{C}} \subset \mathcal{M}(g_1, g_2, \dots, g_r)$, dove $1 \leq g_1 < g_2 < \dots < g_r \leq d$, e per ogni $0 \leq s \leq t$, $c_s \in \mathcal{C} \cap \mathcal{M}^*(g_1, g_2, \dots, g_r)$. Inoltre, f è C^3 su $\mathcal{D}(g_1, g_2, \dots, g_r)$.

L'assunzione 3, e in particolare $c_s \in \mathcal{M}^*(g_1, g_2, \dots, g_r)$, garantisce che differenti gruppi di autovalori non si intersechino in $[0, t]$. Questa condizione è usata principalmente per distribuire il teorema del limite centrale congiunto per funzioni spettrali che dipendono da tutti gli autovalori, anche se non è necessario per alcuni casi speciali, come $\det(\cdot)$ e $\text{Tr}(\cdot)$, che sono lisce ovunque. La condizione di convessità su \mathcal{C} è, in teoria, facile da soddisfare, posto che $\mathcal{M}(g_1, g_2, \dots, g_r)$ può essere incorporato in qualche spazio euclideo di vettori reali. Questa condizione viene imposta per assicurare che il dominio della funzione spettrale possa essere ristretto a un intorno di $\{c_s\}_{0 \leq s \leq t}$, dove la funzione è liscia e il teorema di Lagrange può essere applicato. Questa assunzione non è necessaria se c_t è continuo. Si noti che tutti gli autovalori di c_t sono distinti in un caso speciale, che si rivela abbastanza rilevante nella pratica, come specificato nel Lemma 5. Anche in questo scenario è richiesta l'assunzione 3, perché la funzione autovalore $\lambda(\cdot)$ non è differenziabile ovunque.

Se si è solamente interessati alla funzione spettrale che dipende da un autovalore semplice, ad esempio, il più grande autovalore integrato, $\int_0^t \lambda_1(c_s) ds$, allora basta assicurarsi che $\lambda_1(c_s) > \lambda_2(c_s)$, per $0 \leq s \leq t$, indipendentemente dal fatto che i restanti autovalori siano semplici o meno. Per questo si introduce la seguente assunzione, più debole dell'assunzione 3,

Assunzione 4. Esiste qualche insieme \mathcal{C} aperto e convesso, tale che $\bar{\mathcal{C}} \subset \mathcal{M}(g)$, per qualche $g = 1, 2, \dots, d$, e per ogni $0 \leq s \leq t, c_s \in \mathcal{C}$. Inoltre, f è C^3 su $\mathcal{D}(g)$.

2.2.2 Funzioni spettrali realizzate

Da questa sezione si comincia la costruzione degli stimatori. Per stimare la funzione spettrale integrata, si parte dalla stima della matrice di covarianza istantanea. Si suppone di avere osservazioni equidistanti su X nel periodo $[0, t]$, separate da un intervallo di tempo Δ_n . Si formano blocchi non sovrapposti di lunghezza $k_n \Delta_n$.

Ad ogni $ik_n\Delta_n$ si stima $c_{ik_n\Delta_n}$ dalla formula

$$\hat{c}_{ik_n\Delta_n} = \frac{1}{k_n\Delta_n} \sum_{j=1}^{k_n} (\Delta_{ik_n+j}^n X) (\Delta_{ik_n+j}^n X)^\top \mathbf{1}_{\{\|\Delta_{ik_n+j}^n X\| \leq u_n\}}, \quad (2.5)$$

dove $u_n = \alpha\Delta_n^{\bar{\omega}}$ e $\Delta_l^n X = X_{l\Delta_n} - X_{(l-1)\Delta_n}$. Le scelte di α e $\bar{\omega}$ sono standard in letteratura, Aït-Sahalia e Jacod (2014).

Successivamente sono stimati gli autovalori di $\hat{c}_{ik_n\Delta_n}$ trovando le radici di $|\hat{c}_{ik_n\Delta_n} - \lambda\mathbb{I}| = 0$. Usando la notazione del Lemma 1, si ha $\lambda(\hat{c}_{ik_n\Delta_n}) = \hat{\lambda}_{ik_n\Delta_n}$. Quasi certamente gli autovalori di $\hat{\lambda}_{ik_n\Delta_n}$ sono distinti, Okamoto (1973), in modo tale che si abbia $\hat{\lambda}_{1,ik_n\Delta_n} > \hat{\lambda}_{2,ik_n\Delta_n} > \dots > \hat{\lambda}_{d,ik_n\Delta_n}$. Lo stimatore proposto in Aït-Sahalia e Xiu (2019) è dato da

$$V(\Delta_n, X; F) = k_n\Delta_n \sum_{i=0}^{\lfloor t/(k_n\Delta_n) \rfloor} f(\hat{\lambda}_{ik_n\Delta_n}). \quad (2.6)$$

Teorema 2.2.1. *Si suppongano vere le Assunzioni 1 e 2. Se è vero che $\zeta \leq 1$ o $\zeta > 1$, e sia $\bar{\omega} \in \left[\frac{\zeta - 1}{2\zeta - \gamma}, \frac{1}{2} \right)$, allora lo stimatore 2.6 è consistente. Per $k_n \rightarrow \infty$ e $k_n\Delta_n \rightarrow 0$,*

$$V(\Delta_n, X; F) \xrightarrow{\text{u.c.p.}} \int_0^t F(c_s) ds. \quad (2.7)$$

Ora si passa al teorema del limite centrale per lo stimatore proposto. L'applicazione risulta, però, complicata dal fatto che lo stimatore presenta una distorsione asintotica di secondo ordine, complicazione. La distorsione è caratterizzata come segue:

Proposizione 2.2.1. *Si suppongano vere le Assunzioni 1 - 3. Si supponga inoltre che $k_n \asymp \Delta_n^{-\varsigma}$ e $u_n \asymp \Delta_n^{\bar{\omega}}$ per qualche $\varsigma \in \left(\frac{\gamma}{2}, \frac{1}{2} \right)$ e $\bar{\omega} \in \left[\frac{1 - \varsigma}{2 - \gamma}, \frac{1}{2} \right)$, per $\Delta_n \rightarrow 0$ si ha*

$$k_n \left(V(\Delta_n, X; F) - \int_0^t F(c_s) ds \right) \xrightarrow{p} \frac{1}{2} \sum_{j,k,l,m=1}^d \int_0^t \partial_{jk,lm}^2 F(c_s) (c_{jl,s} c_{jm,s} + c_{jm,s} c_{kl,s}) ds \quad (2.8)$$

La caratterizzazione della distorsione in 2.2.1 suggerisce uno stimatore corretto come:

$$\tilde{V}(\Delta_n, X; F) = k_n \Delta_n \sum_{i=0}^{\lfloor t/(k_n \Delta_n) \rfloor} \left\{ F(\hat{c}_{ik_n \Delta_n}) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^d \partial_{jk,lm}^2 D(\hat{c}_{ik_n \Delta_n}) (\hat{c}_{jl,ik_n \Delta_n} \hat{c}_{km,ik_n \Delta_n} + \hat{c}_{jm,ik_n \Delta_n} \hat{c}_{kl,ik_n \Delta_n}) \right\}. \quad (2.9)$$

Quindi si deriva la distribuzione asintotica dello stimatore corretto:

Teorema 2.2.2. *Si suppongano vere le Assunzioni 1 - 3. Si supponga inoltre che $k_n \asymp \Delta_n^{-\varsigma}$ e $u_n \asymp \Delta_n^{\bar{\omega}}$ per qualche $\varsigma \in \left(\frac{\gamma}{2} \vee \frac{1}{3}, \frac{1}{2}\right)$ e $\bar{\omega} \in \left[\frac{1-\varsigma}{2-\gamma}, \frac{1}{2}\right)$, per $\Delta_n \rightarrow 0$ si ha*

$$\frac{1}{\sqrt{\Delta_n}} \left(\tilde{V}(\Delta_n, X; F) - \int_0^t F(c_s) ds \right) \xrightarrow{\mathcal{L}} \mathcal{W}_t, \quad (2.10)$$

dove \mathcal{W} è un processo continuo definito come un'estensione dello spazio di probabilità originale, che, condizionatamente a \mathcal{F} , è una martingala gaussiana continua centrata con covarianza data da

$$E(\mathcal{W}_{p,t} \mathcal{W}_{q,t} | \mathcal{F}) = \int_0^t \sum_{j,k,l,m=1}^d \partial_{jk} F_p(c_s) \partial_{lm} F_q(c_s) (c_{jl,s} c_{km,s} + c_{jm,s} c_{kl,s}) ds. \quad (2.11)$$

Osservazione 2.2.1. Come fatto notare in precedenza, nel caso in cui la funzione spettrale F dipenda solo da un autovalore semplice λ_g , lo stesso risultato continua a valere sotto l'Assunzione 4 invece che la 3.

Un'implementazione fattibile di questa distribuzione richiede uno stimatore della varianza asintotica, che viene costruito come:

Proposizione 2.2.2. *La varianza asintotica di $\tilde{V}(\Delta_n, C; F)$ può essere stimata in modo consistente da*

$$\begin{aligned}
 & k_n \Delta_n \sum_{i=0}^{\lfloor t/(k_n \Delta_n) \rfloor} \sum_{j,k,l,m=1}^d \partial_{jk} F_p(\hat{c}_{ik_n \Delta_n}) \partial_{lm} F_q(\hat{c}_{ik_n \Delta_n}) \\
 & \quad \times (\hat{c}_{jl, ik_n \Delta_n} \hat{c}_{km, ik_n \Delta_n} + \hat{c}_{jm, ik_n \Delta_n} \hat{c}_{kl, ik_n \Delta_n}) \\
 & \xrightarrow{p} \int_0^t \sum_{j,k,l,m=1}^d \partial_{jk} F_p(c_s) \partial_{lm} F_q(c_s) (c_{jl,s} c_{km,s} + c_{jm,s} c_{kl,s}) ds.
 \end{aligned} \tag{2.12}$$

2.2.3 Autovalori realizzati

Per ottenere gli autovalori realizzati si procede rendendo il teorema enunciato in precedenza più specifico, in modo da ottenere il teorema del limite centrale per gli stimatori degli autovalori realizzati. Con la struttura degli autovalori, Ait-Sahalia e Xiu (2019) fanno uso di una funzione spettrale particolare, F^λ , fatta apposta per costruire la teoria asintotica necessaria:

$$F^\lambda(\cdot) = \left(\frac{1}{g_1} \sum_{j=1}^{g_1} \lambda_j(\cdot), \frac{1}{g_2 - g_1} \sum_{j=g_1+1}^{g_2} \lambda_j(\cdot), \dots, \frac{1}{g_r - g_{r-1}} \sum_{j=g_{r-1}+1}^{g_r} \lambda_j(\cdot) \right)^\top \tag{2.13}$$

Se un gruppo contiene un solo λ_j , allora l'elemento corrispondente di F^λ è uguale al questo singolo autovalore; se in un certo gruppo tutti gli autovalori sono identici, allora il corrispondente valore di F^λ restituisce l'autovalore comune del gruppo.

Corollario 1. *Si supponga $k_n \asymp \Delta_n^{-\varsigma}$ e $u_n \asymp \Delta_n^{\bar{\omega}}$ per qualche $\varsigma \in \left(\frac{\gamma}{2} \vee \frac{1}{3}, \frac{1}{2}\right)$ e $\bar{\omega} \in \left[\frac{1-\varsigma}{2-\gamma}, \frac{1}{2}\right)$.*

(i) *Sotto l'Assunzione 1, lo stimatore del vettore degli autovalori integrati dato da*

$$V(\Delta_n, X; \lambda) = k_n \Delta_n \sum_{i=0}^{\lfloor t/(k_n \Delta_n) \rfloor} \lambda(\hat{c}_{ik_n \Delta_n}) \tag{2.14}$$

è consistente.

(ii) Se vale l'Assunzione 3, lo stimatore corrispondente alla p -esima entrata di F^λ restituito dalla 2.9 può essere riscritto esplicitamente come

$$\begin{aligned} \tilde{V}(\Delta_n, X; F_p^\lambda) &= \frac{k_n \Delta_n}{g_p - g_{p-1}} \sum_{i=0}^{[t/(k_n \Delta_n)]} \sum_{k=g_{p-1}+1}^{g_p} \\ &\times \left\{ \hat{\lambda}_{h, ik_n \Delta_n} - \frac{1}{k_n} \text{Tr} \left((\hat{\lambda}_{h, ik_n \Delta_n} \mathbb{I} - \hat{c}_{ik_n \Delta_n})^+ \hat{c}_{ik_n \Delta_n} \right) \hat{\lambda}_{h, ik_n \Delta_n} \right\}. \end{aligned}$$

Il teorema del limite centrale per $\tilde{V}(\Delta_n, X; F^\lambda) = (\tilde{V}(\Delta_n, X; F_1^\lambda), \dots, \tilde{V}(\Delta_n, X; F_r^\lambda))$ è dato da

$$\frac{1}{\sqrt{\Delta_n}} \left(\tilde{V}(\Delta_n, X; F^\lambda) - \int_0^t F^\lambda(c_s) ds \right) \xrightarrow{\mathcal{L}^{-s}} \mathcal{W}_t^\lambda, \quad (2.15)$$

dove \mathcal{W}^λ è un processo continuo definito su un'estensione dello spazio di probabilità originale, che condizionatamente a \mathcal{F} , è una martingala gaussiana continua con matrice di covarianza diagonale data dalla formula

$$\begin{aligned} &E(\mathcal{W}_t^\lambda (\mathcal{W}_t^\lambda)^\top | \mathcal{F}) \\ &= \begin{pmatrix} \frac{2}{g_1} \int_0^t \lambda_{g_1, s}^2 ds & & & \\ & \frac{2}{g_2 - g_1} \int_0^t \lambda_{g_2, s}^2 ds & & \\ & & \dots & \\ & & & \frac{2}{g_r - g_{r-1}} \int_0^t \lambda_{g_r, s}^2 ds \end{pmatrix} \end{aligned} \quad (2.16)$$

dove $F^\lambda(c_s) = (\lambda_{g_1, s}, \lambda_{g_2, s}, \dots, \lambda_{g_r, s})^\top$.

(iii) Sotto le Assunzioni 1 e 4, lo stimatore rispetto al g -esimo autovalore semplice $\lambda_g(\cdot)$ è dato da

$$\begin{aligned} \tilde{V}(\Delta_n, X; \lambda_g) &= k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \\ &\times \left\{ \hat{\lambda}_{g, ik_n \Delta_n} - \frac{1}{k_n} \text{Tr} \left((\hat{\lambda}_{g, ik_n \Delta_n} - \hat{c}_{ik_n \Delta_n})^+ \hat{c}_{ik_n \Delta_n} \right) \hat{\lambda}_{g, ik_n \Delta_n} \right\}, \end{aligned} \quad (2.17)$$

che soddisfa

$$\frac{1}{\sqrt{\Delta_n}} \left(\tilde{V}(\Delta_n, X; \lambda_g) - \int_0^t F^\lambda(c_s) ds \right) \xrightarrow{\mathcal{L}^{-s}} \mathcal{W}_t^{\lambda_g}, \quad (2.18)$$

dove W^{λ_g} è un processo continuo definito su un'estensione dello spazio di probabilità originale, che condizionatamente a \mathcal{F} , è una martingala gaussiana continua centrata con varianza $\int_0^t \lambda_{g,s}^2 ds$.

Osservazione 2.2.2. Il corollario 1 è, nel contesto preso in esame, l'analogo del risultato classico sulla teoria asintotica per la PCA sviluppato da Anderson (1963), dove si mostrava che

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, 2 \text{diag}(\lambda_1^2, \lambda_2^2, \dots, \lambda_d^2)), \quad (2.19)$$

con $\hat{\lambda}$ e λ vettori di autovalori sulla matrice di correlazione rispettivamente del campione e della popolazione e λ semplice. Si noti la similarità tra 2.19 e 2.16 nel caso particolare in cui gli autovalori siano semplici, e di conseguenza $g_i = i$, e costanti invece che stocastici. Il caso classico richiede che l'assunzione forte per cui la matrice di covarianza deve seguire una distribuzione Wishart; di contro i risultati presentati sono completamente non parametrici e si basano sulle proprietà asintotiche dell'alta frequenza.

Osservazione 2.2.3. Se sono presenti autovalori di c_s pari a 0 per qualsiasi $0 \leq s \leq t$, lo stimatore è super-efficiente. Si tratta di un caso particolare essendo il rango della matrice di covarianza determinato dal numero di autovalori non nulli.

2.2.4 Componenti Principali realizzate

Quando gli autovalori sono semplici, i corrispondenti autovettori sono unicamente determinati. Ne consegue che è possibile stimare l'autovettore insieme all'autovalore, che porta a costruire la corrispondente componente principale:

Proposizione 2.2.3. *Si suppongano vere le Assunzioni 1 e 4. Inoltre, sia $\gamma_{g,s}$ una funzione vettore-valore che fa corrispondere agli autovettori di c_s le rispettive radici*

semplici $\lambda_{g,s}$. Si supponga $k_n \asymp \Delta_n^{-\varsigma}$ e $u_n \asymp \Delta_n^{\bar{\omega}}$ per qualche $\varsigma \in \left(\frac{\gamma}{2}, \frac{1}{2}\right)$ e $\bar{\omega} \in \left[\frac{1-\varsigma}{2-\gamma}, \frac{1}{2}\right)$. Allora si ha per $\Delta_n \rightarrow 0$

$$\sum_{i=1}^{\lfloor t/(k_n \Delta_n) \rfloor - 1} \hat{\gamma}_{g,(i-1)k_n \Delta_n}^\top (X_{(i+1)k_n \Delta_n} - X_{ik_n \Delta_n}) \stackrel{\text{u.c.p.}}{\Rightarrow} \int_0^t \gamma_{g,s}^\top dX_s.$$

2.2.5 Autovettori realizzati

Si calcolano, quindi, i *loadings* delle componenti di X sulle componenti principali, cioè gli autovettori. Essendo l'autovettore stocastico si stima l'autovettore integrato. Tranne che per il segno, un autovettore è unicamente determinato se il suo autovalore associato è semplice. È possibile determinare il segno, e quindi identificare l'autovettore, richiedendo a priori che un certo elemento dell'autovettore sia positivo, ad esempio, il primo elemento non nullo.

