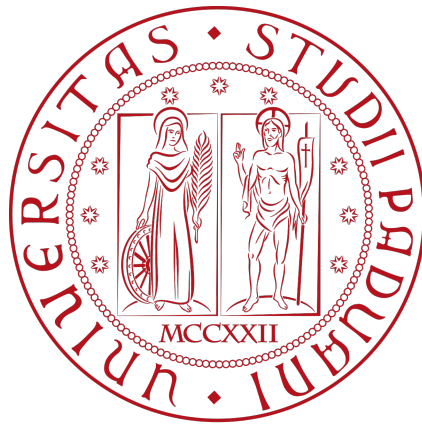


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA MAGISTRALE IN
SCIENZE STATISTICHE



Volatilità Realizzata Parziale e Previsioni Gerarchiche:
un'applicazione sul titolo *Cisco Systems*

Relatore: PROF. MASSIMILIANO CAPORIN

Correlatore: PROF. TOMMASO DI FONZO

Laureando: TOMMASO BONGIOVANNI

Matricola N. 2027570

Anno Accademico 2022/2023

Abstract

Questo studio ha l'obiettivo di applicare metodi di previsione gerarchica alla Volatilità Realizzata, cercando di ottenere risultati migliori rispetto alle previsioni ricavate mediante la modellazione HAR-RV (*Heterogeneous Auto Regression*), prese come riferimento. Le strutture gerarchiche proposte per le previsioni derivano da possibili decomposizioni della Volatilità Realizzata. In particolare, si è interessati a commentare i risultati ottenuti partendo da una specificazione alternativa della Volatilità Realizzata, chiamata *Partition-based RV* o Volatilità Realizzata Parziale, proposta per la prima volta da Caporin (2020) [15]. Nella prima parte dell'elaborato si trattano a livello teorico i due macro-argomenti affrontati: il concetto di Volatilità Realizzata e le sue decomposizioni, le previsioni gerarchiche. Successivamente, le tecniche di previsione gerarchica vengono applicate per predire la Volatilità Realizzata di un particolare titolo finanziario. Dal punto di vista predittivo, i risultati riportati evidenziano un miglioramento ottenuto grazie all'utilizzo di strutture gerarchiche, in particolare quella basata sulla Volatilità Realizzata Parziale.

Indice

Abstract	i
Introduzione	1
1 Volatilità Realizzata	3
1.1 Dalla Volatilità Integrata alla Volatilità Realizzata	4
1.2 Varianza giornaliera e salti nei prezzi	7
1.2.1 Test BNS corretto per la <i>staleness</i>	8
1.2.2 Procedura iterativa di Andersen	12
1.2.3 Periodicità della volatilità e salti nei prezzi	13
1.3 Decomposizione della Volatilità Realizzata	16
1.3.1 Componente continua, discontinua, <i>Good e Bad Volatility</i>	16
1.3.2 Decomposizione basata sulle partizioni: Volatilità Realizzata Parziale	17
2 Previsioni gerarchiche per serie storiche	21
2.1 Serie storiche gerarchiche e raggruppate	22
2.2 Approcci a livello singolo	24
2.2.1 Approccio <i>bottom-up</i>	25
2.2.2 Approccio <i>top-down</i>	25
2.2.3 Approccio <i>middle-out</i>	27
2.2.4 Aspetti positivi e negativi degli approcci	28
2.3 Notazione matriciale	29
2.4 Riconciliazione delle previsioni	31
2.4.1 Correttezza delle previsioni coerenti	31
2.4.2 Riconciliazione GLS	32
2.4.3 Riconciliazione <i>Minimum Trace</i> (MinT)	33
2.4.3.1 Stimatori alternativi di W_h	34
3 Applicazione: previsione della Volatilità Realizzata del titolo Cisco Systems	37
3.1 Contesto e obiettivo dell'analisi	37
3.2 Dati utilizzati: il titolo <i>Cisco Systems</i>	38
3.3 Esercizio di previsione e criteri di raffronto	38
3.4 Analisi descrittive della Volatilità Realizzata	39

3.4.1	Modellazione e previsione della Volatilità Realizzata	42
3.5	Previsioni gerarchiche	43
3.5.1	Gerarchia <i>Good-Bad Volatility</i>	43
3.5.2	Gerarchie temporali e quantiliche	45
3.5.2.1	Decomposizione temporale della Volatilità Realizzata . .	45
3.5.2.2	Decomposizione in partizioni quantiliche della Volatilità Realizzata	47
3.5.3	Decomposizione temporale e quantilica	49
3.5.3.1	Strutture temporali e quantiliche	50
4	Conclusioni	53
4.1	Risultati ottenuti	53
4.2	Considerazioni finali	58
4.3	Possibili miglioramenti	59
	Bibliografia	61

Introduzione

La stima e la previsione della volatilità di un titolo finanziario sono fondamentali per ricavare informazioni utili sulla sua natura e sulla sua rischiosità.

La volatilità, che indica quanto i prezzi di uno strumento finanziario variano nel tempo, difatti, viene utilizzata in numerose applicazioni finanziarie come l'*asset pricing*, il *risk management* e l'allocazione del portafoglio.

Negli anni sono stati compiuti numerosi studi per prevedere in maniera affidabile tale quantità che, a differenza dei rendimenti grezzi, non è direttamente osservabile.

Un approccio possibile per trattare la latenza della volatilità dei rendimenti è quello di utilizzare modelli parametrici come il (*Generalized*) *Autoregressive Conditional Heteroskedasticity* ((G)ARCH) introdotto da Engle (1982) [19] e Bollerslev (1986) [13] e i modelli di volatilità stocastica presentati da Taylor (1986) [37].

La volatilità può essere anche prevista attraverso l'utilizzo di un modello di *pricing* delle opzioni. Questo metodo permette di invertire i prezzi osservati dei derivati in previsioni basate sul mercato della "volatilità implicita" su un orizzonte futuro fisso.

Un'altra procedura, che non fa uso di imposizioni parametriche o ipotesi di mercato, ma tiene solo conto della volatilità passata del titolo, consiste nel valutare la Volatilità Realizzata. La Volatilità Realizzata è una stima della volatilità effettiva dell'attività finanziaria e si basa sulla misurazione della variazione dei rendimenti storici del titolo, che, grazie alla crescente disponibilità di dati ad alta frequenza, può essere calcolata in maniera accurata.

Uno dei modelli più conosciuti per la stima e la previsione di quest'ultima quantità di interesse è sicuramente il modello *Heterogeneous Auto Regression* (HAR-RV), proposto da Corsi nel 2009 [16]. Altri modelli predittivi per la Volatilità Realizzata sono stati realizzati a partire da differenti decomposizioni della stessa, come la decomposizione in componenti di salto e continua, presentata da Andersen, Bollerslev e Diebold nel 2007 [2], o quella in componenti *Good Volatility* e *Bad Volatility*, introdotta da Patton e Sheppard nel 2015 [34].

Un nuovo metodo per la previsione della Volatilità Realizzata di un titolo è l'applicazione di tecniche di previsioni gerarchiche per serie storiche. Questo approccio viene adoperato in presenza di una struttura gerarchica composta da molteplici serie temporali. Per quanto riguarda la Volatilità Realizzata, tali strutture possono essere imposte a partire dalle sue differenti decomposizioni. Questa metodologia sfrutta, quindi, le relazioni e i vincoli di aggregazione presenti tra le diverse serie storiche che compongono la struttura per prevedere in maniera coerente ed accurata la quantità di interesse.

L'obiettivo di questo elaborato è quello di applicare quest'ultima metodologia per la previsione della Volatilità Realizzata di un titolo, partendo da una decomposizione innovativa della Volatilità Realizzata denominata "*Partial Realized Volatility*", proposta da Caporin nel 2020 [15]. Impiegando questa metodologia, si vogliono ottenere risultati predittivi migliori rispetto al modello HAR-RV di Corsi.

Nel *Capitolo 1* vengono presentati il concetto di Volatilità Realizzata e le sue possibili decomposizioni, per trattare poi, nel capitolo successivo, le basi teoriche inerenti alle previsioni gerarchiche di serie storiche; nel *Capitolo 3*, invece, è proposta un'applicazione specifica della metodologia gerarchica per la previsione della Volatilità Realizzata del titolo *Cisco Systems*. Infine, nel *Capitolo 4* sono riportate le conclusioni dello studio compiuto.

Capitolo 1

Volatilità Realizzata

A partire dal nuovo millennio, sono stati effettuati numerosi studi inerenti alla stima e alla modellazione della varianza giornaliera, sfruttando i contemporanei sviluppi tecnologici e la disponibilità di dati ad alta frequenza.

Gli economisti Andersen, Bollerslev, Diebold e Labys (2001) [3] e Barndorff-Nielsen e Shephard (2002) [9] hanno introdotto il concetto di Volatilità Realizzata, una stima non parametrica della varianza giornaliera calcolata come somma dei rendimenti intra-giornalieri al quadrato.

Da lì in poi molti economisti hanno dato contributi importanti, soprattutto riguardo alla stima della Volatilità Realizzata in presenza di salti di prezzo e alla modellazione delle serie di tale quantità di interesse.

In questo capitolo, nella sezione 1.1 vengono presentati i concetti di Varianza Integrata e Variazione Quadratica, fondamentali per comprendere come si ottiene lo stimatore di Volatilità Realizzata; poi, nella sezione 1.2 è illustrata una procedura per l'individuazione dei salti di prezzo; infine, nella sezione 1.3 sono presenti delle possibili decomposizioni della Volatilità Realizzata, tra cui l'innovativa partizione in componenti parziali introdotta da Caporin nel 2020 [15].

I grafici inseriti in questo capitolo fanno riferimento ai rendimenti intra-giornalieri del titolo *American Express* (AXP) nel periodo che va dal 04/01/2021 al 31/12/2021.

1.1 Dalla Volatilità Integrata alla Volatilità Realizzata

La teoria della Volatilità Realizzata [6] è legata alla disponibilità di prezzi e, di conseguenza, di rendimenti di un determinato *asset* con frequenze elevate all'interno del giorno.

È naturale che in primo luogo venga considerata l'evoluzione dei prezzi in un contesto di tempo continuo, così da ottenere un corretto stimatore della volatilità giornaliera, anche se successivamente i dati relativi a prezzi e rendimenti saranno disponibili solo a intervalli discreti.

Il prezzo di uno strumento finanziario segue un moto browniano geometrico esplicitato dalla seguente equazione differenziale stocastica:

$$dp_t = \mu_t dt + \sigma_t dW_t, \quad (1.1)$$

dove p_t è il logaritmo del prezzo del titolo, μ_t è la media (*drift*) del processo, σ_t è la volatilità istantanea e W_t è un moto browniano standard.

L'obiettivo ultimo è quello di ricavare una misura di variazione del rendimento sull'intervallo di tempo giornaliero, che è possibile ottenere riscrivendo il processo (1.1) nel seguente modo:

$$\begin{aligned} p_t &= p_{t-1} + \int_{t-1}^t \mu_s ds + \int_{t-1}^t \sigma_s dW_s, \\ r_t &= p_t - p_{t-1} = \int_{t-1}^t \mu_s ds + \int_{t-1}^t \sigma_s dW_s. \end{aligned} \quad (1.2)$$

Isolando la componente di volatilità, si ottiene la **Varianza Integrata**:

$$IV(t-1, t) = \int_{t-1}^t \sigma_s^2 ds. \quad (1.3)$$

Tuttavia, non avendo a disposizione il prezzo ad ogni istante di tempo, si può pensare di suddividere l'intervallo giornaliero $(t-1, t)$ in m sottointervalli.

Si definisce il rendimento intra-giornaliero come la variazione di prezzo nell'*i-esimo*

periodo all'interno del giorno t :

$$r_{i,t} = p_{i,t} - p_{i-1,t}$$

con $i = 1, \dots, m$.

Barndorff-Nielsen e Shephard (2002) [9] hanno sfruttato tali rendimenti intra-giornalieri per stimare la Varianza Integrata definendo il concetto di **Variatione Quadratica** che, in assenza di salti nei prezzi, coincide con la Volatilità Integrata:

$$QV(t-1, t) = \lim_{m \rightarrow +\infty} \sum_{i=1}^m r_{i,t}^2 = IV(t-1, t). \quad (1.4)$$

Questa quantità può essere approssimata considerando un numero finito di intervalli intra-giornalieri. Si ottiene, quindi, la **Volatilità Realizzata** o **Varianza Realizzata** (termine più appropriato, ma meno utilizzato) che corrisponde alla somma dei rendimenti intra-giornalieri al quadrato:

$$RV_t = \sum_{i=1}^m r_{i,t}^2 \quad (1.5)$$

con

$$RV_t \xrightarrow{m \rightarrow \infty} QV(t-1, t). \quad (1.6)$$

Barndorff-Nielsen e Shephard (2002) [9] hanno ricavato anche la distribuzione asintotica della Volatilità Realizzata:

$$\frac{\sqrt{m} (RV_t^m - QV(t-1, t))}{\sqrt{2IQ(t-1, t)}} \sim N(0, 1), \quad (1.7)$$

dove RV_t^m è la Volatilità Realizzata calcolata con m sottointervalli, $QV(t-1, t)$ è la variazione quadratica giornaliera, IQ è la *Integrated Quarticity* che viene stimata ricorrendo alla *Realized Quarticity*:

$$RQ_t^m = \frac{m}{3} \sum_{i=1}^m r_{i,t}^4.$$

Una volta definiti lo stimatore (1.5) e la sua distribuzione asintotica (1.7), è necessario determinare il numero di intervalli m con cui sottocampionare la giornata borsistica.

Infatti, come si nota dal grafico in *Figura 1.1*, a seconda della frequenza di campionamento giornaliera si ottengono serie di Volatilità Realizzata differenti.

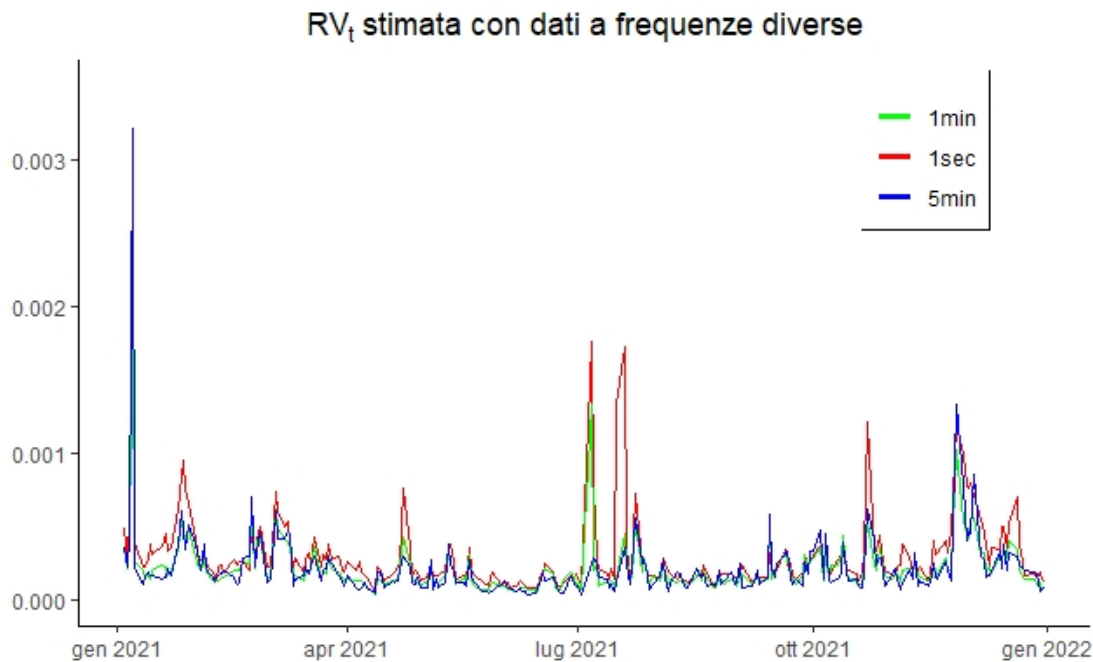


FIGURA 1.1: Sequenze di Volatilità Realizzata del titolo AXP a frequenze diverse

Dalla convergenza (1.6) si può notare che, all'aumentare del numero m di sotto-intervalli giornalieri, la Volatilità Realizzata approssima in maniera migliore la Variazione Quadratica. Tuttavia, esistono alcuni motivi per cui non è conveniente utilizzare una frequenza di campionamento eccessivamente alta.

In primis, è necessario tenere in considerazione che per titoli poco liquidi non si può utilizzare un campionamento elevato poiché introduce un errore di discretizzazione nella misura della Volatilità Realizzata.

In aggiunta, aumentando il numero di intervalli m , i rendimenti osservati possono essere contaminati dal rumore dovuto alla micro-struttura del mercato che induce un'autocorrelazione spuria nelle serie storiche di rendimenti ad altissima frequenza. Sono diversi i fattori che possono causare tale correlazione spuria; alcuni esempi sono i rimbalzi tra domanda e offerta o il cosiddetto “*trading* asincrono” dovuto ai differenti tempi di reazione di titoli con diversa liquidità agli shock del mercato.

Possibili soluzioni adottabili in presenza di tale distorsione sono l'utilizzo di frequenze relativamente basse (1 o 5 minuti) quando si lavora con titoli abbastanza liquidi o l'uso di stimatori robusti in presenza di errori di microstruttura, come il *Two-time scales*

volatility proposto da Zhang, Mykland e Aït-Sahalia (2005) [40].

1.2 Varianza giornaliera e salti nei prezzi

Talvolta l'evoluzione dei prezzi può presentare delle discontinuità o salti che alterano il loro processo continuo. In questo caso l'equazione (1.2) diventa:

$$r_t = p_t - p_{t-1} = \int_{t-1}^t \mu_s ds + \int_{t-1}^t \sigma_s dW_s + \sum_{i=1}^{N_t} J_i^2, \quad (1.8)$$

dove N_t è un processo di conteggio ad attività finita (Poisson) e J_i è un salto.

Tale situazione fa sì che la Variazione Quadratica non coincida più con la Volatilità Integrata come in (1.4), ma si ha

$$QV(t, t-1) = \int_{t-1}^t \sigma_s^2 dW_s + \sum_{i=1}^{N_t} J_i^2 = IV(t, t-1) + \sum_{i=1}^{N_t} J_i^2 \quad (1.9)$$

e di conseguenza:

$$RV(t-1, t) \longrightarrow \int_{t-1}^t \sigma_s^2 dW_s + \sum_{i=1}^{N_t} J_i^2.$$

Per modellare e prevedere la Volatilità Realizzata, quindi, diventa necessario stimare le sue componenti continua e di salto.

Nel corso degli anni sono stati presentati diversi test per l'identificazione di salti e stimatori robusti delle componenti continua e discontinua della Volatilità Realizzata.

I primi a trattare l'argomento sono stati Barndorff-Nielsen e Shephard (2004, 2006) [10] [11] che hanno proposto un test basato sul confronto tra la Volatilità Realizzata e la *Bipower Variation*, stimatore robusto della sola componente continua.