Corollario 2. *Si suppongano vere le assunzioni 1 e 4. Sia, inoltre, $\gamma_{g,s}$ una funzione vettore-valore che faccia corrispondere all'autovettore di c_s la rispettiva radice semplice $\lambda_{g,s}$ per ogni $s \in [0, t]$. Allora per $\Delta_n \rightarrow 0$ si ha*

$$\begin{aligned} & \frac{1}{\sqrt{\Delta_n}} \left(k_n \Delta_n \sum_{i=0}^{\lfloor t/(k_n \Delta_n) \rfloor} \left(\hat{\gamma}_{g,ik_n \Delta_n} \right. \right. \\ & \left. \left. + \frac{1}{2k_n} \sum_{p \neq g} \frac{\hat{\lambda}_{g,ik_n \Delta_n} \hat{\lambda}_{p,ik_n \Delta_n}}{\hat{\lambda}_{g,ik_n \Delta_n} - \hat{\lambda}_{p,ik_n \Delta_n}} \hat{\gamma}_{g,ik_n \Delta_n} \right) - \int_0^t \gamma_{g,s} ds \right) \xrightarrow{L^{-s}} \mathcal{W}_t^\gamma, \end{aligned}$$

dove \mathcal{W}^γ è un processo continuo definito su un'estensione dello spazio di probabilità originale, che condizionatamente a \mathcal{F} , è una martingala gaussiana centrata con matrice di covarianza data da

$$E(\mathcal{W}_t^\gamma (\mathcal{W}_t^\gamma)^\top | \mathcal{F}) = \int_0^t \lambda_{g,s} (\lambda_{g,s} \mathbb{I} - c_s)^+ c_s (\lambda_{g,s} \mathbb{I} - c_s)^+ ds.$$

2.2.6 "PCA" sulla matrice di covarianza integrata

È lecito chiedersi il motivo per il quale si è scelto di stimare gli autovalori integrati sulla matrice di covarianza istantanea piuttosto che gli autovalori sulla matrice di covarianza integrata. Il motivo è che gli autovalori sono funzioni complicate della matrice di covarianza, le due quantità non sono molto collegate, e si ottiene che l'ultima misura è meno informativa della prima. Infatti, il rango della matrice di covarianza istantanea è determinato dal numero di autovalori non nulli istantanei. L'autostruttura della matrice di covarianza integrata è piuttosto opaca a causa dell'aggregazione di matrici di covarianza istantanee. Non di meno è possibile costruire una procedura "PCA" sulla matrice di covarianza integrata, il cui risultato asintotico è disponibile immediatamente attraverso l'uso del metodo Delta, usando il Lemma 2 e il Teorema 13.2.4 di Jacod e Protter (2012).

La corrispondente componente principale è definita come $\gamma_g^T(X_t - X_0)$, dove γ_g è l'autovettore di $\int_0^t c_s ds$. Comunque, il corrispondente autovalore e la componente principale non soddisfano la relazione fondamentale tra autovalore e la varianza della componente principale.

$$\lambda_g \left(\int_0^t c_s ds \right) \neq [\gamma_g^T(X_t - X_0), \gamma_g^T(X_t - X_0)]^c, \quad (2.20)$$

che sarebbe una proprietà desiderata per ogni procedura PCA. Questo fatto rende la seconda procedura molto meno utile rispetto alla prima, basata sull'autovalore integrato della covarianza istantanea. Un altro fattore che distingue il secondo modo di calcolare la PCA sia dal modello classico che dal primo procedimento è che la matrice di covarianza asintotica degli autovalori di $\int_0^t c_s ds$ non è più diagonale.

Capitolo 3

Applicazione ai dati

3.1 Formule e descrizione dei dati

Ora che i fondamenti teorici su cui si basa il resto del lavoro sono stati posti, si procede alla loro applicazione nel contesto dei dati ad alta frequenza rilevati in ambito finanziario

3.1.1 Componenti principali ad alta frequenza nel settore finanziario

In questa sezione si procede brevemente a riassumere le quantità che verranno utilizzate e la loro costruzione, alla luce della natura dei dati presi in esame. Si assume che il vettore X_j contenga i log-prezzi al tempo j di d aziende quotate; il tempo è misurato in intervalli $[0, t]$, e le osservazioni sono equidistanti, separate da un intervallo di tempo Δ_n . I rendimenti sono calcolati nel modo classico come $\Delta^n X = X_{t\Delta_n} - X_{(t-1)\Delta_n}$ campionati per ogni minuto in un orizzonte temporale di una settimana. L'intervallo di tempo $[0, t]$ su cui sono campionati i dati è ottenuto da $t = 1950 / (390 \cdot 252) = 0,0198$ anni in una settimana, dove 1950 sono i minuti di

transazioni sul mercato rilevati in una settimana, 390 sono i minuti di transazioni rilevati in una giornata di mercato aperto e 252 sono i giorni in cui il mercato è aperto nell'anno. Inoltre, Δ_n è l'intervallo di tempo misurato in anni in cui i dati sono separati in $[0, t]$: $\Delta_n = 1/(390 \cdot 252) = 0,000010175$ anni in un minuto. Dati i rendimenti osservati in una settimana, si costruiscono blocchi non sovrapposti di lunghezza $k_n \Delta_n$, ogni uno contenente $k_n = 325$ osservazioni. Per ogni $i k_n \Delta_n$, per $i = 1, \dots, m$, con m il numero di blocchi in una settimana, si stima la covarianza realizzata $c_{i k_n \Delta_n}$ con l'uso della formula 2.5:

$$\hat{c}_{i k_n \Delta_n} = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} (\Delta_{i k_n + j}^n X) (\Delta_{i k_n + j}^n X)^\top.$$

Dalla covarianza realizzata istantanea specifica del blocco, $c_i \equiv c_{i k_n \Delta_n}$, con $i = 1, \dots, m$, si stimano gli autovalori $\lambda_{g,i}$ e gli autovettori $\gamma_{g,i}$, che permettono di costruire le corrispondenti componenti principali. Se $\gamma_{g,i}$ è l'autovettore che corrisponde all'autovalore di $\lambda_{g,i}$, la g -esima componente principale per la settimana presa in considerazione è pari a

$$\sum_{i=0}^{[t/(k_n \Delta)]-1} \gamma_{g,(i-1)k_n \Delta_n}^\top (X_{(i+1)k_n \Delta_n} - X_{i k_n \Delta_n})$$

dove $t/(k_n \Delta_n) = m$, il numero di blocchi nella settimana, pari a 6.

Gli autovalori, come detto, sono stimati su $\hat{c}_{i k_n \Delta_n}$ risolvendo $|\hat{c}_{i k_n \Delta_n} - \lambda \mathbb{I}| = 0$, ottenendo $\lambda(\hat{c}_{i k_n \Delta_n}) = \hat{\lambda}_{i k_n \Delta_n}$. Gli autovalori contenuti nel vettore $\hat{\lambda}_{i k_n \Delta_n}$ sono distinti quasi certamente, quindi sono $\lambda_{1, i k_n \Delta_n} > \lambda_{2, i k_n \Delta_n} > \dots > \lambda_{d, i k_n \Delta_n}$. Lo stimatore per il vettore degli autovalori integrati è dato da

$$V(\Delta_n, X; \lambda) = k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \lambda(\hat{c}_{i k_n \Delta_n}),$$

che è consistente.

Infine si passa alla stima degli autovettori realizzati. Come riportato in precedenza se gli autovalori sono non ripetuti gli autovettori associati sono univocamente

determinati. Lo stimatore per gli autovalori realizzati corrisponde a

$$k_n \Delta_n \sum_{i=0}^{\lfloor t/(k_n \Delta_n) \rfloor} \left(\hat{\gamma}_{g, ik_n \Delta_n} + \frac{1}{2k_n} \sum_{p \neq g} \frac{\hat{\lambda}_{g, ik_n \Delta_n} \hat{\lambda}_{p, ik_n \Delta_n}}{\hat{\lambda}_{g, ik_n \Delta_n} - \hat{\lambda}_{p, ik_n \Delta_n}} \hat{\gamma}_{g, ik_n \Delta_n} \right),$$

definito nel Corollario 2.

3.1.2 Descrizione dei dati

Per le analisi empiriche sono stati utilizzati i prezzi intragiornalieri che vanno dal 10 gennaio 2003 al 23 febbraio 2021, rilevati al minuto. I rendimenti sono stati calcolati, come specificato in precedenza, come $\Delta_l^n X = X_{l\Delta_n} - X_{(l-1)\Delta_n}$. Il dataset preso in esame contiene 89 titoli di cui 73 appartenenti al NYSE (*New York Stock Exchange*) e 15 al NASDAQ (*National Association of Securities Dealers Automated Quotation*). I titoli sono divisi in otto settori:

- 8 titoli appartengono ai *Basic Materials*, ovvero i beni primari;
- 9 titoli fanno parte dei *Consumer Goods*, beni di consumo;
- 14 sono titoli finanziari, *Financial*;
- 13 sono parte del settore sanitario, *Healthcare*;
- 10 sono indicati come *Industrial Goods*, beni industriali;
- 15 fanno parte dei *Services*, i servizi;
- 13 appartengono al settore *Technology*;
- 5, infine, appartengono al settore *Utilities*.

I titoli sono, inoltre, distinti in base alla categoria *Industry*, ovvero il tipo di industria a cui appartengono. Questa classificazione è molto più dettagliata e

comprende un totale di 53 voci, nella tabella A.1 in appendice sono presenti tutti i titoli utilizzati.

Nell'analisi dei titoli, il dataset è stato suddiviso in due periodi, un primo che va dal 10 gennaio 2003 al 16 settembre 2008, data del fallimento di Lehman Brothers, il secondo che va dalla data successiva al fallimento fino alla fine del periodo, 23 febbraio 2021. Per i due periodi si riportano alcune statistiche descrittive dei dati nella tabella 3.1. Dalle statistiche si può notare come per alcuni settori ci sia stata una "stabilizzazione" a seguito della crisi del 2008, mentre per altri è avvenuto l'opposto. In particolare si può notare come i *Consumer Goods* abbiano risentito di un aumento della volatilità, indicato da un aumento della deviazione standard, della skewness e della kurtosi, insieme ai settori dell'*Healthcare*, dei *Basic Materials* e degli *Industrial Goods*, quindi tutti quei settori legati all'economia reale, mentre vi è stata una diminuzione degli indici di volatilità del settore *Financial*, che suggeriscono una maggiore stabilità del settore rispetto al pre-crisi del 2008.

3.2 Autovalori e autovettori realizzati sui dati

Usando la metodologia descritta nel Capitolo 2 e ripresa nella sezione precedente si cominciano le analisi focalizzandosi su 912 autovalori e autovettori settimanali. Per l'interpretazione dei risultati si fa notare che, tradizionalmente, una maggiore percentuale di varianza spiegata dalle prime componenti principali indica un maggior livello di co-movimento tra i titoli. A sua volta, un maggior grado di co-movimento è interpretabile come il grado di fragilità del sistema e governa la resilienza agli shock, con una possibile interpretazione del grado di comovimento in termini di accumulo di rischio sistemico nel mercato. Come detto in Ait-Sahalia e Xiu (2019), i fattori idiosincratici diventano relativamente meno importanti e sono sempre più dominati dai fattori comuni in periodi di crisi.

Rendimenti a un minuto							
	Media	Mediana	Dev. Standard	Skewness	Kurtosi	Massimo	Minimo
Consumer Goods							
Prima	< 0.0001	0	0.0007	0.0667	518.3	0.1098	-0.1017
Dopo	< 0.0001	0	0.0007	-1.0860	1118.2	0.1424	-0.1843
Healthcare							
Prima	< 0.0001	0	0.0009	1.7759	103745.7	0.6936	-0.6880
Dopo	< 0.0001	0	0.0008	-1.0614	424.1	0.1039	-0.1548
Technology							
Prima	< 0.0001	0	0.0009	-0.2736	249.6	0.1146	-0.1193
Dopo	< 0.0001	0	0.0010	-2.0300	750.8	0.1141	-0.2948
Utilities							
Prima	< 0.0001	0	0.0009	-0.8971	331.7	0.0892	-0.0860
Dopo	< 0.0001	0	0.0008	-0.0815	279.1	0.1108	-0.0989
Financial							
Prima	< 0.0001	0	0.0010	1.0003	1287.8	0.1696	-0.2316
Dopo	< 0.0001	0	0.0013	-0.9383	564.5	0.2584	-0.2581
Services							
Prima	< 0.0001	0	0.0009	0.1980	62.5	0.0487	-0.0532
Dopo	< 0.0001	0	0.0008	-0.7907	735.4	0.21370	-0.2152
Basic Materials							
Prima	< 0.0001	0	0.0010	-0.07083	195.5	0.0953	-0.1025
Dopo	< 0.0001	0	0.0012	-1.1917	333.3	0.1112	-0.2508
Industrial Goods							
Prima	< 0.0001	0	0.0009	0.9690	306.5	0.1112	-0.1039
Dopo	< 0.0001	0	0.0011	-0.1053	195744.3	1.1300	-1.1291

Tabella 3.1: Tabella degli indici riassuntivi dei rendimenti nel periodo dal 10 gennaio 2003 al 23 febbraio 2021, suddivisi prima e dopo la data di fallimento di Lehman Brothers, 16 settembre 2008.

3.2.1 Autovalori realizzati

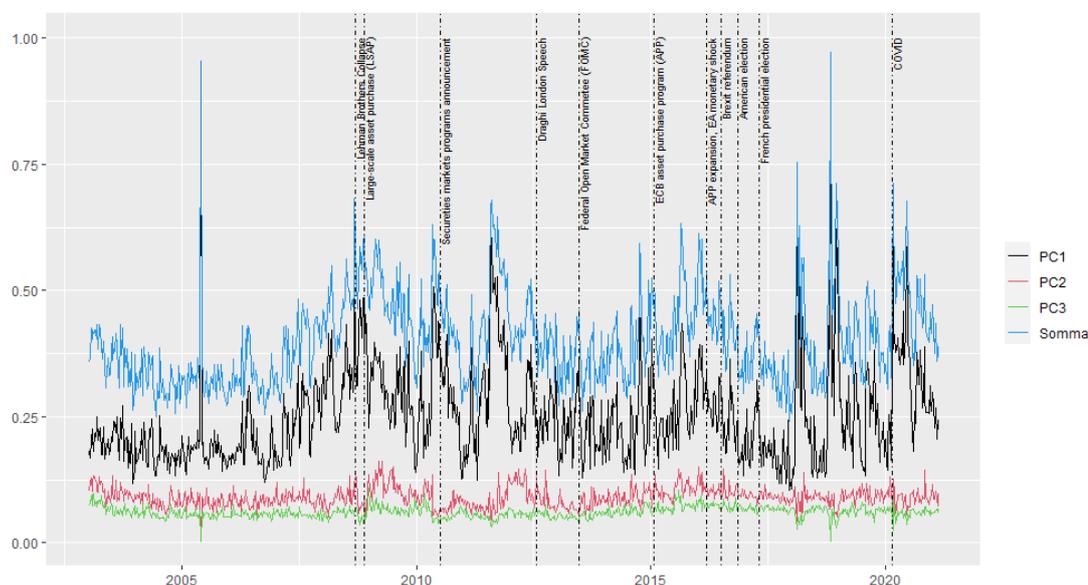


Figura 3.1: Percentuale della varianza spiegata da ciascuna delle prime tre componenti principali (rispettivamente in nero, rosso e verde) e somma della varianza spiegata dalle prime tre componenti principali in blu.

Nella figura 3.1 è visualizzata la percentuale di variazione totale spiegata congiuntamente dalle prime tre componenti principali (in blu) e la percentuale spiegata individualmente da ciascuna delle prime tre componenti principali, rispettivamente in nero, rosso e verde per la prima, seconda e terza componente principale). Sono, inoltre, riportati alcuni eventi importanti per l'andamento del mercato, e non solo, che corrispondono a periodi di forte instabilità; questi sono: il 16 settembre 2008, collasso Lehman Brothers, che corrisponde a un rischio globale, e che divide anche l'analisi dei titoli; il 26 novembre 2008, quanto è stato attuato il Large-Scale Asset Purchase (LSAP), che ha comportato uno shock monetario negli USA; l'8 luglio 2010 è avvenuto l'annuncio del Securities Markets Program come conseguenza della crisi del debito sovrano in Europa, che ha portato uno

shock monetario nell'area Euro; il 27 luglio 2012, discorso di Draghi a Londra e corrispondente shock monetario nell'area Euro; il 20 giugno 2013 ha avuto luogo l'incontro del Federal Open Market Committee (FOMC), che ha comportato uno shock monetario statunitense; il 23 gennaio 2015, annuncio del programma di acquisto dei titoli della banca centrale europea e shock monetario europeo; il 07 marzo 2016 espansione del programma di acquisti della BCE e shock monetario nell'area Euro; 27 giugno 2016 è stato votato il referendum sulla Brexit, questo evento è corrisposto a un rischio nell'area Euro e globale; il 9 novembre 2016 sono state elezioni del presidente statunitense, con un rischio statunitense e globale; il 27 aprile 2017 hanno avuto luogo le elezioni del presidente francese con un rischio nell'area Euro e globale; infine, il 25 febbraio 2020 è avvenuto intensificarsi dell'epidemia di COVID-19 nel mondo, un rischio globale.

Per una lettura semplificata nella figura 3.2 sono riportate le medie annue delle percentuali presentate in figura 3.1. Si osserva, comunque, che la somma della varianza totale spiegata dalle prime tre componenti principali è, fino al 2008, inferiore al 40%, aggirandosi sul 35%. Nel secondo periodo, dopo la Grande Crisi il valore arriva fino al 50%, per poi abbassarsi gradualmente negli anni successivi. Si nota che rimane significativamente sopra il periodo pre-crisi fino al 2012, rimanendo poi stabile, ma comunque sopra la media del primo periodo e presenta un nuovo picco nel 2020, in corrispondenza dell'epidemia da COVID-19.

Si noti che a tutti i fenomeni di rischio sui mercati rappresentati corrisponde un picco della volatilità spiegata dalle componenti principali, ma sono presenti anche vari aumenti di comovimento dei titoli in altri periodi. In particolare dalla figura 3.1 si notano due picchi particolarmente elevati, dove sembra che la prima componente principale spieghi quasi tutta la variabilità. Questi *outlayer* possono essere spiegati come lo scaricarsi sul mercato della variabilità accumulata nel tempo. Una seconda osservazione molto importante è che i picchi negli autovalori

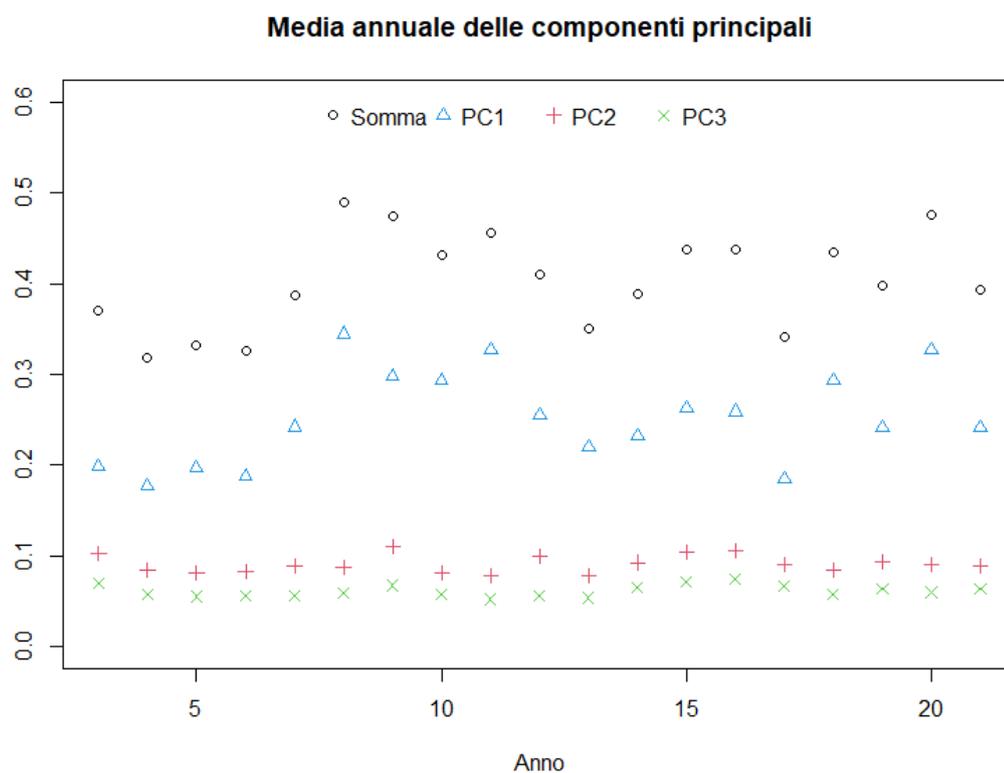


Figura 3.2: Media annuale della varianza spiegata dalle prime tre componenti principali e dalla loro somma

possono essere associati a periodi in cui il mercato agisce come un amplificatore degli shock idiosincratici, anticipando, quindi, un futuro stress finanziario.

3.2.2 Autovettori realizzati

Nguyen, Tran e Nguyen (2018) correlano le componenti dell'autovettore associato all'autovalore più grande, $\gamma_{1,i}$, al grado di correlazione ω_i di ogni marchio, definito come:

$$\omega_i = \sum_{j=1, j \neq i}^N \rho_{ij}, \quad (3.1)$$

dove ρ_{ij} è la correlazione tra la i -esima e la j -esima azienda. Ne consegue che è possibile interpretare la componente dell'autovettore relativa al titolo come una misura di correlazione di quel titolo con gli altri titoli inclusi nell'analisi. Ad esempio, se un titolo ha il più grande elemento nell'autovettore (il *loading* più grande della prima componente principale), sarà, mediamente, più correlato agli altri titoli, avendo un'alta probabilità di essere fortemente sistemico. Esaminando gli autovettori è, allora, possibile identificare il ruolo che ogni settore gioca nella correlazione e quanto contribuisca al comportamento sistemico del mercato: maggiore è la componente del settore, cioè la somma delle componenti associate ai titoli di quel settore, più forte è l'influenza di quel settore sugli altri.

Studiando l'autovettore associato all'autovalore più grande è possibile identificare

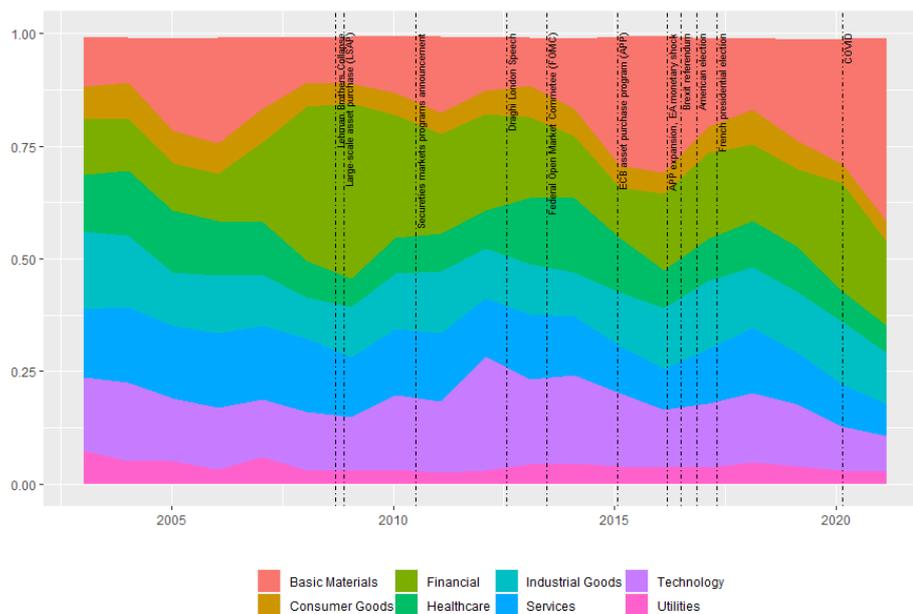


Figura 3.3: Media annuale del contributo percentuale di ogni settore alla prima componente principale

se il comportamento del mercato è dominato da un settore particolare. La figura 3.3 mostra un grafico ad area che rappresenta la media annuale del primo autovet-

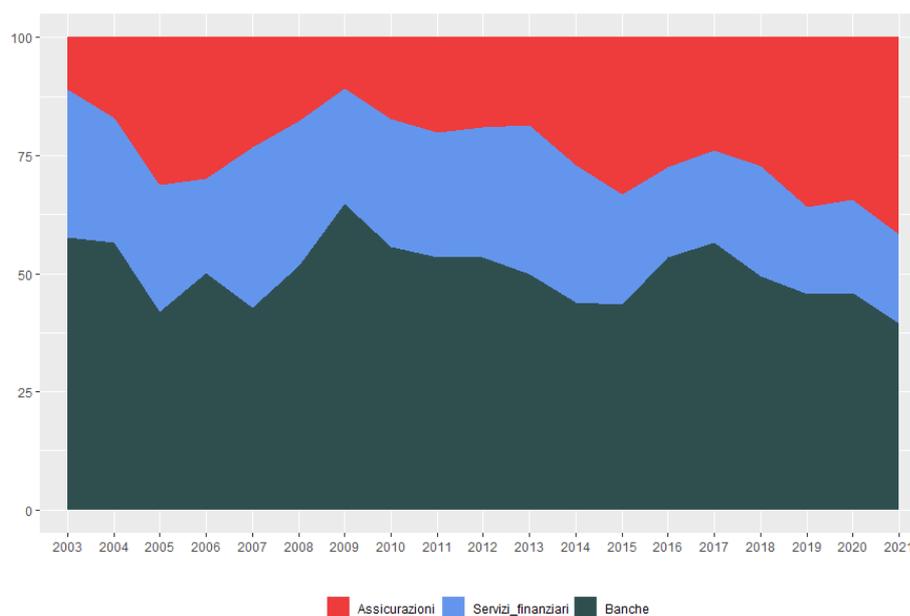


Figura 3.4: Composizione della prima componente principale in riferimento solo al settore finanziario, scomposto in banche, assicurazioni e altri servizi finanziari.

tore integrato normalizzato per settore¹. Allo stesso modo la figura 3.4 evidenzia solo la composizione della componente *Financial* del primo grafico, mettendo in risalto i sotto-settori in cui è suddivisa: bancario, assicurativo e di altri servizi finanziari.

Si nota che con la crisi finanziaria del 2008 il contributo del settore finanziario al rischio del mercato è molto elevato, in particolare vi è un aumento repentino rispetto al pre-2008 sostanziale, passando da un 17,8% nel 2007 a un 34,2% nel 2008, fino a un 39,1% l'anno successivo; ritornando, infine, a dei valori comparabili al pre-crisi solo nel 2013, dove segna un 17,8%. Si può notare, invece, come la crisi del 2020, causata dal COVID-19, abbia prodotto una forte crisi dell'economia

¹A causa della mancata specificazione del settore e dell'industria per il titolo "WBA" il primo grafico non somma propriamente a 100, in ogni caso l'azienda non svolge un ruolo fondamentale nel mercato e non rappresenta un grosso problema

reale, con un forte aumento del contributo alla variabilità spiegata dato dal settore dei *Basic Materials*, mentre, al contrario della crisi precedente, il contributo del settore finanziario rimane relativamente stabile. Infatti per il settore dei *Basic Materials* si passa da un contributo medio del 16,85% fino al 2020 al 41,1% nel 2021. La seconda figura, 3.4, evidenzia come, durante la crisi del 2008, vi sia stato un maggior contributo al rischio delle banche rispetto agli altri attori del panorama finanziario, per poi decrescere nel periodo successivo.

Grazie al contenuto informativo delle componenti principali su dati ad alta frequenza è possibile studiare il contributo medio di ogni settore su base settimanale. La figura 3.5 contiene le distribuzioni dei contributi settimanali alla prima componente principale. Innanzitutto si può notare come nei periodi di crisi vi sia una maggiore concentrazione attorno alla media, ad eccezione del settore *Financial*, che nel periodo 2008-2009 sembra subire un aumento della media e della volatilità. Al contrario, durante i periodi di crisi più contenuti (2012-2013 e 2015-2017) vi è una maggiore volatilità del contributo per settore.

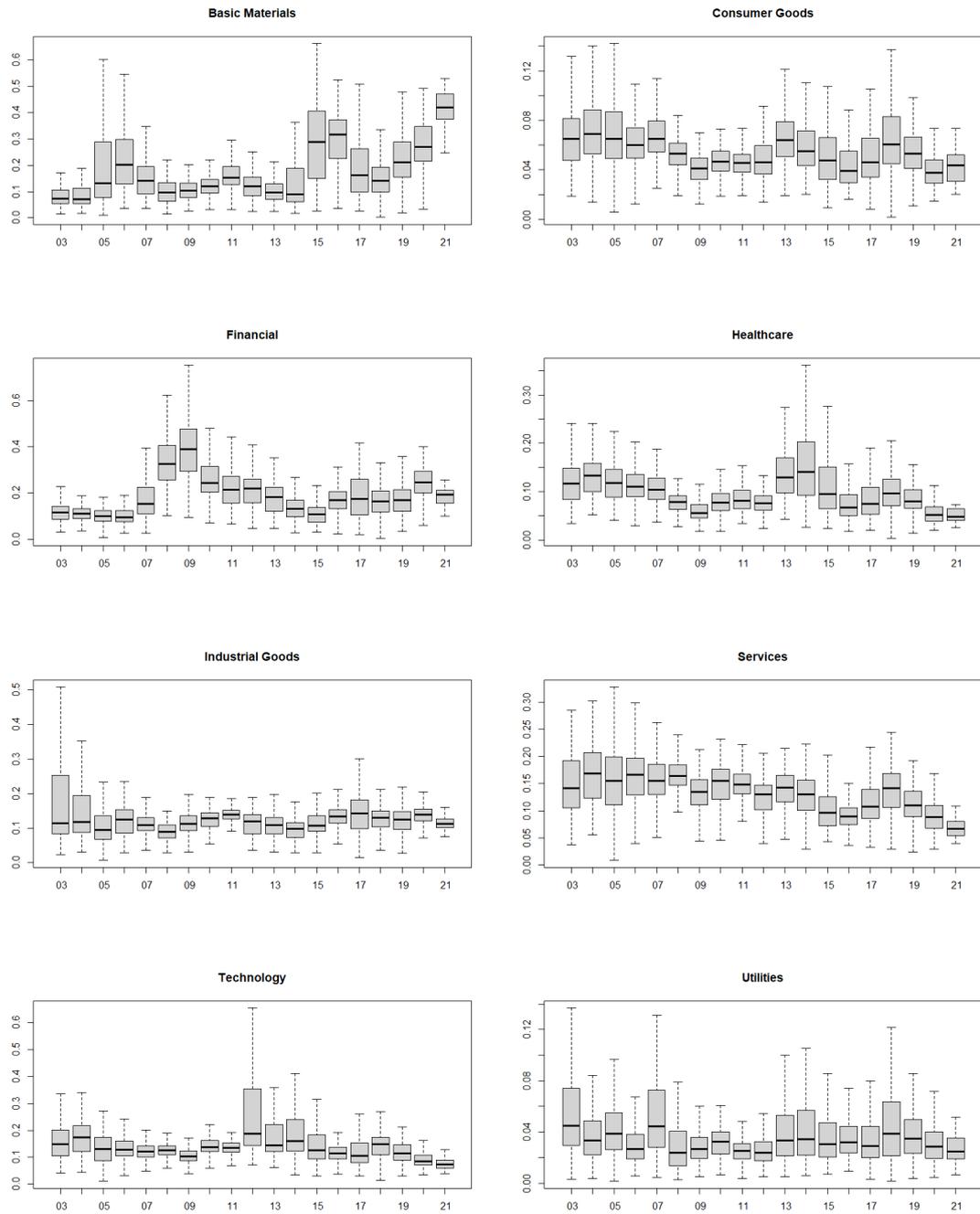


Figura 3.5: Dispersione del contributo percentuale per settore alla prima componente principale per anno.

Capitolo 4

Analisi seriale delle componenti principali

In questa sezione è d'interesse monitorare le caratteristiche della dipendenza seriale che caratterizzano autovalori e autovettori realizzati.

4.1 Analisi degli autovalori

4.1.1 Modelli a effetti moltiplicativi

Le serie degli autovalori realizzati sono, per costruzione, caratterizzate da valori non negativi. Si prestano, di conseguenza, ad essere analizzate tramite modelli a effetti moltiplicativi. Introdotti formalmente da Engle (2002), questi modelli rappresentano un'estensione dei modelli ARCH e GARCH, con l'obiettivo di descrivere meglio dati non negativi, fino ad allora trattati solamente ignorando la non negatività o usando una trasformazione logaritmica, che, però, restituisce un errore quando vi è un'osservazione pari esattamente a zero. La soluzione, in tal caso,

è quella di aggiungere una piccola costante positiva, di contro, questo escamotage restituisce un modello distorto.

I modelli proposti da Engle (2002) sono specificati come

$$x_t = \mu_t \varepsilon_t, \quad \varepsilon_t | \mathfrak{F}_{t-1} \sim D(1, \phi^2), \quad (4.1)$$

dove

$$\begin{aligned} E(x_t | I_{t-1}) &= \mu_t \cdot E(\varepsilon_t | I_{t-1}) = \mu_t, \\ V(x_t | I_{t-1}) &= \mu_t^2 \cdot V(\varepsilon_t | I_{t-1}) = \mu_t^2 \phi^2, \\ \text{con } x_t | I_{t-1} &= D(\mu_t, \mu_t^2 \phi^2) \end{aligned}$$

i cui residui sono $\hat{\varepsilon}_t = x_t / \mu_t$ omoschedastici. Si procede specificando una funzione della media, che viene espressa in modo simile alla varianza di un processo ARCH o GARCH. Supposto che la media sia una funzione lineare traslata di x e vi siano $k \times 1$ vettori di variabili esogene z_t , la specificazione della media diventa

$$\mu_t = \omega + \sum_{j=1}^p \alpha_j x_{t-j} + \sum_{j=1}^q \beta_j \mu_{t-j} + \gamma' z_t. \quad (4.2)$$

Come nel caso dei modelli GARCH è necessario fare delle restrizioni sui parametri per assicurare valori positivi delle medie per tutte le possibili realizzazioni e la distribuzione stazionaria per x . Se le z sono variabili positive, allora è sufficiente assicurare una condizione di non negatività per i parametri. Comunque non è necessario, da Bollerslev, Engle e Nelson (1994) le condizioni sufficienti per la stazionarietà in covarianza di x sono che le z siano stazionarie in covarianza e che

$$\sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j < 1. \quad (4.3)$$

Il risultato è stato formalizzato da Engle e Russell (1998) basandosi su un teorema in Lee e Hansen (1994) per i modelli GARCH. In tal caso, x era la durata tra eventi successivi, ma il teorema è applicabile a qualsiasi processo non negativo.

Nel teorema si assume il processo GARCH di primo ordine con la possibilità di radici sia unitarie che esplosive. Un importante corollario del teorema trovato è che:

Corollario 3. *Se*

i. $E_{t-1}(x_t) \equiv \mu_{0,t} = \omega_0 + \alpha_0 x_{t-1} + \beta_0 \mu_{0,t-1},$

ii. $\varepsilon_t \equiv x_t / \mu_{0,t},$

iii. *il processo è strettamente stazionario, ergodico, non degenerare e ha momento condizionato secondo limitato,*

iv. $\sup_t E[\ln(\beta_0 + \alpha_0 \varepsilon_t) | \mathfrak{F}_{t-1}] < 0,$

v. $\theta_0 \equiv (\omega_0, \alpha_0, \beta_0)$ *è nell'interno di* $\Omega,$

vi. *la verosimiglianza è*

$$L(\theta) = - \sum_{t=1}^T \left(\log(\mu_t) + \frac{x_t}{\mu_t} \right) \quad (4.4)$$

dove $\mu_t = \omega + \alpha x_t + \beta \mu_{t-1}$ *per* $t > 1$ *o* $\mu_t = \omega / (1 - \beta)$ *per* $t = 1,$

Allora

- 1. Il massimizzatore della verosimiglianza (L) è consistente e asintoticamente normale con matrice di covarianza data dalla famiglia di standard error robusti come in Lee e Hansen (1994).*
- 2. Il modello può essere stimato con il software sviluppato per i modelli GARCH, prendendo \sqrt{x} come variabile dipendente e ponendo la media a zero.*
- 3. Gli standard error robusti di Lee e Hansen (1994) coincidono con quelli trovati da Bollerslev e Wooldridge (1992).*

Dei tre risultati elencati, i due più rilevanti per l'analisi della dipendenza seriale, nel caso presente, sono i primi due, perché confermano le usuali proprietà dello stimatore di massima verosimiglianza utilizzato e, operativamente, permettono di utilizzare il software già sviluppato e convalidato per i modelli GARCH con l'accortezza di usare la radice quadrata della variabile presa in considerazione.