Approcci alternativi sono stati quelli di Aït-Sahalia e Jacod (2009) [1] e Corsi et al. (2010) [17], che hanno introdotto metodologie di troncamento nella costruzione delle statistiche per la rilevazione dei salti, o anche Lee e Mykland (2008) [30], che, invece, hanno utilizzato test di tipo non parametrico.

Un contributo notevole è stato apportato anche da Andersen, Bollerslev e Dobrev (2007, 2010) [4] [5] con l'introduzione di un nuovo test di identificazione di salti a livello intragiornaliero. Questi, in aggiunta, hanno proposto una procedura iterativa per l'individuazione dei salti intra-giornalieri molto utilizzata anche con altri test come quello BNS (Barndorff-Nielsen e Shephard, 2006 [11]).

Certamente meritano una menzione anche i lavori di Boudt, Croux e Laurent (2011) [14] che hanno mostrato l'esistenza di una periodicità intra-giornaliera della volatilità che influenza la prestazione dei test sui salti.

Di seguito vengono presentati il test BNS che tiene conto della *staleness*, la procedura iterativa di Andersen per l'individuazione di salti a livello intra-giornaliero e, infine, il filtraggio dei rendimenti per la periodicità della volatilità introdotta da Boudt et al. [14].

1.2.1 Test BNS corretto per la *staleness*

Sia la **Bipower Variation** (Barndorff-Nielsen and Shephard, 2004 [10]) lo stimatore di riferimento per la sola componente continua della Volatilità Realizzata:

$$BPV_t = \mu^{-1} \frac{m}{m-1} \sum_{i=2}^m |r_{i,t}| |r_{i-1,t}| \quad (1.10)$$

con $\mu = \sqrt{\frac{2}{\pi}}$ e m numero di intervalli in cui è stato suddiviso il giorno.

In presenza di salti si ha quindi:

$$\begin{aligned} RV(t-1, t) &\longrightarrow \int_{t-1}^t \sigma_s^2 dW_s + \sum_{i=1}^{N_t} J_i^2, \\ BPV(t-1, t) &\longrightarrow \int_{t-1}^t \sigma_s^2 dW_s, \\ RV(t-1, t) - BPV(t-1, t) &\approx \sum_{i=1}^{N_t} J_i^2. \end{aligned} \quad (1.11)$$

Nella *Figura 1.2* sono presentate le serie di Volatilità Realizzata e *Bipower Variation* calcolate con frequenza di 5 minuti ($m=78$), mentre nella *Figura 1.3* è illustrata la loro differenza.

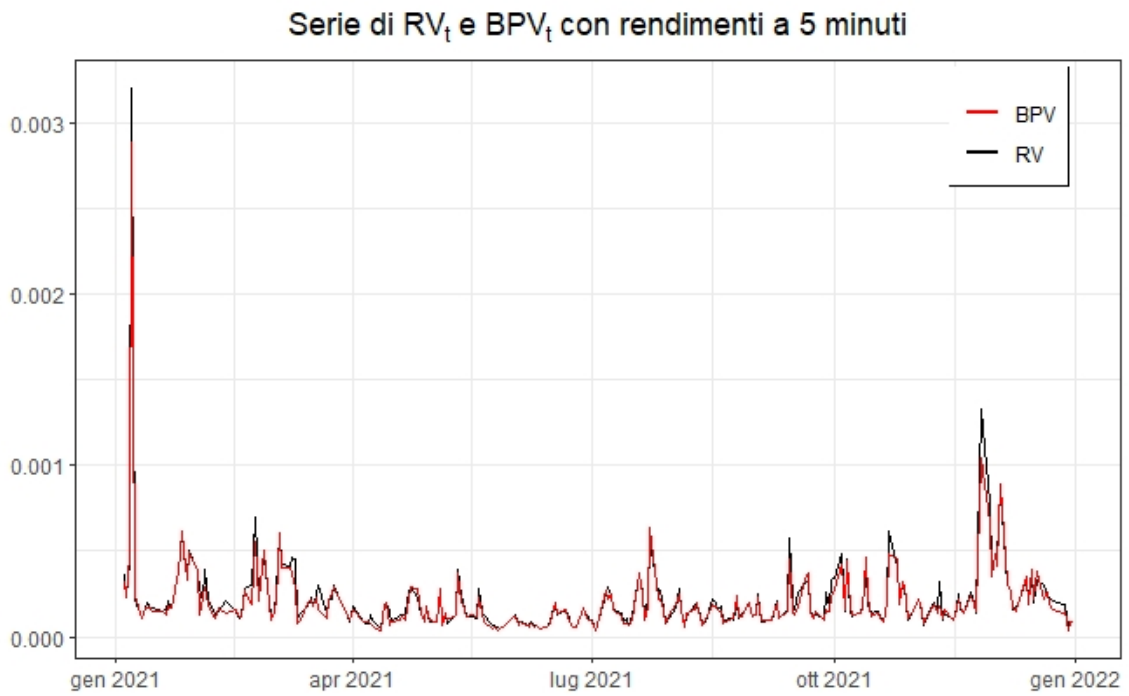


FIGURA 1.2: Sequenze di Volatilità Realizzata e *Bipower Variation* con frequenza di 5 minuti del titolo AXP

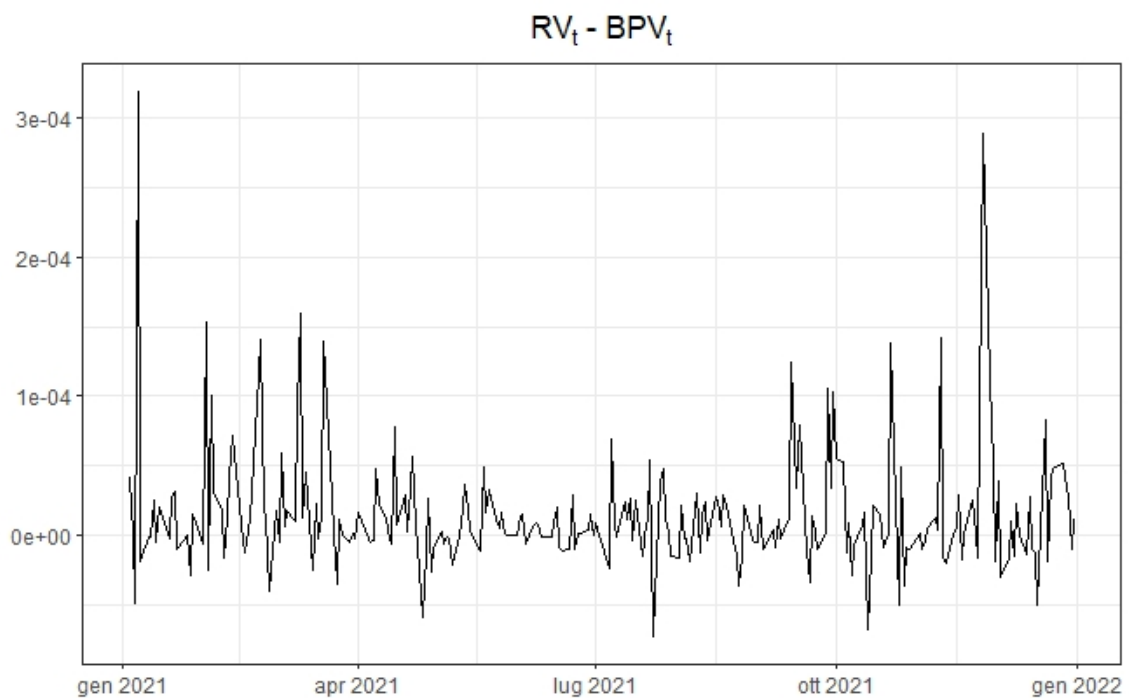


FIGURA 1.3: Differenza tra sequenze di Volatilità Realizzata e *Bipower Variation* con frequenza di 5 minuti del titolo AXP

Proprio dalla differenza tra BPV_t e RV_t si ottiene la componente di salto; il test BNS (Barndorff-Nielsen and Shephard, 2006 [11]) sfrutta questa intuizione per ricavare la statistica test che permette di individuare la presenza di salti nella giornata borsistica t -esima:

$$BNS_t = \frac{(RV_t - BPV_t)RV_t^{-1}}{\sqrt{\theta \max\left(1, \frac{TPV_t}{BPV_t^2}\right)}} \rightarrow N(0, 1) \quad (1.12)$$

con $\theta = \frac{1}{m} \left(\frac{\pi^2}{4} + \pi - 5 \right)$, $TPV_t = \mu_{4/3}^{-3} \frac{m^2}{m-2} \sum_{i=2}^m |r_{i,t}|^{\frac{4}{3}} |r_{i-1,t}|^{\frac{4}{3}} |r_{i-2,t}|^{\frac{4}{3}}$ stimatore robusto della *Integrated Quarticity*, $\mu_a = E[|Z|^a]$ con $Z \sim N(0, 1)$ e $a > 0$.

Tale test si distribuisce asintoticamente come una Normale standardizzata per $m \rightarrow \infty$. Quando $BNS_t > \Phi(1 - \alpha)^{-1}$ con α livello di significatività, allora nel giorno t -esimo è presente un salto.

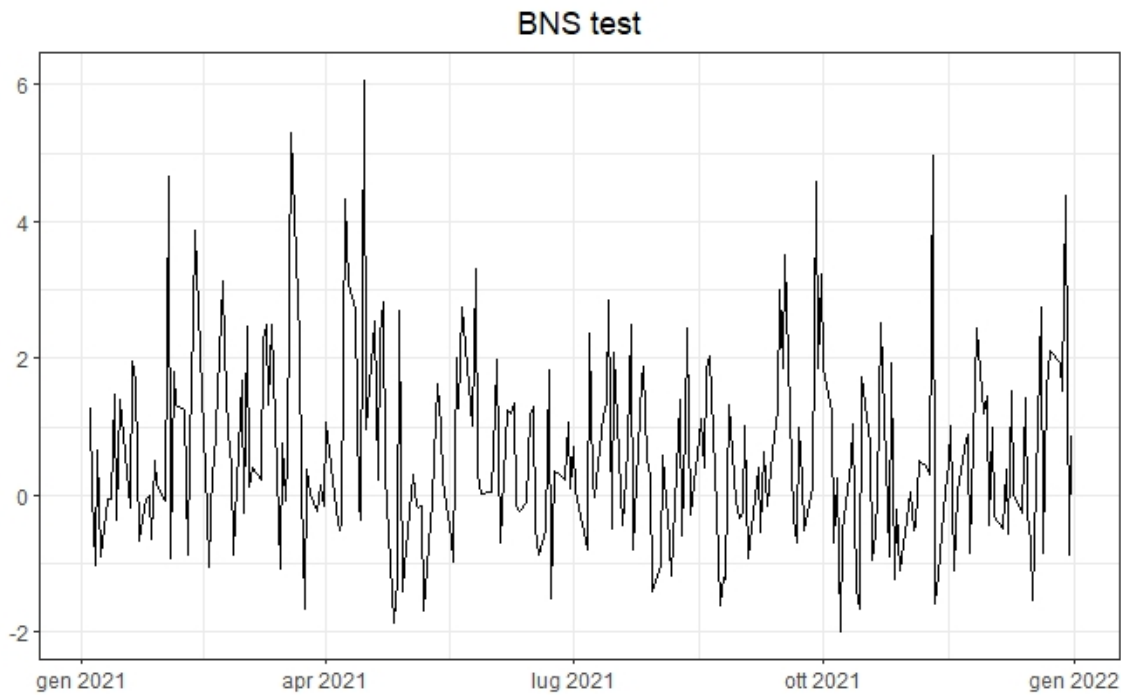


FIGURA 1.4: Valori del test BNS calcolato con rendimenti a 5 minuti del titolo AXP

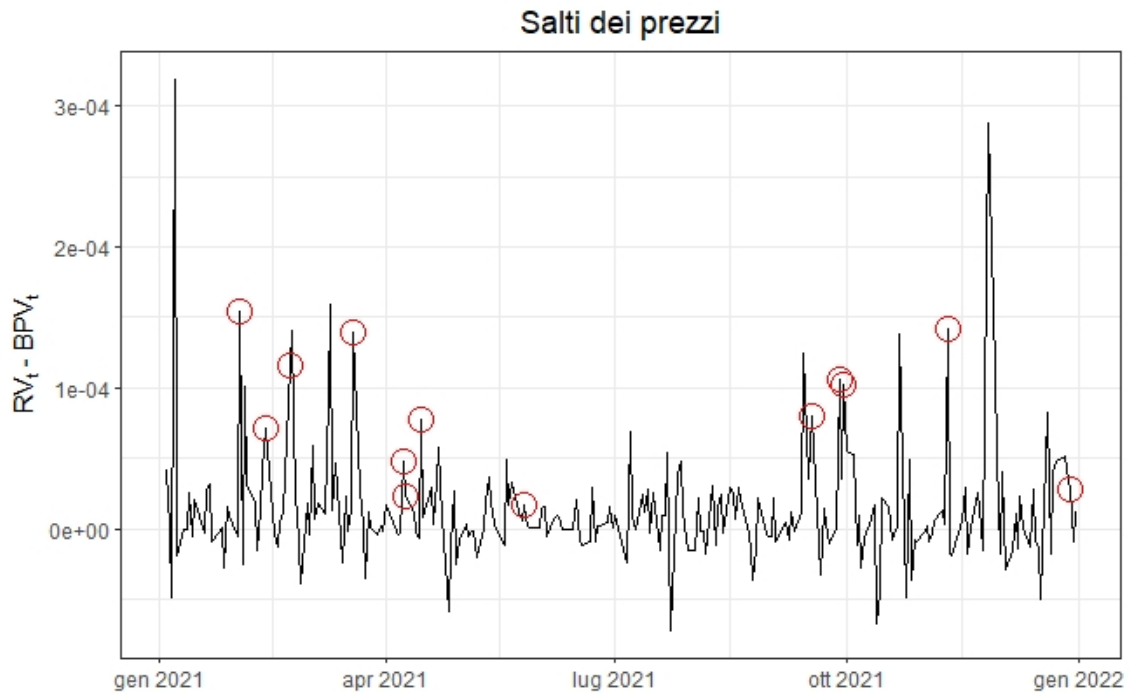


FIGURA 1.5: Individuazione dei giorni in cui è presente un salto di prezzo tramite il test BNS con $\alpha = 0.001$

Kolokolov e Renò [29] hanno proposto una correzione di questo test che tenesse in considerazione la presenza di *staleness*, ovvero di periodi all'interno della giornata in cui sono presenti rendimenti intra-giornalieri nulli.

Per fare ciò, hanno introdotto il seguente stimatore **Multipower Variation** robusto:

$$MV_t^c(\mathbf{e}) = \sum_{i=1}^{m_t-q} |(\Delta(N, i))^{-1/2} \Delta_i P|^{e_1} \dots |(\Delta(N, i+q-1))^{-1/2} \Delta_{i+q-1} P|^{e_q}$$

con m_t numero di rendimenti intra-giornalieri non nulli del giorno t , \mathbf{e} vettore q -dimensionale di esponenti di potenza, $\Delta(N, i) = \tau_i - \tau_{i-1}$ indicatore del tempo passato tra due osservazioni di prezzo con $\tau_0, \tau_1, \dots, \tau_{m_t}$ sottoinsieme della griglia temporale delle osservazioni dei prezzi, $\Delta_i P = \log P_{\tau_i, t} - \log P_{\tau_{i-1}, t}$ rendimento tra due istanti τ successivi.

L'idea è, appunto, di ponderare i rendimenti, utilizzati per il calcolo dello stimatore, per il tempo che intercorre tra le due osservazioni di prezzo differenti.

Il test BNS corretto per la *staleness* diventa:

$$BNS_c = \frac{(RV_t^c - BPV_t^c) RV_t^{c,-1}}{\sqrt{\theta_c \max\left(1, \frac{TPV_t^c}{BPV_t^{c,2}}\right)}}, \quad (1.13)$$

dove $RV_t^c = MV_t^c(2) = RV_t$, $BPV_t^c = \mu_1^{-2}MV_t^c \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \right)$,
 $TPV_t^c = \mu_{4/3}^{-3}MV_t^c \left(\begin{bmatrix} \frac{4}{3} & \frac{4}{3} & \frac{4}{3} \end{bmatrix} \right)$ è lo stimatore dell'*Integrated Quarticity* e $\theta_c = (\frac{\pi^2}{4} + \pi - 5)\bar{\Delta}_N$, con $\bar{\Delta}_N = E[\Delta(N, i)]$ dipendente dal numero di rendimenti non nulli presenti in un dato giorno.

Grazie a questa correzione, si evita di considerare eventuali salti spuri dovuti alla poca liquidità del titolo.

Una volta definito il test BNS_t^c , esso viene utilizzato nella procedura iterativa di Andersen per individuare i rendimenti intra-giornalieri associati a dei salti e definire così le componenti continua e discontinua della Volatilità Realizzata.

1.2.2 Procedura iterativa di Andersen

Il test BNS_t^c permette di individuare se nel giorno t è presente un salto nei prezzi per il titolo considerato, ma non indica quali rendimenti intra-giornalieri sono associati a dei salti. Tale informazione è necessaria per poter decomporre la Volatilità Realizzata nella componente continua e discontinua. Un possibile approccio è quello proposto da Andersen, Bollerslev, Frederiksen e Nielsen (2010) [5], che individua i salti in modo indiretto.

Sia m_j il numero di salti e \mathcal{I} l'insieme di indici di posizione dei salti; tale procedura viene applicata per ogni giornata borsistica t -esima:

1. si valuta la presenza di salti con il test BNS_t^c ; se si rifiuta l'ipotesi nulla, allora è presente almeno un salto e si pone inizialmente $m_j = 1$ e $\mathcal{I} = \{\emptyset\}$;
2. si individua il salto massimizzando la seguente differenza:

$$\max_{i \in 1, 2, \dots, m} r_{t,i}^2 - \frac{1}{m - m_j} \sum_{j=1, j \neq i, j \notin \mathcal{I}}^m r_{t,j}^2;$$

3. il rendimento individuato viene rimosso, l'insieme \mathcal{I} viene aggiornato $\mathcal{I} = \mathcal{I} \cup \{i\}$, lo stimatore della componente continua BPV_t rimane invariato, mentre la RV_t viene ricalcolata:

$$RV_{t,m_j} = \frac{m}{m - m_j} \sum_{i=1, i \notin \mathcal{I}}^m r_{t,i}^2;$$

4. a questo punto, viene rieseguito il test BNS_t^c nello stesso giorno, con il nuovo valore RV_{t,m_j} , per vedere se ci sono ulteriori salti; se si rifiuta l'ipotesi nulla, allora si ricomincia dallo step 2, altrimenti si passa allo step successivo;
5. i rendimenti individuati da tale procedura vanno a formare la componente di salto, mentre i restanti vengono utilizzati per calcolare la componente continua.