4.1.2 Applicazione alla serie degli autovalori

Per l'analisi sui dati reali ci si concentra sulla serie dell'autovalore più grande, ovvero della prima componente principale, che può a tutti gli effetti essere considerata come un indice della volatilità e del rischio del mercato, esprimendo il livello di movimento congiunto dei titoli. La serie è visualizzata nella figura 3.1. Si riportano alcuni grafici e indici riassuntivi della serie nella tabella 4.1 e nelle figure 4.1 e 4.2. Da questi indicatori si conferma la non negatività della serie e si nota

Minimo	1° quantile	Media	Mediana	3° quantile	Massimo	Skewness	Kurtosi
0.10	0.18	0.23	0.25	0.30	0.94	1.71	9.14

Tabella 4.1: Indici riassuntivi della serie dell'autovalore della prima componente principale.

che questa presenta una certa asimmetria e una coda pesante a destra. Inoltre dai grafici dell'autocorrelazione e dell'autocorrelazione parziale si nota che la seconda decade molto velocemente, mentre la prima rimane significativa anche per ritardi lunghi, e sembra stabilizzarsi dopo il ventesimo ritardo. Pertanto, considerata la premessa teorica su cui si basa il modello, si è deciso di procedere analizzando la serie tramite un modello a errori moltiplicativi con due parametri, α_0 e β_0 , avendo, quindi, una caratterizzazione della media come in 3:

$$\mu_t = \omega + \alpha_0 x_t + \beta_0 \mu_{t-1}. \quad (4.5)$$

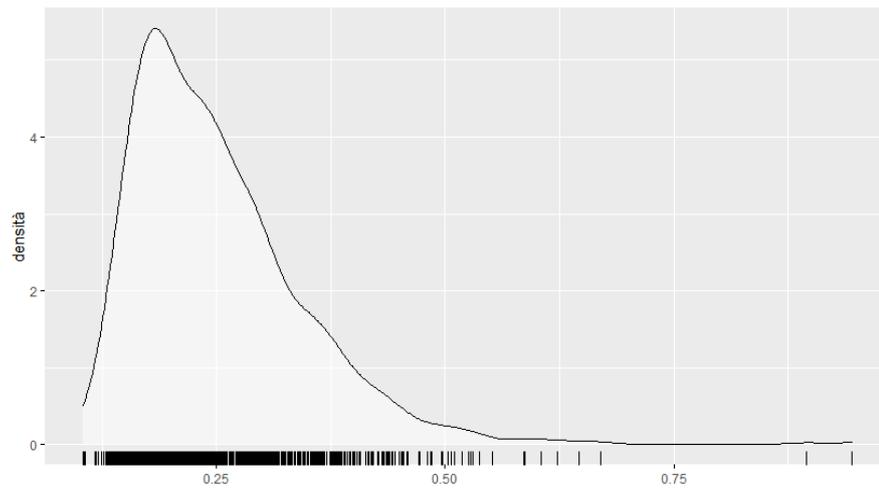


Figura 4.1: Densità dell'autovalore della prima componente principale

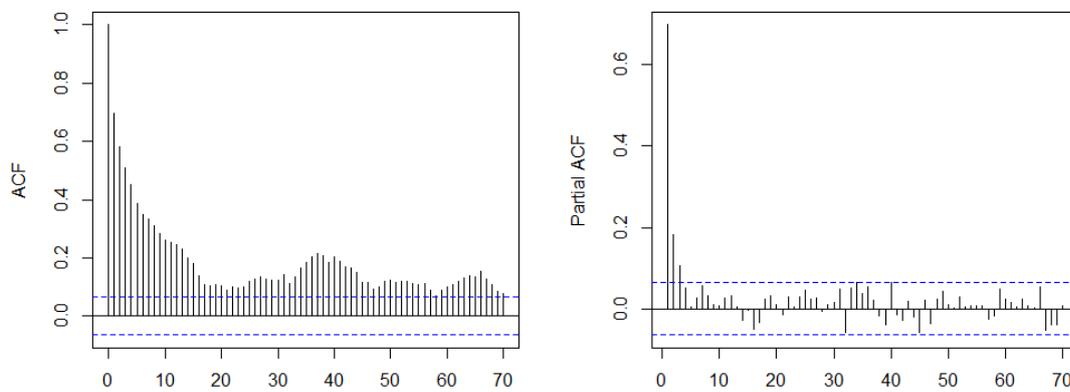


Figura 4.2: ACF e PACF della serie dell'autovalore della prima componente principale

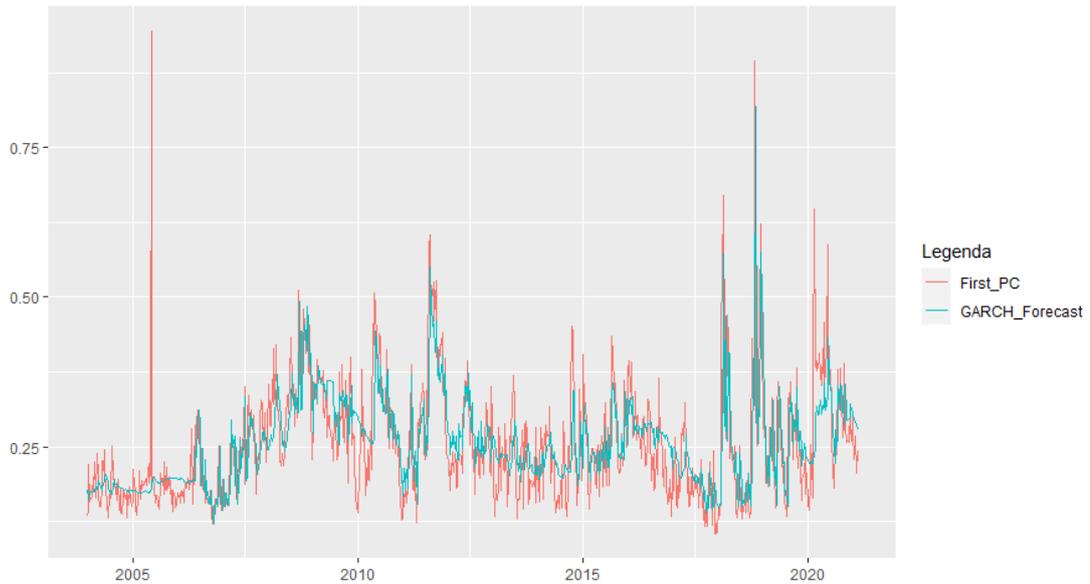


Figura 4.3: Serie dell'autovalore della prima componente principale dal 2004 al 2021 (in arancio) e previsione della serie con modello MEM(1,1) un passo avanti (in blu)

Per guardare al comportamento seriale dell'autovalore, il modello è stato stimato su finestre mobili di 50 settimane, corrispondenti a un anno di osservazioni, e ne è stata fatta la previsione un passo in avanti, per poi confrontarla con l'elemento corrispondente della serie. Il modello utilizzato è stato, quindi, un MEM(1,1), come specificato in 4.5, con distribuzione skew-student (sstd), in modo da tenere conto dell'asimmetria e delle code pesanti.

Il risultato ottenuto è stato visualizzato nella figura 4.3, dove si nota che il modello presenta una buona capacità previsiva un passo avanti (quindi prevedendo a una settimana) con solo un anno di dati. Si nota che con questa procedura è possibile catturare l'andamento della serie ignorando alcuni picchi dovuti allo scaricarsi improvviso della variabilità del mercato. Tramite questo metodo si è ottenuto un errore quadratico medio pari a 22.945. Si evidenzia, di conseguenza, una struttura deterministica che è possibile prevedere tramite modellazione. Si riportano nelle

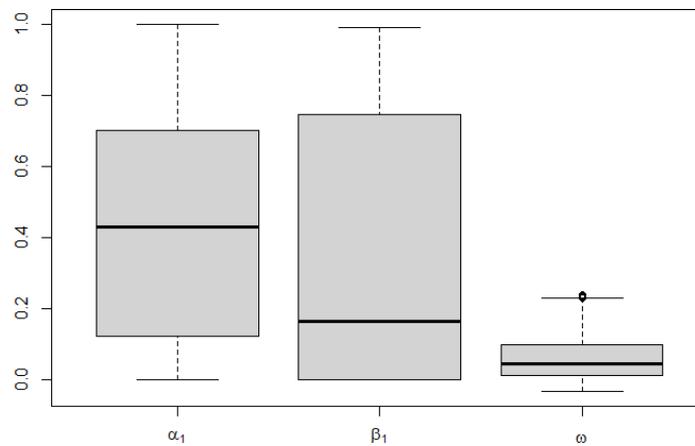


Figura 4.4: Boxplot dei parametri stimati del MEM(1,1) su finestre mobili nel periodo considerato

figure 4.4 e 4.5 i Box-Plot riassuntivi e le serie dei parametri stimati del modello MEM(1,1) su finestre mobili. È interessante notare come a cambiamenti nella stima dei parametri coincidano eventi estremi sul mercato, come la crisi finanziaria del 2008, il discorso di Draghi o la crisi epidemica.

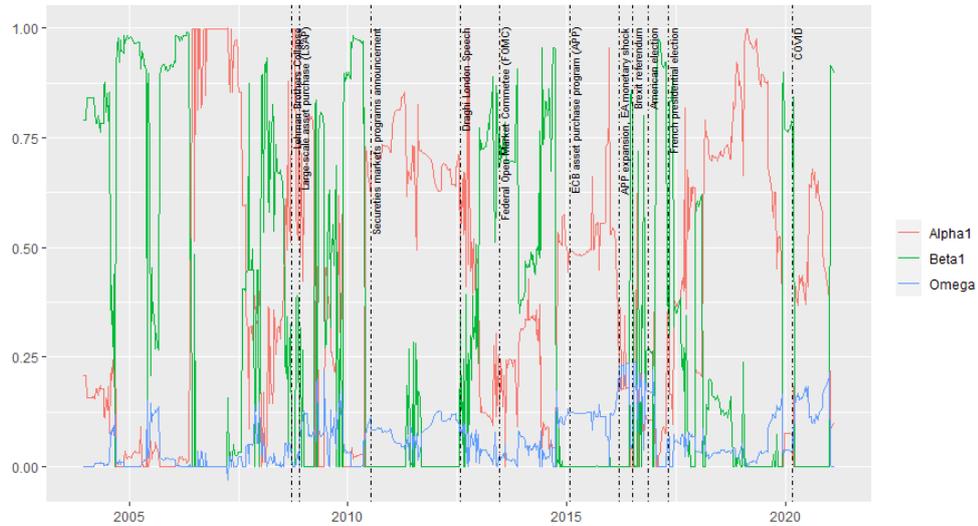


Figura 4.5: Serie dei parametri del modello MEM(1,1) stimati su finestre mobili

4.2 Analisi degli autovettori

Per gli autovettori l'analisi è stata condotta applicando una modellazione VAR, infatti questi non sono sempre risultati positivi e di conseguenza non è possibile beneficiare dei modelli a effetti moltiplicativi visti in precedenza. A causa delle dimensioni degli autovettori (89 componenti per autovettore) condurre un'analisi grafica preliminare completa risulta piuttosto complicato. Sono state riportate alcune analisi grafiche della serie del primo autovettore nelle figure 4.6, 4.7, 4.8 e 4.9 a titolo di esempio. Si può notare che alcune serie presentano una persistenza molto lunga mentre altre decadono velocemente, allo stesso modo è possibile notare come non tutte siano fortemente correlate con le altre. Questo ha portato alla scelta di analizzare le serie utilizzando un VAR. L'ordine del modello è stato scelto attraverso quattro criteri:

AIC(n)	HQ(n)	SC(n)	FPE(n)
1	1	1	1

Tutti i criteri riportati concordano nella scelta di un modello VAR(1), con cui si è proceduto. Si riportano, a titolo di esempio, in Appendice B i risultati ottenuti per il primo elemento del primo autovettore della serie (che corrisponde alla componente del titolo Apple). Come è possibile notare questo dipende principalmente dal suo primo ritardo, con un legame significativo rispetto alle variabili 34 e 40, che corrispondono ai titoli EXC (Exelon Corporation) e GS (The Goldman Sachs Group, Inc.). In generale questo processo può essere eseguito per tutti gli 89 elementi dell'autovettore, ed è possibile, in questo modo, determinare come i titoli si influenzino a vicenda. I Box-Plot riportati nelle figure 4.10 e 4.11 riassumono rispettivamente tutti i parametri stimati e i soli parametri significativi stimati dal modello VAR(1). Questi rappresentano, come detto in precedenza, quali aziende hanno una maggiore influenza sugli autovettori, quindi quali aziende aiutano a prevedere la composizione della prima componente. In generale è possibile affermare che tramite l'analisi delle componenti principali realizzate sui titoli si possono evidenziare delle strutture di correlazione negli autovalori che permettono di identificare come i titoli, e di conseguenza i settori, si influenzino a vicenda.

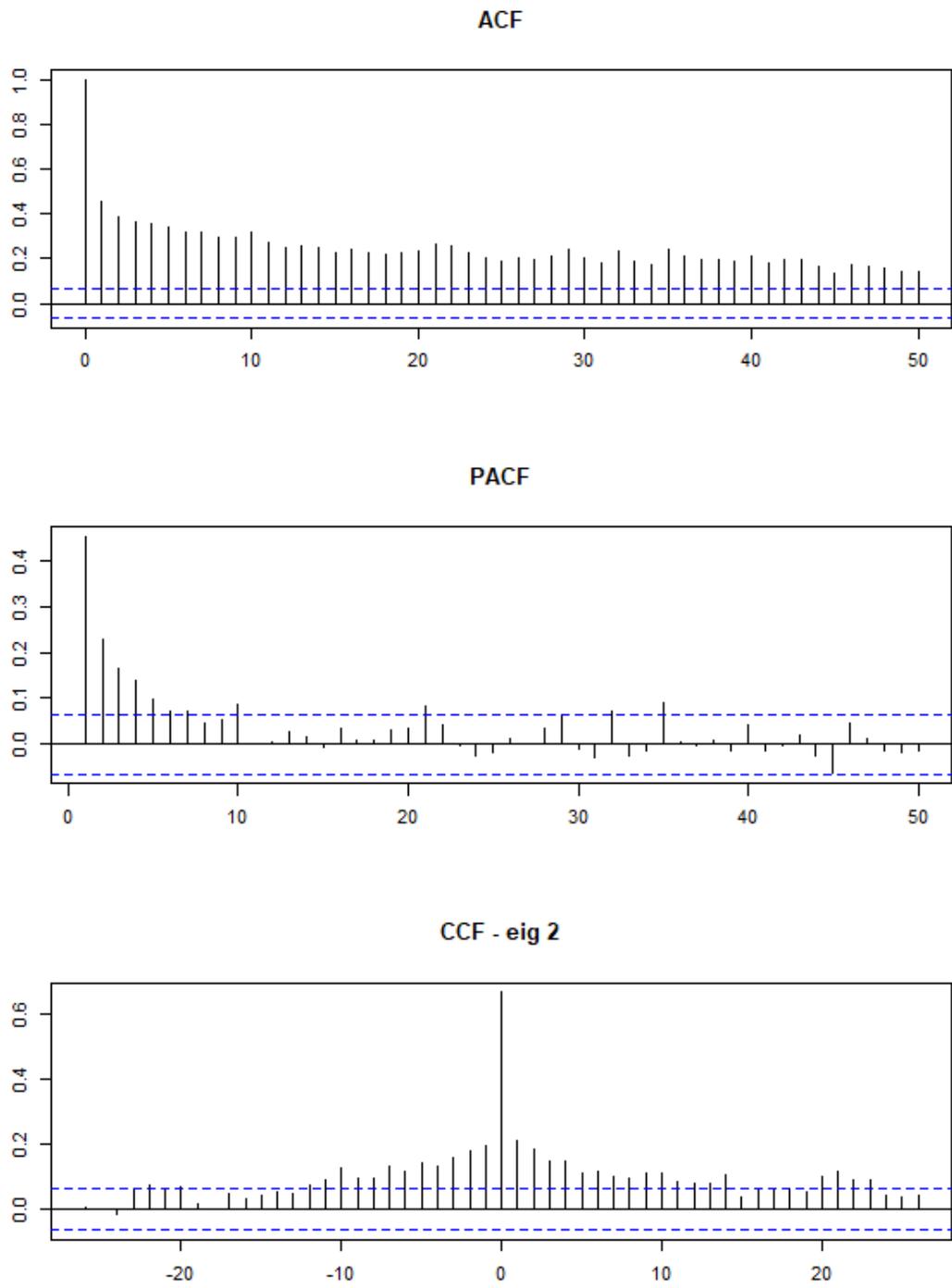


Figura 4.6: Nell'ordine: l'ACF della serie del primo elemento dell'autovettore, la corrispondente PACF e la CCF con il secondo elemento del primo autovettore

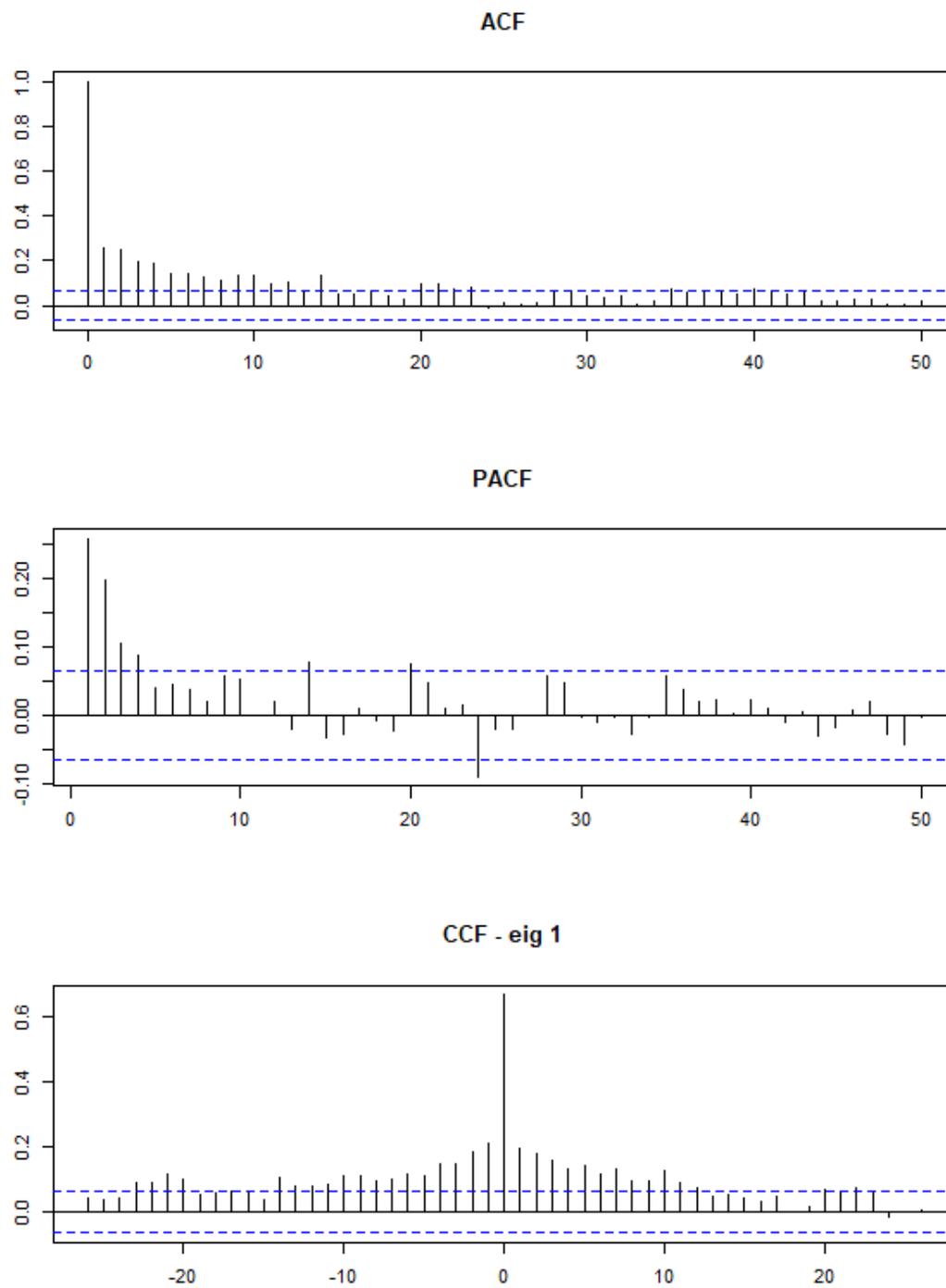


Figura 4.7: Nell'ordine: l'ACF della serie del secondo elemento del primo autovettore, la corrispondente PACF e la CCF con il primo elemento del primo autovettore

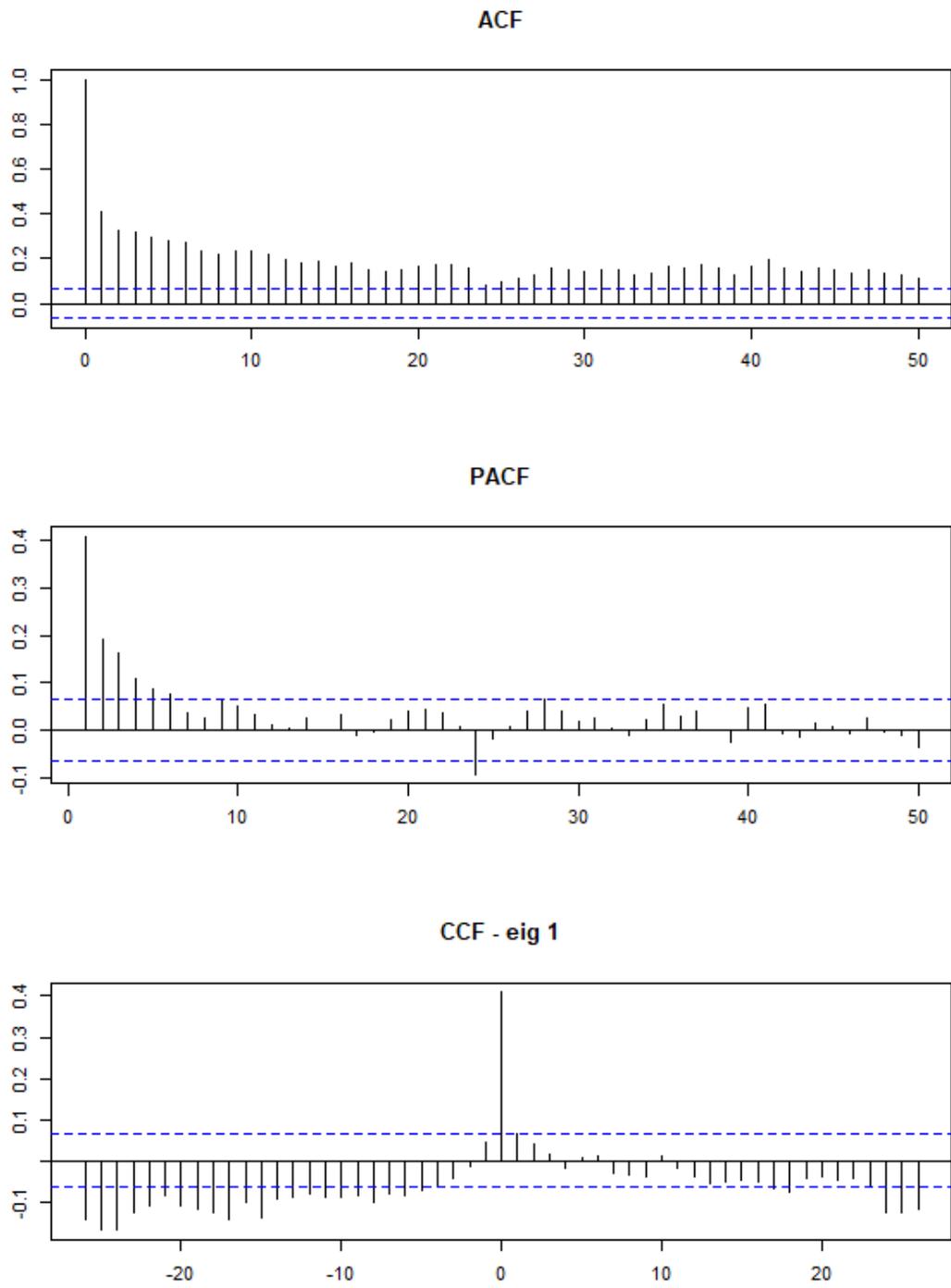


Figura 4.8: Nell'ordine: l'ACF della serie del terzo elemento del primo autovettore, la corrispondente PACF e la CCF con il primo elemento del primo autovettore

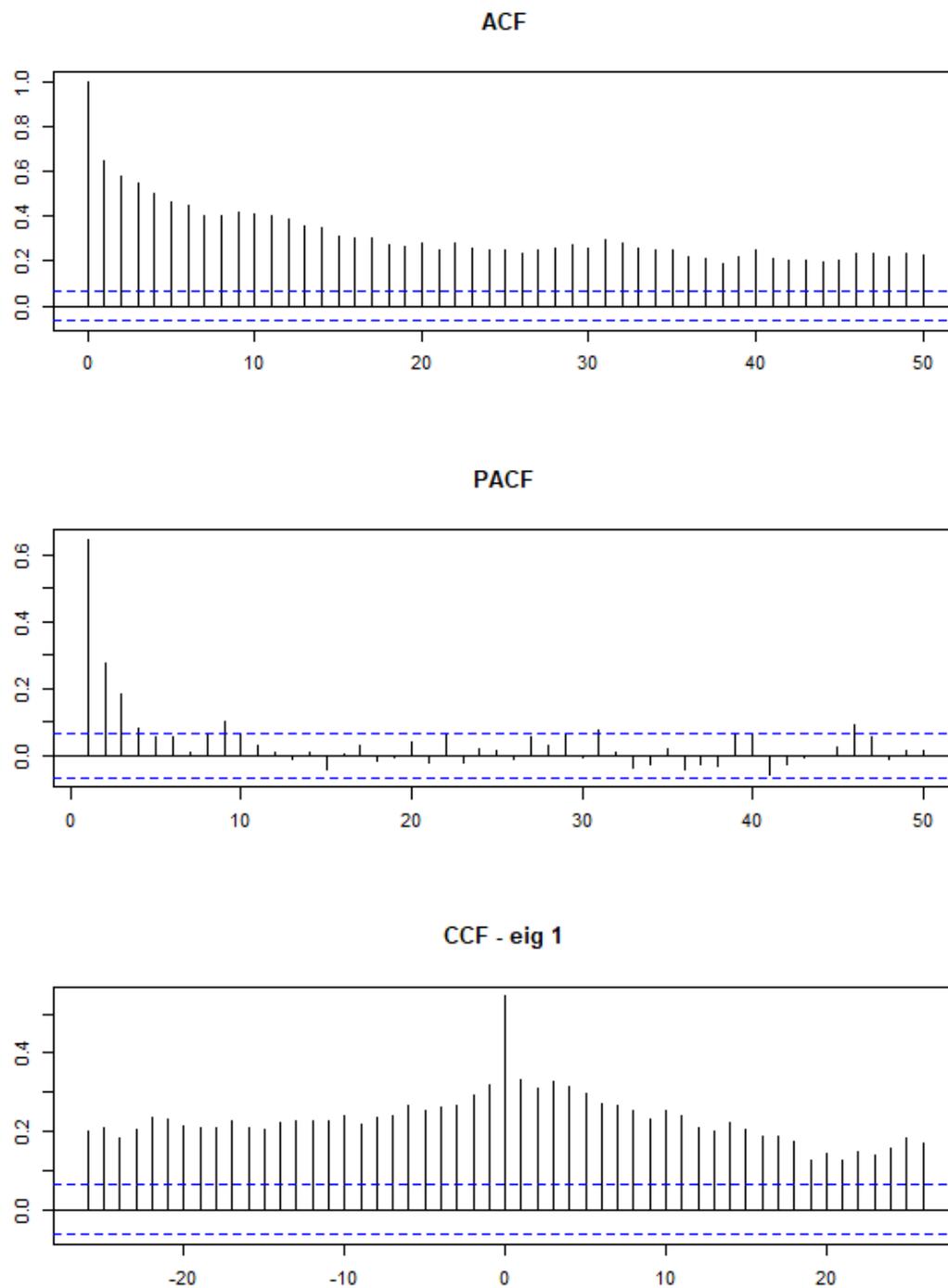


Figura 4.9: Nell'ordine: l'ACF della serie del quarto elemento del primo autovettore, la corrispondente PACF e la CCF con il primo elemento del primo autovettore

Capitolo 5

Spillover

I decisori della politica monetaria monitorano con attenzione le fluttuazioni nei prezzi dei titoli per individuare segnali sulle dinamiche attese sottostanti dell'economia. È stabilito che i prezzi dei titoli contengano un ricco set di informazioni che si propaga nel mercato, lo *spillover*, ma è meno chiaro se e come l'informazione contenuta nel co-movimento dei titoli possa predire future instabilità o crisi finanziarie. Le recenti crisi finanziarie hanno rinnovato l'interesse nella misurazione dello stress finanziario, definito da Kliesen e Smith (2010) come "l'insieme delle condizioni del mercato, in cui gli attori nel mercato sperimentano un aumento dell'incertezza o un cambiamento delle aspettative sulle future perdite finanziarie, valori sottostanti degli *asset*, e dell'attività economica". Entrambi gli aspetti, anche se non direttamente osservabili, si riflettono nel prezzo dei titoli. Secondo l'esperienza pregressa, forti stress finanziari sono preceduti da un aumento della volatilità del mercato, e il meccanismo attraverso il quale la volatilità si propaga tra i settori dell'attività economica ha una forte implicazione per investitori, regolatori e decisori. Questo potrebbe portare all'introduzione di regolamentazioni e regole istituzionali per ridurre l'impatto incrociato nel mercato di movimenti dei prezzi eccessivi, Laborda e Olmo (2021).

Usando la metodologia introdotta da Diebold e Yilmaz (2012), nella presente sezione si analizza la trasmissione degli shock da (e verso) il mercato dei titoli a (e da) altri indicatori del rischio di mercato. Si usa il primo autovalore realizzato come una approssimazione del contenuto informativo delle componenti principali ad alta frequenza. Infatti l'autovalore è la varianza della componente principale, quindi una misura di rischio.

Il sistema studiato include il primo il più grande autovalore realizzato e quattro indici addizionali, comunemente usati come misura del rischio di mercato, stress sistemico e incertezza politica:

- il VIX, o *CBOE Volatility Index*, introdotto dal *Chicago Board Options Exchange* (CBOE), chiamato anche *fear index* o *fear gauge*, è un indice che misura la volatilità attesa del mercato a 30 giorni nel futuro, basato sulle opzioni dell'S&P 500;
- lo SLFSI (*St. Louis Fed Financial Stress Index*), che misura il grado di stress finanziario nel mercato, costruito da una serie di 18 settimane, composte dalla media settimanale di rilevazioni giornaliere: sette indici di interesse, sei differenziali di rendimento e altri cinque indicatori;
- l'EPU (*Economic Policy Uncertainty Index*) costituito da tre componenti: un primo indice ottenuto dalle notizie riportate su 10 testate statunitensi, un secondo indice riferito al *Tax Code Expiration Data* e l'*Economic Forecaster Disagreement*;
- l'NFCI (*National Financial Conditions Index*), rilasciato settimanalmente, provvede una panoramica sulla condizione finanziaria negli Stati Uniti, in particolare sul mercato monetario, sul debito e sul mercato azionario e, inoltre, sul sistema bancario tradizionale e sullo *shadow banking system*, ovvero

il sistema bancario non vincolato alle stesse regole di vigilanza di quello tradizionale.

Gli ultimi tre indici sono pubblicati dalla Federal Reserve con frequenza settimanale per lo SLFSI e l'NFCI, mensile per l'EPU.

5.1 Metodologia

Per condurre un'analisi dello *spillover* ci si pone nel contesto autoregressivo multivariato, dove la scomposizione dell'errore di previsione è invariante rispetto all'ordine, Koop, Pesaran e Potter (1996) e Pesaran e Shin (1998), e dove vengono esplicitamente inclusi *spillover* della volatilità direzionali tra le componenti principali realizzate e gli indici elencati in precedenza, usando la metodologia proposta da Diebold e Yilmaz (2012), introdotta nel seguito.

Si consideri un VAR(p) di N variabili, con rappresentazione a media mobile infinita, $x_t = \Theta(L)\varepsilon_t$, con $\Theta(L) = I + \sum_{j=1}^{\infty} \Theta_j L^j$, e $E(\varepsilon_t \varepsilon_t') = \Sigma$. La scomposizione della varianza dell'errore di previsione permette di decomporre tale varianza come

$$\delta_{i,j} = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} (e_i' \Theta_h \Sigma e_j)^2}{\sum_{h=0}^{H-1} (e_i' \Theta_h \Sigma \Theta_h' e_i)}$$

dove σ_{jj} è il j -esimo della diagonale di Σ e e_j è un vettore di selezione, con il j -esimo elemento pari a uno e gli altri elementi pari a zero. Siccome gli shock nella decomposizione dell'errore di previsione non sono ortogonali, la somma del contributo dell'errore di previsione non è necessariamente pari a uno, $\sum_{j=1}^N \delta_{i,j} \neq 1$. Quindi nel calcolo dell'indice dello *spillover* le righe della matrice di varianza dell'errore di previsione sono normalizzate come

$$\tilde{\delta}_{i,j} = \frac{\delta_{i,j}}{\sum_{j=1}^N \delta_{i,j}},$$

dove per costruzione $\sum_{j=1}^N \tilde{\delta}_{ij} = 1$ e $\sum_{i,j=1}^N \tilde{\delta}_{ij} = N$. La decomposizione della varianza permette di valutare la frazione dell'errore della varianza H passi avanti

causato da shock sulle altre variabili incluse nel sistema. Si fissa a priori l'ordine del modello VAR scegliendo $p = 4$, imponendo un ritardo mensile, e ci si concentra sulla previsione a 2 settimane, $H = 2$, come in Diebold e Yilmaz (2012).

Si definisce $S_{i \leftarrow j} = \tilde{\delta}_{ij}$ come *cross-variance shares* o *spillover*, ovvero la frazione della varianza dell'errore H passi avanti nella previsione x_i dovuta a shock su x_j , per $i, j = 1, 2, \dots, N$, con $i \neq j$. Si noti che $S_{i \leftarrow j} \neq S_{j \leftarrow i}$, quindi sono presenti $N^2 - N$ *spillover*, in questo caso 20.

Si definisce *gross directional volatility spillover*, ricevuto da x_i da tutte le altre x_j per $i \neq j$, $j = 1, 2, \dots, N$, come

$$S_{i \leftarrow \cdot} = \sum_{\substack{j=1 \\ j \neq i}}^N \tilde{\delta}_{ij},$$

e *gross directional volatility spillover*, trasmesso da x_i a tutti gli altri x_j per $i \neq j$, $j = 1, 2, \dots, N$, come

$$S_{\cdot \leftarrow i} = \sum_{\substack{j=1 \\ j \neq i}}^N \tilde{\delta}_{ij}.$$

Sono, quindi, presenti $2N$ *directional volatility spillover*, N ricevuti dalle altre variabili e N trasmessi alle altre variabili. Si ottiene il *net volatility spillover* da x_i a tutte le altre x_j per $i \neq j$, $j = 1, 2, \dots, N$, come la differenza tra il *gross directional volatility spillover* trasmesso e *gross directional volatility spillover* ricevuto,

$$S_i = S_{\cdot \leftarrow i} - S_{i \leftarrow \cdot},$$

sono presenti N *net volatility spillover*.

Infine, si calcola il *total volatility spillover index* che misura il contributo medio della trasmissione degli shock della volatilità tra le N variabili e l'errore di previsione totale,

$$S = \frac{1}{N} \sum_{\substack{i,j=1 \\ j \neq i}}^N \tilde{\delta}_{ij}.$$

5.2 Risultati

La tabella 5.1 mostra gli *spillover* tra il primo autovalore (EIGV1) e le quattro macro-variabili finanziarie elencate in precedenza: VIX, EPU, NFCI e SLFSI per tutto il periodo in cui sono state campionate le osservazioni. Ogni elemento ij della tabella rappresenta il contributo stimato alla varianza dell'errore di previsione della variabile i proveniente dalla variabile j . La riga della tabella chiamata "Agli altri" è la somma delle colonne della tabella senza la diagonale e rappresenta il *directional volatility spillover* trasmesso dalla variabile i a tutte le altre variabili j . La colonna denominata "Dagli altri" è, invece, la somma delle righe della tabella senza la diagonale e rappresenta il *directional volatility spillover* ricevuto dalla variabile i rispetto a tutte le altre variabili j . La riga denominata "Net" è la differenza tra la volatilità lorda trasmessa a e quella ricevuta da tutte le altre variabili. Inoltre nell'angolo in basso a destra compare il *total spillover index*, che rappresenta la percentuale media della varianza dell'errore di previsione delle cinque variabili risultante dallo *spillover*. Può essere calcolata come media sia della riga "Agli altri" che della colonna "Dagli altri".

La matrice diagonale della tabella 5.1 mostra che più del 30% dell'errore di previsione è spiegato dalla variabile stessa per ogni variabile, in particolare la percentuale auto-spiegata è più del 90% per il primo autovalore e per il VIX, mentre scende all'88% per l'EPU e poi al 60% e 30% rispettivamente per l'NFCI e lo SLFSI. Guardando alla riga "Agli altri", gli shock che colpiscono SLFSI e VIX ricadono in gran parte sugli altri indici (64% e 62% rispettivamente), in particolare, però, il contributo alla volatilità al primo autovalore (EIGV1) è solamente dell'1% per lo SLFSI e del 5% per il VIX, che ha il contributo più alto tra gli indici. All'alto contributo degli indici VIX e SLFSI seguono quasi alla pari l'NFCI e l'EIGV1. Guardando, invece, alla colonna "Dagli altri" lo *spillover* della volatilità ricevuto più grande è quello dello SLFSI, con ben il 66%, seguito dall'NFCI (36%) per poi

scendere in ordine all'EPU (11%), EIGV1 (8%) e VIX (7%). Riguardo all'effetto netto, espresso dalla riga "Net" della tabella, i più grandi contributi sono dati dal VIX agli altri (54%) e dal primo autovalore agli altri (20%). Il *total spillover index* è decisamente alto, al 26%, indicando che in media la volatilità dell'errore di previsione per i 5 indici proviene dagli *spillover*. In prospettiva, Diebold e Yilmaz (2012) nel loro lavoro trovano un indice pari al 13%. I risultati ottenuti indicano, di conseguenza, che in questo caso il *total* e il *directional spillover* nel campione è decisamente ampio.

	EIGV1	VIX	EPU	NFCI	SLFSI	Dagli altri
EIGV1	91.33	5.10	0.28	2.21	1.09	8.68
VIX	7.22	92.27	0.10	0.29	0.12	7.73
EPU	4.42	7.08	88.13	0.28	0.08	11.86
NFCI	8.79	24.50	2.52	63.97	0.22	36.03
SLFSI	8.90	25.78	3.21	28.75	33.37	66.63
Agli altri	29.33	62.46	6.11	31.53	64.3	Total Spillover Index
Net	20.65	54.73	-5.75	-4.50	-2.33	26.19

Tabella 5.1: *Spillover* della volatilità tra il primo autovalore (EIGV1) e gli indici di mercato per tutto il periodo.