1.2.3 Periodicità della volatilità e salti nei prezzi

Nel 2011, Boudt, Croux e Laurent [14] hanno apportato un contributo importante per la rilevazione dei salti, notando un'anomalia nella distribuzione di questi all'interno della giornata borsistica. Nell'esempio in *Figura 1.6* si può notare come la maggior parte dei salti si concentri all'inizio della giornata, fatto non coerente con l'ipotesi di distribuzione casuale dei salti.

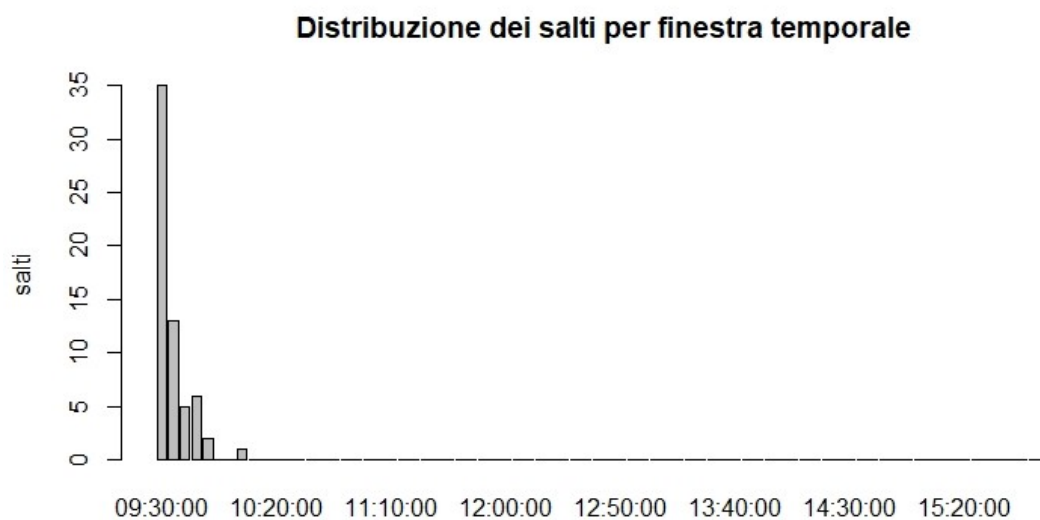


FIGURA 1.6: Distribuzione dei salti a livello intra-giornaliero del titolo AXP nel 2021

Alcuni elementi strutturali dei mercati, come la reazione ad informazioni rilasciate *overnight* o le attività di *day-trading*, influenzano la volatilità intra-giornaliera che non risulta costante, ma presenta picchi ad inizio e fine giornata, come si nota nella *Figura 1.7*, che mostra la componente $\hat{f}_{i,t}^{sd}$ definita successivamente nell'equazione (1.13).

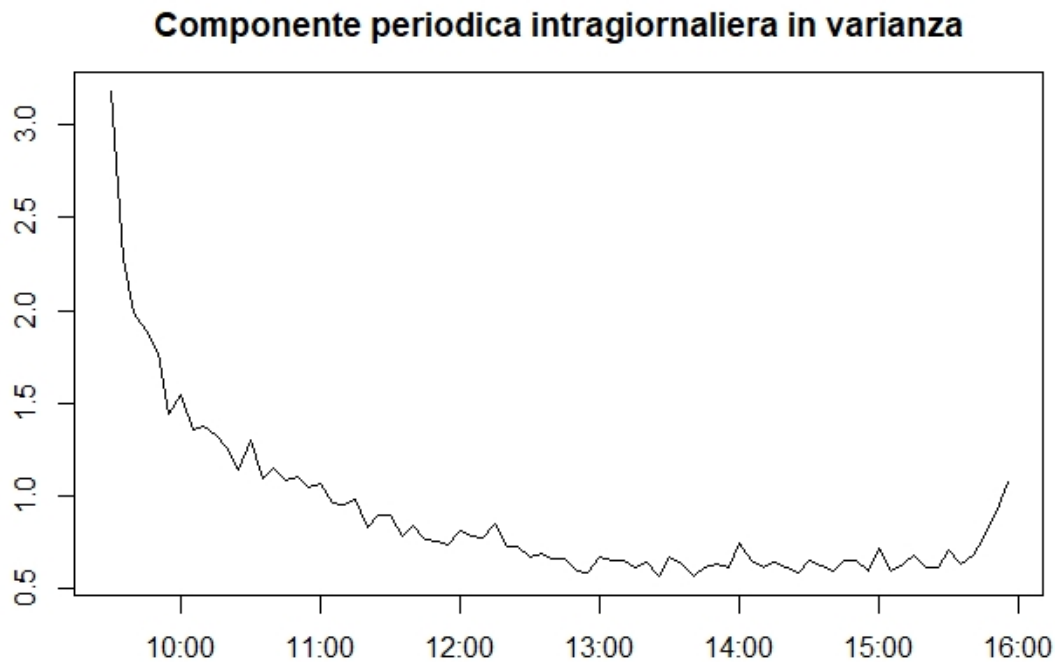


FIGURA 1.7: Componente periodica della varianza all'interno del giorno con dati ogni 5 minuti

Considerare questo andamento della volatilità giornaliera migliora notevolmente l'accuratezza dei metodi di rilevamento dei salti intra-giornalieri. Difatti, si aumenta la capacità di identificare salti relativamente piccoli che si verificano in momenti in cui la volatilità è periodicamente bassa e si riduce sostanzialmente il numero di salti spuri individuati in momenti di volatilità periodicamente elevata.

I rendimenti, perciò, devono essere filtrati da questa periodicità per non distorcere i test di rilevazione dei salti e la stima delle componenti continue e discontinue della volatilità.

Seguendo l'approccio di Boudt, Croux e Laurent [14], si ha il seguente modello generatore dei rendimenti intra-giornalieri:

$$r_{i,t} = f_{i,t}s_{i,t}z_{i,t} + a_{i,t},$$

dove

- $f_{i,t}$ è una componente deterministica che influenza la volatilità;
- $s_{i,t}$ è una componente in volatilità costante all'interno di una finestra;
- $z_{i,t}$ è una innovazione i.i.d. a media nulla e varianza unitaria;
- $a_{i,t}$ è una componente di salto.

L'obiettivo è quello di filtrare i rendimenti per le componenti $s_{i,t}$ e $f_{i,t}$.

Inizialmente, viene calcolata $s_{i,t}$, componente della volatilità che rimane costante all'interno di una finestra temporale scelta, in genere la giornata borsistica.

Questa viene stimata riscaldando la *Bipower Variation*, in quanto bisogna considerare la presenza di salti:

$$\hat{s}_{i,t} = \sqrt{\mu^{-2} \frac{1}{m-1} \sum_{j=2}^m |r_{j,t}| |r_{j-1,t}|}. \quad (1.14)$$

Una volta ottenuta $\hat{s}_{i,t}$, la si rimuove dai dati prima di calcolare $f_{i,t}$:

$$\tilde{r}_{i,t} = \frac{r_{i,t}}{\hat{s}_{i,t}}.$$

Per la stima della componente deterministica che influenza la volatilità esistono diversi approcci, sia parametrici che non. Il più immediato tra gli approcci non parametrici è quello proposto da Taylor and Xu (1997) [38], basato sulla deviazione standard dei rendimenti intra-giornalieri appartenenti alla stessa finestra temporale:

$$SD_i = \sqrt{\frac{1}{T} \sum_{t=1}^T \tilde{r}_{i,t}^2}, \quad (1.15)$$

dove T è la lunghezza del campione di giornate considerate.

Utilizzando questo stimatore, la componente periodica dipende esclusivamente dalla finestra temporale associata al rendimento i -esimo all'interno del giorno. Perciò, tutte le osservazioni calcolate alla stessa ora sono caratterizzate dallo stesso valore della componente periodica.

Tale assunzione può risultare restrittiva, ma enfatizza il fatto stilizzato che la volatilità intra-giornaliera presenti picchi all'inizio e alla fine della giornata.

La stima della periodicità giornaliera si ottiene standardizzando la quantità (1.14):

$$\hat{f}_{i,t}^{sd} = \frac{SD_i}{\frac{1}{m} \sum_{i=1}^m SD_i^2}. \quad (1.16)$$

I rendimenti filtrati delle due componenti della volatilità sono $\tilde{r}_{t,i} = \frac{\tilde{r}_{i,t}}{\hat{f}_{i,t}^{sd}}$ e ad essi si possono applicare i test per l'identificazione dei salti evitando distorsioni dovute alla volatilità giornaliera.

1.3 Decomposizione della Volatilità Realizzata

1.3.1 Componente continua, discontinua, *Good* e *Bad Volatility*

La Volatilità Realizzata viene decomposta per analizzare il diverso contributo apportato dai rendimenti a seconda della loro dimensione o del loro segno. Inoltre, la suddivisione della volatilità giornaliera in componenti può migliorare la sua modellazione e la sua previsione.

Una volta individuati i salti a livello intra-giornaliero, una prima possibile decomposizione può essere quella che suddivide la Volatilità Realizzata in componenti continua e discontinua [2]:

$$RV_t = J_t^2 + CV_t = \sum_{i \in \mathcal{I}} r_{i,t}^2 + \sum_{i \in \bar{\mathcal{I}}} r_{i,t}^2$$

con \mathcal{I} set di indici temporali in cui compaiono salti e $\bar{\mathcal{I}}$ suo complementare, CV_t componente continua e J_t^2 componente discontinua o di salto.

Tale decomposizione mette in luce il contributo dovuto alla diversa dimensione dei rendimenti. Si ricordi che, mentre i salti sono individuati utilizzando i rendimenti filtrati, le componenti sono calcolate successivamente a partire dai rendimenti iniziali.

Un altro aspetto che può essere evidenziato è il diverso contributo alla Volatilità apportato dai rendimenti negativi rispetto ai rendimenti positivi. Patton e Sheppard (2015) [34] hanno introdotto, così, il concetto di *Good volatility* e *Bad volatility* che compongono la Volatilità Realizzata giornaliera nel seguente modo:

$$RV_t = GV_t + BV_t, \tag{1.17}$$

dove

$$GV_t = \sum_{i=1}^m r_{i,t}^2 I(r_{i,t} > 0), \quad BV_t = \sum_{i=1}^m r_{i,t}^2 I(r_{i,t} < 0).$$

Infine, combinando questi due approcci, si ottiene un'unica decomposizione che tiene conto sia del segno che della dimensione dei rendimenti.

$$RV_t = CV_t^+ + J_t^{2,+} + CV_t^- + J_t^{2,-}$$

con

$$CV_t^+ = \sum_{i \in \bar{\mathcal{I}}} r_{i,t}^2 I(r_{i,t} > 0), \quad J_t^{2,+} = \sum_{i \in \mathcal{I}} r_{i,t}^2 I(r_{i,t} > 0),$$

$$CV_t^- = \sum_{i \in \bar{\mathcal{I}}} r_{i,t}^2 I(r_{i,t} < 0), \quad J_t^{2,-} = \sum_{i \in \mathcal{I}} r_{i,t}^2 I(r_{i,t} < 0).$$

1.3.2 Decomposizione basata sulle partizioni: Volatilità Realizzata Parziale

La decomposizione della Volatilità Realizzata giornaliera in componenti continua e discontinua è fortemente condizionata dal modo in cui vengono individuati i salti.

Huang e Tauchen (2005) [27] hanno dimostrato come la variazione della statistica test di salto porti all'identificazione di salti differenti e, quindi, ad un loro diverso impatto relativo sulla Volatilità Realizzata.

Un altro elemento che influisce sull'individuazione dei salti è la scelta del valore critico associato al test che permette di classificare un rendimento estremo come un evento raro. Tale valore viene scelto a discrezione di chi effettua il test e in base alle caratteristiche della distribuzione dei rendimenti del titolo specifico.

Solitamente per i test di salto vengono scelti valori critici α inusuali e più piccoli dei classici valori 0.05 e 0.01. Tuttavia, l'aumento della precisione dei test e l'applicazione di questi valori critici portano ad una riduzione del numero di eventi su cui ci si concentra per la stima dei salti.

Di conseguenza, l'informazione relativa ai rendimenti di grandi dimensioni viene parzialmente persa.

Per ovviare a tale problema, Lee e Wang (2019) [31] hanno provato ad impiegare un valore critico di 0.05, molto più alto dei valori usuali per i test di salto, ottenendo dei risultati empirici interessanti per la stima e la previsione della Volatilità Realizzata giornaliera. Questo differente approccio, si focalizza sull'effetto di rendimenti intra-giornalieri che non possono essere considerati come veri e propri salti nei prezzi, ma

rendimenti di grandi dimensioni.

Tale evidenza ha ispirato alla formulazione di un nuovo metodo per la decomposizione della Volatilità Realizzata, proposto da Caporin (2020) [15].

Questo risulta più flessibile, in quanto scompone la quantità di interesse in gruppi di rendimenti intra-giornalieri e non utilizza alcun test per l'identificazione dei salti.

Data la distribuzione asintotica proposta da Andersen et al. (2007) [4] per i rendimenti filtrati dalla periodicità della volatilità intra-giornaliera

$$\frac{\tilde{r}_{t,i}}{\sqrt{\frac{1}{m}BPV_t}} \sim N(0, 1),$$

sia L il numero di gruppi in cui si vogliono suddividere i rendimenti e q_l con $l = 1, 2, \dots, L$ i quantili della normale standardizzata utilizzati per dividere i gruppi.

L'appartenenza dell' i -esimo rendimento del t -esimo giorno al gruppo denominato \mathcal{N}_l si ha quando:

$$\mathcal{N}_l = \left\{ r_{i,t} : q_{l-1} < \frac{\tilde{r}_{i,t}}{\sqrt{\frac{1}{m}BPV_t}} \leq q_l, \quad l = 1, 2, \dots, L \right\},$$

dove $q_0 = -\infty$, $q_L = +\infty$ e i restanti $q_l = \Phi(\delta_l)^{-1}$ sono scelti arbitrariamente in modo tale che $0 < \delta_1 < \delta_2 < \dots < \delta_{L-1} < 1$.

La Volatilità Realizzata del giorno t viene suddivisa in L partizioni:

$$RV_t = \sum_{l=1}^L PRV_{t,l},$$

dove

$$PRV_{t,l} = \sum_{i \in \mathcal{N}_l} r_{i,t}^2.$$

In questo modo si tiene conto allo stesso tempo sia del valore dei rendimenti, poichè quelli di dimensione assoluta più elevata si collocano nei gruppi estremi e quelli più piccoli nei gruppi centrali, sia del loro segno, in quanto i rendimenti negativi si trovano nei primi gruppi, i positivi negli ultimi e quelli vicini alla neutralità in quelli centrali.

È da precisare, infine, che le partizioni della Volatilità sono state calcolate utilizzando i rendimenti non filtrati, mentre per individuare i gruppi di appartenenza \mathcal{N}_l sono stati

utilizzati quelli filtrati.

Per questo motivo, i rendimenti dopo l'apertura del mercato possono essere inclusi nei gruppi *piccoli* o *medi*, anche se la loro dimensione è maggiore dei rendimenti osservati nel resto della giornata.

Tale partizione, perciò, si svincola da ogni possibile definizione di salto e di statistica test, ma allo stesso tempo prende in considerazione il contributo apportato alla Volatilità dai rendimenti di segno e dimensione differenti.

Capitolo 2

Previsioni gerarchiche per serie storiche

In economia e in finanza molte applicazioni richiedono la previsione di un insieme di serie temporali correlate tra loro che possono essere organizzate in una struttura gerarchica o essere raggruppate in base a delle caratteristiche comuni.

A seconda del tipo di struttura presente si effettua una distinzione degli insiemi di serie. Quando esiste una sola struttura di raggruppamento annidata, la collezione di serie viene denominata “serie temporale gerarchica”. Ad esempio, le vendite di una multinazionale di biciclette possono essere suddivise secondo una gerarchia geografica in paesi, regioni, città e negozi. Infatti, solitamente l’azienda non è interessata a conoscere solamente le previsioni future delle vendite totali, ma anche di quelle per paese, regione e negozio e pretende che queste si sommino in modo appropriato rispettando la gerarchia.

Quando, invece, si è in presenza di raggruppamenti che rispettano vincoli di aggregazione incrociati, più che annidati, la collezione di serie viene denominata “serie temporale raggruppata”.

Riprendendo l’esempio precedente, il produttore di biciclette può essere interessato ad attributi quali le dimensioni del telaio, il tipo di bici (sportiva, da città, da montagna...), la fascia di prezzo, ecc.. In questo caso, tali attributi non si disaggregano naturalmente in un modo gerarchico unico, poiché non sono annidati.

La conoscenza della struttura gerarchica o di vincoli di aggregazione che sono presenti tra le varie serie temporali, è utile per ottenere delle previsioni migliori nei vari livelli della gerarchia. Attraverso l’utilizzo di previsioni gerarchiche, si tiene conto di queste informazioni a priori per ottenere delle previsioni che siano coerenti con la struttura di aggregazione delle serie, cosa che non accadrebbe se si prevedessero tutte le serie in

maniera indipendente.

Inoltre, modellando simultaneamente l'intera gerarchia, si tiene in considerazione l'influenza di cambiamenti di una serie temporale sulle serie dello stesso livello o di livelli superiori e inferiori.

I primi approcci utilizzati sono quelli in cui viene predetto un unico livello della gerarchia che successivamente viene aggregato o disaggregato. Il metodo *bottom-up* aggrega le previsioni delle serie al livello più basso per ottenere i livelli superiori, mentre il metodo *top-down* prevede la serie al livello superiore per poi disaggregarla e ottenere i livelli inferiori; per ultimo, l'approccio *middle-out* parte da un livello intermedio della gerarchia che poi aggrega e disaggrega nelle serie superiori e inferiori. Nessuno di questi metodi, però, tiene conto della struttura di correlazione intrinseca della gerarchia.

Nel 2011, Hyndman et al. [28] hanno proposto un nuovo approccio alla previsione gerarchica che fornisce risultati migliori di quelli prodotti da un approccio *top-down* o *bottom-up*. Il metodo innovativo si basa sulla previsione indipendente di tutte le serie a tutti i livelli della gerarchia e sull'utilizzo di un modello di regressione per combinare tali previsioni in maniera ottimale.

Successivamente, Wickramasuriya et al. (2019) [39] hanno dimostrato che tale approccio risulta impossibile da applicare ed hanno, quindi, presentato un nuovo metodo per la riconciliazione delle previsioni.

In questo capitolo, nella sezione 2.1 viene descritta la differenza tra strutture gerarchiche e raggruppate; successivamente, nella sezione 2.2, vengono presentati gli approcci a livello singolo, mentre nella sezione 2.3 viene utilizzata la notazione matriciale per esprimere il modello predittivo che viene poi impiegato per spiegare i metodi di riconciliazione delle previsioni presenti nella sezione 2.4.