Un ulteriore approfondimento dei risultati porta allo studio dell'interazione dinamica tra le variabili. La figura 5.1 risponde a questa esigenza riportando i *net pairwise spillover* calcolati tramite finestre mobili di 200 settimane (circa 4 anni). I *net pairwise spillover* tra EIGV1 e gli altri indici sono la differenza tra lo *spillover* volatilità da EIGV1 a un altro indice e quello trasmesso da un altro indice a EIGV1. Quindi, un *net pairwise spillover* positivo indica che la ricaduta di volatilità da EIGV1 a un altro *asset* è più grande rispetto a quella ricevuta da EIGV1 rispetto all'altro *asset*. Viceversa se il *net pairwise spillover* è negativo lo

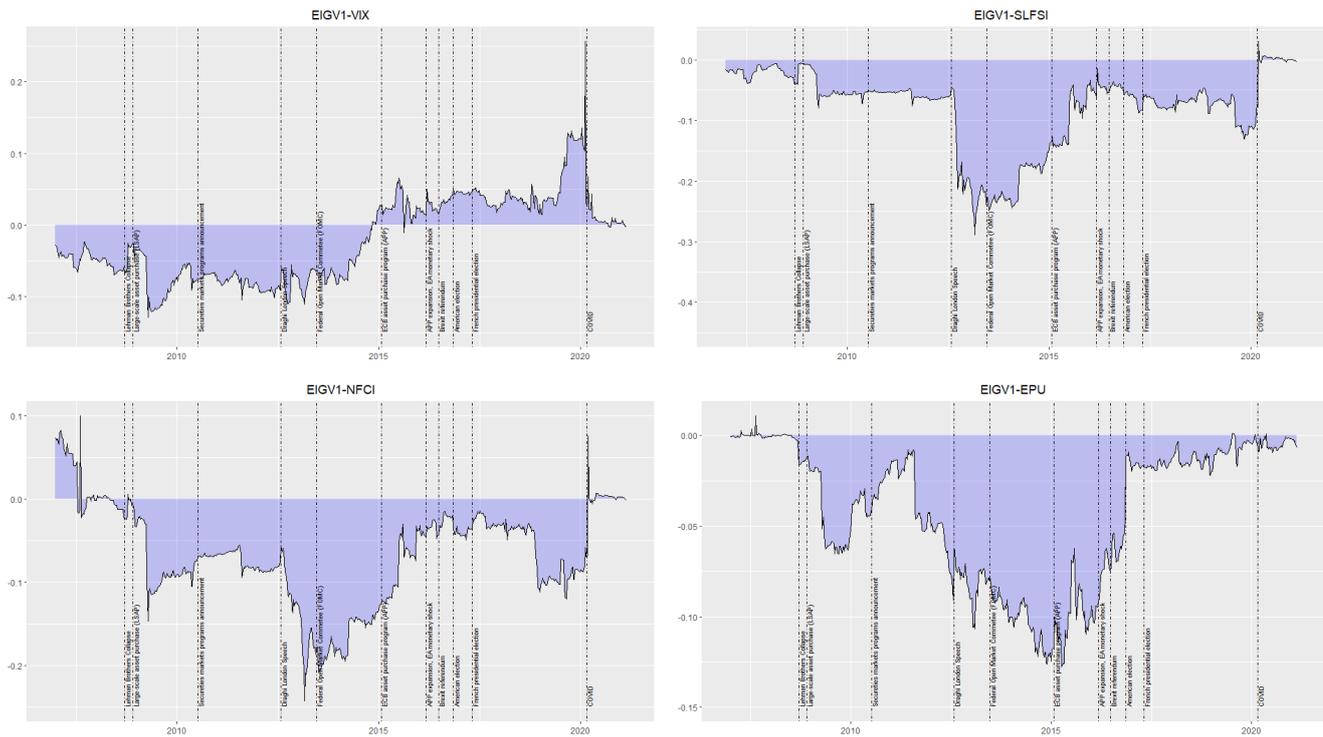


Figura 5.1: *Net pairwise spillovers* tra la radice di EIGV1 e il VIX, lo SLFSI, il NFCI e l'EPU.

spillover ricevuto da EIGV1 rispetto a un altro asset è stato più grande di quello trasmesso.

Il pannello in alto a sinistra nella 5.1 mostra i *net pairwise spillovers* da EIGV1 al VIX. La serie comincia da un valore negativo e si mantiene sotto lo zero fino al 2015, attorno al periodo in cui è avvenuto l'*asset purchase program* (APP), dove ha cambiato di segno, a indicare che fino all'APP lo *spillover* netto è stato ricevuto da EIGV1, con una quota variabile nel periodo, che va da un minimo di circa il 2% fino a un massimo del 12%, successivamente alla crisi del 2008. Dopo il 2015 la serie dello *spillover* si mantiene sopra lo zero, aggirandosi attorno al 3% in media, con un forte rialzo oltre il 10% prima della crisi legata al COVID-19. In corrispondenza della crisi si nota un picco superiore al 20%, per poi stabilizzarsi

nel corso dell'anno attorno allo zero. Passando al secondo pannello in alto a destra è rappresentato il *net pairwise spillover* tra il primo autovalore e lo SLFSI. Si nota che il contenuto informativo è sempre a favore dell'indice, con un incremento dopo la crisi finanziaria del 2008, dove si stabilizza attorno al 5% fino al discorso del presidente della BCE Mario Draghi. Successivamente all'evento il *net pairwise spillover* passa in valore assoluto dal 5% a oltre il 20%, scendendo negli anni successivi. Vi è, infine, un crollo dopo l'inizio dell'epidemia di COVID-19.

Nel pannello in basso a sinistra vi è la relazione tra EIGV1 e l'NFCI, si nota che nel primo anno l'indice è un ricevitore netto della volatilità da EIGV1, successivamente, prima della crisi del 2008, sembra non esserci alcun contributo da entrambe le parti, per poi, dopo la crisi finanziaria, avere un lungo periodo, fino all'epidemia del 2020, in cui l'indice trasmette al primo autovalore più varianza di quanta ne riceve, con un aumento sostanziale dopo il discorso di Draghi, da poco sotto il 9% fino al 20%, per poi calare gradualmente fino al 2016. Infine, dopo la crisi pandemica, sembra non esserci un contributo da entrambe le parti.

Nell'ultimo pannello, è riportato il *net pairwise spillover* tra EIGV1 e l'EPU, dove si vede un contributo nullo fino alla crisi del 2008, per poi avere un primo aumento nel 2009 e in seguito un calo durante gli anni 2010 e 2011, dopo di che vi è un risalita al 10% in corrispondenza del discorso di Draghi e un continuo aumento fino al programma di acquisto degli *asset*. Dopo di che vi è un continuo calo in corrispondenza della Brexit e delle elezioni statunitensi del 2016. Per tutto questo periodo il contributo alla varianza netto è sempre negativo, quindi ricevuto da EIGV1.

La tabella 5.2 mostra gli *spillovers* tra il secondo autovalore (EIGV2) e i quattro indici macroeconomici per tutto il periodo. Si nota un andamento simile a quello visto nella tabella 5.1, in particolare la varianza dell'errore di previsione del secondo autovalore e dell'indice VIX è quasi interamente spiegata dalle variabili stesse, con

uno *spillover* tra i due indici pari solamente all'1%. Il contributo netto di EIGV2 è quasi nullo, pari, in valore assoluto, solo allo 0.14%, mentre continua a rimanere molto elevato per il VIX e in questo caso, in valore assoluto, è molto alto anche per lo SLFSI. Si nota, guardando alla colonna "Dagli altri" che quest'ultimo riceve da tutte le altre variabili un importante contributo alla variabilità, pari al 65%, seguito dall'indice NFCI, che però riceve solo il 34%. Osservando la riga "Agli altri" si nota come il contributo maggiore alla trasmissione di variabilità verso gli altri sia dato dal VIX (69%), seguito dall'NFCI (32%). Infine, il *total spillover index* continua a essere elevato, al 22,88%.

	EIGV2	VIX	EPU	NFCI	SLFSI	Dagli altri
EIGV2	96.40	1.04	0.54	1.13	0.88	3.60
VIX	0.85	98.51	0.15	0.42	0.07	1.49
EPU	0.80	8.73	89.97	0.42	0.08	10.03
NFCI	1.03	29.79	3.11	65.78	0.28	34.22
SLFSI	0.77	30.04	4.11	30.16	34.92	65.08
Agli altri	3.45	69.61	7.92	32.13	1.31	Total Spillover Index
Net	-0.14	68.12	-2.11	-2.09	-63.78	22.88

Tabella 5.2: *Spillover* della volatilità tra il secondo autovalore (EIGV2) e gli indici di mercato per tutto il periodo.

Lo studio del *net pairwise spillover* tramite la figura 5.2 mostra un pattern diverso rispetto a quello analizzato nel punto precedente. In questo caso per tutto il periodo considerato il *net pairwise spillover* tra EIGV2 e il VIX si alterna di segno, stabilizzandosi per un certo periodo dopo il 2010 e fino al 2012. Si torna poi a una situazione di instabilità, dove il contributo alla varianza si alterna tra EIGV2 e VIX, per poi avere un periodo in cui il secondo autovalore diventa un ricevitore netto, dal 2014 ai primi mesi del 2016. Dopo il 2017 vi è un primo periodo in cui il

secondo autovalore è ancora un ricevitore netto, per poi diventare un contributore netto alla variabilità fino alla crisi pandemica. In generale, comunque, il contributo assoluto non supera quasi mai il 2,5%. L'analisi EIGV2 - SLFSI mostra come i due indici non si influenzino in modo marcato per tutta la durata delle analisi. Guardando al rapporto tra EIGV2 e NFCI si osserva che, oltre a un primo periodo nel 2007, il contributo tra le due variabili si mantenga quasi nullo, se non dopo il referendum sulla Brexit, quando l'NFCI è un contributore netto, per circa un anno. Infine, guardando al *net pairwise spillover* tra EIGV2 e EPU si nota che il secondo è un contributore netto fino alla crisi del 2008, dopo di che il contributo è nullo fino al discorso di Draghi, dove lo *spillover* torna ad essere negativo, ma non fortemente. Si ha un picco oltre il 2,5% solo dopo la crisi del COVID-19, dove l'autovalore preso in considerazione è un contributore netto alla varianza. In generale tutti i contributi sono bassi, rimanendo sotto la soglia del 5%.

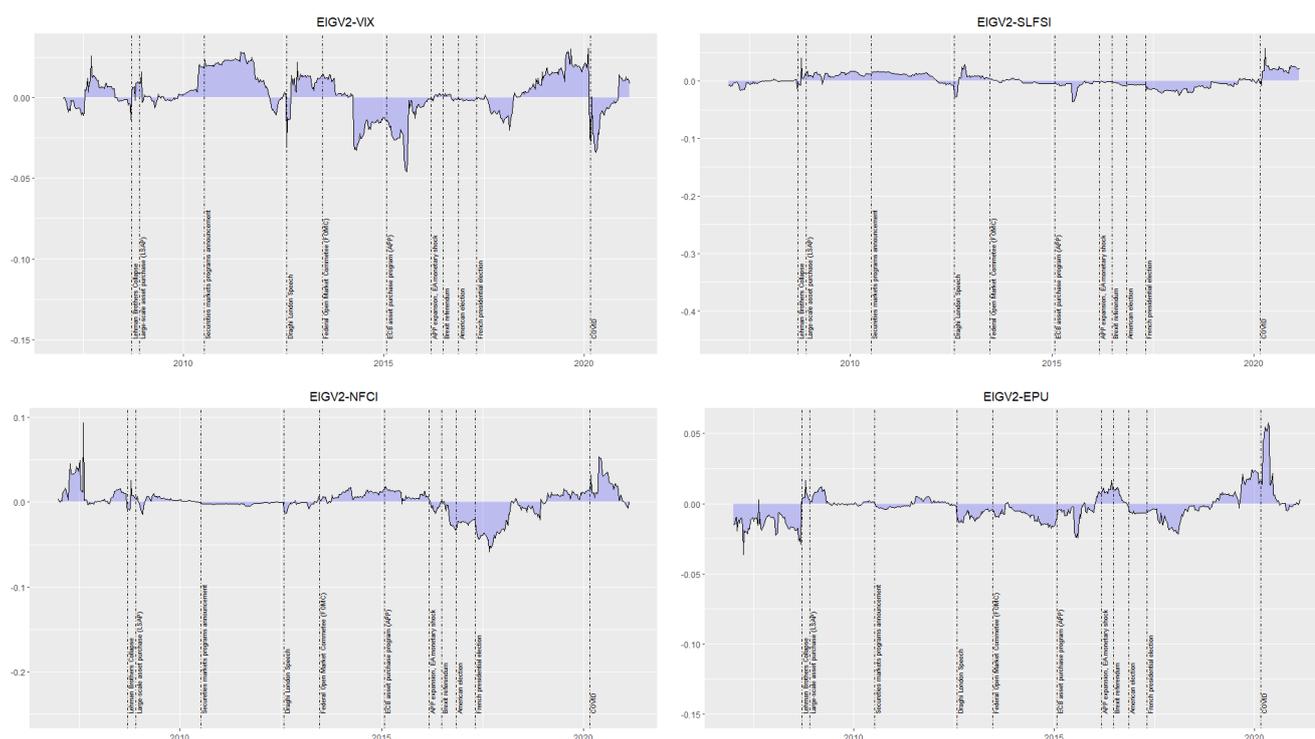


Figura 5.2: *Net pairwise spillovers* tra la radice di EIGV2 e il VIX, lo SLFSI, il NFCI e l'EPU.

La tabella 5.3, infine, riassume lo *spillover* tra il terzo autovalore EIGV3 e i quattro indici. Anche qui il terzo autovalore e il VIX spiegano quasi per intero la loro volatilità, con un contributo del VIX a EIGV3 pari quasi al 4%. Di nuovo l'indice che riceve un contributo maggiore dagli altri è lo SLFSI, seguito dall'NFCI. In termini di contributo dato alle altre variabili è ancora il VIX che si attesta come maggior contributore lordo (70.56%) che netto (67.81%), mentre il maggior ricevitore netto è lo SLFSI. Il *total spillover index* è ancora alto, attestandosi al 23.81%.

	EIGV3	VIX	EPU	NFCI	SLFSI	Dagli altri
EIGV3	93.42	3.73	0.50	1.49	0.87	6.58
VIX	2.09	97.26	0.18	0.36	0.12	2.74
EPU	1.08	8.85	89.58	0.41	0.08	10.41
NFCI	1.38	29.22	3.16	65.99	0.25	34.01
SLFSI	2.40	28.76	4.05	30.10	34.69	65.31
Agli altri	6.94	70.56	7.90	32.35	1.32	Total Spillover Index
Net	0.36	67.81	-2.52	-1.66	-63.99	23.81

Tabella 5.3: *Spillover* della volatilità tra il terzo autovalore (EIGV3) e gli indici di mercato per tutto il periodo.

La figura 5.3 riassume allo stesso modo degli altri il *net pairwise spillover*. Al contrario del secondo autovalore, qui i contributi sembrano essere, soprattutto per il VIX, generalmente più alti, in valore assoluto. Si noti come nel caso del VIX, questo sembra essere un ricevitore detto della varianza nel periodo che va dal 2017 fino alla crisi pandemica.

Si conclude con una modellazione sia sulle serie degli autovalori che sulla serie del VIX per osservare meglio la dipendenza tra le variabili. La prima modellazione è stata fatta sulle tre serie degli autovalori, per osservare se gli indici utilizzati sono in grado di aiutare nel fare previsione. Le tabelle 5.4, 5.5 e 5.6 riportano i parametri stimati e il loro errore, come si può notare in tutti i casi il VIX è l'unico indice che, ritardato di un periodo, è significativo nell'aiutare a prevedere i valori degli autovalori.

Quindi si analizza la relazione esistente tra l'indice VIX e il primo autovalore, EIGV1. Per fare questo sono stati stimati due modelli ARMA, le cui stime sono contenute nelle tabelle 5.7 e 5.8. Il primo modello è un ARMA(1,0), scelto guardando prima ai grafici delle autocorrelazioni del VIX, figura 5.4, dove sem-

	Estimate	Std. Error	t value	Pr(> t)
α_1	0.5351	0.0439	12.2000	0.0000
β_1	0.0496	0.0528	0.9387	0.3479
VIX	0.0013	0.0003	3.9853	0.0001
SLFSI	0.0000	0.0001	0.0781	0.9377
NFCI	0.0000	0.0000	0.4511	0.6519
EPU	0.0000	0.0023	0.0000	1.0000
skew	0.3814	0.0120	31.6972	0.0000
shape	9.7883	0.9250	10.5822	0.0000
ω	0.0009			

Tabella 5.4: Parametri stimati, sul primo autovalore, del modello MEM(1,1) con tutti gli indici come regressori al primo ritardo.

	Estimate	Std. Error	t value	Pr(> t)
α_1	0.22357	0.0204	10.9612	0.0000
β_1	0.62696	0.0145	43.2501	0.0000
VIX	0.00003	0.0000	2.6352	0.0084
SLFSI	0.00000	0.0000	0.0011	0.9992
NFCI	0.00000	0.0000	0.0000	1.0000
EPU	0.00000	0.0001	0.0002	0.9998
skew	0.31318	0.0127	24.7189	0.0000
shape	13.21240	1.8349	7.2005	0.0000
ω	0.00071			

Tabella 5.5: Parametri stimati, sul secondo autovalore, del modello MEM(1,1) con tutti gli indici come regressori al primo ritardo

	Estimate	Std. Error	t value	Pr(> t)
α_1	0.223565	0.0204	10.9612	0.0000
β_1	0.626958	0.0145	43.2501	0.0000
VIX	0.000029	0.0000	2.6352	0.0084
SLFSI	0.000000	0.0000	0.0011	0.9992
NFCI	0.000000	0.0000	0.0000	1.0000
EPU	0.000000	0.0001	0.0002	0.9998
skew	0.313176	0.0127	24.7189	0.0000
shape	13.212405	1.8349	7.2005	0.0000
ω	0.000715			

Tabella 5.6: Parametri stimati, sul terzo autovalore, del modello MEM(1,1) con tutti gli indici come regressori al primo ritardo

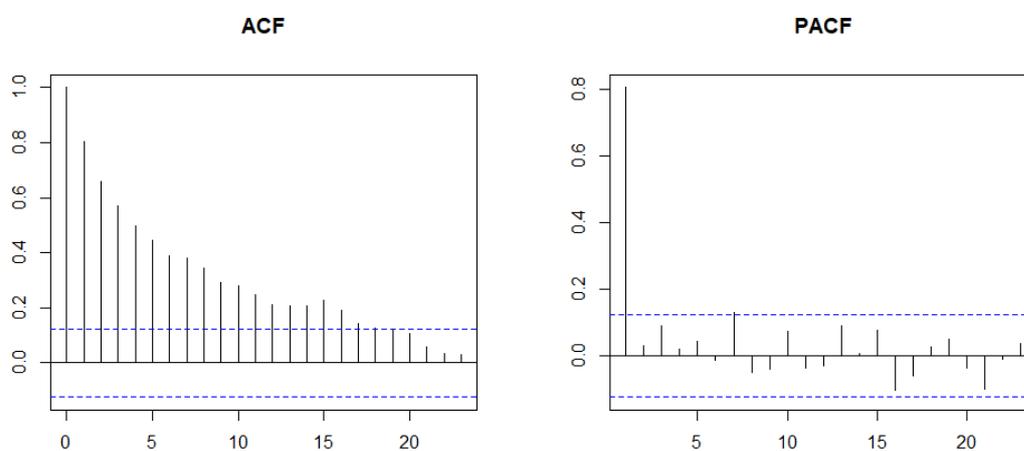


Figura 5.4: Grafico dell'autocorrelazione e autocorrelazione parziale dell'indice VIX.