2.1 Serie storiche gerarchiche e raggruppate

Nella *Figura 2.1* è rappresentato un semplice esempio di **struttura gerarchica**. In cima alla gerarchia si trova il Totale (TOT), ovvero il livello più aggregato dei dati. La *t-esima* osservazione del totale viene indicata con y_t per $t = 1, \dots, T$. Si noti che il Totale è suddiviso in due serie che a loro volta si disaggregano rispettivamente in tre e due serie che compongono il livello inferiore. Al di sotto del livello superiore, si usano $y_{j,t}$ per indicare l'osservazione t della serie corrispondente al nodo j . Ad esempio, $y_{A,t}$ indica l'osservazione t della serie corrispondente al nodo A , mentre $y_{AB,t}$ indica l'osservazione t della serie corrispondente al nodo AB e così via.

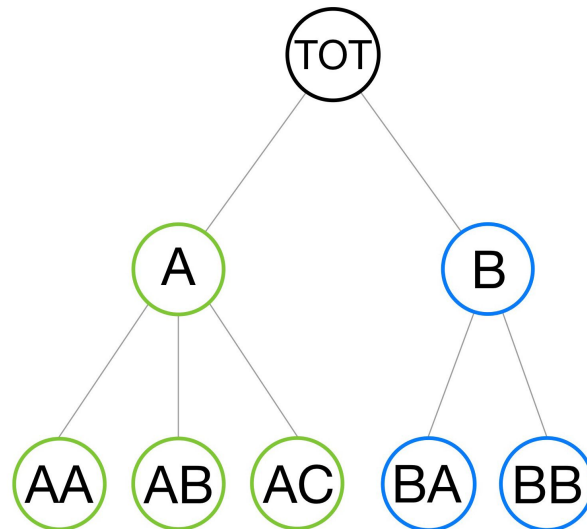


FIGURA 2.1: Rappresentazione di serie storiche gerarchiche

In questo esempio, il numero totale di serie della gerarchia è $n = 1 + 2 + 5 = 8$, mentre il numero di serie al livello inferiore è $m = 5$, con n maggiore di m in tutte le gerarchie: in ogni istante t le serie di un livello si vanno a sommare per costituire le serie del livello superiore. In questo caso si ha:

$$y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t},$$

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t}, \quad y_{B,t} = y_{BA,t} + y_{BB,t}$$

e di conseguenza $y_t = y_{A,t} + y_{B,t}$.

In presenza di **serie storiche raggruppate**, la struttura dei dati non si disaggrega naturalmente in un unico schema gerarchico. Tale caratteristica si nota nell'esempio in *Figura 3.2*. In cima alla struttura raggruppata si trova sempre il Totale, il livello più aggregato dei dati, rappresentato da y_t . Il Totale può essere disaggregato per attributi (A, B), formando le serie $y_{A,t}$ e $y_{B,t}$, o per attributi (X, Y), formando le serie $y_{X,t}$ e $y_{Y,t}$. Al livello inferiore, troviamo le stesse serie disaggregate per entrambi gli attributi.

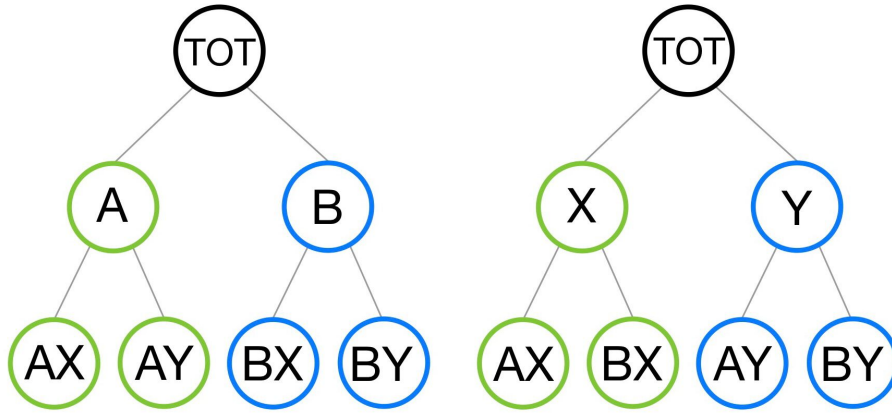


FIGURA 2.2: Rappresentazione di serie storiche raggruppate

Questo esempio mostra che esistono percorsi di aggregazione alternativi per le strutture raggruppate. Per calcolare il totale al tempo t si ha sempre:

$$y_t = y_{AX,t} + y_{AY,t} + y_{BX,t} + y_{BY,t}. \quad (2.1)$$

Tuttavia, il primo livello della struttura raggruppata è composto da

$$y_{A,t} = y_{AX,t} + y_{AY,t}, \quad y_{B,t} = y_{BX,t} + y_{BY,t}, \quad (2.2)$$

ma anche

$$y_{X,t} = y_{AX,t} + y_{BX,t}, \quad y_{Y,t} = y_{AY,t} + y_{BY,t}. \quad (2.3)$$

Le serie temporali raggruppate possono essere considerate come serie temporali gerarchiche che non impongono una struttura gerarchica unica, infatti non esiste un singolo ordine con cui le serie possono essere raggruppate.

2.2 Approcci a livello singolo

I primi approcci proposti per predire le serie temporali gerarchiche o raggruppate sono quelli in cui si effettuano previsioni per un solo livello della gerarchia. Queste vengono poi aggregate per ottenere le previsioni dei livelli superiori o disaggregate per ottenere quelle dei livelli inferiori. Tali metodi permettono di ottenere, dunque, previsioni coerenti, in quanto rispettano la struttura di aggregazione presente nei dati.

2.2.1 Approccio *bottom-up*

Un primo approccio utilizzato per ottenere previsioni coerenti è quello in cui vengono generate previsioni indipendenti, denominate “previsioni di base”, solamente per le serie storiche al livello inferiore, che poi vengono sommate per ottenere quelle ai livelli superiori.

Riprendendo l'esempio in *Figura 2.1*, innanzitutto vengono generate previsioni h passi in avanti per le serie al livello inferiore:

$$\hat{y}_{AA,h}, \hat{y}_{AB,h}, \hat{y}_{AC,h}, \hat{y}_{BA,h}, \hat{y}_{BB,h},$$

dove il generico \hat{y}_h indica, per comodità, la previsione di base $\hat{y}_{T+h|T}$.

Sommandole si ottengono le previsioni coerenti per le serie ai livelli superiori:

$$\begin{aligned} \tilde{y}_h &= \hat{y}_{AA,h} + \hat{y}_{AB,h} + \hat{y}_{AC,h} + \hat{y}_{BA,h} + \hat{y}_{BB,h}, \\ \tilde{y}_{A,h} &= \hat{y}_{AA,h} + \hat{y}_{AB,h} + \hat{y}_{AC,h}, \\ \tilde{y}_{B,h} &= \hat{y}_{BA,h} + \hat{y}_{BB,h}, \end{aligned} \tag{2.4}$$

dove la “tilde” viene utilizzata per le previsioni coerenti.

2.2.2 Approccio *top-down*

Questo metodo, a differenza di quello precedente, genera previsioni per la serie totale che poi disaggrega nella serie di livello inferiore. Riprendendo l'esempio in *Figura 2.1*, siano p_1, \dots, p_m le proporzioni di disaggregazione che determinano come le previsioni della serie totale sono distribuite per ottenere le previsioni al livello più basso della struttura gerarchica. Si ottengono, così, le seguenti previsioni coerenti per il livello inferiore:

$$\tilde{y}_{AA,h} = p_1 \hat{y}_h, \quad \tilde{y}_{AB,h} = p_2 \hat{y}_h, \quad \tilde{y}_{AC,h} = p_3 \hat{y}_h, \quad \tilde{y}_{BA,h} = p_4 \hat{y}_h, \quad \tilde{y}_{BB,h} = p_5 \hat{y}_h.$$

Queste, poi, vengono sommate appropriatamente per ottenere le previsioni delle serie ai livelli intermedi. A questo punto è necessario scegliere un metodo per calcolare le proporzioni di disaggregazione.

Proporzioni storiche medie

$$p_j = \frac{1}{T} \sum_{t=1}^T \frac{y_{j,t}}{y_t}$$

con $j = 1, \dots, m$.

Ogni proporzione p_j riflette la media delle proporzioni storiche delle serie di livello inferiore $y_{j,t}$ rispetto al valore dell'aggregato totale y_t .

Proporzioni delle medie storiche

$$p_j = \sum_{t=1}^T \frac{y_{j,t}}{T} / \sum_{t=1}^T \frac{y_t}{T}$$

con $j = 1, \dots, m$.

Ogni proporzione p_j rappresenta il valore storico medio della serie al livello inferiore $y_{j,t}$ rispetto al valore medio dell'aggregato totale y_t .

Questi primi due approcci sono semplici da applicare e garantiscono buone prestazioni come risulta nello studio di Gross, Sohl e Jeffrey (1990) [23], ma non considerano la possibilità che le proporzioni possano cambiare nel tempo.

Proporzioni di previsione

Nel 2009, Athanasopoulos et al. [7] hanno proposto l'utilizzo di proporzioni che si basassero su previsioni future piuttosto che su dati storici. Per spiegare come vengono ottenute, si consideri il caso più semplice di una gerarchia avente un solo livello.

Per prima cosa vengono effettuate le previsioni di base h passi in avanti di tutte le serie al livello inferiore. Ovviamente, le previsioni ottenute non si sommano coerentemente. Dopodichè si calcola la proporzione tra ogni previsione di base h passi in avanti al livello inferiore e la somma di tutte le previsioni di base a questo livello. Si ottengono, quindi, le cosiddette "proporzioni di previsione", che vengono poi utilizzate per disaggregare le previsioni di base del livello superiore h passi in avanti, ottenendo quelle coerenti per l'intera gerarchia.

Generalizzando ad una gerarchia di K livelli, questo processo viene ripetuto per ogni nodo dal livello più alto a quello più basso; si ottiene così la seguente regola:

$$p_j = \prod_{\ell=0}^{K-1} \frac{\hat{y}_{j,h}^{(\ell)}}{\hat{S}_{j,h}^{(\ell+1)}} \quad (2.5)$$

con $j = 1, \dots, m$, $\hat{y}_{j,h}^{(\ell)}$ previsione di base h passi in avanti della serie corrispondente al nodo situato ℓ livelli sopra a quello j e $\hat{S}_{j,h}^{(\ell+1)}$ somma delle previsioni di base delle serie

sottostanti al nodo situato l livelli sopra a quello j . Una volta ricavate, le proporzioni vengono utilizzate per disaggregare le previsioni h passi in avanti della serie totale e ottenere le previsioni h passi in avanti delle serie inferiori.

Si riprenda come esempio la gerarchia in *Figura 2.1* e si supponga di aver ricavato le previsioni per tutte le serie della gerarchia, ricordando che per gli approcci *top-down* si ha $\hat{y}_h = \tilde{y}_h$. Vengono riportati di seguito gli elementi per calcolare le proporzioni secondo la formula (2.1):

- $\hat{y}_{A,h}^{(1)} = \hat{y}_{B,h}^{(1)} = \hat{y}_h = \tilde{y}_h$;
- $\hat{y}_{AA,h}^{(1)} = \hat{y}_{AB,h}^{(1)} = \hat{y}_{AC,h}^{(1)} = \hat{y}_{A,h}$;
- $\hat{y}_{AA,h}^{(2)} = \hat{y}_{AB,h}^{(2)} = \hat{y}_{AC,h}^{(2)} = \hat{y}_{BA,h}^{(2)} = \hat{y}_{BB,h}^{(2)} = \hat{y}_h = \tilde{y}_h$;
- $\hat{S}_{AA,h}^{(1)} = \hat{S}_{AB,h}^{(1)} = \hat{S}_{AC,h}^{(1)} = \hat{y}_{AA,h} + \hat{y}_{AB,h} + \hat{y}_{AC,h}$;
- $\hat{S}_{AA,h}^{(2)} = \hat{S}_{AB,h}^{(2)} = \hat{S}_{AC,h}^{(2)} = \hat{S}_{A,h}^{(1)} = \hat{S}_{B,h}^{(1)} = \hat{S}_h = \hat{y}_{A,h} + \hat{y}_{B,h}$.

Concentrandosi sulla parte sinistra della gerarchia si ha:

$$\begin{aligned}\tilde{y}_{A,h} &= \left(\frac{\hat{y}_{A,h}}{\hat{S}_{A,h}^{(1)}} \right) \tilde{y}_h = \left(\frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(2)}} \right) \tilde{y}_h, \\ \tilde{y}_{AA,h} &= \left(\frac{\hat{y}_{AA,h}}{\hat{S}_{AA,h}^{(1)}} \right) \tilde{y}_{A,h} = \left(\frac{\hat{y}_{AA,h}}{\hat{S}_{AA,h}^{(1)}} \right) \left(\frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(2)}} \right) \tilde{y}_h.\end{aligned}$$

Di conseguenza

$$p_1 = \left(\frac{\hat{y}_{AA,h}}{\hat{S}_{AA,h}^{(1)}} \right) \left(\frac{\hat{y}_{AA,h}^{(1)}}{\hat{S}_{AA,h}^{(2)}} \right).$$

Le altre proporzioni si ottengono allo stesso modo.

2.2.3 Approccio *middle-out*

L'approccio *middle-out* combina le due tecniche viste in precedenza e può essere utilizzato solo con strutture strettamente gerarchiche. Inizialmente vengono effettuate le previsioni di base delle serie di un livello intermedio, dopodichè le previsioni coerenti per le serie a livelli superiori si ottengono tramite l'approccio *bottom-up*, mentre quelle a livello inferiore tramite l'approccio *top-down*.

2.2.4 Aspetti positivi e negativi degli approcci

Negli ultimi decenni del '900, la maggior parte della letteratura previsionale in questo settore ha esaminato le prestazioni di questi approcci cercando di capire quale fosse il migliore in termini predittivi. Già nel 1960, Grunfeld e Griliches [24] hanno apportato un primo contributo sostenendo che i dati disaggregati sono soggetti ad errori e per questo motivo le previsioni *top-down* risultano più accurate. Conclusioni simili sono state tratte successivamente da Fogarty et al. (1991) [21], Narasimhan et al. (1995) [32] e Fliedner (1999) [20].

D'altra parte, Orcutt et al. (1968) [33] e Edwards et al. (1969) [18] hanno sostenuto che con la disaggregazione vengono perse informazioni importanti, perciò il metodo *bottom-up* fornisce previsioni più accurate. Anche Shlifer e Wolff (1979) [36] hanno confrontato le prestazioni previsionali di entrambi i metodi e hanno concluso che quello *bottom-up* è preferibile a determinate condizioni sulla struttura della gerarchia e sull'orizzonte di previsione. Schwarzkopf et al. (1988) [35], esaminando il bias e la robustezza dei due metodi, hanno concluso che il metodo *bottom-up* è migliore, tranne quando i dati a livelli più bassi sono scarsi o inaffidabili.

Le tabelle successive riassumono i principali aspetti positivi e negativi degli approcci *bottom-up* e *top-down*.

Pro	Contro
<ul style="list-style-type: none"> • Non si ha perdita di informazione dovuta alla disaggregazione • Le previsioni coerenti ottenute non sono distorte 	<ul style="list-style-type: none"> • Produce previsioni non accurate con dati ai livelli inferiori di scarsa qualità • Necessita della previsione di più serie di base

TABELLA 2.1: Pro e contro del metodo *bottom-up*

Pro	Contro
<ul style="list-style-type: none"> • Non è influenzato dal rumore presente nei dati • Necessita della previsione della sola serie totale 	<ul style="list-style-type: none"> • Si ha una perdita di informazioni dovuta alla disaggregazione • Le previsioni coerenti ottenute sono distorte

 TABELLA 2.2: Pro e contro del metodo *top-down*

2.3 Notazione matriciale

La notazione matriciale [39] viene utilizzata per rappresentare in maniera più efficiente i vincoli di aggregazione che intercorrono tra le varie serie della struttura gerarchica o raggruppata. Si prendano come esempio la struttura raggruppata in *Figura 2.2* e le equazioni di aggregazione (2.1), (2.2), (2.3). I vincoli di aggregazione che intercorrono tra le serie possono essere rappresentati nella seguente forma matriciale:

$$Y_t = \mathbf{S}b_t$$

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{X,t} \\ y_{Y,t} \\ y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AX,t} \\ y_{AY,t} \\ y_{BX,t} \\ y_{BY,t} \end{bmatrix},$$

dove $\mathbf{S} \in \mathbb{R}^{n \times m}$ è la matrice somma che determina il modo in cui le serie di livello inferiore si aggregano, Y_t è un vettore di dimensione n composto dal valore di tutte le serie storiche della gerarchia al tempo t , mentre b_t è un vettore m -dimensionale costituito dai valori delle serie situate nel livello inferiore della gerarchia al tempo t .

Si noti che la prima riga della matrice \mathbf{S} rappresenta l'equazione in (2.1), la seconda e

la terza le equazioni in (2.2) e la quarta e la quinta le equazioni in (2.3). Le ultime m righe sono sostanzialmente una matrice identità \mathbf{I}_m , tale che le serie del livello inferiore nel lato destro dell'equazione siano uguali a loro stesse nel lato sinistro.

Questa notazione matriciale permette di rappresentare tutti i metodi di previsione gerarchica per le serie temporali. Infatti, tutti gli approcci di previsione coerente per strutture gerarchiche o raggruppate possono essere riscritti nella seguente maniera:

$$\tilde{Y}_h = \mathbf{S}\mathbf{G}\hat{Y}_h, \quad (2.6)$$

dove $\tilde{Y}_h = \tilde{Y}_{T+h|T}$ rappresenta il vettore delle previsioni coerenti di tutte le serie h passi in avanti, $\hat{Y}_h = \hat{Y}_{T+h|T}$ è il vettore di previsioni di base h passi in avanti ottenute in maniera indipendente per tutte le serie della gerarchia, \mathbf{S} è la matrice somma vista in precedenza che indica i vincoli di aggregazione della struttura e \mathbf{G} rappresenta la matrice che mappa le previsioni di base in previsioni disaggregate di livello inferiore che vengono poi sommate da \mathbf{S} .

La matrice \mathbf{G} viene definita in base all'approccio che si vuole applicare.

Prendendo di nuovo in considerazione l'esempio di struttura raggruppata in *Figura 2.2*, se si volesse adoperare l'approccio *bottom-up* per ottenere le previsioni coerenti, la matrice \mathbf{G} sarebbe la seguente:

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Le prime cinque colonne della matrice \mathbf{G} sono composte da tutti zeri, in quanto annullano le previsioni di base delle serie al di sopra di quelle inferiori, mentre la restante matrice identità m -dimensionale seleziona solo le previsioni di base del livello inferiore, che vengono poi sommate dalla matrice \mathbf{S} per ottenere le previsioni coerenti delle serie ai livelli superiori.

Se si decidesse di usare, invece, un approccio *top-down*, la matrice \mathbf{G} diventerebbe:

$$\mathbf{G} = \begin{bmatrix} p_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (2.7)$$

La prima colonna comprende l'insieme di proporzioni che serve a disaggregare il livello più alto della gerarchia per ottenere le previsioni coerenti del livello più basso. Queste, poi, vengono sommate opportunamente dalla matrice \mathbf{S} per ottenere le restanti previsioni di serie. Le altre colonne azzerano le previsioni di base al di sotto del livello di aggregazione più alto.