Tabella 5.7: Output del modello ARMA(1,0) applicato al VIX.

<i>Variabili dipendenti:</i>	
ar1	0.805*** (0.037)
intercetta	19.990*** (1.588)
Osservazioni	247
Log-verosimiglianza	-745.622
σ^2	24.418
Akaike Inf. Crit.	1,497.244

Note: *p<0.1; **p<0.05; ***p<0.01

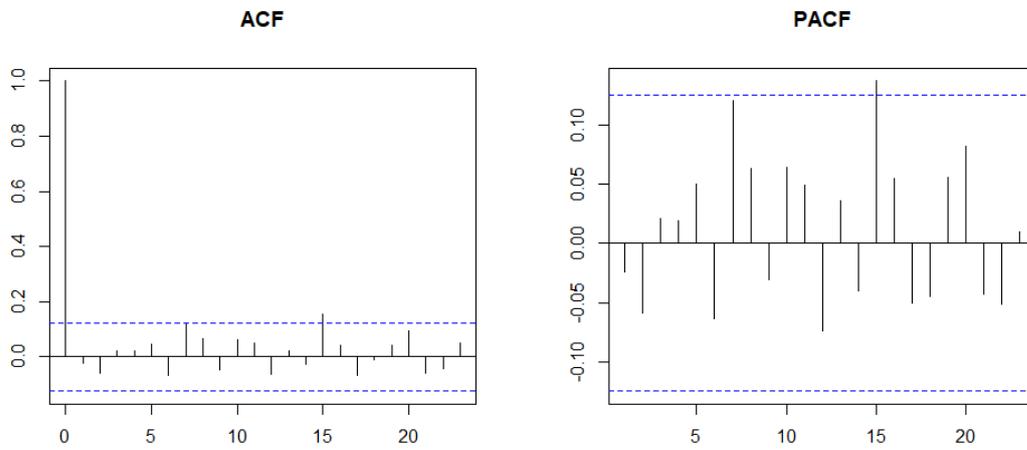


Figura 5.5: Grafici dell'autocorrelazione dei residui del modello ARMA(1,0) applicato al VIX senza regressori.

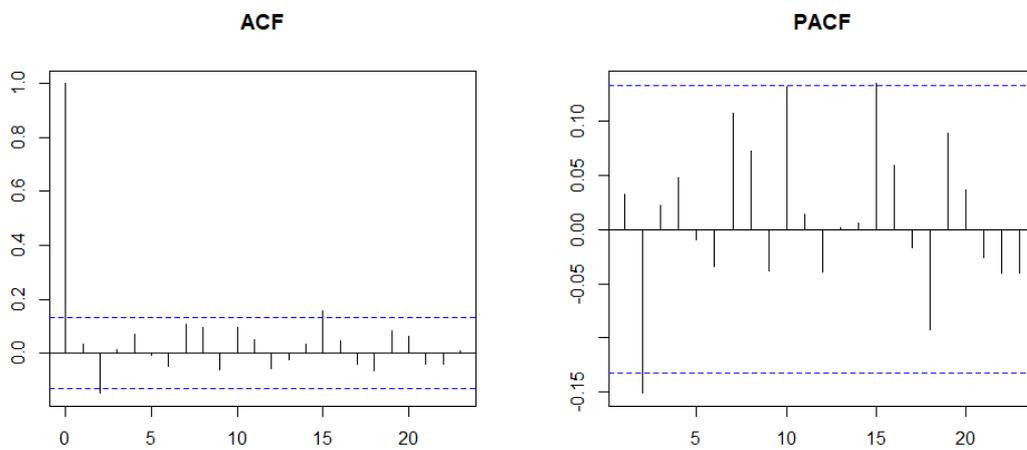


Figura 5.6: Grafici dell'autocorrelazione dei residui del modello ARMA(1,0) applicato al VIX utilizzando come ulteriore regressore il primo autovalore ritardato.

Tabella 5.8: Output del modello ARMA(1,0) applicato al VIX con EIGV1 utilizzato come regressore al primo ritardo.

<i>Variabili dipendenti:</i>	
ar1	0.809*** (0.041)
intercetta	18.780*** (2.145)
Prima_pc	2.506 (5.203)
Osservazioni	218
Log-verosimiglianza	-654.814
σ^2	23.682
Akaike Inf. Crit.	1,317.628

Note: *p<0.1; **p<0.05; ***p<0.01

Capitolo 6

Conclusioni

Dalle osservazioni fatte nel Capitolo 3, Capitolo 4 e Capitolo 5 si può concludere che le componenti principali realizzate applicate a dati ad alta frequenza nel contesto finanziario si rivelino essere uno strumento molto potente. Innanzitutto permettono di riassumere nelle prime componenti principali una grande quantità di informazioni riguardanti l'andamento della volatilità del mercato e la sua composizione. In particolare l'andamento dei primi autovettori, che a loro volta rappresentano la percentuale di volatilità spiegata dalle singole componenti, può essere utilizzato come una approssimazione della volatilità del mercato. La prima componente principale riassume il co-movimento dei titoli nel mercato, agendo, di conseguenza, sia come una misura di rischio di mercato che come un indice di fragilità del sistema finanziario. Il Capitolo 3 mette in evidenza nella sottosezione 3.2.1 l'andamento delle componenti principali, come i picchi di queste coincidano con momenti di stress finanziario internazionale, ad esempio il default di Lehman Brothers, i piani di *buy-back* delle banche centrali o la crisi epidemica, e, infine, come le medie annuali della volatilità spiegata dalle componenti principali siano più alte negli anni di crisi. La sottosezione 3.2.2 mette, invece, in evidenza la composizione delle componenti principali, permettendo di analizzare quale settore contribuisce

maggiormente alla volatilità.

I risultati del Capitolo 4 sono ancora più interessanti. Una volta appurato per via empirica la consistenza degli stimatori si procede all'analisi della loro correlazione seriale, per notare se vi sia la possibilità di fare previsioni sul rischio futuro. Questo viene confermato nella sottosezione 4.1.2, dove attraverso la modellazione MEM si è in grado di mostrare che la previsione un passo avanti risulta essere molto buona. Inoltre, la sezione 4.2 mostra come vi sia la presenza di correlazione seriale negli autovettori e che questa possa essere descritta tramite modelli VAR.

Infine il Capitolo 5 mette in evidenza come le serie degli autovalori delle prime componenti principali siano influenzate e, a loro volta, sembrano influenzare, l'andamento di alcuni indici macro-economici, trovando che a tutti gli effetti, anche se la maggior parte della volatilità è spiegata dalla variabile stessa, sia presente l'influenza delle altre variabili, in particolare del VIX.

Appendice A

Titoli utilizzati

Nella seguente tabella sono riportati tutti i titoli utilizzati nelle analisi e le loro caratteristiche. Nelle analisi sono stati utilizzate solo le osservazioni rilevate a partire dal 10 gennaio 2003 e fino al 23 febbraio 2021. Si faccia attenzione che per il titolo "WBA", *Walgreens Boots Alliance*, non è stata fornita alcuna classificazione e non è stato possibile integrarla successivamente. Per ogni titolo è stata riportata la corrispondente sigla, la data di inizio rilevazione, la descrizione del titolo (il nome esteso dell'azienda), il mercato il cui il titolo è scambiato, l'industria e il settore.

#	Symbol	StartDate	Description	Exchange	Industry	Sector
1	AAPL	1998-02-01	Apple Inc.	NASDAQ	Electronic Equipment	Consumer Goods
2	ABT	1998-02-01	Abbott Laboratories	NYSE	Drugs - Generic	Healthcare
3	ACN	7/19/2001	Accenture plc	NYSE	Information Technology Services	Technology
4	AEP	1998-02-01	American Electric Power Co., Inc.	NYSE	Electric Utilities	Utilities
5	AES	1998-02-01	The AES Corporation	NYSE	Electric Utilities	Utilities
6	AIG	1998-02-01	American International Group, Inc.	NYSE	Property & Casualty Insurance	Financial
7	ALL	1998-02-01	The Allstate Corporation	NYSE	Property & Casualty Insurance	Financial
8	AMGN	1998-02-01	Amgen Inc.	NASDAQ	Biotechnology	Healthcare
9	AMZN	1998-02-01	Amazon.com Inc.	NASDAQ	Catalog & Mail Order Houses	Services
10	APA	1998-02-01	Apache Corp.	NYSE	Independent Oil & Gas	Basic Materials
11	ATI	11/30/1999	Allegheny Technologies Inc.	NYSE	Metal Fabrication	Industrial Goods
12	AXP	1998-02-01	American Express Company	NYSE	Credit Services	Financial
13	BA	1998-02-01	The Boeing Company	NYSE	Aerospace/Defense Products & Services	Industrial Goods
14	BAC	1998-02-01	Bank of America Corporation	NYSE	Money Center Banks	Financial
15	BAX	1998-02-01	Baxter International Inc.	NYSE	Medical Instruments & Supplies	Healthcare
16	BIIB	1998-02-01	Biogen Idec Inc.	NASDAQ	Biotechnology	Healthcare
17	BK	1998-02-01	The Bank of New York Mellon Corporation	NYSE	Asset Management	Financial
18	BMJ	1998-02-01	Bristol-Myers Squibb Company	NYSE	Drug Manufacturers - Major	Healthcare
19	C	1998-02-01	Citigroup Inc.	NYSE	Money Center Banks	Financial

20	CAT	1998-02-01	Caterpillar Inc.	NYSE	Farm & Construction Machinery	Industrial Goods
21	CL	1998-02-01	Colgate-Palmolive Co.	NYSE	Personal Products	Consumer Goods
22	CMCSA	1998-02-01	Comcast Corporation	NASDAQ	Entertainment - Diversified	Services
23	COF	1998-02-01	Capital One Financial Corporation	NYSE	Credit Services	Financial
24	COP	2002-03-09	ConocoPhillips	NYSE	Independent Oil & Gas	Basic Materials
25	COST	1998-02-01	Costco Wholesale Corporation	NASDAQ	Discount, Variety Stores	Services
26	CPB	1998-02-01	Campbell Soup Company	NYSE	Processed & Packaged Goods	Consumer Goods
27	CSCO	1998-02-01	Cisco Systems, Inc.	NASDAQ	Networking & Communication Devices	Technology
28	CVS	1998-02-01	CVS Health Corporation	NYSE	Health Care Plans	Healthcare
29	CVX	2001-10-10	Chevron Corporation	NYSE	Major Integrated Oil & Gas	Basic Materials
30	DIS	1998-02-01	The Walt Disney Company	NYSE	Entertainment - Diversified	Services
31	EBAY	9/24/1998	eBay Inc.	NASDAQ	Specialty Retail, Other	Services
32	EMR	1998-02-01	Emerson Electric Co.	NYSE	Industrial Electrical Equipment	Industrial Goods
33	ETR	1998-02-01	Entergy Corporation	NYSE	Electric Utilities	Utilities
34	EXC	10/23/2000	Exelon Corporation	NYSE	Diversified Utilities	Utilities
35	FCX	1998-02-01	Freeport-McMoRan Inc.	NYSE	Copper	Basic Materials
36	FDX	1998-02-01	FedEx Corporation	NYSE	Air Delivery & Freight Services	Services
37	GD	1998-02-01	General Dynamics Corporation	NYSE	Aerospace/Defense Products & Services	Industrial Goods
38	GE	1998-02-01	General Electric Company	NYSE	Diversified Machinery	Industrial Goods
39	GILD	1998-02-01	Gilead Sciences Inc.	NASDAQ	Biotechnology	Healthcare

40	GS	1999-04-05	The Goldman Sachs Group, Inc.	NYSE	Investment Brokerage - National	Financial
41	HAL	1998-02-01	Halliburton Company	NYSE	Oil & Gas Equipment & Services	Basic Materials
42	HD	1998-02-01	The Home Depot, Inc.	NYSE	Home Improvement Stores	Services
43	HON	1998-02-01	Honeywell International Inc.	NYSE	Diversified Machinery	Industrial Goods
44	HPQ	1998-02-01	Hewlett-Packard Company	NYSE	Diversified Computer Systems	Technology
45	IBM	1998-02-01	International Business Machines Corporation	NYSE	Information Technology Services	Technology
46	INTC	1998-02-01	Intel Corporation	NASDAQ	Semiconductor - Broad Line	Technology
47	JNJ	1998-02-01	Johnson & Johnson	NYSE	Drug Manufacturers - Major	Healthcare
48	JPM	1998-02-01	JPMorgan Chase & Co.	NYSE	Money Center Banks	Financial
49	KO	1998-02-01	The Coca-Cola Company	NYSE	Beverages - Soft Drinks	Consumer Goods
50	LLY	1998-02-01	Eli Lilly and Company	NYSE	Drug Manufacturers - Major	Healthcare
51	LMT	1998-02-01	Lockheed Martin Corporation	NYSE	Aerospace/Defense Products & Services	Industrial Goods
52	LOW	1998-02-01	Lowe's Companies Inc.	NYSE	Home Improvement Stores	Services
53	MCD	1998-02-01	McDonald's Corp.	NYSE	Restaurants	Services
54	MDLZ	6/13/2001	Mondelez International, Inc.	NASDAQ	Confectioners	Consumer Goods
55	MDT	1998-02-01	Medtronic, Inc.	NYSE	Medical Appliances & Equipment	Healthcare
56	MET	2000-05-04	MetLife, Inc.	NYSE	Life Insurance	Financial
57	MMM	1998-02-01	3M Company	NYSE	Diversified Machinery	Industrial Goods
58	MO	1998-02-01	Altria Group Inc.	NYSE	Cigarettes	Consumer Goods
59	MRK	1998-02-01	Merck & Co. Inc.	NYSE	Drug Manufacturers - Major	Healthcare

60	MSFT	1998-02-01	Microsoft Corporation	NASDAQ	Business Software & Services	Technology
61	NKE	1998-02-01	Nike, Inc.	NYSE	Textile - Apparel Footwear & Accessories	Consumer Goods
62	NSC	1998-02-01	Norfolk Southern Corporation	NYSE	Railroads	Services
63	ORCL	1998-02-01	Oracle Corporation	NYSE	Application Software	Technology
64	OXY	1998-02-01	Occidental Petroleum Corporation	NYSE	Independent Oil & Gas	Basic Materials
65	PEP	1998-02-01	Pepsico, Inc.	NYSE	Beverages - Soft Drinks	Consumer Goods
66	PFE	1998-02-01	Pfizer Inc.	NYSE	Drug Manufacturers - Major	Healthcare
67	PG	1998-02-01	The Procter & Gamble Company	NYSE	Personal Products	Consumer Goods
68	QCOM	1998-02-01	QUALCOMM Incorporated	NASDAQ	Communication Equipment	Technology
69	RF	2002-10-05	Regions Financial Corporation	NYSE	Regional - Southeast Banks	Financial
70	S	1998-02-01	Sprint Corporation	NYSE	Wireless Communications	Technology
71	SBUX	1998-02-01	Starbucks Corporation	NASDAQ	Specialty Eateries	Services
72	SLB	1998-02-01	Schlumberger Limited	NYSE	Oil & Gas Equipment & Services	Basic Materials
73	SO	1998-02-01	Southern Company	NYSE	Electric Utilities	Utilities
74	SPG	1998-02-01	Simon Property Group Inc.	NYSE	REIT - Retail	Financial
75	T	1998-02-01	AT&T, Inc.	NYSE	Telecom Services - Domestic	Technology
76	TGT	1/31/2000	Target Corp.	NYSE	Discount, Variety Stores	Services
77	TXN	1998-02-01	Texas Instruments Inc.	NASDAQ	Semiconductor - Broad Line	Technology
78	UNH	1998-02-01	UnitedHealth Group Incorporated	NYSE	Health Care Plans	Healthcare
79	UNP	1998-02-01	Union Pacific Corporation	NYSE	Railroads	Services

80	UPS	1999-10-11	United Parcel Service, Inc.	NYSE	Air Delivery & Freight Services	Services
81	USB	1998-02-01	U.S. Bancorp	NYSE	Regional - Midwest Banks	Financial
82	VZ	2000-03-07	Verizon Communications Inc.	NYSE	Telecom Services - Domestic	Technology
83	WBA	1998-02-01				
84	WFC	1998-02-01	Wells Fargo & Company	NYSE	Money Center Banks	Financial
85	WMB	1998-02-01	Williams Companies, Inc.	NYSE	Oil & Gas Pipelines	Basic Materials
86	WMT	1998-02-01	Wal-Mart Stores Inc.	NYSE	Discount, Variety Stores	Services
87	WY	1998-02-01	Weyerhaeuser Co.	NYSE	Lumber, Wood Production	Industrial Goods
88	XOM	1999-01-12	Exxon Mobil Corporation	NYSE	Major Integrated Oil & Gas	Basic Materials
89	XRX	1998-02-01	Xerox Corporation	NYSE	Information Technology Services	Technology