Quindi, l'equazione (2.5) mostra che per ottenere le previsioni coerenti è sufficiente premoltiplicare le previsioni di base per le opportune matrici \mathbf{SG} , dove la matrice \mathbf{G} cambia a seconda dell'approccio utilizzato. Tuttavia, i metodi visti in precedenza sono piuttosto limitati in quanto utilizzano le previsioni di base di un solo livello per ottenere le previsioni coerenti delle altre serie, perdendo così l'informazione ricavabile dalle restanti previsioni di base e non tenendo conto delle correlazioni presenti tra le diverse serie temporali. Nel 2011, Hyndman et al. [28] hanno proposto un nuovo approccio in cui viene ricavata la matrice \mathbf{G} ottimale per ottenere previsioni coerenti più accurate.

2.4 Riconciliazione delle previsioni

La riconciliazione delle previsioni è il processo che modifica le previsioni di base per renderle coerenti alla struttura di aggregazione delle serie temporali. I metodi di riconciliazione delle previsioni vengono, quindi, adoperati per individuare la matrice \mathbf{G} ottimale e ottenere previsioni riconciliate non distorte che siano le più accurate possibili. Usando questi metodi, si tiene conto della struttura di correlazione presente all'interno della gerarchia, cosa che non accade quando si utilizzano gli approcci a livello singolo trattati precedentemente.

2.4.1 Correttezza delle previsioni coerenti

Innanzitutto, si vogliono ottenere delle previsioni coerenti che non siano distorte. Hyndman et al. [28] hanno dimostrato che non tutti gli approcci rispettano tale condizione.

Riprendendo la formula (2.5), con la quale possono essere riscritti tutti i metodi di previsione gerarchica,

$$\tilde{Y}_h = \mathbf{SG}\hat{Y}_h,$$

si assuma che le previsioni di base indipendenti siano non distorte $E[\hat{Y}_h] = E[Y_h]$.

Per ricavare previsioni coerenti non distorte, è necessario che $E[\tilde{Y}_h] = E[Y_h] = \mathbf{SE}[b_h]$, con $b_h = b_{T+h}$ vettore m -dimensionale di tutte le serie del livello più basso della gerarchia al tempo $T + h$.

Si supponga che $\beta_h = E[b_h|Y_1, \dots, Y_T]$ sia il valore atteso futuro del livello inferiore alla gerarchia. Perciò, secondo la formula (2.5), si ha che $E[\tilde{Y}_h] = \mathbf{SG}E[\hat{Y}_h] = \mathbf{SGS}\beta_h$. Quindi, le previsioni coerenti risultano non distorte imponendo un vincolo alla matrice \mathbf{G} in modo tale che

$$\mathbf{SGS} = \mathbf{S}.$$

Tale condizione si verifica, per esempio, quando si utilizza un approccio *bottom-up*, in quanto la matrice \mathbf{G} è la (2.6), mentre non si verifica con l'approccio *top-down* con \mathbf{G} in (2.7), qualsiasi siano le p nella prima colonna.

2.4.2 Riconciliazione GLS

Nel 2011, Hyndman et al. [28] hanno proposto un metodo di riconciliazione basato su un semplice modello lineare che combinasse in maniera ottimale le previsioni di base di tutte le serie.

Le previsioni ottenute con questo approccio si sommano in maniera appropriata nelle gerarchie, non sono distorte e hanno minima varianza tra tutte le combinazioni di previsioni possibili.

Il loro approccio, denominato anche approccio della “combinazione ottimale”, si basa sul seguente modello di regressione:

$$\hat{Y}_h = \mathbf{S}\beta_h + \epsilon_h,$$

dove β_h rappresenta il valore atteso sconosciuto delle serie al livello inferiore della gerarchia al tempo $T+h$ e ϵ_h è l'errore di coerenza con media zero e matrice di covarianza

$$\text{Var}(\epsilon_h|\mathcal{I}_t) = \Sigma_h. \quad (2.8)$$

Se Σ_h è nota, è possibile utilizzare la stima dei minimi quadrati generalizzati per

ottenere la stima non distorta e con varianza minima di β_h , pari a

$$\hat{\beta}_h = (\mathbf{S}'\Sigma_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\Sigma_h^{-1}\hat{Y}_h,$$

da cui si ottengono le seguenti previsioni riconciliate:

$$\tilde{Y}_h = \mathbf{S}\hat{\beta}_h = \mathbf{S}\mathbf{G}\hat{Y}_h,$$

dove

$$\mathbf{G} = (\mathbf{S}'\Sigma_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\Sigma_h^{-1}. \quad (2.9)$$

Tali previsioni risultano non distorte e la loro varianza è:

$$Var[\tilde{Y}_h] = \mathbf{S}(\mathbf{S}'\Sigma_h^{-1}\mathbf{S})^{-1}\mathbf{S}'.$$

Il metodo esposto, però, presenta un importante punto critico. Nel 2019, Wickramasuriya et al. [39] hanno dimostrato che l'approccio proposto da Hyndman et al. [28] risulta impraticabile, in quanto la matrice di covarianza Σ_h non rispetta la condizione di identificabilità del modello.

Per questo motivo, hanno introdotto un nuovo metodo per la riconciliazione, denominato MinT.

2.4.3 Riconciliazione *Minimum Trace* (MinT)

L'approccio di riconciliazione *Minimum Trace* (MinT), presentato da Wickramasuriya et al. nel 2019 [39], stima la matrice \mathbf{G} ottimale, minimizzando la somma delle varianze degli errori di previsioni riconciliate sotto il vincolo di non distorsione.

Gli errori di previsioni riconciliate o coerenti h passi in avanti sono definiti nella seguente maniera:

$$\tilde{e}_T(h) = Y_{T+h} - \tilde{Y}_h, \quad (2.10)$$

dove \tilde{Y}_h sono le previsioni riconciliate h passi in avanti ottenute utilizzando le informazioni fino al tempo T compreso e Y_{T+h} sono i valori osservati di tutte le serie al tempo $T + h$.

Di conseguenza, la matrice di varianza-covarianza degli errori di previsioni coerenti h passi in avanti è:

$$V_h = Var[Y_{T+h} - \tilde{Y}_h] = \mathbf{S}\mathbf{G}\mathbf{W}_h\mathbf{G}'\mathbf{S}', \quad (2.11)$$

dove $W_h = \text{Var}[Y_{T+h} - \hat{Y}_h] = E[\hat{e}'_T(h)\hat{e}_T(h)]$ è la matrice di varianza-covarianza dei corrispondenti errori di previsione di base

$$\hat{e}_T(h) = Y_{T+h} - \hat{Y}_h. \quad (2.12)$$

Secondo l'approccio di Wickramasuriya et al. [39], la matrice \mathbf{G} ottimale è quella che minimizza le varianze di errore delle previsioni coerenti. Tali varianze si trovano nella diagonale della matrice V_h dell'equazione (2.10), perciò la somma delle varianze di errore equivale alla traccia di tale matrice.

Wickramasuriya et al. [39] hanno mostrato che la matrice \mathbf{G} che minimizza la traccia di V_h , tale che $\mathbf{S}\mathbf{G}\mathbf{S} = \mathbf{S}$, è pari a

$$\mathbf{G} = (\mathbf{S}'W_h^{-1}\mathbf{S})^{-1}\mathbf{S}'W_h^{-1}, \quad (2.13)$$

perciò le previsioni riconciliate ottimali sono:

$$\tilde{Y}_h = \mathbf{S}(\mathbf{S}'W_h^{-1}\mathbf{S})^{-1}\mathbf{S}'W_h^{-1}\hat{Y}_h. \quad (2.14)$$

Si noti che l'unica differenza tra la soluzione GLS in (2.7) e quella MinT in (2.10) è la matrice di covarianza che entra negli stimatori. Nella soluzione GLS è presente Σ_h , ovvero la matrice di covarianza degli errori di coerenza (2.7), che non può essere identificata, mentre in quella MinT compare la matrice W_h di covarianza degli errori di previsione di base (2.11).

Nonostante W_h non soffra di una mancanza di identificazione, rimane difficile da stimare, in particolar modo quando $h > 1$. Al suo posto vengono utilizzate diverse approssimazioni, discusse da Wickramasuriya et al. [39], che vengono scelte a seconda delle assunzioni fatte sulla struttura dei dati presenti.

2.4.3.1 Stimatori alternativi di W_h

1. $W_h = k_h\mathbf{I}$, $\forall h$, dove $k_h > 0$ è una costante. Adoperando questa semplice approssimazione, lo stimatore MinT collassa a quello dei minimi quadrati OLS di Hyndman et al. [28].

In questo caso, si assume semplicemente che la matrice \mathbf{G} sia indipendente dai dati. Tale soluzione comporta sostanziali vantaggi computazionali, ma non tiene conto né delle differenze di scala tra i vari livelli della struttura, né delle relazioni

che intercorrono tra le varie serie temporali. Quindi, risulta una buona approssimazione quando gli errori di previsione di base sono equivarianti e incorrelati, fatto che raramente accade in presenza di una struttura gerarchica o raggruppata.

2. $W_h = k_h \text{diag}(\hat{W}_1) \forall h$, dove $k_h > 0$ è una costante e

$$\hat{W}_1 = \frac{1}{T} \sum_{t=1}^T \hat{e}_t(1) \hat{e}_t(1)',$$

è lo stimatore distorto di covarianza campionaria degli errori di previsione di base un passo avanti calcolati nel campione. Con questa specificazione il MinT può essere considerato uno stimatore dei minimi quadrati pesati, applicando una varianza scalata (WLS_v).

3. $W_h = K_h \Lambda$, $\forall h$ dove $k_h > 0$ e $\Lambda = \text{diag}(\mathbf{S1})$ con $\mathbf{1}$ un vettore colonna unitario di dimensione m . Questa soluzione, proposta per la prima volta da Athanasopoulos et al. [8], assume che gli errori di previsione di base di livello inferiore abbiano una varianza k_h e siano non correlati tra i nodi. Quindi, ogni elemento della matrice diagonale Λ contiene il numero di varianze di errore di previsione che contribuiscono a quel livello di aggregazione. Questo stimatore dipende esclusivamente dalla struttura di aggregazione delle serie e non dai dati, pertanto viene denominato WLS_s, ovvero uno stimatore che applica lo *scaling* strutturale. Rispetto a quello OLS, ha il vantaggio di assumere errori di previsione equivarianti solo al livello inferiore della struttura e non a tutti i livelli. L'applicazione di questa ponderazione strutturale è particolarmente utile nei casi in cui i residui non siano disponibili e quindi non è possibile applicare il WLS_v; ad esempio, nei casi in cui le previsioni di base sono generate da previsioni basate sul giudizio (previsioni fatte in assenza di dati storici).
4. $W_h = k_h \hat{W}_1 \forall h$, con $k_h > 0$ [39]. In questo caso si assume solo che le matrici di covarianza degli errori siano proporzionali tra loro e viene stimata direttamente la matrice di covarianza completa un passo in avanti \hat{W}_1 senza alcuna restrizione. Il modo più ovvio per la stima sarebbe quello di utilizzare la covarianza campionaria che, tuttavia, non è un buono stimatore nei casi in cui il numero di serie di livello inferiore m è elevato rispetto alla lunghezza T della serie. In tali condizioni, viene impiegato uno stimatore *shrinkage* che comprime la covarianza del campione a una matrice diagonale.

5. $W_h = k_h \hat{W}_{1,D}^* \forall h$, dove $k_h > 0$; $\hat{W}_{1,D}^* = \lambda_D \hat{W}_{1,D} + (1 - \lambda_D) \hat{W}_1$ è uno stimatore *shrinkage* con target diagonale, $\hat{W}_{1,D}^*$ è una matrice diagonale costituita dai valori diagonali di \hat{W}_1 e λ_D è il parametro di intensità dello *shrinkage*. Pertanto, gli elementi fuori dalla diagonale di \hat{W}_1 sono compressi verso lo zero, mentre gli elementi lungo la diagonale, ovvero le varianze, non cambiano.

Capitolo 3

Applicazione: previsione della Volatilità Realizzata del titolo Cisco Systems

3.1 Contesto e obiettivo dell'analisi

In questo capitolo viene applicata la teoria delle previsioni gerarchiche, trattata in precedenza, per prevedere la Volatilità Realizzata di un titolo finanziario.

Lo scopo dell'analisi è quello di ottenere delle previsioni gerarchiche coerenti che siano più accurate di quelle ricavate a partire dal modello HAR-RV (*Heterogeneous Auto Regression*) di Corsi [16], che rimane uno dei modelli più utilizzati per la previsione della Volatilità Realizzata.

Le previsioni gerarchiche sono state ricavate specificando quattro strutture di aggregazione della Volatilità Realizzata, ognuna delle quali è stata creata a partire da decomposizioni differenti di tale quantità di interesse. Per rendere più completo l'esperimento predittivo, sono state utilizzate strutture con complessità varie. In questo modo si riesce a comprendere se l'imposizione a priori di diversi vincoli di aggregazione porti esiti significativamente migliori nella previsione della Volatilità Realizzata. Tra i risultati dell'analisi viene posto maggiore interesse su quelli ottenuti a partire da strutture che si riferiscono alla decomposizione della Volatilità Realizzata basata sulla partizione quantilica descritta nella sezione 1.3.2.

3.2 Dati utilizzati: il titolo *Cisco Systems*

I dati utilizzati nelle analisi fanno riferimento ai rendimenti ad un minuto del titolo *Cisco Systems* (CSCO) dal giorno 02/08/2010 al 30/06/2022, per un totale di 3000 giornate borsistiche, e sono stati scaricati dal sito *www.kibot.com*.

Cisco Systems è una società americana che produce attrezzature di rete, software e servizi per l'industria delle telecomunicazioni. La società è stata fondata nel 1984 ed è diventata un'importante forza nell'industria delle telecomunicazioni. *Cisco Systems* è stata quotata in borsa per la prima volta il 16 febbraio 1990; in quell'occasione, le azioni della società sono state messe in vendita tramite un'offerta pubblica iniziale (IPO) sulla borsa NASDAQ, uno dei principali mercati azionari degli Stati Uniti. Nel corso degli anni, le azioni di *Cisco Systems* hanno subito fluttuazioni a seconda di una serie di fattori, tra cui le prestazioni finanziarie dell'azienda, le notizie relative all'industria delle telecomunicazioni, le tendenze del mercato azionario e altri eventi geopolitici.

3.3 Esercizio di previsione e criteri di raffronto

Per confrontare la capacità predittiva del modello di Corsi con quella dei metodi gerarchici è stato considerato il seguente esercizio di previsione: si è scelto di adottare un approccio *rolling* per la previsione della Volatilità Realizzata. È stata, quindi, fissata una finestra mobile di 1000 osservazioni, all'incirca 4 anni, mentre il periodo *out-of-sample* è composto dalle restanti 2000 osservazioni, che vanno dal giorno 23/07/2014 al 30/06/2022.

Per ogni modello si ottengono, così, 2000 previsioni un passo in avanti della Volatilità Realizzata che vengono comparate in termini di *Mean Absolute Error*, di *Mean Squared Error* e applicando il *Model Confidence Set*.

L'esercizio, poi, è stato ripetuto per verificare se i risultati dell'analisi sono sensibili all'ampiezza della finestra *rolling* e del periodo *out-of sample*. Questa volta la finestra *rolling* è stata fissata a 250 osservazioni e quella *out-of-sample* a 2750, dal 28/07/2011 al 30/06/2022.

Il *Mean Absolute Error* e il *Mean Squared Error* sono due metriche di valutazione della bontà di un modello di previsione e si calcolano attraverso le seguenti formule:

$$MAE = \frac{1}{N} \sum_{t=1}^N \left| RV_t - \widehat{RV}_t \right|, \quad MSE = \frac{1}{N} \sum_{t=1}^N \left(RV_t - \widehat{RV}_t \right)^2,$$

dove \widehat{RV}_t è la serie della Volatilità Realizzata prevista un passo in avanti per ogni modello e N il numero di osservazioni *out-of-sample*.

Il MSE tende a penalizzare gli errori più grandi in modo più significativo rispetto al MAE poiché gli errori sono elevati al quadrato. D'altra parte, il MAE è meno sensibile a possibili *outliers*. Per le esigenze specifiche di questo esercizio, si è deciso di tener conto principalmente dei risultati in termini di MSE, in quanto si vuole dare più peso agli errori maggiori poiché essi portano a valutazioni completamente errate della Volatilità del titolo.

Per effettuare un confronto affidabile tra più modelli, è stato adottato anche il *Model Confidence Set* (MCS) di Hansen et al. (2003, 2011) [26] [25], cioè un metodo flessibile che esclude da un insieme di modelli solamente quelli che producono delle previsioni significativamente peggiori.

Per comparare i modelli concorrenti, viene effettuato un test creato a partire da una funzione di perdita. In particolare, possono essere effettuati due diversi test statistici: il test R e il test SQ. Entrambi associano ai modelli confrontati un *p-value* relativo alla probabilità di inclusione di ciascuno di essi nell'insieme dei modelli statisticamente equivalenti. L'MCS, poi, include nel set di modelli quelli che superano una soglia (*p-value*) relativamente alta.

In questa applicazione, il test utilizzato per il *Model Confidence Set* è il test SQ, basato sulla funzione di perdita del MSE; inoltre, il *p-value* minimo affinché un modello possa entrare nell'insieme è stato fissato a 0.20.

Per maggiori dettagli su questa procedura e sulla sua implementazione in R, si consultino gli articoli di Hansen et al. (2011) [25] e Bernardi e Catania (2018) [12].

3.4 Analisi descrittive della Volatilità Realizzata

La Volatilità Realizzata è stata calcolata attraverso l'equazione (1.5)

$$RV_t = \sum_{i=1}^m r_{i,t}^2,$$

dove $r_{i,t}$ è l'*i-esimo* rendimento al minuto della giornata t e di conseguenza $m = 390$, dato che il mercato americano apre alle ore 09:30 USA e chiude alle ore 16:00 USA.

Il grafico in *Figura 3.1* mostra la serie della Volatilità Realizzata del titolo *Cisco Systems* durante il periodo considerato.

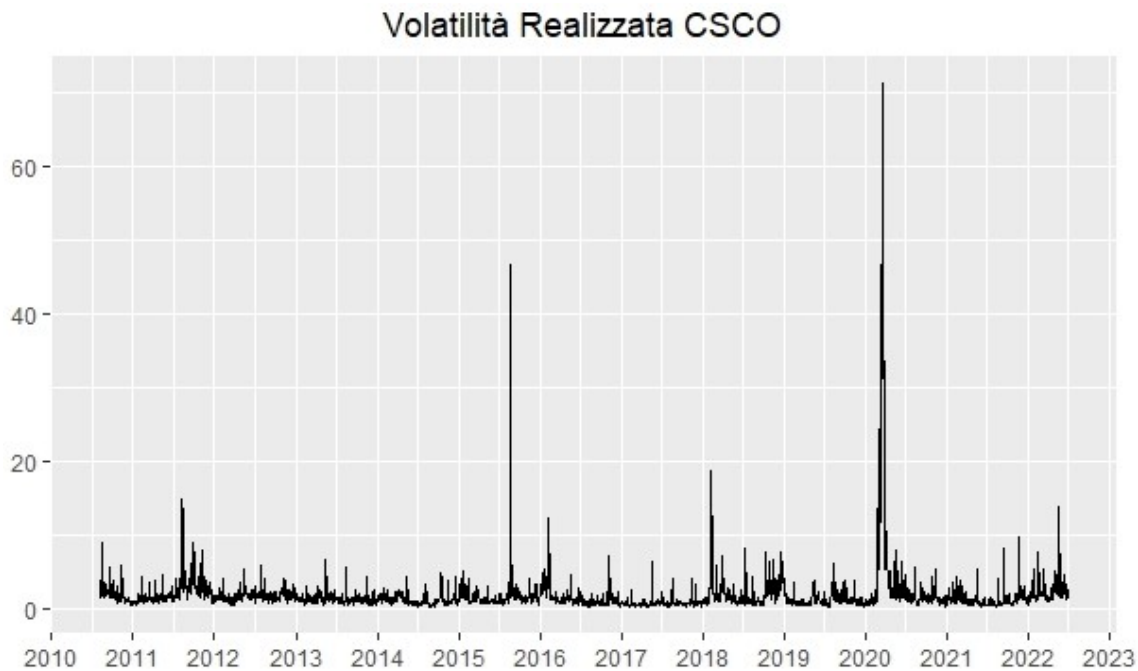


FIGURA 3.1: Andamento della Volatilità Realizzata con rendimenti al minuto del titolo CSCO

Si può notare come la Volatilità non sia costante nel tempo: sono presenti periodi di Volatilità minore, come nel 2017, e fasi di maggiori fluttuazioni, come all’inizio della pandemia (dopo il 2020). Inoltre, è evidente la presenza di un considerevole picco il 24 agosto 2015, giorno in cui è avvenuto il crollo delle borse cinesi che ha scatenato una vendita massiccia di azioni in tutto il mondo e un conseguente aumento di volatilità dei titoli.

Di seguito vengono riportati anche l’istogramma della distribuzione della Volatilità Realizzata (*Figura 3.2*) e una tabella riassuntiva dei quantili della serie di interesse (*Tabella 3.1*).

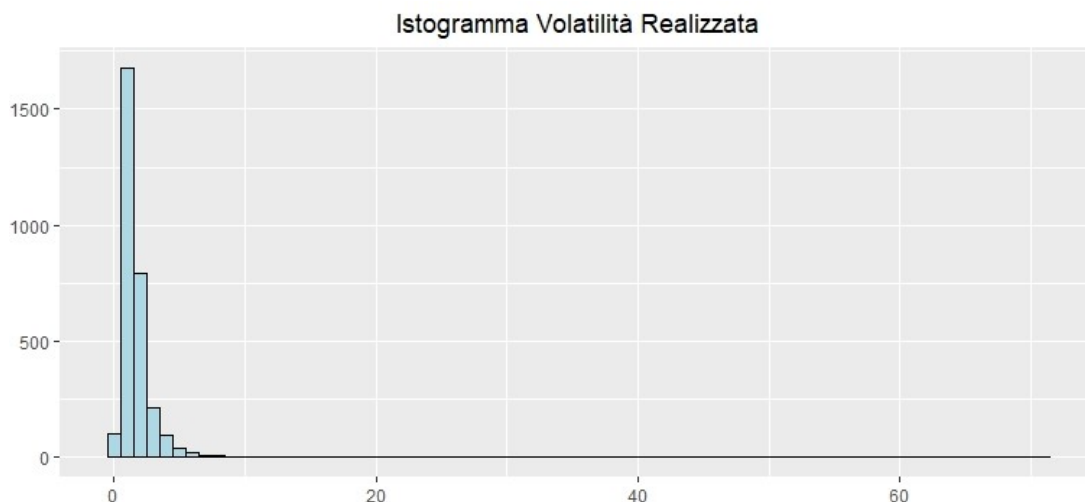


FIGURA 3.2: Istogramma delle Volatilità Realizzate del titolo CSCO

min	max	media	q. 10%	q. 30%	q. 50%	q. 70%	q. 90%
0.237	71.198	1.844	0.646	0.954	1.321	1.758	2.997

TABELLA 3.1: Statistiche riassuntive della Volatilità Realizzata

Dai risultati ottenuti emergono le caratteristiche empiriche della distribuzione della Volatilità Realizzata calcolata con frequenze elevate. Essa risulta, infatti, asimmetrica e fortemente leptocurtica, in quanto assume valori positivi e molto concentrati intorno alla media, ma presenta anche dei picchi che comportano code lunghe alla destra della distribuzione.

Nella *Figura 3.3* sono rappresentate le funzioni di autocorrelazione globale e parziale: la serie risulta persistente poiché, in generale, l'andamento della volatilità in un periodo di tempo tende a influenzare l'andamento della volatilità nei periodi successivi, formando i cosiddetti *clusters* di volatilità. Si noti che la persistenza della Volatilità Realizzata si mantiene per poco più di un mese borsistico, composto da 22 giornate.

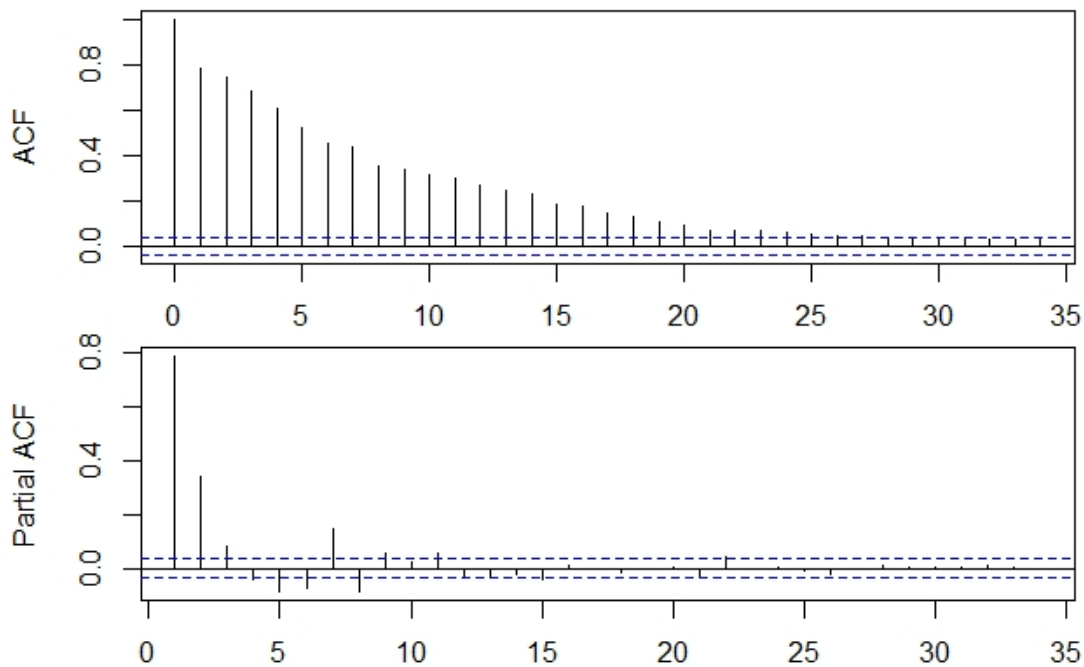


FIGURA 3.3: Funzioni di autocorrelazione globale e parziale della Volatilità Realizzata di CSCO

3.4.1 Modellazione e previsione della Volatilità Realizzata

Uno dei modelli più utilizzati per la stima e la previsione della Volatilità Realizzata è sicuramente l'*Heterogeneous Auto Regression* o HAR-RV, presentato da Corsi (2009) [16], che è specificato nella seguente maniera:

$$RV_t = \alpha + \beta_D RV_{t-1} + \beta_W RV_{t-1:t-5} + \beta_M RV_{t-1:t-22} + \epsilon_t,$$

dove $RV_{t-1:t-k} = \frac{1}{k} \sum_{j=1}^k RV_{t-j}$.

Tale definizione del modello si basa su elementi di natura economica, ovvero la presenza di *traders* che operano nel mercato con orizzonti temporali differenti, dal giornaliero al settimanale, fino al mensile. Questo approccio approssima la presenza di memoria lunga, in quanto corrisponde ad un modello AR(22) con restrizioni sui parametri. I parametri da valutare, infatti, sono solo tre e vengono stimati tramite OLS utilizzando *standard errors* robusti a causa della presenza di eteroschedasticità nei residui.

Le previsioni ottenute a partire da questo modello, utilizzando l'approccio *rolling* precedentemente descritto, fungono da *benchmark* per il confronto con le previsioni gerarchiche.

3.5 Previsioni gerarchiche

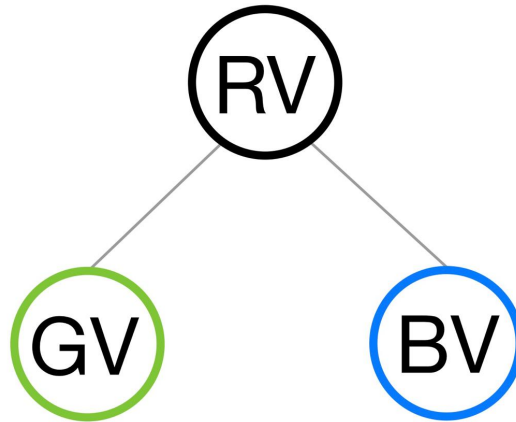
Le previsioni gerarchiche sono state ricavate a partire da quattro differenti strutture di aggregazione per la Volatilità Realizzata. Per ogni struttura è stato scelto l'approccio che producesse un MSE minore nell'*out-of-sample* tra il *bottom-up* e il MinT con le cinque specificazioni della matrice W_h definite nella sezione 2.5.3.

Per gli approcci MinT, l'utilizzo di una finestra *rolling* comporta il ricalcolo della matrice \mathbf{G} ottimale ad ogni ciclo.

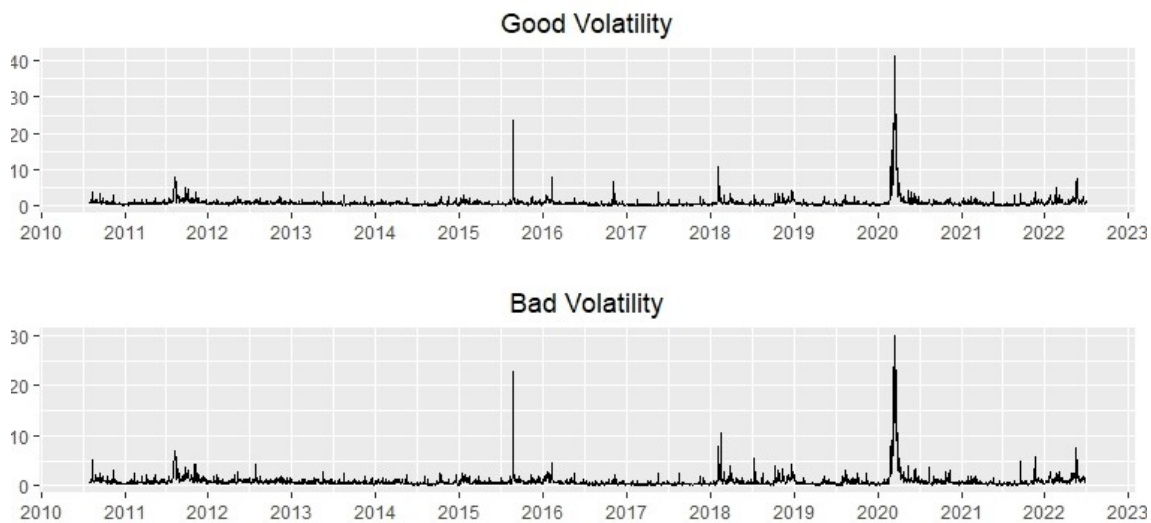
Tutte le previsioni di base delle componenti di ogni struttura sono state realizzate a partire dal modello di Corsi, applicato sempre con approccio *rolling*. Per tutte le procedure di riconciliazione si è utilizzato il pacchetto R FoReco (Di Fonzo e Girolimetto 2022 [22]).

3.5.1 Gerarchia *Good-Bad Volatility*

La prima struttura gerarchica che viene proposta è quella basata sulla decomposizione della Volatilità Realizzata in componenti *Good* e *Bad*, presentata da Patton e Sheppard nel 2015 [34]. Tale gerarchia risulta la più semplice tra quelle trattate in quanto è composta da un solo livello, ovvero quello inferiore in cui compaiono le serie delle componenti *Good* e *Bad*, calcolate come nella (1.17), che si aggregano per formare la Volatilità Realizzata (*Figura 3.4*). Tale gerarchia è stata denominata “**Gerarchia GBV**”.

FIGURA 3.4: Gerarchia *Good-Bad Volatility*

Le due componenti, ricavate a partire dal segno dei rendimenti, hanno caratteristiche simili a quelle della Volatilità Realizzata. Infatti, seguono lo stesso andamento, hanno distribuzione asimmetrica con code a destra e hanno le funzioni di autocorrelazione e correlazione incrociata persistenti ad indicare non solo una dipendenza seriale, ma anche una dipendenza tra le due serie nel tempo.

FIGURA 3.5: Andamento delle componenti *Good* e *Bad Volatility* del titolo CSCO

	min	max	media	q. 10%	q. 30%	q. 50%	q. 70%	q. 90%
GV	0.122	41.280	0.936	0.322	0.479	0.651	0.888	1.516
BV	0.110	29.918	0.639	0.310	0.465	0.639	0.886	1.484

TABELLA 3.2: Statistiche riassuntive delle componenti *Good* e *Bad Volatility*

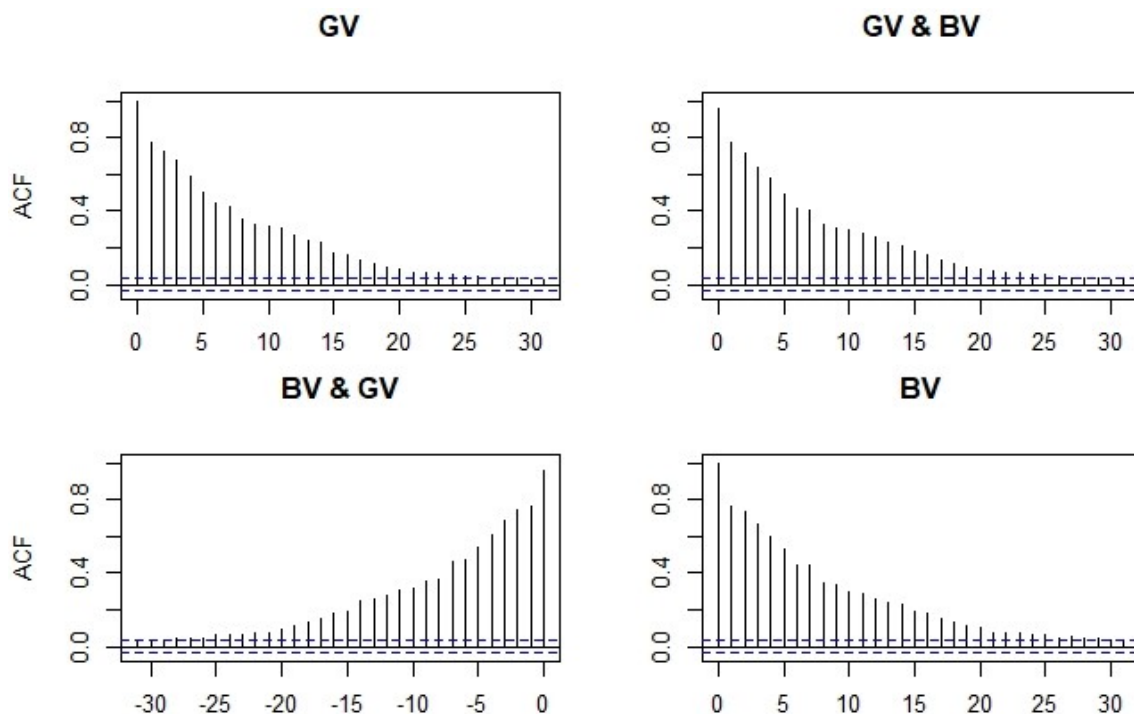


FIGURA 3.6: Funzioni di autocorrelazione e correlazione incrociata delle componenti *Good Volatility* e *Bad Volatility*

3.5.2 Gerarchie temporali e quantiliche

Le restanti gerarchie sono state strutturate a partire da una decomposizione della Volatilità Realizzata sia a livello temporale che sfruttando la partizione quantilica dei rendimenti (sezione 1.3.2).

3.5.2.1 Decomposizione temporale della Volatilità Realizzata

La decomposizione temporale mette in luce il diverso apporto alla Volatilità Realizzata dei rendimenti in fasi differenti della giornata borsistica.

Difatti, all'interno della giornata borsistica si hanno periodi di volatilità maggiore, soprattutto all'inizio, a causa delle informazioni *overnight*, e periodi di volatilità minore. È stato deciso di suddividere la giornata borsistica composta da 390 minuti in 5 parti da 78 minuti ciascuna.

Si sono così ottenute le seguenti componenti della Volatilità Realizzata:

$$\begin{aligned}
 RV_t^A &= \sum_{i=1}^{78} r_{i,t}^2 \text{ calcolata con i rendimenti osservati dalle ore 9:30 alle 10:48;} \\
 RV_t^B &= \sum_{i=79}^{156} r_{i,t}^2 \text{ calcolata con i rendimenti osservati dalle ore 10:48 alle 12:06;} \\
 RV_t^C &= \sum_{i=157}^{234} r_{i,t}^2 \text{ calcolata con i rendimenti osservati dalle ore 12:06 alle 13:24;} \\
 RV_t^D &= \sum_{i=235}^{312} r_{i,t}^2 \text{ calcolata con i rendimenti osservati dalle ore 13:24 alle 14:42;} \\
 RV_t^E &= \sum_{i=313}^{390} r_{i,t}^2 \text{ calcolata con i rendimenti osservati dalle ore 14:42 alle 16:00.}
 \end{aligned}$$

Quindi, la Volatilità Realizzata del giorno *t-esimo* può essere scomposta in:

$$RV_t = RV_t^A + RV_t^B + RV_t^C + RV_t^D + RV_t^E.$$

Nella *Figura 3.7* sono rappresentate le serie delle componenti temporali, mentre nella *Tabella 3.3* le statistiche riassuntive per ogni componente. Da tali analisi descrittive si nota che la componente RV^A presenta valori dei quantili molto più alti rispetto alle altre componenti e quindi apporta un contributo maggiore alla Volatilità Realizzata del titolo.

Un esempio lampante è il picco della RV del 24 agosto 2015 che è causato quasi esclusivamente da fluttuazioni della prima componente, come si nota nei grafici delle serie storiche in *Figura 3.7*. Dalla *Tabella 3.3* si evince anche che le componenti RV^B e RV^E hanno quantili leggermente più alti di quelle RV^C e RV^D ; inoltre le componenti RV^D e RV^E hanno come valore minimo della serie 0 perchè in alcune giornate borsistiche il mercato americano chiude anticipatamente (festività).