Tabella A.1: Lista dei titoli utilizzati nello studio

Appendice B

Output di R

Si riporta un esempio di output relativo alla modellazione degli autovettori nella sezione 4.2

```
##  
## VAR Estimation Results:  
## =====  
## Endogenous variables: y1, y2, y3, y4, y5, y6, y7, y8, y9,  
## y10, y11, y12, y13, y14, y15, y16, y17, y18, y19, y20, y21,  
## y22, y23, y24, y25, y26, y27, y28, y29, y30, y31, y32, y33,  
## y34, y35, y36, y37, y38, y39, y40, y41, y42, y43, y44, y45,  
## y46, y47, y48, y49, y50, y51, y52, y53, y54, y55, y56, y57,  
## y58, y59, y60, y61, y62, y63, y64, y65, y66, y67, y68, y69,  
## y70, y71, y72, y73, y74, y75, y76, y77, y78, y79, y80, y81,  
## y82, y83, y84, y85, y86, y87, y88, y89  
## Deterministic variables: const  
## Sample size: 911  
## Log Likelihood: Inf  
## Roots of the characteristic polynomial:  
## 0.885 0.867 0.8349 0.7973 0.791 0.7506 0.7506 0.7269 0.7136
```

```

## 0.6644 0.6163 0.5981 0.5981 0.5725 0.5537 0.5537 0.5298 0.5298
## 0.4858 0.4715 0.4715 0.4556 0.4556 0.4389 0.4215 0.4215
## 0.4053 0.4003 0.4003 0.3715 0.3715 0.342 0.342 0.3334
## 0.3143 0.3143 0.3111 0.3111 0.2956 0.2956 0.2927 0.292
## 0.292 0.2895 0.2895 0.289 0.289 0.2765 0.2578 0.2578 0.2535
## 0.2535 0.244 0.244 0.2299 0.2299 0.2287 0.2287 0.2266 0.2264
## 0.2264 0.224 0.2074 0.2074 0.2001 0.1967 0.1967 0.1869
## 0.1869 0.1867 0.1867 0.1547 0.1547 0.1516 0.1516 0.1448
## 0.1341 0.1341 0.132 0.132 0.1196 0.1196 0.1018 0.1014 0.0693
## 0.0652 0.0652 0.01491 0.01491
## Call:
## vars::VAR(y = egien1_storico.matrix, p = 1, type = c("const"),
##          exogen = NULL, lag.max = NULL, ic = c("AIC", "HQ", "SC"))
##
##
## Estimation results for equation y1:
## =====
## y1 = y1.l1 + y2.l1 + y3.l1 + y4.l1 + y5.l1 + y6.l1 + y7.l1
## + y8.l1 + y9.l1 + y10.l1 + y11.l1 + y12.l1 + y13.l1 + y14.l1
## + y15.l1 + y16.l1 + y17.l1 + y18.l1 + y19.l1 + y20.l1
## + y21.l1 + y22.l1 + y23.l1 + y24.l1 + y25.l1 + y26.l1
## + y27.l1 + y28.l1 + y29.l1 + y30.l1 + y31.l1 + y32.l1
## + y33.l1 + y34.l1 + y35.l1 + y36.l1 + y37.l1 + y38.l1
## + y39.l1 + y40.l1 + y41.l1 + y42.l1 + y43.l1 + y44.l1
## + y45.l1 + y46.l1 + y47.l1 + y48.l1 + y49.l1 + y50.l1
## + y51.l1 + y52.l1 + y53.l1 + y54.l1 + y55.l1 + y56.l1
## + y57.l1 + y58.l1 + y59.l1 + y60.l1 + y61.l1 + y62.l1
## + y63.l1 + y64.l1 + y65.l1 + y66.l1 + y67.l1 + y68.l1
## + y69.l1 + y70.l1 + y71.l1 + y72.l1 + y73.l1 + y74.l1

```

```

## + y75.l1 + y76.l1 + y77.l1 + y78.l1 + y79.l1 + y80.l1
## + y81.l1 + y82.l1 + y83.l1 + y84.l1 + y85.l1 + y86.l1
## + y87.l1 + y88.l1 + y89.l1 + const
##
##          Estimate Std. Error t value Pr(>|t|)
## y1.l1    4.451e-01  7.131e-02   6.241 6.94e-10 ***
## y2.l1    1.646e-01  1.137e-01   1.447  0.14819
## y3.l1   -1.713e-02  7.944e-02  -0.216  0.82936
## y4.l1    1.724e-01  1.476e-01   1.168  0.24295
## y5.l1    6.191e-02  5.174e-02   1.197  0.23177
## y6.l1    1.370e-02  1.409e-02   0.972  0.33116
## y7.l1    1.932e-01  1.168e-01   1.654  0.09848 .
## y8.l1    5.223e-02  9.635e-02   0.542  0.58793
## y9.l1    6.353e-02  7.401e-02   0.858  0.39091
## y10.l1  -8.874e-02  5.911e-02  -1.501  0.13370
## y11.l1  -1.327e-02  2.431e-02  -0.546  0.58543
## y12.l1  -2.706e-01  1.100e-01  -2.460  0.01411 *
## y13.l1  -1.001e-01  9.615e-02  -1.041  0.29813
## y14.l1  -1.186e-01  8.598e-02  -1.379  0.16817
## y15.l1   1.926e-02  8.644e-02   0.223  0.82378
## y16.l1  -1.037e-01  5.442e-02  -1.905  0.05714 .
## y17.l1  -1.676e-02  8.401e-02  -0.200  0.84188
## y18.l1  -3.008e-02  1.024e-01  -0.294  0.76913
## y19.l1  -1.284e-01  6.757e-02  -1.900  0.05785 .
## y20.l1   3.034e-02  8.899e-02   0.341  0.73322
## y21.l1   2.267e-01  1.345e-01   1.685  0.09240 .
## y22.l1  -1.600e-01  7.415e-02  -2.157  0.03127 *
## y23.l1   9.075e-03  9.174e-02   0.099  0.92123
## y24.l1  -3.959e-02  1.013e-01  -0.391  0.69588

```

```
## y25.11 -1.382e-01  1.219e-01 -1.133  0.25738
## y26.11 -6.683e-02  6.917e-02 -0.966  0.33425
## y27.11  1.425e-01  1.133e-01  1.258  0.20867
## y28.11 -4.981e-02  6.879e-02 -0.724  0.46922
## y29.11 -7.100e-02  1.533e-01 -0.463  0.64327
## y30.11 -4.689e-02  1.177e-01 -0.398  0.69050
## y31.11  1.629e-02  8.260e-02  0.197  0.84374
## y32.11 -9.658e-02  1.058e-01 -0.912  0.36178
## y33.11 -1.289e-01  1.460e-01 -0.883  0.37739
## y34.11 -3.265e-01  1.125e-01 -2.903  0.00379 **
## y35.11  2.855e-02  3.602e-02  0.793  0.42830
## y36.11 -3.153e-02  1.038e-01 -0.304  0.76130
## y37.11  1.579e-01  1.217e-01  1.298  0.19481
## y38.11 -9.702e-02  7.433e-02 -1.305  0.19221
## y39.11 -3.972e-02  7.376e-02 -0.538  0.59039
## y40.11  2.795e-01  9.832e-02  2.843  0.00459 **
## y41.11  1.511e-01  6.834e-02  2.210  0.02736 *
## y42.11  5.672e-02  1.168e-01  0.486  0.62739
## y43.11 -3.354e-02  1.112e-01 -0.302  0.76299
## y44.11 -1.361e-02  7.016e-02 -0.194  0.84619
## y45.11  2.340e-01  1.620e-01  1.444  0.14901
## y46.11  2.333e-02  1.083e-01  0.215  0.82944
## y47.11 -2.831e-01  1.539e-01 -1.840  0.06614 .
## y48.11  5.846e-02  1.449e-01  0.404  0.68667
## y49.11 -5.389e-02  1.570e-01 -0.343  0.73154
## y50.11 -1.947e-01  1.275e-01 -1.527  0.12702
## y51.11  1.417e-01  1.351e-01  1.049  0.29461
## y52.11 -1.112e-01  1.035e-01 -1.073  0.28336
## y53.11  1.808e-01  1.457e-01  1.241  0.21513
```

```
## y54.11 -1.875e-01 9.450e-02 -1.984 0.04755 *
## y55.11 1.722e-02 1.155e-01 0.149 0.88148
## y56.11 -1.089e-02 7.977e-02 -0.137 0.89142
## y57.11 1.933e-01 1.404e-01 1.377 0.16899
## y58.11 1.448e-02 9.699e-02 0.149 0.88135
## y59.11 2.359e-01 1.049e-01 2.249 0.02476 *
## y60.11 -2.812e-02 1.196e-01 -0.235 0.81425
## y61.11 1.080e-01 1.023e-01 1.056 0.29141
## y62.11 -1.273e-01 7.752e-02 -1.643 0.10087
## y63.11 2.520e-02 1.002e-01 0.252 0.80141
## y64.11 5.116e-02 9.333e-02 0.548 0.58372
## y65.11 -2.527e-02 1.666e-01 -0.152 0.87949
## y66.11 2.250e-02 1.152e-01 0.195 0.84520
## y67.11 -1.978e-01 1.443e-01 -1.371 0.17076
## y68.11 -7.195e-02 7.960e-02 -0.904 0.36632
## y69.11 -1.204e-02 5.641e-02 -0.213 0.83107
## y70.11 1.122e-02 2.032e-02 0.552 0.58111
## y71.11 6.526e-02 1.026e-01 0.636 0.52477
## y72.11 -2.291e-01 9.368e-02 -2.446 0.01467 *
## y73.11 2.874e-01 1.559e-01 1.843 0.06567 .
## y74.11 -2.186e-02 6.078e-02 -0.360 0.71923
## y75.11 1.712e-02 1.382e-01 0.124 0.90145
## y76.11 6.316e-02 8.473e-02 0.745 0.45621
## y77.11 1.209e-03 8.577e-02 0.014 0.98876
## y78.11 8.811e-02 6.988e-02 1.261 0.20772
## y79.11 1.532e-01 8.828e-02 1.736 0.08303 .
## y80.11 -1.071e-01 1.396e-01 -0.767 0.44324
## y81.11 -1.560e-01 1.087e-01 -1.436 0.15152
## y82.11 2.760e-01 1.391e-01 1.984 0.04762 *
```

```
## y83.l1  1.545e-03  8.651e-02  0.018  0.98576
## y84.l1  2.882e-01  1.223e-01  2.357  0.01866 *
## y85.l1  1.107e-02  4.578e-02  0.242  0.80904
## y86.l1 -1.606e-01  1.502e-01 -1.069  0.28522
## y87.l1 -1.718e-01  7.976e-02 -2.154  0.03156 *
## y88.l1  1.586e-01  1.649e-01  0.962  0.33621
## y89.l1 -1.247e-01  5.276e-02 -2.363  0.01836 *
## const  -1.041e-03  7.941e-05 -13.114  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.0008855 on 821 degrees of freedom
## Multiple R-Squared: 0.3788, Adjusted R-squared: 0.3115
## F-statistic: 5.626 on 89 and 821 DF, p-value: < 2.2e-16
```

Bibliografia

- [1] Acharya, Viral V. et al. «Measuring Systemic Risk». In: *The Review of Financial Studies* 30.1 (ott. 2016), pp. 2–47.
- [2] Allen, Linda, Bali, Turan G e Tang, Yi. «Does systemic risk in the financial sector predict future economic downturns?» In: *The Review of Financial Studies* 25.10 (2012), pp. 3000–3036.
- [3] Anderson, T. W. «Asymptotic Theory for Principal Component Analysis». In: *The Annals of Mathematical Statistics* 34.1 (1963), pp. 122 –148.
- [4] Anderson, Theodore Wilbur. *An introduction to multivariate statistical analysis*. A cura di statistics., Wiley publications in. John Wiley Sons Inc, 1958.
- [5] Anderson, Theodore Wilbur. *An introduction to multivariate statistical analysis*. Rapp. tecn. Wiley New York, 1962.
- [6] Aït-Sahalia, Yacine e Jacod, Jean. *High-Frequency Financial Econometrics*. A cura di Princeton, NJ. Princeton University Press, 2014.
- [7] Aït-Sahalia, Yacine e Xiu, Dacheng. «Principal Component Analysis of High-Frequency Data». In: *Journal of the American Statistical Association* 114.525 (2019), pp. 287–303.

- [8] Aït-Sahalia, Yacine e Xiu, Dacheng. «Using principal component analysis to estimate a high dimensional factor model with high-frequency data». In: *Journal of Econometrics* 201.2 (2017), pp. 384–399.
- [9] Bai, Jushan. «Inferential Theory for Factor Models of Large Dimensions». In: *Econometrica* 71.1 (2003), pp. 135–171.
- [10] Benoit, Sylvain et al. «Where the Risks Lie: A Survey on Systemic Risk». In: *Review of Finance* 21.1 (giu. 2016), pp. 109–152.
- [11] Benoit, Sylvain et al. «Where the Risks Lie: A Survey on Systemic Risk». In: *Review of Finance* 21.1 (2017), pp. 109–152.
- [12] Billio, Monica et al. «Econometric measures of connectedness and systemic risk in the finance and insurance sectors». In: *Journal of Financial Economics* 104.3 (2012), pp. 535–559.
- [13] Bisias, Dimitrios et al. «A Survey of Systemic Risk Analytics». In: *Annual Review of Financial Economics* 4.1 (2012), pp. 255–296.
- [14] Black, F. «Studies of Stock Price Volatility Changes». In: *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association* (1976), pp. 171–181.
- [15] Bollerslev, Tim, Engle, Robert F. e Nelson, Daniel B. «Chapter 49 Arch models». In: vol. 4. *Handbook of Econometrics*. Elsevier, 1994, pp. 2959–3038.
- [16] Bollerslev, Tim e Wooldridge, Jeffrey M. «Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances». In: *Econometric Reviews* 11.2 (1992), pp. 143–172.
- [17] Borio, Claudio. «The Great Financial Crisis: setting priorities for new statistics». In: *Journal of Banking Regulation* 14.3 (2013), pp. 306–317.

- [18] Borio, Claudio, Lowe, Philip et al. «Assessing the risk of banking crises». In: *BIS Quarterly Review* 7.1 (2002), pp. 43–54.
- [19] Borio, Claudio EV e Drehmann, Mathias. «Assessing the risk of banking crises—revisited». In: *BIS Quarterly Review, March* (2009).
- [20] Borio, Claudio EV e Lowe, Philip William. «Asset prices, financial and monetary stability: exploring the nexus». In: (2002).
- [21] Brownlees, Christian e Engle, Robert F. «SRISK: A conditional capital shortfall measure of systemic risk». In: *The Review of Financial Studies* 30.1 (2017), pp. 48–79.
- [22] Brunnermeier, Markus Konrad. *Asset pricing under asymmetric information: Bubbles, crashes, technical analysis, and herding*. Oxford University Press on Demand, 2001.
- [23] Caporin, Massimiliano et al. «Systemic risk and severe economic downturns: A targeted and sparse analysis». In: *Journal of Banking & Finance* 134 (2022), p. 106339.
- [24] Delbaen, Freddy e Schachermayer, Walter. «A General Version of the Fundamental Theorem of Asset Pricing (1994)». In: *Mathematische Annalen* 300 (set. 1994), pp. 463–520.
- [25] Diebold, Francis X. e Yilmaz, Kamil. «Better to give than to receive: Predictive directional measurement of volatility spillovers». In: *International Journal of Forecasting* 28.1 (2012). Special Section 1: The Predictability of Financial Markets Special Section 2: Credit Risk Modelling and Forecasting, pp. 57–66.
- [26] Egloff, Daniel, Leippold, Markus e Wu, Liuren. «The term structure of variance swap rates and optimal variance swap investments». In: *Journal of Financial and Quantitative Analysis* 45.5 (2010), pp. 1279–1310.

- [27] Engle, Robert. «New frontiers for arch models». In: *Journal of Applied Econometrics* 17.5 (2002), pp. 425–446.
- [28] Engle, Robert F. e Russell, Jeffrey R. «Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data». In: *Econometrica* 66.5 (1998), pp. 1127–1162.
- [29] Friedland, Shmuel. «Convex spectral functions». In: *Linear and Multilinear Algebra* 9.4 (1981), pp. 299–316.
- [30] Giglio, Stefano, Kelly, Bryan e Pruitt, Seth. «Systemic risk and the macroeconomy: An empirical evaluation». In: *Journal of Financial Economics* 119.3 (2016), pp. 457–471.
- [31] Huang, Xin, Zhou, Hao e Zhu, Haibin. «A framework for assessing the systemic risk of major financial institutions». In: *Journal of Banking Finance* 33.11 (2009). Financial Globalisation, Risk Analysis and Risk Management, pp. 2036–2049.
- [32] Jacod, J. e Protter, P. *Discretization of Processes*. A cura di Springer. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2012.
- [33] Jiménez-Martin, Juan-Angel, Caporin, Massimiliano e Garcia-Jorcano, Laura. «Monitoring Financial Stress Spillovers with High-Frequency Principal Components». In: *Available at SSRN 4119746* (2022).
- [34] Kiyotaki, Nobuhiro e Moore, John. «Credit cycles». In: *Journal of political economy* 105.2 (1997), pp. 211–248.
- [35] Kliesen, Kevin L. e Smith, Douglas C. «Measuring financial market stress». In: *Economic Synopses* (2010).

- [36] Koop, Gary, Pesaran, M.Hashem e Potter, Simon M. «Impulse response analysis in nonlinear multivariate models». In: *Journal of Econometrics* 74.1 (1996), pp. 119–147.
- [37] Laborda, Ricardo e Olmo, Jose. «Volatility spillover between economic sectors in financial crisis prediction: Evidence spanning the great financial crisis and Covid-19 pandemic». In: *Research in International Business and Finance* 57 (2021), p. 101402.
- [38] Lee, Sang-Won e Hansen, Bruce E. «Asymptotic Theory for the Garch(1,1) Quasi-Maximum Likelihood Estimator». In: *Econometric Theory* 10.1 (1994), 29–52.
- [39] Litterman, Robert e Scheinkman, Jose. «Common factors affecting bond returns». In: *Journal of fixed income* 1.1 (1991), pp. 54–61.
- [40] Magnus, Jan R. e Neudecker, Heinz. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Second. John Wiley, 1999.
- [41] Nguyen, Hieu, Tran, Phuong e Nguyen, Quang. «An Analysis of Eigenvectors of a Stock Market Cross-Correlation Matrix». In: *Studies in Computational Intelligence* (gen. 2018), pp. 504–513.
- [42] Nucera, Federico et al. «The information in systemic risk rankings». In: *Journal of Empirical Finance* 38 (2016), pp. 461–475.
- [43] Okamoto, Masashi. «Distinctness of the Eigenvalues of a Quadratic form in a Multivariate Sample». In: *The Annals of Statistics 1973-jul vol. 1 iss. 4 1* (4 1973).
- [44] Pesaran, H.Hashem e Shin, Yongcheol. «Generalized impulse response analysis in linear multivariate models». In: *Economics Letters* 58.1 (1998), pp. 17–29.

- [45] Rodríguez-Moreno, María e Peña, Juan Ignacio. «Systemic risk measures: The simpler the better?» In: *Journal of Banking & Finance* 37.6 (2013), pp. 1817–1831.
- [46] Segoviano Basurto, Miguel e Goodhart, Charles A.E. «Banking Stability Measures». In: *IMF Working Paper No. 09/4* (gen. 2009).
- [47] Silva, Walmir, Kimura, Herbert e Sobreiro, Vinicius Amorim. «An analysis of the literature on systemic financial risk: A survey». In: *Journal of Financial Stability* 28 (2017), pp. 91–114.
- [48] Tao, Terence. *Topics in Random Matrix Theory*. 2012.
- [49] Zhou, Chen. «Why the micro-prudential regulation fails?» In: *The impact on systemic risk by imposing a capital requirement. Netherlands Central Bank, Research Department* (2010).