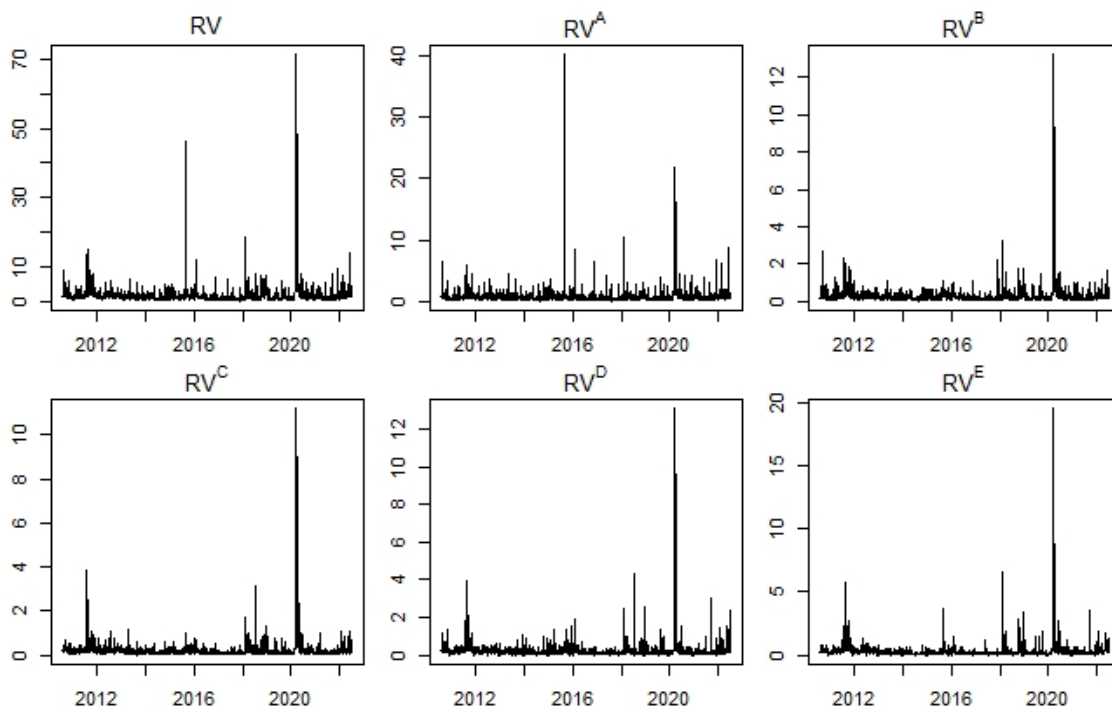


FIGURA 3.7: Grafici della serie storica della Volatilità Realizzata e delle sue componenti temporali

	min	max	media	q. 10%	q. 30%	q. 50%	q. 70%	q. 90%
RV^A	0.106	40.073	0.799	0.269	0.417	0.571	0.771	1.320
RV^B	0.033	13.184	0.294	0.099	0.150	0.207	0.287	0.497
RV^C	0.015	11.173	0.220	0.069	0.105	0.146	0.209	0.373
RV^D	0.000	13.084	0.222	0.059	0.097	0.139	0.203	0.377
RV^E	0.000	19.529	0.309	0.079	0.124	0.181	0.269	0.509

TABELLA 3.3: Statistiche riassuntive delle componenti temporali della Volatilità Realizzata

3.5.2.2 Decomposizione in partizioni quantiliche della Volatilità Realizzata

Un altro modo per scomporre la Volatilità Realizzata è quello di applicare la partizione quantilica dei rendimenti come illustrato nella sezione 1.3.2. In questa applicazione sono stati utilizzati i quantili 10%, 30%, 70% e 90% per la formazione delle diverse componenti nella seguente maniera:

$$\begin{aligned}
PRV_{t,1} &= \sum_{i=1}^n r_{i,t}^2 I \left(\frac{\tilde{r}_{i,t}}{\sqrt{\frac{1}{n}BPV_t}} < \Phi(0.10)^{-1} \right), \\
PRV_{t,2} &= \sum_{i=1}^n r_{i,t}^2 I \left(\Phi(0.10)^{-1} \leq \frac{\tilde{r}_{i,t}}{\sqrt{\frac{1}{n}BPV_t}} < \Phi(0.30)^{-1} \right), \\
PRV_{t,3} &= \sum_{i=1}^n r_{i,t}^2 I \left(\Phi(0.30)^{-1} \leq \frac{\tilde{r}_{i,t}}{\sqrt{\frac{1}{n}BPV_t}} < \Phi(0.70)^{-1} \right), \\
PRV_{t,4} &= \sum_{i=1}^n r_{i,t}^2 I \left(\Phi(0.70)^{-1} \leq \frac{\tilde{r}_{i,t}}{\sqrt{\frac{1}{n}BPV_t}} < \Phi(0.90)^{-1} \right), \\
PRV_{t,5} &= \sum_{i=1}^n r_{i,t}^2 I \left(\frac{\tilde{r}_{i,t}}{\sqrt{\frac{1}{n}BPV_t}} \geq \Phi(0.90)^{-1} \right).
\end{aligned}$$

Grazie a questa partizione, la Volatilità Realizzata del giorno t viene suddivisa nel contributo di cinque componenti:

$$RV_t = PRV_{t,1} + PRV_{t,2} + PRV_{t,3} + PRV_{t,4} + PRV_{t,5}.$$

La scelta di questi quantili specifici permette, così, di ottenere due gruppi calcolati a partire da rendimenti filtrati di grandi dimensioni negativi e positivi ($PRV_{t,1}$, $PRV_{t,5}$), due gruppi da rendimenti filtrati di medie dimensioni negativi e positivi ($PRV_{t,2}$, $PRV_{t,4}$) e un gruppo da rendimenti di piccole dimensioni sia positivi che negativi ($PRV_{t,3}$).

L'analisi grafica in *Figura 3.8* confronta gli andamenti delle partizioni quantiliche e della Volatilità Realizzata nell'orizzonte di tempo considerato. Come ci si poteva aspettare, le serie delle componenti quantiliche presentano i picchi nelle stesse posizioni della Volatilità Realizzata. Un'eccezione può essere rappresentata sempre dal picco della RV corrispondente al 24 agosto 2015, evidente anche nella PRV_A e nella PRV_E , ma meno nelle altre componenti.

Osservando la *Tabella 3.4*, si nota che le componenti PRV_1 e PRV_5 hanno valori elevati dei diversi quantili della distribuzione, le PRV_2 e PRV_4 , invece presentano una distribuzione quantilica spostata su valori leggermente più bassi, mentre la componente PRV_3 presenta valori inferiori in tutte le statistiche riassuntive, a dimostrazione del fatto che apporta un minor contributo alla Volatilità Realizzata.

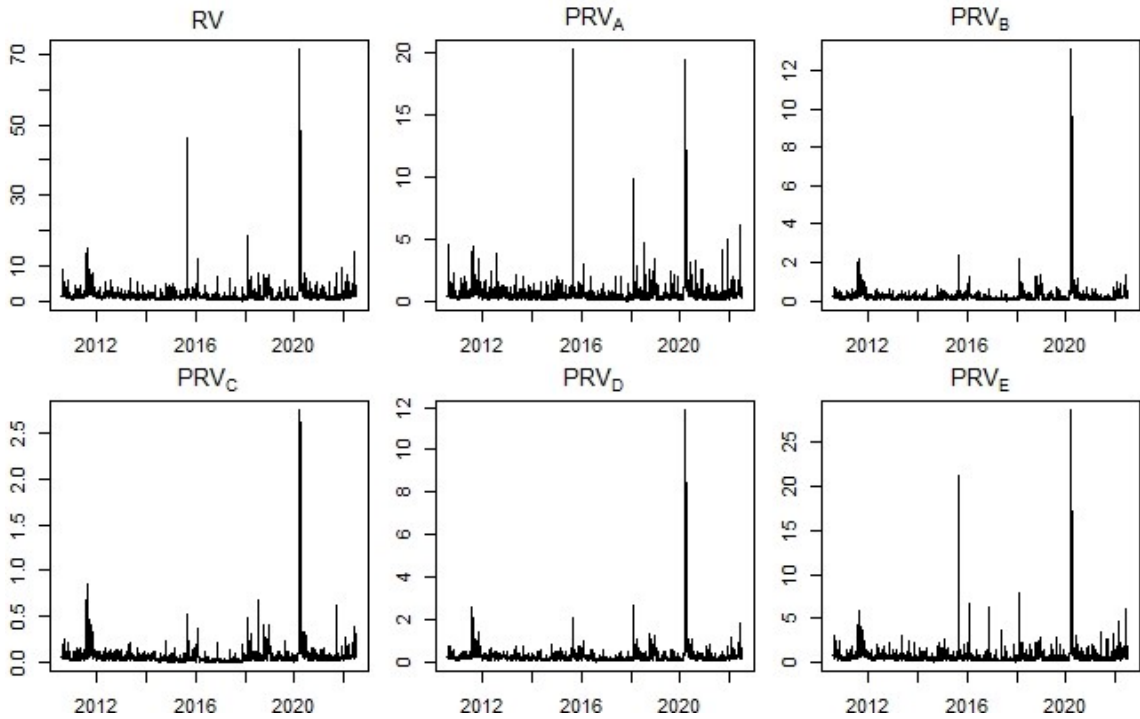


FIGURA 3.8: Grafici della serie storica della Volatilità Realizzata e delle sue componenti quantiliche

	min	max	media	q. 10%	q. 30%	q. 50%	q. 70%	q. 90%
PRV_1	0.096	20.257	0.638	0.225	0.330	0.453	0.635	1.043
PRV_2	0.008	13.077	0.243	0.070	0.117	0.165	0.228	0.414
PRV_3	0.000	2.752	0.054	0.003	0.014	0.031	0.054	0.100
PRV_4	0.004	11.841	0.240	0.072	0.120	0.167	0.226	0.409
PRV_5	0.100	28.645	0.669	0.230	0.341	0.465	0.636	1.112

TABELLA 3.4: Statistiche riassuntive delle componenti quantiliche della Volatilità Realizzata

3.5.3 Decomposizione temporale e quantilica

Le due decomposizioni appena trattate, quella temporale e quella quantilica, possono essere combinate in una scomposizione della Volatilità Realizzata che, quindi, tiene conto di entrambi gli aspetti. Si ottengono, così, venticinque componenti che si sommano nella seguente maniera per ricavare la Volatilità Realizzata:

$$RV_t = \sum_{j=1}^5 \sum_{i=A}^E PRV_{t,j}^i, \quad (3.1)$$

dove la componente $PRV_{t,j}^i$ è stata calcolata con rendimenti osservati nel periodo i -esimo della giornata t -esima che fanno parte del gruppo j -esimo della partizione quantilica.

Le componenti ricavate da questa decomposizione più dettagliata vanno a costituire il livello inferiore delle restanti tre strutture gerarchiche della Volatilità utilizzate per ottenere le previsioni riconciliate.

3.5.3.1 Strutture temporali e quantiliche

La seconda struttura che viene utilizzata per ottenere le previsioni gerarchiche della Volatilità Realizzata è denominata “**Gerarchia temporale RV**” o, per abbreviare, “**Gerarchia RV**”. Essa è costituita da un livello inferiore composto dalle venticinque serie temporali e quantiliche trattate in precedenza (PRV_j^i), che poi si aggregano secondo vincoli temporali per ottenere un livello intermedio caratterizzato dalle componenti temporali della Volatilità Realizzata (RV^i). Il livello più alto è sempre occupato dalla Volatilità Realizzata complessiva.

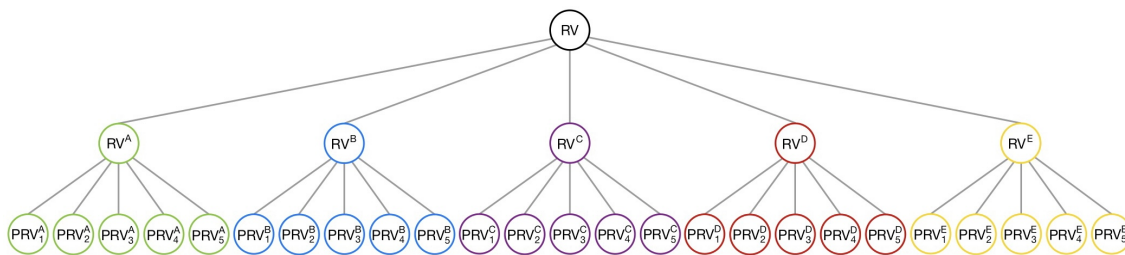


FIGURA 3.9: Gerarchia temporale RV

La terza struttura presentata è simile a quella precedente, ma varia nello strato intermedio. Difatti, le serie di livello inferiore si aggregano secondo la partizione quantilica andando a formare i cinque gruppi PRV_j . Questa organizzazione gerarchica viene chiamata “**Gerarchia quantilica PRV**” oppure “**Gerarchia PRV**”.

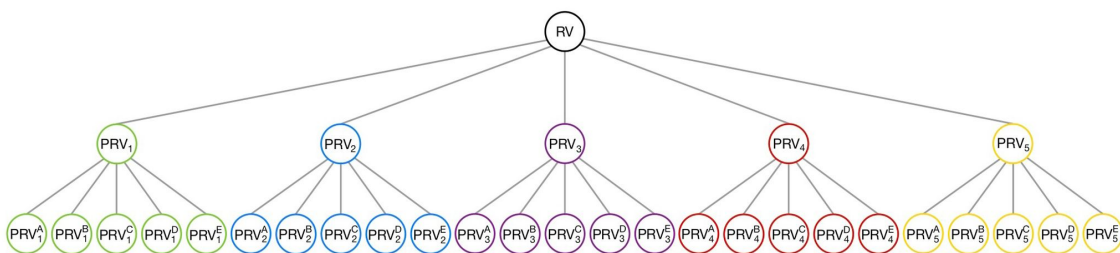


FIGURA 3.10: Gerarchia quantilica PRV

L'ultima struttura è la più complessa delle quattro utilizzate e rappresenta un esempio di serie storiche raggruppate viste nella sezione 2.1. Infatti, tale raggruppamento viene ottenuto considerando allo stesso tempo le due gerarchie viste in precedenza. Il livello inferiore è sempre formato dalle venticinque serie che, in questo caso, si aggregano nei due modi differenti, sia secondo i vincoli temporali, sia rispettando la partizione quantilica. Quindi, lo strato intermedio è formato dalle componenti RV^i e dalle componenti PRV_j . Tale struttura viene nominata “**Raggruppamento temporale-quantilico RVPRV**” o semplicemente “**Gerarchia RVPRV**”.

Capitolo 4

Conclusioni

4.1 Risultati ottenuti

Il primo esercizio consiste nella previsione ad un passo in avanti della Volatilità Realizzata con finestra *rolling* di 1000 osservazioni e orizzonte predittivo che va dal 23/07/2014 al 30/06/2022. L'andamento della Volatilità Realizzata del titolo *Cisco Systems* nella finestra di valutazione considerata è descritto dal seguente grafico.

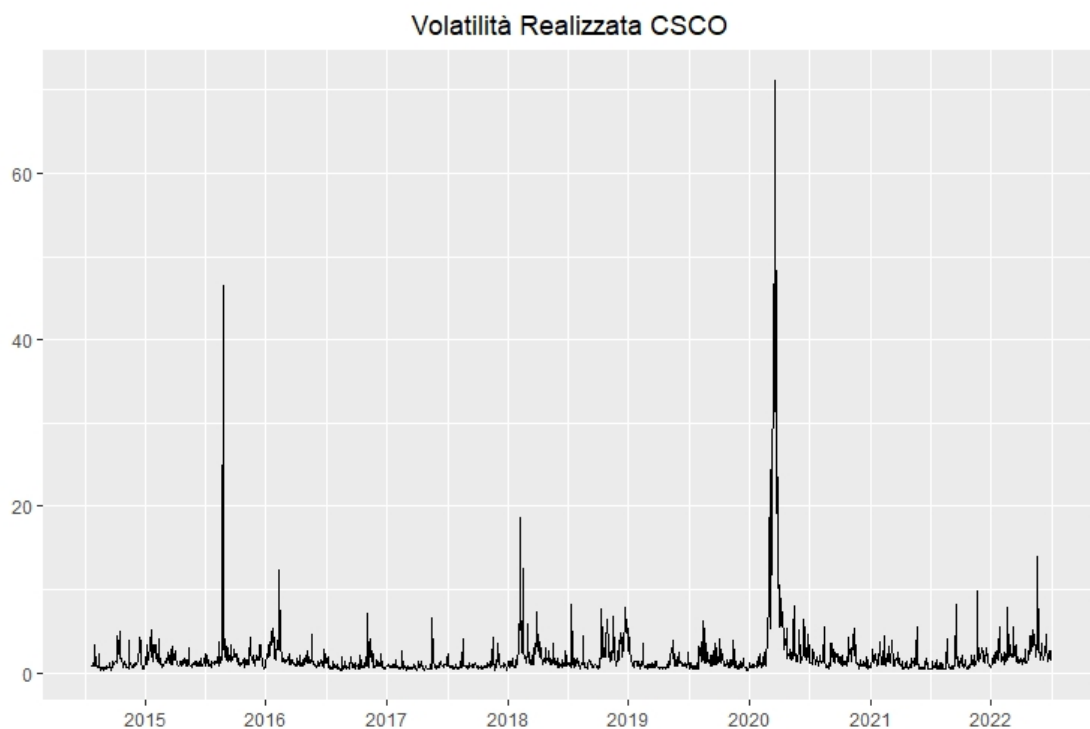


FIGURA 4.1: Andamento della Volatilità Realizzata del titolo CSCO

Inizialmente si è dovuto stabilire l'approccio gerarchico da utilizzare per ogni struttura, scegliendo tra il *bottom-up* (BU) e quello MinT con le cinque diverse specificazioni della matrice W_h elencate nella sezione 2.4.3.1. In particolare, è stato scelto il metodo che minimizzasse il MSE nell'orizzonte predittivo. Nella *Tabella 4.1*, in cui vengono riassunti i risultati ottenuti, i numeri riportati di fianco agli approcci che compongono la prima riga fanno riferimento alle matrici W_h della sezione 2.4.3.1.

	BU	OLS(1)	WLS _s (3)	WLS _v (2)	SHR(5)	SAM(4)
Ger. GBV	10586.80	10543.23	10551.93	10558.61	10557.47	10524.44
Ger. RV	10672.74	10380.12	10254.85	10333.89	10045.53	9296.88
Ger. PRV	10672.74	10511.30	10452.67	10521.17	10513.32	10779.48
Ger. RVPRV	10672.74	10412.30	10347.19	10390.28	10140.20	11598.45

TABELLA 4.1: MSE dei diversi approcci applicati alle gerarchie su finestra *out-of-sample* di 2000 osservazioni

Si nota che l'approccio SAM, il quale utilizza la matrice di varianza-covarianza campionaria degli errori di previsione un passo in avanti, risulta il migliore sia per la **Gerarchia GBV** sia per la **Gerarchia RV**, mentre per le altre due gerarchie ottiene MSE abbastanza alti. Utilizzando lo *shrinkage* per la matrice W_h , invece, si conseguono dei buoni risultati per tutte le gerarchie e, nello specifico, il migliore per la **Gerarchia RVPRV**, ovvero quella che ha una struttura più complessa.

Al contrario, il modello *bottom-up* è quello che, in generale, performa in maniera peggiore, mentre tra gli approcci che utilizzano una matrice W_h diagonale, il WLS_s è quello che porta a risultati più vantaggiosi, soprattutto per la **Gerarchia PRV**.

Nella *Tabella 4.2* sono riportati il MSE e il MAE del modello di **Corsi** e dei modelli di previsione gerarchica considerati.

	MSE	MAE
Corsi	10530.24	1544.76
Ger. GBV	10524.44	1548.16
Ger. RV	9296.88	1510.00
Ger. PRV	10452.67	1546.60
Ger. RVPRV	10140.20	1531.14

TABELLA 4.2: MSE e MAE dei cinque modelli su finestra *out-of-sample* di 2000 osservazioni

Si osserva che il modello basato sulla **Gerarchia RV** con approccio MinT(SAM) presenta valori migliori sia in termini di MSE che di MAE. La gerarchia temporale, quindi, trae maggiore vantaggio dalle relazioni che sussistono tra le serie dei dati.

Allo stesso modo, con la **Gerarchia RVPRV** si ottengono buoni risultati in confronto a quelli del modello di **Corsi**, mentre quella **PRV** è preferibile al modello di **Corsi** in termini di MSE, ma non di MAE.

D'altra parte, l'applicazione di una struttura gerarchica semplice come la **Gerarchia GBV** non apporta sostanziali miglioramenti in questa situazione.

Gli andamenti del MSE dei modelli adoperati, mostrati nella *Figura 4.2*, sono esplicativi riguardo a quanto presentato in precedenza. Infatti, per il modello basato sulla **Gerarchia RV** si evidenzia un MSE finale minore rispetto agli altri modelli. Questo è dovuto al fatto che il primo performa meglio degli altri nei giorni in cui la Volatilità Realizzata presenta un picco, quindi quelli in cui il MSE aumenta vertiginosamente.

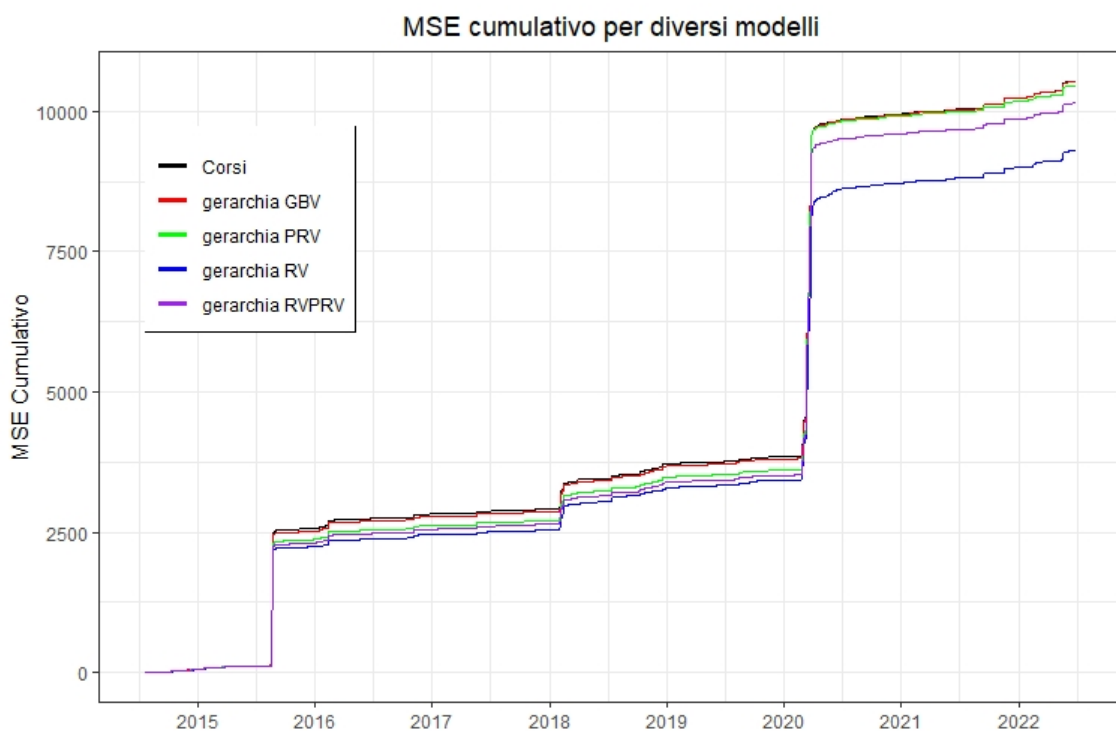


FIGURA 4.2: Andamento del MSE dei cinque modelli su finestra *out-of-sample* di 2000 osservazioni

Per confrontare la bontà predittiva dei modelli è stato applicato anche il *Model Confidence Set*, secondo la procedura descritta nella sezione 3.3. Il test ha escluso dall'insieme di modelli quello di **Corsi** e quello basato sulla **Gerarchia GBV**, a conferma del fatto

che l'applicazione di previsioni gerarchiche con strutture quantiliche e temporali apporta sostanziali miglioramenti nella previsione della Volatilità Realizzata in termini di perdita quadratica.

Le capacità predittive dei vari metodi sono state confrontate anche su orizzonti temporali minori. La finestra *out-of-sample* lunga 2000 osservazioni è stata suddivisa in quattro parti, per ognuna delle quali sono stati calcolati il MSE e il MAE dei cinque modelli confrontati. Ogni sezione valutativa è caratterizzata da oscillazioni della Volatilità più o meno marcate. I quattro periodi considerati sono:

- dal 23/07/2014 al 31/12/2015, con il primo picco di Volatilità;
- dal 04/01/2016 al 29/12/2017, che presenta meno oscillazioni della Volatilità;
- dal 02/01/2018 al 31/12/2019, con più oscillazioni rispetto al precedente;
- dal 02/01/2020 al 30/06/2022, caratterizzato da molte oscillazioni dovute alla pandemia.

I risultati ottenuti sono riportati nella *Tabella 4.3*, in cui sono inseriti intervalli di tempo convenzionali che fanno riferimento ai periodi sopracitati.

	2014-2016		2016-2018	
	MSE	MAE	MSE	MAE
Corsi	2570.20	247.54	347.00	248.03
Ger. GBV	2524.09	244.86	348.08	251.40
Ger. RV	2242.79	232.08	301.47	217.36
Ger. PRV	2370.40	239.74	340.97	247.71
Ger. RVPRV	2313.41	234.70	334.41	243.91

	2018-2020		2020-2022	
	MSE	MAE	MSE	MAE
Corsi	926.66	353.63	6686.38	695.56
Ger. GBV	935.04	354.54	6717.24	697.35
Ger. RV	884.21	348.37	5868.42	712.19
Ger. PRV	897.27	353.12	6844.03	706.03
Ger. RVPRV	867.12	350.02	6625.26	702.52

TABELLA 4.3: MSE e MAE nei quattro periodi considerati

Nonostante i quattro periodi non siano della stessa durata, si nota facilmente che per tutti i modelli il MSE è molto alto nel primo e nel quarto intervallo di tempo per via dei picchi inaspettati che li caratterizzano.

Anche utilizzando più finestre per la valutazione, il modello con la **Gerarchia RV** risulta preferibile in tutti i periodi, eccetto che nel terzo in termini di MSE e nell'ultimo rispetto al MAE.

Nel complesso, a partire dalla **Gerarchia PRV** e dalla **Gerarchia RVPRV** si ottengono risultati migliori rispetto al modello di **Corsi** e a quello basato sulla **Gerarchia GBV**, tranne che nell'ultima fase, in cui il modello costruito con la **Gerarchia PRV** risulta il peggiore.

Il secondo esercizio predittivo è quello che utilizza una finestra *rolling* di 250 osservazioni e un periodo *out-of sample* che va dal 28/07/2011 al 30/06/2022. Per questa applicazione vengono presentate solo la tabella per la scelta dell'approccio gerarchico da applicare ad ogni modello (*Tabella 4.4*) e quella che riporta le due misure di accuratezza dei modelli di previsione considerati (*Tabella 4.5*).

	BU	OLS(1)	WLS _s (3)	WLS _v (2)	SHR(5)	SAM(4)
Ger. GBV	13887.13	14106.58	14046.26	13989.73	13992.36	14363.47
Ger. RV	12792.54	13706.86	12705.66	12595.12	12423.93	15214.93
Ger. PRV	12792.54	13947.53	13042.71	12978.41	13006.46	18933.64
Ger. RVPRV	12792.54	13659.83	12891.28	12800.98	12684.68	27692.12

TABELLA 4.4: MSE dei diversi approcci applicati alle gerarchie su finestra *out-of-sample* di 2750 osservazioni

Differentemente da quanto osservato nel primo esercizio predittivo, applicando l'approccio MinT(SAM) alle diverse gerarchie si ottengono MSE molto alti, specialmente quando la struttura è più complessa, come nel caso della **Gerarchia RVPRV**. Al contrario, lo *shrinkage* produce risultati migliori, infatti è consigliabile sia per la **Gerarchia RV** che per la **Gerarchia RVPRV**. Un altro aspetto da sottolineare è che per la **Gerarchia GBV** e per la **Gerarchia PRV** è preferibile utilizzare l'approccio *bottom-up*, il quale produce risultati migliori con finestre *rolling* più piccole.

Dalla *Tabella 4.5* risulta che il modello associato alla **Gerarchia RV** è nuovamente quello che prevede la Volatilità Realizzata più accuratamente, sia in termini di MSE che di MAE. Inoltre, anche gli altri modelli basati sulle due gerarchie create a partire

dalle decomposizioni temporali e quantiliche della Volatilità, ottengono risultati migliori rispetto al modello di **Corsi** e al modello gerarchico *Good-Bad Volatility*.

	MSE	MAE
Corsi	14238.15	2011.86
Ger. GBV	13887.13	2011.23
Ger. RV	12423.93	2005.98
Ger. PRV	12792.54	2066.57
Ger. RVPRV	12684.68	2003.90

TABELLA 4.5: MSE e MAE dei cinque modelli su finestra *out-of-sample* di 2750 osservazioni

4.2 Considerazioni finali

Le prime considerazioni che si possono fare a partire dai risultati dei due esercizi riguardano la scelta dell'approccio da applicare per i modelli gerarchici. L'utilizzo del metodo MinT con matrice SAM è preferibile in presenza di campioni di grandi dimensioni perché può trovare difficoltà nella stima della matrice di varianza-covarianza; al contrario, il *bottom-up* funziona meglio con finestre di stima meno ampie. In generale, il metodo MinT che applica lo *shrinkage* rimane l'approccio preferibile, sia perché produce dei buoni risultati sia perché evita problemi nella stima della matrice di varianza-covarianza.

Passando al confronto sulla bontà di previsione dei modelli considerati, si nota, innanzitutto, che il modello gerarchico basato sulla **Gerarchia RV** produce le previsioni più accurate della Volatilità Realizzata in termini di MSE e MAE. Anche i modelli con **Gerarchia PRV** e **Gerarchia RVPRV** sono preferibili al modello di **Corsi**, in particolar modo secondo la misura del MSE. Ciò suggerisce che, in questo caso, le informazioni ricavabili dalle strutture temporali e quantiliche della Volatilità portino ad una migliore previsione di tale quantità d'interesse. La previsione delle diverse componenti della Volatilità Realizzata, infatti, permette di rilevare eventuali cambiamenti strutturali della stessa, soprattutto in periodi con maggiori fluttuazioni. D'altra parte, il modello costruito a partire dalla **Gerarchia GBV** non produce esiti particolarmente diversi da quelli del modello di **Corsi**, probabilmente a causa della minima specificazione della Volatilità Realizzata.

In conclusione, l'applicazione di modelli gerarchici per la previsione della Volatilità Realizzata permette di migliorare i risultati ottenuti dal modello HAR-RV, soprattutto quando si inseriscono dei vincoli temporali e quantilici nella struttura della Volatilità. L'utilizzo di questi modelli, però, comporta un maggiore onere computazionale, in quanto necessita di previsioni di più serie temporali. Per questo motivo, non sempre si preferisce applicare i modelli gerarchici, soprattutto quando il miglioramento nei risultati non è rilevante.

4.3 Possibili miglioramenti

Per capire se effettivamente l'approccio gerarchico sia preferibile al modello di Corsi, l'esercizio predittivo dovrebbe essere ripetuto per più titoli differenti, in modo tale da compiere delle valutazioni trasversali. Inoltre, si potrebbe verificare la robustezza dei risultati ottenuti utilizzando un maggior numero di finestre *rolling* e *out-of-sample* di ampiezze diverse.

Altri aspetti che potrebbe essere interessante esaminare sono la distribuzione delle previsioni gerarchiche ottenute e gli intervalli di previsione. Difatti, negli esercizi svolti sono state adoperate solo previsioni puntuali.

Gli intervalli di previsione possono essere ricavati facilmente se si ipotizza che le previsioni di base siano normalmente distribuite, ma ciò risulta poco ragionevole; quindi, è necessario calcolarli tramite tecniche *bootstrap*.

Ulteriori sviluppi possono essere raggiunti adoperando strutture gerarchiche della Volatilità Realizzata diverse dalle quattro proposte o applicando differenti metodi per ottenere le previsioni di base, come, ad esempio, il modello autoregressivo vettoriale.

Bibliografia

- [1] Yacine Aït-Sahalia and Jean Jacod. Testing for jumps in a discretely observed process. *The Annals of Statistics*, pages 184–222, 2009.
- [2] Torben G Andersen, Tim Bollerslev, and Francis X Diebold. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics*, 89(4):701–720, 2007.
- [3] Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Paul Labys. The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453):42–55, 2001.
- [4] Torben G Andersen, Tim Bollerslev, and Dobrislav Dobrev. No-arbitrage semimartingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics*, 138(1):125–180, 2007.
- [5] Torben G Andersen, Tim Bollerslev, Per Frederiksen, and Morten Ørregaard Nielsen. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics*, 25(2):233–261, 2010.
- [6] Torben G Andersen and Timo Teräsvirta. Realized volatility. In *Handbook of financial time series*, pages 555–575. Springer, 2009.
- [7] George Athanasopoulos, Roman A Ahmed, and Rob J Hyndman. Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, 25(1):146–166, 2009.
- [8] George Athanasopoulos, Rob J Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos. Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1):60–74, 2017.

-
- [9] Ole E Barndorff-Nielsen and Neil Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002.
- [10] Ole E Barndorff-Nielsen and Neil Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics*, 2(1):1–37, 2004.
- [11] Ole E Barndorff-Nielsen and Neil Shephard. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics*, 4(1):1–30, 2006.
- [12] Mauro Bernardi and Leopoldo Catania. The model confidence set package for r. *International Journal of Computational Economics and Econometrics*, 8(2):144–158, 2018.
- [13] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.
- [14] Kris Boudt, Christophe Croux, and Sébastien Laurent. Robust estimation of intra-week periodicity in volatility and jump detection. *Journal of Empirical Finance*, 18(2):353–367, 2011.
- [15] Massimiliano Caporin. Do jumps matter in realized volatility modeling and forecasting? empirical evidence and a new model. *Empirical Evidence and a New Model (July 14, 2020)*, 2020.
- [16] Fulvio Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009.
- [17] Fulvio Corsi, Davide Pirino, and Roberto Reno. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276–288, 2010.
- [18] John B Edwards and Guy H Orcutt. Should aggregation prior to estimation be the rule? *The Review of Economics and Statistics*, pages 409–420, 1969.
- [19] Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, pages 987–1007, 1982.

-
- [20] Eugene B Fliedner and Barry Lawrence. Forecasting system parent group formation: An empirical application of cluster analysis. *Journal of Operations Management*, 12(2):119–130, 1995.
- [21] Donald W Fogarty, John H Blackstone Jr, and TR Hoffmann. Production and inventory management. south, 1991.
- [22] Daniele Girolimetto and Tommaso Di Fonzo. Foreco: Point forecast reconciliation. url: [https://cran.r-project.org/package= FoReco](https://cran.r-project.org/package=FoReco), 2022.
- [23] Charles W Gross and Jeffrey E Sohl. Disaggregation methods to expedite product line forecasting. *Journal of forecasting*, 9(3):233–254, 1990.
- [24] Yehuda Grunfeld and Zvi Griliches. Is aggregation necessarily bad? *The review of economics and statistics*, pages 1–13, 1960.
- [25] Peter R Hansen, Asger Lunde, and James M Nason. The model confidence set. *Econometrica*, 79(2):453–497, 2011.
- [26] Peter Reinhard Hansen, Asger Lunde, and James M Nason. Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics*, 65:839–861, 2003.
- [27] Xin Huang and George Tauchen. The relative contribution of jumps to total price variance. *Journal of financial econometrics*, 3(4):456–499, 2005.
- [28] Rob J Hyndman, Roman A Ahmed, George Athanasopoulos, and Han Lin Shang. Optimal combination forecasts for hierarchical time series. *Computational statistics & data analysis*, 55(9):2579–2589, 2011.
- [29] Aleksey Kolokolov and Roberto Renò. Jumps or staleness? Available at SSRN 3882775, 2021.
- [30] Suzanne S Lee and Per A Mykland. Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies*, 21(6):2535–2563, 2008.
- [31] Suzanne S Lee and Minho Wang. The impact of jumps on carry trade returns. *Journal of Financial Economics*, 131(2):433–455, 2019.
- [32] Seetharama L Narasimhan, Dennis W McLeavey, and Peter Billington. *Production planning and inventory control*. Pearson College Division, 1995.

-
- [33] Guy H Orcutt, Harold W Watts, and John B Edwards. Data aggregation and information loss. *The American Economic Review*, 58(4):773–787, 1968.
- [34] Andrew J Patton and Kevin Sheppard. Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3):683–697, 2015.
- [35] Albert B Schwarzkopf, Richard J Tersine, and John S Morris. Top-down versus bottom-up forecasting strategies. *The International Journal Of Production Research*, 26(11):1833–1843, 1988.
- [36] Eli Shlifer and Ronald W Wolff. Aggregation and proration in forecasting. *Management Science*, 25(6):594–603, 1979.
- [37] Stephen J Taylor. *Modelling financial time series*. Wiley, 1986.
- [38] Stephen J Taylor and Xinzhong Xu. The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance*, 4(4):317–340, 1997.
- [39] Shanika L Wickramasuriya, George Athanasopoulos, and Rob J Hyndman. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*, 114(526):804–819, 2019.
- [40] Lan Zhang, Per A Mykland, and Yacine Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.

